Genomic and metabolomic studies for better understanding of osteoarthritis pathogenesis

by

©Salem Werdyani

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Division of Biomedical Sciences (Genetics), Faculty of Medicine, Memorial University, St. John's, Newfoundland and Labrador, Canada.

May 2024

GENERAL ABSTRACT:

Osteoarthritis is the most common form of arthritis and one of the ten most disabling diseases in developed countries. The main objective of my thesis was to employ genomic and metabolomic approaches to improve our understanding of the OA pathogenesis.

I carried out a metabolomics analysis and identified three distinct endotypes of OA patients. Butyrylcarnitine, arginine, and a number of glycerophosphlipids were the major contributing metabolites for the differentiation between the three endotypes suggesting that the primary OA patients can be classified as muscle weakness, arginine deficient, and low inflammatory OA. Using the same metabolomics approach, I found that the elevated blood level of the ADMA and uric acid were associated with the muscle weakness over 10-years and may elevate the study participants risk for developing OA.

Additionally, I conducted an independent GWAS analysis in OA patients from NL and identified novel genes significantly associated with OA. These genes are involved in cartilage deterioration; inflammatory signaling; innate immune pathway; abnormal bone growth and remodeling; panic disorder; and pain mechanisms.

Further, I performed a genetic variant annotation study using WES data and identified deleterious variants in the *IGSF3, ZNF717, PRSS1, AQP7,* and *ESRRA* genes in OA patients that have not been reported in previous OA GWAS studies. These genes act in the ECM homeostasis and degradation. With the same WES data, I conducted a genome-

wide digenic interaction test in OA patients and identified aggregated variants in each of the *CDH19*, *SOGA1*, *MORC4*, *TMTC4*, and *ANK3* genes to be significantly interacting with rs56158521 in the *HLA-DRB1* gene. Our findings suggested the implication of the immunoinflammatory pathway in the pathogenesis of OA.

Also, I conducted a GWAS analysis and found variants in the *MC5R* gene to be significantly increasing the TJR pain, and variants adjacent to the *TPTE* gene to be significantly increasing the TJR disability. These genes are involved in immunoinflammatory reactions and may play a significant role in the pain and function mechanisms following TJR.

While confirmation is required, these findings provided new insights into better understanding of the OA pathogenesis and hold promising as druggable targets for developing OA therapies.

ACKNOWLEDGEMENTS:

In the name of Allah, the Most Compassionate and the Most Merciful. All praises, gratitude and glory to Almighty Allah who gave me the strengths, courage, and blessing throughout my research work to complete this thesis successfully, and peace and blessing of Allah be upon our beloved Prophet Mohammad (S).

I would like to express my heartfelt gratitude and unrestrained appreciation to my supervisor, Dr. Guangju Zhai, whose encouragement, supervision, guidance, and unlimited support throughout my thesis project have led to the success of this research. I also gratefully acknowledge the financial support of the School of Graduate Studies, Research and Graduate Studies at the Faculty of Medicine during my PhD program. Moreover, special appreciation goes to the Canadian Institutes of Health Research (CIHR), the Research and Development Corporation of Newfoundland and Labrador; and Memorial University of Newfoundland, and the National Health and Medical Research Council (NHMRC, Australia). Furthermore, I deeply appreciate the financial support of the Canadian Arthritis Society that partially funded my PhD stipend with the training graduate PhD Salary award for 2019–2022 academic years. Moreover, I would like to thank the Metabolomics Innovation Centre (TMIC) for the metabolic profiling analysis of this thesis.

My sincere thanks to those people who in one way or another contributed and extended their support and valuable assistance in the preparation and completion of this academic

work, including my supervisory committee members (Dr. Proton Rahman, Dr. Zhiwei Gao, and Dr. Andrew Furey); I also thank the members of the Zhai lab, specifically Ms. Ming Liu and Ms. Christie Costello for their coordination. Moreover, I would like to thank all the study participants who made the current thesis possible; and I thank all staff in the operation room who assisted with the sample collection, all the research team in the Newfoundland Osteoarthritis Study (NFOAS), and the staff of the Tasmanian Older Adult Cohort (TASOAC).

I would also like to take this opportunity to express my thankfulness to Dr. Sevtap Savas, the coordinator of the Human Genetics Graduate Program for her unlimited support and guidance. Even after completing my master's degree at her lab, Dr. Savas's office door has been open to me all the time for guidance and helpful advice over the years of my graduate studies, both from a personal and scientific perspective. Also, special thanks to Dr. Ann Dorward, the Associated Dean of the Graduate Studies for her time, patience, help, and guidance over my PhD years.

My gratitude and thanks are extended to my Beloved wife, Dr. Rahaf Alkhateb who never failed to provide me with love, warmth, patience, and care during my thesis project. Not forgotten, a special thanks to my dearest sons, Adnan and Mohamed-Nour Werdyani, who have been my inspiration as I overcome all the obstacles in the completion of this thesis project. Also, I would like to take this opportunity to express the grateful appreciation to my late father Adnan Werdyani, and my mother Lamah Werdyani for their infinite sacrifice, encouragement, prayers, and blessing. I would also thank my

parents in law, Yehia Alkhateb and Nahed Haj Ali, as well as all members of my family for their unlimited support and blessings.

Words can’t describe my gratitude to Mrs. Gayatri Raghwa for her encouragement and blessing from thousands of kilometres away. Gayatri, always believed in me, even when I didn't believe in myself. She is the first person who has ever seen my potential and helped me to realize it. her faith in me is an extraordinary gift that I will never forget!

Last, but not the least, I would like to thank my best friend in St. John’s, Mr. Eyad Sakkar who has always been there for me, and I’ve always been able to count on him in the difficult times.

SALEM WERDYANI.

TABLE OF CONTENTS

GENERAL ABSTRACT: I

ACKNOWLEDGEMENTS: III

TABLE OF CONTENTS VI

LIST OF FIGURES X

LIST OF TABLES XIV

LIST OF APPENDICES XVII

LIST OF ABBREVIATIONS XVIII

RESEARCH OUTPUT AND AWARDS XXVII

- **ABSTRACTS: ..... XXVII**
- **PUBLICATIONS: ..... XXXI**
- **AWARDS: ..... XXXI**

Chapter 1: General introduction 1

1.1 WHAT IS OSTEOARTHRITIS (OA)? 2

1.1.1 OA definition: 2

1.1.2 OA pathogenesis: 5

1.1.3 Symptoms and signs of OA: 13

1.1.4 Diagnosis of OA: 16

1.2 OA EPIDEMIOLOGY: 23

1.2.1 Site specific prevalence and incidence of OA: 26

1.2.1.1 Knee OA: 26

1.2.1.2 Hip OA: 27

1.2.1.3 Hand OA: 28

1.2.1.4 Multiple Joint OA (MJOA): 29

1.2.2 Economic and social burdens of OA: 29

1.2.3 OA aetiologies and risk factors: 31

1.2.3.1 Age: 31

1.2.3.2 Sex: 32

1.2.3.3 Obesity: 33

1.2.3.4 Metabolic syndrome: 35

1.2.3.5 Osteoporosis: 37

1.2.3.6 Muscle weakness (Sarcopenia): 39

1.2.3.7 Joint injury: 40

1.2.3.8 Occupation: 40

1.2.3.9 Sports and physical activities: 42

1.2.3.10 Joint Alignment: 43

1.2.3.11 Ethnicity and race role in OA: 43

1.2.3.12 Diet: 44

1.2.3.13 Smoking: 46

1.2.3.14 Genetic architecture of OA: 46

1.2.3.14.1 Familial aggregation studies: 48

1.2.3.14.2 Twin studies: 50

1.2.3.14.3 Genome-wide linkage scan: 51
1.2.3.14.4 Candidate gene studies: 53
1.2.3.14.5 GWAS analysis of OA: 55
1.3 OA MANAGEMENT: 64
1.3.1 Non-pharmacologic interventions: 65
1.3.2 Pharmacologic treatment: 66
1.3.3 Surgical treatment: 74
Chapter 2: Thesis objectives 76
Chapter 3: Endotypes of primary osteoarthritis identified by plasma metabolomics analysis. 80
3.1 Co-authorship statement: 81
3.2 Copyright and License information: 81
3.3 Abstract: 81
3.4 Introduction: 84
3.5 Materials and Methods: 87
3.5.1 Study participants: 87
3.5.2 Demographic and medical information: 87
3.5.3 Metabolic profiling: 88
3.5.4 Statistical analysis: 92
3.6 Results: 97
3.7 Discussion: 113
Chapter 4: Metabolomic signatures for the longitudinal reduction of muscle strength over 10 years 119
4.1 Co-authorship statement: 120
4.2 Copyright and License information: 120
4.3 Abstract: 121
4.4 Introduction: 124
4.5 Materials and Methods: 127
4.5.1 Study participants: 127
4.5.2 Demographic information: 127
4.5.3 Muscle strength measurements: 128
4.5.4 Metabolic profiling: 130
4.5.5 Statistical analysis: 133
4.6 Results: 135
4.7 Discussion: 151
Chapter 5: A genome-wide association study identified novel genes associated with osteoarthritis 158
5.1 Co-authorship statement: 159

5.2 Abstract: 159

5.3 Introduction: 163

5.3.1 The structure of Newfoundland founder population: 163

5.4 Materials and Methods: 166

5.4.1 The Newfoundland Osteoarthritis Study (NFOAS): 166

5.4.2 Study participants: 167

5.4.3 Demographic information: 168

5.4.4 Genome-wide genotyping: 169

5.4.5 Pre-association QC of the raw SNP genotyping data: 171

5.4.6 Genotype imputation analysis: 173

5.4.7 Genome-wide association analysis: 178

5.4.8 Compare the associated genetic variants with OA in the NFOAS with the publicly available OA GWAS datasets: 179

5.5 Results: 180

5.6 Discussion: 196

Chapter 6: Whole exome sequencing analysis identified novel genes and digenic interaction of osteoarthritis in Newfoundland & Labrador population 209

6.1 Co-authorship statement: 210

6.2 Abstract: 210

6.3 Introduction: 214

6.4 Materials and methods: 218

6.4.1 Whole Exome Sequencing (WES): 218

6.4.2 Raw Exome sequencing data processing: 220

6.4.3 Variant Annotation analysis: 228

6.4.4 Case-only digenic interaction analysis: 231

6.5 Results: 234

6.6 Discussion: 245

Chapter 7: Novel genes for poor outcome of the total joint replacement surgery in the osteoarthritis patients identified by a genome-wide association study 261

7.1 Co-authorship statement: 262

7.2 Abstract 262

7.3 Introduction: 264

7.4 Materials and methods: 267

7.4.1 Study participants: 267

7.4.2 Assessment of total joint replacement outcome: 267

7.4.3 Demographic information: 269

7.4.4 Genome-wide genotyping: 270

7.4.5 Pre-association QC and imputation analysis of the genotyping data: 270

7.4.6 Genome-wide association analysis: 271

7.5 Results: 271

7.6 Discussion: 316
Chapter 8: General discussion and conclusion 325
8.1 General discussion: 326
8.2 Strengths 337
8.3 Limitations: 339
8.4 Conclusion: 340
8.5 Future directions: 341
REFERENCES: 344
APPENDIX 412

LIST OF FIGURES

Figure 1.1: Schematic showing the anatomical structure of the synovial joint. 7

Figure 1.2: Schematic showing the differences between the normal synovial joint and the OA joint. 11

Figure 1.3: Knee X-ray images showing the grades of the Kellgren and Lawrence (K/L) grading system. 20

Figure 1.4: Prevalence (%) of diagnosed OA by sex and age group in Canada during 2016–2017. 25

Figure 1.5: Incidence of diagnosed OA (per 1,000 persons-year), by sex and age group in Canada during 2016–2017. 25

Figure 1.6: Age-standardized prevalence (%) and incidence (per 1,000 persons-year) of diagnosed OA among Canadians aged 20 years and older, by sex, Canada, from 2007–2008 to 2016–2017. 26

Figure 3.1: Estimation of the factor numbers from the concentrations of 162 metabolites of 615 OA patients. 99

Figure 3.2: Estimation of the factor and cluster numbers from the concentrations of 162 metabolites of 615 OA patients. 101

Figure 3.3: The three distinct clusters of OA patients identified by K-means clustering method coupled with metabolomics data. 102

Figure 3.4: The ROC curve analysis results of each identified cluster. 106

Figure 3.5: Estimation of the cluster number in the training dataset of 308 OA patients. 108

Figure 3.6: Estimation of the cluster number in the validation dataset of 307 OA patients. 109

Figure 4.1: Muscle strength measurements including A) Hand-grip strength test, B) Knee extension strength measurement, and C) Leg muscle strength test. 129

Figure 4.2: A histogram showing the 409 study participants age distribution at baseline. 135

Figure 4.3: Plots of the fitted lines for changes over 10-year follow-up time for A) hand grip strength, B) knee extension, and C) leg muscle strength 137

Figure 4.4: The relationship between muscle strength and the study participants' age at the baseline, A) hand grip strength, B) knee extension strength, and C) leg muscle strength changes. 139

Figure 4.5: The differences between males and females in muscle strength changes over 10-year follow-up time: A) hand grip strength, B) knee extension strength, and C) leg muscle strength. 140

Figure 4.6: Volcano plots of the association results between metabolites and the A) hand grip strength change, B) knee extension strength change, and C) leg muscle strength changes over 10-year follow-up time. 145

Figure 4.7: GWAS for the asymmetric dimethylarginine (ADMA) in the 77 OA patients from the NFOAS. 150

Figure 5.1: A flow chart illustrates the GWAS pre-association quality control analysis steps. 172

Figure 5.2: A flow chart illustrates step-by-step genome-wide phasing using ShapeIT4 and imputation analysis using IMPUTE5. 174

Figure 5.3: A Manhattan plot of the GWAS results showing the most significantly associated genetic variations with OA in the 557 OA patients from the NFOAS in comparing with the 118 OA free controls from the same population. 182

Figure 5.4: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the OA. 183

Figure 5.5: The regional association plots of the significantly associated variants with the OA. 188

Figure 5.6: A Manhattan plot of the GWAS results showing the most significantly associated genetic variations with knee OA in 386 OA patients from the NFOAS in comparing with 118 OA free controls from the same population. 192

Figure 5.7: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the knee OA. . 193

Figure 5.8: A Manhattan plot of the GWAS results showing the associated genetic variations with hip OA in 171 OA patients from the NFOAS compared to 118 OA free controls from the same population. 194

Figure 5.9: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the hip OA. 195

Figure 6.1: A flow chart illustrates the steps of the GATK Best Practices pipeline performed to process the WES raw data. 221

Figure 6.2: A screenshot of the FASTQ file format. 222

Figure 6.3: A screenshot of the SAM file format. 223

Figure 6.4: A screenshot of the BAM file format. 225

Figure 6.5: A screenshot of the VCF file format. 227

Figure 6.6: Bar chart illustrates the number of OA samples carrying both interacting genes. 244

Figure 7.1: A Manhattan plot of the GWAS results showing the significantly associated genetic variations with non-responders to TJR pain in 441 OA patients from the NFOAS. 280

Figure 7.2: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the non-responders to pain TJR in the 441 OA patients. 281

Figure 7.3: The regional association plot for the chromosomal region showing the most significantly associated genetic variants with the TJR pain non-responders. 285

Figure 7.4: A Manhattan plot of the GWAS results showing the associated rs4797006 and rs146474469 on chr18 with the TJR pain non-responders in 315 knee OA patients. 289

Figure 7.5: A Manhattan plot of the GWAS results showing no association of the genetic variations at $p<5*10^{-8}$ with the non-responders to TJR pain in 126 hip OA patients in our study. 290

Figure 7.6: A Manhattan plot of the GWAS results showing the significantly associated genetic variations with non-responders to TJR function in 441 OA patients from the NFOAS. 292

Figure 7.7: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the non-responders to function TJR in the 441 OA patients. 293

Figure 7.8: The regional association plot for the chromosomal region showing the most significantly associated variants on chr21 with the TJR function non-responders. 296

Figure 7.9: A Manhattan plot of the GWAS results showing the associated rs28971219 on chr21 with the TJR function non-responders in 315 knee OA patients. 297

Figure 7.10: A Manhattan plot of the GWAS results showing no association of the genetic variations at $p<5*10^{-8}$ with the non-responders to TJR function in 126 hip OA patients in our study. 298

Figure 7.11: A Manhattan plot of the GWAS results showing the significantly associated genetic variations with non-responders to TJR both (pain and function) in 441 OA patients from the NFOAS. 300

Figure 7.12: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the non-responders to both (pain and function) TJR in the 441 OA patients. 301

Figure 7.13: A Manhattan plot of the GWAS results showing the significantly associated genetic variants with the TJR both (pain and function) non-responders in 315 knee OA patients. 305

Figure 7.14: A Manhattan plot of the GWAS results showing no association of the genetic variations at $p<5*10^{-8}$ with the non-responders to TJR both (pain and function) in 126 hip OA patients in our study. 307

Figure 7.15: A Manhattan plot of the GWAS results showing the significantly associated genetic variations with non-responders to TJR either (pain or function) in 441 OA patients from the NFOAS. 309

Figure 7.16: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the non-responders to either (pain or function) TJR in the 441 OA patients. 310

Figure 7.17: A Manhattan plot of the GWAS results showing the significantly associated genetic variants with the TJR either (pain or function) non-responders in 315 knee OA patients. 314

Figure 7.18: A Manhattan plot of the GWAS results showing no association of the genetic variations at $p<5*10^{-8}$ with the non-responders to TJR either (pain or function) in 126 hip OA patients. 315

LIST OF TABLES

Table 1.1: the clinical features of OA used in the diagnosis of the symptomatic OA. 18

Table 1.2: Kellgren and Lawrence (K/L) grading system 20

Table 1.3: The American College of Rheumatology (ACR) criteria for the diagnosis of knee, hip, and hand OA 22

Table 3.1: Biocrates AbsoluteIDQ kit list of 186 metabolite concentrations. 90

Table 3.2: The characteristics of the 615 OA patients and 237 OA-free controls. 97

Table 3.3: 17 distinct factors from 93 metabolites that had rotated factor loading > |0.3| in only one factor. 100

Table 3.4: Most significant metabolites contributing to the classification of endotypes of primary OA patients. 105

Table 3.5: The most significant metabolites contributing to the classification of endotypes of primary OA patients in the training and validation datasets. 110

Table 3.6: Significant epidemiological factors and comorbidities associated with each of identified clusters. 112

Table 4.1: TMIC Prime Metabolomics Profiling Assay list of 143 metabolite concentrations. 131

Table 4.2: The characteristics of the study participants (n=409). 136

Table 4.3: The number and percentage of the study participants that completed the muscle strength measurements at each of the follow-up time points. 136

Table 4.4: Muscle strength differences between males and females over 10-years follow-up period in whole cohort (n=409). 138

Table 4.5: Summary statistics for the 129 metabolites that passed the quality checks and were included in the analysis. 141

Table 4.6: Top metabolites associated with the muscle strength change measures in the study participants (n=409) over 10-years follow up time 146

Table 4.7: Top metabolites associated with the muscle strength change measures in three different age groups over 10-years follow up time 148

Table 5.1: The list of the 22 datasets across 11 OA phenotypes from the recent largest GWAS-meta-analysis of 13 cohorts from 9 populations with a total of 826,690 individuals. 179

Table 5.2: The characteristics of the 557 OA patients and 118 OA-free controls. 181

Table 5.3: The 29 genetic variants that were identified to be significantly associated with OA in the 557 patients from the NFOAS at the genome-wide significance level ($p=5*10^{-8}$)..... 184

Table 6.1: The 2*2 contingency table used for the interaction analysis between the aggregated rare variates in each gene and the common variants from other gene. 234

Table 6.2: List of the 55 variants that were found in at least 80% of the OA patients from the NFOAS with MAF ≥ 0.4 for both knee and hip OA. 236

Table 6.3: List of the variants that were prioritized in our WES analysis to be associated with OA. 241

Table 6.4: The digenic interaction results listing the significant interaction between the aggregated rare variants with the rs56158521 common missense variant in the HLA-DRB1 gene. 243

Table 7.1: The characteristics of the 315 knee OA patients and 126 hip OA patient of this study..... 273

Table 7.2: Classification of the OA patients of this study to responders and non-responders to the TJR surgery based on the absolute WOMAC change score and the OMERACT-OARSI classification criteria. 275

Table 7.3: Differences between responders and non-responders of TJR patients in each classification method, and the differences of non-responders to TJR between the absolute WOMAC change score and the OMERACT-OARSI classification criteria. 276

Table 7.4: The most significantly associated genetic variants with the TJR pain non-responding in the 441 OA samples. 284

Table 7.5: The most significantly associated genetic variants with the TJR pain non-responding in the 315 knee OA samples. 288

Table 7.6: The most significantly associated genetic variants with the TJR function non-responding in the 441 OA samples. 295

Table 7.7: The most significantly associated genetic variants with non-responding to both (pain and function) in the 441 OA samples. 303

Table 7.8: The most significantly associated genetic variants with non-responding to both (pain and function) of the TJR in the 315 knee OA patients. 306

Table 7.9: The most significantly associated genetic variants with non-responding to either (pain or function) of the TJR in the 441 OA samples. 312

LIST OF APPENDICES

Appendix A: A copy of the Western Ontario and McMasters Universities Osteoarthritis Index (WOMAC). 412

Appendix B: A list of genetic variations that were found by far to be significantly associated with OA in 11 OA phenotypes. This table was adapted from the Boer et al. (2021). 413

Appendix C: A copy of the updated study ethical approval that was acquired from the Health Research Ethics Authority of NL with the reference number 11.311. 417

Appendix D: A copy of the consent taken from the NFOAS participants to take a part in the research study. 418

Appendix E: The self-administered questionnaire that was used in the NFOAS to collect the OA patients' demographic and medical information including age, sex, BMI, and comorbidities. 422

Appendix F: A copy of the TASOAC study participants' interview questionnaire. 436

Appendix G: Copyright approval to adapt and use the leg muscle strength test using a dynamometer figure from Scott et al. (2009). 448

Appendix H: A list of the significantly associated variants with OA in the recent largest GWAS-meta-analysis (Boer et al. Cell 2021) that were identified to be associated with OA in all 557 OA patients, knee OA (n=386) patients, and hip OA (n=171) patients from the NFOAS with $p \leq 0.05$. 449

LIST OF ABBREVIATIONS

A	
AAPC	Adenomatous polyposis coli
AChR	Nicotinic acetylcholine receptors
ACL	Anterior cruciate ligament
ACLT	Anterior cruciate ligament transection
ACR	The American College of Rheumatology
ADAM	Disintegrin and metalloprotease domain protein family
ADAM8	ADAM Metallopeptidase Domain 8
ADMA	Asymmetric dimethylarginine; NG,NG-dimethyl-L-arginine.
AGEs	Advanced glycation end products
AGRN	Agrin
ANXA3	Annexin A3
APC	Adenomatosis Polyposis Coli
APCDD1L	APC Down-Regulated 1 Like
arcOGEN	The Arthritis Research UK Osteoarthritis Genetics
ARVD5	Arrhythmogenic right ventricular dysplasia type 5
ASIR	Age standardised incidence rate
ASPN	Asporin
ASTN1	Astrotactin 1
ASTN2	Astrotactin 2
ASU	Avocado soybean unsaponifiables
AUC	Area under the curve
AZU1	Azurocidin 1
B	
BANP	CTG3 Associated Nuclear Protein
BBS	Bardet-Biedl syndrome
BCAP29	B-Cell Receptor-Associated Protein 29
BMP-7	Bone morphogenetic protein 7
BMD	Bone mineral density
BMI	Body mass index
BML	Bone marrow lesion
BMP5	Bone morphogenetic protein 5
BRSK2	BR Serine/Threonine Kinase 2
BSS	Between Sum of Squares
BTNL2	Butyrophilin Like 2

C	
C4	Butyrylcarnitine
CAST	Cohort allelic sums test
CA5A	Carbonic Anhydrase 5A
CCDSS	The Canadian Chronic Disease Surveillance System
CCP	Cyclic citrullinated peptide antibodies
CDC5L	CDC5 Cell Division Cycle 5-Like
CDH1	Cadherin 1
CHADL	Chondroadherin like
CHD	Coronary heart disease
Chr	Chromosome
CHST11	Carbohydrate Sulfotransferase 11
95% CI	95% Confidence interval
CMC joints	Carpometacarpal
CNV	Copy number variant
CODING	The Complex Diseases in the Newfoundland population: Environment and Genetics
COG5	Component of Oligometric Golgi Complex 5
COL11A1	Collagen type XI alpha 1 chain
COL12A1	Collagen type XII alpha 1
COL9A3	Collagen Type IX Alpha 3 Chain
CRP	C-reactive protein
CTSK	Cathepsin K
CV	Coefficient of variation
D	
DAGLA	Diacylglycerol Lipase Alpha
DDAH-1	Dimethylarginine dimethylaminohydrolase-1
DDAH-2	Dimethylarginine dimethylaminohydrolase-2
DDH	Developmental dysplasia of the hip
DEGs	Differentially expressed genes
DEIRGs	Differentially expressed immune-related genes
DGCR6L	DiGeorge Syndrome Critical Region Gene 6 Like
DGKB	Diacylglycerol kinase beta gene
DIDO1	The Death Inducer-Obliterator 1
DIP joints	The distal interphalangeal
DMOADs	Disease-modifying OA drugs
DMP1	Dentin matrix protein 1
DOT1L	Disruptor of telomeric silencing 1-like

DPF1	Double PHD Fingers 1
DUS4L	Dihydrouridine Synthase 4-Like
DVWA	Double Von Willebrand Factor Type A domain
DZ twins	Dizygotic

E

ECM	The extracellular matrix
EGF receptor	Epidermal growth factor
ELANE	Elastase, Neutrophil Expressed
EQTN	Equatorin sperm-acrosome-associated
ESR	Erythrocyte sedimentation rate
ExAC database	Exome Aggregation Consortium

F

F8	Coagulation factor VIII
FADS gene	Fatty acid desaturase
FADS2	Fatty Acid Desaturase 2
FAI	Femoroacetabular impingement
FAM101A	Filamin-Interacting Protein, Refilin A
FAM20C	FAM20C Golgi Associated Secretory Pathway Kinase
FAPs	Fibro-adipogenic progenitors.
FDA	Food and Drug Administration
FFAs	Free fatty acids
FGF18	Fibroblast growth factor 18
FILIP1	Filamin A interacting protein 1
FLS	Fibroblast-like synoviocytes
FRZB	Frizzled-related protein 3
FSS	Familial short stature
FTO	Fat Mass and Obesity Associated

G

GBD	The Global Burden of Diseases
GDA	Illumina Infinium Global Diversity 8 v1.0 genotyping arrays
GDF5	Growth differentiation factor5
GID8	GID Complex Subunit 8 Homolog
GLT8D1	Glycosyltransferase 8 Domain Containing 1
GNL3	Guanine Nucleotide Binding Protein-Like 3
GO Consortium	The Genetics of Osteoarthritis

GOA	Generalized OA
GPCRs	The G-protein-coupled receptors
GPR123	G-Protein Coupled Receptor 123
GPR22	G Protein-Coupled Receptor 22
GTP	Guanoside triphosphate
GWAS	Genome-wide association studies

H

HA	Hyaluronic acid
HDAC9	Histone Deacetylase 9
HEMA	hemophilia A
HFE	Homeostatic iron regulator
HLA	Human Leukocyte Antigen
HMG box 9	High mobility group
HMM	Hidden Markov Model
HNPCC	Hereditary non-polyposis colorectal cancer syndrome
HPB1	HMG-box Transcription Factor 1
HRC	Human Reference Consortium
HWE	Hardy Weinberg disequilibrium

I

IBD	Identity-by-descent
IBS	Identity-by-state
Ig	immunoglobulin
IL-1β	Interlukin-1 beta
IL-6	Interleukin-6
IL1RN	Interluekin 1 receptor antagonist
ILD	Interstitial lung disease
INDELs	Insertion and deletion
ITIH1	Inter-alpha-trypsin inhibitor heavy chain 1

J

JSN	Joint space narrowing

K

K/L	Kellgren and Lawrence radiographic grading scheme
Kg	Kilogram
KIR	Killer Cell Immunoglobulin-Like Receptor

KLHDC5	Kelch Domain Containing 5
KMO	The Kaiser-Meyer-Olkin

L	
LD	Linkage disequilibrium
LMX1B	LIM homeobox transcription factor 1 beta
LNMMA	NG-monomethyl-L-arginine.
LOD	Limit of detection
LPS	lipopolysaccharide
LSM	The Li and Stephens model
LTBP1	Latent transforming growth factor beta binding protein 1
LTBP3	Latent transforming growth factor beta binding protein 3
lysoPC a C16:0	Lyso-phosphatidylcholine

M	
MAF	Minor allele frequency
MAML3	Mastermind Like Transcriptional Coactivator 3
MAP2K6	Mitogen-Activated Protein Kinase Kinase 6
MATN3	Matrilin 3
MaxSpSe	the maximum sensitivity and specificity simultaneously
MC5R gene	Melanocortin5 receptor
MCF2L	Protein encoding MCF.2 cell line derived transforming sequence-like
MCP joints	Metacarpophalangeal
MCs	Mast cells
MED16	Mediator Complex Subunit 16
MEN-1	Endocrine neoplasia type 1
MetS	Metabolic syndrome
MHC	Major Histocompatibility Complex
MJOA	Multiple joint OA
MLR	Mixed logistic regression
MMP-13	The matrix metalloproteinases 13
MMPs	The matrix metalloproteinases
MOB3B	MOB Kinase Activator 3B
MPa	Megapascal
MRI	Magnetic resonance imaging
MSA	Measure of sampling adequacy
MUC5B	Oligomeric Mucus/Gel-Forming
MuSCs	Muscle stem cells.

MZ twins	Monozygotic
N	
NCOA3	Nuclear Receptor Co-Activator 3
NDRG1	N-Myc Downstream Regulated 1
Nell-1	Neural epidermal growth factor-like 1
Nf-kB	Nuclear factor-kappa B
NFOAS.	Newfoundland Osteoarthritis Study
NMJ	Neuromuscular junction
NO	Nitric Oxide
NOS1	Nitric Oxide Synthase 1
NPEPL1	Aminopeptidase Like 1
NSAIDs	Non-steroidal anti-inflammatory drugs
NTSR1	Neurotensin Receptor 1
O	
OA	Osteoarthritis
OARSI	The Osteoarthritis Research Society International
OF	Ossification front
OGFR	Opioid Growth Factor Receptor
OMERACT	the Outcome Measures in Arthritis Clinical Trials
OP	Osteoporosis
OR4F16	Olfactory Receptor Family 4 Subfamily F Member 16
OR4F29	Olfactory Receptor Family 4 Subfamily F Member 29
OR4F3	Olfactory Receptor Family 4 Subfamily F Member 3
P	
PAG	Periaqueductal gray
PBWT	Positional Burrows–Wheeler transform
PCA	Principal component analysis
PC aa C32:2	Diacyl-phosphatidylcholines with 32 carbons and two double bonds
PC ae C38:2	acyl-alkyl Phosphatidylcholine with 38 carbons and two double bonds
PC ae C40:3	acyl-alkyl phosphatidylcholine with 40 carbons and three double bonds
PERM1	PPARGC1 And ESRR Induced Regulator, Muscle 1
PGM5P2	Phosphoglucomutase 5 Pseudogene 2
PHAC	The Public Health Agency of Canada
PIP joints	The proximal interphalangeal
PLEC	plectin

PLPPR3	Phospholipid Phosphatase Related 3
PPP1R14A	Protein Phosphatase 1 Regulatory Inhibitor Subunit 14A
PRKAR2B	Protein Kinase cAMP-dependent Regulatory Type II Beta
PRMT-I	Protein arginine methyltransferase type I.
PRMT-II	Protein arginine methyltransferase type II.
PROs	Patient-reported outcomes
PRTN3	Proteinase 3
PSI	Pounds per square inch.
PSMD8	Proteasome 26S Subunit, Non-ATPase 8
PTHLH	Parathyroid Hormone-Like Hormone
Q	
QC	Quality Control
R	
RA	Rheumatoid arthritis
Ran	Ras-related nuclear protein
RANBP1	RAN Binding Protein 1
REML	Restricted maximum likelihood.
RF	Rheumatoid factor
ROC	Receiver operating characteristic
ROS	The reactive oxygen species
RTN4R	Reticulon 4 Receptor
RUNX2	Runt Related Transcription Factor 2
RWDD2B	RWD-domain-containing 2B
S	
SBNO1	Strawberry notch homolog 1
SCAD	short-chain acyl-CoA dehydrogenase deficiency
SD	Standard deviation
SDMA	Symmetric dimethylarginine; NG,N’G-dimethyl-L-arginine.
SENP6	Sentrin specific peptidase 6
sGAG	Sulfated-glycosaminoglycans
SIPA1L3	Signal Induced Proliferation Associated 1 Like
SMAD3	SMAD family member 3
SMO	Smoothened, frizzled class receptor
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant

SRY	Sex-determining region Y
ST3GAL1	ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 1.
STAT3	Signal transducer and activator of transcription 3
STX16	Syntaxin 16
SUPT3H	Suppressor of Ty3 Homolog
SV	Structural Variant

T

TASOAC	Tasmanian Older Adult Cohort.
TEKT4P2	Tektin 4 Pseudogene 2
TENS	Transcutaneous nerve stimulation
TFF	Trefoil factor family peptides
TGF-β	Transforming growth factor beta
TNF-α	Transforming growth factor alpha
THR	Total hip replacement
TJR	Total joint replacement
TKR	Total knee replacement
TMEM 132D	Transmembrane Protein 132D
TMEM179	Transmembrane Protein 179
TMEM191B	Transmembrane Protein 191B
TMEM43	Transmembrane Protein 43
TMIC	The Metabolomics Innovation Centre
TNC	Tenascin C
TNF	Tumor necrosis factor
TOLLIP	Toll Interacting Protein
TP63	Tumor Protein p63
TPTE gene	Transmembrane Phosphatase With Tensin Homology
TRPV4	Transient receptor potential vanilloid 4
TSHR	Thyroid Stimulating Hormone Receptor
TSS	Total Sum of Squares
TUBGCP2	Tubulin Gamma Complex Associated Protein 2

U

UHPLC	Ultra-high-performance liquid chromatography system.
UTR	Untranslated region

V

VAPB	VAMP Associated Protein B And C

VCF files	Variant call format
VEGF	Vascular Endothelial Growth Factor A
VEGFR2	Vascular endothelial growth factor receptor 2
VEGFR3	Vascular endothelial growth factor receptor 3
VWC2	Von Willebrand factor C domain containing 2 gene

W

WAT	White adipose tissue
WES	Whole exome sequencing
WISP1	WNT1 Inducible Signaling Pathway Protein 1.
Wnt	Wingless/Integrated
WOMAC	The Western Ontario and McMasters Universities Osteoarthritis Index
WWP2	WW domain containing E3 ubiquitin protein ligase 2

Z

ZC3H11B	Zinc-finger CCCH-type containing 11B pseudogene
ZFAT	Zinc Finger And AT-Hook Domain Containing.
ZNF345	Zinc-finger-protein 345

RESEARCH OUTPUT AND AWARDS

ABSTRACTS:

A. Oral presentations:

1 **Salem Werdyani**, Ming Liu, Andrew Furey, Zhiwei Gao, Proton Rahman, Guangju Zhai. *A genome-wide association study identified novel genes for poor outcome of the total joint replacement surgery in osteoarthritis patients* was presented as a podium presentation during the Canadian Orthopedic Research Society (CORS) Top Canadian Research Paper Session: Founders' Award Candidates at The Canadian Orthopedic Association Annual meeting, 8-11 June 2022, Quebec City, Canada.

2 **Salem Werdyani**, Ming Liu, Andrew Furey, Zhiwei Gao, Proton Rahman, Guangju Zhai. *Novel genes for poor outcome of the total joint replacement surgery in osteoarthritis patients identified by a genome-wide association study*. The Annual Research Presentation Days January 25th and 31st, 2022 of the Canadian Arthritis Research Conference.

3 **Salem Werdyani**, Quan Li, Ming Liu, Amanda Dohey, Dianne Codner, Darren O'Rielly, Proton Rahman, Guangju Zhai. *Novel genes found to be associated with osteoarthritis by whole exome sequencing analysis*. The Virtual Canadian Connective Tissue Conference (CCTC), June - August 2021.

4 **Salem Werdyani**, Dawn Aitken, Zhiwei Gao, Ming Liu, Edward W. Randell, Proton Rahman, Graeme Jones, Guangju Zhai. *Metabolomic signatures for the longitudinal reduction of muscle strength over 10 years*. The Osteoarthritis Society Research International (OASRI 21) Virtual World Congress, April 29th- May 1st, 2021.

5 **Salem Werdyani**, Dawn Aitken, Zhiwei Gao, Ming Liu, Edward W. Randell, Proton Rahman, Graeme Jones, Guangju Zhai. *Metabolomic signatures for the longitudinal reduction of muscle strength over 10 years*. The original research presentation day at the Arthritis Society and Canadian Rheumatology Virtual conference 2021, 5th February 2021.

6 **Salem Werdyani**, Ming Liu, Zikun Xie, Andrew Furey, Zhiwei Gao, Proton Rahman, Guangju Zhai. *Genes related to muscle strength, behavioral trait, pain response, and inflammation are associated with poor outcome of the total joint replacement therapy in primary osteoarthritis patients*. The Top Canadian Research Paper Session of the 2020 Virtual Canadian Orthopaedic Research Society (CORS) Annual Meeting, 19-21 June 2020.

7 **Salem Werdyani**, Ming Liu, Guang Sun, Andrew Furey, Edward Randell, Proton Rahman, Guangju Zhai. *Plasma metabolomics identified three distinct endotypes of primary osteoarthritis patients*. Accepted for a plenary oral presentation at the Osteoarthritis Research Society International (OARSI 2020) World Congress on

Osteoarthritis, April 30–May 3, 2020, at the Messe Wien Exhibition & Congress Center in Vienna, Austria. The conference was canceled due to the COVID-19 pandemic.

8 **Salem Werdyani**, Ming Liu, Andrew Furey, Zhiwei Gao, Proton Rahman, Guangju Zhai. *Possible genetic predictors for poor outcome of total joint replacement therapy in primary osteoarthritis patients*. Invited keynote speaker in the seniors living well with arthritis in Newfoundland conference, Arthritis society, November 03, 2019, Capital Hotel, St. John's, Newfoundland, Canada.

9 **Salem Werdyani**, Ming Liu, Andrew Furey, Edward Randell, Proton Rahman, Guangju Zhai. *Endotypes of primary osteoarthritis identified by a metabolomics approach*. The Canadian Connective Tissue Conference (CCTC), May 29-31, 2019, McGill University, Montreal, Québec, Canada.

B. Poster presentations:

1. **Salem Werdyani**, Ming Liu, Andrew Furey, Zhiwei Gao, Proton Rahman, Guangju Zhai. *A genome-wide association study identified novel genes associated with osteoarthritis*. The Osteoarthritis Society Research International (OASRI 22) Virtual World Congress, April 7th-10th, 2022.

2. **Salem Werdyani**, Ming Liu, Andrew Furey, Zhiwei Gao, Proton Rahman, Guangju Zhai. *Novel genes for poor outcome of the total joint replacement surgery in*

osteoarthritis patients identified by a genome-wide association study. The Canadian Arthritis Research Conference 2022, February 7th 8th, 2022.

3. **Salem Werdyani**, Quan Li, Dawn Aitken, Ming Liu, Amanda Dohey, Dianne Codner, Darren O'Rielly, Proton Rahman, Guangju Zhai. *Whole exome sequencing analysis identified five novel genes associated with Osteoarthritis*. The virtual American Society of Human Genetics (ASHG), October 18-22, 2021.

4. **Salem Werdyani**, Dawn Aitken, Zhiwei Gao, Ming Liu, Edward W. Randell, Proton Rahman, Graeme Jones, Guangju Zhai. *Metabolomic signatures for the longitudinal reduction of muscle strength over 10 years*. The Arthritis Society and Canadian Rheumatology Virtual conference 2021, 16-17 February 2021.

5. **Salem Werdyani**, Ming Liu, Zikun Xie, Andrew Furey, Zhiwei Gao, Proton Rahman, Guangju Zhai. *Genes related to muscle strength, behavioral trait, pain response, and inflammation are associated with poor outcome of the total joint replacement therapy in primary osteoarthritis patients*. The Canadian Arthritis Research Conference, February 25-26, 2020, at the Fairmont Empress Hotel, Victoria, British Columbia, Canada.

C. Other presentations as a co-author:

1. Christie Costello, Ming Liu, **Salem Werdyani**, Proton Rahman, Edward W. Randell, Guangju Zhai. *Genome-wide association study identified MC5R as a*

candidate gene for refractory knee pain in knee osteoarthritis patients. The Canadian Arthritis Research Conference 2022, February 7th 8th, 2022.

- **PUBLICATIONS:**

1. **Salem Werdyani**, Ming Liu, Hongwei Zhang, Guang Sun, Andrew Furey, Edward W. Randell, Proton Rahman, Guangju Zhai. *Endotypes of primary osteoarthritis identified by plasma metabolomics analysis*. Rheumatology (Oxford), 2020. PMID: 33159799, DOI: 10.1093/rheumatology/keaa693.

2. **Salem Werdyani**, Dawn Aitken, Zhiwei Gao, Ming Liu, Edward W. Randell, Proton Rahman, Graeme Jones, Guangju Zhai. *Metabolomic signatures for the longitudinal reduction of muscle strength over 10 years*. BMC Skeletal Muscles Journal. PMID: 35130970, DOI: 10.1186/s13395-022-00286-9.

- **AWARDS:**

1. I had been awarded the Canadian Arthritis Society Training Graduate PhD Salary Award for three years (September 2019– September 2022).

2. My study entitled "*Endotypes of primary osteoarthritis identified by plasma metabolomics analysis*" has been selected as a part of the Arthritis Society top 10 research advances of 2021.

https://mail.google.com/mail/u/0/#search/arthritis+society/FMfcgzGmtNjLfzXJKjFFTNVmwpnGJFMM, https://gazette.mun.ca/research/personalized-treatment/.

3. I was selected among the Top Canadian Researchers: Founders' Award Candidates at The Canadian Orthopedic Association Annual meeting, 8-11 June 2022, Quebec City, Canada.

4. I received the division of Biomedical Sciences (Genetics) research forum award, May 2020.

5. I received the Canadian Connective Tissue Conference (CCTC2019) travel award to attend the CCTC, May 29-31, 2019, McGill University, Montreal, Québec, Canada.

Chapter 1: General introduction

1.1 WHAT IS OSTEOARTHRITIS?

Osteoarthritis (OA) is the most common chronic rheumatic disabling disorder that affects about 10% of world population aged 60 years or older (1,2). OA is an age-related degenerative musculoskeletal disease that consists of heterogeneous groups of overlapping distinct conditions with different aetiologies, but it has similar biological, morphological, and clinical presentations (3). Clinical manifestation of OA ranges from asymptomatic to incidental finding on clinical and/or radiographic examination to rapidly progressive disabling disorder (4). While OA may affect any joint, it most commonly affects the knees, hips, hands, joints of the lower cervical, and lower lumbar spine, yet less commonly affects the elbow, wrist, shoulder, and ankle (5). The disease affects the entire joint and deteriorates its compartments including the articular cartilage, subchondral bone, synovial membrane, ligaments, capsule, and periarticular muscles leading to joint pain, stiffness, reduced range of movement, joint grating, and limited joint function (6,7).

1.1.1 OA definition:

Clinical and molecular heterogeneity of OA is the major obstacle to developing a universal definition of the disease (8). Initially, OA was defined as gradual damage and loss of articular cartilage at the synovial joints due to aging wear and tear processes (9). This definition was accepted based on the United States Food and Drug Administration (FDA) disease classification systems that define diseases primarily based on their signs and symptoms (10). However, this definition classified OA based on its morphological,

and clinical manifestations, it did not classify distinct subtypes of OA based on their underlying molecular mechanisms (10). As a result, this disease definition failed to optimally incorporate the new OA biological insights and was a fundamental hurdle of developing disease-modifying OA drugs (DMOADs). Subsequently, the US National Academy of Sciences created a new disease taxonomy system to define and describe diseases based on their intrinsic biology as well as signs and symptoms (10). Thus, several different strategies have been proposed for describing OA phenotypes based on modern imaging and pathophysiological mechanisms (11). Based on this definition criteria, OA was estimated to be a diverse group of conditions caused by a variety of biochemical, hormonal, and mechanical factors that lead to overlapping morphological, and clinical presentations (11). Then, the most recent definition of OA that is globally recognized for classifying the disease has been proposed by the Osteoarthritis Research Society International (OARSI) (10). The OARSI defines OA as "a disorder involving movable joints characterized by cell stress and extracellular matrix degradation initiated by micro- and macro-injury that activates maladaptive repair responses including pro-inflammatory pathways of innate immunity. The disease manifests first as a molecular derangement (abnormal joint tissue metabolism) followed by anatomic, and/or physiologic derangements (characterized by cartilage degradation, bone remodeling, osteophyte formation, joint inflammation, and loss of normal joint function), that can culminate in illness (10)."

Furthermore, OA has been classified into primary and secondary OA based on the disease aetiologies. Primary OA, also known as idiopathic OA, is a form of the condition that

develops over time without a known cause, however biochemical and biomechanical factors including obesity, muscle weakness and neurologic dysfunction may be implicated in the primary OA aetiology and pathogenesis (12). It affects women more often than men, especially after menopausal age. It usually affects knees, hips, hands, and big toes (13,14). On the other hand, secondary OA has a known underlying cause such as trauma or joint injury, metabolic disorders, congenital joint disorders, avascular necrosis, and other types of arthritis (13).

Symptomatic and radiographic OA are the most frequently used OA definitions. The symptomatic OA classification criteria are based on experiencing the disease symptoms, as well as the findings from the clinical examinations including pain and at least three of the following six symptoms: age >45 years, morning stiffness duration <30 minutes, crepitus on active motion, tenderness of the bony margins of the joint, bony enlargement noted on examination, and a lack of palpable warmth of the synovium (15–17). While radiographic OA investigates the nonuniform joint space narrowing (JSN), osteophyte formation, cyst formation, and subchondral sclerosis on radiographic imaging including X-ray and magnetic resonance imaging (MRI). Multiple classification methods are used to define the degree of radiographic severity of OA. For example, the American College of Rheumatology (ACR) classification of OA was a widely used classification system (15), but it has several limitations, including: a) artificial separation of nodal and erosive (non-nodal) OA as two distinct subsets of hand OA; b) inclusion of intervertebral disc degeneration and Forestier's disease as a subset of spinal OA, even though OA is pathologically confined to synovial joints; and c) the guideline development group has

recognized that these distinctions are arbitrary and have yet to be validated (18). Thus, the Kellgren and Lawrence (K/L) radiographic grading scheme that will be mentioned in detail in the diagnosis of OA section became for more than four decades the most used radiographic classification system for clinical applications and research. This scoring system assesses OA based on five radiographic grades from 0 to 4, and defines OA by the presence of a definite osteophyte in grade 2 and above, and more severe grades by the presumed successive appearance of the JSN, sclerosis, cysts, and deformity in the subchondral bone (19).

OA frequently affects multiple joints. Hence, it is also defined as generalized OA (GOA) or multiple joint OA (MJOA) due to the polyarticular involvement of OA (20). This form of polyarticular OA expresses a high clinical and disease burden in affected individuals (21). It may represent a distinct aetiology from mono-articular OA (22) that should be considered separately from single joint OA when assessing the risk factors and associated disease (23).

1.1.2 OA pathogenesis:

The underlying pathogenesis of OA remains largely unknown. However, OA pathology has been referred to as repetitive joint overloading, leading to an excessive biomechanical joint stress (24). Moreover, a combination of genetic, metabolomic, and environmental factors are thought to be implicated in the disease development and progression (6). In the past, OA was deemed to simply be a degenerative ‘wear and tear’ process in the synovial

joints (9). Following over two decades of OA research, the degenerative 'wear and tear' concept has been modified, and OA is explained in the current dogma as 'a whole joint degenerative disease' that involves a deterioration continuum of multiple joint components (25).

OA affects synovial joints that are the most common and main functional joints of the body (26). These joints are called diarthroses due to their free movement (27). There are several tissues that are essential components of synovial joints. Each tissue has its own composition and structure which plays critical functional role in effectively dealing with mechanical loads encountered throughout the life. The main structural component of a synovial joint is the presence of a joint cavity (26,27). The cavity is surrounded by the articular capsule, which is a fibrous connective tissue where the articulating surfaces of the bones contact each other, **Figure 1.1**.

Hyaline articular cartilage covers the entire articulating surface of each bone and distributes an impressive load across the joint to allow the articulating bones to move smoothly against each other without damaging the underlying bone tissue (28). Articular cartilage is composed of chondrocyte cells surrounded by the extracellular matrix (ECM). The ECM is comprised of water, 15-20% collagen type II, 4-7% proteoglycans, and a small amount of calcium salt, but it does not have blood vessels, nerves, or lymphatics (29). The turnover rate of collagen is relatively slow in comparison with the rapid turnover of the proteoglycan (30). The ECM provides elasticity and high tensile strength to the articular cartilage. It allows a frictionless movement and functions as a biological

shock absorber of mechanical forces that are distributed via the underlying subchondral bone (31).

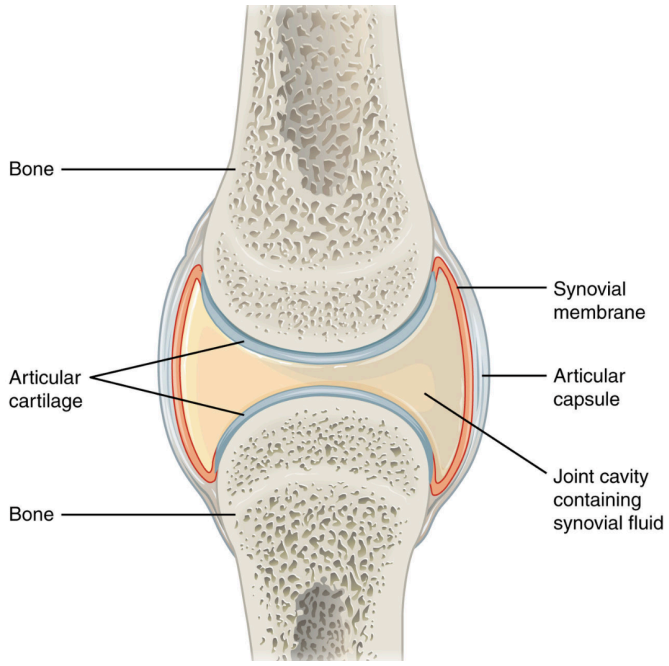


Figure 1.1: Schematic showing the anatomical structure of the synovial joint that has articular capsule with synovial membrane, synovial cavity containing synovial fluid, wide joint space, smooth hyaline articular cartilage covering the surface of the subchordal bone, allowing the articulating bones to move smoothly against each other without damaging the underlying bone tissue. This image was created by the OpenStax College and downloaded from WIKIMEDIA COMMONS that is licensed under Creative Commons Attribution-Share Alike 3.0 un-ported License.

Chondrocytes are the only cell type in the articular cartilage that represent about 5% of its total volume, they are enormously important for the production of the ECM components, modulate their turnover, and produce the proteolytic enzymes responsible for their breakdown (32). In turn, chondrocytes are themselves influenced by growth factors and

cytokines, structural and physical stimuli, and even the components of the ECM (33). Chondrocytes also play a central role in the regulation of the ECM water content; they produce the negatively charged sulfated-glycosaminoglycans (sGAG) that gives the cartilage a fixed charge density (34). This fixed charge density allows large amounts of positive ions and water to enter the cartilage leading to a high hydrostatic pressure within the collagen network. The water is squeezed out of the cartilage with each compressive force during joint movement (34,35). Due to the continuous presence of the sGAG in the ECM, water is immediately attracted back into the ECM when loads are reduced. This mechanism gives the cartilage a spongelike structure, allows it to continuously absorb forces between 10-20 megapascal (MPa) (36), and protects the underlying bone (35,37,38). Some synovial joints of the body also have a fibrocartilage structure located between the articulating bones, such as menisci which is large C-shaped found between articulating bones to smooth their movement (39).

Subchondral bone is bony components underlies the articular cartilage at the synovial joints (40,41). It consists of subchondral bone plate, the underlying trabecular bone, and bone marrow space. The subchondral bone plate is a thin dome-like cortical lamella that is made of cortical bone and is separated from the articular cartilage by the zone of calcified cartilage (42). It is penetrated by channels that provide a direct link between articular cartilage and subchondral trabecular bone. An extremely high number of arterial and venous vessels, as well as nerves scattered through the channels with tiny branches reach calcified cartilage (40,43). Subchondral trabecular bone has significant shock-absorbing and supporting functions in normal joints, it is also thought to have an

important role in cartilage nutrient supply and metabolism (44). Subchondral trabecular bone is more porous and metabolically active, containing blood vessels, sensory nerves, and bone marrow than the subchondral bone plate (45). Thus, Subchondral bone has a central role in evenly distributing forces from weight bearing and impact during physical activity, which protects the cartilage from high peak stresses and the ECM damage, as well as it contributes to the cartilage homeostasis.

Other components of the synovial joints include synovial membrane (synovium) that lines the articular capsule, and the synoviocyte cells of the synovial membrane secrete the synovial fluid that fills the joint cavity (46). Synovial fluid consists of the hyaluronic acid (HA) that mainly lubricates, reduces friction, and nourishes the articular cartilage at the joint. Also, synovial joints contain ligament and tendons. Where ligaments connect two bones and stabilize organs, and tendons connect muscles to the bone at the joint (39).

Normal joint structure and function depend on the ability of constituent tissues to respond to stress and load. Usually, large loads or forceful impacts at joints stimulate repair mechanisms to repair and restore normal function of the joint and ensure that the joint continues to dissipate load correctly (47). However, when the damage at the joint exceeds the repair rate, chondrocytes fail to maintain normal homeostasis between synthesis and degradation of the ECM components (32). For instance, excessive loading and injuries produce microfractures at the affected joint. In response to the microfractures, chondrocytes synthesize more inflammatory cytokines including tumor necrosis factor-α (TNF-α) and interleukin-1β (IL-1β). If these cytokines exceeded the normal concentration

and were not swept over by the synovial macrophages, they become mediators of inflammation that bind to chondrocyte receptors leading to release of destructive enzymes like the matrix metalloproteinases (MMPs) (24). In turn, the MMPs inhibit collagen type II production (48). Specifically, the matrix metalloproteinases 13 (MMP-13) cleaves collagen type II and degrades other ECM proteins including collagen types IV and IX, aggrecan, and perlecan (49). Loss of homeostasis yields elevation in water level that results in dilution of the proteoglycans concentration and increases apoptosis of chondrocytes (50). Shortly after the initiation of cartilage degradation cascade at the OA joint, the rate of anabolic and catabolic activity increases. To primarily maintain the structure and function of the articular cartilage, synthesis of the ECM collagen, proteoglycans, and hyaluronate raises. Then, proliferation of chondrocytes expedites leadings to expansion of the cartilage size. As a result, cartilage softens, fibrillates, diminishes, and fissures (51). Next, the water content of the cartilage decreases gradually due to the reduction of the sGAG production (24,49), and over time cartilage thins resulting in JSN leaving the underlying subchondral bone plate completely exposed in which is called chondropathy (52), **Figure 1.2**.

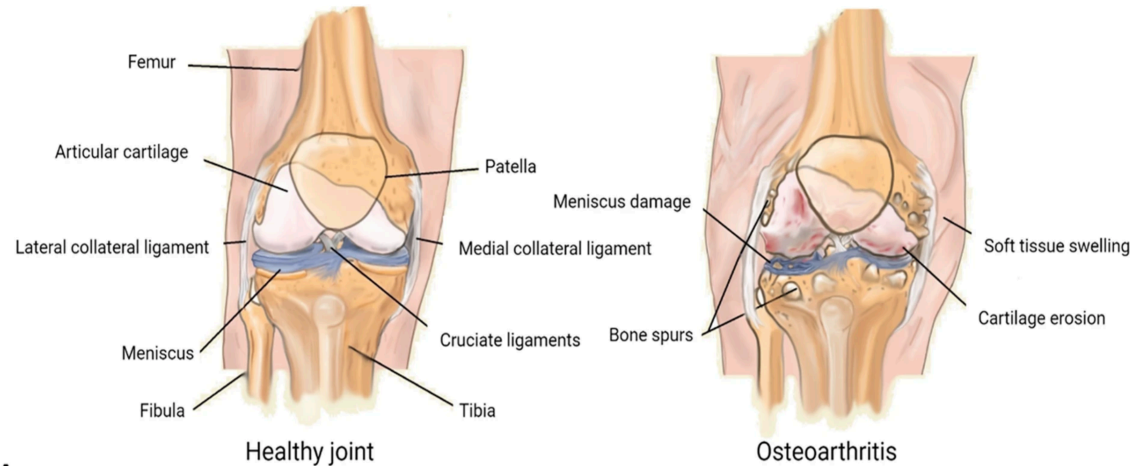


Figure 1.2: Schematic showing the differences between the normal synovial joint and the OA joint. OA Joint has cartilage degradation, joint space narrowing, abnormal bone growth (osteophytes), and inflammation of the synovial tissues in comparison with the healthy joint. This image was adapted from Zhang *et al.* (2022) that is licensed under a Creative Commons Attribution 4.0 International License (53).

Deterioration of the articular cartilage is intercorrelated with subchondral bone remodeling in the OA joint. Since the subchondral bone changes are detected by the MRI at the early stages of OA (54), it is not fully understood if the subchondral bone changes are intimately implicated in the initiation of OA or are consequences of the articular cartilage damage (55). It is estimated that the exposure of the subchondral bone leads to angiogenesis of blood vessels accompanied with sensory nerves that penetrate the cartilage. As a result, the subchondral plate thickness increases, the architecture of subchondral trabecular bone is modified, and outgrowth of the bone and cartilage called osteophytes occur at the joint margins to decrease the load across the joint by increasing its surface area (40). Subsequently, advanced bone remodeling processes occur in the severe cartilage degradation regions leading to aseptic bone necrosis and invasion of the

synovial fluid to the bone marrow leading to the formation of bone cysts that are frequently seen in late stage of OA (56). These cysts are pouches of fibrous tissues filled with a semi-liquid fluid and are connected to each other and to the joint space through microfractures in the bone.

Afterward, chondral, osseous, or osteochondral fragments break-down from the bone surfaces that is covered with hyaline cartilage and form intra-articular loose bodies in the joint cavity. Loose bodies that move freely in the joint cavity are predisposed to be entrapped between the articular surfaces causing intermittent joint locking, limitation of motion, pain, and intra-articular effusion (57). Also, these loose bodies are thought to be implicated in the development of synovitis in OA. Products of loose bodies that are released into the synovial fluid are phagocytosed by synovial cells called synoviocytes. As a result, synoviocytes increase in number leading to raise in the synovial lining cell layer thickness and increase vascularity and inflammatory cell infiltration of the synovial membranes at late stages of OA (58). Activated synoviocytes in the inflamed synovium produce large quantities of proteolytic enzymes including the MMP1, MMP3, MMP9, and MMP13 (59). Also, synoviocytes secrete pro-inflammatory cytokines, such as TNF-α, IL-1β, and interleukin-6 (IL-6) leading to vascular hyperplasia in osteoarthritic synovial membranes (60). These destructions in the OA joint are followed by deterioration of ligaments, capsule, menisci, periarticular muscles, peripheral nerves, and alteration of the biochemical and biomechanical joint's properties leading to irreversible damage to the joint as whole, pain, stiffness, and disability in affected individuals (25,61).

1.1.3 Symptoms and signs of OA:

OA takes years to reach a severe stage. Where it is primarily asymptomatic at the early stages, its clinical symptoms and signs appear gradually as it progresses. Symptoms of OA vary significantly from person to person. For instance, some patients are greatly affected by their symptoms, but their X-rays fail to detect joint structural changes. Other patients may have moderate to very few symptoms, while their X-rays appearance include JSN, osteophytes formation, and sclerosis (62).

Joint pain is the strongest OA symptom that forces the patients to seek medical advice. Pain affects one or few joints at a time that usually develops gradually and progresses slowly over long period of time with variable intensity (55). It usually initiates from deep within the joint and is described as aching, sharp or a burning pain that is poorly localized. Pain is aggravated by activity and relieved by rest. It is also often described as mechanical that is worsen with activity such as walking or climbing stairs due to putting pressure and weight on the affected joint (62).

At late disease stages when the structure of the joint is badly damaged, pain occurs at rest and night-time, its severity may wake the OA patients while sleeping (63). Since the assessment of pain is highly subjective and its intensity is variable, evaluation of OA pain became a challenging task. Thus, using self-reported questionnaires like the Western Ontario and McMasters Universities Osteoarthritis Index (WOMAC) (64) that can record intensity of pain, severity of disease, and impairment of function are preferred in

epidemiological studies (65). **Appendix A** contains a copy of the WOMAC Osteoarthritis Index. The WOMAC score consists of 24 items that are divided into pain, stiffness, and physical function subscales. While five pain items assess pain during walking, using stairs, in bed, sitting or lying, and standing; two stiffness scales evaluate limitations of motion after first waking and later in the day; and physical function measures estimate difficulties in stair use, rising from sitting, standing, bending, walking, getting in/out of a car, shopping, putting on/taking off socks, rising from bed, lying in bed, getting in/out of bath, sitting, getting on/off toilet, heavy and light household duties. Each response weighs four points, thus the WOMAC gives pain score (0-20), stiffness (0-8), and function score (0-68), with zero represents no pain, stiffness, or function difficulties.

Stiffness of the joint that is described as 'gelling' is another symptom of OA, it is reported in most OA patients with different degrees ranging from slowness of the joint to pain with initial joint motion. The joint stiffness is usually at its worst in the morning following to first rising from bed that lasts for ≤30 minutes (62). Also, it can be troublesome following to joint rest during the day. Moving the joint or doing some exercise for a few minutes can help restore the joint function. Function impairment and difficulties in performing the daily activities due to joint stiffness are usually linked to lower quality of life of the OA patients including social isolation and loss of job opportunities (66).

As the disease progresses, its symptoms and signs become more noticeable, one of several OA signs is limitation of joint motion and loss of joint flexibility. Reduction in the

normal range of motion within the joint can be noticed when the joint becomes less movable, and eventually it may not be possible to fully straighten or bend it (62).

Moreover, ligamentous instability, significant joint destruction, and decline in the strength and function of the joint muscles lead to instability and insecurity of OA patients (62,67). Also, OA patients can feel and hear crunching, creaking, crackling, grating, or grinding sounds and sensation on movement of the joint that is called ‘crepitus’ due to roughening of the normally smooth cartilage surfaces inside the joint (62). As well as most of the OA patients experience tenderness with discomfort feeling when pressing on the affected joint even in the absence of obvious signs of inflammation. Joint line tenderness could be due to articular pathology, while tenderness away from joint is possibly due to periarticular pathology (55,62).

In later stages of OA, visible bone swelling can be noticed in the affected joint due to cartilage distraction and bone remodeling on either side of the joint, as well as the formation of small osteophytes at the joint margins. Also, swelling can be soft and warm due to extra synovial fluid that is called effusion resulting from the accumulation of excess fluid in the joint space (62).

Another sign of OA is bone growth or deformity, bone growth at advanced stages of OA can be seen under the skin near joints due to the articular cartilage loss, formation of osteophytes at the bone margins, and bone cysts at the affected joint (62).

OA does not affect all joints equally (62). Since knee OA patients may have a collection of joint fluid in the hollow at the back of the knee called a Baker's cyst. They sometimes walk with difficulty, and it can worsen as the joint degenerates. The pain of knee OA is often gets worse during physical activities such as walking, squatting, setting or getting off a chair and climbing stairs (68,69). In hip OA patients, pain develops slowly in the groin and on the outside of the hips, or sometimes in the buttocks. The pain may also radiate to the knee, making the diagnosis unclear (62). Hip OA patients often have a restricted range of motion particularly when trying to rotate the hip and walk with a limp, because they must slightly turn the affected leg to avoid the pain. The pain gets worse during physical activities such as walking, getting in and out of the car, and putting socks or stockings on (69). Hand OA occurs most often in older women and may be inherited within families. Hand OA causes the formation of hard bony enlargements (nodes) of the small joints of the fingers. The characteristic appearances of these finger nodes can be used in diagnosing hand OA (70). OA of the hand may cause enlargements of the distal interphalangeal (DIP) joints called Heberden's nodes, it may also cause enlargements of the proximal interphalangeal (PIP) joints called Bouchard's nodes (62). Moreover, it frequently damages the base of the thumb, which may give the hand a squared appearance (62,70).

1.1.4 Diagnosis of OA:

The diagnosis of OA is made clinically based on the typical disease symptoms, signs, and physical examination in the age of onset (> 40 years) group. However, when the clinical

diagnosis is unclear, radiography and/or laboratory investigations can be used to assess the presence and severity of the disease (71).

In clinical diagnosis of OA, peripheral OA joints may be diagnosed confidently with persistent joint pain at one or few joints along with morning stiffness for ≤30 minutes (72). Also, the presence of additional clinical signs of OA including joint swelling and redness, osteophytes formation, and the pattern of joint involvement are used in the symptomatic diagnosis of OA (73) as shown in **Table 1.1**.

This OA clinical diagnosis method is used regardless of the presence or absence of the radiographical structural changes and vice versa. For example, patients with a robust clinical diagnosis of OA may have normal radiographs (74,75). Although, there may be a significant correlation between radiographic features and OA symptoms in the affected joint, radiographic diagnosis fails to correspond directly with symptoms and disability, and implementation of radiographic criteria alone would not be accurate in clinical research studies of OA (75).

Radiographic examination may be used to support the diagnosis of OA but is not a routine procedure to explain the clinical symptoms. When there is still diagnostic uncertainty regarding the cause of the joint pain, imaging with plain radiographs, CT scan, MRI, or ultrasonography may be used (74).

Table 1.1: the clinical features of OA used in the diagnosis of the symptomatic OA. The table was adapted from the Abhishek and Doherty *et al.* (2013) (73).

Patient characteristics	
Age of onset	>40 years
Symptoms	
Pain	▪ Affects one or few joints at a time. ▪ Insidious onset - slow progression over years. ▪ Variable intensity. ▪ May be intermittent. ▪ Increased by joint use and relieved by rest. ▪ Night pain in severe OA.
Stiffness	▪ Short-lived (<30 minutes) and early morning- or inactivity-related.
Swelling	▪ Some (eg, nodal OA) patients present with swelling and/or deformity.
Constitutional symptoms	▪ Absent.
Physical exam findings	
Appearance	▪ Swelling (bony overgrowth ± fluid/synovial hypertrophy). ▪ Attitude. ▪ Deformity. ▪ Muscle wasting (global - all muscles acting over the joint).
Palpation	▪ Absence of warmth. ▪ Swelling (effusion if present is usually small and cool). ▪ Joint line tenderness. ▪ Periarticular tenderness (especially knee, hip).
Range of motion	▪ Crepitus (knee, thumb bases). ▪ Reduced range of movement. ▪ Weak local muscles.

Traditional radiography using X-ray is the most widely implemented imaging method in OA diagnosis that allows for detection of OA characteristic features including marginal osteophytes, subchondral sclerosis and cysts, and JSN as indicator of cartilage degradation (75). Also, it is used to evaluate patient's need for total joint replacement (TJR) surgery. However, X-ray is insensitive, particularly with early OA stage, and it often correlates poorly with the disease symptoms (76).

MRI is not necessary for most of the OA patients with symptoms suggestive of OA and/or typical radiographic features, however the utility of MRI is increasing due to its ability in identifying OA at earlier stages (74). Since MRI can detect cartilage damages and bone marrow lesions at the affected joint, and it can also be used to assess pathology in other structures of the joint including effusions, synovium, and ligaments that are not seen by traditional X-ray testing (75,76).

Although, ultrasonography became a useful tool in OA diagnosis to identify the OA-associated structural changes including synovial inflammation, effusion, and osteophytosis, ultrasound cannot be used to assess deeper articular structures and subchondral bone (77).

A common and widely used method of Radiographic classifying OA severity is the K/L scale that grades radiographs by the presence of osteophyte, joint space narrowing, sclerosis and joint deformity (78), **Figure 1.3**. The classification ranges from 0 to 4, where 0 is none and 4 is severe OA of the joint respectively (78,79) as shown in **Table 1.2**.

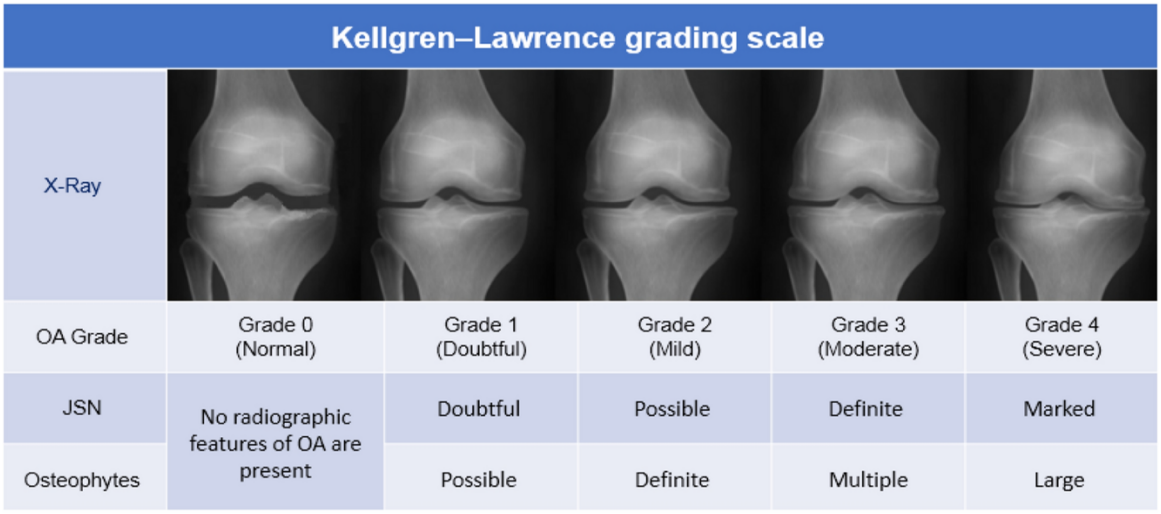


Figure 1.3: Knee X-ray images showing the grades of the Kellgren and Lawrence (K/L) radiographic grading system. **Grade 0** corresponds to the normal knee without radiographic findings of OA; **Grade 1** corresponds to minute osteophytes of doubtful clinical significance of OA; **Grade 2** corresponds to mild OA with definite osteophytes and possible JSN <3 mm (<25%); **Grade 3** corresponds to moderate OA with definite osteophytes and minor JSN <5 mm (50-75%); and **Grade 4** corresponds to severe OA with definite osteophytes with moderate JSN 5-15 mm (>75%). This image was adapted from Bany Muhammad *et al*. (2021) that is licensed under Creative Commons CC BY License (80).

Table 1.2: Kellgren and Lawrence (K/L) grading system. The table was adapted from the Abhishek and Doherty *et al*. (2013) (73).

Grade of the OA	Description
Grade 0	No radiographic findings of OA.
Grade I	Minute osteophytes of doubtful clinical significance.
Grade II	Definite osteophytes with possible JSN <3 mm (<25%).
Grade III	Definite osteophytes with minor JSN <5 mm (50-75%).
Grade IV	Definite osteophytes with moderate JSN 5-15 mm (>75%).
Grade V	Definite osteophytes with sever JSN >15 mm, subchondral sclerosis, and subchondral cysts.

OA: osteoarthritis, JSN: joint space narrowing, mm: millimeter.

Also, the ACR proposed different criteria for the diagnosis of knee, hip, and hand OA. For the OA of knee: the existence of knee pain in addition to at least three of the following six features: a) older than 50 years, b) stiffness lasting less than 30 minutes, c) crepitus on active motion of the knee, d) bony tenderness, e) bony enlargement, or f) no palpable warmth (15). Whereby, diagnosis of hip OA investigates: the existence of hip pain in addition to at least two of the following three features: a) erythrocyte sedimentation rate (ESR) less than 20 mm/hour, b) femoral and/or acetabular osteophytes evident on X-ray, or c) JSN shown on X-ray (71). Finally, detection of hand OA contains: the existence of hand pain in addition to at least three of the following four features: a) hard tissue enlargement of two or more of 10 selected joints (2nd and 3rd DIP, 3rd PIP, and first carpometacarpal (CMC) joints), b) deformity of at least two of the above ten selected joints, c) hard enlargement of two or more DIP joints, or d) fewer than three swollen metacarpophalangeal (MCP) joints (71) as shown in **Table 1.3**.

Table 1.3: The American College of Rheumatology (ACR) criteria for the diagnosis of knee, hip, and hand OA *(71)*.

Joint affected	symptoms	Additional features
Knee OA	Knee pain	At least three of the following six features: ▪ Greater than 50 years of age. ▪ Stiffness lasting less than 30 minutes. ▪ Crepitus on active motion of the knee. ▪ Bony tenderness. ▪ Bony enlargement. ▪ No palpable warmth.
Hip OA	Hip pain	At least two of the following three features: ▪ Erythrocyte sedimentation rate (ESR) less than 20 mm/hour. ▪ Femoral and/or acetabular osteophytes evident on X-ray. ▪ Joint space narrowing evident on X-ray.
Hand OA	Hand pain	At least three of the following six features: ▪ Hard tissue enlargement of two or more of 10 selected joints (2nd and 3rd DIP, 3rd PIP, first CMC joints). ▪ Deformity of at least two of the above ten selected joints. ▪ Hard enlargement of two or more DIP joints. ▪ Fewer than three swollen MCP joints.

However, laboratory testing is not being used as a routine clinical diagnostic procedure for OA (81). Synovial fluid examination can be examined in the lab to differentiate between OA and the rheumatoid arthritis (RA) (82), since synovial fluid is normally non-inflammatory or moderately inflammatory in the OA with 2000 white blood cells/mm^3, as well as the presence of the predominantly mononuclear cells (75,82). Further laboratory testing in OA diagnosis may include investigation of the ESR or C-reactive protein

(CRP), rheumatoid factor (RF), and anti-cyclic citrullinated peptide (CCP) antibodies as indicators of inflammatory (82).

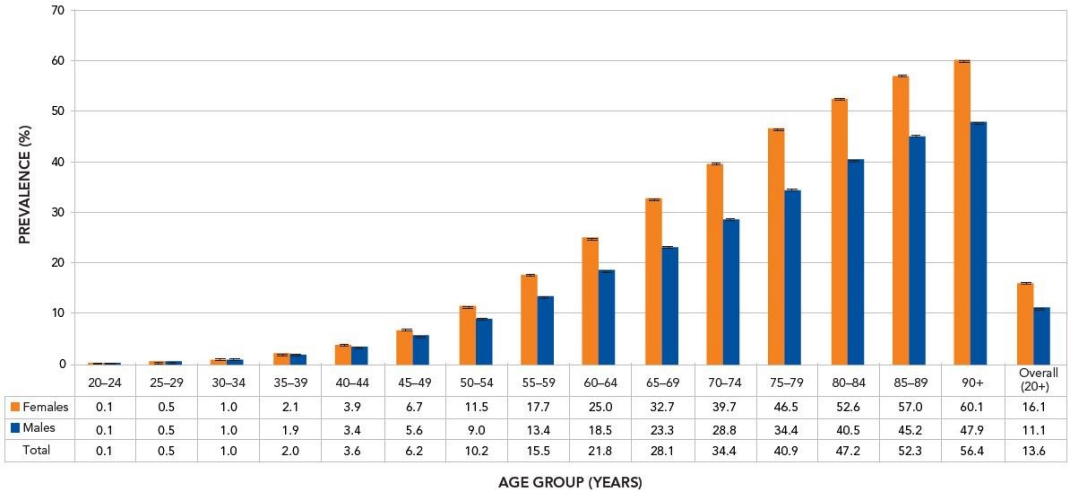
Also, it is worth mentioning that appropriate imaging and laboratory investigations of OA should be carried out in young individuals with atypical OA symptoms and signs such as unusual site of involvement; joint inflammation; and pain at rest during night-time, and rapidly progressive pain (72,75).

1.2 OA EPIDEMIOLOGY:

The difficulties in defining OA and determining its aetiology and onset led to complexity in estimating its epidemiology. Therefore, the prevalence and incidence of OA varies according to the OA definition, the affected joint, and the population of each study (83). Moreover, since OA is a chronic disease affecting the elderly, competing risk factors or death from other diseases complicates direct estimation of the cumulative incidence of OA (84). Nevertheless, studies consistently report a high global OA prevalence, with a greater burden in older peoples, females, specific ethnical groups, and among peoples with lower socioeconomic conditions (85). This burden is expected to rise with the aging population and at higher rates of obesity (4). The rapid increase in prevalence of OA will lead to a growing impact and major challenges for health care and public health systems (86). A report from the Global Burden of Diseases (GBD), Injuries and Risk Factors Study estimated the global age standardized point incidence of the symptomatic, radiographically confirmed OA in 2017 to be 3.8% (3754.2/100,000 person-years), with

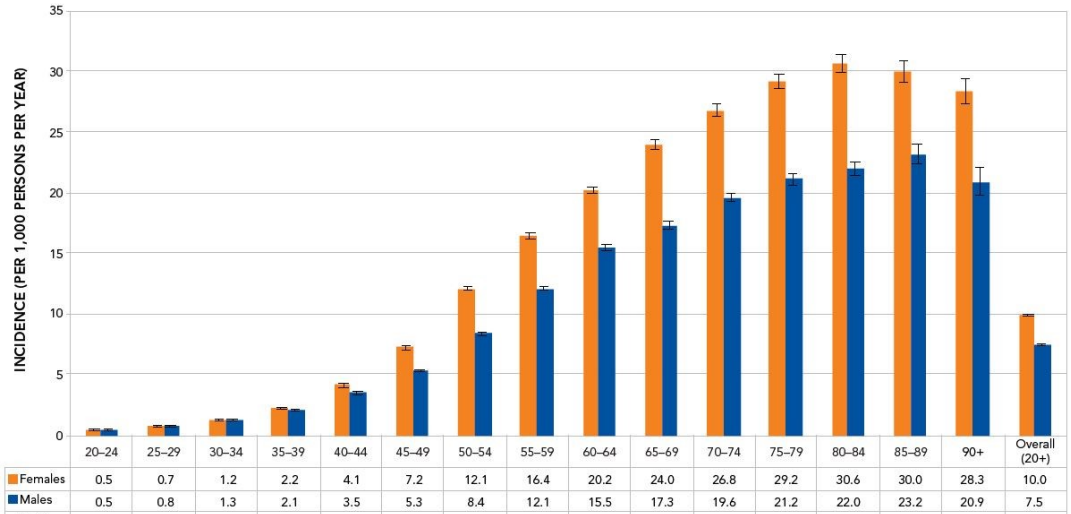
an increase of 9.3% from 1990. Rates being higher in women (4.8%) compared to men (2.8%) (85). Also, the global age standardized annual incidence rate was 181.2/100,000 person-years and 8.2% increases since 1990 (85). Also, Jin *et al.* (2020) used the GBD data and found that the global annual increase of OA was 0.32% (95% Confidence interval (95% CI)=0.28-0.36) in age standardised incidence rate (ASIR) with about 102% increase in crude incidence rate between 1990-2017 (87). The increase in the OA incidence could be due to the aging of the global population and the westernized lifestyle that are associated with increased age-related muscle weakness, lack of physical activities and exercises, increase of obesity, and comorbidities. Moreover, a UK-based study of 494,716 OA patients found that the incidence of symptomatic OA was 6.8/1000 person-years between 1997-2017 (88).

Limited number of studies investigated the prevalence and incidence of OA in Canada. Based on the data from the Canadian Chronic Disease Surveillance System (CCDSS) and the report on OA from the Public Health Agency of Canada (PHAC, 2019), the prevalence and incidence of OA increase with age and are higher in females than their estimates in males in Canada (89). The prevalence of OA was 16.1% in females and 11.1% in males, **Figure 1.4**. Also, the incidence was 10.0 per 1,000 persons-year in females and 7.5 per 1,000 persons-year in males (89), **Figure 1.5**. During the 2007–2008 and 2016–2017 period, the age-standardized prevalence of diagnosed OA increased by 0.7% in females and by 1% in males, while the age-standardized incidence of diagnosed OA decreased in females from 13.0 to 10.9 per 1,000 persons-year and in males from 9.2 to 8.1 per 1,000 persons-year as shown in **Figure 1.6** (89).



	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90+	Overall (20+)
Females	0.1	0.5	1.0	2.1	3.9	6.7	11.5	17.7	25.0	32.7	39.7	46.5	52.6	57.0	60.1	16.1
Males	0.1	0.5	1.0	1.9	3.4	5.6	9.0	13.4	18.5	23.3	28.8	34.4	40.5	45.2	47.9	11.1
Total	0.1	0.5	1.0	2.0	3.6	6.2	10.2	15.5	21.8	28.1	34.4	40.9	47.2	52.3	56.4	13.6

Figure 1.4: Prevalence (%) of diagnosed OA by sex and age group in Canada during 2016–2017. The figure was adapted from the Public Health Agency of Canada statistical report on OA.



	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90+	Overall (20+)
Females	0.5	0.7	1.2	2.2	4.1	7.2	12.1	16.4	20.2	24.0	26.8	29.2	30.6	30.0	28.3	10.0
Males	0.5	0.8	1.3	2.1	3.5	5.3	8.4	12.1	15.5	17.3	19.6	21.2	22.0	23.2	20.9	7.5
Total	0.5	0.8	1.3	2.1	3.8	6.3	10.2	14.2	17.8	20.5	23.0	25.0	26.3	26.9	25.7	8.7

Figure 1.5: Incidence of diagnosed OA (per 1,000 persons-year), by sex and age group in Canada during 2016–2017. The figure was adapted from the Public Health Agency of Canada statistical report on OA.

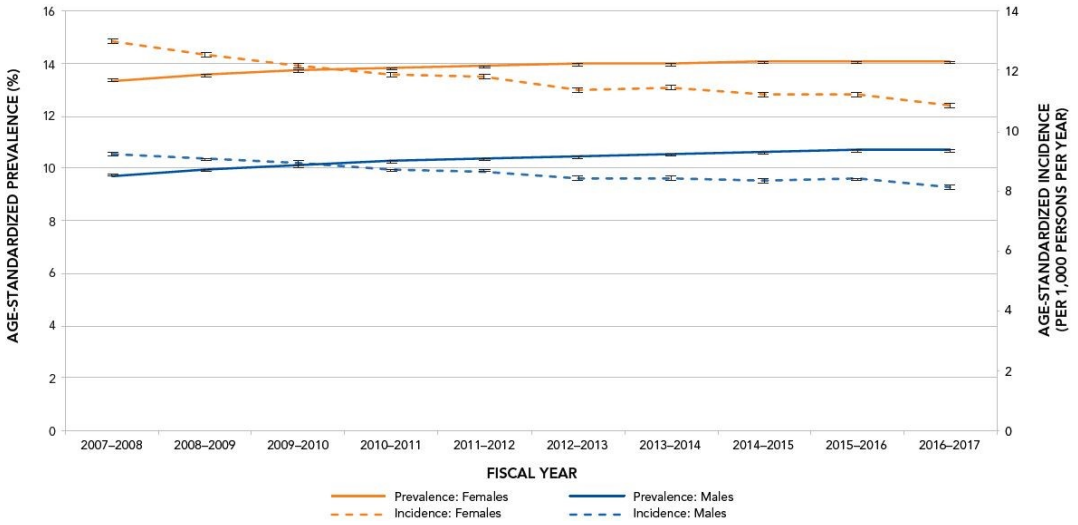


Figure 1.6: Age-standardized prevalence (%) and incidence (per 1,000 persons-year) of diagnosed OA among Canadians aged 20 years and older, by sex, Canada, from 2007–2008 to 2016–2017. The figure was adapted from the Public Health Agency of Canada statistical report on OA.

1.2.1 Site specific prevalence and incidence of OA:

Since knees, hips, and hands are the most common joints affected by OA, the majority of epidemiological studies have focused on OA of these sites. A population-based study of healthcare records from northeast Spain with more than three million OA patients reported that incidence rates per 1000 person-years was for knee OA 6.5 (95% CI=6.4-6.6), hip OA 2.1 (95% CI=2.0-2.1), and hand OA 2.4 (95% CI=2.4-2.4) (5).

1.2.1.1 Knee OA:

Several studies investigated the prevalence and incidence of knee OA. A study in Malmo, Sweden reported that the prevalence of radiographic knee OA was 25.4% (95% CI=24.1-26.1) among individuals aged from 56 to 84 years, whereas the prevalence of symptomatic knee OA was 15.4% (95% CI=14.2-16.7) in adults aged 56 to 84 years old

(90,91). In the US-based Johnston County OA project, the prevalence of the symptomatic knee OA was 17% among adults ≥ 45 years old (92,93). Another population study, the Framingham OA study in the USA identified the symptomatic knee OA prevalence to be 7% in the same age group. However, this study reported the age and body mass index (BMI) adjusted prevalence of knee pain and symptomatic knee OA, but not radiographic knee OA, approximately doubled among women and tripled among men in adults ≥ 70 years old between 1983-2005 (94). A meta- analysis based on the studies from China estimated that the symptomatic knee OA prevalence was 14.6% (95) in Chinese population aged 40-74 years. Moreover, the Chingford study reported that the cumulative 5-year incidence of typical radiographic knee OA in women aged 45-64 years was 17.6%, and the incidence of accelerated radiographic knee OA was 3.7% (96). Also, by using different cohorts and methodologies, the lifetime risk of symptomatic knee OA was estimated to be between 14% and 45% (97,98). Additionally, in the Korean National Health and Nutrition Examination Survey, the weighted prevalence of radiographic knee OA was 35.1% in adults ≥ 50 years old (99).

1.2.1.2 Hip OA:

Pertaining to the few epidemiological studies of hip OA, the US-based Johnston County OA study identified the prevalence of the radiographic hip OA to be 27% in patients ≥ 45 years old (83). However, among the same age group, this study reported the symptomatic hip OA to be 10% (83,100). In the Framingham Community Cohort, the age-standardized prevalence of radiographic hip OA among adults ≥ 50 years old was 19.6% (95%

CI=16.7-23.0), and the prevalence of symptomatic hip OA was 4.2% (95% CI=2.9, 6.1%) (101). Another study showed that the incidence rate of radiographic hip OA in adults ≥ 23 years was 5.6 per 1,000 person-years for men and 8.4 per 1,000 person-years in women (102). While the age- and sex-standardized incidence rate of symptomatic hip OA was identified to be 88 per 100,000 person-years in adults ≥ 50 years old (150).

1.2.1.3 Hand OA:

There was a high difference in the prevalence of hand OA due to differences between the radiographic and symptomatic status, variable disorder definitions, and study populations. The weighted prevalence of hand OA in a Spanish study of adults ≥ 40 years old that used the ACR clinical criteria was 7.7% (85). Another study of three cohorts of OA patients ≥ 50 years old from England reported that the weighted prevalence of radiographic and symptomatic hand OA was 22%, with first CMC joint OA being the most common subtype of the disorder (104). Also, the primary investigations of the Framingham OA study reported the age-standardized prevalence of symptomatic hand OA to be 14% in women and 7% in men (105); these measures were increased to 26% among women and 13% among men in patients ≥ 71 years old (106). The same study showed that the nine-years incidence of radiographic OA at any joint of the hand was about 35%, and the incidence of symptomatic OA at one or more of hand joints being 4% in men and 10% in women (105). Furthermore, in the US-based Johnston County OA Project, the incidence of radiographic hand OA among adults age ≥ 45 years old was 60%, and the incidence of symptomatic hand OA was 13% throughout a 12-year follow-up time (107). In the same

cohort, the weighted lifetime risk of symptomatic hand OA was 40% (108). Also, the age- and sex-standardized incidence rate of symptomatic hand OA was 100 per 100,000 person-years, and this value increases significantly after age of 50 in adults in a community health plan (103).

1.2.1.4 Multiple Joint OA (MJOA):

The presence of OA in multiple joints is an often-overlooked component of the burden of OA (109). Overall, MJOA has been associated with poorer OA-related outcomes compared with single joint involvement. Previous studies estimated that the prevalence of MJOA range from 5% to 25% (23). A population-based study from Korea suggested that almost 11% of men and 23% of women had more than two joints involved in the disease among individuals older than 50 years (110). In a Swedish population, almost 27% of prevalent OA cases report OA in multiple joints (111).

1.2.2 Economic and social burdens of OA:

OA is the most common form of arthritis and the eleventh most disabling disease in the developed countries (1). Worldwide estimates indicated that 240 million individuals have symptomatic OA, including 10% of men and 18% of women aged over 60 years (112). About 80% of OA patients have limitations in movement, and 25% of them cannot perform their major daily activities (113). In addition to pain and disability, OA patients may also experience low grade of health, psychological distress, and reduced quality of life. Although OA is not a life threatening, its social and economic burden is very

significant. The OA burden was reported as the 6th in the East Asia and high-income East Pacific countries, 10th in North America, 7th in Eastern Europe, and 13th in Western Europe (114). These burdens are elevated by significant co-morbidities, such as metabolic syndrome, diabetes, cardiovascular disease, and the adverse effects of non-steroidal anti-inflammatory drugs (NSAID) treatments (103). For instance, a population-based study of a large men and women sample with symptomatic and radiographically confirmed knee or hip OA identified more severe walking disability and greater the risk of death. This increased risk may be due to low-grade systemic inflammation or a lack of physical activity (115). Another study reported that the risk of cardiovascular events was associated with only greater baseline walking disability (116). Also, in a US population-based cohort study of 9,704 white women aged ≥ 65 years, about 25% of the increased risk of cardiovascular disease mortality among women with hip OA could be explained by poor physical function (117).

Additionally, costs associated with OA may include costs for outpatient visits and hospitalization, adaptive aids and devices, medications, and time lost from employment. In general, the annual direct costs of OA were estimated to be US $10,124 for male and US $12,534 for female (85). Particularly, the costs occurring during outpatient visits and hospitalization were accounting for 40–93% of the total costs. Also, the annual medication costs were consistently estimated between US $1,062- $3,624 per individual (95,97). Moreover, the number of days absent from work due to OA varied from 3.3 to 40.6 days with annual indirect costs ranging US $2,298-5,715 (118,119).

In Canada, about 3.9 million (13.6%) Canadians aged ≥ 20 years live with OA, and 219,000 (8.7 per 1,000 persons-year) were newly diagnosed with OA in 2016–2017 (89). Of these, 12% experience pain in their hip(s); 29% in their knee(s), and 29% in both joints. OA is estimated to be the major source of joint pain and disability in Canada (66,120). Sharif *et al.* (2015) predicted that the burden of OA in Canada will increase from approximately 2.9 billion Canadian Dollar in 2010 to almost 7.6 billion Canadian Dollar in 2031, a 2.6-fold increase from 2011. In which hospitalization will account for 38.2% of the total costs (121). Thus, OA has an extremely high burden in terms of health and economics on individuals, communities, and healthcare systems; This burden is largely attributable to the effects of disability, comorbid disease, and the expenses of OA treatment (122).

1.2.3 OA aetiologies and risk factors:

It became obvious that OA is a heterogeneous disorder, and its pathophysiology is driven by many systemic, intrinsic, and existing risk factors (123). Interactions between two or more of these factors may increase the susceptibility of OA development and progression (4,124).

1.2.3.1 Age:

Age is the strongest, non-modifiable risk factor for OA. The incidence and prevalence of OA are highly associated with age (125). Approximately, 14% of adults aged 25 and older have symptomatic OA of at least one joint, and 30-50% of adults ≥ 65 years old

develop OA (18-20). Then, leveling-off occurs at all joint sites around the age of 80 years (126). However, many older individuals at advanced age live without OA signs, and some young individuals could be diagnosed with OA at early stages of their life (127). The relationship between the age and the risk of OA is multifactorial that can be consequence of physiological processes of aging including oxidative damage, thinning of cartilage, loss of normal bone structure, increased stiffness of ligaments and tendons, meniscal degeneration, muscle weakening (sarcopenia), and a reduction in proprioception (128). These factors may affect the joint function and predispose the joint to OA. Moreover, basic cellular mechanisms that maintain tissue homeostasis decline with aging. For example, disruption of the catabolic and anabolic processes that are linked to aging lead to loss of the cartilage plate, decline in the chondrocytes number, and reduction of ECM water equilibrium (114). As a result, the risk of joint deterioration increases, and the chance of OA initiation and development elevates. Thus, OA could be a good marker for long life expectancy, people with OA could have other comorbidities and thereby die earlier than those without OA, and the reduction of OA incident rates around the age of 80 years is possibly due to withdrawal from studies or death of OA patients at older ages.

1.2.3.2 Sex:

Sex is another key risk factor for OA. Generally, there is no significant difference in the OA prevalence and incidence between males and females before the age of 50. For individuals older than 50 years, women are more often affected with hand, foot, and knee OA. A meta-analysis of sex differences prevalence, incidence, and severity of OA

reported that women older than 50 years are at a higher risk of severe knee OA (129), and these measures elevate dramatically after the menopause age (126,130). The Framingham Knee OA study reported a 1.7 (95% CI=1.5-2.7) times higher incidence of knee OA in women aged 40-80 years compared with men. Women were estimated to have a lifetime increased risk of symptomatic knee OA of 47% (131). In addition, females live with OA tend to have more severe symptoms, faster progression, and more prone to undergo TJR surgery and use of healthcare than men (132). Multiple studies suggested an implication of sex hormones and menopause in the development of OA in women, but the results of clinical and epidemiologic studies are conflicting (133−135). However, a recent study reported that there was no clear association between sex hormones and radiographic knee, hip, or hand OA in women (135). Although, the reason for the increased risk of OA in women remains unclear, it was hypothesized that other factors including reduced volume of cartilage, bone loss, or muscle strength difference between both sexes may lead to this difference in OA prevalence and incidence (90).

1.2.3.3 Obesity:

Individuals with BMI > 30 kg/m^2 are categorized as obese. Obesity represents one of the most important modifiable risk factors for the incidence and progression of OA. It has been associated with early articular cartilage damage identified by MRI before the development of OA symptoms, suggesting a causation between obesity and OA, rather than obesity being developed due to lifestyle changes associated with OA (136).

The effect of obesity on the development and progression of OA is possibly due to increased overloading during weight bearing activities, decreased muscle strength, and altered biomechanics on the knee and hip joints (137,138). Studies showed obese individuals were 6.8 times more likely to develop knee OA than normal-weight controls (139). A study in patients with new-onset knee pain identified that 24.6% of cases were overweight or obese individuals (124). Also, a large population-based study found that overweight individuals had a 46%, and obese subjects had a 93% increased risk of developing hip OA (140). Furthermore, studies of BMI and risk of OA showed that a five-unit increase in BMI was associated with a 35% raised risk of knee OA and 11% increased risk of hip OA (141,142).

As a modifiable risk factor, weight loss have been demonstrated in many studies in the relief of pain, delay of joint structural damage, restore joint's function, and decrease inflammatory associated with OA (143). The Framingham Study estimated that weight reduction by 5 kgs decreased the risk of developing knee OA by 50% (144). Likewise, a reduction in weight by 7.7% was reported to clinically improve joint's function scores (145). Also, a multicenter observational study in 640 adults who were overweight or obese with mild to moderate OA identified that a decrease in BMI ≥ 5% over a 48-month period is strongly correlated with a decrease in rate of progression of knee cartilage and meniscal damage seen on MRI (146).

Interestingly, obesity has been also associated with the risk of developing hand OA (147).

As hands are non-weight-bearing joints, and the association between body weight and hand OA is inconsistent and weaker than its estimates in knee and hip OA (114). In this case, the effect of obesity may be due to not only just biomechanical, but also metabolic diseases or inflammatory processes (148). Previous studies reported that metabolic factors linked to obesity like circulating adipocytokines, adipokines, and lipid abnormalities may impact the joints of obese individuals (149). For example, adipokines has been identified to stimulate similar chondrocyte activation in a similar manner to the mechanical stress and pro-inflammatory cytokines (150). Likewise, excess leptin generated due to adiposity may lead to the ECM synthesis reduction and degradation of cartilage (137). It has been also shown that white adipose tissue (WAT) secretes inflammatory mediators into the systemic circulation and effects cartilage degeneration (147). Additionally, bone marrow of the cancellous bone at the femoral head of end-stage hip OA patients had higher fat mass (151). Thus, obesity and adiposity have a serious effect on the hip bone and cartilage. Furthermore, obesity as a low-grade inflammatory state may significantly impacts joint's cartilage, synovium, and bone (152).

1.2.3.4 Metabolic syndrome:

In addition to obesity, other components of metabolic syndrome (MetS) including high cholesterol, hypertension, cardiovascular disease, and diabetes may contribute to OA pathophysiology (153). The association between these MetS components and OA is possibly due to the shared pathogenic mechanism involving metabolic abnormalities and systemic inflammation. For instance, the Korean National Health and Nutrition

Examination Survey III study reported an increased risk of metabolic syndrome by 5.26-folds in OA patients at the age of 43.8 years than older individuals (154). Eckstein F, *et al.* (2013) found that patients with severe spine OA had statistically higher composition of MetS factors than patients with early spine OA (155).

Moreover, hyperlipidemia was shown to possibly affect chondrogenesis, osteogenesis, and mesenchymatous cell differentiation (156). For example, increased circulation levels of cholesterol are implicated in the hypertension. In turn, hypertension was hypothesized as an important driver for the initiation and progression of OA (157). It has been identified that venous blocking due to hypertension may lead to episodic reduction in blood flow through the micro vessels in the subchondral bone. Therefore, the interstitial fluid flow within subchondral bone is reduced leading to subchondral bone ischemia, nutrient, and oxygen supplies decline (154). Thus, apoptosis of the subchondral osteocytes with subsequent abnormalities in subchondral bone remodeling occurs and results in OA development (158). Additionally, individuals with OA are at higher risk of physical inactivity and the use of the NSAIDs that increases the risk of cardiovascular disease which may further explain the association between these two conditions (154). However, the link between diabetes and OA is not fully understood, the association between these two conditions has been traditionally attributed to underlying shared risk factors of age and obesity (159). A meta-analysis study of 28 cross-sectional, 11 cohort, and 28 case-control studies found that the mean OA prevalence was 29.5±1.2% in 5,788 diabetes patients from 17 studies, and the prevalence of diabetes was 14.4 ±0.1% in 645,089 OA patients from 31 studies. Moreover, OA and diabetes were significantly

associated, the overall risk of OA in diabetic patients studies found the risk of OA to be greater in the diabetes patients than individuals without diabetes with OR=1.46 (95% CI=1.08 to 1.96), and the risk of diabetes in OA patients' studies reported the risk of diabetes to be higher in the OA patients than individuals without OA diagnosis with OR=1.41 (95% CI=1.21 to 1.65) (160). It has been proposed that alterations in lipid metabolism and excessive glucose concentration (hyperglycemia) might have a direct impact on cartilage health and subchondral bone. Emerging evidence suggests that high intracellular glucose concentration in diabetic patients promotes the production of reactive oxygen species and advanced glycation end products (AGEs) (160,161). In turn, accumulation of the AGEs compounds in articular cartilage give rise to a cascade of events that promote release of pro-inflammatory factors, such as TNF-α and activate transcription factors including nuclear factor-kappa B (Nf-kB), which cause inflammation and oxidative stress intracellularly and might promote cartilage degradation leading to OA pathogenesis (161).

1.2.3.5 Osteoporosis:

Osteoporosis (OP) is another common age-related skeletal condition. It develops when bone mineral density (BMD) declines leading to a decreased bone strength and elevated risk of fractures. OP is one of the major causes of fractures in postmenopausal women and in older men. Fractures can occur in any bone but happen mostly in bones of the hip, vertebrae in the spine, and wrist (162). The association between OA and OP is more complicated than what previously thought. The inverse association between OA and OP is not well understood and has been a subject of argument, these two diseases are not

mutually exclusive (162,163). It has been found that increased BMD is often associated with elevated risk of radiographic OA and increased number of affected joints (164). For example, a systematic review and meta-analysis of the knee OA risk factors in older adults identified a significant association between increased BMD and the onset of knee OA in women (165). Also, subjects with high BMD were found to have a higher prevalence of self-reported joint replacement and use of NSAIDs compared with unaffected controls, suggesting an increased risk of OA (164). However, the pathophysiological mechanisms of this association between OA and OP are not fully known, it is estimated that individuals with high BMD have a tendency for bone-forming, which elevates their susceptibility to OA. One study showed that individuals with high BMD are at higher susceptibly of hip OA and formation of multiple osteophytes (166). Furthermore, the studies of high BMD and OA in non-weightbearing joints in hands referred this association to osteophyte formation rather than cartilage loss and JSN (167). Moreover, the inverse association of OA with OP can be explained by shared genetic components between OA and OP (168). Also, a small number of studies in end-stage knee OA women have found sclerotic bone with a bone marrow lesion (BML) suggesting implication of BMLs in the relationship between OA and OP (169). Additionally, studies reported involvement of muscle weakness in the pathogenesis of OA and OP. Thus, the inter-correlation of BMLs, muscle weakness, OA, and OP may represent different mechanism that explains the relationship of OA and OP (170).

1.2.3.6 Muscle weakness (Sarcopenia):

Skeletal muscles are essential components of the neuromuscular skeletal system that has an integral role in the biomechanical structure and function of the synovial joints, since they produce movement, absorb loading, and provide dynamic joint stability (4,171). Muscle weakness and atrophy, which is also called sarcopenia, is one of the earliest signs of aging and an important geriatric condition (172). It became obvious that the age-related reduction of muscle strength and OA share similar pathologic processes and clinical features including changes of the joint anatomy, joint pain, decreased function, joint instability, periarticular muscle weakness, and fatigue (173). However, the underlying mechanisms between age-related muscle weakness and the pathogenesis of OA are not fully understood, it has been estimated that muscle weakness and OA are intercorrelated, and muscle weakness may represent an important risk factor for OA in the weightbearing joints. Limited number of studies have been conducted to investigate the contribution of muscle strength to the pathogenesis of OA, and the majority of the reported studies found strong correlation between muscle weakness and initiation of OA (174–176). It has been reported that individuals with reduced skeletal muscle mass and strength are at higher risk of falls (177), fractures (178), and loss of function (179). On the other hand, OA patients have significant loss of muscle mass, function, and strength due to reduced physical activities (180–185). For example, men and women with symptomatic and radiographic knee OA have been identified to have weakness in quadriceps compared to those without OA. Also, a large cohort study of knee OA patients reported that quadriceps weakness in women, but not in men, was associated with increased risk for tibiofemoral and whole knee JSN over 30 months' follow-up (186). This quadriceps weakness could be due to

atrophy with disuse (187). There is evidence that exercise to strengthen the lower limb may improve OA symptoms (188). Moreover, findings from the OA Initiative found significantly lower isometric strength in symptomatic versus asymptomatic knees (189). In another study examined ~8,000 patients, lower knee extensor strength was associated with an increased risk of pain and functional impermeant (190).

1.2.3.7 Joint injury:

Joint injury is one of the strongest risk factors for the development of OA and is commonly called posttraumatic OA (114). It is obvious that meniscal and cruciate tears, fractures, and dislocations, can result in an increased risk of OA development and musculoskeletal symptoms (128). The knee is one of the most frequently injured joints. Damage of the anterior cruciate ligament (ACL) is implicated in deterioration of cartilage, subchondral bone, collateral ligaments and/or menisci observed in approximately 65–75% of ACL-injured knees (191,192). Studies found that early onset in 13% of knee OA cases after 10-15 years is due to tearing of the ACL (193). However, injury of hip is less frequent than knee injury, individuals with hip injury were reported to have a five-fold risk of incidence and progression of hip OA (194,195).

1.2.3.8 Occupation:

Heavy workload is one of the most common occupational risk factors for OA. However, the mechanism of the association between the occupation and OA is not fully understood. It is thought that excessive joint loading and repetitive damage of the joints over time are

implicated in the development of OA at several anatomic sites (128). Most studies of occupation and OA are limited to knee OA. Knee OA was more frequently observed in people working in construction, firefighting, agriculture, fisheries, forestry, and mining (196) that required prolonged squatting, knee bending, kneeling (197), regular stair climbing, crawling, whole-body vibration, and repetitive movements (198). For example, one British study identified that workers ≥ 55 years old who were exposed to more than 25 kgs heavy lifting and kneeling/squatting or climbing stairs are at more than a five-fold greater risk of developing knee OA (199). Another study of the occupational exposure duration revealed that men who worked for 11 to 30 years in building and construction work had a 3.7-fold greater risk of developing knee OA (200). Occupation was also linked to the susceptibility of hip OA. Heavy physical workload involving bending, twisting, and reaching has been shown to contribute to hip OA (124). It was estimated that participation in such occupations doubling the risk of hip OA in men and increasing by approximately 40% in women (201). Like knee OA, the number of years in the occupation was also contributing to the increased risk of hip OA. For instance, farming for up to nine years increasing the risk of hip OA by 4.5 times; while farming 10 or more years increased the risk 9.3 times (114,128).

Moreover, heavy physical workload is considered as an occupational risk factor for spine and neck OA (196). Also, due to frequent extreme positions or high load being placed on the cervical spine, significantly higher risk of cervical spondylosis was reported in meat carriers, dentists, and miners in comparing to office workers (202). Additionally, workers

whose jobs required repeated pincer grip and increased manual dexterity have increased the risk of radiographic hand OA, particularly in the DIP joint (90,203).

1.2.3.9 Sports and physical activities:

The relationship between physical activity and subsequent OA remains complex and controversial (204). It is unclear if the associations are due to the involvement of the sport activities itself, or consequences of injury associated with the sport activities participation (205). There are some evidences suggesting that highly repetitive, intense and high-impact physical activity may increase the risk of radiographic hip and knee OA development (90,204,206). For example, a meta-analysis identified an increased risk of OA in soccer and elite soccer players, but lower and not significant risk in runners (207). Another study reported that middle-aged runners that were followed up for a five-year period did not develop symptomatic or radiographic knee OA compared with age matched controls from the general population (90,114). Similarly, observational studies using the Framingham cohort data did not show any additional risk or benefit associated with incidence of radiographic knee OA and participation in walking or jogging, even in subjects with a BMI >30 kg/m^2 (207). Moreover, other studies estimated that in the absence of joint injury, moderate daily recreational sport activities, whatever is the type of sport, do not appear to increase the risk of developing clinical or radiographic hip or knee OA (208), but could reduce the risk of various age-related chronic conditions (209).

1.2.3.10 Joint Alignment:

Anatomic factors such as joint shape and alignment have been also associated with the development of knee and hip OA. For instance, the mechanical alignment of the knee modulates the distribution of load across the articular surfaces. About 65% of weight-bearing load of a normally aligned knee is transmitted through the medial compartment. Shift in a valgus or varus direction affects the load distribution and increases stress on joint structures including articular cartilage and leads to degenerative change in the joint (128) . In a prospective observational study of 230 knee OA patients, medial and lateral progression of knee OA was four-fold greater in patients with varus alignment and five-fold greater in those with valgus alignment, respectively (116). Also, a meta-analysis of radiographic OA confirmed that knee malalignment is an independent risk factor for progression of radiographic knee OA (210). Similarly, hip joint shape alterations such as femoroacetabular impingement (FAI) and severe developmental dysplasia of the hip (DDH) have been associated with high risk of early-onset hip OA, with FAI being the more prevalent of these malalignment (91).

1.2.3.11 Ethnicity and race role in OA:

Different ethnic groups were found to have different prevalence of OA and patterns of joint involvement. In general, the prevalence of OA is higher in the populations of the developed countries including the Caucasian population of Europe and America, as well as East Asians population in Japan compared to the less developed countries (211,212). While, the Chinese from Beijing were identified to have significantly lower radiographic

hip and hand OA in comparing with the white Americans from the Framingham Study. In contrast, Chinese women had a higher prevalence of radiographic and symptomatic knee OA than white Americans (197). From other point of view, Johnston County OA Project reported similar prevalence of the radiographic hip OA in African and white American women, but the African American men were more likely to have superior or medial JSN and lateral osteophytes, and they were less likely to have axial JSN than white men (213).

1.2.3.12 Diet:

It had been found that some diets have anti-inflammatory capabilities which can help reduce OA symptoms, while other foods may increase them. For instance, Omega-6 fatty acids are known as precursors of pro-inflammatory eicosanoids that accumulate in the OA joints at high concentrations (214) and regulate inflammation, leading to impaired chondrocyte structure and cartilage degradation (215). While total plasma omega-3 decreases production of pro-inflammatory eicosanoids, reactive oxygen and nitrogen species, and cytokines. Thus, omega-3 generates anti-inflammatory mediators (216) and have protective affect against cartilage damage (217). Western diet contains high amount of meat and vegetable oils, hence higher omega-6 ratio than omega-3 (218,219) that promote inflammatory, bone-marrow lesions and increase the risk of OA development (220).

Serum cholesterol has also been thought as a systemic OA risk factor (221,149). Accumulation of cellular cholesterol induces cytotoxicity (222), and

hypercholesterolaemia increases arachidonic acid formation and production of pro-inflammatory eicosanoids (223) in human OA cartilage (218). Dietary strategies and weight reduction are recommended to lower plasma cholesterol concentrations and protect against OA (224).

Reactive oxygen species and reactive nitrogen species may be implicated in the pathophysiology of OA, and therefore antioxidants might delay its onset and progression (225,226). The antioxidant vitamins including vitamins D, C and K have been found to protect against OA (227). Vitamin D has multiple roles in the musculoskeletal system, it is thought to regulate bone metabolism and calcium homeostasis (228). It is also believed to modulate inflammation and cytokine synthesis (229). Studies determined that individuals with lower levels of vitamin D could be at higher risk to develop OA (218), and those with moderate vitamin D deficiency are more likely to have pain at the hip and knee joints (218). A number of trial studies have shown that vitamin D supplementation has positive effects on muscle strength (230); Other studies found protective role of vitamin C against OA development, cartilage deterioration, and knee pain (218). Also, vitamin K that is a group of fat-soluble compounds, is involved in bone and cartilage mineralization (231), and insufficient vitamin K intake may lead to abnormalities in bone and cartilage that are seen in the OA patients (218).

1.2.3.13 Smoking:

The association between smoking and OA is somewhat unclear and there have been conflicting reports on the role of smoking OA (128). Some studies have found a protective association between smoking and OA, but others in contrast, reported that smoking may be associated with a greater risk of cartilage loss and knee pain in OA (128). As an example, a large meta-analysis of 38 independent observational studies including 481,744 participants reported an inverse association between smoking and the risk of OA in the knee joint (232). A significantly lower risk of knee OA development was identified in those who had ever smoked in comparing to those who had never smoked. The decreased risk for knee OA was more obvious in male smokers than female smokers. Furthermore, a dose-response analysis showed a linear decrease in knee OA with increased number of smoked cigarettes (232). A meta-analysis of 48 observational studies confirmed a negative association between smoking and OA; however, when analyses were performed in hospital-based case-control studies, no association was observed among the cohort or cross-sectional studies (233). This false protective effect of smoking was likely related to selection bias, as many studies have been conducted in a hospital setting where control subjects have smoking-related conditions (234). Additionally, the rate of smoking may be higher in study participants from hospital settings than in the community and general population (234).

1.2.3.14 Genetic architecture of OA:

OA is a multifactorial condition that is caused by interplay between many environmental and genetic factors. Although, the involvement of genetic factors in the development and

progression of OA is complex as associations can vary depending on factors such as the joint involved, history of joint injury, and gender; OA in all of its forms appears to be strongly genetically determined (129). Identification of genes involved in OA risk and progression enables us to detect individuals at high risk of OA, allows us to better monitor disease progression, improves our understanding of OA molecular pathogenesis, and represents novel targets for OA therapeutic intervention (235). Evidence of genetic implication in OA has been determined by the epidemiological studies of family history and family clustering, twin studies, linkage scan and candidate genes studies, and genome-wide association studies (GWAS) analysis. The heritable contribution to primary OA risk has been established in twin and sibling studies to range from 40% to 65% with genetic factors being stronger for hand and hip OA than knee OA (129). Due to the multifactorial inheritance of OA, no single gene is involved in the development of OA, and multiple genes could contribute to the disease onset that are transmitted in a non-Mendelian manner. It is likely that different genes contribute to the OA development and its pain (23). Moreover, it is possible that different genes play roles for specific sites rather than a generalized OA phenotype. Also, GWAS analysis found differences between different ethnic groups including a significant association of a signal in a region on chromosome 7q22 with OA in European-descent samples but not in Asian populations (129,236). Similarly, two single nucleotide polymorphisms (SNPs) on a region of human leukocyte antigens (HLA) locus on chromosome 6p were associated with knee OA in a Japanese population, but not in Han Chinese or European populations (129).

Most established OA-associated genetic variants are represented by common SNPs with minor allele frequency (MAF) >5% that have moderate to small effect sizes with odd ratio (OR) ≤1.3, while low frequency and rare variants with MAF <1%, epigenetic changes, structural variants and gene-environment interactions may contribute to the missing heritability of OA (237).

1.2.3.14.1 Familial aggregation studies:

Family studies investigate whether the disease or trait run in families. Where individuals with a family history of the disease will have a higher disease risk ratio compared with the disease prevalence in the general population (238). The familial aggregation of OA has been reported at the earliest descriptions of the disease. Charcot *et al.* (1881) reported in his clinical lectures on senile and chronic diseases that multiple members of a family may have Heberden's nodes of the fingers that may often appear with disease of the hip or knee, proposing that OA could be hereditable and runs in families (239). Similarly, Duckworth *et al.* (1890) detected the Heberden's nodes in a multi-generation family with a female predominance (240). Then, familial clustering of Heberden's nodes was first formally studied by Stecher *et al.* in the 1940s who found the condition twice as frequently in the mothers and three times as frequently in the sisters of affected women as in the general population. He postulated that the nodes were inherited as a Mendelian dominant in females and as a recessive in males (241). Subsequently, family clustering of hand and knee OA had been confirmed in the epidemiological studies. For instance the study conducted by Kellgren *et al.* (1963) in the United Kingdom (242) found that definite OA in five or more joints was found in 36% of the male relatives, compared with

an expected rate of 17%, and in 49% of the female relatives, compared with an expected rate of 26% (242). Also, when the relatives were divided into two groups according to the presence or absence of Heberden's nodes in the proband, definite multiple OA was equally prevalent in both groups, but there was a distinctly higher prevalence of severe multiple OA in the relatives of the nodal probands (242). Moreover, clinically-observed Heberden's nodes of moderate or severe grade were found in 45% of female relatives of "nodal" probands compared with an expected rate of 10%, and in 13% of female relatives of "non-nodal" probands compared with an expected rate of 10% (242). More recently large community-based studies in the USA including the Baltimore Longitudinal Study of Aging (152) and the Framingham Offspring Study (243) reported that there is a clustering of cases of hand, knee, and hip OA with significant increases in the disease risk for a relatives of OA patient within affected families. Specifically, first degree relatives to OA patients have 2-3 fold increase of the disease risk. Also, the sibling OA recurrence risk is 2.08-2.31 for radiographic knee OA and 2.27-5.07 for radiographic hip OA (235,244). Other multiple radiological case reports and case series of sciatica, cervical spondylosis, and herniated discs reported family clustering of OA (245–249) and estimated that sibling recurrence risk of radiographic knee OA is 2.08–2.31, and the risk of the total knee replacement (TKR) ranges from 2.8 to 4.8. Likewise, the risk of radiographic hip OA is 4.27–5.07, and the risk of the total hip replacement (THR) ranges from 1.78 to 8.5 (12). These estimates are relatively low due to the effect of environment and similar lifestyle choices of individuals within the same family (152).

However, the familial aggregation studies were successful to show the recurrence risk of OA, they failed to match age of probands and their relatives, since age matching is particularly important in OA studies (152). Moreover, family studies were ineffective in clearly quantify genetic involvement in OA, since they do not permit differentiation of clustering that is due to a shared environment in the same family (12,152). Furthermore, population data on hip and spine OA are limited, making it difficult to determine the expected rates of OA at these sites for comparative purposes. Hence, twin studies were conducted to compare identical to non-identical twins after adjustment for environmental and other confounders and identify the implication of genetic factors and heritability of the OA.

1.2.3.14.2 Twin studies:

Generally, twin studies compare the occurrence of trait or disease between monozygotic (MZ) and dizygotic (DZ) twins (250). This study design distinguishes between the effects of genetic factors and the shared family environment with matching for age (251). Since MZ twins inherit identical genetic composition, any variation between them can be due to environmental factors. On the other hand, the DZ twins share 50% of their genes on average, and any intrapair similarity may be attributed to a combination of both environmental and genetic factors (250,251). Thus, twin studies perform comparison of the disease occurrence between MZ and DZ twins and quantify genetic and environmental contributions to the disease and disease-related traits in a population (235,250).

In a twin study investigated the involvement of genetic and environmental factors in OA in 130 MZ and 120 DZ female twins aged 48 to 70 years from the St. Thomas' UK Adult Twin Registry, the effect of genetic factors in radiographic OA of the hand and knee was between 39% and 65% respectively, independent of known environmental or demographic confounding factors (252). Another study in a larger sample of twins, reported the heritability of the hip JSN at 60% (253). Additional study evaluated the extent of genetic influences on disc degeneration in 172 MZ and 154 DZ twins unselected for back pain or disc disease using MRI and reported overall degeneration score heritability of 73% at the cervical spine and 74% at the lumbar spine (152,254). Overall, twin studies suggested that the effect of genetic factors is approximately 40% for knee OA, 60% for hip OA, 65% for hand OA, and about 70% for spine OA, after controlling for known environmental and demographic confounding factors (152). These estimates suggested an overall heritability of OA to be approximately 50% (range 40%-65%) that is transmitted in a non-Mendelian mode of inheritance (235,254). Generally, twin studies have proved that genetic factors are implicated in the susceptibility of OA. However, as a complex disease, OA is influenced by multiple loci and each locus has a small effect (OR<1.2). Hence, for genetic mapping of OA, a very large sample size would be needed to discover any rare large-effect variant (255).

1.2.3.14.3 Genome-wide linkage scan:

Linkage analyses investigates a set of genetic markers across the whole genome and map a trait or disease to a genomic location by demonstrating co-segregation of the disease

with genetic markers of known chromosomal location throughout extended families with multiple affected individuals (256).

A number of genome-wide linkage analysis based on small families or twins of hand, hip, or knee OA affected relatives have been conducted and have identified multiple relatively large chromosomal regions on chromosomes (Chrs) 2, 4, 6, 7, 11, 16, 19, and X that may comprise OA susceptibility genes (244). For example, a meta-analysis of OA genome-wide linkage scans including 893 families with 3,000 affected individuals from Iceland, the United Kingdom, and the USA had concluded that chromosomal regions including 7q34–7q36.3, 11p12–11q13.4, 6p21.1–6q15, 2q31.1–2q34, and 15q21.3–15q26.1 were the most likely to harbor OA susceptibility genes (257). Further investigation of these genomic intervals using candidate gene or SNP association studies led to the discovery of genetic variants associated with OA in the *IL1* gene cluster in Chr 2q11-q13, *matrilin 3* (*MATN3*) gene in Chr 2p24.1, *IL4R* gene in Chr 16p12.1, the *secreted frizzled-related protein 3* (*FRZB*) gene in Chr 2q32.1, and the *bone morphogenetic protein 5* (*BMP5*) gene in Chr 6p12.1 (258).

However, a genetic linkage scan identified large linked chromosomal regions that may include the disease candidate genes, and they have been very successful in identifying rare variants involved in monogenic disorders that are inherited in a Mendelian fashion, they failed to detect specific disease genes, and they have largely failed in detecting the complex disease susceptibility genes (259). Limited ability of linkage studies to identify

genes involved in common complex diseases, such as OA could be due to the disease genetic heterogeneity, lack of power, inherent complexity, and poor phenotypic definition (244).

1.2.3.14.4 Candidate gene studies:

Candidate gene or gene-centric studies focus on a pre-specified set of markers, based on the known biological aetiology of a disease, using *a priori* hypothesis. These analyses are carried out by resequencing either the entire candidate gene or a set of SNPs selected in patients and controls from population-based cohorts and identifying variants that are associated with the phenotype of interest in the affected individuals compared with unaffected controls (260,261). Few candidate gene studies have been conducted based on the limited knowledge available of the OA aetiology (235,244). These studies investigated variants in the *SMAD family member 3* (*SMAD3*), *Asporin* (*ASPN*), *FRZB*, *Collagen type XI alpha 1 chain* (*COL11A1*), *Vascular Endothelial Growth Factor A* (*VEGF*), *Interluekin 1 receptor antagonist (IL1RN)*, and *growth differentiation factor5* (*GDF5*) genes that are already hypothesized to be involved in OA (235,262,263). A variant rs143383 (T>C) in the 5' untranslated region (UTR) of the *GDF5* gene was the most convincing and robust variant linked to the risk of OA (264–266). The *GDF5* gene was originally discovered in gene-centric studies and has been associated with OA in many studies in different populations (264,266,267). This gene is a bone morphogenetic family member that plays a role in joint development (265). It affects chondrogenesis and joint compartments formation during skeletal development, consistent with the protein

expression in the cartilage and joint inter zone (264,265). Mutations in this gene are thought to predispose to OA due to altered joint shape. Studies documented the important role of this gene in the development and progression of OA (264–273). The first candidate gene study which was conducted in 2007 reported that the allele T of rs143383 was associated with hip OA in two independent Japanese populations (266). The rs143383 T risk allele is associated with a reduction of the gene transcription in chondrogenic cells leading to reduced protein expression (266). Furthermore, this identified locus was found to be associated with OA in the European populations with a lower effect size (268). Also, a meta-analyses showed that the rs143383 risk allele was associated with knee OA and was found to have a stronger effect for knee and hip OA in the Asian populations (273). In a study of individuals from the Caucasians population in Europe, this variant was also found to associate with developmental dysplasia of the hip at the genome wide association level (270). Additional studies suggested association of rs143383 variant with severe radiographic OA (274,275). An epigenetic study of human chondrocytes showed that the *GDF5* gene expression is influenced by CpG methylation at rs143383 (276). Genetic variations in the *GDF5* gene were also found to associate with height (277,278). Animal model studies identified that the *GDF5* gene has a critical functional role in bone and joint disease in mice (279). This study has suggested that lower levels of the *GDF5* gene product in mice include mechanisms of altered loading and changes in subchondral bone, exhibiting developmental failure of the condyles and the articular ligaments (279–281).

Although, candidate gene studies were able to discover important OA genes, they have several limitations including dependency on assumptions, incomplete variant coverage, lack of power, phenotypic and genetic heterogeneity, population stratification, lack of replication, and exclusion of discovering novel genes influencing the OA (282,283).

1.2.3.14.5 GWAS analysis of OA:

Since OA is multifactorial condition, it is estimated that multiple common genetic variations may have small contribution in the disease risk. Based on the limited knowledge about the genetic aetiology of OA, GWAS analysis was the approach of choice that can be performed on a small or large scale to unlock the genetic bases of OA (284). Thus, GWAS have been suggested as the best approach when studying OA (284). GWAS is a hypothesis-free approach that investigates hundreds of thousands of genetic variants across the whole genome and identifies the association between these genetic variants and the disease locus based on the linkage disequilibrium (LD) (136,284). These large scale human genome studies compare the genotyped allele frequencies of genetic variants between cases and unaffected population controls (285). Based on each study research question, different genotyping microarrays with variable number of genotyped genetic variants (> a million) are available for the GWAS analysis. Then, pre-association quality control (QC) filtering is conducted to check the genotyping data quality of study samples and genetic variants. Subsequently, genome-wide imputation analysis can be performed to improve the coverage of the genetic variation by allowing the assessment of more SNPs from a greater allelic frequency spectrum and increase the power of the GWAS analysis (286,287). Variables such as sample age, sex, BMI, and relatedness

principal components can be taken into account while investigating the association (288). Then, due to the many false positive discoveries in the complex diseases reported by candidate gene studies, the conventional significance threshold of $p \leq 5*10^{-8}$ became widely used in GWAS analysis to account for multiple testing and limit false positive association findings (289). Replication of a discovery study's findings in an independent cohort is recommended to report robust associations between the genetic variations and phenotype of study. In the complex disease studies, thousands of SNPs have been robustly associated with disease risk at genome wide statistical significance (290). Most associations are at common loci and their effect sizes are small to moderate with typically OR<1.4 (291). In GWAS analysis, sample size and distinct phenotype definitions are important considerations to maximize study power and to avoid case misclassification, respectively (292,293). Also, to improve power of associations, meta-analysis of several GWAS analyses is performed by either replicating and meta-analyzing signals taken forward from the discovery study with $p<1*10^{-5}$ in the replication cohort using summary statistics, or by genome-wide meta-analysis between studies (294,295).

Multiple GWAS analyses have been conducted for OA. For instance, a study that was performed in 2008 for knee OA patients from a Japanese population investigated about 100,000 SNPs and identified a SNP rs7639618 to be associated with knee OA. This SNP is a missense variant in the *Double Von Willebrand Factor Type A domain* (*DVWA*) gene located on Chr 3p24.3. SNP rs7639618 was also replicated ($p= 7.3*10^{-11}$) in the Japanese and Han Chinese combined cohorts (296). The *DVWA* gene was reported to be highly expressed in cartilage compared to other human tissues suggesting an involvement of the

gene in metabolism of cartilage in humans. Also, the DVWA protein that consists of 276 amino acid binds to β-tubulin and modulates its chondrogenic function (297). Further analyses in Korea, UK, Netherlands, Spain, and Greece cohorts have failed to independently replicate this signal, suggesting ancestry-specific effects at this locus, maybe due to differences in study populations (298–300).

Another GWAS analysis in 4,800 Japanese participants with knee OA identified SNPs rs7775228 (OR=1.34, 95% CI=1.21-1.49; $p=2.43*10^{-8}$) and rs10947262 (OR=1.32, 95% CI=1.19- 1.46; $p=6.73*10^{-8}$) on Chr6 to be significantly associated with OA. These genetic variants locate in a region containing the Human Leukocyte Antigen (HLA) class II/III genes including the *HLA-DQB1* and the *Butyrophilin Like 2* (*BTNL2*). These results suggested an involvement of immunologic mechanisms defects in the development and progression of OA. These findings were also replicated in the European population (301), but further analyses were unsuccessful to replicate these findings in the Han-Chinese (302) and other European populations (303).

Moreover, a GWAS that included 1,341 cases and 3,496 controls from Dutch Caucasian population investigated the association of 500,510 genetic variants with knee and hand OA. This study identified rs3815148 in intron 12 of the *Component of Oligometric Golgi Complex 5* (*COG5*) gene on Chr7q22 to be associated with a 1.14-fold increased risk of knee and hand OA ($p=8*10^{-8}$) (304). Replication of the analysis in 14,938 cases and 39,000 controls identified variants in the *COG5* gene to be in complete LD with

rs3757713 variant that is adjacent to five genes containing the *Protein Kinase cAMP-dependent Regulatory Type II Beta* (*PRKAR2B*), *HMG-box Transcription Factor 1* (*DUS4L*), and *B-Cell Receptor-Associated Protein 29* (*BCAP29*) genes. Further, SNP rs3757713 was reported to associate with the *GPR22* gene expression levels in lymphoblast cell lines ($p=4* 10^{-12}$). Immunohistochemistry experiments identified that the *GPR22* gene expression was absent in the articular cartilage and synovial fluid of normal mice but present in instability-induced OA mice (304). Moreover, in a large meta-analysis and replication study, a SNP rs4730250 in the *DUS4L* gene was significantly associated with knee OA ($p=9.2*10^{-9}$). This finding indicates that any of these highly linked genes may contribute to the risk of developing knee OA (141). Validated expression of these genes in the joint environment was also confirmed by functional analysis and gene expression studies using cartilage tissues from OA cases and controls, suggested the implication of the *HBP1* gene in the pathogenesis of OA (305).

In a GWAS of 6,523 individuals, the G allele of rs12982744 on Chr19p13.3 was associated with a 5% larger hip endophenotype space width ($P= 4.8*10^{-10}$). The association was replicated in 4,442 individuals from the European population with an overall meta-analysis $p=1.1*10^{-11}$. SNP rs12982744 falls in the *Disruptor of telomeric silencing 1-like* (*DOT1L*) gene that is conserved histone methyltransferase. Lately, the *DOT1L* gene was estimated to play an enzymatic role for the Wingless/Integrated (Wnt) target-gene activation in leukemia. Functional analysis in mice identified a role for the *DOT1L* gene in chondrogenesis (306). The same locus was previously associated with height (277) and skeletal development (307). Another GWAS using a larger sample size

from the European origin individuals also identified an association of rs1292744 with hip OA in males (OR=1.17, 95% CI=1.11-1.23, $p=7.8*10^{-9}$) (308).

In 2014, a large GWAS meta-analysis of 11,277 cases of radiographic and symptomatic hip OA identified SNP rs6094710 (MAF=0.04) to be associated with hip OA with an OR of 1.28 (95% CI=1.18-1.39, $p=7.9*10^{-9}$) (309). The identified SNP is close to the *Nuclear Receptor Co-Activator 3* (*NCOA3*) gene on Chr20q13. Reduced expression of the *NCOA3* gene was detected in OA damaged articular cartilage compared to macroscopically non affected cartilage of the same joint. However, the role of the *NCOA3* gene in the OA pathogenesis is not fully understood, it has been observed to modulate hormonal regulation of bone turnover, such as thyroid hormones and steroids (310,311). Also, it is thought to be involved in chondrocyte mechano-transduction (312).

Furthermore, the Arthritis Research UK Osteoarthritis Genetics (arcOGEN) consortium GWAS study that included 3,177 hip and knee OA cases and 4,894 OA-free controls from all over the UK was conducted in a two-stages in 2011. Initial and replication analysis did not identify any loci to be associated with OA at the genome-wide significance level, which could be due to the small sample size, and heterogeneity of the disease (313). Subsequently, imputation analysis of the initial GWAS data was performed using the 1000 Genomes Project dataset (314) and meta-analysis was implemented. As a result, a SNP rs11842874 in the *protein encoding MCF.2 cell line derived transforming sequence-like* (*MCF2L)* gene on Chr13q34 was associated with knee OA at genome wide significance level (OR=1.17, 95% CI=1.11-1.23, $p=2.1*10^{-8}$) (315). The *MCF2L* gene

plays a role in cell motility of the nervous system, indicating that this gene affects nociception. Functional studies in zebrafish concluded that expression of the *MCF2L* plays a role in skeletal system development (316).

Then, the final arcOGEN GWAS analysis that contained 7,410 OA cases and 11,009 un-affected controls from the UK identified five novel loci to be associated with OA at the genome-wide significance level (317). The top SNPs were rs6976 in the 3' UTR of the *Glycosyltransferase 8 Domain Containing 1* (*GLT8D1*) gene, and a missense variant rs11177 in the *Guanine Nucleotide Binding Protein-Like 3* (*GNL3*) gene, both variants locate on Chr3p21.1 and in almost perfect LD with each other with OR of 1.12 (95% CI=1.08-1.16, and $p<1.3*10^{-10}$) (317). These two SNPs are associated with hip and knee OA in both sexes. Also, the association was stronger in subjects that underwent the TJR (317). Approximately 80% of the studied OA cases had undergone the THR and/or TKR, indicating a severe OA phenotype definition (317). Further significantly associated variants included rs4836732 that was identified to be associated with hip OA in women with the TJR (OR=1.20, 95% CI=1.13-1.27, $p=6.11*10^{-10}$) (317). SNP rs4836732 falls in intron one of the *Astrotactin 2* (*ASTN2*) gene on Chr9q33.1 that has been associated with neurological disorders (318) and migraine (319). Also, it is suggested to act in the regulation of the Astrotactin 1 (ASTN1) neuronal protein (320). Another SNP (rs835487) was associated with hip OA in the TJR subjects at the genome-wide significance level (OR of 1.13, 95% CI 1.09 to 1.18, $p=1.64*10^{-8}$) for both men and women. SNP rs835487 locates in intron two of the *Carbohydrate Sulfotransferase 11* (*CHST11*) gene on Chr12q23. The *CHST11* gene has a role in cartilage development (321) and is

differentially expressed in OA and normal cartilage (322). Also, SNP rs9350591 was significantly associated with hip OA (OR=1.18, 95% CI=1.12-1.25, $p=2.42*10^{-9}$) (317). SNP rs9350591 falls between the *filamin A interacting protein 1* (*FILIP1*) and the *sentrin specific peptidase 6* (*SENP6*) genes on chr6. However these two genes are not previously linked to OA, the *collagen type XII alpha 1* (*COL12A1*) gene resides nearby and is known to play a role in bone formation (323). The last significantly associated variant, rs10492367 was also established as a risk locus for hip OA (OR=1.14, 95% CI=1.09-1.20, $p=1.48*10^{-8}$) (317). This variant lies between the *Kelch Domain Containing 5* (*KLHDC5*) and the *Parathyroid Hormone-Like Hormone* (*PTHLH* also known as *PTHrP*) genes. Deletion of the gene encoding PTHrP has been reported to play a role in mice skeletal development (324,325). Three additional novel variants were identified to be associated with OA at borderline in the arcOGEN study including rs8044769 that locates in an intron of the *Fat Mass and Obesity Associated* (*FTO*) gene; the intronic variant rs12107036 in the *Tumor Protein p63* (*TP63*) gene; and the intergenic rs10948172 that lies between the *Suppressor of Ty3 Homolog* (*SUPT3H*) and the *CDC5 Cell Division Cycle 5-Like* (*CDC5L*) genes (317). *FTO* gene is an established obesity risk locus (326,327). The phenotypic intercorrelation between OA and obesity has been genetically verified, and the role of the *FTO* gene in OA has been confirmed to be modified through obesity (328). The functional role of the *TP63* gene in the pathogenesis of OA is unclear, although it is reported to be involved in facial shape development (329). Also. the functional roles of the *SUPT3H* and *CDC5L* genes remain undefined in OA, while an adjacent *Runt Related Transcription Factor 2* (*RUNX2*) gene in extended LD with these genes is thought to play a regulatory role in bone development (330).

Since 2018, four GWAS meta-analysis of OA were performed and identified more than 100 genetic variations to be associated with different forms of OA. Firstly, A GWAS meta-analysis of OA that was conducted by Zengini *et al.* (2018) in about 30,727 OA cases and 297,191 OA-free controls from the UK Biobank dataset (331). This study identified three genetic variants to be associated with hip OA including rs11780978 in the *plectin* (*PLEC*) gene on Chr8, rs2521349 in the *Mitogen-Activated Protein Kinase Kinase 6* (*MAP2K6*) gene on Chr17, and rs6516886 falls 1 kb upstream of the *RWD-domain-containing 2B* (*RWDD2B*) gene on Chr21 (331). Also, three variants were associated with knee OA containing rs11335718 in an intron of the *annexin A3* (*ANXA3*) gene on Chr4, rs116882138 falls between the *MOB Kinase Activator 3B* (*MOB3B*) and the *equatorin sperm-acrosome-associated* (EQTN) genes on Chr9, and rs375575359 in intron three of the *zinc-finger-protein 345* (*ZNF345*) gene on Chr19 (331). These genes act in multiple pathways including the bone morphogenetic proteins and Wnt-beta catenin-signaling pathways which play a central role in the joint development and bone differentiation (265,332). Thus, if OA is initiated from bone, these pathways may be implicated in the OA development and progression. Moreover, this GWAS analysis identified two variants to be associated with OA at any joint. These genetic variants include SNP rs2820436 that resides within a region including multiple metabolic- and anthropometric-trait-associated variants, it is also adjacent to the *long-noncoding-RNA* gene RP11-392O17.1 and the *zinc-finger CCCH-type containing 11B pseudogene* (*ZC3H11B*) genes on Chr1. Also, rs3771501 that is in intron three of the *transforming growth factor alpha* (*TGFA*) gene on

Chr2. The *TGFA* regulates the conversion of cartilage to bone during the process of endochondral bone growth (333).

Another GWAS meta-analysis of OA was performed by Styrkarsdottir *et al.* (2018) in 41,028 (17,151 hip, and 23,877 knee OA patients) and more than 562,000 controls from the UK Biobank and Iceland (386). This study identified 23 variants at 22 loci to be significantly associated with OA and are located in or adjacent to the *COL11A1*; *Histone Deacetylase* 9 (*HDAC9*); *Smoothened, frizzled class receptor* (*SMO*); *tenascin C* (*TNC*); *LIM homeobox transcription factor 1 beta* (*LMX1B*); *Latent transforming growth factor beta binding protein 1* (*LTBP1*); *Latent transforming growth factor beta binding protein 3* (*LTBP3*); *Filamin-Interacting Protein, Refilin A* (*FAM101A*)*; IL11*; *Inter-alpha-trypsin inhibitor heavy chain 1* (*ITIH1*); *FILIP1*; *RUNX2*; *ASTN2*; *SMAD3*; *Homeostatic iron regulator* (*HFE*); *Chondroadherin like* (*CHADL*); *Strawberry notch homolog 1* (*SBNO1*); *WW domain containing E3 ubiquitin protein ligase 2* (*WWP2*); and the *GDF5* genes (386). The genes detected by this study are linked to formation of collagen, chondrocytes proliferation, and anabolic/catabolic processes of the ECM (387,388); osteoblast development and proper bone formation; OA pain (372); transforming growth factor beta (TGF-β)/SMAD3 pathway that is involved in the OA risk, cartilage maintenance and repair; and chondrogenesis and joint compartments formation during skeletal development (311,312,389).

Moreover, one of the largest GWAS meta-analysis of OA was conducted by Tachmazidou *et al.* (2019) and tested the genome-wide association between genetic

variants and four OA phenotypes including knee, hip, knee and/or hip, and any OA in 77,052 cases and 378,169 unaffected controls from the UK Biobank dataset (196). This study detected 64 signals, 52 of which are novel. This study identified putative effector genes including the *TGF-β1* , *fibroblast growth factor 18* (*FGF18*), *cathepsin K* (*CTSK*), and *IL11* that underly monogenic forms of bone development diseases, the collagen formation, and ECM organization biological pathways (196,334).

The most recent and largest OA GWAS meta-analysis was performed by Boer *et al.* (2021) and investigated the association of the whole genome genetic variations with 11 OA phenotypes in 177,517 OA patients and 649,173 OA-free controls from the UK Biobank dataset (334). This study identified ~100 independently associated risk variants with different forms of OA, 52 of them have not been associated with the disease before as shown in the table in **Appendix B**. Interestingly, this study identified high-confidence effector genes and provided evidence for genetic correlation with phenotypes related to pain, the main disease symptoms, and identified likely causal genes linked to neuronal processes (335). Furthermore, the study findings provided insights into key molecular players in the disease processes and highlighted attractive drug targets to accelerate translation (334).

1.3 OA MANAGEMENT:

While there is no known way to prevent OA, controlling the OA modifiable risk factors including weight loss in overweight OA patients, preventing age-related muscle weakness

and wasting, avoiding joint injuries, managing comorbidities, and eating healthy food minimise the risk of OA and slow its progression (336).

As OA progresses toward advanced stages, it is unstoppable and irreversible. Regrettably, OA does not have an available cure thus far that has been shown to reverse, slow down or stop its progression (337,338). Hence, the management of OA focuses on managing its pain, disability, and delay its progression (339). Mild to moderate OA symptoms are usually well driven by a combination of non-pharmacologic and pharmacologic treatments. When These treatments do not relief OA pain and disability at severe OA stage, doctors may recommend surgical treatments for OA patients (340).

1.3.1 Non-pharmacologic interventions:

Physical, occupational, exercise therapy, and weight management practices are the most common non-pharmacological treatments of OA (341). Personalized physiotherapy and exercise programs such as swimming, water aerobics, and low-impact muscle strength training are widely recommended by clinical guidelines. Multiple studies reported an inverse correlation between these physical activity regimes and the amount of pain and disability that OA patients experience (342,343). In contrary, extreme exercise programs were found to elevate the OA symptoms and potentially accelerate the progression of the disease. Since obesity is the most important modifiable risk factor for OA (12,344), weight loss with physical activities and exercise programs in overweight OA patients, especially in women, has been identified to reduce stress and the amount of pain in weight-bearing joints, as well as reduce the inflammatory processes that contribute to OA

(345). A meta-analysis reported that a 5% decrease in weight over a 20-week period is beneficial for knee OA (144). furthermore, a clinical trial study revealed that 10% weight loss due to diet and exercise may improve OA symptoms by about 50% (346).

Additionally, medical devices are recommended by healthcare providers to reduce pain and improve function of the affected joints. For instance, knee braces, splints, custom foot orthotics and inserts, canes, or walkers can help to decrease pressure, diminish pain, and improve function of the OA joints (347–349). Similarly, intermittent heating pads, hot packs, hot showers, and cold packs may be beneficial to relief joint pain and stiffness (350).

1.3.2 Pharmacologic treatment:

Unlike other forms of arthritis, progress on DMOADs has been much slower, and there are no medications yet to reverse or slowdown the disease progression. Hence, medications used in OA are focused on managing the OA symptoms. Non-steroidal anti-inflammatory drugs (NSAIDs) are considered as the first-line drug treatment for OA pain (351,352). Previous studies have reported that NSAIDs are favorable for rest and general OA pain (353). However heavy use of NSAIDs was associated with increased risk for cardiovascular problems (354), and gastrointestinal bleeding (355). Due to the side effects of NSAIDs, four grams of acetaminophen per day has been suggested to reduce mild to moderate pain of OA. Acetaminophen medications has a $C_8H_9NO_2$ chemical structure that is called acetaminophen or Tylenol in the USA and Japan, while it is called paracetamol

or Panadol in Europe and most of the world areas. A meta-analysis found low-level effects of acetaminophen for pain management in OA (341,356), and a randomised controlled trial found four grams of acetaminophen per day was preferable over placebo for knee OA (356). However, an increased risk of gastrointestinal, cardiovascular complications, and multi-organ failure were reported with supratherapeutic doses of acetaminophen drugs (354,355). Thus, due to safety concerns pertaining to the use of acetaminophen, increased awareness of their negligible and non-clinically significant effects on pain, these medications are no longer considered the first-line analgesic for the treatment of knee and hip OA by clinical guidelines, especially for patients with comorbidities (341,357).

Moreover, topical drugs including NSAIDs, and capsaicin were suggested to manage OA pain and improve joint function, however these interventions have local adverse effects like rash, burning and itching. Topical NSAIDs, such as diclofenac, are valuable analgesic for knee and hand OA as local drug delivery that reduce gastrointestinal adverse reactions of oral NSAIDs (358,359). Also, efficacy of the topical NSAIDs was found to be greater than placebo and oral NSAIDs (359). Since most of the previous studies focused on knee OA patients, the benefits of topical NSAIDs on patients with multiple-joint OA remain uncertain. Despite this, topical NSAIDs are increasingly being considered as a first-line pharmacological option, especially in patients with an increased risk of adverse events. Similarly, topical capsaicin can be used as an alternative or as an adjunct to standard drug treatment. Reviews of randomised controlled trials found that

topical capsaicin is beneficial for knee OA that reduces pain by 50%, however withdrawal of the study was higher in capsaicin than placebo because of the adverse event (360,361).

Additional medications used to manage OA symptoms include intra-articular corticosteroid injections. These injections have shown to provide short-term (up to three months) pain relief and improve joint motion for OA patients having acute exacerbations with joint effusions and local inflammation. Furthermore, intra-articular injections of hyaluronic acid have been reported to diminish OA symptoms that extend beyond eight weeks (362). The drawback of the intra-articular injections is that they are given more frequently than once every four months, which may lead to cartilage and joint deterioration (363,364), as well as increased risk of infection.

Opioids including codeine and propoxyphene are alternative drugs for short-term use in patients with severe OA and comorbidities (365). Overall, oral and transdermal opioids were more effective compared to placebo in relieving pain and improving function in knee and hip OA patients (365,366). However, their benefits were small to moderate along with possible adverse events including fractures, cardiovascular complications, and all-cause mortality, as well as nausea, dizziness, drowsiness caused many patients to withdraw from the studies, especially in the older adults (366). Also, the effectiveness of opioids in the long term is limited (365).

Alternative treatment for OA includes glucosamine and chondroitin (367,368). Studies concluded that glucosamine may possibly reduce or slow down the progression of

cartilage loss and other joint's structural changes (369). Similarly, chondroitin has been reported to play a role in reducing the rate of the JSN (370). Also, studies found that taking supplements of glucosamine and chondroitin for two years had a statistically significant reduction in the JSN compared to placebo. However, no statistical difference was found with individual treatment alone (371). Other nutritional supplements used for OA treatment may include vitamin D (230,372), diacerein (373), and avocado soybean unsaponifiables (ASU) (374). Also, supplements of fish oil for OA is gaining attention for its possible role in reducing the expression of degradative enzymes and inflammatory cytokines in the cartilage of OA animal models (375,376). However, one clinical study showed fish oil at low or high doses did not delay structural progression of symptomatic knee OA (377). Furthermore, alternative therapies that have been implemented in the treatment of OA contain acupuncture (378), traditional Chinese medicine (379,380), and transcutaneous nerve stimulation (TENS) (381).

Regrettably, existing pharmacological treatments of OA are merely pain-relieving drugs. Thus, finding DMOADs that can induce the repair and regeneration of articular tissues would be important. However, there are no identifiable DMOADs by far, some latest developments have been made to investigate the impact of anti-cytokine therapy, enzyme inhibitors, growth factors, gene therapy, and peptides that are emerging pharmaceutical therapies for OA (382).

Anti-cytokine therapies that are recently used in OA treatment include tanezumab, AMG 108, adalimumab, etanercept, and anakinra. Tanezumab is a monoclonal nerve growth

factor inhibitor that relives hip and knee pain in OA patients (383,384). Oral administration of 5 mg for 24-56 weeks was recommended for OA patients (383,385), however it may cause osteonecrosis of the knee, rapid illness progression, augmented prevalence of total joint arthroplasty particularly when combined with NSAIDs (386–388). AMG 108 and Anakinra are another anti-cytokine drugs that are recently investigated to relive OA pain (389,390). AMG 108 is antibody against IL-1 receptor type 1 that restrains the IL-1α and IL-1β activities (389), while anakinra is an IL-1 receptor antagonist (390). Previous studies did not find significant difference in pain relive between OA patients taking AMG 108/anakinra and placebo (389,390). Adalimumab and etanercept are also of the anti-cytokine therapies that are tested lately for OA patients (391,392). Adalimumab is an antibody against TNF-α. One knee OA study identified that the intra-articular utilization of adalimumab reduced pain. Another study found that the utilization of adalimumab relieved pain of hand erosive OA (391). Also, etanercept was the first TNF-α inhibitor that was approved by the FDA and the European Medicine Agency for the treatment of moderate to severe RA (392). Inflammatory hand OA clinical trial showed that etanercept did not improve pain and function at 24 weeks or 1 year in OA patients (393). However, on subgroup analyses of participants with active inflammation, such as the presence of soft tissue swelling or power Doppler signals, etanercept revealed an improvement in radiographic scores (392). In addition, etanercept reduced serum MMP3-levels but no other soluble biomarkers of inflammation, cartilage, and bone damage (392,393).

Enzyme inhibitors of MMP such as M6495, doxycycline, cindunistat, and PG-116800 have been demonstrated to have a chondroprotective impact in OA studies (394–396). M6495 is a novel anti-ADAMTS-5 inhibiting nanobody that showed a protective effect on articular cartilage degradation and inhibited aggrecan turnover ex vivo in a dose-dependent manner (394). Although, doxycycline has broad-spectrum activity as a matrix MMP inhibitor, oral doxycycline has not been shown to be efficacious in managing OA and has produced significant adverse effects (395). Cindunistat is another enzyme inhibitor that is linked to the progressive degeneration occurred in OA and implemented in the condition management (382). Cindunistat is a selective inducible nitric oxide synthase inhibitor with a controversial role in clinical practice. In an experimental study on dogs, cindunistat inhibition caused a decrease in the catabolic effectors. However, a study comparing oral cindunistat with placebo demonstrated no difference in pain alleviation in individuals with knee OA (395). furthermore, in a multicentre double-blind placebo-controlled study of oral cindunistat (50 or 200 mg/day), the drug did not decrease the rate of the JSN versus placebo (397). However, the MMP inhibitor PG-116800 did not modify matrix structure in OA patients, it had unexpected side effects on muscle and skeleton; it limited joint mobility, and caused arthralgia, hand oedema, palmar fibrosis, Dupuytren's contracture and persistent tendon thickness or nodules (398).

Growth factors including bone morphogenetic protein 7 (BMP-7) and FGF18 have also been studied for managing OA (399,400). The intra-articular use of BMP-7 in individuals with knee OA has been shown to be safe and well tolerated. However, its effectiveness has not yet been shown (382,399). FGF18 or as it is called sprifermin was reported to

reduce type I collagen expression and had no hypertrophic effect (400). In a study of single or multiple intra-articular injections of FGF18 with a 1-year follow-up in knee OA patients, sprifermin resulted in less cartilage volume loss and augmented joint width in the lateral compartment than in the placebo group (401). However, pain alleviation was greater than in the placebo cohort. Therefore, sprifermin became a promising DMOAD, and 30-100 μg intra-articular injection every 6 months for 18 months is recommended for knee OA patients (402).

Moreover, several gene therapies using plasmid DNA, mRNA, and short oligonucleotides have been implemented to slow down the degeneration of the affected joint parts and simultaneously improve their repair and regeneration (403). Also, miRNA levels have been encountered to be increased in osteoarthritic cartilage (404–406). For instance, miRNA-140 has been found to play a fundamental role in the development of cartilage. Several experimental studies have reported that intra-articular injections of miRNA-140 have an anti-inflammatory effect and slow the progression of OA (407,408). Moreover, pre-clinical studies have shown the possible silencing of miRNA181a-5p by the antisense oligonucleotide (405).

Although pre-clinical studies have stated that calcitonin has a protective impact on cartilage and bone, a clinical trial conducted on OA patients did not demonstrate the drug's clinical efficacy (409). It is indicated that this was possibly due to low exposure to the drug (409,410).

Further potential DMOADs include SM04690 that is a Wnt signaling pathway inhibitor, several clinical trials on OA have reported the role of SM04690 in relieving pain and restoring function in OA patients based on the WOMAC osteoarthritis index (411,412).

Another OA therapy that targets the mechanisms of senescence includes the UBX101 (senolitic) (382,413). In an experimental study, UBX101 was demonstrated to eliminate senescent cells and slow down illness progression in OA mice (414). Also, clinical trials are being performed to assess the safety and tolerability of UBX101 in individuals with OA (415). This drug increases the activity of p53 tumor suppressor that regulates the cell cycle and induces apoptosis in senescent cells (382,415).

Moreover, transient receptor potential vanilloid 4 (TRPV4) that belongs to the TRPV subfamily of transient receptor potential ion channels was found to play a crucial role in the TGF-β signalling in chondrocytes, and therefore it became an attractive goal of the DMOADs (382). A study performed by Atobe *et al.* (2019) identified that local injection of TRPV4 agonist is a potential treatment for OA (416).

Another study identified that intra-articular injection of IL-1β induced more severe inflammation and cartilage degradation in the knee of neural epidermal growth factor-like 1 (Nell-1) mice than wildtype animals (417). Therefore, neural EGFL-like 1 (Nell-1) is considered as promising DMOAD with pro-chondrogenic and anti-inflammatory effects (382,417).

Likewise, Kjelgaard-Petersen *et al* (2019) reported that TPCA-1 (the κB kinase inhibitor) and tofacitinib (the Janus kinase inhibitor) maintain and help conserve cartilage ECM under inflammatory conditions and could be potential DMOADs for inflammation-driven OA (418).

Additional studies have been conducted to investigate the role of lorecivivint that modulates the Wnt signaling pathway in OA treatment (419). These studies showed that intra-articular injections of 0.07 mg of lorecivivint are effective on patient-related outcomes in knee OA patients (420).

Furthermore, immunohistochemical data have shown that quercitrin exerts an anti-osteoarthritic effect by deferring ECM degradation (421). Therefore, quercitrin could be a prospective DMOADs to prevent and manage the early stages of OA (382,421).

1.3.3 Surgical treatment:

When non-surgical conservative therapies fail to relief OA pain and restore joint function at severe OA stage, doctors may recommend surgical treatments for OA patients (422). Several types of techniques are employed. Arthroscopic surgery is a common outpatient procedure for knees and shoulders (423). In this procedure, doctors can repair the surfaces of damaged joints, removing loose cartilage, repairing cartilage tears, and smoothing bone surfaces (423). In contrary, the TJR or arthroplasty is performed for patients with advanced knee and hip OA. In the TJR, parts of an arthritic or damaged joint are removed

and replaced with a metal, plastic, or ceramic devices called prosthesis that mimic the shape and movement of a natural joint. Although hip and knee replacements are the most common performed TJR, joint replacement is also done for shoulders, fingers, elbows, and back joints (424). While they have risks, joint replacement surgery can be effective to restore some joint function and relief pain for majority of OA patients (425).

In conclusion, OA is a multifactorial musculoskeletal disorder that is estimated to be the most prevalent forms of chronic and painful arthritis. It is one of the ten most disabling disorders that affects ~10% of the worlds' population aged 60 years or older. Pathogenesis of OA involves all of the joint tissues with active anabolic and catabolic processes. Thus, OA is now defined as a disease of the whole joint involving articular cartilage, subchondral bone, synovium, ligaments, menisci, peri-articular muscles, and peripheral nerves. Despite the high prevalence and social burden of OA, it does not have a cure yet. Thus, there is an urgent need to identify novel targets for better understanding of the OA pathogenesis and establish OA intervention strategies to improve the life of people living with OA.

Chapter 2: Thesis objectives

OA is the most prevalent musculoskeletal disease and a leading cause of disability that affects about 10% of the world population aged 60 years and older. Heterogeneity of OA is the major hurdle of developing DMOADs. Efforts have been made to classify subtypes of OA patients based on epidemiological factors or joint structural changes seen on MRI, but significant overlapping features among OA subtypes defined by these methods limit their clinical application. Hence, there is an urgent need to develop new tools that can classify different subtypes of OA. Given that metabolites are the functional molecular intermediates that are the closest molecular products to the observed phenotypes, we hypothesized in chapter 3 that different subtypes of OA have unique metabolic profiles which can be identified by a metabolomics approach and tested our hypothesis in the well-established Newfoundland Osteoarthritis Study (NFOAS).

Age-related muscle weakness and wasting has been associated with several health and socioeconomical consequences including OA. However, the underlying mechanisms and intercorrelation between age-related muscle weakness and the pathogenesis of OA are not fully understood, previous studies documented that patients with OA have significant loss of muscle mass, function, and strength. Likewise, it became obvious that the age-related reduction of muscle strength and OA share similar pathologic processes and clinical features including changes of the joint anatomy, joint pain, decreased function, joint instability, periarticular muscle weakness, and fatigue. Moreover, our recent study on endotypes of OA came in agreement with these studies and found that an endotype of OA patients could be classified as muscle weakness OA. Thereby, the study of muscle strength has been increasingly recommended for better understanding of the OA

pathogenesis. In chapter 4, we hypothesized that metabolic profiling can be used to investigate the longitudinal reduction of muscle strength over 10 years and its possible implication in the pathogenesis of OA in older adults in a community-based older adult cohort from Southern Tasmania, Australia.

Although the recent largest GWAS-meta-analysis of 13 cohorts from 9 populations with a total of 826,690 individuals identified about 100 independent associated risk variants across 11 OA phenotypes, and the GWAS meta-analyses that were conducted to date have identified about 140 OA genetic risk variants, these genetic variants can only account for ~10% of OA heritability (334). Hence, further analysis is needed to identify the genetic component and explain the missing heritability of OA in non-studied populations like the NL population. We hypothesized in chapter 5 that novel OA susceptibility loci can be detected in the NL founder population to clarify whether the contribution of these loci extends to this European ancestry population. Therefore, a series of independent GWAS analyses in 557 primary OA patients from the NFOAS and 118 unaffected controls from the same population were conducted to identify novel genetic variants associated with OA in the NL population that were not identified previously.

However, previous GWAS studies identified about 140 OA genetic risk variants, the aetiology and pathogenesis of OA still not fully understood. Thereby, we hypothesized in the first part of chapter 6 that genes for OA can be detected by whole exome sequencing (WES) analysis, and I conducted a genetic variation annotation analysis in a WES data of

200 OA patients from the well-established NFOAS to investigate the possible putative genes for knee and hip OA. Furthermore, due to the involvement of multiple genes in the development of the OA, we hypothesized in the second part of chapter 6 that two-locus (gene*gene interaction) may identify digenic implication in the development and progression of OA and performed a genome-wide case-only test for detecting digenic interaction in the WES data of 200 OA patients from the NFOAS.

Despite the high prevalence and societal burden of OA, there is no cure for it yet. If nonsurgical treatments like medications and physiotherapy do not relieve pain and disability, doctors may recommend TJR surgery. Hip and knee replacements are the most commonly performed joint replacements, however up to 23% of THR and 34% of TKR patients either do not achieve improvement or get worse after the operation. Accordingly, it is important, to identify factors to predict responders and non-responders to TJR for clinical application and educate patients on their surgery expectations reliably. A number of potential non-genetic predictors for the outcome of TJR have been investigated, however the results were either inconclusive or with very limited predictive power. Also, none of the previously performed studies considered the genetic factors as predictor for TJR non-responders. We hypothesized in chapter 7 that genetic factors may play a role and could be used to predict poor outcome of the TJR in the NFOAS. The work conducted in chapter 7 of this thesis included a series of GWAS analyses to identify the genetic variations associated with non-responding of TJR in comparing with responding of TJR surgery from NFOAS.

Chapter 3: Endotypes of primary osteoarthritis identified by plasma metabolomics analysis.

Salem Werdyani[1], Ming Liu[1], Hongwei Zhang[2], Guang Sun[2], Andrew Furey[3], Edward W. Randell[4], Proton Rahman[2], Guangju Zhai[1]

[1]Division of Biomedical Sciences (Genetics),
[2]Discipline of Medicine,
[3]Discipline of Surgery,
[4]Department of Laboratory Medicine,
Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada.

A shorter version of this chapter is published in Rheumatology 2021; 60: 2735-2744.

3.1 Co-authorship statement:

GZ and SW were responsible for the study conception and design; statistical expertise; analysis and interpretation of the results; and drafting of the article. GS, HZ, AF, ML, and GZ participated in specimens; data collecting; and assembly. GS, AF, PR, and GZ were responsible for the provision of study materials or patients. ML, GS, AF, ER, PR, and GZ performed the critical revision of the article for important intellectual content. SW, ML, AF, ER, PR, and GZ provided the final approval of the article. GZ obtained the study funding. ML, PR, GZ provided the administrative, technical, and logistic support.

3.2 Copyright and License information:

3.3 Abstract:

Background: OA is the most common type of arthritis that affects about 10% of world population aged over 60 years. Heterogeneity of OA is the major hurdle of developing DMOADs. Efforts have been made to classify subtypes of OA patients based on epidemiological factors or joint structural changes seen on MRI, but significant overlapping features among OA subtypes defined by these methods limit their clinical

application. Hence, there is an urgent need to develop new tools that can classify different subtypes of OA. Given that metabolites are the functional molecular intermediates that are the closest to the observed phenotypes, we hypothesized that different subtypes of OA have unique metabolic profiles which can be identified by a metabolomics approach and tested our hypothesis in the well-established NFOAS.

Patients and Methods: The study participants were primary OA patients who underwent total hip or knee replacement surgery before 2017 in St. John's, Canada. Study controls were unaffected adults from the Complex Diseases in the Newfoundland population: Environment and Genetics (CODING) study. Diagnosis was made based on the ACR-OA criteria and post-surgery pathology reports on cartilage. Patients' plasma samples were collected before their surgery after at least 8 hours fasting, and metabolomic profiling was performed using the Biocrates AbsoluteIDQ p180 kit. Subsequently, quality control filtering was conducted to exclude low quality metabolite values. Then, common factor analysis was utilized to reduce the dimensionality of the metabolite's concentrations, and the identified factors were used for the following clustering analysis. Next, the optimal number of clusters was determined, the K-means clustering was applied to identify the endotypes of OA patients, and the logistic regression was used to identify the most significant metabolites that contribute to the classification of the endotypes of OA patients. Finally, sixteen clinical and epidemiological variables were examined in relation to the identified endotypes.

Results: A total of 615 primary OA patients and 237 controls were included in the study. Among the 186 metabolites measured, 162 passed the quality control filtering and included in the subsequent analysis. Factor analysis identified 17 distinct factors in 93 metabolites that had rotated factor loading > |0.3| in only one factor. Subsequently, the 615 OA patients were categorized in three main clusters (A = 66, B = 200, and C = 349). Multivariable logistic regression detected butyrylcarnitine (C4), arginine, and a number of glycerophosphlipids to be the major contributing factors for the differentiation between the clusters. Cluster A patients were characterized by a significantly higher concentration of C4 than other two clusters ($p = 2.04*10^{-4}$) and unaffected controls ($p = 4.08*10^{-8}$), but lower concentration of phosphatidylcholine acyl-alkyl (PC ae C40:3) in comparison with other OA patients ($p = 0.012$) and controls ($p = 1.30*10^{-11}$). The ratio of these two metabolites PC ae C40:3 to C4 had a great discriminatory power to classify patients in cluster A from controls as indicated by the area under the curve (AUC) = 0.92 from the receiver operating characteristic (ROC) analysis with sensitivity of 0.92 and specificity of 0.64. Elevated C4 is associated with generalized muscle weakness. Cluster B patients had a significantly lower arginine concentration than clusters A and C ($p = 3.44*10^{-16}$) and controls ($p = 7.98*10^{-11}$). The cluster C was distinguished from other two clusters by the lower concentration of lyso-phosphatidylcholine (lysoPC a C16:0, $p = 1.42*10^{-12}$) and controls ($p = 3.79*10^{-6}$), but higher phosphatidylcholine acyl-alkyl (PC ae C38:2) concentrations than other patients ($p = 5.8*10^{-3}$), and lower than controls ($p = 1.33*10^{-7}$). These phospholipids are thought to have pro-inflammatory effects. Further, we found that 55% of cluster A patients were diabetic in comparing to other clusters (13%; $p = 5.05*10^{-13}$), about 8% of cluster B patients had coronary heart disease (CHD) compared to other

OA patients (3.4%; p = 0.003), and 41% of cluster C patients had higher cholesterol concentration than other study participants (38%; p = 0.02).

Conclusion: Our data demonstrated that at least three distinct endotypes existed in primary OA, suggesting muscle weakness OA, arginine deficient OA, and inflammatory OA that can be distinguished by specific blood metabolic markers. While confirmation is needed, these findings provide us better understanding of OA pathogenesis and hold promising in developing personalized tools for OA management.

3.4 Introduction:

OA is the most common chronic progressive type of arthritis and one of the ten most disabling diseases in developed countries (426,427). The prevalence of OA has doubled since the mid-20th century, and about 240 million people worldwide have been diagnosed with OA (428). Global estimates of the symptomatic OA show that 9.6% of men and 18.0% in women aged over 60 years live with OA (2,113). Pain is the strongest symptom and sign that forces OA patients to seek medical assistance. Also, 80% of those patients with OA have limitations in movement, and 25% of them cannot perform their major daily activities (429). An estimated 37% of Canadians aged 20 years or older who had been diagnosed with arthritis reported OA as their only form of the condition (430). Of these, 12% experienced pain in their hip(s); 29%, in their knee(s), and 29%, in both joints (430). Thus, OA is considered to be the major source of joint pain and disability throughout Canadians (121,431), which it is expected to affect six million Canadians and

result in direct costs of CAD $7.6 billion annually in Canada by 2031, a 2.6-fold increase from 2011 (432).

Although OA has long been considered as a disease of the wear and tear of the articular protective cartilage at the load-bearing joints (16,73). This theme has been changed lately, and OA have been found to deteriorate all components of the arthritic joint causing loss of articular cartilage, JSN, remodeling of subchondral bone, formation of osteophytes, laxity of ligamentous, inflammation of synovial membrane, and weakening of the periarticular muscles (433). The destruction of the affected joint is mostly found in hands, knees, hips, shoulders and spine and leads to joint pain, swelling, bone deformity, and muscle weakness (434).

OA is a heterogeneous group of overlapping distinct conditions that have different aetiologies, but similar clinical manifestations (73). Its heterogeneity represents the major pitfall to efficacy detection of the disease modifying OA treatments. Efforts have been made to classify subtypes of OA patients based on the epidemiological factors and structure changes on MRI including disease mechanism of onset and pathophysiology (435), anatomical components (436), clinical manifestations (437), disease stage (438), affected joints (439,440), and degree of inflammation (441), but significant overlapping features among the OA subtypes defined by these methods limit their clinical application (442). Hence, there is an urgent need for novel biomarkers that can provide better understanding of the disease status and suggest suitable treatments for OA patients (442,443).

Lately, metabolomics study has become potential filed of research that investigates small-molecule metabolites in the body fluids or tissues of the biological system (443). Metabolites are the low molecular weight intermediates and downstream products of the genome, transcriptome, proteome, which are required at optimum concentration and steady state to maintain normal function of the cellular processes (442,444). Several studies recently proposed the involvement of metabolites in diseases including age-related metabolic dysfunction, fatigue syndrome, obesity, hypertension, diabetes, cardiovascular, and dyslipidemia (437,442,445–447). Also, the association of OA with obesity, diabetes, systemic low-grade inflammation, and oxidative stress has been lately reported (167,448–450), which suggest the involvement of metabolites in initiation and progression OA (443).

Previous study by our research team reported that the OA patients can be clearly classified into two distinct groups based on the significant difference in the metabolite concentrations from OA patient synovial fluid samples (442), with no significant differences were identified in either epidemiological risk factors or structural changes on MRI. We therefore hypothesized in this study that endotypes of OA patients exist and can be identified by a metabolomics approach and tested our hypothesis in the well-established NFOAS (71).

3.5 Materials and Methods:

3.5.1 Study participants:

The study subjects were primary OA patients who were recruited to the NFOAS (450) and comprised THR or TKR between November 2011 and September 2017 in St Clare's Mercy Hospital and Health Science Centre General Hospital in St John's, NL, Canada (441). OA diagnosis was made based on The ACR-OA clinical diagnostic criteria (71). Pathology reports on cartilage were investigated post-surgery to confirm the OA diagnoses in these patients. Study controls were OA-free adults from the same geographical region who were recruited to the on-going large-scale nutrigenomics cross-sectional CODING study (451). Subjects of the CODING study gave their written consent and completed a screening questionnaire to assess their physical characteristics and health status. Participants who were between 20 and 79 years old who were born and lived in NL and at least being third generation of a NL family; and healthy without serious metabolic, cardiovascular, or endocrine disease were eligible to participate in the CODING study (452). Then, our study ethical approval was acquired from the Health Research Ethics Authority of NL (reference number 11.311, **Appendix C**), and a written consent was obtained from all study participants. **Appendix D** includes a blank consent form.

3.5.2 Demographic and medical information:

A Self-administered questionnaire was used to collect the patients' demographic and medical information including age, sex, BMI, and comorbidities, **Appendix E**. A

patient's age was calculated at the TJR surgery by subtracting the birth date from the surgery date, then divided by 365.25 to get the number of years. BMI was calculated by dividing patient's weight in kilograms by the squared height in meters. Comorbidities including diabetes, hypertension, high cholesterol, coronary heart disease, gout, osteoporosis, breast cancer, colon cancer, skin cancer, melanoma, basal cell carcinoma, and squamous cell carcinoma were self-reported by the OA patients (441).

3.5.3 Metabolic profiling:

Patients' blood samples were collected with EDTA tubes after minimum of 8 hours fasting. Plasma was extracted from blood by centrifugation at 1500 rcf for 10 minutes at 4°C. Then, plasma was aliquoted and stored at -80^{o}C until use (453). Metabolomic profiling was performed on plasma samples obtained from study participants using the Biocrates *AbsoluteIDQ* p180 kit (BIOCRATES Life Sciences AG, Innsbruck, Austria), which measures the concentration of 186 metabolites, including 21 amino acids, 19 biogenic amines, 40 acylcarnitines, 90 glycerophospholipids, 15 sphingolipids, and one sugar hexose (>90% is glucose). **Table 3.1** provides the full list of the metabolites measured in the study. The profiling was done at the Metabolomics Innovation Centre (TMIC) using an API4000 QtrapVR tandem mass spectrometry instrument (Applied Biosystems/MDS Analytical Technologies, Foster City, CA, USA) equipped with an Agilent 1100 HPLC system (Agilent Technologies, Santa Clara, CA, USA) at the Metabolomics Innovation Centre (https://www.metabolo micscentre.ca). The complete analytical process of targeted metabolite concentrations was performed using the MetIQ software package, which is an integral part of the AbsoluteIDQVR p180 kit, and the

concentrations were reported in micromolar (454). Our in-house reproducibility of the assay was performed in 23 samples as previously describe (455); the mean coefficient of variation (CV) for all metabolites was 0.07 (0.05) **µ**M.

Table 3.1: Biocrates AbsoluteIDQ kit list of 186 metabolite concentrations.

Metabolite class	Number	Metabolite name or abbreviation	Biological relevance (selected examples)
Amino acids	21	Alanine, arginine, asparagine, aspartate, citrulline, glutamine, glutamate, glycine, histidine, isoleucine, leucine, lysine, methionine, ornithine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine	Amino acid metabolism, urea-cycle, activity of gluconeogenesis and glycolysis, insulin sensitivity, neurotransmitter metabolism, oxidative stress
Carnitine	1	C0	Energy metabolism, fatty acid transport and mitochondrial fatty acid oxidation, ketosis, oxidative stress, mitochondrial membrane damage
Acylcarnitine	25	C2, C3, C3:1, C4, C4:1, C5, C5:1, C6(or C4:1-DC), C6:1, C8, C9, C10, C10:1, C10:2, C12, C12:1, C14, C14:1, C14:2, C16, C16:1, C16:2, C18, C18:1, C18:2	
Hydroxy- and dicarboxyacylcarnitines	14	C3-OH, C4-OH(or C3-DC), C5:1-DC, C5-DC(or C6-OH), C5-M-DC, C5-OH(or C3-DC-M), C7-DC, C12-DC, C14:1-OH, C14:2-OH, C16:1-OH, C16:2-OH, C16-OH, C18:1-OH	
Biogenic amines	19	acetylornithine, asymmetric dimethylarginine, total dimethylarginine, alpha-Aminoadipic acid, carnosine, creatinine, histamine, kynurenine, methioninesulfoxide, nitrotyrosine, hydroxyproline, phenylethylamine, putrescine, sarcosine, serotonin, spermidine, spermine, taurine	
Lyso-phosphatidylcholines	14	lysoPC a C14:0/C16:0/C16:1/C17:0/C18:0/C18:1/C18:2/ C20:3/C20:4/C26:0/C26:1/C28:0/C28:1	Degradation of phospholipids, membrane damage, signalling cascades, fatty acid profile
Diacyl-phosphatidylcholines	38	PC aa C24:0/C26:0/C28:1/C30:0/C30:2/C32:0/C32:1/C32:2/ C32:3/C34:1/C34:2/C34:3/C34:4/C36:0/C36:1/C36:2/C36:3/C36:4/ C36:5/C36:6/C38:0/C38:1/C38:3/C38:4/C38:5/C38:6/C40:1/C40:2/ C40:3/C40:4/C40:5/C40:6/C42:0/C42:1/C42:2/C42:4/C42:5/C42:6	Dyslipidaemia, membrane composition and damage, fatty acid profile, activity of desaturases
Acyl-alkyl-phosphatidylcholines	38	PC ae C30:0/C30:2/C32:1/C32:2/C34:0/C34:1/C34:2/C34:3/ C36:0/C36:1/C36:2/C36:3/C36:4/C36:5/C38:0/C38:1/C38:2/C38:3/ C38:4/C38:5/C38:6/C40:1/C40:2/C40:3/C40:4/C40:5/C40:6/C42:0/ C42:1/C42:2/C42:3/C42:4/C42:5/C44:3/C44:4/C44:5/C44:6	
Sphingomyelines	10	SM C16:0, SM C16:1, SM C18:0, SM C18:1, SM C20:2, SM C22:3, SM C24:0, SM 24:1, SM C26:0, SM C26:1	Signalling cascades, membrane damage (eg,

Hydroxysphingomyelines	5	SM (OH) C14:1, SM (OH) C16:1, SM (OH) C22:1, SM (OH) C22:2, SM (OH) C24:1	neurodegeneration)
Hexose	1	H1	Carbohydrate metabolism

aa: acyl-acyl; ae, acyl-alkyl; a: lyso; Cx:y: where x is the number of carbons in the fatty acid side chain, y is the number of double bonds in the fatty acid side chain; DC: decarboxyl; M: methyl; OH: hydroxyl; PC: phophatidylcholine; SM: sphingomyelin; C: carbon; lysoPC: lysophosphatidylcholine; H: hexose.

3.5.4 Statistical analysis:

Quality control filtering was conducted on the raw metabolites data resulted from the *AbsoluteIDQ* p180 kit. Subsequent to combining the metabolite profiling data for all patients, metabolites with missing values or having values below the limit of detection (LOD) in more than 10% of OA patients were excluded from further analysis to minimize the false positive results as a standard practice in metabolomics studies (456). For metabolites with missing values or having values below the LOD in <10% of samples, the values were imputed by the mean of the given metabolites. Principal component analysis (PCA) demonstrated that we did not have any batch effect in our experiment; therefore, no correction for batch effects was performed.

The Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy were used to evaluate the factorability of the data and to determine if there were meaningful latent factors and variables inter-correlation within the metabolomic data (457). The Bartlett's test evaluated the statistical significance between the observed correlation matrix and the identified matrix in the metabolomic data (457). Also, the KMO examined if the partial correlations within the metabolites data suggest any latent factor underlying the metabolite concentrations based on the minimum acceptable value of 0.50 before undertaking a factor analysis. Then, the number of factors was determined using the scree plot, and parallel analysis (457). In the scree plot, the eigenvalues were plotted for all factors, and the number of factors was represented by the point where the eigenvalues dropped off sharply on the plot (458). Likewise, parallel analysis calculated the eigenvalues from randomly generated correlation matrices, which were then

compared with the observed eigenvalues extracted from the metabolite's dataset. Thus, the number of the observed eigenvalues generated from the metabolite's dataset with higher values than the corresponding random eigenvalues were more likely to form meaningful factors and represented the number of the retained factors (459). In contrary, the observed eigenvalues with lower values than their corresponding random eigenvalues do not produce true factors. Parallel analysis appears to be among the best methods for deciding how many factors to extract or retain (457–459). Based on the suggested number of factors in the metabolite data, common factor analysis was performed to reduce the dimensionality of the metabolites concentration data (457,460). Factor analysis identified the latent variables that were underlying the set of metabolite concentrations (457). Then, the identified factors were utilized in the subsequent clustering analysis if their corresponding eigenvalues were > 1.0 along with rotated absolute factor loading for each metabolite > 0.3 detected in only one factor (461). Next, factor scores for each OA patient were calculated by adding up all the metabolite concentrations in each identified factor and used in the clustering analysis (461).

Subsequently, the Hopkins statistic was used to assess the clustering tendency of the calculated factor scores from the 615 samples (462). Clustering tendency is measuring to what degree clusters exist in the data to be clustered, and may be performed as an initial test, before attempting clustering (462). The Hopkins statistic tested the spatial randomness of the dataset, checked if the dataset was generated by a uniform data distribution, this implies that there is underlying structure in the data and therefore meaningful results to be gained from further clustering analysis, and investigated whether

the 615 OA samples are clusterable and contains any inherent grouping structure based on the metabolite data. Hence, the dataset is considered to be significantly clusterable and used in the clustering analysis, if the identified Hopkins statistic value is close to zero (far below 0.5) (462).

Choosing which clustering method to use and the optimal number of clusters can be a daunting task. The optimal number of clusters was determined using the silhouette width and elbow plot methods (463). The average silhouette width approach measured the quality of clustering, and the optimal number of clusters was determined by maximizing the average silhouette over a range of possible values (464). Also, elbow plot method verified the number of the clusters by using the F-test to check and plot the percentage of variance between expected clusters, and at a certain point with a lot of variance between subject groups it gave an angle in the graph used to determine the optimum number of clusters (463).

Next, the clValid package in R was used for simultaneously compare multiple clustering algorithms including hierarchical, self-organizing maps, K-means, self-organizing tree algorithm, and model-based in a single function call for identifying the best clustering approach and the optimal number of clusters that should not only have good statistical properties, but also give results that are biologically relevant (465). This function used the internal evaluation silhouette score to evaluate the clustering probertites of our dataset by assigning the best score to the algorithm that produces clusters with high similarity within

a cluster and low similarity between clusters (464). The silhouette score is a metric that measures how cohesive and separated are the clusters. It ranges from -1 to 1, where a higher value indicates that the points are well matched to their own cluster and poorly matched to the neighboring clusters. The silhouette score is calculated by taking the average of the silhouette coefficients for each point, which are defined as the difference between the average distance to the points in the same cluster and the average distance to the points in the nearest cluster, divided by the maximum of these two distances (464,466). The silhouette score works well with *k*-means clustering and can help choosing the optimal number of clusters, by comparing the scores for different values of K (464). The clValid function estimated that the K-means clustering algorithm is the best algorithm to cluster our 615 OA patients (465).

Then, the factor scores from all of the 615 OA patients were utilized in the clustering analysis using K-means clustering algorithm to identify the endotypes of OA patients. K-means clustering is the most commonly used unsupervised machine learning algorithm for categorizing a given dataset into clusters based on the similarity and minimized variance between the clustered subjects (463). The K-Means algorithm calculates the goodness of classification and assigns individuals to clusters such that the overall within-cluster variance is minimized (internal cohesion) and the between-cluster variance is maximized (external separation) with the between Sum of Squares (BSS)/total Sum of Squares (TSS) ratio should approach 1 (463).

Then, multivariable logistic regression was performed to identify the most significant metabolites that contribute to the classification of the endotypes of OA patients and differentiate the patients in each group from healthy controls. Metabolite ratios between those identified metabolites as proxies for enzymatic reaction were examined to identify the most likely metabolic pathways (467). ROC analysis was conducted to evaluate the performance of identified metabolites and ratios in the classification of each endotype. The AUC was calculated, and optimal cut-off values were determined using the maximum sensitivity and specificity simultaneously (MaxSpSe) method.

Furthermore, we examined 15 clinical and epidemiological variables in relation to the identified endotypes, including age, sex, BMI, and comorbidities. Joint specificity (knee/hip) for the identified endotypes was also tested. Significance level was defined as $P \leq 0.0003$ after correction of multiple testing of 162 metabolites with the Bonferroni method. All the analyses were performed in R version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) with psych, GPArotation, corpcor, Factoextra, NbClust, cluster, clValid, stats, pROC, caret, OptimalCutpoints, dplyr and ggplot2 packages.

To complement our data analytic method, we randomly split our cohort equally into training and validation datasets with similar distribution of age, sex and, BMI between the two datasets using the CreateDataPartition function in the caret package in R (468). Then, the exact same methods described above were applied to the training dataset and replicated the results in the validation dataset.

3.6 Results:

A total of 615 primary OA patients and 237 OA-free controls were included in the study. OA patients were significantly older (p=0.0001) and had a higher BMI than controls (p=0.0001), but there was no difference in sex distribution between OA patients and controls (p=0.27), **Table 3.2**.

Table 3.2: The characteristics of the 615 OA patients and 237 OA-free controls.

	OA patients n=615	**OA-free controls n=237**	**P-value**
Sex (% for Females)	55.28%	59.49%	0.265
Age (yrs)	66.05□8.55	49.23□12.71	0.0001
BMI (kg/m^2)	33.74□6.86	29.03□5.05	0.0001

Values are mean ± SD for age and BMI continuous variable, and percentage for sex. P-values were obtained from Chi squared test for sex distribution and Student's t-test for continuous variables.

Among the 186 metabolite concentrations measured, 162 metabolites passed the QC criteria and were included in the analysis. The evaluation of the factorability of the metabolomic data using Bartlett's test of sphericity showed statistically significance (p=0.0001), suggesting significant difference of the correlation matrix from the identity matrix and thereby indicating that the metabolomic data is factorable. Also, the KMO test indicated that the metabolomic data was adequate for factor analysis with overall measure

of sampling adequacy (MSA) = 0.93. Scree plot (**Figure 3.1A)** and parallel analysis (**Figure 3.1B)** estimated that the metabolomic data can be categorized by 17 factors based on their correlation and relationship.

Accordingly, common factor analysis retrieved a rotated factor matrix of 17 distinct factors from 93 metabolites that had a rotated factor loading >|0.3| in only one factor. These 93 metabolites included 12 amino acids, 3 biogenic amines, 22 acylcarnitines, 48 glycerophospholipids, 7 sphingolipids, and 1 sugar hexose (>90% is glucose), **Table 3.3**.

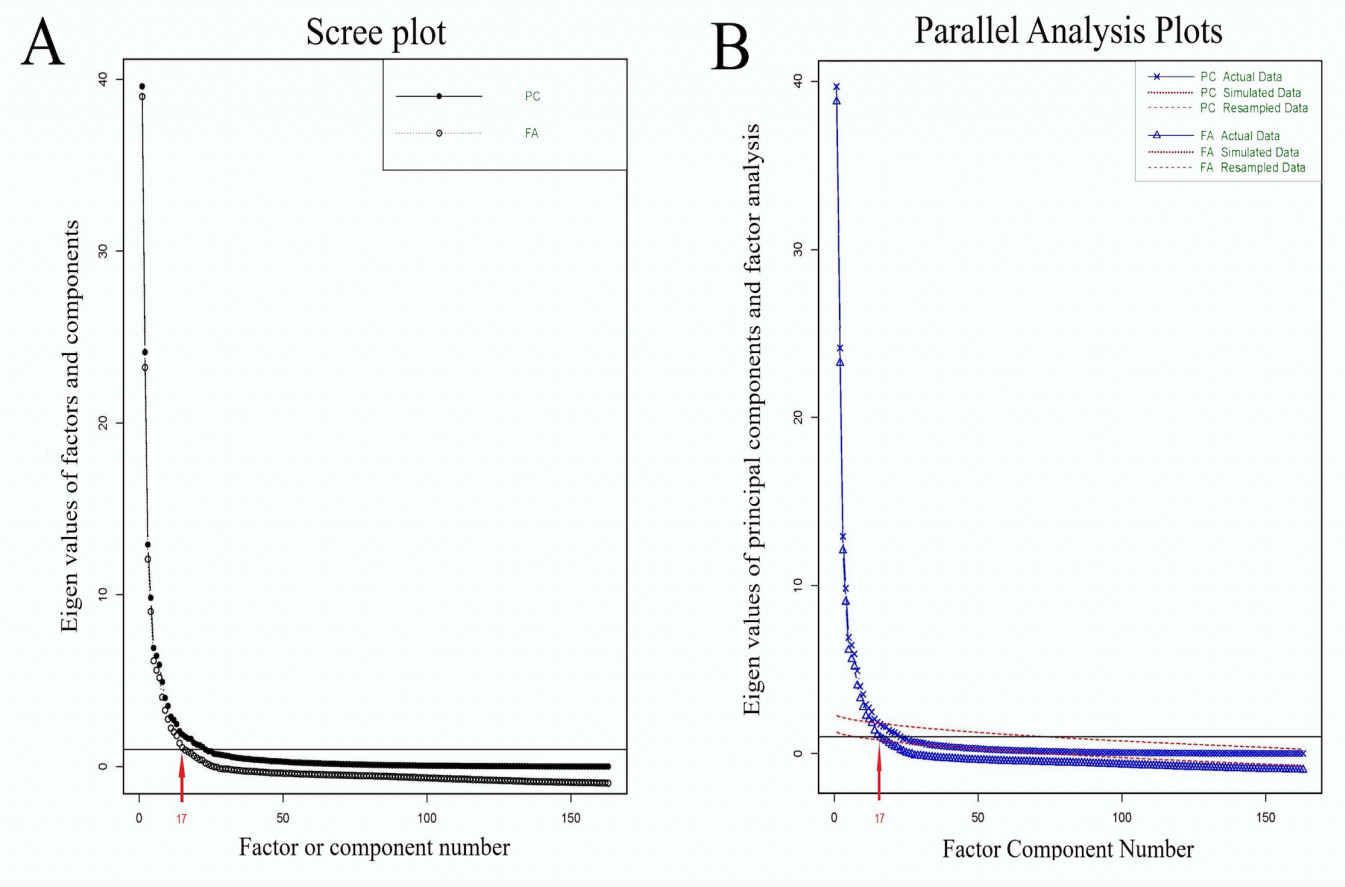


Figure 3.1: Estimation of the factor numbers from the concentrations of 162 metabolites of 615 OA patients. The red arrows in the Scree plot. (**A**) and parallel analysis plot (**B**) indicating that the metabolite data could be categorized by 17 factors based on their correlation and relationship. FA: factor analysis; PC: principal component analysis.

Table 3.3: 17 distinct factors from 93 metabolites that had rotated factor loading > |0.3| in only one factor.

Factors	Factor Loading	Factor items (metabolites)
Factor 1	7.63	PC ae C36.4, PC.ae.C36.5
Factor 2	18.33	C3-DC/C4-OH, C3-OH, C3.1, C4, C4:1, C5, C5-DC/C6-OH, C5-M-DC, C5-OH/C3-DC-M, C5:1, C5:1-DC, C6:1, C7-DC, C9, C10:2, C12-DC
Factor 3	9.46	Glycine, lysoPC a C16:0, lysoPC a C16:1, lysoPC a C17:0, lysoPC a C18:0, lysoPC a C18:1, lysoPC a C18:2, lysoPC a C20:3, H1
Factor 4	9.52	PC aa C24:0, PC aa C32:3, PC aa C40:2, PC ae C30:1, PC ae C30:2, PC ae C38:1, PC ae C38:2, PC ae C40:3
Factor 5	4.78	PC aa C30:0, PC aa C32:1, PC aa C34:1
Factor 6	7.27	Isoleucine, Leucine, Lysine, Phenylalanine, Proline, Tryptophan, Tyrosine, Valine, Kynurenine
Factor 7	6.72	SM-OH C22:1, SM-OH C22:2, SM C20:2, SM C22:3
Factor 8	7.92	C6/C4:1-DC, C8, C10, C14:1, C14:2, C16:2
Factor 9	8.02	PC aa C36:5, PC aa C36:6, PC aa C38:6, PC aa C40:6, PC aa C42:2, PC ae C38:0
Factor 10	5.26	lysoPC a C24:0, lysoPC a C26:0, lysoPC a C26:1, lysoPC a C28:0
Factor 11	6.96	Taurine, PC aa C42:5, PC ae C34:1, PC ae C36:0, PC ae C42:1
Factor 12	5.64	SM-OH C14:1, SM-OH C16:1, SM C16:1
Factor 13	5.76	PC aa C36:4, PC aa C38:4, PC aa C40:4, PC aa C40:5
Factor 14	10.42	PC ae C40:5, PC ae C42:4, PC ae C42:5, PC ae C44:3, PC ae C44:4, PC ae C44:5, PC ae C44:6
Factor 15	5.72	Arginine
Factor 16	7.06	PC aa C34:2, PC aa C36:2, PC aa C36:3
Factor 17	3.65	Citruline, Glutamine, Serine

aa: acyl-acyl; ae, acyl-alkyl; a: lyso; Cx:y: where x is the number of carbons in the fatty acid side chain, y is the number of double bonds in the fatty acid side chain; DC: decarboxyl; M: methyl; OH: hydroxyl; PC: phophatidylcholine; SM: sphingomyelin; C: carbon; lysoPC: lysophosphatidylcholine; H: hexose.

Then, the Hopkins statistic indicated that the metabolomic dataset of OA patients was significantly clusterable with Hopkins statistic value = 0.18. The minimum silhouette width (**Figure 3.2A**) and elbow plot (**Figure 3.2B**) anticipated that the 615 OA patients could be classified into three main clusters based on the metabolomic factor scores. K-means clustering analysis determined three distinct main clusters (cluster A=66, cluster B=200, and cluster C=349) of OA patients with BSS/TSS = 77.0 %, **Figure 3.3**.

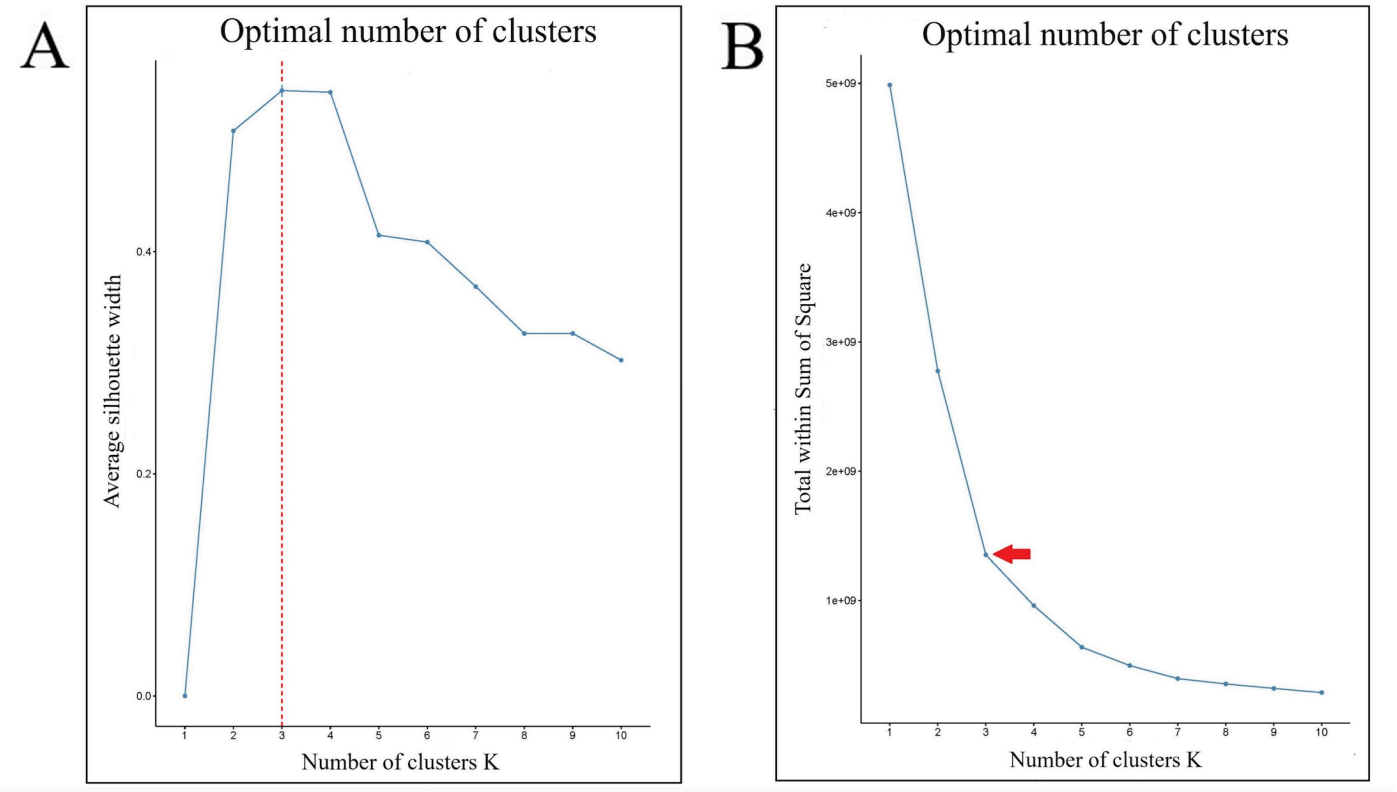


Figure 3.2: Estimation of the factor and cluster numbers from the concentrations of 162 metabolites of 615 OA patients. The red vertical dash line in the minimum silhouette width plot (**A**) and the red arrow in the elbow plot (**B**) indicate that the 615 OA patients can be classified into three main clusters based on the calculated metabolite factor scores from the 17 identified factors.

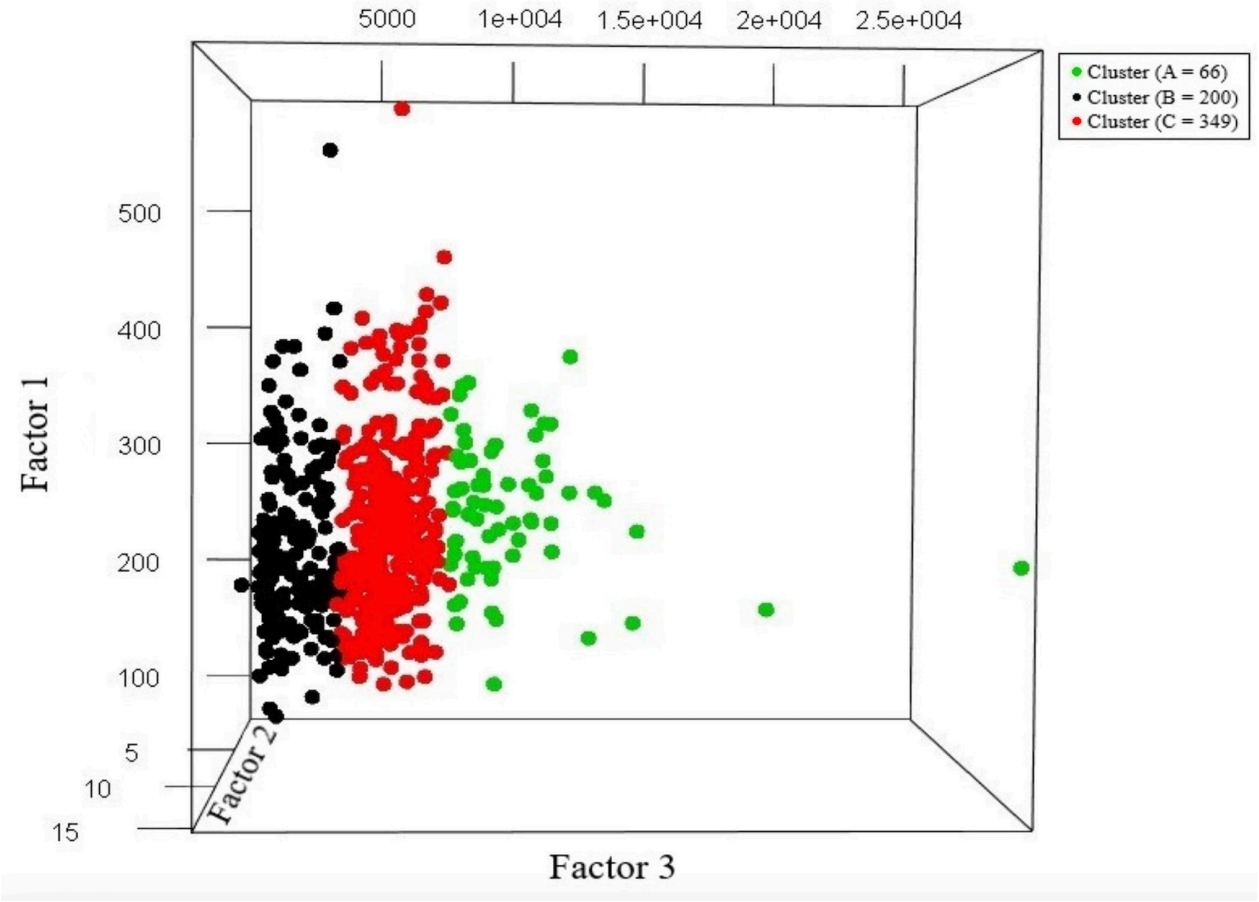


Figure 3.3: The three distinct clusters of OA patients identified by K-means clustering method coupled with metabolomics data.

Logistic regression identified butyrylcarnitine (C4), arginine, and lysophosphatidylcholine with palmitic acid (lysoPC a C16:0) to be the major contributing factors for the classification and differentiation between cluster A, B, and C. The cluster A patients were characterized by a significantly higher concentration of C4 than the other two clusters ($p=2.04*10^{-4}$) and OA-free controls ($p=4.08*10^{-8}$), **Table 3.4**. While it did

not reach the pre-defined significance, the second metabolite was a phosphatidylcholine with 40 carbons and three double bonds (PC ae C40:3) for which the patients in cluster A had a lower concentration than other OA patients (p=0.01) and controls ($p=1.30*10^{-11}$), **Table 3.4**. When the ratio between these two metabolites was considered, the significance became stronger. The PC ae C40:3 to C4 ratio was significantly lower in cluster A patients than in patients of other clusters ($p=4.89*10^{-7}$) and controls ($p=1.31*10^{-11}$), **Table 3.4**. The ROC analysis showed that the ratio had an AUC of 0.92 (95% CI: 0.89 to 0.95) to distinguish cluster A patients from OA-free controls with a sensitivity of 0.83 and specificity of 0.83 at the optimal cut-off value of 4.75. The ratio had a moderate discriminatory power to distinguish patients in cluster A from patients in other clusters with an AUC of 0.68 (95% CI: 0.62 to 0.75), **Figure 3.4A**.

The cluster B patients had a significantly lower arginine concentration than clusters A and C ($p=3.44*10^{-16}$) and controls ($p=7.98*10^{-11}$), **Table 3.4**. The ROC curve analyses showed that arginine had an AUC = 0.82 (95% CI: 0.78 to 0.85), and a sensitivity of 0.74 and a specificity of 0.75 to discriminate patients in cluster B from patients in other two clusters at an optimal cut-off value of 28.30μM; and an AUC = 0.99 (95% CI: 0.98 to 1.00) with a sensitivity of 0.95 and a specificity of 0.96 to distinguish patients in cluster B from controls at the optimal cut-off value of 56.10 μM, **Figure 3.4B**.

Patients in cluster C were distinguished from patients in the other two clusters and OA-free controls by a lower concentration of lysoPC a C16:0 (all $p \leq 3.79*10^{-6}$), **Table 3.4**. Although it did not reach the pre-defined significance level, the second top associated

metabolite was a phosphatidylcholine with 38 carbons and two double bonds (PC ae C38:2, p≤0.006), **Table 3.4**. When examine the ratio of these two metabolites, we found that the significance became stronger and lysoPC a C16:0 to PC ae C38:2 ratio was significantly lower in patients of cluster C than in patients of other clusters (p=0.0001) but significantly higher than in OA-free controls (p=0.0001), **Table 3.4**. The ratio was also significantly higher in patients of cluster A and B than in controls (p=0.0001). The ROC curve analyses showed that the ratio had an AUC = 0.74 (95% CI: 0.70 to 0.78), and sensitivity of 0.66 and specificity of 0.70 to discriminate patients in cluster C from patients in other clusters at an optimal cut-off value of 63.28, and an AUC = 0.87 (95% CI: 0.85 to 0.90) with a sensitivity of 0.80 and specificity of 0.81 to distinguish cluster C patients from OA-free controls with the optimal cut-off value of 32.01, **Figure 3.4C**.

Table 3.4: Most significant metabolites contributing to the classification of endotypes of primary OA patients.

OA endotypes	Metabolite	Concentration (Mean±SD)	Concentration in other OA patients		Concentration in controls	
			Mean±SD	P-value	Mean±SD	P-value
Cluster A (n=66)	C4 (μM)	0.34±0.25	0.26±0.16	0.0002	0.23±0.14	$4.08*10^{-8}$
	PC ae C40:3 (μM)	0.85±0.32	0.97±0.37	0.01	1.67±0.58	$1.30*10^{-11}$
	PC ae C40:3 to C4 ratio	3.11±1.57	4.51±2.39	$4.89*10^{-7}$	8.81±4.51	$1.31*10^{-11}$
Cluster B (n=200)	Arginine (μM)	21.39±17.83	41.76±19.48	$3.44*10^{-16}$	115.08±50.77	$7.98*10^{-11}$
Cluster C (n=349)	LysoPC a C16:0 (μM)	79.02±25.46	130.47±65.10	$1.42*10^{-12}$	82.61±37.02	$3.79*10^{-6}$
	PC ae C38:2 (μM)	1.94±1.44	1.74±1.20	0.006	4.54±2.30	$1.33*10^{-7}$
	lysoPC a C16:0 to PC ae C38:2 ratio	53.15±25.09	90.93±50.67	0.0001	22.34±12.13	0.0001

C4 – butyrylcarnitine; PC ae C40:3 - Phosphatidylcholine acyl-alkyl with 40 carbons and 3 double bonds; lysoPC a C16:0 - Lyso-phosphatidylcholine with 16 carbons and no double bond; PC ae C38:2 - Phosphatidylcholine acyl-alkyl with 38 carbons and 2 double bonds.

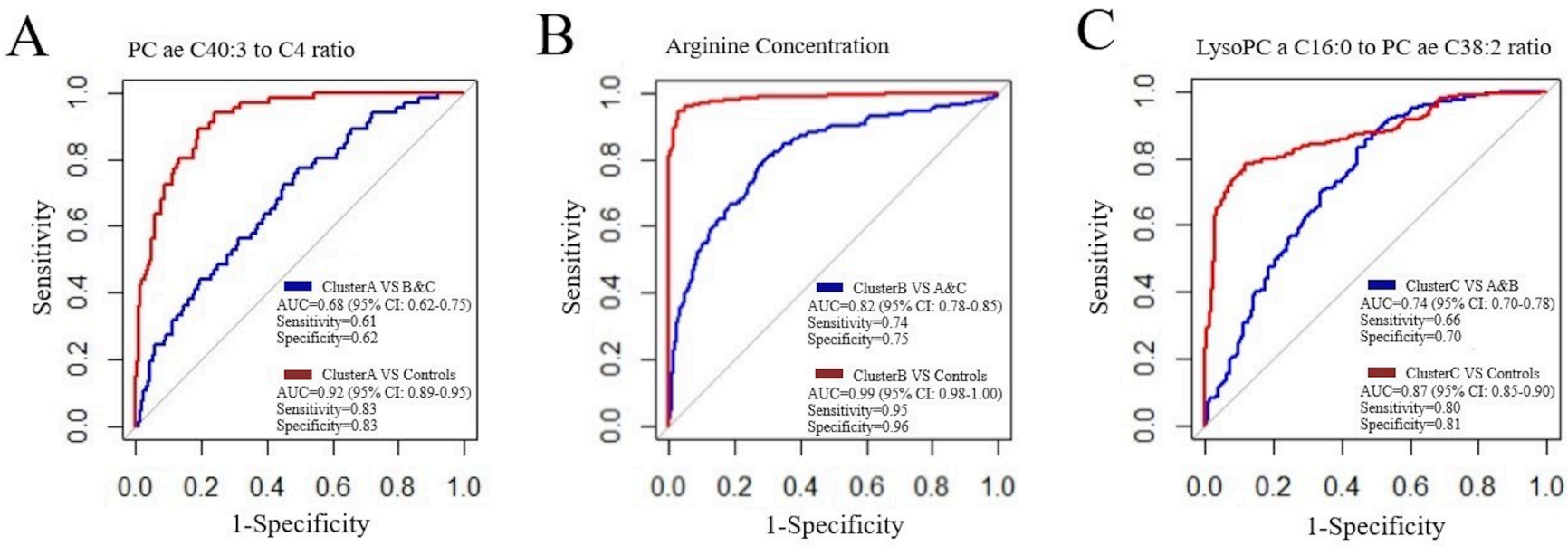


Figure 3.4: The ROC curve analysis results of each identified cluster: **A)** the PC ae C40:3 to C4 ratio in relation to distinguish patients in cluster A from other clusters and controls. **B)** Arginine in relation to distinguish patients in cluster B from other clusters and controls. **C)** LysoPC a C16:0 to PC ae C38:2 ratio in relation to distinguish patients in cluster C from other clusters and controls.

In the split sample analysis, a total of 615 OA patients were randomly divided into a training dataset (n= 308) and a validation dataset (n= 307). The minimum silhouette width (**Figure 3.5A**) and elbow plot (**Figure 3.5B**) estimated that the training dataset could be classified into three main clusters, which was validated in the validation dataset, **Figure 3.6A** and **B**. K-means clustering analysis identified three distinct clusters (TA=39, TB=95, and TC=174) of OA patients in the training dataset, which was also validated in the validation dataset (VA=10, VB=112, and VC=185). Logistic regression analysis identified exactly same metabolites- arginine ($p=4.28*10^{-8}$) and lysoPC a C 16:0 ($p=4.79*10^{-10}$) that were the key contributors for cluster B and C as found in the entire cohort analysis, which were also confirmed in the validation dataset with $p=1.38*10^{-7}$ and $p=2.98*10^{-9}$ for arginine and lysoPC a C16:0 respectively, **Table 3.5**. C4 was also identified as the key contributor for cluster A in the training dataset and confirmed in the validation dataset, but the P-value=0.01 did not reach the pre-defined significance level because of the small sample size, **Table 3.5**.

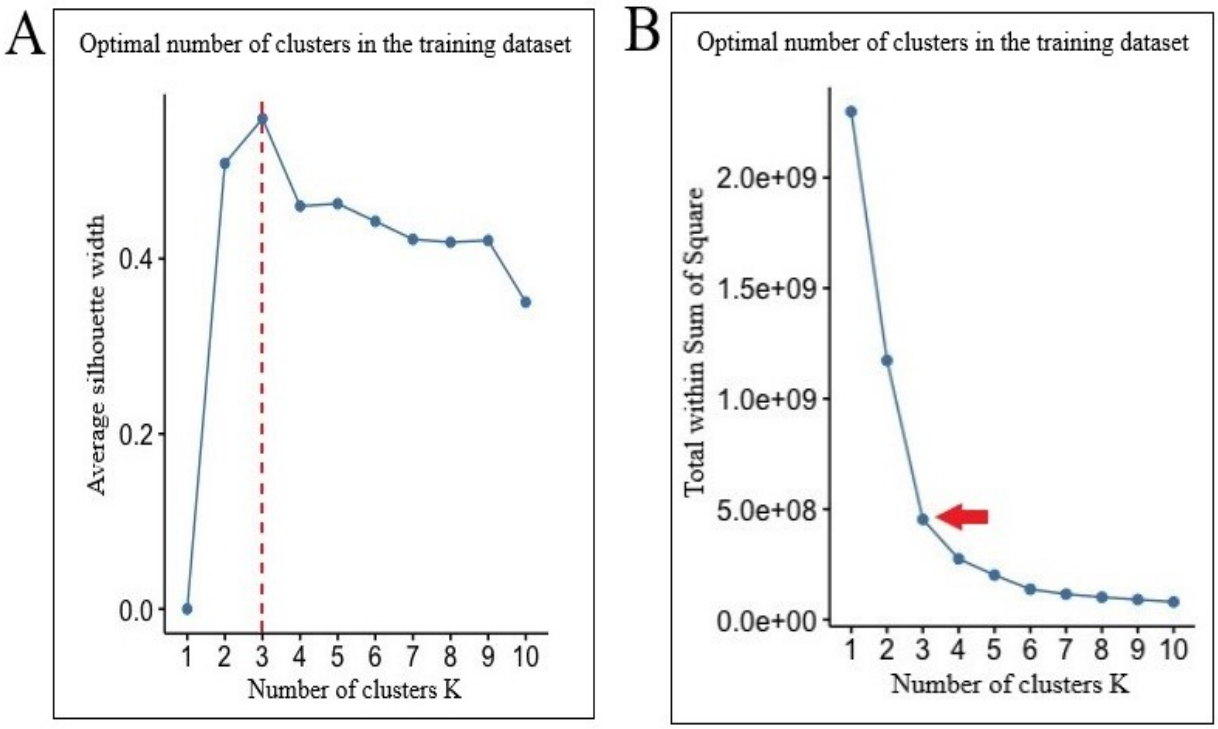


Figure 3.5: Estimation of the cluster number in the training dataset of 308 OA patients, **A)** the minimum silhouette width plot, and **B)** the elbow plot showing that the 308 OA patients in the test dataset can be classified into three main clusters based on the similarity and relationship between the metabolites.

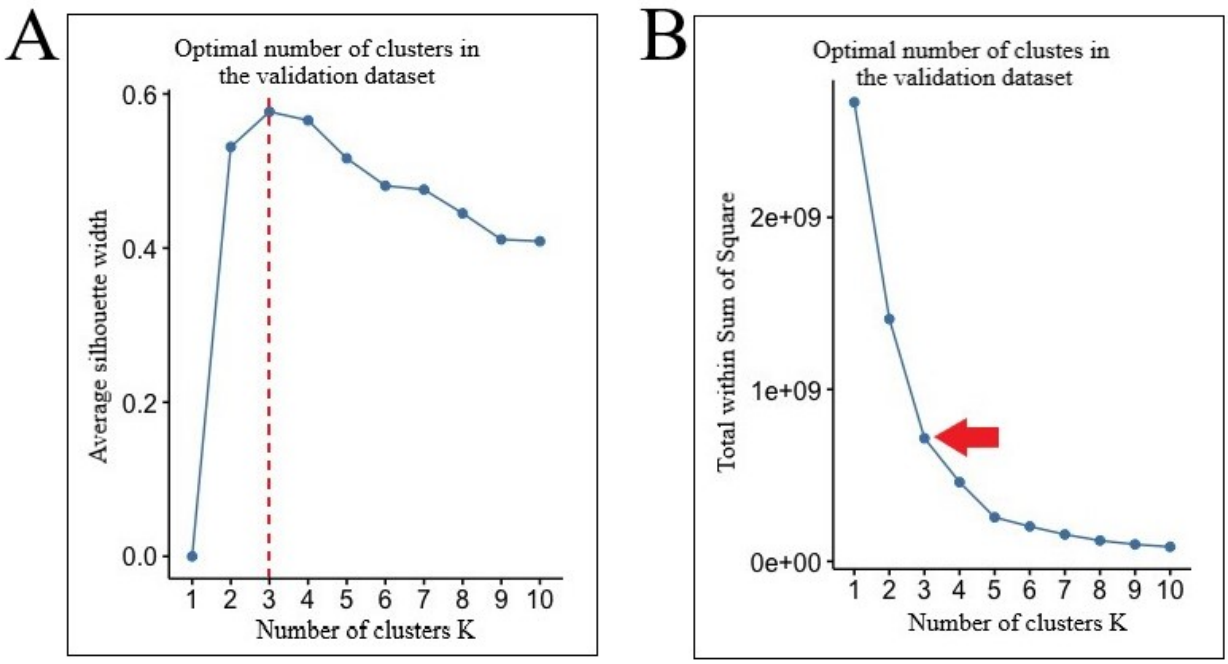


Figure 3.6: Estimation of the cluster number in the validation dataset of 307 OA patients, **A)** the minimum silhouette width plot, and **B)** the elbow plot showing that the 307 OA patients in the validation dataset can be classified into three main clusters based on the similarity and relationship between the metabolites.

Table 3.5: The most significant metabolites contributing to the classification of endotypes of primary OA patients in the training and validation datasets.

	Training dataset				**Validation dataset**			
	Clusters	Concentration (Mean±SD µM)	Concentration in other clusters (Mean±SD µM)	P-value	Clusters	Concentration (Mean±SD µM)	Concentration in other clusters (Mean±SD µM)	P-value
C4	Cluster A (n=39)	0.34±0.22	0.27±0.22	0.011	Cluster A (n=10)	0.48±0.15	0.25±0.15	0.011
Arginine	Cluster B (n=95)	22.25±16.76	43.35±18.64	$4.28*10^{-8}$	Cluster B (n=112)	20.65±18.32	40.75±20.13	$1.38*10^{-7}$
LysoPC a C16:0	Cluster C (n=174)	79.26±25.36	133.41±69.36	$4.79*10^{-10}$	Cluster C (n=185)	78.55±26.41	131.83±59.75	$2.98*10^{-9}$

C4 – butyrylcarnitine; lysoPC a C16:0 - Lyso-phosphatidylcholine with 16 carbons and no double bond.

With regards to the clinical and epidemiological factors, there was no sex difference among the three identified clusters (all $p>0.22$), but patients in cluster A had a higher BMI and a higher prevalence of diabetes than the other two clusters (all $p\leq0.0009$), as well as significantly older than cluster B ($p=0.018$) and had a higher prevalence of CHD than cluster C patients ($p=0.039$) (**Table 3.6**). Patients in cluster B had a significantly higher prevalence of CHD than cluster C ($p=0.003$), whereas patients in cluster C had a significantly higher prevalence of osteoporosis than cluster B ($p=0.009$). The prevalence of osteoporosis in cluster C was also higher than in cluster A (22.06% vs. 15.15%) but was not statistically significant (**Table 3.6**).

In addition, the study cohort included 68% of knee OA patients and 32% of hip OA patients. However, we found that there was no difference in the distribution of knee and hip OA among the three identified OA clusters ($p=0.43$). The proportion of knee OA in cluster A, B, and C was 72.73%, 62.32%, and 69.86%, respectively.

Table 3.6: Significant epidemiological factors and comorbidities associated with each of identified clusters.

Factors	Cluster A	Cluster B	Cluster C	P-value		
				A vs. B	***A vs. C***	***B vs. C***
Age (years)	67.78±7.37	65.03±8.32	66.32±8.82	0.018	0.207	0.094
BMI (kg/m^2)	36.57±7.27	33.29±6.70	33.46±6.74	0.0009	0.0008	0.776
Diabetes (%)	54.55	13.50	12.89	$1.039*10^{-11}$	$4.906*10^{-15}$	0.840
CHD (%)	7.57	8.00	2.58	0.912	0.039	0.003
Osteoporosis (%)	15.15	13.00	22.06	0.658	0.206	0.009

*Values are mean ± SD for age and BMI, percentage for diabetes, CHD, and osteoporosis. CHD – coronary heart disease. P-values were obtained from Chi-squared test or Student t-test wherever appropriate.

3.7 Discussion:

In this metabolomic analysis with a large sample size, we were able to identify three distinct endotypes of OA patients. The significant metabolite contributors to each of the three endotypes implied that the primary OA patients can be classified as muscle weakness, arginine deficient, and low inflammatory OA. The findings provide new insights into the pathogenesis of primary OA and could help to develop personalized tools for OA management.

In a preliminary study with only 80 OA patients, we previously found that the OA patients can be classified into two distinct groups (469). 11% of the 80 patients were classified into one group characterized by high concentration of acetylcarnitines in the synovial fluid. The findings of the current study were consistent with it, and we found that 11% of the 615 primary OA patients was classified into cluster A based on plasma metabolic profiles, and a specific acetylcarnitine - C4 was the key contributor to the clustering. The patients in cluster A had on average 33% increase in C4 compared to the OA patients in other two clusters and 52% increase compared to OA-free controls.

Acetylcarnitines are used to transport fatty acids from cytosol into mitochondrial matrix for energy production (441). C4 is a short-chain acetylcarnitine and responsible for the transfer of short-chain fatty acids. It had been reported that accumulation of C4 reflects the abnormal concentration of tissue butyryl CoA due to defect or inhibition of short-chain acyl-CoA dehydrogenase (SCAD), which is the key enzyme involved in short-chain

fatty acid beta-oxidation pathway in mitochondria, leading to energy pathway defect and generalized muscle weakness (470). Elevated concentration of C4 in blood is one of the diagnostic parameters for SCAD deficiency. SCAD deficiency is a clinically heterogeneous disorder with variable clinical phenotypes ranging from fatal metabolic decompensation in early life to subtle adult onset with asymptomatic phenotypes in some patients (471). Adult patients are more likely to have problems related to muscle weakness and wasting (471,472). Muscle weakness has been associated with OA (473). Thus, our findings suggest that patients in cluster A might have weakened muscle strength which makes them susceptible to OA. To the best of our knowledge, elevated plasma concentration of C4 had not been reported in OA patients previously, however, it had been associated with disorders including diabetes, obesity, and cardiovascular diseases (474,475). Interestingly, the analysis of fifteen clinical and epidemiological factors in relation to the identified clusters found that the majority of cluster A patients were diabetic and having higher BMI than other study participants. This cluster patients also had a significantly higher prevalence of CHD than cluster C patients.

Moreover, we found previously that both knee OA patients and diabetic patients had a lower concentration of two phosphatidylcholines (PC ae C34:3 and PC ae C36:3) than controls (476). In the current study, we found that patients in cluster A were also associated with a phosphatidylcholine - PC ae C40:3. Although the statistical test only reached the pre-defined significance when comparing cluster A with controls, the strength of association became much stronger when inspect the ratio of these two top metabolites, supporting a hypothesis that cluster A had an impaired fatty acid oxidation for energy

production leading to insufficient energy levels for muscle functions. Thus, OA patients in cluster A would benefit from interventions or therapies that improve muscle strength.

Cluster B with 200 OA patients was characterized by a significant reduction of plasma free arginine concentration. Previous studies have reported that blood arginine levels were reduced by 24 - 31% in OA patients (477,478). Animal model of OA showed a significant reduction of arginine concentration after anterior cruciate ligament transection (ACLT) in the ACLT rabbit model with a negative association between the post-ACLT arginine concentration and severity of OA (479), suggesting the potential mechanisms of the reduced arginine concentration in OA patients is due to increased demand of arginine for cartilage repair in OA (480) and inability of the body to meet the demand.

Further, arginine has antihypertensive and antioxidant properties, which influences blood viscosity and coagulation system, and affects the metabolism of glucose, lipids and proteins (481). Evidences show that arginine intake in cardiovascular patients reverses endothelial dysfunction associated with major cardiovascular risk factors, such as hypercholesterolemia, smoking, hypertension, diabetes, obesity, insulin resistance, and aging (482). Also, multiple studies reported an important role of arginine in improving blood flow in the arteries of the heart that may improve symptoms of clogged arteries, chest pain or angina, and CHD (483). Interestingly, our analysis came in agreement with the findings of these studies and showed that cluster B patients had a significantly higher prevalence of CHD especially than cluster C. Hence, supplementation of arginine might be beneficial to OA patients, particularly in cluster B patients. Further, arginine is a

natural inhibitor of cathepsin such as cathepsin B and K, a protease that breaks down cartilage (484), thus, the depletion of arginine in cluster B patients reduces its inhibitory effect on cathepsin and leads to over activity of cathepsin which in turn leads to cartilage breakdown. Functional studies are needed to confirm this.

Significant differences in the concentration of lysoPC a C16:0 contributed to the categorization of cluster C of 349 OA patients from other OA patients and controls. LysoPCs are bioactive lipids that contribute to a variety of cellular functions (485,486), such as proliferation, apoptosis, smooth muscle contraction, wound healing, and tumor cell invasiveness (487), and they have been reported to stimulate pro-inflammatory cytokines, such as IL-1β, TNF, and IL-6 leading to the initiation and progression of OA (488). The other possible pathway to produce lysoPCs from phosphatidylcholines uses the reactive oxygen species (ROS) in neutrophils (453). Under the oxidative stress that is implicated in the pathogenesis of OA, the majority of OA joint cells can produce large amounts of ROS and nitric oxide (NO) in response to biomechanical or biochemical stimuli. Then, the mixture of proteolytic enzymes released from neutrophils combined with different ROS may stimulate the cartilage damage at the OA joints (453,489).

Although the second metabolite PC ae C38:2 only reached the pre-defined significance when comparing patients in cluster C with controls, the ratio of lysoPC a C16:0 to PC ae C38:2 was statistically significant. The ratio was significantly higher in all three identified OA clusters than in OA-free controls which is in agreement with our previous studies, where we documented a significant increase of lysoPCs to PCs ratio to be

associated with knee OA risk (453), knee cartilage volume changes in two years follow-up (490), and an increased risk for undergoing TKR in 10 years follow-up (453). The ratio has also been found to be able to predict OA patient's response to symptomatic drugs (491). The elevated ratio indicated that the conversion pathway of PC to lysoPC was overactivated in OA and led to the stimulation of inflammation. Thus, lysoPCs to PCs ratio has been suggested as a possible biomarker for monitoring anti-inflammatory treatment in RA as well (492). Between identified clusters, lysoPC a C16:0 to PC ae C38:2 ratio was significantly lower in cluster C patients compared to OA patients in other clusters, suggesting that cluster C might have a lower level of inflammation. Further analysis found a significantly lower percentage of cluster C patients to have diabetes, CHD, but higher percentage of these patients had osteoporosis compared to other two clusters. These findings support a proposed hypothesis that bone loss might be an initiation factor for OA development at least for patients in cluster C (493). Thus, identifying cluster C patients for clinical trials of anti-osteoporotic drugs in OA would be helpful for detecting its efficacy.

There are a number of limitations in this study. The metabolic profiling was done with a commercially available metabolomics assay kit that has limited coverage of metabolites. Thus, we might miss some metabolites that may contribute to the endotypes of OA. Also, study participants included both knee and hip OA patients. While knee and hip OA share a number of risk factors, the aetiology might be different between knee and hip OA. However, we did not find a different distribution of knee and hip OA in the identified three clusters. Further studies of knee and hip OA cohorts with sufficient sample sizes are

needed to confirm these findings. Similarly, prevalence of OA particularly knee OA was different between men and women, however, we did not find a sex difference in the three identified endotypes of OA. Different endotypes might have different severity of the disease or different observed characteristics such as muscle weakness in cluster A, but we did not have data on the severity or muscle strength, further studies are needed to confirm these results. Lastly, all the study participants were from NL, which is a genetically/ethnically homogeneous population that may limit the generalizability of our findings to other populations.

In conclusion, our data demonstrated that at least three distinct endotypes existed in primary OA, suggesting muscle weakness, arginine deficient, and low inflammatory OA subtypes that can be distinguished by specific blood metabolic markers. While confirmation is needed, these findings provide new insights into the understanding of OA pathogenesis and hold promise in developing personalized tools for OA management toward reduction of economic burden and better quality of life for OA patients.

Chapter 4: Metabolomic signatures for the longitudinal reduction of muscle strength over 10 years

Salem Werdyani[1], Dawn Aitken[5], Zhiwei Gao[2], Ming Liu[1], Edward W. Randell[3], Proton Rahman[4], Graeme Jones[5], Guangju Zhai[1]

[1]Division of Biomedical Sciences (Genetics),
[2]Division of Community Health and Humanities,
[3]Discipline of Laboratory Medicine,
[4]Discipline of Medicine,
Faculty of Medicine, Memorial University of Newfoundland, St. John's, Canada.
[5]Menzies Research Institute, University of Tasmania, Hobart, Australia.

A shorter version of this chapter is published in Skeletal Muscles 2022; 12:4.

4.1 Co-authorship statement:

GZ, ZG, and SW were responsible for the study conception and design; statistical expertise; and analysis and interpretation of the results. GZ and SW drafted the article. DA, GJ, PR, and GZ were responsible for the provision of study materials or patients. DA, ZG, ML, ER, PR, GJ, and GZ performed the critical revision of the article for important intellectual content. SW, DA, ZG, ML, ER, PR, GJ, and GZ provided the final approval of the article. DA, GJ, and GZ obtained the study funding. ML, PR, GZ provided the administrative, technical, and logistic support.

4.2 Copyright and License information:

4.3 Abstract:

Purpose: Skeletal muscles are essential components of the neuromuscular skeletal system that have an integral role in the structure and function of synovial joints. Muscle weakness and atrophy is one of the earliest signs of aging. Limited number of studies have been performed to investigate the implication of muscle weakness in the pathogenesis of OA. Since metabolites are the functional molecular intermediates that are the closest to the observed phenotypes, we undertook this study to identify the baseline metabolomic signatures for the longitudinal reduction of muscle strength over 10 years in the well-established community-based Tasmanian Older Adult Cohort (TASOAC).

Methods: The study was conducted as part of the TASOAC Study, a prospective, population-based study aimed at identifying the environmental, genetic, and biochemical factors associated with the development and progression of OA. Older adults who were 50-79 years old were selected randomly with an equal number of men and women from the roll of electors in southern Tasmania. Muscle strength measurements, including hand grip, knee extension, and leg muscle strength were conducted at baseline, 2.6-, 5-, and 10-year follow-up time points. Blood samples were collected at the 2.6-year follow-up point after at least 8 hours fasting, and then serum was separated and metabolically profiled using the TMIC Prime Metabolomics Profiling Assay which measures a total of 143

metabolites. Then, the generalized linear mixed effects model with restricted maximum likelihood (REML) method implemented in R package nlme was used to identify the metabolites that were associated with the longitudinal reduction of hand grip strength, knee extension, and leg strength over 10 years with adjustment for age, sex, and BMI. Significance level was defined at α=0.0004 after correcting multiple testing of 129 metabolites with Bonferroni method. Further, a GWAS analysis on the identified metabolomic markers was performed in 77 individuals who had both GWAS and metabolomic data available from the NFOAS to explore the potential mechanisms of the association between the metabolomic markers and the longitudinal reduction of muscle strength over 10 years.

Results: A total of 409 older adults (50% of them were females) were included in this study. Study participants had a mean age of 60.93□6.50 years, and mean BMI of 27.12□4.18 kg/m^2 at baseline. BMI did not change significantly (p=0.06) over the 10 years follow-up period. Although hand grip, knee extension, and leg strength measurements in males were significantly higher than in females (p=0.0001) at baseline, there was no significant difference (p=0.24) in muscle strength reduction between males and females over the 10 years follow-up period. Muscle strength declined by 0.09 psi, 0.02 kg, and 2.57 kg per year for hand grip, knee extension, and leg strength, respectively. Among the 143 metabolites measured, 129 passed the quality control filtering and were included in the subsequent analysis. We found that the elevated level of asymmetric dimethylarginine (ADMA) was significantly associated with the reduction of average

hand grip strength over 10 years (beta= -0.13 psi/per year per log μM, SE= 0.05; p=0.0003). The total dimethylarginine was the second associated metabolite for hand grip strength reduction but the p-value did not reach the pre-defined significance level (p=0.0005). The increased concentration of this metabolite was associated with a decline in average hand grip strength (beta=-0.12 psi/per year per log μM, SE=0.05). This suggests ADMA but not the symmetric dimethylarginine to play a role in hand grip strength reduction. ADMA was also the top metabolite associated with the reduction in knee extension strength (beta= -0.13 kg/per year per log μM, SE= 0.05; p=0.008). A similar trend was found for the reduction in leg strength but was not statistically significant. The GWAS analysis found that SNP rs1125718 was associated with ADMA concentration at the GWAS significance level ($p=4.394*10^{-8}$). This SNP is in the gene desert region on chr8 but adjacent to the *WNT1 Inducible Signaling Pathway Protein 1* (*WISP1*) gene that plays important roles in maintenance of the extracellular matrix of connective tissues. Although it did not reach the GWAS significant, the second most associated SNP rs816296 ($p=2.03*10^{-06}$) on chr12 is very interesting as it is in intron 2 of the *nitric oxide synthase 1* (*NOS1*) gene which is expressed abundantly in skeletal muscle tissues and plays a role in synthesizing nitric oxide from arginine. Further, we found that the increased serum concentration of uric acid was significantly associated with the decline of leg strength over 10 years (beta= -0.64 kg/per year per log μM, SE= 0.17; p = 0.0001) but not with hand grip or knee extension strength. While they did not reach the pre-defined significance level, methionine, creatinine, and diacyl-phosphatidylcholines with 32 carbons and two double bonds (PC aa C32:2) were found to be associated with

the reduction of muscle strength over 10 years with $p<0.01$, warranting further investigation.

Conclusion: Our results demonstrated that elevated serum concentrations of dimethylarginine, especially ADMA, and uric acid at baseline were significantly associated with age-dependent muscle strength reduction which may make study participants more susceptible for the development and progression of OA. While confirmation is needed, these findings provided new insights into the pathogenesis of age-related muscle strength decline and novel targets for developing strategies to prevent muscle strength loss over time.

4.4 Introduction:

OA is an age-related multifactorial degenerative condition that damages the whole joint, with pathological changes to multiple musculoskeletal tissues including cartilage, meniscus, ligament, and synovium (494). It is characterised by deterioration of articular cartilage, subchondral bone sclerosis, and the formation of osteophytes at the joint margin and synovitis (495).

Skeletal muscles are essential components of the neuromuscular skeletal system that has an integral role in the biomechanical structure and function of the synovial joints (496). Skeletal muscles attach to the bone by tendons and produce all body movements. They also resist gravity to absorb loading and maintain posture (497,498). Small, constant

adjustments of the skeletal muscles are needed to hold a body upright or balanced in any position. Muscles also prevent excess movement of the bones and joints, maintain skeletal and joint stability (499), and prevent skeletal structure damage or deformation (500,501).

Muscle weakness and atrophy is one of the earliest signs of aging and an important geriatric condition (172). As we get older, loss in skeletal muscle mass and strength increases (502). Previous studies reported that after the age of 30 muscle mass decline about 3–8% per decade and this rate of reduction is even higher after the age of 60 (503,504). Although, the true mechanisms that lead to muscle strength reduction onset and progression are not entirely understood, various age-related mechanisms including endocrine dysfunctions, neurodegenerative diseases, insufficient nutrition, and muscle disuse and immobility may contribute to the decline of muscle strength (505). Also, mitochondrial dysfunction, oxidative stress, pro-inflammatory status, or metabolic inefficiencies are certainly implicated in the muscle weakness and wasting in older individuals (506). Age-related muscle weakness has been associated with several health and socioeconomical consequences (507). It leads to higher risk of falls, fractures (178), loss of function, and disability in older adults (508). This involuntary loss of muscle mass, strength, and function is a fundamental cause and contributor to disability in older people (504).

Although, limited number of studies have been conducted to investigate the contribution of muscle strength to the pathogenesis of OA, the majority of the performed researches detected strong correlation between muscle weakness and initiation of OA (509–511). For

example, Cicuttini *et al.* (2005) reported that loss of muscle mass and strength was associated with an increased loss of medial and lateral tibial cartilage over two years (512). Another study identified that hip OA patients had lower limb muscle strength and volume deficits. Also, our recent study on endotypes of OA (Chapter 3) showed that muscle weakness indicated by an elevated C4 serum level might be responsible for a subset of OA patients (513). Thus, understanding the potential mechanisms of the age-related muscle strength reduction would provide an avenue to develop OA intervention strategies and improve the quality of life in older adults.

Since metabolites are considered to be the intermediates and end products of cellular processes that affect or are affected by a set of biological systems, genetics, lifestyle, and environmental changes, their concentrations provide a functional information about the physiological state of the human observed phenotypes (514). Recent advances in the metabolic analysis offered new opportunities to measure diverse cell or body fluid metabolites, and it greatly improved our knowledge about the molecular mechanisms underlying metabolism and corresponding human traits and diseases (515). Thus, we undertook this study to investigate the metabolomic signatures for the longitudinal reduction of muscle strength over 10 years in a well-established community based TASOAC Study.

4.5 Materials and Methods:

4.5.1 Study participants:

This study was conducted as part of the TASOAC Study, a prospective, population-based study aimed at identifying the environmental, genetic, and biochemical factors associated with the development and progression of OA (516). Ethics approval was obtained from the Southern Tasmanian Health and Medical Human Research Ethics Committee. Older adults who were 50-79 years old at recruitment were selected randomly with an equal number of men and women from the roll of electors in southern Tasmania, Australia and completed a written informed consent at the baseline between March 2002 and September 2004 (516).

4.5.2 Demographic information:

Demographic, joint symptoms, and daily physical activities information were obtained by a self-administered questionnaire, and anthropometric data including height and weight were measured at subjects' clinical interview (516), **Appendix F** contains a copy of the interview questionnaire. Participants' age at baseline was calculated by subtracting the birth date from the time of recruitment, and then divided by 365 to get the numbers of years. BMI was calculated by dividing weight in kilograms by squared height in meters at the baseline, and 10 years follow-up.

4.5.3 Muscle strength measurements:

Hand grip, knee extension, and leg muscle strength measurements were performed upon baseline, 2.6-, 5-, and 10-year follow-up time points. Hand grip strength was measured using a pneumatic bulb dynamometer (North Coast^{TM} bulb dynamometer; adult 0-30 psi, model no.70154). Participants were seated straight on a chair and held the dynamometer with their elbow at a 90□angle, and their opposite arm resting on their lap (517,518), **Figure 4.1A**. Then, the participants were asked to squeeze as hard as they could for three seconds. The test was conducted twice for each hand interchangeably, with a 30 second rest between trials. Measurements were taken twice for each hand, and the mean score to the nearest pounds per square inch (psi) was used in the analyses (517).

Knee extension strength of the dominant leg was measured by isometric contraction of knee extensors to the nearest kilogram (kg) (518). Subjects were seated on a custom dynamometer chair having a 100 kg pocket balance connected to the back of the chair with their hips and knees at 90° angle. They were asked to keep their backs straight and grip the chair throughout the test. A strap was placed 10 cm above participants' lateral malleolus and attached to the dynamometer that recorded maximum contractile force whilst participants attempted to extend their leg, **Figure 4.1B**. Measurement was taken twice and the average was used in the analyses (518).

Leg muscle strength was measured to the nearest kg in both legs simultaneously using a dynamometer (TTM Muscular Metre, Tokyo, Japan) (518). Study subjects were seated on

the back of a dynamometer platform while their knees were flexed by 115° angle, and their backs were rested on a wall (518). Then, participants were instructed to lift the dynamometer bar that was attached to the dynamometer to the maximum contractile force, using their legs while their head and neck constant, **Figure 4.1C**. Measurement was taken twice, then the best reading was used in the analyses (518).

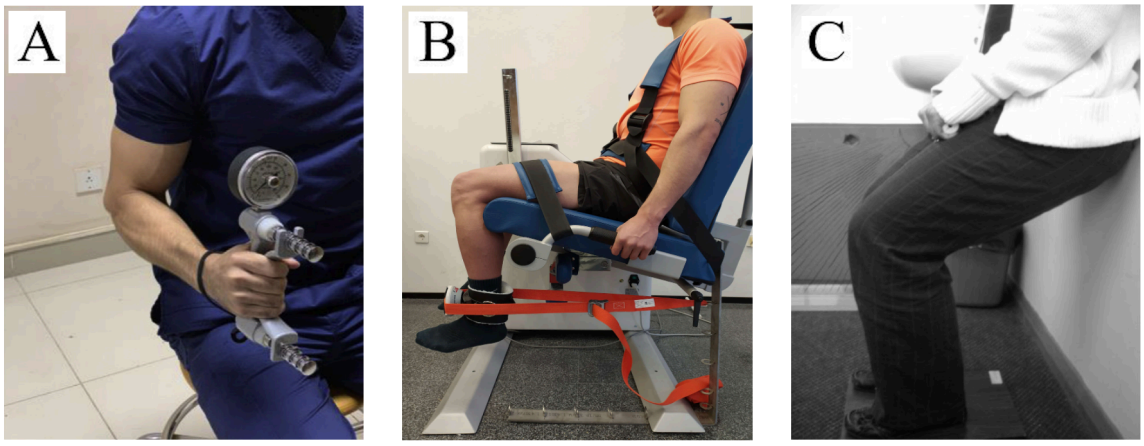


Figure 4.1: Muscle strength measurements including **A)** Hand-grip strength test using the hand-grip dynamometer, participants were seated straight on a chair and held the dynamometer with their elbow flexion at a 90□angle, and their opposite arm resting on their lap. This image was adapted from El-gohary *et al.* (2019) under the Creative Commons CC BY-NC-ND 4.0 License (519). **B)** Knee extension strength measurement using a custom dynamometer chair with participant's hips and knees at 90° angle, their backs straight, and they grip the chair throughout the test. A strap was attached to the dynamometer and placed 10 cm above participants' lateral malleolus to record maximum contractile force whilst participants attempted to extend their leg. This image was adapted from Martín-San Agustín *et al.* (2020) under the Creative Commons CC BY 4.0 License (520). **C)** Leg muscle strength test using a dynamometer. Study subjects stand on the back of a dynamometer platform while their knees flex by 115° angle, and their backs rest on a wall. Then, they lift the dynamometer bar that is attached to the dynamometer to the maximum contractile force, using their legs while their head and neck constant. This image was adapted from Scott et al. (2009), **Appendix G** contains the copyright permission to use this picture.

4.5.4 Metabolic profiling:

Blood samples were collected at 2.6-year follow-up point after at least 8 hours fasting, then the blood was centrifuged at 2500 rpm for 5 minutes, and the serum was transferred into microcentrifuge tubes and stored at -80ºC until analysis. We used the samples collected at this time point because the baseline collected samples were depleted. It was the closest time point to the baseline, thus could be considered as baseline surrogate. Targeted metabolic profiling was performed using the TMIC Prime Metabolomics Profiling Assay which quantifies 143 compounds including 40 acylcarnitines, 25 amino acids and derivatives, 23 biogenic amines, one amine oxides, one carboxylic acid, one monosaccharide, 17 organic acids, 34 phospho-and sphingolipids, and one vitamin and cofactor, **Table 4.1**. The profiling was done at the TMIC using an AB SCIEX QTRAP®4000 mass spectrometer (Sciex Canada, Concord, ON, Canada) equipped with an Agilent 1260 series ultra-high-performance liquid chromatography (UHPLC) system (Agilent Technologies, Palo Alto, CA, USA). The Analyst software 1.6.2 (Concord, ON, Canada) was used to control the entire assay's workflow and the metabolite concentrations were reported in μM. The CV for the metabolites ranged between 1.16-15.93%.

Table 4.1: TMIC Prime Metabolomics Profiling Assay list of 143 metabolite concentrations.

Metabolite class	Number	Metabolite name or abbreviation
Carnitine	1	C0
Acylcarnitine	25	C2, C3, C3:1, C4, C4:1, C5, C5:1, C6(or C4:1-DC), C6:1, C8, C9, C10, C10:1, C10:2, C12, C12:1, C14, C14:1, C14:2, C16, C16:1, C16:2, C18, C18:1, C18:2
Hydroxy- and dicarboxyacylcarnitines	14	C3-OH, C4-OH(or C3-DC), C5:1-DC, C5-DC(or C6-OH), C5-M-DC, C5-OH(or C3-DC-M), C7-DC, C12-DC, C14:1-OH, C14:2-OH, C16:1-OH, C16:2-OH, C16-OH, C18:1-OH
Amino Acids	22	Alanine, Arginine, Asparagine, Aspartate, Citrulline, Glutamine, Glutamate, Glycine, Histidine, Isoleucine, Leucine, Lysine, Methionine, Ornithine, Phenylalanine, Proline, Serine, Threonine, Tryptophan, Tyrosine, Valine, Betaine
Amino Acid Derivatives	3	Creatine, Phosphocreatine, Methylhistidine
Biogenic Amines	23	Acetylornithine, Asymmetric dimethylarginine, Total dimethylarginine, Alpha-Aminoadipic acid, Carnosine, Creatinine, Dihydroxyphenylalanine (DOPA), Dopamine, Histamine, Kynurenine, Methioninesulfoxide, Hydroxyproline (c4-OH-Pro), Hydroxyproline (t4-OH-Pro), Nitrotyrosine, Phenylethylamine, Putrescine, Sarcosine, Serotonin, Spermidine, Spermine, Diacetylspermine, Taurine, Tyramine
Amine Oxide	1	Trimethylamine N-oxide (TMAO)
Carboxylic Acid	1	Homocysteine
Monosaccharides	1	Glucose
Organic Acids	17	Lactic acid, Beta-hydroxybutyric acid, Alpha-ketoglutaric acid, Citric acid, Butyric acid, Propionic acid, HPHPA, Para-hydroxyhippuric acid, Succinic acid, Fumaric acid, Pyruvic acid, Isobutyric acid, Hippuric acid, Methylmalonic acid, Homovanillic acid, Indole acetic acid, Uric acid
Diacyl-phosphatidylcholines	8	PC aa C32:2/ C36:0/ C36:6/ C38:0/ C38:6/ C40:1/ C40:2/ C40:6
Acyl-alkyl-phosphatidylcholines	2	PC ae C36:0/ C40:6
Lyso-phosphatidylcholines	14	LysoPC a C14:0/ C16:0/ C16:1/ C17:0/ C18:0/ C18:1/ C18:2/ C20:3/ C20:4/ C24:0/ C26:0/ C26:1/ C28:0/ C28:1
Sphingomyelines	5	SM C22:3, SM C24:0, SM 24:1, SM C26:0, SM C26:1

Hydroxysphingomyelines	5	SM (OH) C14:1, SM (OH) C16:1, SM (OH) C22:1, SM (OH) C22:2, SM (OH) C24:1
Vitamins & Cofactors	1	Choline

C: carbon; DC: decarboxyl; M: methyl; OH: hydroxyl; TMAO: Trimethylamine N-oxide; HPHPA: 3-(3-hydroxyphenyl)-3-hydroxypropionic acid; PC: phophatidylcholine; aa: acyl-acyl; ae, acyl-alkyl; lysoPC: lysophosphatidylcholine; SM: sphingomyelin.

4.5.5 Statistical analysis:

Metabolomic and muscle strength data underwent a set of quality control filtering. Metabolites with missing values or with concentration below the LOD in more than 10% of the samples were removed from the subsequent analysis to minimize the false positive results as a standard practice in metabolomics studies (456). For metabolites with missing values or having concentration below the LOD in less than 10% of samples, values were imputed by the mean of the given metabolites. Then, log transformation was used to make the metabolite concentration data less skewed and normally distributed for the subsequent analysis. The average hand grip strength was calculated from the right- and left-hand grip strength measurements and was used in the analysis. Then, the metabolomic and muscle strength datasets were combined and transformed from wide to long format using long_panel function from the panelr package in R. Subsequently, generalized linear mixed effects model with REML method implemented in R package nlme (522) was used to identify the metabolites that were associated with the longitudinal reduction of hand grip strength, knee extension, and leg muscle strength over 10-years. An interaction term between each metabolite and a follow-up time variable was also introduced into the multiple regression models as a predictor for longitudinal changes in muscle strength, and the beta coefficient for the interaction term was interpreted as the rate of muscle strength change per year over the follow-up time in relation to a given metabolite concentration unit.

Random effects of the y-intercept (sample-ID) and slope (muscle strength reduction over the follow-up time) were utilized to account for the excess variation implicit in the study design between and within study subjects, respectively. The analyses were adjusted for age, sex, and BMI as potential confounders, as shown in the below linear mixed effects model equation of our analyses:

```
library(nlme)
lme(Muscle Strength ~ follow-up time + Sex + Age +  BMI + Metabolite +
(follow-up time* Metabolite), data=TASOAC_Data, random=~ follow-up
time|SampleID, control=list(opt = "optim"), method= "REML",
na.action=na.omit)
```

Then, the significance level was defined at $\alpha=0.0004$ after correcting for multiple testing of 129 metabolites with Bonferroni method.

Since the metabolomics profiling was conducted at the 2.6-years follow-up point only, and the age range of the study participants spanned ~30 years, additional analysis using a multiple linear regression was performed to investigate the cross-sectional association of the identified metabolites with the muscle strength measurements at the 2.6-years follow-up phase across individuals of different ages.

Further, a GWAS was performed on 77 individuals from the NFOAS whose metabolomic and genome-wide SNP genotyping data were available from previous studies (513,523).

The GWAS analysis was conducted to explore the potential mechanisms of the association between the metabolomic markers and the longitudinal reduction of muscle strength over 10-years using the commonly accepted GWAS significance threshold ($p<5*10^{-8}$).

4.6 Results:

A total of 409 subjects (50% females) were included in this study. The subjects were followed up for three-time phases with the mean follow-up time of 2.60□0.40, 5.06□0.48, and 10.73□0.67 years, respectively. Mean age was 60.93□6.50 years **(Figure 4.2)** and mean BMI was 27.12□4.18 kg/m^2 at baseline. Males were older than females (p=0.02) but there were no significant differences in BMI between males and females at baseline, as well as at each of the follow-up phases, **Table 4.2**. Further, BMI did not change significantly over the 10-years for both sexes (p=0.06).

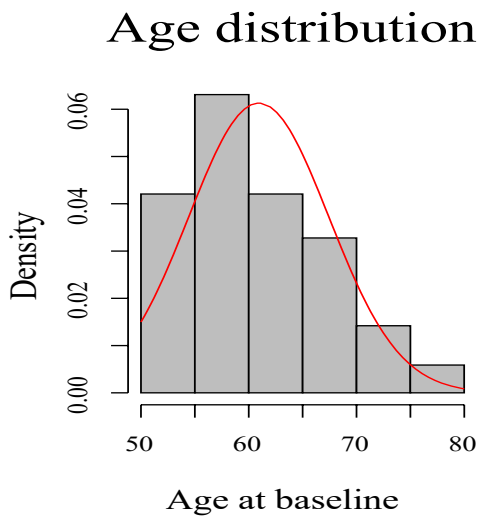


Figure 4.2: A histogram showing the 409 study participants age distribution at baseline.

Table 4.2: The characteristics of the study participants (n=409).

	Male	Female	P-value
Sex (number (%))	205 (50.12)	204 (49.88)	1
Age (years)	64.27□□.91	62.79□□.94	0.02
Baseline BMI (kg/m^2)	27.12□3.54	27.12□4.74	0.99
2.5-year BMI (kg/m^2)	27.18□3.64	28.59□16.99	0.25
5-year BMI (kg/m^2)	27.44□□□□	28.96□17.29	0.22
10-year BMI (kg/m^2)	27.57□4.10	27.64□5.44	0.87

Values are mean ± SD for continuous variables and percentage for sex. P-values were obtained from Chi squared test for sex distribution and Student's t-test for continuous variables. BMI: body mass index.

Similar to other longitudinal analysis, our study had few individuals that did not complete all four follow-up muscle strength measurements. **Table 4.3** lists the numbers and percentages of the study participants that completed the hand grip, knee extension, and leg muscle strength measurements at the baseline and each of the follow-up time points.

Table 4.3: The number and percentage of the study participants that completed the muscle strength measurements at each of the follow-up time points.

	Baseline number (%)	2.6-years point number (%)	5-years point number (%)	10-years point number (%)
Hand grip	409 (100)	408 (99.76)	408 (99.76)	407 (99.51)
Knee extension	408 (99.76)	407 (99.51)	407 (99.51)	403 (98.53)
Leg strength	399 (97.56)	396 (96.82)	386 (94.38)	377 (92.18)

Our data showed that hand grip declined by 0.09 psi per year (p=0.0002), and leg muscle strength decreased by 2.57 kg per year ($p=8.49*10^{-15}$), while the decrease in knee extension of 0.02 kg per year was not statistically significant (p=0.58), estimated by the generalized linear mixed affects model **(Figure 4.3; Table 4.4)**.

Furthermore, our findings showed a significant association between higher age and lower hand grip ($p=1*10^{-15}$), knee extension ($p=4.5*10^{-15}$), and leg muscle strength ($p=3.4*10^{-12}$) at the baseline (**Figure 4.4)**.

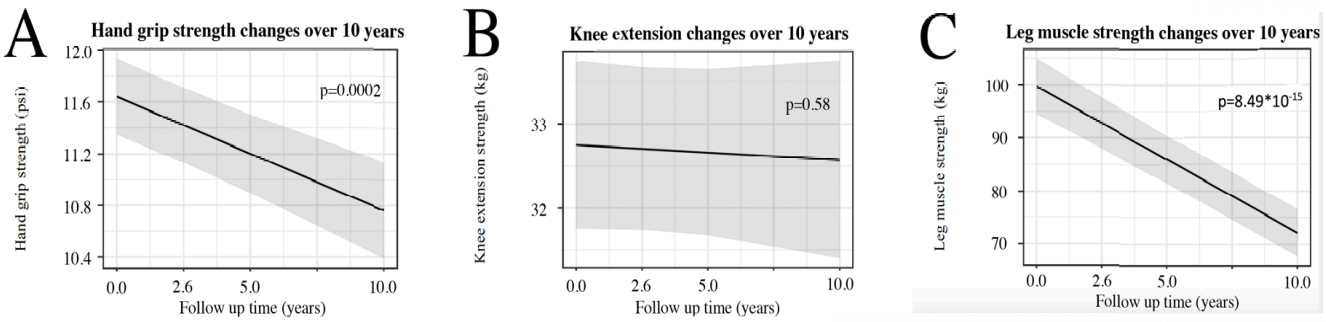


Figure 4.3: Plots of the fitted lines for changes over 10-year follow-up time for **A)** hand grip strength, **B)** knee extension, and **C)** leg muscle strength estimated by linear mixed regression model done with nlme R package with the function lme(Muscle Strength ~ follow-up time, data=TASOAC_ Data, random=~ follow-up time|SampleID, control=list(opt = "optim"), method= "REML", na.action=na.omit). The gray shaded areas are 95% CI.

Table 4.4: Muscle strength differences between males and females over 10-years follow-up period in whole cohort (n=409).

		Whole dataset	Male	Female	Male VS Female P-value
Hand Grip (psi)	**Baseline**	12.3±2.99	14.4±2.53	10.3±1.77	$2.76*10^{-14}$
	2.6-years	10.1±2.93	11.9±2.59	8.43±2.15	$1.03*10^{-4}$
	5-years	11.7±4.35	13.9±4.36	9.37±2.92	$2.13*10^{-5}$
	10-years	10.8±3.64	12.8±3.62	8.91±2.45	$3.12*10^{-4}$
decline/year (psi/year)		0.09	0.09	0.09	
P-value		0.0002	$1.2*10^{-14}$	$1.46*10^{-14}$	
Knee extension (kg)	**Baseline**	31.55±11.3	38.2±9.19	24.8±9.12	$2.74*10^{-18}$
	2.6-years	33.2±11.0	38.8±9.17	27.6±9.71	$2.36*10^{-12}$
	5-years	34.3±10.3	39.5±8.69	29.2±9.2	$1.52*10^{-14}$
	10-years	31.54±11.3	38.5±12.6	24.7±9.32	$8.28*10^{-7}$
decline/year (Kg/year)		0.02	0.002	0.067	
P-value		0.58	0.50	0.20	
Leg strength (kg)	**Baseline**	96.0±52.6	132±46.3	59.6±27.7	$1.82*10^{-4}$
	2.6-years	96.7±53.4	133±46.5	60.1±29.5	$5.36*10^{-4}$
	5-years	87.3±54.4	124±47.6	50.1±30.1	$3.48*10^{-5}$
	10-years	70.4±43.9	95.3±43.6	45.4±26.8	$3.71*10^{-4}$
decline/year (Kg/year)		2.57	3.68	1.47	
P-value		$8.49*10^{-15}$	$2.0*10^{-16}$	$2.09*10^{-8}$	

Values are mean ± SD for muscle strength measurement at baseline, 2.6-, 5-, and 10-years follow-up phases. P-values for muscle strength differences between males and females at each follow-up points were obtained by using the linear regression (lm) method in R; psi: pounds per square inch; kg: kilogram.

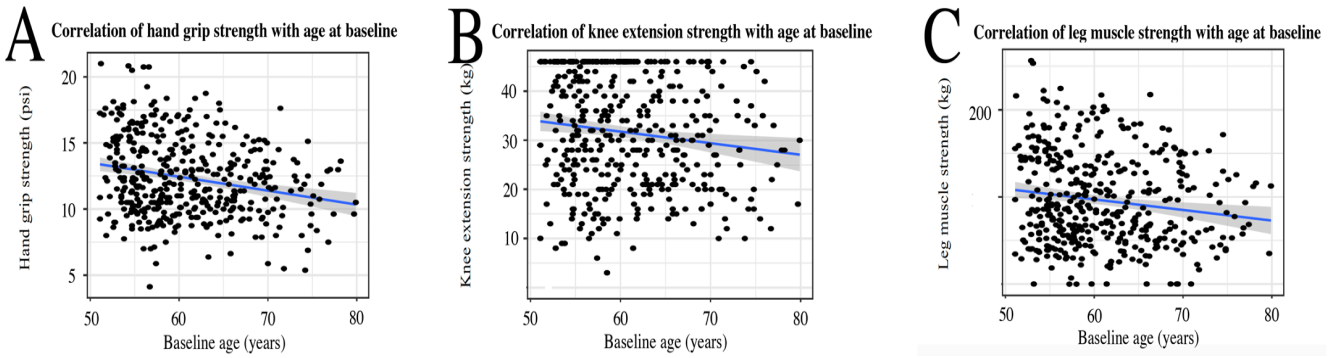


Figure 4.4: The relationship between muscle strength and the study participants' age at the baseline, **A**) hand grip strength, **B**) knee extension strength, and **C**) leg muscle strength changes.

Although all muscle strength measurements in males were significantly higher than that in females over the 10-years follow-up time ($p=2*10^{-15}$), at the baseline, and at each of the follow-up phase (all $p=5.36*10^{-4}$, **Table 4.4**), there was no significant difference ($p=0.24$) in the muscle strength change rate between males and females over the 10-years follow-up period, **Figure 4.5**. Interestingly, higher BMI in the 409 study participants was associated with higher knee extension strength ($p=0.036$).

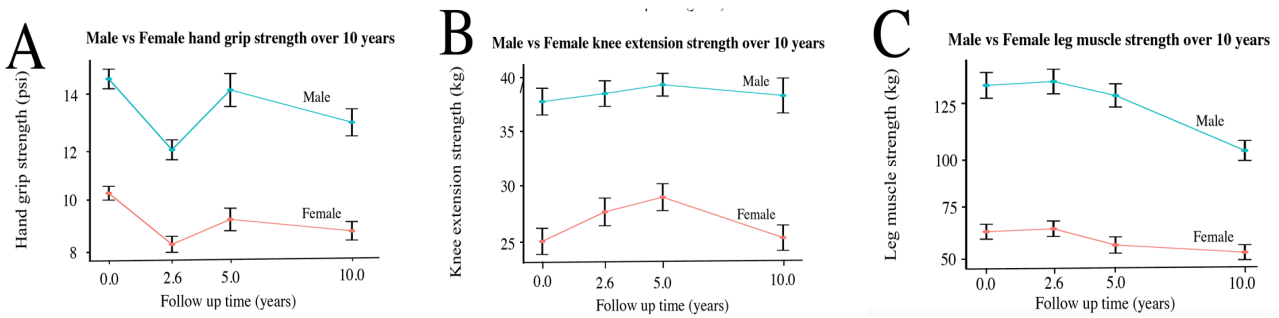


Figure 4.5: The differences between males and females in muscle strength changes over 10-year follow-up time: **A**) hand grip strength, **B**) knee extension strength, and **C**) leg muscle strength. *Bars represent 95% CI.

Among the 143 metabolites measured, 129 passed the quality checks and were included in the subsequent analysis, **Table 4.5**. The volcano plot in the **Figure 4.6A** shows the results of the association between the reduction of the hand grip strength and each of the 129 metabolites. While four metabolites had $p<0.05$, only one metabolite - asymmetric dimethylarginine (ADMA) was associated with the reduction in the hand grip strength at the pre-defined significance level ($p=0.0003$). Per log µM increase in ADMA was associated with 0.05±0.02 psi/year reduction rate in the hand grip strength **(Table 4.6)**. The total dimethylarginine was the second top metabolite associated with the hand grip strength reduction, but the P-value ($p=0.0006$) did not reach the pre-defined significance, **Figure 4.6A**. Per log µM increase of this metabolite was associated with 0.05±0.01 psi/year reduction rate in the hand grip strength**, Table 4.6**. Taurine (beta=0.03±0.01 psi/year per log µM; $p=0.015$) and lactic acid (beta=0.03±0.01 psi/year per log µM; $p=0.04$) were also potentially associated with the reduction in hand grip strength over 10-years, **Table 4.6**.

Table 4.5: Summary statistics for the 129 metabolites that passed the quality checks and were included in the analysis.

#	**Metabolites**	**Mean±SD**	**Min**	**Max**
1	Creatinine	95.25±30.52	27.7	212
2	Glycine	323.39±96.86	161	892
3	Alanine	496.25±113.63	268	970
4	Serine	136.28±25.42	68.5	229
5	Proline	222.82±65.68	85.4	516
6	Valine	267.14±52.76	139	465
7	Threonine	174.27±39.69	83.9	338
8	Taurine	113.89±35.71	46.4	237
9	Putrescine	0.26±0.13	0.0521	0.972
10	trans-Hydroxyproline	13.83±6.30	5.11	40.9
11	Leucine	173.90±40.50	92.7	317
12	Isoleucine	80.06±18.20	39.4	134
13	Asparagine	46.81±10.29	28.1	93.5
14	Aspartic acid	16.33±6.61	3.02	51.8
15	Glutamine	751.19±124.18	419	1230
16	Glutamic acid	58.72±25.80	12.8	164
17	Methionine	32.27±6.60	18.6	53.6
18	Histidine	109.68±29.18	62.6	511
19	alpha-Aminoadipic acid	0.71±0.36	0.0271	3.52
20	Phenylalanine	84.07±15.04	55.3	194
21	Methionine-sulfoxide	0.88±0.36	0.183	2.19
22	Arginine	156.79±33.97	88.2	283
23	Acetyl-ornithine	0.95±0.79	0.0669	10.5
24	Citrulline	44.24±12.34	9.75	141
25	Asymmetric dimethylarginine	0.72±0.76	0.331	14.3
26	Serotonin	1.27±0.73	0.0302	4.71
27	Tyrosine	90.09±22.16	45.1	170
28	Kynurenine	3.08±1.06	1.05	7.27
29	Total dimethylarginine	2.10±0.84	1.14	13.7
30	Tryptophan	75.71±15.32	29.1	131
31	Ornithine	112.97±26.41	52	217
32	Lysine	257.92±61.44	135	444

#	**Metabolites**	**Mean±SD**	**Min**	**Max**
33	Sarcosine	1.78±0.84	0.203	7.21
34	Spermidine	0.18±0.07	0.0249	0.458
35	Spermine	0.13±0.04	0.0498	0.436
36	Creatine	42.33±22.19	8.79	131
37	Betaine	42.19±15.47	12.2	107
38	Choline	12.29±3.40	4.87	27
39	Trimethylamine N-oxide	8.86±9.57	0.432	93.6
40	Methylhistidine	15.92±10.69	3.72	70.8
41	Lactic acid	1778.59±585.23	516	4100
42	beta-Hydroxybutyric acid	89.18±98.63	13.2	736
43	alpha-Ketoglutaric acid	8.46±2.42	4.06	29.3
44	Citric acid	144.94±43.19	66.4	382
45	Butyric acid	0.72±0.32	0.161	2.48
46	Propionic acid	2.20±14.62	0.13	297
47	HPHPA	0.18±0.21	0.0101	1.44
48	Succinic acid	2.04±0.49	1.1	3.64
49	Fumaric acid	1.29±0.40	0.579	3.12
50	Pyruvic acid	58.44±20.35	13.7	149
51	Isobutyric acid	0.60±0.19	0.201	1.87
52	Hippuric acid	6.12±5.09	0.59	51.6
53	Methylmalonic acid	0.19±0.14	0.0247	1.73
54	Indole acetic acid	2.34±1.83	0.255	19.6
55	Uric acid	341.78±82.41	144	572
56	Glucose	5234.08±991.52	3571	13425
57	LysoPC a C14:0	2.90±0.94	1.0659	6.9832
58	LysoPC a C16:0	74.76±19.26	36.1693	164.2629
59	LysoPC a C16:1	3.03±1.16	0.9811	12.9972
60	LysoPC a C17:0	1.86±0.70	0.5198	5.8315
61	LysoPC a C18:0	22.87±6.76	9.0653	50.5419
62	LysoPC a C18:1	18.17±6.72	6.8234	79.2064
63	LysoPC a C18:2	22.07±8.21	7.3367	62.6617
64	LysoPC a C20:3	2.85±1.02	0.9418	10.4587
65	LysoPC a C20:4	4.54±1.68	1.6613	14.6625
66	LysoPC a C24:0	0.10±0.03	0.0314	0.2355
67	LysoPC a C26:0	0.11±0.04	0.0328	0.2671
68	LysoPC a C26:1	0.05±0.02	0.0097	0.1206

#	**Metabolites**	**Mean±SD**	**Min**	**Max**
69	LysoPC a C28:0	0.21±0.07	0.0674	0.588
70	LysoPC a C28:1	0.27±0.10	0.0621	0.6177
71	SM C16:0	130.43±30.02	60.8714	231.5374
72	SM C16:1	19.36±4.82	8.3687	37.9455
73	SM C18:0	32.40±8.14	15.4762	59.3029
74	SM C18:1	14.07±4.20	5.185	28.7409
75	SM C20:2	1.67±0.52	0.6583	4.0601
76	SM (OH) C14:1	9.43±2.81	3.2538	18.4442
77	SM (OH) C16:1	5.62±1.60	1.8069	11.2314
78	SM (OH) C22:1	18.99±5.17	6.663	41.0485
79	SM (OH) C22:2	17.04±4.74	6.7205	35.1672
80	SM (OH) C24:1	2.84±0.79	1.0599	6.3808
81	PC aa C32:2	6.74±1.95	2.4347	14.5156
82	PC aa C36:0	6.85±1.92	2.8197	14.6563
83	PC aa C36:6	1.20±0.51	0.275	4.2165
84	PC aa C38:0	2.96±1.17	0.9397	8.4584
85	PC aa C38:6	73.95±27.16	22.1437	224.4749
86	PC aa C40:1	0.25±0.07	0.1151	0.5082
87	PC aa C40:2	0.34±0.08	0.1573	0.6621
88	PC aa C40:6	22.01±8.43	5.7374	62.3499
89	PC ae C36:0	1.70±0.47	0.5595	3.4495
90	PC ae C40:6	4.37±1.43	1.8528	9.777
91	C0	54.08±13.14	21.6175	98.7145
92	C2	12.02±4.22	3.172	30.0657
93	C3	0.37±0.12	0.1308	1.0289
94	C3:1	0.01±0.00	0.0017	0.0292
95	C4	0.29±0.13	0.0801	1.2048
96	C4:1	0.02±0.01	0.0062	0.1398
97	C5	0.15±0.06	0.0586	0.4868
98	C5:1	0.02±0.01	0.0077	0.065
99	C6(or C4:1-DC)	0.08±0.04	0.0276	0.5666
100	C6:1	0.02±0.01	0.0073	0.048
101	C8	0.25±0.22	0.0519	2.9411
102	C9	0.08±0.05	0.0183	0.4074
103	C10	0.57±0.46	0.1078	6.0561
104	C10:1	0.32±0.12	0.125	0.9399

#	Metabolites	Mean±SD	Min	Max
105	C10:2	0.06±0.02	0.0267	0.1799
106	C12	0.14±0.07	0.0439	0.8476
107	C12:1	0.17±0.06	0.0685	0.5337
108	C14	0.06±0.02	0.019	0.2495
109	C14:1	0.20±0.09	0.0686	0.9702
110	C14:2	0.06±0.03	0.0194	0.2514
111	C16	0.15±0.04	0.0504	0.3142
112	C16:1	0.06±0.02	0.0257	0.2616
113	C16:2	0.02±0.01	0.0067	0.0674
114	C18	0.06±0.02	0.0276	0.1372
115	C18:1	0.19±0.06	0.0581	0.4725
116	C18:2	0.06±0.02	0.0164	0.1283
117	C3-OH	0.03±0.01	0.0098	0.0647
118	C4-OH(or C3-DC)	0.04±0.02	0.0163	0.1363
119	C5-OH(or C3-DC-M)	0.03±0.01	0.0118	0.081
120	C5-DC(or C6-OH)	0.01±0.01	0.0034	0.0356
121	C5:1-DC	0.03±0.01	0.011	0.0789
122	C5M-DC	0.03±0.01	0.0109	0.1407
123	C12-DC	0.01±0.00	0.0041	0.016
124	C14:1-OH	0.03±0.01	0.0121	0.0731
125	C14:2-OH	0.02±0.00	0.0053	0.0383
126	C16-OH	0.01±0.00	0.0041	0.0247
127	C16:1-OH	0.02±0.01	0.0084	0.047
128	C16:2-OH	0.01±0.00	0.0055	0.0243
129	C18:1-OH	0.01±0.00	0.0048	0.0368

C: carbon; DC: decarboxyl; M: methyl; OH: hydroxyl; TMAO: Trimethylamine N-oxide; HPHPA: 3-(3-hydroxyphenyl)-3-hydroxypropionic acid; PC: phophatidylcholine; aa: acyl-acyl; ae, acyl-alkyl; lysoPC: lysophosphatidylcholine; SM: sphingomyelin.

The volcano plot in the **Figure 4.6B** presents the results of the association between the change of knee extension strength over time and each of the 129 metabolites. While total

dimethylarginine and ADMA were the top associated metabolites with the reduction of knee extension with $p<0.05$ level, both of them did not reach the pre-defined significance level, **Table 4.6**.

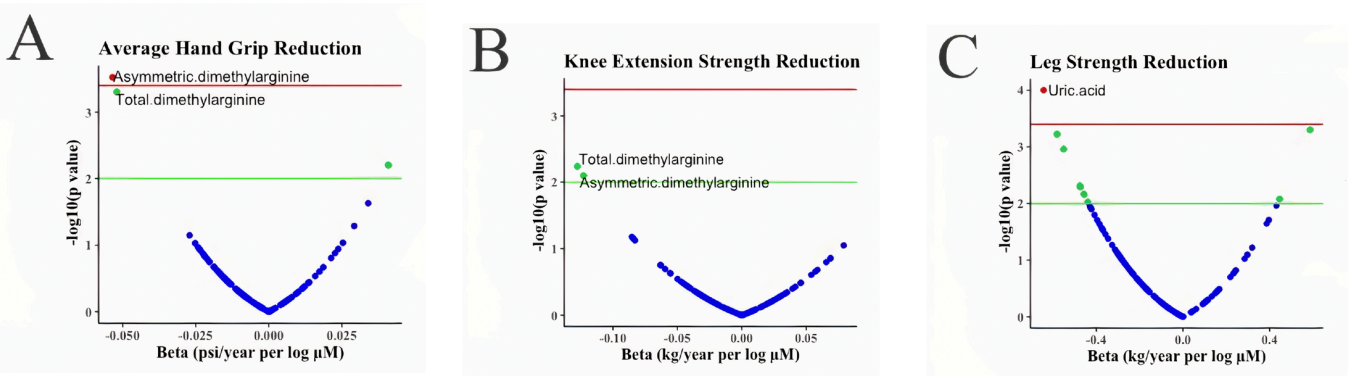


Figure 4.6: Volcano plots of the association results between metabolites and the changes of muscle strength: **A)** hand grip strength change, **B)** knee extension strength change, and **C)** leg muscle strength changes over 10-year follow-up time. X-axis is the effect size in betas obtained from the generalized linear mixed effects model, and y-axis is minus log transformed P-values. The green line indicates P-value = 0.01, and the red line shows the pre-defend significance level at $\alpha=0.0004$ after correction of multiple testing of 129 metabolites with Bonferroni method. A negative value of beta in the X-axis refers to a reduction in muscle strength per unit increase for a given metabolite, while a positive value of beta in the X-axes refers to an increase in muscle strength per unit increase for a given metabolite over the 10-year follow-up time.

Table 4.6: Top metabolites associated with the muscle strength change measures in the study participants (n=409) over 10-years follow up time.

Muscle strength change	(Metabolite*follow-up time)‡	**Beta**	**SE**	**P-value**
A) Hand grip	Asymmetric dimethylarginine	-0.053	0.015	0.0003
	Total dimethylarginine	-0.049	0.014	0.0006
	Taurine	0.033	0.014	0.015
	Lactic acid	0.029	0.014	0.04
B) Knee extension	Total dimethylarginine	-0.119	0.044	0.007
	Asymmetric dimethylarginine	-0.119	0.045	0.008
	Methylmalonic acid	-0.085	0.046	0.07
	C3:1	-0.084	0.046	0.07
C) Leg strength	Uric acid	-0.633	0.157	0.0001
	PC aa 32:2	0.586	0.168	0.0005
	Creatinine	-0.581	0.168	0.0006
	Methionine	-0.551	0.168	0.001

‡An interaction term between metabolite values and follow-up time used as predictors for longitudinal changes in muscle strength. C3:1 - acrylylcarnitine; PC aa 32:2 - phosphatidylcholine acyl-acyl with 32 carbons and 2 double bonds.

The volcano plot in the **Figure 4.6C** presents the results of the association between the reduction in leg muscle strength and each of the 129 metabolites. A total of 24 metabolites were significantly associated with the reduction in leg muscle strength at $p<0.05$ level, but uric acid was the only metabolite that reached the pre-defined significance ($p=0.0001$). Per log μM increase in uric acid was associated with 0.63±0.16 kg/year reduction in leg muscle strength, **Table 4.6**. We also tested the interaction between sex and uric acid which was not statistically significant ($p=0.83$). In addition, diacyl-phosphatidylcholines with 32 carbons and two double bonds (PC aa C32:2;

beta=0.59±0.17 kg/year per log μM; p=0.0005), creatinine (beta=-0.58±0.17 kg/year per log μM; p=0.0006), and methionine (beta=-0.55±0.17 kg/year per log μM; p=0.001) were potentially associated with the reduction rate of leg muscle strength over 10-years follow-up period **(Table 4.6)**.

Further, the increased serum concentration of ADMA and total dimethylarginine were associated with difficulty in run errands and shopping ($p \leq 0.008$), vacuuming ($p \leq 0.001$), and bathing (p=0.0003). The elevated uric acid level was associated with hardship of putting on socks (p=0.04) and climbing up five steps (p=0.046) over 10-years follow-up period.

Since the metabolomic profiling was performed on the serum samples collected at 2.6-year follow-up point, we also performed a cross-sectional association test for the visit at 2.6 years and found that the higher levels of ADMA (p=0.027) and total dimethylarginine (p=0.01) were associated with a lower knee extension, but not with hand grip strength (all p=0.42). There was no cross-sectional association between uric acid and leg muscle strength (p=0.56) at the 2.6-years follow-up point.

Further, we divided the cohort of 409 individuals into three different age groups based on the sample size distribution: younger group included 140 individuals who had age range 51.1-57 years (mean±SD= 54.44±1.52), middle age group included 146 individuals who had age range 57.2-63.9 years (mean±SD= 60.12±1.99), and older group included 123 individuals who had age range 64.1-79.9 years (mean±SD= 69.26±3.89) and examined

whether the identified metabolite associations were stronger in the older group. The results showed that the effect size of ADMA and total dimethylarginine on hand grip and knee extension in the older age group was greater than that in the younger groups, **Table 4.7**. Also, the effect size of the uric acid on leg strength was larger in the middle age group than the younger and the older groups, **Table 4.7**.

Table 4.7: Top metabolites associated with the muscle strength change measures in three different age groups over 10-years follow up time.

		(Metabolite*follow-up time)‡	**Beta**	**SE**	**P-value**
Hand Grip	**Younger Age Group** (51.1-57 years)	Asymmetric dimethylarginine	-0.049	0.13	0.69
		Total dimethylarginine	-0.01	0.05	0.79
	Middle Age Group (57.2-63.9 years)	Asymmetric dimethylarginine	-0.056	0.01	0.0005
		Total dimethylarginine	-0.05	0.02	0.003
	Older Age Group (64.1-79.9 years)	Asymmetric dimethylarginine	-0.13	0.12	0.29
		Total dimethylarginine	-0.03	0.04	0.56
Knee Extension	**Younger Age Group** (51.1-57 years)	Asymmetric dimethylarginine	-0.51	0.41	0.22
		Total dimethylarginine	-0.06	0.16	0.70
	Middle Age Group (57.2-63.9 years)	Asymmetric dimethylarginine	-0.10	0.05	0.03
		Total dimethylarginine	-0.1	0.05	0.06
	Older Age Group (64.1-79.9 years)	Asymmetric dimethylarginine	-0.60	0.34	0.075
		Total dimethylarginine	-0.21	0.12	0.083
Leg Strength	**Younger Age Group** (51.1-57 years)	Uric acid	-0.64	0.26	0.02
	Middle Age Group (57.2-63.9 years)	Uric acid	-0.73	0.3	0.01
	Older Age Group (64.1-79.9 years)	Uric acid	-0.52	0.26	0.05

‡An interaction term between metabolite values and follow-up time used as predictors for longitudinal changes in muscle strength.

The GWAS analysis was performed on the metabolites that reached the pre-defined significance level. **Figure 4.7A** shows the Manhattan plot of the GWAS results for ADMA. We found that the SNP rs1125718 (G>A, with MAF=0.29 on chr8 was associated with elevated ADMA concentrations at GWAS significance level ($p=4.39*10^{-8}$). This SNP is located in the gene desert on chr8 but adjacent to several genes including N-Myc Downstream Regulated 1 (*NDRG1*), *WISP1,* ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 1 (*ST3GAL1*), and Zinc Finger and AT-Hook Domain Containing (*ZFAT*), **Figure 4.7B**. Although it did not reach the GWAS significance, the second most associated SNP rs816296 (C>A, MAF=0.17, $p=2.03*10^{-6}$) on chr12 is located in intron 1 of the *NOS1* gene, **Figure 4.7C**. Data on uric acid was not available in the NFOAS, thus no GWAS analysis was performed for uric acid.

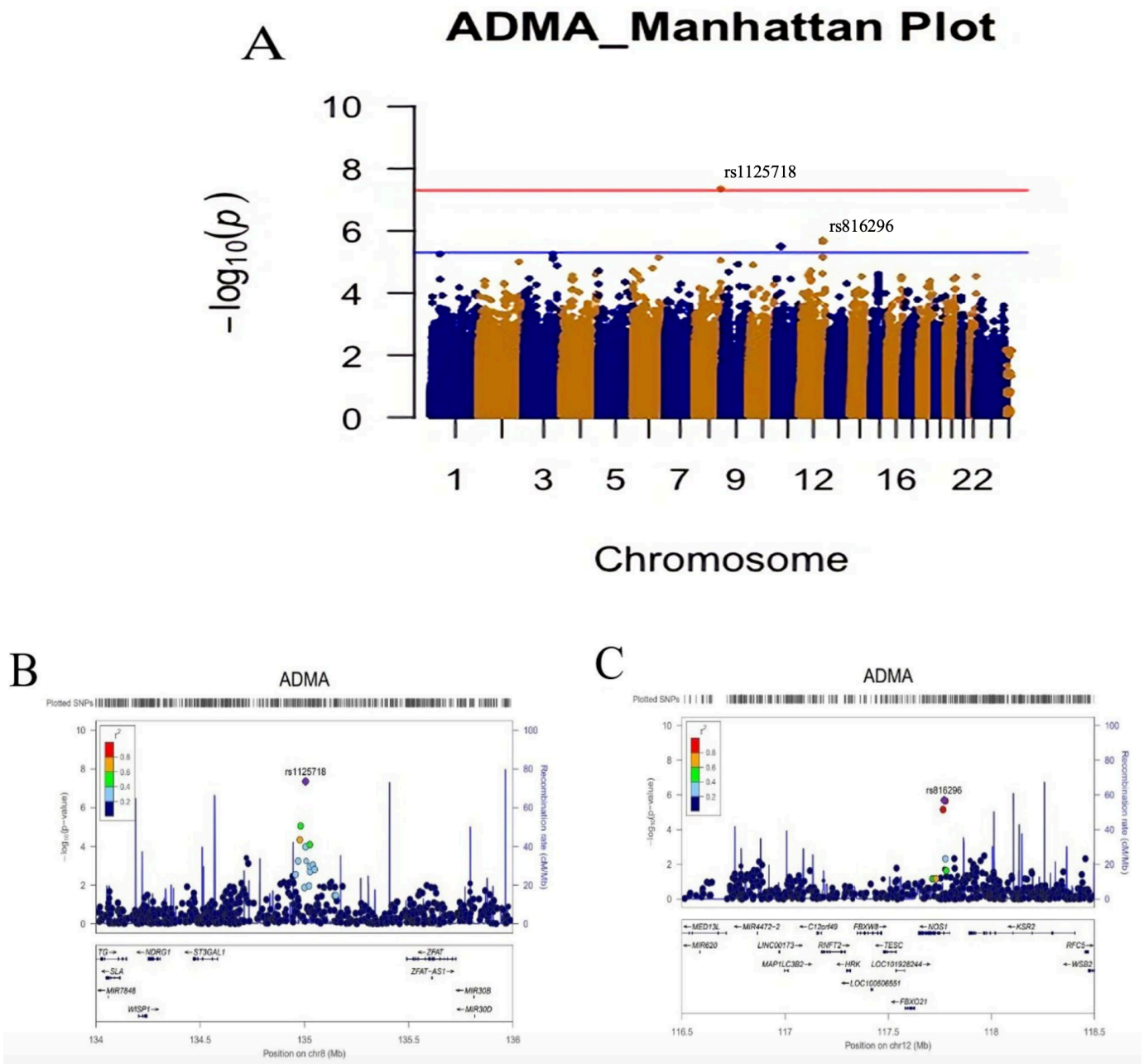


Figure 4.7: GWAS for the asymmetric dimethylarginine (ADMA) in the 77 OA patients from the NFOAS: **A)** a Manhattan plot of the GWAS results. The red line indicates the GWAS significance at $\alpha = 5*10^{-8}$, **B)** the regional association plot for the chromosomal region around the most significant SNP rs1125718 on chromosome 8 ($p=4.394*10^{-8}$) showing adjacent NDRG1, WISP1, ST3GAL1, and ZFAT genes, and **C)** the regional association plot for the chromosomal region around the second most significant SNP rs816296 on chromosome 12 ($p=2.03*10^{-6}$). This SNP is an intronic variant located in intron 1 of NOS1 gene.

4.7 Discussion:

To the best of our knowledge, this was the first population-based study that investigated the relationship between the serum metabolome and the longitudinal reduction rates in hand grip, knee extension, and leg muscle strength over 10-years follow-up period in a large sample size of older adults that were randomly selected from a general population. The reduction rate of the hand grip and leg muscle strength in the current study were comparable with previous studies (524–526), but the reduction rate of the knee extension was lower in the current study than the previous report (527), which might be due to the difference in study populations, follow-up times, study designs, and measurement methods.

We reported that elevated serum concentrations of dimethylarginines, especially ADMA, were significantly associated with the longitudinal reduction rate in the hand grip and knee extension strength. Interestingly, the elevated ADMA and total dimethylarginine blood levels were also associated with functional impairments including run errands and shopping, vacuuming, and bathing over the 10-years follow-up period. We also found that elevated uric acid concentration was significantly associated with the decline rate in leg muscle strength over a 10-years follow-up period. The increased uric acid level was also associated with the longitudinal complication of putting on socks and climbing up five steps over 10-years follow-up period.

However, there is still much more in common between a leg muscle strength and a knee extension, than between hand grip and knee extension, it is interesting that both hand grip and knee extension had negative correlation with ADMA and total dimethylarginine blood levels, while leg muscle strength decline had a significant association with the elevated blood level of uric acid. This might be due to the hand grip and knee extension common functionalities and mechanisms. Wrist and fingers flexion are mostly initiated by the muscles in the anterior and posterior compartments of the forearm (extrinsic muscles), and only the thin tendons of these muscles are found directly in the hand. The flexor tendons of the forearm anterior compartments run in the anterior of the hand through the palms to the tips of the fingers to facilitate flexing of the wrist and fingers leading to wrist flexion and hand grip force production (528). Moreover, the extensor tendons of the forearm posterior compartments used for wrist extension and hand grip relaxation run through the back of the hand to the figures (529). While the extrinsic muscles of the hand are responsible for stronger movements of the wrist and hand, the intrinsic muscles of the hand have no direct effect on wrist action but can contribute to grip force via the extensor mechanism (530). The intrinsic muscles produce finer, more controlled movements and play important roles in rotating fingers toward the palm to maintain and improve hand grip (531). Thus, hand grip strength is relatively specific for muscles in the anterior compartment of the forearm that are involved in finger/wrist flexion. Similarly, while the quadriceps femoris in the anterior compartment of the thigh are activated to extend the knee in the knee extension strength test, the hamstrings in the posterior compartment of the thigh are predominantly involved to flex the knee (532). Also, while the hand grip and knee extension tests are mostly used to assess the upper and lower body's muscle

strength and power, leg muscle strength test is implemented to evaluate the body balance and risk of fall in older adults, because balance consists of multiple body systems including the ability to align different body segments and to generate multi-joint movements to effectively control body position and movement (533). Moreover, as neural decrements present earlier than loss of strength this would likely affect more complex movements more drastically than measures of specific, relatively isolated muscle groups performed in a stable setting. Indeed, the change in leg strength was more pronounced than grip strength.

Data on ADMA and muscle strength are sparse in the literature. In the cross-sectional study of 550 individuals (534), high serum level of ADMA was associated with lower hand grip, quadriceps strengths, and slower gait speed (534). Cancer patients (535) were found to have higher levels of ADMA in the skeletal muscle compared with healthy controls, suggesting that increased levels of ADMA may contribute to impaired muscle protein synthesis in cancer cachexia. In the longitudinal setting, our data documented that the elevated ADMA level was associated with the reduction of muscle strength over time, especially hand grip strength and knee extension. Further studies to investigate the causal relationship between ADMA and muscle strength reduction is warranted. The increased blood concentration of the total dimethylarginine was also associated with the strength reduction in the hand grip and knee extension over the follow-up period. However, the effect size was similar to that of ADMA, suggesting that the association was most likely driven by ADMA rather than symmetric dimethylarginine (SDMA).

Dimethylarginines are products of degraded methylated proteins. Two enzymes - protein arginine methyltransferase type I and II (*PRMT-I, PRMT-II*), are involved in the methylation of arginine residues within proteins or polypeptides with the methyl groups derived from *S*-adenosylmethionine (536). PRMT-I catalyzes the formation of NG-monomethyl-L-arginine (LNMMA) and NG,NG-dimethyl-L-arginine (ADMA) while PRMT-II methylates proteins to release NG,N'G-dimethyl-L-arginine (SDMA) and LNMMA. Free dimethylarginines are released into the cytoplasm during proteolytic breakdown of proteins and can be detected in blood and eliminated from the body by renal excretion (537). ADMA, but not SDMA, is metabolized via hydrolytic degradation to citrulline and dimethylamine by the dimethylarginine dimethylaminohydrolase-1 (DDAH-1), and -2 (DDAH-2) enzymes. Thus, the increased ADMA levels could be due to increased PRMT-I activity, reduced elimination by the kidney, decreased DDAH-1 and 2 enzymatic activities, or a combination. However, our GWAS analysis did not find any association between ADMA and these genes including PRMT-I and DDAH-1 and 2, suggesting that the increased ADMA level may not be genetic. Instead, we found that SNP rs1125718 on chromosome 8 was associated with ADMA concentration at GWAS significance level. This SNP is located in a gene desert and has not been reported to be associated with any disease or traits yet. However, several genes are located in the nearby region including *NDRG1, WISP1, ST3GAL1,* and *ZFAT*. Among them, *WISP1* gene is interesting because a study showed that *WISP1* as fibro-adipogenic progenitors (FAPs)-derived matricellular signal is lost during aging. WISP1 is required for efficient muscle regeneration, and it controls the expansion and asymmetric commitment of tissue-resident muscle stem cells (MuSCs) through Akt signaling (538). Also, previous studies showed

that nitric oxide (NO) level positively correlated with *WISP1* gene expression, and elevated levels of NO increased the *WISP1* mRNA and protein expression levels through a beta-catenin signaling (539). Interestingly, ADMA is known as an endogenous competitive inhibitor of NO synthase (540). Our GWAS analysis showed that the second most associated SNP with ADMA was rs816296 which is located in the intron 1 of the *NOS1* gene. Thus, we hypothesize that possible age-related muscle protein breakdown may lead to an increased release of ADMA which in turn inhibits NO production. The reduced NO synthesis may result in lower expression of WISP1 which leads to the matricellular signals in the skeletal muscle stem cell niche being disturbed (538,541). Hence, the MuSCs number, activity, adhesion, migration, proliferation, self-renewal, and differentiation in skeletal muscle regeneration could be considerably deteriorated leading to the reduction of muscle strength (542–544).

Uric acid is an enzymatic waste end-product from the degradation of purine nucleosides that is renally excreted. Uric acid plays both protective and damaging roles in the skeletal muscles (545), most likely due to its strong antioxidant properties at low levels and pro-inflammatory effect at high levels (541). It has been proposed that oxidative stress might contribute to muscle weakness and wasting. Uric acid at a low level might stabilise the excessive production of free radicals that causes muscle protein damage leading to muscle strength reduction (546). However, at high levels, uric acid stimulates the pro-inflammatory pathway and elevates the production of pro-inflammatory cytokines including IL-1, IL-6, and TNF, which have an impact on muscle mass and function in

aged muscles (547,548). While we did not find a significant cross-sectional association between uric acid and leg muscle strength (p=0.56) at the 2.6-years follow-up phase, we found that there was a positive association between uric acid concentrations and leg muscle strength at the baseline time point. This is consistent with previous studies (545,546). We also documented a negative association between uric acid levels and longitudinal leg muscle strength, consistent with previous studies (547,548). Thus, our findings suggested the importance of maintaining optimal levels of uric acid in the blood for muscle strength (546).

The strength of the current study was its longitudinal nature which allowed us to detect significant metabolite associations for muscle strength changes overtime within an individual. This cannot be achieved in a cross-sectional analysis. Indeed, when we analyzed the data cross-sectionally for the 2.6-years follow-up point, the significance for the identified metabolites became weaker or even non-significant. The current study also underscored the importance of the longer follow-up time with multiple time point measurements as it could minimize the effect of fluctuating variability on the measurements and provide more accurate estimate of changes over time. However, there are a number of limitations in this study. The present study used a commercially available metabolomics assay kit that offers limited coverage of metabolome. Thus, we might miss some metabolites that may contribute to the longitudinal reduction of muscle strength. Also, since metabolomics profiling was performed at only 2.6-years follow-up point, we cannot make any inference regarding the association between the changes in metabolite profiles over time and the muscle strength decline over time. Thus, further studies with

multiple time point metabolomic profiling are needed. However, our study answered an important question whether the baseline metabolic markers can predict muscle strength reduction in 10 years. Loss to follow-up might have influenced our results, especially for leg muscle strength as we had 6 – 8% of missing values at phase3 and phase4 follow-up points. Indeed, those lost to follow-up had a lower leg muscle strength measurement at baseline than those included in the analysis (data not shown). However, there was no difference in uric acid levels between those included and excluded in the final analysis, suggesting that loss to follow-up was unlikely to bias the observed association. We cannot rule out the potential confounding effect of gout on the association between uric acid and leg muscle strength as we did not have data on gout in our cohort. However, gout mostly affects big toes and associated with reduced muscle strength of ankle and foot, not leg muscle strength, suggesting the observed association was less likely to be biased. Finally, our results may not be generalized to populations that have different area-specific socioeconomic indexes and health provisions than that in Tasmania, Australia.

In conclusion, our data demonstrates that baseline elevated serum concentrations of ADMA and uric acid were associated with age-dependent muscle strength reduction. Confirmation of these findings would establish new insights into the pathogenesis of age-related muscle strength decline and uncover novel targets for developing strategies to prevent muscle strength loss over time.

Chapter 5: A genome-wide association study identified novel genes associated with osteoarthritis

Salem Werdyani[1], Ming Liu[1], Hongwei Zhang[2], Guang Sun[2], Zhiwei Gao[3], Andrew Furey[4], Proton Rahman[2], Guangju Zhai[1]

[1]Division of Biomedical Sciences (Genetics),
[2]Discipline of Medicine,
[3]Division of Community Health and Humanities,
[4]Discipline of Surgery,
Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada.

An abstract of this chapter is published in the osteoarthritis and cartilage 2022; 30:1.

5.1 Co-authorship statement:

GZ and SW were responsible for the study conception and design; statistical expertise; analysis, and interpretation of the results. GS, HZ, AF, ML, and GZ participated in specimens; data collecting; and assembly. GS, AF, PR, and GZ were responsible for the provision of study materials or patients. GZ obtained the study funding. ML, PR, GZ provided the administrative, technical, and logistic support.

5.2 Abstract:

Purpose: Osteoarthritis (OA) is the most common form of arthritis and one of the ten most disabling diseases in developed countries. Although the recent largest GWAS-meta-analysis of 13 cohorts from 9 populations with a total of 826,690 individuals identified 100 independent associated risk variants across 11 OA phenotypes, and previous genetic studies in general identified about 140 genetic loci to be associated with OA, these genetic variants can only account for approximately 10% of OA heritability. Thus, we conducted an independent GWAS in the well-established NFOAS which was not included in the previous GWAS analyses to identify novel genetic variants to be associated with OA.

Methods: Study participants were total hip or knee replacement patients due to primary OA who were recruited to the NFOAS before 2017 in St John's, NL, Canada. Self-reported OA-free controls were derived from the same source population who were originally recruited to the CODING study. DNA samples were extracted from whole

blood and genotyped using the Illumina HumanOmni2.58 and Infinium Global Diversity 8 v1.0 genotyping microarrays. Then, pre-association QC filtering and population structure analyses were performed for the study participants based on the genotyping data. Individuals were excluded from the study if they: a) had a heterozygosity rate beyond the mean ±3SD, b) had discordant sex information; c) were non-Caucasian; and d) had call rate < 95%. Also, variants QC filtering excluded SNPs that: a) had call rate <95%; b) had MAF <5%; and c) were deviated from the Hardy Weinberg Equilibrium (HWE, $p<0.0001$). Genome-wide imputation was performed using ShapeIT4 phasing and Impute5 imputation programs for both HumanOmni2.58 and the Infinium GDA 8 v1.0 SNP genotyping datasets separately to predict the genotypes of the un-genotyped SNPs. Subsequently, both OA imputed datasets were merged, and post imputation QC filtering was performed to exclude duplicated samples and eliminate rare SNPs with MAF<0.01 that are deviated from the HWE ($p < 0.0001$). Next, the GWAS analysis was performed using the Mixed Logistic Regression (MLR) implemented in the MilorGWAS package in R version 4.1.2 to investigate the association between the whole-genome genetic variants and OA, knee OA, and hip OA patients.

Results: A total of 557 OA patients (386 knee and 171 hip OA patients) and 118 unaffected controls were included in this study. OA patients were significantly older ($p=0.0001$) and had a higher BMI ($p=2.95*10^{-07}$) than controls, while there was no significant difference in sex distribution of OA patients compared with controls. Association analysis identified 29 genetic variants to be significantly associated with OA in the 557 patients at the genome-wide significance level ($p\leq5*10^{-8}$). Out of the 29

identified variants, 28 were not reported previously. Our analysis detected 11 variants including rs115049241, rs9331225, rs115413462, rs114389542, rs116106396, rs114879738, rs115449966, rs139005472, rs191270495, rs112039851, and rs9697293 in the *Agrin* (*AGRN*) gene on chr1 to be significantly associated (all $p<1.94*10^{-9}$) with OA. The *AGRN* gene play a central role in the formation, maintenance, activation, and synaptic stabilisation of the neuromuscular junction. Moreover, The SNP rs563964217 on chr10 was the most significantly associated (OR=0.05, $p=2.29*10^{-14}$) variant with OA. This SNP is a stop-gain variant (c.CAG>TAG, p.Q>*) in exon 3 of the *GPR123* gene that encodes a protein belonging to the adhesion family of the GPCRs. Two other variants including rs111704643 and rs2298121on chr10 were significantly associated with OA. These variants are ~57 Kbps apart spanning the *ADAM8* gene. Previous studies identified a strong correlation between overexpression of the *ADAM8* gene and the OA pathogenesis. Additionally, SNPs rs200222877 and rs560586208 were significantly associated with OA (all $p< 4.20*10^{-8}$) and locate in the *MUC5B* and *FADS2* genes on chr11, respectively. The *MUC5B* gene is the major gel-forming mucin in mucus, while the *FADS2* gene regulates unsaturation of fatty acids. Likewise, rs72298257, rs57991786, rs1552423, and rs201085636 were significantly associated ($p=2.65*10^{-8}$) with OA. These variants fall in intron 5 of the *TMEM132D* gene on chr12 that is associated with panic-disorder and anxiety related behaviour. Also, SNP rs142897716 was significantly associated ($p= 1.91*10^{-9}$) with OA. This SNP is adjacent to the *MMP17* gene on chr12. *MMP17* gene is involved in breakdown of extracellular matrix in normal physiological processes. Moreover, SNP rs78941287 on chr16 was significantly associated ($p= 4.96*10^{-9}$) with OA. The variant rs78941287 is intergenic that is adjacent to the *CA5A*

and *BANP* genes playing roles in bone resorption and cell cycle regulation. Additionally, the insertion rs569627639 on chr19 was significantly associated ($p= 1.09*10^{-8}$) with OA. This variant is in the *MED16* gene that is involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. Another significantly associated variant with OA on chr20 is rs76236562 ($p= 1.38*10^{-8}$). This SNP is adjacent to the *NTSR1*, *OGFR*, *COL9A3*, *DIDO1*, and the *GID8* genes that are involved in cell signaling, cell proliferation, type IX collagen formation, cell apoptosis, and positive regulation of canonical Wnt signaling pathway. Finally, rs181536033, rs560452957, and rs532097186 on chr22 were significantly associated with OA ($p=1.24*10^{-10}$). These variants are harboring the *RANBP1*, *RTN4R*, *DGCR6L*, and *TMEM191B* genes that have roles in regulating the cell cycle, regulating axonal growth regeneration in the adult central nervous system, cell attachment and migration, and component of cell membrane.

Conclusion: The majority of the identified genetic variants associated with OA in our study were not reported previously and estimated to be specific for the NL population. The identified variants play a critical role in the immune system, pro-inflammatory pathway, OA pain, cartilage degradation, subchondral bone remodeling, and skeletal muscle weakness. While confirmation is required, these findings provided new insights into the pathogenesis of OA and novel targets for developing OA drugs.

5.3 Introduction:

Although the recent largest GWAS-meta-analysis of 13 cohorts from 9 populations with a total of 826,690 individuals identified about 100 independent associated risk variants across 11 OA phenotypes, these genetic variants can only account for a small fraction of OA heritability (334). Thus, we took advantage of the NL founder population and the availability of the NFOAS and conducted an independent GWAS in a well-established NFOAS which was not included in the recent meta-analysis to identify novel genetic variants to be associated with OA.

5.3.1 The structure of Newfoundland founder population:

The island of Newfoundland is a part of the most easterly province of Canada in the Atlantic Ocean off Canada's east coast. It was first discovered by continental Europeans, when the Italian explorer John Cabot arrived and landed on Cape Bonavista in 1497 (549). Then, the seasonal fishing trips of English and Irish fishermen that came to the island to harvest the summer cod fishery led to settlements of small groups of immigrants along the island coastline during the 1600s. Due to the expansion of the fishing industry in the mid 1700s, the fisheries became very profitable, and the permanent immigration to the island of Newfoundland took place (549). Thus, hundreds of fishing communities of about 20,000 individuals of English Protestants from south-west England and Irish Catholics from south-east Ireland were established along 10,000 kilometers of the island coastline, and the population grew through natural expansion to 200,000 by the late 19th century (549). Based on the information from statistics Canada report in 2019, there are

recently approximately 477,787 residents on the Newfoundland Island, 98% of them are of English or Irish descent, and 60% of whom live in small communities of less than 2,500 residents along the island's coastline (550,551).

Geographical distance between communities and religious isolation have contributed to the development of several Newfoundland genetic isolates that exhibit high inbreeding coefficients and genetic homogeneity (552,553). These isolation factors resulted in three distinct populations in Newfoundland including Protestant, Roman Catholic, and a relatively small population of North American Indigenous (554). These properties of the Newfoundland population make it a remarkable resource for studying genetic disorders. Moreover, the Newfoundland population is considered to be one of the few founder populations in the world that also include the Canadian Mennonites, French Northeastern in Québec, Hutterites, Icelandic, Dutch, and Finnish populations (555). The founder effect in these founder populations is defined as the loss of genetic variation that occurs when a new population is established by a very small number of individuals from a larger population (556). Compared to admixture populations that have a more heterogenous gene pool, homogenous founder populations present many advantages in gene discovery studies, including both Mendelian and complex disorders. Relative to other founder populations, NL is comparatively young (<20 generations). Thus, genetic drift due to founder effect in the Newfoundland population retains many advantages in the mapping and identifying disease genes (557) and providing an ideal research environment to identify the genetic basis that underlies many diseases (551).

The founder effect has been reported for many diseases in the Newfoundland population. An excellent example of this was identified in 1998, when a founder nonsense mutation in the *endocrine neoplasia type 1* (*MEN1*) gene was identified in affected individuals from four large families with familial multiple endocrine neoplasia type 1 (MEN-l) from the Burin peninsula/Fortune Bay area (558). Also, five attenuated adenomatous polyposis coli (AAPC) families were reported in 1999 to have the same ancestral *Adenomatosis Polyposis Coli* (*APC*) gene splice mutation that resulted in the deletion of exon 4 (559). There have also been 12 large Newfoundland hereditary non-polyposis colorectal cancer syndrome (HNPCC) families reported to carry the same *MutS Homolog* (*MSH2*) gene founder mutation (560,561), three large Newfoundland families with a novel and clinically variable spastic ataxia and supranuclear gaze palsy that carry the same disease haplotype on chr12pl3 (562,563), and four families with diffuse gastric cancer from the south-east coast of Newfoundland that carry the same ancestral *Cadherin 1* (*CDH1*) gene mutation (564). The *coagulation factor VIII* (*F8*) gene variant that causes a mild form of hemophilia A (HEMA) that is highly prevalent in the Newfoundland population has also been determined to be a founder mutation (565). Moreover, the characteristics of Newfoundland population has led to multiple gene discoveries among Newfoundland families including the hereditary hearing loss (566), hereditary sensory and autonomic neuropathy type 2 (567,568), and IL-1 receptor antagonist deficiency (569). Also, a heterozygous missense variant in the *Transmembrane Protein 43* (*TMEM43*) gene that causes the lethal arrhythmogenic right ventricular dysplasia type 5 (*ARVD5*) leading to sudden death in young people was identified in 25 families from Newfoundland population (570). Furthermore, the Newfoundland population, lead to gene discovery

efforts for Bardet-Biedl syndrome (BBS), for which Newfoundland families aided in the discovery of four of the 14 known *BBS* loci or genes. Five years after the discovery of the first *BBS* locus (571).

Since OA is a multifactorial condition with strong genetic components, the NFOAS was established to benefit from the Newfoundland population unique characteristics and identify novel genomic, epigenomic, and metabolomic biomarkers in OA patients from the Newfoundland population.

5.4 Materials and Methods:

5.4.1 The Newfoundland Osteoarthritis Study (NFOAS):

The NFOAS is a patient-based study that was initiated in 2011 by Dr. Guangju Zhai and colleagues at the Division of Biomedical Sciences, Faculty of Medicine, Memorial University of Newfoundland (572–574). The main objectives of the NFOAS were to identify novel biomarkers associated with the development and progression of OA, as well as the outcome of the TJR therapy. The NFOAS was a major initiative in NL to create a biobank of human joint tissues. 1,369 total hip or knee replacement patients due to primary OA were recruited to the NFOAS between November 2011 and September 2017 in St Clare's Mercy Hospital and Health Science Centre General Hospital in St John's, the capital city of NL, Canada (441). OA diagnosis was made based on the ACR-OA criteria and post-surgery pathology reports on cartilage.

Patients' blood and joint tissue including cartilage and synovial fluid were collected during surgery. Then, DNA was extracted from patients' blood samples. About 76.4% of the subjects were primary knee or hip OA patients, 2.3% were post-traumatic OA, 2.4% were RA, 7% were other diseases, and 11.8% were hip fracture. Most of patients (n=1,086) completed their outcome assessment by 2017 using the WOMAC Likert version 3.0 at 3.99±1.38 years post-surgery (441,454,572–574).

The WOMAC Likert version 3.0 pain and function subscales scored 0–20 and 0–68 respectively, with 0 represent no pain or functional difficulties, **Appendix A**. To investigate the outcome of the TJR, the same outcome measures used in other studies were used in the NFOAS (575). Furthermore, pre-surgery WOMAC information including extensive demographic, anthropologic, epidemiological data, and medical information were collected at baseline. Participants were classified as having achieved the minimum clinically important difference (MCID) (576) on the pain subscale based on achieving a change score of at least 7 points (of 20 points total) for the WOMAC pain from pre-surgery to post-surgery. For function, participants achieved the MCID if the change score was 22 points or greater (of a total of 68) from pre-surgery to post-surgery.

5.4.2 Study participants:

This study was a part of the NFOAS. Study participants were THR and TKR patients due to primary OA who were recruited to the NFOAS before 2017 (577) in St Clare's Mercy Hospital and Health Science Centre General Hospital in St John's, NL, Canada (578). Hip and knee OA diagnosis was made based on the ACR-OA clinical diagnostic criteria

(579). Pathology reports on cartilage and osteophytic irregularities were investigated following surgery to confirm OA diagnose. Ethic approval for the study was obtained from the Health Research Ethics Authority of NL (HREA, reference number 11.311, **Appendix C**), and a written consent was received from all study participants, **Appendix D**.

Self-reported OA-free controls were originally recruited to the CODING study which is an on-going large-scale nutrigenomics cross-sectional cohort performed in the NL population. Subjects of the CODING study were self-selected through a poster campaign advertising. Individuals who were interested in participating in the study gave their written consent and completed a screening questionnaire that included basic personal information, such as physical characteristics and health status. Participants who were between 20 and 79 years old who were born and lived in NL and at least third generation of a NL family; and healthy without serious metabolic, cardiovascular, or endocrine disease were eligible to participate in the CODING study. The CODING study was approved by the Human Investigation Committee of the Faculty of Medicine, Memorial University of Newfoundland (452).

5.4.3 Demographic information:

Self-administered questionnaire was used to collect the patients' demographic and medical information including height, weight, sex, and BMI, **Appendix E** includes a copy of the questionnaire. Patient's age was calculated by subtracting the birth date from the

TJR surgery date, then divided by 365.25 to get the number of years. BMI was calculated by dividing patient's weight in kilograms by the squared height in meters (513,578).

Likewise, the 29-items validated short habitual physical activity questionnaire developed by Baecker *et al.* (1982) was used to estimate physical activity and collect the general characteristics such as age, sex, smoking, medication use, and menopausal status of the CODING study subjects (580). The chi-square test was used to analyze nominal variables; and multiple linear regressions were used to test for associations between genotype and energy, macronutrient intake, or physical activity with adjustment for age, sex, alcohol intake, physical activity, alcohol intake, energy intake, and medication use that was susceptible of affecting energy intake or body composition or weight.
All statistical analyses were performed using SAS software (version 9.3; SAS Institute Inc) (451).

5.4.4 Genome-wide genotyping:

Genomic DNA was isolated from Study participants' whole blood and genotyped using either Illumina HumanOmni2.58 or Illumina Infinium Global Diversity 8 v1.0 genotyping arrays (GDA) (581). The Illumina HumanOmni2.58 is a BeadChip microarray that offers an optimal and comprehensive set of approximately 2,381,000 markers of the human genome including both common and rare SNP content from the 1000 Genomes Project (MAF>2.5%), high throughput and optimized tag SNP content, full support of copy number variants (CNVs), Insertions and deletions (INDELs), and Structural Variants (SVs). The design of this platform enables the parallel genotyping of eight DNA samples, decreases

the amount of DNA required to 200 ng/sample, and provides the fastest and the most cost efficient platform for identification of disease causing/associated genetic variations from diverse world populations (581). The Infinium GDA-8 v1.0 BeadChip is an eight DNA samples microarray that contains about 1,825,277 markers of the human genome and provides the most cost-effective/variant coverage throughout the Illumina human array portfolio. It is built on a high-density SNP global backbone optimized for cross-population imputation coverage of the genome. It enables polygenic risk score development and characterization of genetic architecture in diverse populations. The combination of a high-density SNP backbone and updated, relevant clinical research variant coverage offers exceptional value per genotype by delivering insights for both discovery and screening applications (582). Moreover, it includes enhanced tagging in exonic regions and enriched coverage to map GWAS loci with previously identified disease or trait associations with precision. More than 400,000 markers of exome content were gathered from 36,000 individuals of diverse ethnic groups, including African Americans, Hispanics, Pacific Islanders, East Asians, and individuals of mixed ancestry (582). This Infinium GDA-8 v1.0 BeadChip also features diverse exonic content from the Exome Aggregation Consortium (ExAC) database including both cross-population and population-specific markers with either functionality or strong evidence for association. The Infinium GDA-8 v1.0 BeadChip provides coverage of variants selected from the National Human Genome Research Institute (NHGRI)-GWAS catalog, representing a broad range of phenotypes and disease classifications. This content provides a powerful opportunity to test and validate associations previously found in European populations (582). Also, variants included on the Infinium GDA-8 v1.0 BeadChip consist of markers

with known disease association based on ClinVar, the PharmGKB, and the NHGRI-EBI databases. Furthermore, this BeadChip array provides imputation-based tagSNPs for HLA alleles, extended Major Histocompatibility Complex (MHC) region, the *Killer Cell Immunoglobulin-Like Receptor* (*KIR*) gene, and exonic content from the gnomAD database (582).

5.4.5 Pre-association QC of the raw SNP genotyping data:

The raw genome-wide genotyping data may contain low quality of DNA samples, insufficient DNA hybridization to the array, poorly performing genotype probes, and sample mix-ups or contamination. Therefore, it is an essential practice in GWAS analysis to conduct pre-association QC filtering (583). The QC flow chart in **Figure 5.1** illustrates the PLINK (V1.7) (584) pre-association QC filtering steps that were performed separately for the Illumina HumanOmni2.58 and the Infinium GDA 8 v1.0 SNP genotyping datasets to : a) exclude SNPs and individuals with very high levels of missingness (genotyping call rate <95%); b) exclude individuals with sex discordance between the self-reported in the dataset and their sex based on the X chromosome heterozygosity/homozygosity rates that can be resulted from sample handling errors; c) eliminate individuals deviated ±3 SDs from the samples' autosomal heterozygosity rate mean, since deviations higher and lower than 3SD of heterozygosity rate mean indicates sample contamination and inbreeding, respectively; d) avoid bias resulted by population stratification by excluding non-Caucasian individuals using the PCA; e) exclude SNPs with MAF <0.01 that lead to lack of SNP-phenotype association power; and f) exclude SNPs deviate from the HWE ($p < 0.0001$) (583,585).

Then, the list of genetic variations rsIDs of the Illumina HumanOmni2.58 and the Infinium GDA 8 v1.0 SNP microarrays were downloaded from Illumina product support website and the genetic variations were updated using PLINK (V2.0) (586).

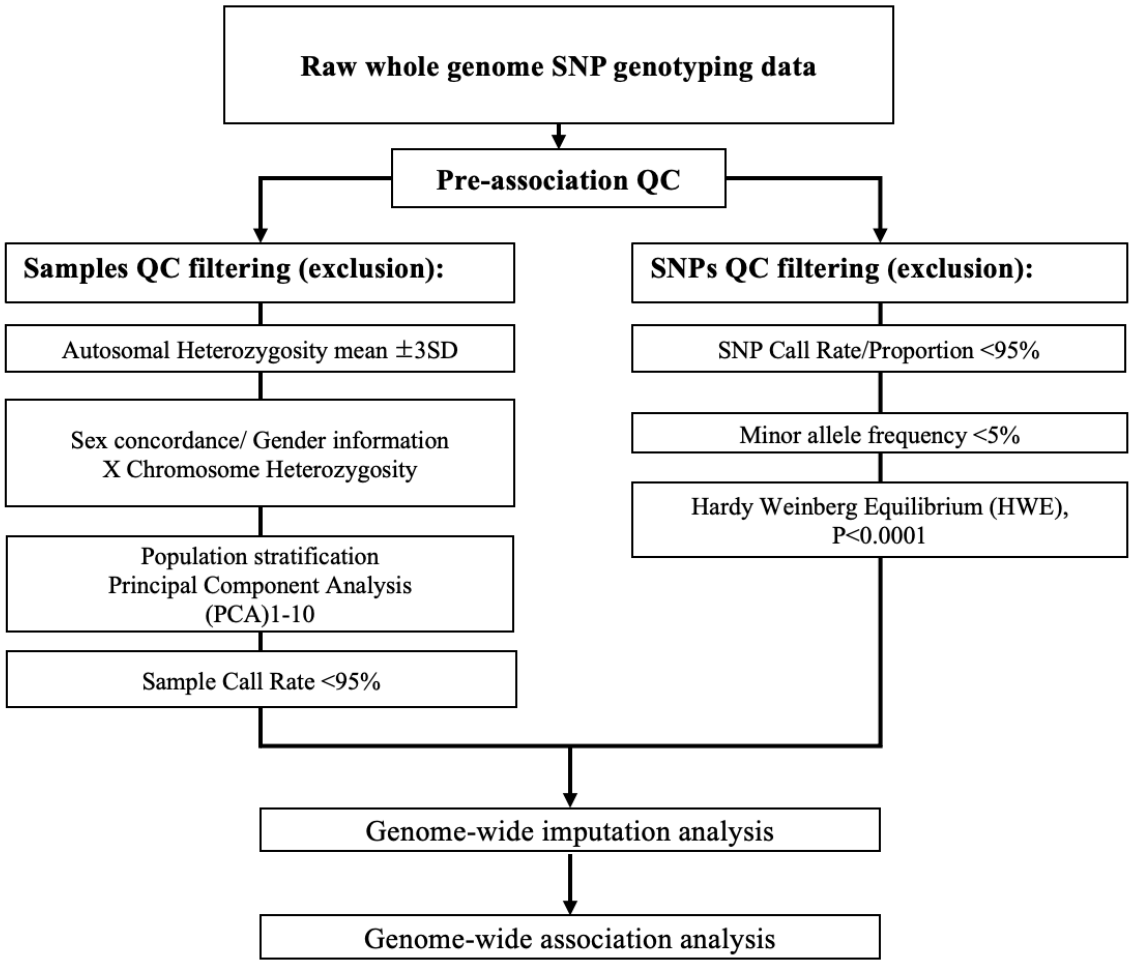


Figure 5.1: A flow chart illustrates the GWAS pre-association quality control analysis steps performed in this study to exclude samples and SNPs with low quality genotyping data from further analysis. Also, this chart shows the subsequent genome-wide imputation and association analysis.

5.4.6 Genotype imputation analysis:

Availability of the imputation models in the last two decades allowed researchers to obtain nearly whole-genome data from SNP genotyping microarrays data at reasonable cost (587).

To improve the SNP coverage and increase the statistical power of our GWAS analysis, genome-wide imputation was performed for both HumanOmni2.58 and the Infinium GDA 8 v1.0 SNP genotyping datasets separately to predict the genotypes of the un-genotyped SNPs. Imputation analysis was performed on the Compute Canada computing cluster (588) using ShapeIT4 phasing (589) and Impute5 imputation (590) programs. Specifically, strands, alleles, and positions of genetic variants in both SNP genotyping datasets were checked and corrected based on the human reference genome map panel build GRCh37/hg19 using BCFtools (591). Then, PLINK (V2.0) was used to convert the clean autosomal SNP genotyping datasets in PLINK binary BED file format into variant call format (VCF) files (586), **Figure 5.2**. Moreover, the phased 1000 Genomes Project phase III dataset that contains the genomes of 2,504 individuals from 26 populations and characterized ~86 million genetic variants was downloaded and used as reference genome throughout the imputation analysis (592). Next, the im5Chunker_1.1.5_static program of the Impute5 was used to split the OA datasets and the 1000 Genomes Project reference genome to small chromosomal chunks of ~5MBs with 250KBs overlapping buffer regions, **Figure 5.2**.

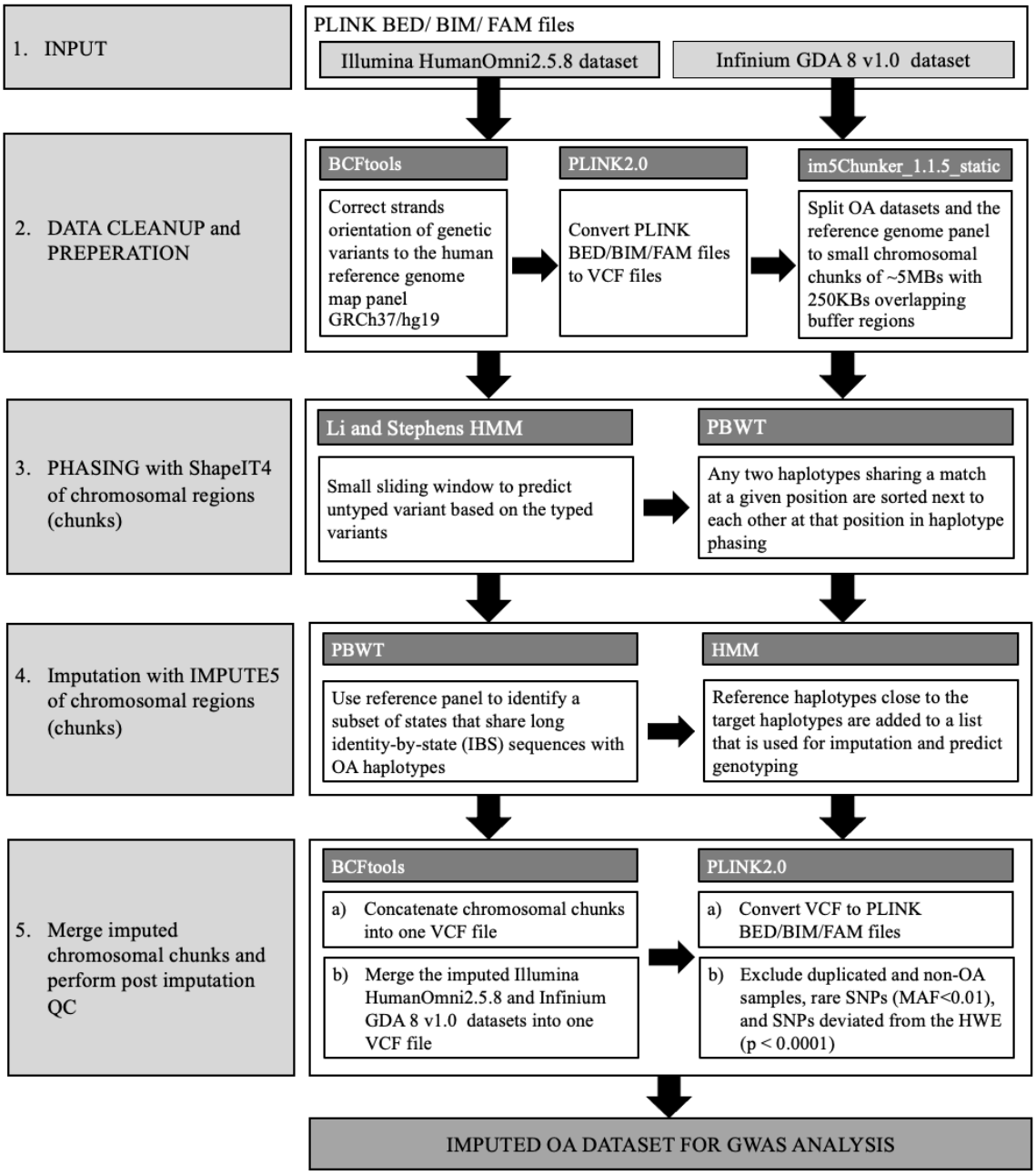


Figure 5.2: A flow chart illustrates step-by-step genome-wide phasing using ShapeIT4 and imputation analysis using IMPUTE5.

Subsequently, phasing analysis of the OA chromosome chunks was performed using the ShapeIT4 (589). Nowadays, haplotype phasing became commonly performed in imputation analysis based on large reference panels of haplotypes. However, haplotype estimation has two main challenges including computational efficiently to accurately process large-scale datasets and data integration to exploit simultaneously large reference panels of haplotypes and long sequencing reads (589). ShapeIT4 statistical haplotype estimation represents efficient solutions to these challenges. It allows processing of SNP genotyping array and sequencing data accurately with running times that are sub-linear with sample size, therefore making it well suited for very large-scale datasets. Furthermore, it facilitates the integration of additional phasing information such as reference haplotypes, long sequencing reads, and sets of pre-phased variants altogether to boost the quality of the resulting haplotypes (589). This phasing method offers the possibility to integrate a reference panel of haplotypes, phase information contained in sequencing reads, and subsets of genotypes at which the phase is known a priori haplotype scaffold. So that, all these layers of information can be conveniently and simultaneously used (593). ShapeIT4 phasing method uses the Li and Stephens hidden Markov model (HMM) and the Positional Burrows–Wheeler Transform (PBWT) to quickly assemble small sets of informative haplotypes to condition on when estimating haplotypes. The HMM treats the sequence of each chromosome as a mosaic of the haplotypes from the reference panel, where each untyped variant is imputed using the typed variants in a small sliding window of one centimorgan around itself (594). The PBWT is a generic approach that encapsulates a large amount of local linkage disequilibrium information to encode binary matrices, especially useful in the case of

haplotypes at a set of binary markers, each with two alleles arbitrarily coded as 0 and 1. A PBWT of haplotypes is a data structure in which any two haplotypes sharing a match at a given position are sorted next to each other at that position (595). Thus, ShapeIT4 examines the PBWT arrays for every eight variants within 2 Mb overlapping genomic regions to get the haplotypes that share the longest match with the current haplotype estimates at that position (589). It maintains a PBWT of all the haplotype estimates so that long matches between haplotypes can be identified in constant time. This method provides highly accurate haplotype estimates for large-scale datasets. Beyond phasing, this PBWT-based approach is believed to have the potential to speed up computations involved by other haplotype-based models used in population genetics for admixture mapping, identity-by-descent (IBD) mapping or genotype imputation (596). Thus, ShapeIT4 was used in our analysis to quickly select a small set of informative conditioning haplotypes in the OA genotyping data based on the human reference genome map panel build GRCh37/hg19, **Figure 5.2**.

Afterward, imputation analysis was conducted using the Impute5 program (590), Impute5 uses the PBWT of the reference panel at target markers to identify a subset of states that share long identity-by-state (IBS) sequences with target haplotypes. Initially, each target haplotype is located in the PBWT, and the PBWT of the reference panel is calculated at the target markers sequentially by moving from left to right across the region being imputed and the state selection occurs at the same time (590). So, after one pass through the full dataset, the state selection is performed for all the target haplotypes. Then, the reference haplotypes close to the target haplotypes in the PBWT are added to a list, and

this list is then used as the copying set of states in the HMM. Thus, there is no need to store the full PBWT of the reference panel in the memory to save computation time and memory usage (287). In our analysis, Impute5 used the PBWT of the 1000 Genomes Project reference genome panel at study target markers to identify a subset of states that share long IBS sequences with the OA target haplotypes, **Figure 5.2**.

Following to the genotype imputation analyses, chromosome chunks for each chromosome were concatenated into one VCF file. Then, all VCF files for chromosomes 1-22 were concatenated into one VCF file for each OA dataset using the BCFtools (591).

In the same time, the genome_wide imputation analysis was conducted for the OA Illumina HumanOmni2.58 and the Infinium GDA 8 v1.0 SNP genotyping datasets separately on the Wellcome Trust Sanger Imputation Service (597) to validate our inhouse imputation analysis. Where phasing of the OA genome-wide data was performed with Eagle2 program, and imputation analysis was done with the PBWT algorithm based on the 1000 Genomes Project phase III dataset.

Subsequently, both OA SNP imputed datasets were merged into one VCF file using BCFtools (591) and converted to PLINK binary BED file format with PLINK(V2.0) (586). Post imputation QC filtering was performed using PLINK(V2.0) (584) to exclude duplicated and non-OA samples, and exclude rare SNPs with MAF<0.01 that are deviated from the HWE ($p < 0.0001$), **Figure 5.2**.

5.4.7 Genome-wide association analysis:

The GWAS analysis was performed to investigate the association between the whole-genome genetic variants and OA in the whole OA cohort, knee OA, and hip OA patients. This GWAS analysis was conducted using MilorGWAS package in R version 4.1.2 (598).

MilorGWAS is a computational efficient approach that uses the mixed logistic regression (MLR) for genome-wide association analysis. This approach estimates SNP variant effect size for structured populations with a stratified quantile-quantile (Q-Q)-plot, enhancing the diagnosis of p-values inflation or deflation when population strata are not clearly identified in the study sample (598). MilorGWAS proposed two MLR methods including the Approximate Maximum Likelihood Estimate (AMLE) and Offset methods that were constructed with conceptually simple mathematical principles. The AMLE is based on a first-order approximation of the MLR, which leads to an approximation of the SNPs effect. The association is tested by a Wald test for binomial count data using a Markov Chain Monte Carlo-based approach, and thus the conclusions drawn for the MLR apply regarding type I error (598). The second method is called the Offset method that have similar performances to AMLE but is slightly over-conservative in the presence of strong familial effects (598). The two methods give similar p-values, but AMLE produces slightly higher p-values for the most associated SNPs. These two proposed methods of the MilorGWAS include the top 10 PCs as fixed effects in the MLR to completely correct for population structure in GWAS analysis (598). Moreover, stratified Q-Q-plot created by Chen *et al*. (2016) was implemented to control type I error rates (599).

To correct for variance inflation resulting from genotyping errors or subtle subpopulation structure, if any, the observed χ^2 statistic was adjusted using the genomic control approach based on the median test statistic (600). A Q-Q-plot was also carried out to check whether the distribution of the inflation corrected p-values deviated from the expected distribution under the null hypothesis of no genetic association. Finally, the genome-wide significance level was defined at ($p < 5*10^{-8}$) to control for multiple testing.

5.4.8 Compare the genetic variants associated with OA in the NFOAS with the publicly available OA GWAS datasets:

Furthermore, a list of 22 publicly available GWAS datasets across 11 OA phenotypes (**Table 5.1**) were downloaded from the Genetics of Osteoarthritis (GO) Consortium GWAS: (https://msk.hugeamp.org/dinspector.html?dataset=Boer2021_OA_Mixed_Main). Then, the identified genetic variants to be significantly associated ($p \leq 5*10^{-8}$) with OA in our NFOAS cohort were compared with the genetic variants that were reported previously to be associated ($p \leq 0.05$) with OA from the recent largest GWAS-meta-analysis (334). Also, the SNPs that were identified to be significantly associated ($p \leq 5*10^{-8}$) with OA in the recent largest GWAS-meta-analysis were compared with the associated genetic variants ($p \leq 0.05$) with OA in the NFOAS (334).

Table 5.1: The list of the 22 datasets across 11 OA phenotypes from the recent largest GWAS-meta-analysis of 13 cohorts from 9 populations with a total of 826,690 individuals.

	GWAS datasets across 11 OA phenotypes
1	KP.Format.ThumbOA.gwama.plotting.gz.minus9sout.gz
2	KP.Format.TKR.gwama.plotting.gz.minus9sout.gz
3	KP.Format.TJR.gwama.plotting.gz.minus9sout.gz
4	KP.Format.THR.gwama.plotting.gz.minus9sout.gz
5	KP.Format.SpineOA.gwama.plotting.gz.minus9sout.gz
6	KP.Format.KneeOA.gwama.plotting.gz.minus9sout.gz
7	KP.Format.KneeHipOA.gwama.plotting.gz.minus9sout.gz
8	KP.Format.HipOA.gwama.plotting.gz.minus9sout.gz
9	KP.Format.HandOA.gwama.plotting.gz.minus9sout.gz
10	KP.Format.GO.RAW.final.meta.results.Early_AllOA.txt.gz
11	KP.Format.GO.FILTER.GW.ThumbOA.FULL.09052019.txt.gz
12	KP.Format.GO.FILTER.GW.TKR.FULL.09052019.txt.gz
13	KP.Format.GO.FILTER.GW.TJR.FULL.09052019.txt.gz
14	KP.Format.GO.FILTER.GW.THR.FULL.09052019.txt.gz
15	KP.Format.GO.FILTER.GW.SpineOA.FULL.09052019.txt.gz
16	KP.Format.GO.FILTER.GW.KneeOA.FULL.09052019.txt.gz
17	KP.Format.GO.FILTER.GW.KneeHipOA.FULL.09052019.txt.gz
18	KP.Format.GO.FILTER.GW.HipOA.FULL.09052019.txt.gz
19	KP.Format.GO.FILTER.GW.HandOA.FULL.09052019.txt.gz
20	KP.Format.GO.FILTER.GW.AllOA.FULL.09052019.txt.gz
21	KP.Format.FingerOA.gwama.plotting.gz.minus9sout.gz
22	KP.Format.AllOA.gwama.plotting.gz.minus9sout.gz

5.5 Results:

In total, 95 subjects and 1,466,378 autosomal genetic variants from the Illumina HumanOmni2.58 SNP genotyping dataset, and 482 OA patients and 823,710 genetic variants from the Illumina Infinium GDA 8 v1.0 dataset passed pre-association QC check and were used in the genome-wide imputation analysis that detected 9,617,754 autosomal genetic variants.

A number of 557 OA patients (mean age 66.5±8.64 years, mean BMI 33.53±7.23 kg/m^2, and 54.94% females) and 118 unaffected controls (mean age 56.42±8.92 years, mean BMI 29.58±4.81 kg/m^2, and 59.32% females) passed the post imputation QC check and were included in the association analysis, **Table 5.2**. OA patients were significantly older ($p=0.0001$) and had a higher BMI ($p=2.95*10^{-07}$) than controls. Also, knee OA patients (n=386) were significantly older ($p=0.0001$) and had a higher BMI ($p=1.57*10^{-12}$) than controls, while hip OA patients (n=171) were significantly older ($p=6.84*10^{-14}$) than controls, but there was no significant difference in BMI between them and controls ($p=0.221$). Furthermore, there was no significant difference in sex distribution of OA patients ($p=0.384$), knee OA ($p=0.65$), and hip OA ($p=0.131$) patients compared with controls, **Table 5.2**.

Table 5.2: The characteristics of the 557 OA patients and 118 OA-free controls.

			OA-free controls (n=118, Female %)	**P-value**
Sex (% for Females)	All OA patients (n=557)	54.94%	59.32%	0.384
	Knee OA patients (n=386)	56.99%		0.65
	Hip OA patients (n=171)	59.32%		0.131
Age (yrs)	All OA patients (n=557)	66.5□8.64	56.42□8.92	0.0001
	Knee OA patients (n=386)	65.99□7.82		0.0001
	Hip OA patients (n=171)	67.64□10.17		$6.84*10^{-14}$
BMI (kg/m^2)	All OA patients (n=557)	33.53□7.23	29.58□4.81	$2.95*10^{-07}$
	Knee OA patients (n=386)	34.85□6.7		$1.57*10^{-12}$
	Hip OA patients (n=171)	30.55□7.49		0.221

Values are mean ± SD for continuous variable and percentage for sex. P-values were obtained from Chi squared test for sex distribution and Student's t-test for age and BMI continuous variables.

MilorGWAS revealed that 29 genetic variants on chromosomes 1,7, 10, 11, 12, 16, 19, 20, and 22 were significantly associated with OA in the 557 patients at the genome-wide significance level ($p \leq 5*10^{-8}$) as shown in the Manhattan plot (**Figure 5.3**) with genomic inflation factor of p-value that was estimated at (lambda=1.0). The Q-Q plot in **Figure 5.4** shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the OA. The most significantly associated variants with OA were also listed in **Table 5.3**.

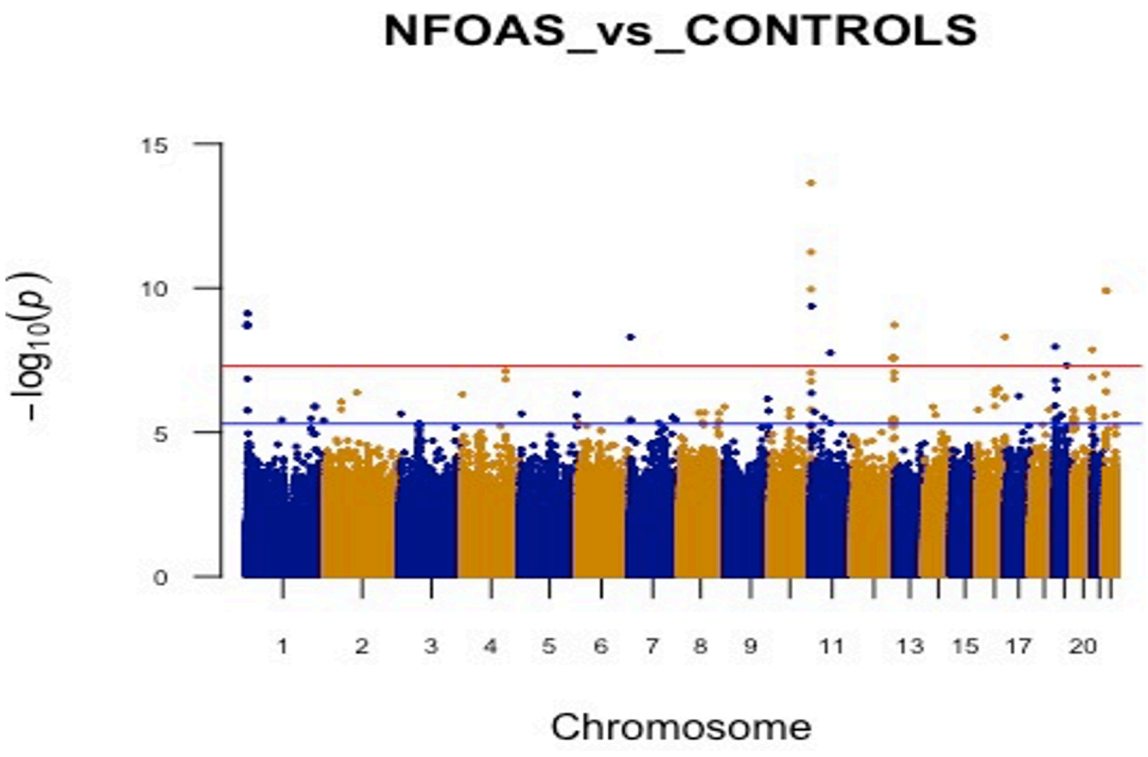


Figure 5.3: A Manhattan plot of the GWAS results showing the most significantly associated genetic variations with OA in the 557 OA patients from the NFOAS in comparing with the 118 OA free controls from the same population. The red line indicates the GWAS significance at $\alpha = 5*10^{-8}$.

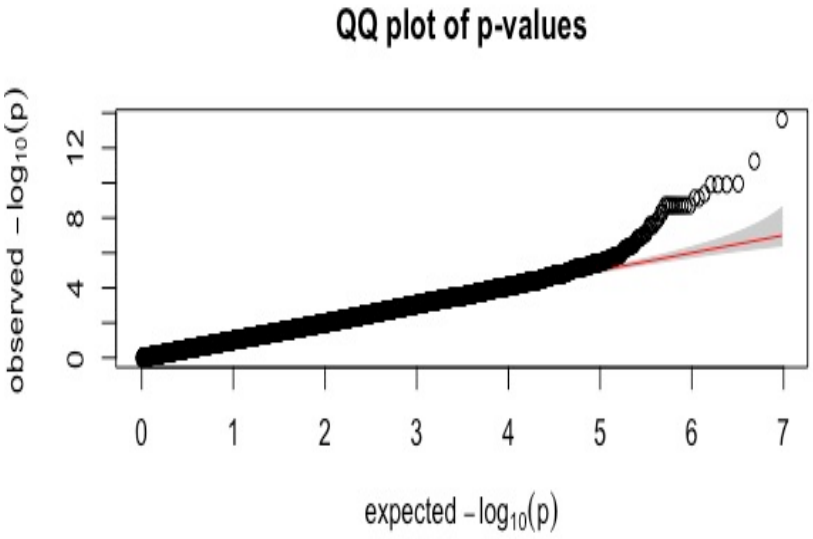


Figure 5.4: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the OA.

Table 5.3: The 29 genetic variants that were identified to be significantly associated with OA in the 557 patients from the NFOAS at the genome-wide significance level ($p=5*10^{-8}$).

chr	position	Variant ID	A2	A2 freq	OR	L95%	U95%	P-value
1	974988	rs115049241	T	0.01	0.08	0.02	0.35	$7.55*10^{-10}$
1	976554	rs191270495	G	0.01	0.08	0.02	0.35	$7.55*10^{-10}$
1	978046	rs9331225	T	0.01	0.12	0.03	0.51	$1.94*10^{-9}$
1	979690	rs115413462	A	0.01	0.12	0.03	0.51	$1.94*10^{-9}$
1	979835	rs114389542	A	0.01	0.12	0.03	0.51	$1.94*10^{-9}$
1	980077	rs116106396	T	0.01	0.12	0.03	0.51	$1.94*10^{-9}$
1	980189	rs114879738	A	0.01	0.12	0.03	0.51	$1.94*10^{-9}$
1	980276	rs115449966	T	0.01	0.12	0.03	0.51	$1.94*10^{-9}$
1	980824	rs112039851	C	0.01	0.12	0.03	0.51	$1.94*10^{-9}$
1	980955	rs139005472	A	0.01	0.12	0.03	0.51	$1.94*10^{-9}$
1	981131	rs9697293	G	0.01	0.12	0.03	0.51	$1.94*10^{-9}$
7	289138	rs141053204	T	0.15	0.34	0.22	0.52	$4.91*10^{-9}$
10	134886621	rs563964217	T	0.01	0.05	0.01	0.21	$2.29*10^{-14}$
10	135056001	rs111704643	C	0.01	0.03	0.003	0.24	$5.68*10^{-12}$
10	135113056	rs2298121	T	0.01	0.12	0.03	0.49	$1.12*10^{-10}$
11	1253904	rs200222877	T	0.02	0.08	0.02	0.25	$4.2*10^{-10}$
11	61627203	rs560586208	C	0.06	0.21	0.11	0.42	$1.79*10^{-8}$
12	129586987	rs72298257	T	0.01	0.10	0.03	0.42	$2.65*10^{-8}$
12	129587049	rs57991786	G	0.01	0.10	0.03	0.42	$2.65*10^{-8}$
12	129587535	rs1552423	G	0.01	0.10	0.03	0.42	$2.65*10^{-8}$
12	129588957	rs201085636	C	0.01	0.10	0.03	0.42	$2.65*10^{-8}$
12	132340280	rs142897716	A	0.01	0.04	0.01	0.30	$1.91*10^{-9}$
16	88166883	rs78941287	G	0.01	0.08	0.02	0.38	$4.96*10^{-9}$
19	869605	rs569627639	AT	0.01	0.01	0.002	0.09	$1.09*10^{-8}$
19	38706886	rs34837947	A	0.03	0.27	0.11	0.66	$4.81*10^{-8}$
20	61410484	rs76236562	T	0.02	0.16	0.05	0.52	$1.38*10^{-8}$
22	20265515	rs181536033	T	0.02	0.10	0.03	0.30	$1.24*10^{-10}$
22	20265527	rs560452957	A	0.02	0.10	0.03	0.30	$1.24*10^{-10}$
22	20266657	rs532097186	T	0.02	0.10	0.03	0.30	$1.24*10^{-10}$

Chr: chromosome, A2: minor allele, A2 freq: minor allele frequency in the OA patients from the NFOAS, OR: Odd Ratio of the minor allele, L95%: Lower 95% confidence interval, U95%: Upper 95% confidence interval.

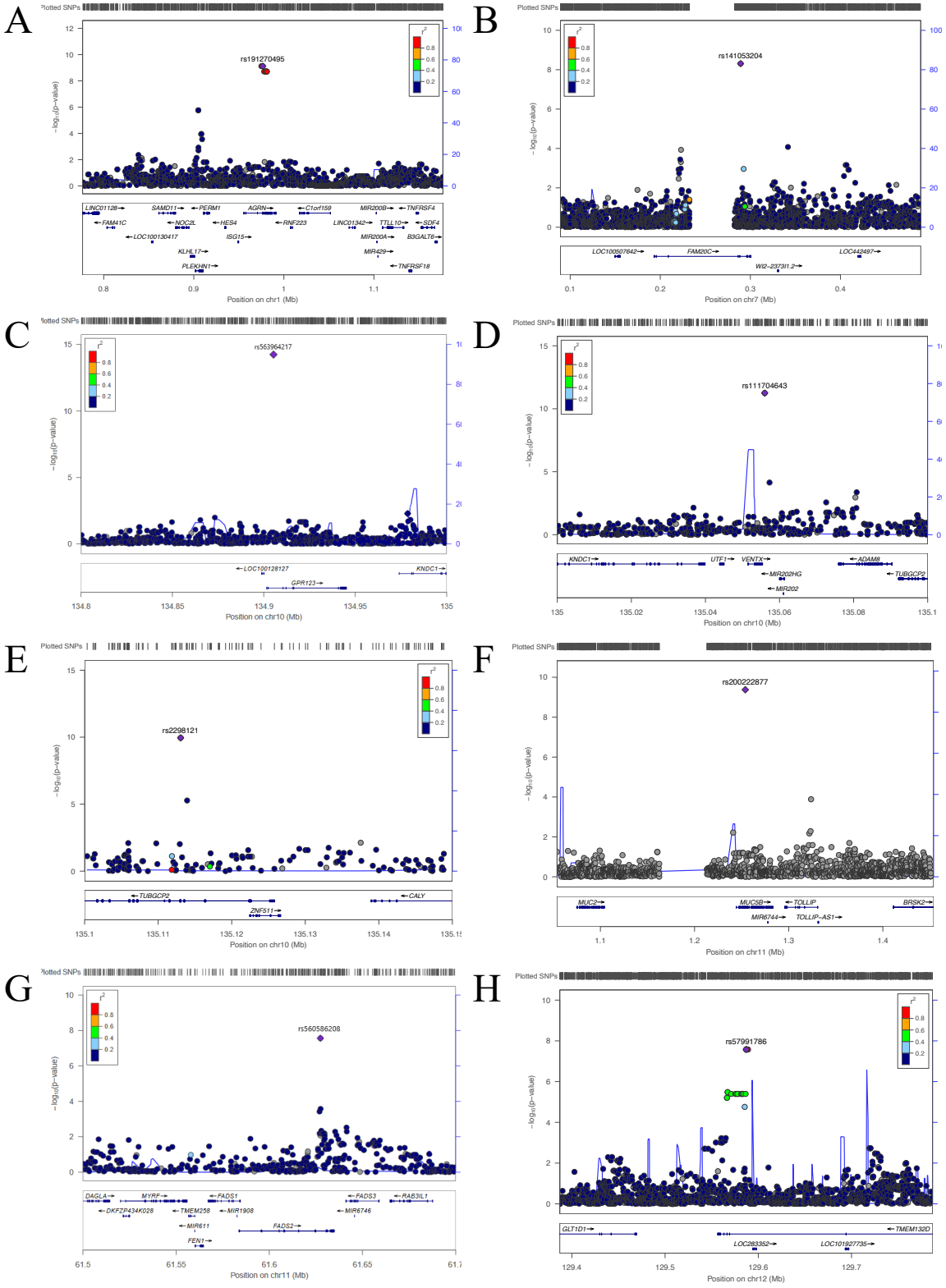
We found eleven variants in the *Agrin* (*AGRN*) gene on chr1 to be significantly associated with OA (**Table 5.3**), eight of them are intronic variants including rs115049241 (OR=0.08, $p=7.55*10^{-10}$), rs9331225, rs115413462, rs114389542, rs116106396, rs114879738, rs115449966, and rs139005472 (all OR=0.12, $p=1.94*10^{-9}$); two synonymous variants including rs191270495 (OR=0.08, $p=7.55*10^{-10}$) and rs112039851 (OR=0.12, $p=1.94*10^{-9}$); and a missense variant rs9697293 (OR=0.12, $p=1.94*10^{-9}$), **Figure 5.5A**. Analysis showed that theses variants are in LD and are correlated with $r^2>0.96$. Thus, any of them can be used as a tagSNP of the other identified SNPs in this haplotype. Also, this chromosomal region includes the *PPARGC1 And ESRR Induced Regulator, Muscle 1* (*PERM1*) gene. Another SNP (rs141053204) in the intron of the *FAM20C Golgi Associated Secretory Pathway Kinase* (*FAM20C*) gene on chr7 (**Figure 5.5B**) was significantly associated (OR=0.34, $p=4.91*10^{-9}$) with OA as shown in **Table 5.3**.

We also found that the T allele of rs563964217 was the most significantly associated with OA (OR=0.05, $p=2.29*10^{-14}$), **Table 5.3**. This SNP is a stop-gain variant (c.CAG>TAG, p.Q>*) in exon 3 of the *G-Protein Coupled Receptor 123* (*GPR123*) gene on chr10 as shown in the regional association plot in **Figure 5.5C**. The two other significantly associated SNPs on chr10 included rs111704643 (OR=0.03, $p=5.68*10^{-12}$), and rs2298121 (OR=0.12, $p=1.12*10^{-10}$), **Table 5.3**. SNP (rs111704643) is intergenic in a gene rich region that is located ~ 170 Kbps downstream of rs563964217 and adjacent to the *ADAM Metallopeptidase Domain 8* (*ADAM8*) gene, **Figure 5.5D**. While SNP

(rs2298121) is a missense variant (c.GCT>ACT, p.A>T) in the *Tubulin Gamma Complex Associated Protein 2* (*TUBGCP2*) gene, **Figure 5.5E**.

Additionally, two variants on chr11 were significantly associated with OA including rs200222877 and rs560586208 as shown in **Table 5.3**. SNP rs200222877 (OR=0.08, p= $4.20*10^{-10}$) is a missense variant in the *Oligomeric Mucus/Gel-Forming* (*MUC5B*) gene. Also, this chromosomal locus includes the *Toll Interacting Protein* (*TOLLIP*), and the *BR Serine/Threonine Kinase 2* (*BRSK2*) genes, **Figure 5.5F**. While SNP rs560586208 (OR=0.21, p= $4.20*10^{-8}$) is an intronic variant in the *Fatty Acid Desaturase 2* (*FADS2*) gene on chr11, **Figure 5.5G**.

Furthermore, two loci were found to be significantly associated with OA on chr12, **Table 5.3**. The first locus contains four variants (rs72298257, rs57991786, rs1552423, and rs201085636) that are associated with OA at p=$2.65*10^{-8}$ (OR=0.10). These variants are ~2 Kbps apart in intron 5 of the *Transmembrane Protein 132D* (*TMEM 132D*) gene and arencorrelated in LD with r^2=1, **Figure 5.5H**. Whereas the second locus contains rs142897716 variant that was significantly associated (OR=0.04, p= $1.91*10^{-9}$) with OA. This SNP is intergenic adjacent to the *Matrix Metallopeptidase 17* (*MMP17*) gene, **Figure 5.5I**.


A
rs191270495
Position on chr1 (Mb)
B
rs141053204
Position on chr7 (Mb)
C
rs563964217
Position on chr10 (Mb)
D
rs111704643
Position on chr10 (Mb)
E
rs2298121
Position on chr10 (Mb)
F
rs200222877
Position on chr11 (Mb)
G
rs560586208
Position on chr11 (Mb)
H
rs57991786
Position on chr12 (Mb)

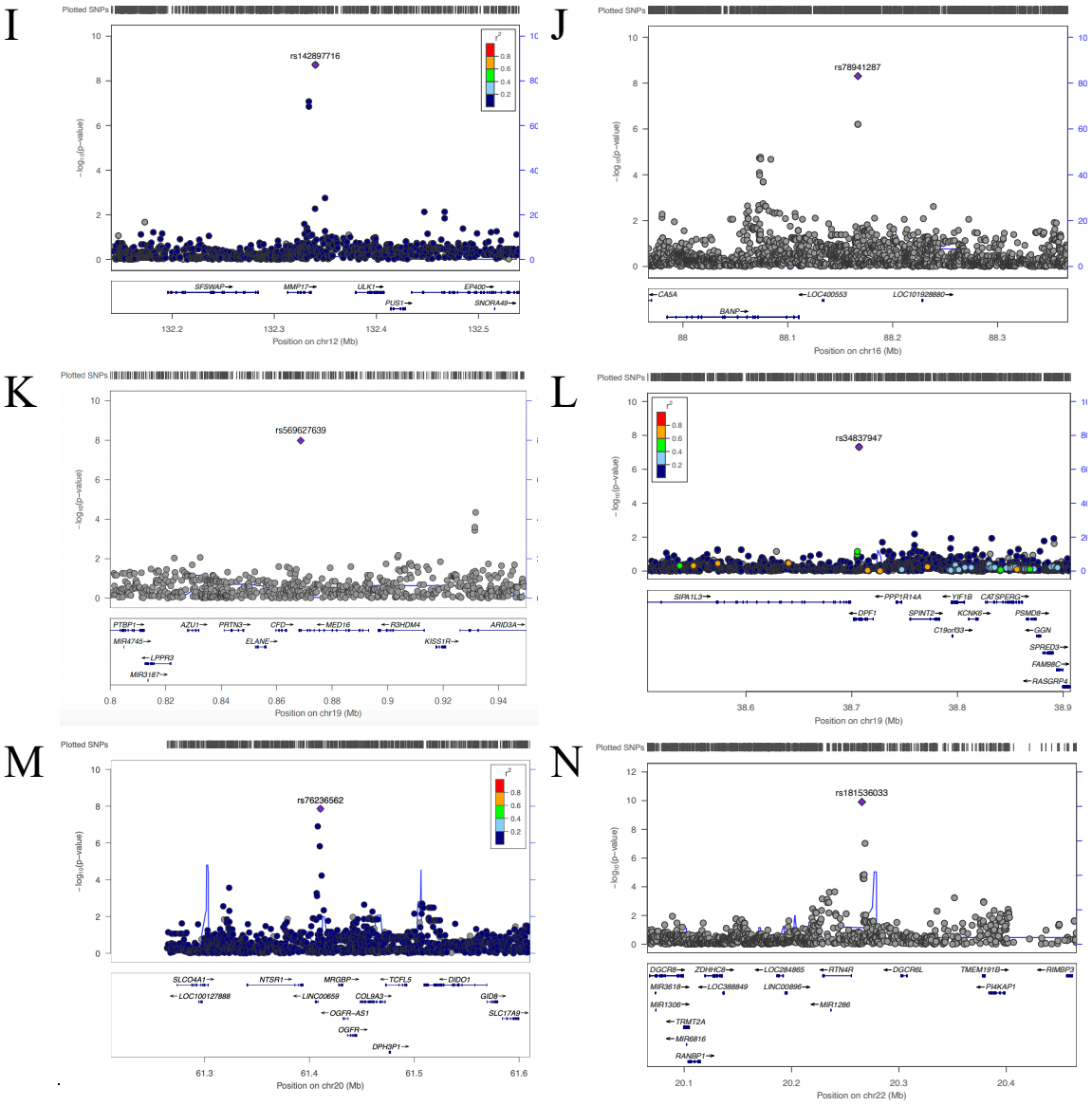


Figure 5.5: The regional association plots of the significantly associated variants with the OA showing: **A)** the chromosomal region around the most significant variants on chr1 (all $p<1.94*10^{-9}$) including rs115049241, rs9331225, rs115413462, rs114389542, rs116106396, rs114879738, rs115449966, rs139005472, rs191270495, rs112039851, and rs9697293 variants in the *Agrin* (*AGRN*) gene; **B)** the chromosomal region around rs141053204 ($p=4.91*10^{-9}$) that is intronic variant in the *FAM20C Golgi Associated Secretory Pathway Kinase* (*FAM20C*) gene on chr7; **C)** the chromosomal region around the most significantly identified SNP rs563964217 on chr10 ($p=2.29*10^{-14}$) showing the location of this variant in the *Adhesion G Protein-Coupled Receptor A1* (*GPR123*) gene;

D) the chromosomal region around the significant SNP rs111704643 on chr10 ($p=5.68*10^{-12}$) showing the adjacent genes including the *ADAM Metallopeptidase Domain 8* (*ADAM8*) gene; **E)** the chromosomal region around the significant SNP rs2298121 on chr10 ($p=1.12*10^{-10}$) showing the location of the variant in the *Tubulin Gamma Complex Associated Protein 2* (*TUBGCP2*) gene; **F)** the chromosomal region around the missense variant rs200222877 in the *MUC5B, Oligomeric Mucus/Gel-Forming* (*MUC5B*) gene on chr11 that was identified to be significantly associated with OA ($p=4.20*10^{-10}$); **G)** the chromosomal region showing the second top SNP rs560586208 that was detected to be significantly associated with OA (p= 4.20*10-8) on chr11. This SNP is an intronic variant located in intron 7 of the *Fatty Acid Desaturase 2* (*FADS2*) gene; **H)** the chromosomal region around the first significantly associated locus (all $p=2.65*10^{-8}$) with OA on chr12 that includes rs72298257, rs57991786, rs1552423, and rs201085636; **I)** the second associated locus with OA ($p= 1.91*10^{-9}$) on chr12 that includes the intergenic SNP rs142897716. This SNP is nearby to the *Matrix Metallopeptidase 17* (*MMP17*) gene; **J)** the chromosomal region around the most significant SNP rs78941287 ($p= 4.96*10^{-9}$) on chr16, this SNP is intergenic that is adjacent to the *Carbonic Anhydrase 5A* (*CA5A*) and the *BTG3 Associated Nuclear Protein* (*BANP*) genes; **K)** the top associated chromosomal region with OA on chr19 that contains the INDEL rs569627639 that was significantly associated ($p= 1.09*10^{-8}$) with OA, this INDEL is an intronic variant in the *Mediator Complex Subunit 16* (*MED16*) gene and harbouring the *Azurocidin 1* (*AZU1*); the *Proteinase 3* (*PRTN3*); and the *Elastase, Neutrophil Expressed* (*ELANE*) genes; **L)** the chromosomal region around the most significant SNP rs34837947 ($p= 4.81*10^{-8}$) on chr19 that is a synonymous variant the *Double PHD Fingers 1* (*DPF1*) gene, this genetic region includes the *Signal Induced Proliferation Associated 1 Like* (*SIPA1L3*), the *Protein Phosphatase 1 Regulatory Inhibitor Subunit 14A* (*PPP1R14A*), and the *Proteasome 26S Subunit, Non-ATPase 8* (*PSMD8*) genes; **M)** the chromosomal region around the most significant SNP rs76236562 ($p= 1.38*10^{-8}$) on chr20 that is intergenic variant adjacent to the *Neurotensin Receptor 1* (*NTSR1*), the *Opioid Growth Factor Receptor* (*OGFR*), the *Collagen Type IX Alpha 3 Chain* (*COL9A3*), the *Death Inducer-Obliterator 1* (*DIDO1*), and the *GID Complex Subunit 8 Homolog* (*GID8*) genes; **N)** the most significantly associated variants with OA on chr22 ($p=1.12*10^{-10}$) including rs181536033, rs560452957, and rs532097186 that are adjacent to the *RAN Binding Protein 1* (*RANBP1*), the *Reticulon 4 Receptor* (*RTN4R*), the *DiGeorge Syndrome Critical Region Gene 6 Like* (*DGCR6L*), and the *Transmembrane Protein 191B* (*TMEM191B*) genes.

One SNP rs78941287 on chr16 was also significantly associated (OR=0.08, p= $4.96*10^{-9}$) with OA, **Table 5.3**. rs78941287 is intergenic variant that harbors the *Carbonic Anhydrase 5A* (*CA5A*) and the *BTG3 Associated Nuclear Protein* (*BANP*) genes, **Figure 5.5J**. Additionally, two chromosomal regions on chr19 were detected to be significantly associated with OA, **Table 5.3**. Firstly, the INDEL, rs569627639 was significantly associated (OR=0.01, p= $1.09*10^{-8}$) with OA. This insertion is an intronic variant in the *Mediator Complex Subunit 16* (*MED16*) gene. Other adjacent genes in this locus include the *Phospholipid Phosphatase Related 3* (*PLPPR3*); the *Azurocidin 1* (*AZU1*); the *Proteinase 3* (*PRTN3*); and the *Elastase, Neutrophil Expressed* (*ELANE*) genes, **Figure 5.5K**. Secondly, the SNP rs34837947 on chr19 was significantly associated (OR=0.27, p= $4.81*10^{-8}$) with OA. This SNP is a synonymous variant (c.CCC>CCT) in the *Double PHD Fingers 1* (*DPF1*) gene. This genetic region has also other genes, such as the *Signal Induced Proliferation Associated 1 Like* (*SIPA1L3*), the *Protein Phosphatase 1 Regulatory Inhibitor Subunit 14A* (*PPP1R14A*), and the *Proteasome 26S Subunit, Non-ATPase 8* (*PSMD8*) genes, **Figure 5.5L**.

Furthermore, the significantly associated variant with OA on chr20 was the rs76236562 (OR=0.16, p= $1.38*10^{-8}$), **Table 5.3**. This SNP is an intergenic variant located in gene rich region that includes the *Neurotensin Receptor 1* (*NTSR1*), the *Opioid Growth Factor Receptor* (*OGFR*), the *Collagen Type IX Alpha 3 Chain* (*COL9A3*), the *Death Inducer-Obliterator 1* (*DIDO1*), and the *GID Complex Subunit 8 Homolog* (*GID8*) genes, **Figure 5.5M**.

Finally, three genetic variants including rs181536033, rs560452957, and rs532097186 on chr22 were significantly associated with OA (OR=0.10, $p=1.24*10^{-10}$), **Table 5.3**. These three intergenic variants span 1,142 bps in LD with $r^2=1$ and harboring the *RAN Binding Protein 1 (RANBP1)*, *Reticulon 4 Receptor (RTN4R)*, *DiGeorge Syndrome Critical Region Gene 6 Like* (*DGCR6L*), and the *Transmembrane Protein 191B* (*TMEM191B*) genes, **Figure 5.5N**.

MilorGWAS analysis of knee OA patients (n=386) versus 118 unaffected controls identified three variants on chr10 to be significantly associated with knee OA including rs563964217 (OR= 0.07, $p=2.64*10^{-10}$), rs111704643 (OR= 0.05, $p=1.59*10^{-8}$), and rs2298121 (OR= 0.13, $p=1.90*10^{-8}$). These variants were also detected to be significantly associated with OA in the whole 557 OA patient cohort. Although the rest of the significantly associated variants with OA in the 557 OA patients did not reach the GWAS significance level, they were associated with knee OA at ($p\leq5*10^{-6}$) as shown in the Manhattan plot in **Figure 5.6**. The genomic inflation of P-value estimated at (lambda= 0.962) as shown in the Q-Q plot in **Figure 5.7**.

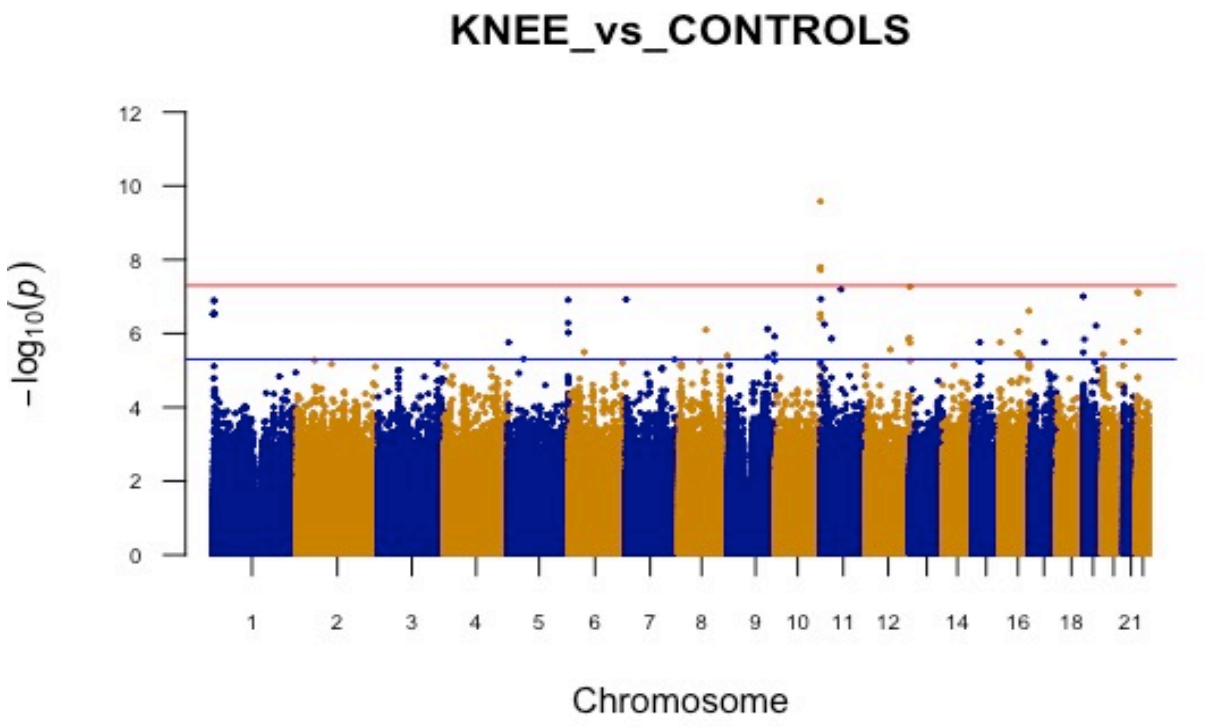


Figure 5.6: A Manhattan plot of the GWAS results showing the most significantly associated genetic variations with knee OA in 386 OA patients from the NFOAS in comparing with 118 OA free controls from the same population. The red line indicates the GWAS significance at $\alpha = 5*10^{-8}$.

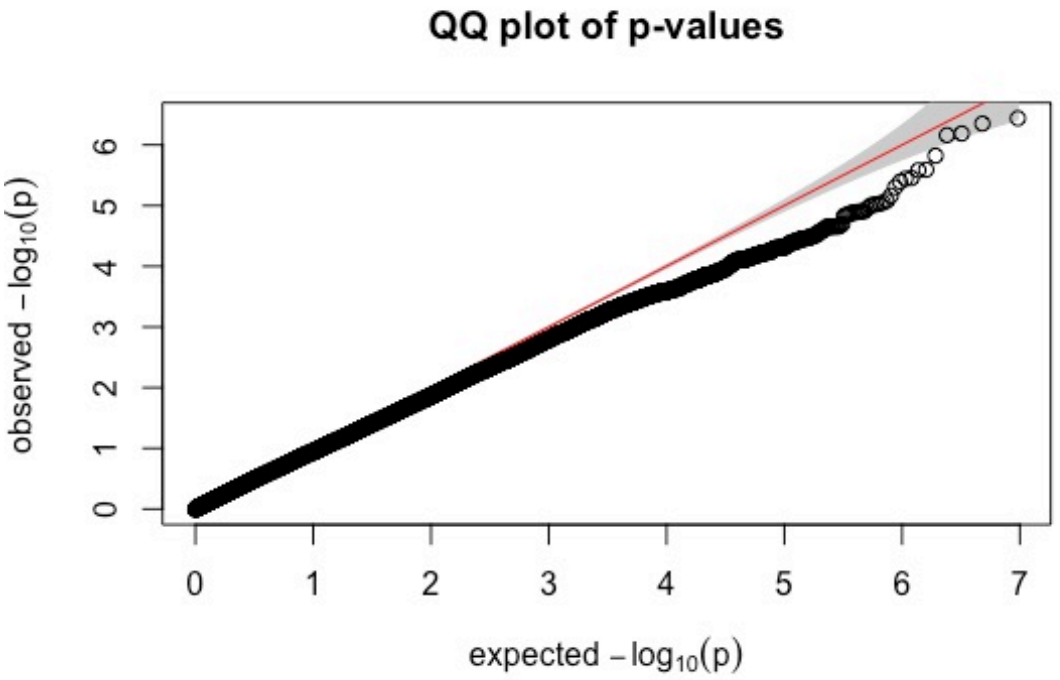


Figure 5.7: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the knee OA.

However, we did not find any associations with the GWAS significance level in MilorGWAS analysis of hip OA patients (n=171) compared to 118 unaffected controls. rs563964217 was associated with hip OA at (OR= 0.10, $p=6.06*10^{-7}$) as shown in the Manhattan plot in **Figure 5.8**. The genomic inflation of P-value estimated at (lambda=1.02) as shown in the Q-Q plot in **Figure 5.9**.

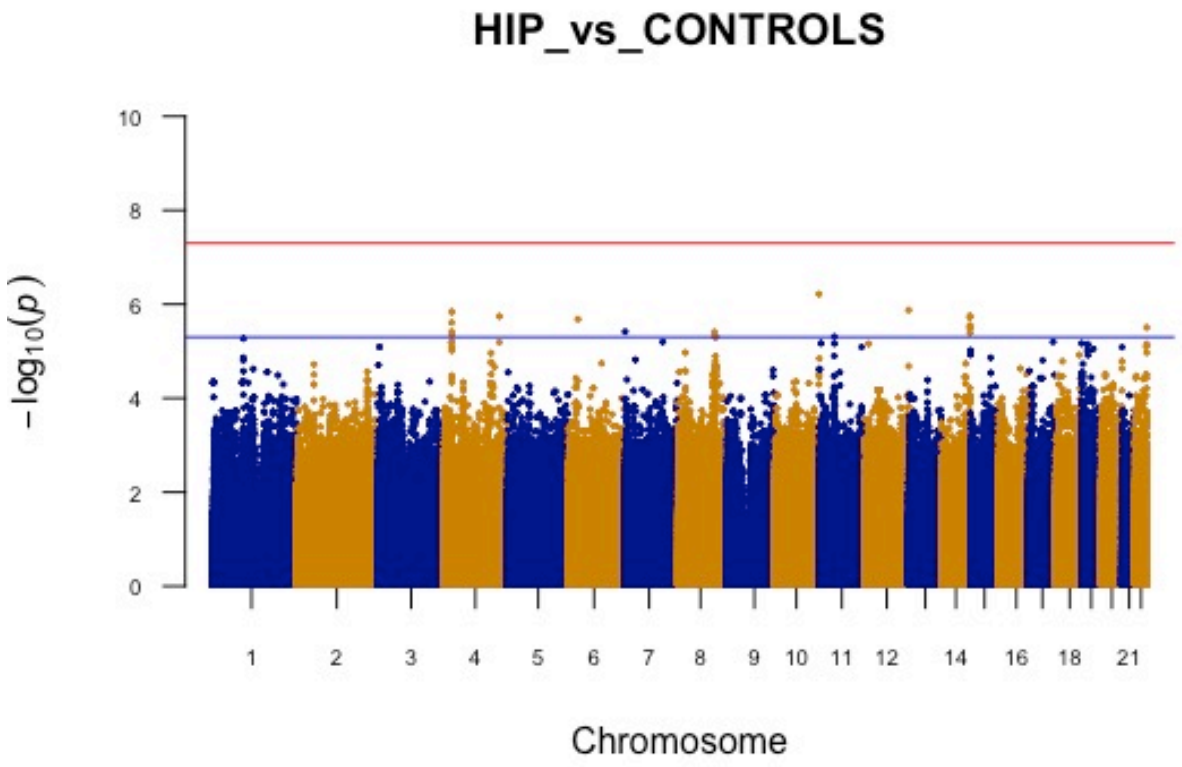


Figure 5.8: A Manhattan plot of the GWAS results showing the associated genetic variations with hip OA in 171 OA patients from the NFOAS compared to 118 OA free controls from the same population. The red line indicates the GWAS significance at $\alpha = 5*10^{-8}$.

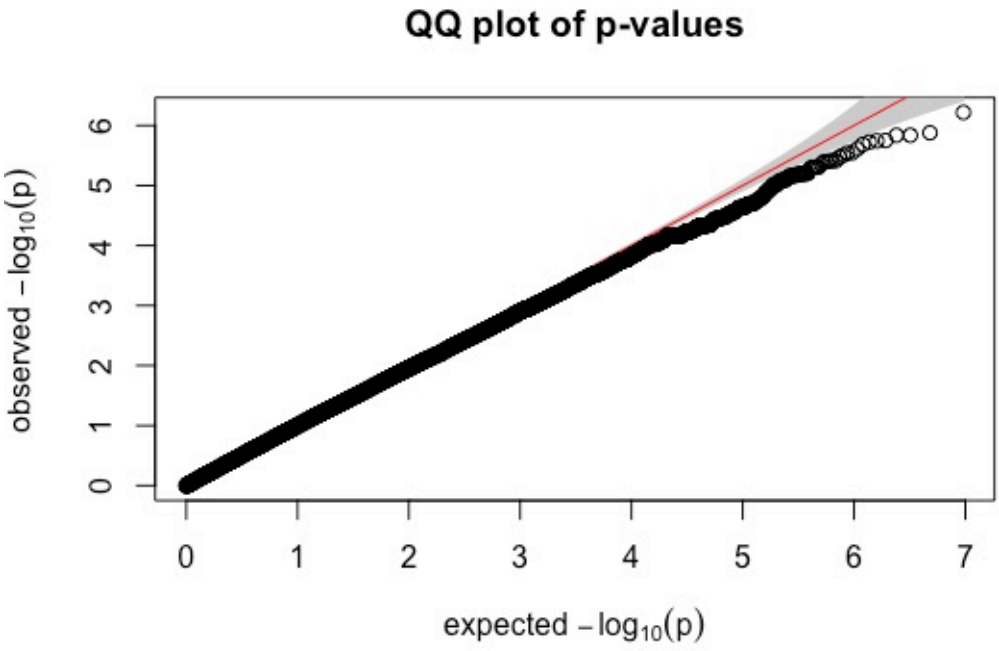


Figure 5.9: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the hip OA.

Interestingly, out of the 29 variants that were significantly associated with OA ($p \leq 5*10^{-8}$) in our NFOAS, only rs141053204 on chr7 was reported in the recent largest GWAS-meta-analysis to be associated with OA at p=0.048 in patients underwent TKR therapy, while the other 28 significantly associated variants with OA in our study were novel and were not identified and reported previously. Furthermore, four significantly associated variants with OA based on the recent largest GWAS-meta-analysis (Boer *et al. Cell 2021*) results were also associated with OA in our NFOAS at $p \leq 0.05$ as shown in the table in the **Appendix H**.

5.6 Discussion:

Since ~50% of the OA heritability is due to genetic components, and previous genetic studies have identified more than 140 associated loci that traversing knee, hip and hand OA with limited overlap, the OA heritability explained by these loci is relatively low and can only account for ~10% (235,254). The distinct characteristics of the Newfoundland population as a genetic isolate formed and have since helped in several gene discoveries in Mendelian and complex diseases. In this study, we benefit from the Newfoundland population structure and the well-established NFOAS and conducted an independent GWAS analysis in a cohort 557 OA (386 knee and 171 hip OA patients) and 118 OA-free controls from the same population to identify novel genetic variants to be associated with OA.

The association analysis in this study was conducted using the MLR in genome-wide association studies that was implemented in the MilorGWAS package in R version 4.1.2 and included the top 10 PCs to completely correct for population structure (601). It is worth mentioning that all the identified genetic variants to be significantly associated with OA using the MilorGWAS were imputed variants. Interestingly, 28 out of 29 identified genetic variants to be associated with OA in this study were not reported previously to be associated with any OA phenotype, which suggests that these variants are specific to the Newfoundland population. The ROC analysis was conducted to address the imbalance between the cases and un-affected controls from the Newfoundland population. Also, the AUC was calculated using the maximum sensitivity and specificity simultaneously (MaxSpSe) method. As a result, the identified variants had a moderate discriminatory

power to distinguish the OA patients from the OA-free controls as indicated by the AUC ≥ 0.70 with an average sensitivity of 0.68 and specificity of 0.72.

The top SNP that was identified to be significantly associated ($p=2.29*10^{-14}$) with OA was rs563964217 on chr10. This SNP is a stop-gain variant (c.CAG>TAG, p.Q>*) in exon 3 of the *GPR123* gene. However, *GPR123* gene had not been previously linked to OA, it encodes a protein belongs to the adhesion family of the G-protein-coupled receptors (GPCRs). Different types of GPCRs are implicated in cartilage degeneration, subchondral bone sclerosis, and chronic pain of OA (91). Members of the adhesion GPCRs family have been reported to be activated by free fatty acids (FFAs) (602,603). Animal model studies stated that mutations in these genes stimulate chondrocytes to secrete more inflammatory mediators and reduce anabolism upon IL-1β treatment leading to progression of OA in the knee induced OA mice (604). Also, mice displayed an accelerated progression of ACLT surgery-induced OA (605). In contrary, agonist of *GPR* genes block degeneration of type II collagen and aggrecan by reducing the expression of matrix-degrading enzymes and pro-inflammatory cytokines in vitro (606). Moreover, they prevent IL-1β-induced reduction of ECM through sex-determining region Y (SRY)-related high mobility group (HMG) box 9 (SOX9) mediated expression of collagen II and aggrecan in ATDC5 cell lines chondrocytes (607). Previous studies reported a central role for these genes in stabilizing cartilage homeostasis and reducing expression of pro-inflammatory mediators (cyclooxygenases 2, nitric oxide synthase 2, and IL-6), pro-inflammatory adipokines (lipocalin-2 and nesfatin-1), and adhesion molecule (Vascular cell adhesion molecule 1 and Intercellular adhesion molecule 1) in OA patients (608,609).

Due to the involvement of GPCRs in wide range of physiological and pathophysiological processes, they became exciting targets for drug discovery in several diseases including OA (610,611). Our results proposed the implication of the *GPR123* gene in the pro-inflammatory and cartilage degradation of OA.

The two other variants on chr10 that were significantly associated with OA included the intergenic SNP (rs111704643, $p=5.68*10^{-12}$) and the missense mutation (rs2298121, $p=5.68*10^{-12}$) in the *TUBGCP2* gene. Interestingly, these variants are ~57 Kbps apart spanning the *ADAM8* gene. *ADAM8* encodes a member of the disintegrin and metalloprotease domain (ADAM) protein family (612). Previous studies identified a strong correlation between overexpression of the *ADAM8* gene and the development of OA (613), and knockdown of the *ADAM8* by gene silence in chondrogenic OA cell model resulted in significantly reduction of OA phenotype, including the suppression of MMP-9, TNF-α, IL-6 expression, the restore of Collagen II, Aggrecan expression, the stimulation of cell proliferation, and the inhibition of cell apoptosis. These findings suggested a critical implication of the *ADAM8* gene in OA pathogenies including the production of inflammatory factors, over expression of MMP-9, and degradation of ECM, as well as proliferation and apoptosis. The *ADAM8* gene acts as an ectodomain sheddase of receptors or ligands which were anchored on cell surface (614). The cleavage of the receptors from cell surface by the ADAM8 resulted in a releasing and promotion of soluble form of receptors, and a suppression of respective intracellular signaling via a blocking of specific ligands. A study performed by Duan *et al*. (2019) identified that overexpression of the *ADAM8* gene in osteoarthritic chondrocytes leads to the cleavage

and release of the membrane anchored epidermal growth factor (EGF) receptor ligands. Access of soluble form of this growth factor activates ERK1/2 and NF-κB to promote the *MMP-9* gene expression, which in turn facilitates the degradation of ECM and OA progression (615,616). Moreover, Notch signaling was reported to play a crucial role in the OA pathogenesis (617,618). It was identified that the inhibition of the Notch signaling leads to the reduction of the *ADAM8* and *MMP-9* expression. The animal model studies in OA rat revealed that the inhibition of Notch signaling or knockdown of the *ADAM8* led to suppression of OA phenotypes, including promotion of Collagen II and Aggrecan expression, down-regulation of Notch1, Hes1, *ADAM8* and *MMP-9* expression, as well as EGF-ERK/ NF-κB signaling (619). Importantly, the glycoprotein in the joint cartilage was found to be promoted by Notch inhibition or the *ADAM8* knockdown. While the overexpression of the *ADAM8* resulted in promotion of OA phenotypes in OA rats, including suppression of Collagen II, Aggrecan, and glycoprotein level, and suppression of Notch1, Hes1, ADAM8 and MMP-9 expression, along with the EGF-ERK/ NF-κB signaling. Therefore, these results demonstrated that Notch-ADAM8 positive feedback loop facilitated OA progression in rat animal model (615). Our findings came in concordance with these studies and proved the possible implication of the *ADAM8* gene in the development and progression of OA in the Newfoundland population.

The identified genetic variants (rs181536033, rs560452957, and rs532097186) on chr22 to be significantly associated with OA ($p=1.24*10^{-10}$) are in complete LD ($r^2=1$). These variants are harboring the *RANBP1*, *RTN4R*, *DGCR6L*, and *TMEM191B* genes. *RANBP1* gene encodes a protein that forms a complex with Ras-related nuclear protein (Ran) and

metabolizes guanosine triphosphate (GTP). This complex participates in the regulation of the cell cycle by controlling transport of proteins and nucleic acids into the nucleus (620). *RTN4R* gene encodes a receptor that mediates axonal growth inhibition and may play a role in regulating axonal regeneration and plasticity in the adult central nervous system (621). *DGCR6L* gene plays a role in cell attachment and migration (622), while the *TMEM191B* gene has been predicted to be an integral component of the cell membrane (623). However, none of these genes have been identified previously to associate with OA, they play central role in cell membrane structure, cell junction, cell cycle, energy, and central nervous system.

Furthermore, we were able to identify eleven variants in the *AGRN* gene on chr1 to be significantly associated (all $p \leq 1.94*10^{-9}$) with OA. Theses variants are in LD ($r^2>0.96$). *AGRN* gene encodes the proteoglycan agrin protein that contains several laminin G, Kazal type serine protease inhibitor and epidermal growth factor domains (624,625). The encoded protein is involved in the formation, maintenance, activation, and synaptic stabilisation of the neuromuscular junction (NMJ) by mediating MuSK activation (626). The activation of MUSK in myotubes induces the formation of NMJ by regulating different processes including the transcription of specific genes and the clustering of Nicotinic acetylcholine receptors (AChR) in the postsynaptic membrane (627–629). Mutations in *AGRN* have been identified to be implicated in the congenital myasthenic syndrome characterised by fatigable weakness of skeletal muscle (630). Since muscle

strength weakness is a risk factor of the OA. These results suggest the involvement of *AGRN* in the OA pathogenesis and pain mechanisms.

SNP rs141053204 that was significantly associated ($p=4.91*10^{-9}$) with OA on chr7 is intronic variant in the *FAM20C* gene that encodes a member of the family of secreted protein kinases. The encoded protein binds calcium and phosphorylates proteins involved in bone mineralization, mutations in *FAM20C* gene are associated with lethal osteosclerotic bone dysplasia (631). Mouse model studies reported that mechanical loading through increased body weight and/or muscle contractions triggers the mechanoreceptor PIEZO1 that play a key role in the regulation of skeletal development and bone maturation to enhance the production of FAM20C kinase in osteoblasts (632–636), FAM20C induces a burst in dentin matrix protein 1 (DMP1) secretion into extracellular matrix. Large amounts of extracellular DMP1 inhibit the VEGF signalling in the ossification front (OF) active bone growth by preventing the phosphorylation of the *vascular endothelial growth factor receptor 2* (*VEGFR2*) leading to reduced activation of VEFGR2 and expression of the *vascular endothelial growth factor receptor 3* (*VEGFR3*) on the tip cells of type H endothelium (637). Thus, highly-angiogenic type H vessels are transformed into quiescent type L vasculature to prevent bone growth activity and lead to rapid matrix mineralisation and strengthening of long bones (637). These findings suggested new options for the treatment of different kinds of bone pathologies featuring abnormal bone growth and vessels such as osteoporosis, osteosarcoma, and OA.

Additionally, rs200222877 and rs560586208 were significantly associated with OA on chr11. SNP rs200222877 ($p= 4.20*10^{-10}$) is a missense variant in the *MUC5B* gene that is the major gel-forming mucin in mucus. It is a major contributor to the lubricating and viscoelastic properties of whole saliva, normal lung mucus and cervical mucus (638). Multiple studies reported the association of rs35705950 promoter variant of *MUC5B* gene with RA associated with interstitial lung disease (ILD) (639–642). Also, this locus includes the *TOLLIP* and *BRSK2* genes. The *TOLLIP* gene regulates inflammatory signaling and is involved in interleukin-1 receptor trafficking and in the turnover of IL1R-associated kinase (643). A study conducted by Ye *et al.* (2020) identified the *TOLLIP* gene to be prospective candidate for cartilage regeneration associated gene that can be potential therapeutic target of cartilage regeneration in OA patients (644). While the *BRSK2* gene enables ATP binding activity; ATPase binding activity; and protein kinase activity. It is also involved in cellular protein metabolic process; intrinsic apoptotic signaling pathway in response to endoplasmic reticulum stress; and regulation of insulin secretion involved in cellular response to glucose stimulus (645). Moreover, rs560586208 ($p= 4.20*10^{-8}$) is intronic variant in the *FADS2* gene on chr11. This gene is a member of the *fatty acid desaturase* (*FADS*) gene family. Desaturase enzymes encoded by the members of this gene family regulate unsaturation of fatty acids through the introduction of double bonds between defined carbons of the fatty acyl chain and play important role in fatty acid metabolism (646). This chromosomal region includes other genes, such as the *Diacylglycerol Lipase Alpha* (*DAGLA*) gene that encodes a diacylglycerol lipase. The encoded enzyme by the *DAGLA* gene is involved in the biosynthesis of the

endocannabinoid 2-arachidonoyl-glycerol (647). Also, the adjacent *RAB3A Interacting Protein Like 1* (*RAB3IL1*) gene encodes a guanine nucleotide exchange factor for the ras-related protein Rab3A. The encoded protein binds Rab3a and the inositol hexakisphosphate kinase InsP6K1. This gene is associated with Inflammatory Bowel Disease 2 (648). The identified loci on chr11 to be significantly associated with OA in the Newfoundland population contain possible candidate genes that may be implicated in the pro-inflammatory, cartilage deterioration, fatty acids metabolism, and proteolysis that may increase the risk of OA development and progression.

Furthermore, the first locus found to be significantly $p=2.65*10^{-8}$ associated with OA on chr12 contains four variants (rs72298257, rs57991786, rs1552423, and rs201085636). These variants are ~2 Kbps apart and in LD correlated with $r^2=1$ in intron 5 of the *TMEM132D* gene. This gene may serve as a cell-surface marker for oligodendrocyte differentiation and neurotransmission that regulates expression and anxiety-related behavior (649). Previous studies identified variants in the *TMEM132D* gene to be associated with panic-disorder (650) and anxiety related behaviour (651). While SNP rs142897716 on the second locus of chr12 that was significantly associated ($p= 1.91*10^{-9}$) with OA is intergenic falls adjacent to the *MMP17* gene. The *MMP17* gene is involved in breakdown of extracellular matrix in normal physiological processes, such as embryonic development, reproduction, and tissue remodeling, as well as in disease processes, such as arthritis and metastasis (652). Elevated expression of the *MMP17* gene has been observed in OA and multiple human cancers (653). A study performed by Davidson *et al.* (2006) identified *MMP17* gene to be significantly upregulation

(P=0.0146) in the synovium and cartilage tissues of OA patients compared to the fracture in the neck of femur (NOF). These findings suggested a role of the *MMP17* gene in the cartilage degradation in OA patients (654).

SNP rs78941287 on chr16 was also significantly associated ($p= 4.96*10^{-9}$) with OA. This SNP is intergenic variant that is adjacent to the *CA5A* and *BANP* genes. The *CA5A* gene participates in a variety of biological processes, including respiration, calcification, acid-base balance, bone resorption, and the formation of aqueous humor, cerebrospinal fluid, saliva, and gastric acid (655). While *BANP* acts as a tumor suppressor and cell cycle regulator, it encodes a protein that binds to matrix attachment regions and forms a complex with p53 and negatively regulates p53 transcription (656). These genes have not previously been linked to OA mechanisms.

Additionally, INDEL, rs569627639 on chr19 was significantly associated ($p= 1.09*10^{-8}$) with OA. This insertion is an intronic variant in the *MED16* gene that is involved in the regulated transcription of nearly all RNA polymerase II-dependent genes (657). However, this gene has not reported to be associated with OA, other genes in this chromosomal region include the *PLPPR3*; *AZU1*; *PRTN3*; and *ELANE* genes may play an important role in the OA pathogenesis. The *PLPPR3* gene encodes an integral membrane protein modulating bioactive lipid phosphates including phosphatidate, lysophosphatidate, and sphingosine-1-phosphate in the context of cell migration, neurite retraction, and mitogenesis (658). *AZU1* acts in the Innate Immune pathway and is an important multifunctional inflammatory mediator (659). Previous studies identified the *AZU1* gene

to be hypomethylated that acts as activator in the inflammatory process in RA (660). *PRTN3* enables enzyme binding activity; serine-type endopeptidase activity; and signaling receptor binding activity. It is involved in several processes, including mature conventional dendritic cell differentiation; membrane protein ectodomain proteolysis; and neutrophil extravasation (661). *ELANE* forms a subfamily of serine proteases that hydrolyze many proteins in addition to elastin. This enzyme may play a role in degenerative and inflammatory diseases through proteolysis of collagen-IV and elastin (662). Bioinformatics analysis using protein-protein interaction identified upregulation of the *ELANE* gene in OA patient and suggested this gene to be among the potential diagnostic biomarkers for OA (663). SNP rs34837947 was also significantly associated ($p= 4.81*10^{-8}$) with OA. This SNP is a synonymous variant (c.CCC>CCT) in the *DPF1* gene on chr19. The *DPF1* gene may have an important role in developing neurons by participating in regulation of cell survival that acts as a neuron specific transcription factor (664). This genetic region has also other genes, such as the *SIPA1L3*, *PPP1R14A*, and *PSMD8* genes. The *SIPA1L3* gene has been implicated in regulation of cell adhesion, cell polarity, and organization of the cytoskeleton (665). The *PPP1R14A* gene encodes a protein that is an inhibitor of smooth muscle myosin phosphatase and has higher inhibitory activity when phosphorylated. Inhibition of myosin phosphatase leads to increased myosin phosphorylation and enhanced smooth muscle contraction (666). The *PSMD8* gene plays a key role in the maintenance of protein homeostasis by removing misfolded or damaged proteins, which could impair cellular functions, and by removing proteins whose functions are no longer required. Therefore, the proteasome participates in numerous cellular processes, including cell cycle progression, apoptosis, or DNA damage

repair (667). A study performed by Yuan *et al.* (2021) used differentially expressed genes (DEGs) and differentially expressed immune-related genes (DEIRGs) analysis of the immune cells in the OA tissues and found 15 genes including the *PSMD8* gene to be high diagnostic biomarkers of OA (668).

Finally, the significantly associated variant with OA on chr20 was rs76236562 ($p= 1.38*10^{-8}$). This SNP is an intergenic variant locates in gene reach region that includes the *NTSR1*, *OGFR*, *COL9A3*, *DIDO1*, and the *GID8* genes. The *NTSR1* gene has a role in diacylglycerol regulated cell signaling pathways that may play an important role in the final stages of mast cells (MCs) development and mast cells differentiation. The MCs are long-living multifunctional innate immune cells that originate from hematopoietic precursors and specifically differentiate in the destination tissue, e.g., skin, respiratory mucosa, intestine, where they mediate immune cell recruitment and antimicrobial defense (669). The *OGFR* gene is a negative regulator of cell proliferation and tissue organization in a variety of processes. Also, it seems to be involved in growth regulation (670). The *COL9A3* gene encodes one of the three alpha chains of type IX collagen, the major collagen component of hyaline cartilage. Type IX collagen, a heterotrimeric molecule, is usually found in tissues containing type II collagen, a fibrillar collagen. The *COL9A3* gene acts in the collagen chain trimerization and Extracellular matrix organization pathways (671). Multiple studies reported significant association of the *COL9A3* cartilage collagen gene variants with knee and hip OA suggesting important role of this gene in cartilage degradation and progression of OA (672–675). While the *DIDO1* gene is an

important apoptosis and Tumor suppressor gene (676), and the *GID8* gene is involved in positive regulation of canonical Wnt signaling pathway and positive regulation of cell population proliferation (677).

However, the association analysis of knee OA patients only identified three genetic variations on chr10 to be significantly associated with knee OA at the GWAS significance level, the identified variants to be the associated with OA in the knee OA were also associated with OA in the whole NFOAS cohort. These results suggested that the association of these genetic variants with OA in our study is due to knee OA. While the association of genetic variants with hip OA did not identify variations to be significantly associated with hip OA due to the small hip OA sample size (n=171).

This study has a number of limitations, the small sample size (557 OA patients and 118 unaffected controls) is the main study's limitation, giving that NL population is a small isolated population, and assembling large OA sample size for a GWAS analysis was not possible. Thus, there is a possibility of false positive rate inflation that may explain the identification of 28 novel variants to be significantly associated with OA in the NL population, while only one variant was reported in previous OA GWAS studies in Caucasian populations, and the NL population is not that different than other Caucasian populations to account for all these differences. However, our association analysis was conducted using the MilorGWAS approach that uses the MLR and estimates SNP variant effect size for structured populations with a stratified QQ-plot, enhancing the diagnosis of *p*-values inflation or deflation when population strata are not clearly identified in the

study sample. The analysis included the top 10 PCs as fixed effects in the MLR model to control type I error and completely correct for population structure (598) which makes it less likely to be false positive association between the identified variants and OA. Also, the study OA-free controls were self-reported, and a misclassification among these unaffected controls may only lead to null results. This situation is not a major issue for the significant results, but we may miss other signals that might be significant if there was not misclassification in controls. Moreover, due to the heterogeneity of OA, our analysis was performed on the whole OA, knee, and hip OA cohort without taking in consideration the possible endotypes or sub-phenotypes that also may complicate the analysis. Although, knee and hip OA share a number of risk factors, the aetiology might be different between knee and hip OA. Lastly, all the study participants were from NL, which is a genetically/ethnically homogeneous population that may limit the generalizability of our results to other populations.

In conclusion, our study identified novel genetic variations to be potentially associated with OA that play a critical role in the cartilage degradation, subchondral bone remodeling, skeletal muscle weakness, immune system, pro-inflammatory pathway, and OA pain. While confirmation is required, these findings provided new insights into the pathogenesis of OA and novel targets for developing OA management strategies.

Chapter 6: Whole exome sequencing analysis identified novel genes and digenic interaction of osteoarthritis in Newfoundland & Labrador population

Salem Werdyani[1], Quan Li[4], Ming Liu[1], Amanda Dohey[1], Dianne Codner[1], Darren O'Rielly[3], Proton Rahman[2], Guangju Zhai[1]

[1]Division of Biomedical Sciences (Genetics),
[2]Discipline of Medicine,
Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada.
[3]Molecular Genetics and Translational Genomics Laboratory, Eastern Health, St. John's, NL, Canada.
[4]Princess Margaret Cancer Centre, University Health Network, University of Toronto, Toronto, Ontario, Canada.

6.1 Co-authorship statement:

GZ and SW were responsible for the study conception and design; bioinformatics and statistical expertise; analysis, and interpretation of the results. ML, PR, and GZ participated in specimens; data collecting; and assembly. PR, and GZ were responsible for the provision of study materials or patients. AD, DO, and PR whole exome sequencing analysis. GZ obtained the study funding. ML, PR, GZ provided the administrative, technical, and logistic support.

6.2 Abstract:

Background: Although GWAS is the approach of choice for discovering the genetic basis of complex diseases, it has several challenges. GWAS analysis requires development of robust study designs, sufficient sample sizes, rigorous phenotypes, comprehensive maps, accurate high-throughput genotyping technologies, sophisticated IT infrastructure, rapid algorithms for data analysis, and rigorous assessment of genome-wide signatures. Moreover, despite that GWAS cover hundreds of thousands of individuals, it still has a degree of missing heritability. To date, the GWAS meta-analyses have identified more than 140 genetic risk variants to be associated with OA, but the aetiology and pathogenesis of OA remains elusive. Since about 85% of disease-causing variants are within protein coding or gene regulatory regions, whole exome sequencing (WES) approach is often used to identify genetic variants in protein coding regions. Due to the multifactorial inheritance of OA, no single gene is involved in the development of the disease, and multiple genes could contribute to its onset. Thus, we

conducted the current genetic variation annotation study to identify novel genes for OA by the WES analysis and performed a genome-wide case-only test for detecting digenic interaction in the WES data of 200 OA patients from the NFOAS.

Methods: Study participants were total knee or hip replacement patients due to primary OA who were recruited to the NFOAS before 2017 in St. John's, Canada. Patients' blood DNA was extracted and sequenced by the Illumina® NovaSeq 6000 sequencing facility at the Faculty of Medicine, Memorial University of Newfoundland. The GATK Best Practices pipeline was followed to align the WES raw paired end reads to the GRCh37 human reference genome, produce master binary alignment map (BAM) files, and create variants calling format (VCF) files. The resulted VCF files were used for the functional variant annotation using ANNOVAR software. Quality control (QC) filtering was set to exclude genetic variants with ≤ 20X overall depth of coverage and ≤ 5X for alternative allele coverage. Further, variants having minor allele frequency (MAF) ≤ 0.01 and ≥ 0.1 in the general population based on the 1000 genomes project, ExAC, and gnomAD databases were excluded. Finally, only variants present in ≥ 80% of OA patients based on the MAF were considered as potential OA associated genetic variants.

Additional to variant annotation analysis, a genome-wide case-only analysis was performed in this project to investigate a two-locus genome-wide digenic interaction in the WES data of the 200 OA patients. This approach tests all pairs of genes to detect gene-pair interactions underlying the OA condition. The master BAM files were used in the FreeBayes software (1.3.6) from the Broad Institute with base quality ≥ 10 and

mapping quality ≥ 10 cut off threshold to call for all possible genotypes in the WES data and create the VCF files. Then, the OA WES data in the form of VCF file was annotated with the GRCh37.75 reference genome dataset using the SnpEff tool and the GnomAD dataset with SnpSift program, and the QC filtering was performed using SnpEff to: a) retain variants that have depth of coverage (DP) > 8X; b) genotype quality (GQ) > 20; c) minor read ratio (MRR) < 0.2; and d) call-rate > 95%. Also, only diallelic variants from autosomal regions with missense or loss of function (LOF), including frameshift, stop_gained, stop_lost, start_lost, splice_donor, and splice_acceptor with the gnomAD bioinformatics "PASS" filter status were used for further analysis. Moreover, variants with a MAF < 5% in the 1000 genomes project database were included in the analysis. Finally, all the information was provided in the 2*2 contingency table for OA patients, the analysis followed a dominant mode of inheritance for each gene, and the OA patients carrying at least one copy of at least one allele within a gene were considered carriers of the gene variation using the generalized linear regression (glm) to test for gene*gene interaction for every pairwise combination of 17,803 genes using the SeqArray and gdsfmt packages in R (4.1.2). Also, to control for the type I error inflation in this case-only test, variants and genes separated $\delta < 2$ Mb were excluded after the LD analysis, and three principal components were used to adjust for population stratification and to increase the study power.

Results: 144 knee and 56 hip OA patients from the NFOAS (mean age 62.83□7.61 years, and 51.5% of them were females) were included in the study. A total of 526,459 raw

genetic variants were identified in the cohort. After the QC, 92 of them were identified in ≥ 80% of the OA patients with a MAF of ~ 0.49, for both knee and hip OA. The MAFs of these variants in all public available databases were ≤ 0.06. Ten exonic nonsynonymous single nucleotide variants (SNVs) located in *IGSF3, ZNF717, PRSS1, AQP7,* and *ESRRA* were estimated to have significant damaging effects on the related proteins' structure and function by > 4 functional and pathogenic prediction tools. These genes play a central role in metabolism of water-soluble vitamins and cofactors, aquaporin-mediated transport, glucagon signaling in metabolic regulation, regulation of lipolysis in adipocytes, and the extracellular matrix degradation. Interestingly, these genes have not been reported in previous OA GWAS studies. In the genome-wide case-only digenic interaction analysis, rs56158521 common exonic nonsynonymous SNV in the *HLA-DRB1* gene on chr6 was detected to be significantly (all $P \leq 7.70 * 10^{-5}$) interacting with aggregated rare variants in each of the *CDH19*, *SOGA1*, *MORC4*, *TMTC4*, and *ANK3* genes.

Conclusion: The identified genes in this study play central roles in bone mineral density, skeletal muscles' cells differentiation, angiogenesis, morphogenesis, cellular communication, cellular proliferation, oxidative stress, pro-inflammatory cytokines, and immune response in the articular cartilage and synovial fluid of OA patients. While confirmation is required, these findings provided new insights into better understanding of OA pathogenesis and hold promise as druggable targets for developing OA therapies.

6.3 Introduction:

In the past few years, the approach of choice was the GWAS analysis that is a powerful method for discovering the genetic basis of complex diseases (678,679). This approach is a hypothesis-free methodology that aims to identify associations of genotypes on a microarray with phenotypes by testing for differences in the allele frequency of genetic variants between individuals who are ancestrally similar but having different phenotypes. GWAS scans hundreds of thousands or millions of genetic variants across the whole genome to find those statistically associated genetic variations with a specific trait or disease. It typically reports genomic risk loci of correlated SNPs that all show a statistically significant association with the trait of interest (680). GWAS results have accounted for confounding genetic group differences in epidemiological studies, gaining insight into a phenotype's underlying biology, estimating a phenotype's heritability, calculating genetic correlations, and making clinical risk predictions (681). Moreover, it has become accepted as the way forward in the search for susceptibility loci of common diseases, such as autoimmune diseases and cancer and allows for the interrogation of lesser studied regulatory regions (682,683). However, the GWAS has several challenges. The GWAS analysis requires the development of robust study designs to ensure high power to detect genes of modest risk while minimizing the potential of false association signals due to testing large numbers of markers. Moreover, the GWAS needs sufficient sample sizes, rigorous phenotypes, comprehensive maps, accurate high-throughput genotyping technologies, sophisticated IT infrastructure, rapid algorithms for data analysis, and rigorous assessment of genome-wide signatures (678). Also, despite the

GWAS covering hundreds of thousands of individuals, it still has a degree of missing heritability (684). Furthermore, previous GWAS analysis have shown that most traits are influenced by thousands of causal variants that individually confer very little risk, are often associated with many other traits and are correlated with causal and non-causal variants that are physically close in LD, making direct biological, causal inferences complicated. Additionally, genetic associations may differ across ancestries, complicating direct comparisons between groups of individuals. Some of these limitations hamper drawing unambiguous conclusions about the biological meaning of GWAS results, sometimes limiting their utility to produce mechanistic insights or to serve as starting points for drug development (681). Hence, in addition to the common disease-common variant hypothesis there are other potential mechanisms, such as rare genetic variants gene*gene interactions that has been applied to complex human diseases (685).

The human genome comprises of approximately 3,200 Mbps, only 1% of them code for proteins. Based on the information from the 1000 genomes project, individuals differ from the reference human genome (GRCh37) at about 4.1 to 5.0 million sites (592). The majority of the genetic variants found naturally among populations are not disease causing and are found in non-protein coding regions. While it has been estimated that 85% of disease-causing variants are within protein coding or gene regulatory regions (686). A fraction of exonic variants (n=11,000) contains synonymous variants that do not lead to change in amino acid sequence in the polypeptides (592), while about 10,000 to 12,000 variants alter or truncate protein sequence (592). Therefore, when we look for

disease causing variants, whole exome sequencing (WES) approach is often used to identify genetic variants in protein coding regions.

Back in 1965, Fred Sanger and his co-workers developed a technique to sequence RNA fragments (Sanger, Brownlee and Barrell, 1965). This technique used two-dimensional (2D) fractionation and radiolabelling to detect partial-digestion fragments (687). This 2D fractionation method was used in 1972 to produce the first protein-coding gene sequence, followed by the complete genome sequence of bacteriophage MS2 (688). After several rapid developments in the field of genetic sequencing, Sanger developed the 'chain-termination' approach in 1977, using deoxyribonucleoside triphosphates (dNTPs) to produce sequences of increasing length that are labeled at the final termination with dideoxyribonucleotide for detection and therefore deduce the targeted DNA sequence (689). This method was called Sanger sequencing and formed the basis for first-generation sequencing technology that has been recognized as the gold standard of DNA sequencing for nearly 40 years.

The next-generation sequencing (NGS) approach have evolved from the traditional fragment-based Sanger sequencing technologies. The NGS is massively parallel shotgun sequencing technology, it generates flow cells to bind DNA strands that allows for the creation of paired-end sequencing, which reads 100s of millions of short reads, providing greater accuracy when mapping to reference sequences (690). The NGS methods have been continually developing, advances in the technology gradually led to the development of longer read sequencing, and importantly, lower costs at a greater read depth, making

the technology viable in a research and clinical environment with larger scales of genomic sequences available to analyze than ever previously known (691). The NGS technologies can sequence genomes with millions to billions of base pairs in days to weeks versus the first-generation technology that took 13 years to sequence an estimated 3.3 billion base pair human genome (691,692). In addition to its lower costs when compared to whole genome sequencing (WGS), aligned and merged WES data are more compact. An average aligned whole exome binary alignment map (BAM) file (at 60x coverage) takes about 4GB compared to a whole genome BAM file (at 60x coverage) which takes about 150GB of disk space. These factors have made WES more practical in a clinical and diagnostics setting, where potential candidate genes may not be covered by a targeted panel, or for heterogeneous diseases such cardiac diseases and intellectual disabilities (692). Thus, as sequencing costs continue to drop, it became feasible to perform genome-wide studies on WGS data. Several studies make use of lower coverage WGS to detect rare variants or structural variants, then impute these into pre-existing GWAS data (693). This opens new avenues for exploring rare variants in large cohorts and may provide a balance between rare variant detection and the affordability of genotyping.

To date, the GWAS meta-analyses have identified more than 140 genetic risk variants to be associated with OA (196,331,334,694), but the aetiology and pathogenesis of OA still remains elusive. Thus, we conducted the current study to identify novel genes for knee and hip OA by the WES analysis from the well-established NFOAS.

Also, due to the multifactorial inheritance of OA, it is hypothesized that no single gene is involved in the development of the disease, and multiple genes could contribute to OA onset. Evidence of genetic architecture in OA has been determined by the epidemiological studies of family history and clustering, twin studies, linkage scan and candidate genes studies, and GWAS analysis. However, OA is often based on enrichment of individual rare variants or their aggregate burden in affected individuals. Based on the information from the Digenic Diseases DAtabase (DIDA), none of the previous studies investigated the digenic (gene*gene) interaction implication in the development and progression of OA. We hypothesized that two-genes interaction may identify digenic implication in the OA pathogenesis and performed a genome-wide case-only test for detecting digenic interaction in the WES data of 200 OA patients from the NFOAS.

6.4 Materials and methods:

This study was performed as part of the NFOAS, detailed description of the study participates, and their demographic information were stated in the materials and methods of chapter 5, sections 5.4.1-3.

6.4.1 Whole Exome Sequencing (WES):

Genomic DNA was extracted from patients' whole blood. Exome sequencing was performed at the Rahman/O'Rielly lab, Faculty of Medicine, Memorial University of Newfoundland on libraries prepared from 100ng genomic DNA using the Agilent SureSelect XT Human All Exon V7 with Low Input Fragmentation Kit (Agilent

Technologies). Briefly, DNA was enzymatically fragmented, ends repaired and dA-Tailed. Unique molecular-barcoded adaptors were ligated to the fragments and samples purified with AMPure XP beads (Beckman Coulter). Individual P7 Index primers were then added during amplification of the adaptor-ligated library. Following purification, the quantity and quality was assessed using Tape Station 2200 and D1000 reagents (Agilent Technologies). Hybridization of 1000ng of the gDNA library to target-specific probes was followed by capture of the hybridized DNA using Dynabeads TM MyOne TM Streptavidin T1 beads (Thermo Fisher Scientific). Captured libraries were further amplified, purified and quantity and quality assessed with the Tape Station 2200 and High Sensitivity D1000 reagents. Individual libraries were size adjusted and normalized before pooling in batches of 100 samples for sequencing on the Illumina© NovaSeq 6000 with S2, 200 cycle, V1.0 reagent kits for 2x100bp reads along with PhiX spiked in at 1% to aid in assessing sequencing quality.

This paired-end sequencing method sequences both ends of a fragment and generates high-quality, alignable sequence data. In addition to producing twice the number of reads for the same time and effort in library preparation, it facilitates accurate read alignment to detect common DNA rearrangements such as insertions, deletions, and inversions, as well as gene fusions and novel transcripts, which is not possible with single-read data (695). Also, paired-end DNA sequencing reads provide high-quality alignment across DNA regions containing repetitive sequences and produce long contigs for *de novo* sequencing by filling gaps in the consensus sequence (695).

6.4.2 Raw Exome sequencing data processing:

The genome analysis toolkit (GATK) Best Practices pipeline was followed to perform the WES data pre-processing that is required prior to all variant annotation and discovery analysis, **Figure 6.1**. Initially, the Picard's FastqToSam program was used to read the WES raw paired end reads in form of FASTQ files and produce the un-aligned BAM files. Then, the Picard's MarkIlluminaAdapters program was used to mark the Illumina adapter sequencing by adding the existing adapter-trimming tags (XT) to a read record, marking the 5' start position of the specified adapter sequence, and producing a metrics file. Some of the marked adapters come from concatenated adapters that randomly arise from the primordial soup of the polymerase chain reaction (PCR), others represent read-through to 3' adapter ends of reads and arise from insert sizes that are shorter than the read length.

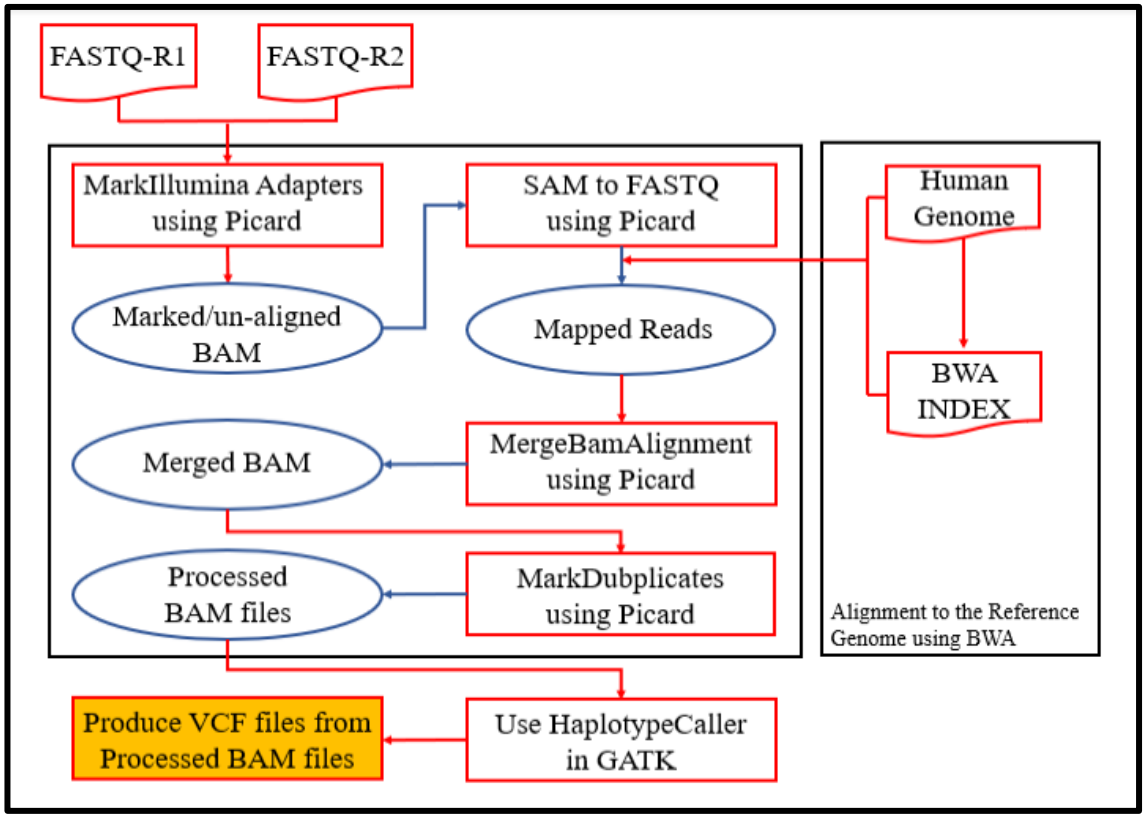


Figure 6.1: A flow chart illustrates the steps of the GATK Best Practices pipeline performed to process the WES raw data from the paired-end sequencing sequence FASTQ to VCF file format.

In some instances, read-though can affect the majority of reads in a sample, and make these reads unmappable by certain aligners. Picard's SamToFastq tool was used to read the un-aligned BAM files and take the XT tag, read identifiers, read sequences, and base quality scores to effectively remove adapter sequence contribution to downstream read alignment and alignment scoring metrics. Then, it produced the Sanger FASTQ files that are text-based format commonly used to store meta sequencing data, such as read group information, alignment information, flag, and tag values. A FASTQ file normally uses

four lines per sequence. The first line begins with a '@' character and is followed by a sequence identifier and an optional description like a FASTA title line, the second line contains the raw sequence letters, the third line begins with a '+' character and is optionally followed by the same sequence identifier and again any description, and the fourth line encodes the quality values for the sequence in line 2, and it contains the same number of symbols as letters in the sequence, **Figure 6.2**.

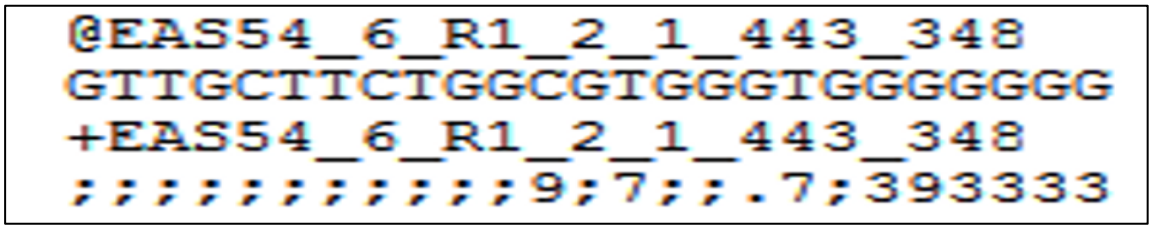


Figure 6.2: A screenshot of the FASTQ file format showing the sequence identifier, and an optional description in the first line, the raw sequence letters in the second line, the same sequence identifier and again any description in the third line, and the quality values for the sequence from the second line is listed in the fourth line.

Subsequently, the BWA-MEM algorithm version 0.7.17 was used to align the sequencing reads data for each individual in the format of FASTQ files to the GRCh37 human reference genome, which is a synthetic single-stranded representation of common genome sequence that is intended to provide a common coordinate framework for all genomic analysis (696). Hence, the alignment process revealed aligned sequencing alignment/map (SAM) files that are TAB-delimited text file format storing large nucleotide short read sequence alignments. SAM file consists of a header and alignment

sections. If present, the header must be prior to the alignments part and starts with '@' character, while the alignment lines do not. The header is followed by the read alignment information, an automatically generated program group record and reads sorted in the same order as the input FASTQ file. Each alignment line has 11 mandatory fields including query template name, bitwise flag, reference sequence name, 1-based leftmost mapping position, mapping quality, CIGAR string, reference name of the mate/next read, position of the mate/next read, observed template length, segment sequence, and ASCII of phred-scaled base quality+33, **Figure 6.3**.

```
@HD VN:1.5 SO:coordinate
@SQ SN:ref LN:45
r001   99 ref  7 30 8M2I4M1D3M = 37  39 TTAGATAAAGGATACTG *
r002    0 ref  9 30 3S6M1P1I4M *  0   0 AAAAGATAAGGATA    *
r003    0 ref  9 30 5S6M       *  0   0 GCCTAAGCTAA       * SA:Z:ref,29,-,6H5M,17,0;
r004    0 ref 16 30 6M14N5M    *  0   0 ATAGCTTCAGC       *
r003 2064 ref 29 17 6H5M       *  0   0 TAGGC             * SA:Z:ref,9,+,5S6M,30,1;
r001  147 ref 37 30 9M         =  7 -39 CAGCGGCAT         * NM:i:1
```

Figure 6.3: A screenshot of the SAM file format showing the header lines that start with @ character followed by the read alignment information that has 11 mandatory fields including query template name, bitwise flag, reference sequence name, 1-based leftmost mapping position, mapping quality, cigar string, ref. name of the mate/next read, position of the mate/next read, observed template length, segment sequence, and ascii of phred-scaled base quality.

Then, the GATK MergeBamAlignment algorithm was used to merge defined information from the un-aligned BAM files with that of the aligned BAM files to conserve read data, including original read information and base quality scores. This tool created a sequence

dictionary file from a reference sequence provided in FASTA format, which is required by many processing and analysis tools. The MergeBamAlignment program also generated additional meta information based on the information generated by the BWA-MEM aligner, which may alter aligner-generated designations, such as mate information and secondary alignment flags. The tool then made adjustments to match all meta information, including read and mate strand information based on proper mate designations.

Moreover, the GATK MarkDuplicates program was utilized to identify read pairs that are likely to have originated from duplicates of the same original DNA fragments through some artifactual processes. These are considered to be non-independent observations, so the program tags all but a single read pair within each set of duplicates, causing the marked pairs to be ignored by default during the variant discovery process. Also, the resulting BAM files were sorted using Picard tools, and INDEL realignment was performed by GATK to produce a master BAM files (697). These BAM files are compressed binary format of SAM files that are increasingly used as a space-saving alternative of the FASTQ files for containing the short raw read data. The BAM files contain a header and alignments sections. The header starts with '@' character and contains information about the entire file, such as sample name and sample length. While the alignments section contains read name, read sequence, read quality, and custom tags, **Figure 6.4**.

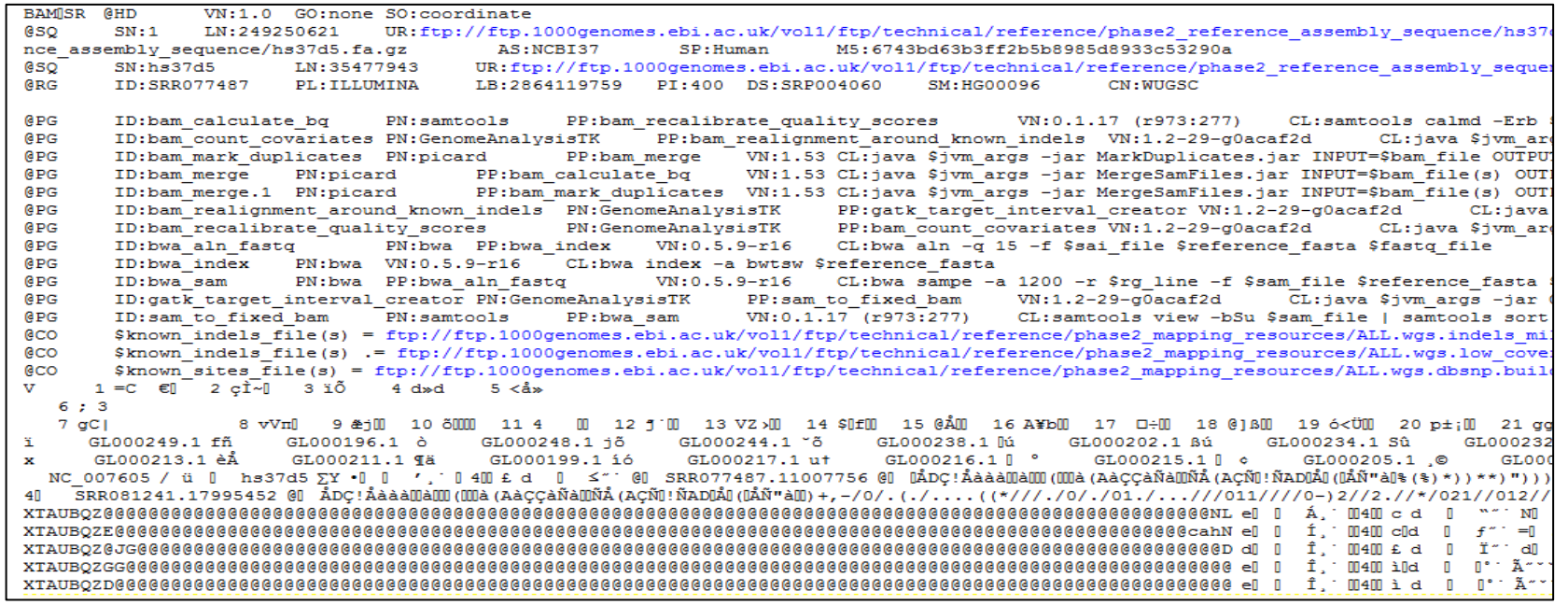


Figure 6.4: A screenshot of the BAM file format showing the header that starts with @ character and contains information about the entire file, such as sample name and sample length, and the alignments section contains read name, read sequence, read quality, and custom tags in binary format.

Variants calling format (VCF) files were also created using the GATK HaplotypeCaller program that call variants based on their precise alignment (696). The VCF is a text format file that is used in bioinformatics analysis for storing gene sequence variations. It contains meta information lines and describes the INFO, FILTER, and FORMAT entries used in the body of the VCF file, Header line, and data lines containing variant information, such as, chromosome number, position, reference allele, and alternated allele, **Figure 6.5**.

Afterward, machine learning analysis was utilized to detect and correct for patterns of systematic errors in the base quality scores, which are confidence scores emitted by the sequencer for each base (696). Base quality scores play an important role in weighing the evidence for or against possible variant alleles during the variant discovery process, so it's important to correct any systematic bias observed in the data. Biases can originate from biochemical processes during library preparation and sequencing, from manufacturing defects in the chips, or instrumentation defects in the sequencer. The recalibration procedure involves collecting covariate measurements from all base calls in the dataset, building a model from those statistics, and applying base quality adjustments to the dataset based on the resulting model. The initial statistics collection was parallelized by scattering across genomic coordinates, typically by chromosome or batches of chromosomes, but this can be also broken down further if needed to boost throughput (696). Then, the per-region statistics were gathered into a single genome-wide model of covariation; that was not parallelized, but it was computationally trivial, and therefore not a bottleneck. Finally, the recalibration rules derived from the model were applied to the

```
##fileformat=VCFv4.0
##fileDate=20110705
##reference=1000GenomesPilot-NCBI37
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=.,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID         REF  ALT      QUAL FILTER INFO                              FORMAT      Sample1         Sample2         Sample3
2      4370    rs6057     G    A        29   .      NS=2;DP=13;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:52,51 1|0:48:8:51,51
1/1:43:5:.,.
2      7330    .          T    A        3    q10    NS=5;DP=12;AF=0.017               GT:GQ:DP:HQ 0|0:46:3:58,50 0|1:3:5:65,3    0/0:41:3
2      110696  rs6055     A    G,T      67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2    2/2:35:4
2      130237  .          T    .        47   .      NS=2;DP=16;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:56,51 0/0:61:2
2      134567  microsat1  GTCT G,GTACT  50   PASS   NS=2;DP=9;AA=G                   GT:GQ:DP    0/1:35:4        0/2:17:2        1/1:40:3
chr1    45796269         .       G       C
chr1    45797505         .       C       G
chr1    45798555         .       T       C
chr1    45798901         .       C       T
chr1    45805566         .       G       C
chr2    47703379         .       C       T
chr2    48010488         .       G       A
chr2    48030838         .       A       T
chr2    48032875         .       CTAT    -
chr2    48032937         .       T       C
```

Figure 6.5: A screenshot of the VCF file format that is a text format file. VCF file contains meta information lines that starts with ## characters and describe the INFO, FILTER and FORMAT entries used in the body of the VCF file. Also, the VCF contains the header line that starts with # character and data lines containing variant information, such as, chr number, position, reference, and alternated variant, variant sequencing quality, filter, and information.

original dataset to produce a recalibrated dataset. This was parallelized in the same way as the initial statistics collection, over genomic regions, then followed by a final file merging to produce a single analysis ready VCF file per sample (696).

6.4.3 Variant Annotation analysis:

The resulted VCF files were then used for the functional annotation of candidate variants to acquire genetic information using Annotate Variation (ANNOVAR) software. ANNOVAR program facilitates fast and easy update to date information to functionally annotate genetic variant in a form of SNVs and INDELs. It identifies variants functional consequence throughout the available biological databases. Hence, it eliminates the genetic variants that are unlikely to be disease causing and identifies the putative genes implicated in disease throughout gene-based, region-based, and filter-based annotations analysis using the command-line tool (698).

Gene-based annotation analysis examines variants functional consequence on genes base on the information from Reference Sequence (RefSeq) genes (699); University of California, Santa Cruz (UCSC) genes (700); the European Bioinformatics (ENSEMBL) genes (701); genome research of the Encyclopedia of DNA Elements (GENCODE) genes (702); or the molecular annotations of the transcripts (AceView) genes (703). It identifies whether SNP, CNV, or INDEL cause protein coding changes and detects the possibly affected amino acids (698).

Region-based annotation analysis of ANNOVAR infers cytogenetic bands that are containing the identified variants and identifies variants in specific genomic regions, including conserved regions among 44 species, predicted transcription factor binding sites, segmental duplication regions, GWAS hits, database of genomic variants, DNAse I hypersensitivity sites, the Encyclopedia of DNA Elements (ENCODE) H3K4Me1/H3K4Me3/H3K27Ac/CTCF sites, and ChIP-Seq peaks, RNA-Seq peaks (698).

Also, ANNOVAR filter-based annotation analysis finds whether the identified variants are documented in specific databases, such as single nucleotide polymorphism database (dbSNP) (704), and it identifies variants' MAF in the general population based on the 1000 genomes project (592), exome aggregation consortium (ExAC) (705), and genome aggregation database (gnomAD) (706). Also, it reports the functional importance scores of identified variants based on the pathogenicity prediction bioinformatics tools including the Sorting Intolerant From Tolerant (SIFT) algorithm (707), Polymorphism Phenotyping v2 (PolyPhen-2) method (708), Likelihood Ratio Test (LRT) (709), MutationTaster application score (710), MutationAssessor score (711), the Functional Analysis through Hidden Markov Models (FATHMM) (712), Protein Variation Effect Analyzer (PROVEAN), phylogenetic within PHAST (PhyloP) conservation score (713,714), the Genomic Evolutionary Rate Profiling (GERP++) (715), and the Combined Annotation-Dependent Depletion (CADD) methods (716,717).

Recently, CADD score became a widely recommended and used measure that uses a machine learning model to effectively prioritize the deleteriousness of genetic variants in form of SNVs and INDELs in the genetic analyses (717). CADD score is based on diverse genomic features derived from surrounding sequence context including introns or UTRs (717,718), evolutionary score from SIFT (707) and GERP++ (715), and functional predictions information on variants in both coding from PolyPhen-2 (708) and non-coding epigenetic measurements from ENCODE (719) and Roadmap Epigenomics Mapping Consortium (720).

ANNOVAR analysis begins with converting the list of genetic variations included in the VCF files into the ANNOVAR input (*.avinput) file format using (convert2annovar.pl) Perl program from the ANNOVAR package; then, conducting ANNOVAR annotation by using the (table_annovar.pl) Perl program from the ANNOVAR package to annotate the genetic variants; and performing the QC filtering using inhouse developed Perl programs pipeline. The QC analysis was performed to: a) retain the genetic variants with ≥ 20X overall and with ≥ 5X alternative allele depth of coverage; b) include the SNVs and INDELs with MAF ≥ 0.01 and ≤ 0.1 in the general population based on the 1000 genomes project (592), ExAC (705), and gnomAD databases (706); c) retain intronic, splicing, and exonic variants. For the exonic variants, missense and nonsense changes that are frameshift insertions/deletions, non-frameshift insertions/deletions, stop_gain, stop_loss, or nonsynonymous SNV were retained; d) retain possibly deleterious variants based on at least four pathogenicity prediction bioinformatics tools (721) including SIFT

score > 0.95 (707), PolyPhen-2 score > 0.5 (708), LRT score 0-1 (709), MutationTaster score > 0.5 (710), MutationAssessor score > 0.65 (711), FATHMM score ≥ 0.45 (712), PROVEAN score < -2.5 ,PhyloP score > 1.6 (713,714), and GERP++ score > 4.4 (715), in addition to CADD score > 15 (716,717); and e) include variants identified in at least 80% of the OA patients to be considered in further analyses.

6.4.4 Case-only digenic interaction analysis:

Digenic interaction is the simplest genetic model with alleles at two different loci being necessary and sufficient to determine disease status (722). This approach used the gene as the unit of analysis and tests all pairs of genes to detect pairwise (gene*gene) interactions underlying a disease (723–725).

Additional to variant annotation analysis, a genome-wide case-only analysis was performed in this project to investigate two-locus genome-wide digenic interaction in the WES data of 200 OA patients using the case-only digenic interaction in the WES data pipeline developed by Kerner *et al* (2020) (726). This approach uses a classic variant aggregation approach to combine multiple rare variants within a gene as the unit of analysis to address the lack of power inherent to studies of rare variants Then, it assumes that aggregated rare variants at a first locus (first gene) interact with a common variant at a second locus (second gene) to detect gene-pair interactions underlying diseases, and the cohort allelic sums test (CAST) approach was used to perform burden tests, allowing for further improved statistical power (726).

To conduct the case-only digenic interaction test, the master clean BAM files that were produced by the GATK Best Practices pipeline were used in the FreeBayes software (1.3.6) from the Broad Institute (727) with base quality ≥ 10 and mapping quality ≥ 10 cut off threshold to call for all possible genotypes in the WES data and create the VCF files.

FreeBayes is a haplotype-based Bayesian genetic variant calling software that uses short-read alignments BAM files with Phred+33 encoded quality scores for any number of individuals from a population and a reference genome FASTA file to determine the most-likely combination of SNVs, and INDELs genotypes for the population at each position in the FASTA reference genome. Then, it reports positions which are putatively polymorphic in VCF format (727).

Following the creation of VCF files by the FreeBayes software, the OA WES data in the form of VCF files was annotated with the GRCh37.75 reference genome dataset using the SnpEff tool and the GnomAD dataset with the SnpSift program (728).

Subsequently, the QC filtering was performed using the SnpEff software to: a) retain variants that have depth of coverage (DP) > 8; b) genotype quality (GQ) > 20; c) minor read ratio (MRR) < 0.2; and d) call-rate > 95%. Also, only diallelic variants from autosomal regions to be missense or loss of function (LOF), including frameshift, stop_gained, stop_lost, start_lost, splice_donor, and splice_acceptor with the gnomAD bioinformatics “PASS” filter status were kept for further analysis (706). Moreover, rare

variants, those with a MAF < 0.05 in the 1000 genomes project database were included in further analysis as suggested by the case-only digenic test pipeline developer. Finally, the analysis followed a dominant mode of inheritance for each gene, and OA patients carrying at least one copy of at least one allele within a gene were considered carriers of the gene variation (726).

Also, to control for the type I error inflation in this case-only test, the pairs of variants and genes separated $\delta < 2$ Mb were excluded after the LD analysis, and three principal components were identified by PLINK(V2.0) and used to adjust for population stratification and to increase the study power based on the pipeline developer recommendations (586). Therefore, this case-only test yielded reasonable type I error rates like those for the analogous case-control approach by limiting the analysis to pairs of genes with $\delta > 2$ Mb to avoid the LD and adjustment for population stratification (726).

Furthermore, the SnpSift tool (728) was used to create a VCF file for rare variants with MAF < 0.0001 for the first locus, and a VCF file for common variants having MAF > 0.01 for the second locus based on the GnomAD allele frequency in the general population (706). For the first locus, only genes with at least 0.05 rare variant carriers were retained, to ensure sufficient study power (726). While for the second locus, sliding windows of 100 kb were used to identify variants in strong LD ($r^2 > 0.6$) and remove them from the analysis (726).

Finally, all the information was provided in the 2*2 contingency table for OA patients, **Table 6.1**. Also, the digenic analysis using generalized linear regression (glm) was performed to test for gene*gene interaction for every pairwise combination of 17,803 genes using the SeqArray and gdsfmt packages in R (4.1.2) (729,730).

Table 6.1: The 2*2 contingency table used for the interaction analysis between the aggregated rare variants in each gene and each variant from other genes under the assumption of dominant inheritance using the generalized linear regression (glm) in R (4.1.2).

	Gene 2	
Gene 1	Carriers	Non-carriers
Carriers	Gene 1 and 2	Gene 1 but not 2
Non-carriers	Not gene 1 but 2	Neither gene 1 and 2

6.5 Results:

144 knee and 56 hip OA patients from the NFOAS (mean age 62.83□7.61 years, mean BMI 34.24□7.16 kg/m^2, and 51.5% of them were females) were included in this study. A total of 526,459 raw genetic variants were identified in the cohort after the variant annotation analysis using ANNOVAR software. We found 401,815 variants including 196,621intronic; 150,416 exonic; 14,618 intergenic; 12,433 UTR3'; 9,278 UTR5'; 9,115

ncRNA_intronic; 3,857 ncRNA_exonic; 2,909 upstream; 1,426 downstream; 865 splicing; 154 upstream/downstream; 51 exonic/splicing; 38 ncRNA_splicing; 29 UTR5'/UTR3'; and 5 ncRNA_exonic/splicing variants that passed the QC filtering and were used in the further variant prioritizing analysis.

Out of these variants, 55 were identified in ≥ 80% of the OA patients with a MAF of ~ 0.4 for both knee and hip OA (**Table 6.2)**, and the MAFs of these variants in all public available databases including the 1000 genomes project (592), ExAC (705), and gnomAD databases (706) were ≤ 0.097.

Table 6.2: List of the 55 variants that were found in at least 80% of the OA patients from the NFOAS with MAF ≥ 0.4 for both knee and hip OA.

Chr	Position	Variant ID	Ref_allele	Alt_allele	Variant Function	Exonic Variant Function	Gene	OA MAF	Global MAF	Functional prediction tools	CADD score
1	12907400	rs149302457	C	T	exonic	nonsynonymous SNV	HNRNPCL1; 3; and 4	0.49	0.05	4	23.8
1	16901757	rs3872317	C	A	intronic	.	NBPF1	0.49	0.05		.
1	117142613	rs76151115	G	A	exonic	nonsynonymous SNV	IGSF3	0.47	0.04	5	15.93
1	117142641	rs76417519	C	T	exonic	nonsynonymous SNV	IGSF3	0.48	0.03	8	26.4
1	143767628	rs6604514	C	T	exonic	nonsynonymous SNV	PPIAL4G; PPIAL4H	0.40	0.06		.
1	143767878	rs7513869	C	T	UTR5	.	PPIAL4G; PPIAL4H	0.46	0.00		.
1	145295539	rs3872115	A	G	ncRNA_intronic	.	NBPF25P	0.48	0.00		.
1	146250905	rs587717447	G	C	intronic	.	NBPF10; 19; and 20	0.43	0.06		.
2	96517824	rs76217973	C	T	intronic	.	ANKRD36C	0.45	0.01		.
2	132905664	rs201427646	G	T	UTR3	.	ANKRD30BL	0.44	0.08		.
2	132905682	rs199535616	T	C	UTR3	.	ANKRD30BL	0.49	0.04		.
2	132905693	rs150889759	T	C	UTR3	.	ANKRD30BL	0.49	0.09		.
3	75786252	rs140641854	C	A	exonic	nonsynonymous SNV	ZNF717	0.46	0.00	6	23.5
7	142247530	rs764426432	GC	-	exonic	frameshift deletion	TRBV73	0.45	0.01		.
7	142247535	rs751581456	-	AA	exonic	frameshift deletion	TRBV73	0.47	0.01		.
7	142247540	rs748246110	G	A	exonic	.	TRBV73	0.47	0.01		.
7	142247546	rs575564328	C	A	exonic	.	TRBV73	0.47	0.01		.
7	142247567	rs372792593	A	G	UTR5	.	TRBV73	0.48	0.04		
7	142247605	rs370122363	G	A	intronic	.	MIR11400; MTRNR2L6	0.47	0.06		.
7	142458873	rs373388800	A	C	intronic	.	PRSS1	0.50	0.07		.
7	142458881	rs376074412	A	T	intronic	.	PRSS1	0.50	0.02		.

7	142458884	rs370405149	C	G	intronic	.	PRSS1	0.50	0.03		.
7	142458886	rs373808731	T	A	intronic	.	PRSS1	0.50	0.02		.
7	142458887	rs368632869	T	C	intronic	.	PRSS1	0.50	0.02		.
7	142458898	rs371109898	A	T	intronic	.	PRSS1	0.50	0.01		.
7	142458906	rs374242167	T	G	intronic	.	PRSS1	0.50	0.01		.
7	142458967	rs367567937	A	T	intronic	.	PRSS1	0.50	0.05		.
7	142458972	rs371766102	T	C	intronic	.	PRSS1	0.50	0.05		.
7	142458987	rs375094420	C	G	intronic	.	PRSS1	0.50	0.02		.
7	142460238	rs749152254	G	A	intronic	.	PRSS1	0.48	0.02		.
7	142460251	rs374503586	AC	-	intronic	.	PRSS1	0.49	0.02		.
7	142460255	rs376880201	C	-	intronic	.	PRSS1	0.49	0.02		.
7	142460690	rs370873919	-	TCTT	intronic	.	PRSS1	0.47	0.02		.
7	142460744	rs150930992	G	C	exonic	nonsynonymous SNV	PRSS1	0.47	0.09	8	22.5
7	142499836	rs782572802	T	C	ncRNA_exonic	.	TCRVB	0.45	0.02		.
7	142499854	rs782379154	T	C	ncRNA_exonic	.	TCRVB	0.45	0.04		.
7	142499862	rs782010780	C	T	ncRNA_exonic	.	TCRVB	0.45	0.03		.
9	33385690	rs139024279	C	A	exonic	nonsynonymous SNV	AQP7	0.48	0.05	9	25.4
9	33385698	rs145516206	T	C	exonic	nonsynonymous SNV	AQP7	0.46	0.05	9	24.1
9	33794798	rs773515866	GA	-	exonic	frameshift deletion	PRSS3	0.46	0.04		.
9	33794907	rs761203400	TGAAACA	-	intronic	.	PRSS3	0.41	0.07		.
9	33794916	rs754499022	TCC	-	intronic	.	PRSS3	0.42	0.06		.
10	126678063	rs767562982	G	A	UTR3	.	CTBP2	0.47	0.02		.
11	1017035	rs754561621	-	AT	exonic	frameshift insertion	MUC6	0.42	0.10		.
11	64083320	rs201072913	T	C	exonic	nonsynonymous SNV	ESRRA	0.44	0.03	9	25
11	64083328	rs79204587	C	T	exonic	nonsynonymous SNV	ESRRA	0.44	0.05	9	26.6
11	64083331	rs80310817	C	T	exonic	nonsynonymous SNV	ESRRA	0.44	0.05	9	25.2

16	70010517	rs372253115	TGT	-	ncRNA_exonic	.	PDXDC2P-NPIPB14P	0.50	0.05		.
16	70977695	rs12149139	T	C	intronic	.	HYDIN	0.44	0.02		.
16	70998808	rs8056408	G	A	intronic	.	HYDIN	0.50	0.02		.
16	74425703	rs200974911	G	A	exonic	nonsynonymous SNV	NPIPB15	0.42	0.04	1	20.9
17	45234430	rs78072949	A	G	exonic	nonsynonymous SNV	CDC27	0.48	0.01	3	20.8
19	8999386	rs77049866	A	G	intronic	.	MUC16	0.41	0.06		.
19	8999449	rs76798407	G	T	exonic	nonsynonymous SNV	MUC16	0.45	0.05	2	16.3
20	29623147	rs777364323	C	T	ncRNA_exonic	.	FRG1BP; FRG1DP	0.41	0.05		.

Ref_allele: Reference allele; Alt_allele: Alternative allele; OA MAF: Minor allele frequency in the OA patients; Global MAF: Minor allele frequency in the global population; CADD: The Combined Annotation-Dependent Depletion methods.

Ten exonic nonsynonymous SNVs were found to be deleterious based on at least four pathogenicity prediction bioinformatics tools and CADD score > 15, **Table 6.3**. These variants included rs149302457 (c.C743T, p.A248V) in exon1 of the *Heterogeneous Nuclear Ribonucleoprotein C Like 1* (*HNRNPCL1*) gene on chr1 that was found in 196 (98%) of the OA patients with MAF = 0.49, while its MAF in the general population was 0.05. This variant was predicted to be damaging based on four functional prediction tools and CADD score = 23.8. This analysis also identified rs76151115 (c.G2039A:p.R680Q) in exon8 and rs76417519 (c.C1951T:p.R651W) in exon7 of the *Immunoglobulin Superfamily Member 3* (*IGSF3*) gene on chr1. Interestingly, these two variants are 28 bps apart. Variant rs76151115 was detected in 191 (95.5%) of the OA cohort with MAF = 0.47 in OA patients and MAF = 0.04 in the general population, This SNV was predicted to be deleterious based on five pathogenicity prediction bioinformatics tools and CADD score = 15.93, while the missense mutation (rs76417519) was found in 192 (96%) of the OA samples with MAF = 0.48 in OA patients and global MAF = 0.03. Based on the score from six functional prediction tools and CADD score = 26.4, rs76417519 was predicted to be damaging variant. Moreover, we identified rs140641854 (c.C2522A, p.P841H) in exon5 of the *Zinc Finger Protein 717* (*ZNF717*) gene on chr3, this SNV was detected in 183 (91.5%) of the OA patients with MAF of 0.46 in the OA cohort and MAF = 0.003 in the general population. It was predicted to be damaging based on four pathogenicity prediction bioinformatics tools and CADD score = 23.5. Also, rs150930992 (c.G617C, p.C206S) in exon5 of the *Serine Protease 1* (*PRSS1*) gene on chr7 was found in 189 (94.5%) of the OA patients with MAF = 0.47 in the OA samples and global MAF = 0.086. The SNV rs150930992 was predicted to be damaging based on

seven functional prediction tools and CADD score = 22.5. Moreover, variants rs139024279 (c.C529A, p.R177S) and rs145516206 (c.T521C, p.L174P) in exon6 of the *Aquaporin 7* (*AQP7*) gene on chr9 were identified to associate with OA. SNV rs139024279 was found in 190 (95%) of the OA patients with MAF = 0.48 in the OA samples and approximately 0.05 globally, while rs145516206 was identified in 183 (91.5%) of the OA patients with MAF = 0.46 in the OA cohort and 0.048 in the general population. The identified variants in the *AQP7* gene were predicted to be damaging based on nine pathogenicity prediction bioinformatics tools and CADD score ≥ 24.1. Additionally, rs201072913 (c.T1154C, p.L385P), rs79204587 (c.C1162T, p.L388F), and rs80310817 (c.C1165T, p.R389C) in exon7 of the *Estrogen Related Receptor Alpha* (*ESRRA*) gene on chr11 were found in 176 (88%) of the OA patients with MAF of 0.44 in the OA samples and ≤ 0.05 in the global population. These three variants were assumed to be deleterious based on nine functional prediction tools and CADD score ≥ 25. This analysis also detected two frameshift deletions and two frameshift insertions to be associated with OA, **Table 6.3**. These variants included the deletion rs764426432 (c. 35_36del, p.C12Sfs*17) and the insertion rs751581456 (c.30_31insTT, p.L11Ffs*29) in exon1 of the *T Cell Receptor Beta Variable 7-3* (*TRBV73*) gene on chr7. The deletion rs764426432 was identified in 187 (93.5%) of OA samples with MAF = 0.45 in OA patients, while its MAF in the general population was 0.012, while the insertion rs751581456 was detected in 189 (94.5%, MAF = 0.47) of OA samples and global MAF = 0.01. Furthermore, the deletion rs773515866 (c.9_10del, p.E5Dfs*28) in exon2 of the of the *Serine Protease 3* (*PRSS3*) gene on chr9 was detected in 194 (97%) of the OA patients with MAF=0.46 in OA cohort, but MAF = 0.04 in the general population.

Table 6.3: List of the variants that were prioritized in our WES analysis to be associated with OA.

Chr	Position	Variant ID	Ref_allele	Alt_allele	Variant Function	Exonic Variant Function	Gene	OA MAF	Global MAF	Functional prediction tools	CADD score
1	12907400	rs149302457	C	T	exonic	nonsynonymous SNV	HNRNPCL1;3; and 4	0.49	0.05	4	23.8
1	117142613	rs76151115	G	A	exonic	nonsynonymous SNV	IGSF3	0.47	0.04	5	15.93
1	117142641	rs76417519	C	T	exonic	nonsynonymous SNV	IGSF3	0.48	0.03	6	26.4
3	75786252	rs140641854	C	A	exonic	nonsynonymous SNV	ZNF717	0.46	0.00	4	23.5
7	142247530	rs764426432	GC	-	exonic	frameshift deletion	TRBV73	0.45	0.01	.	.
7	142247535	rs751581456	-	AA	exonic	frameshift insertion	TRBV73	0.47	0.01	.	.
7	142460744	rs150930992	G	C	exonic	nonsynonymous SNV	PRSS1	0.47	0.09	7	22.5
9	33385690	rs139024279	C	A	exonic	nonsynonymous SNV	AQP7	0.48	0.05	9	25.4
9	33385698	rs145516206	T	C	exonic	nonsynonymous SNV	AQP7	0.46	0.05	9	24.1
9	33794798	rs773515866	GA	-	exonic	frameshift deletion	PRSS3	0.46	0.04	.	.
11	1017035	rs754561621	-	AT	exonic	frameshift insertion	MUC6	0.42	0.10	.	.
11	64083320	rs201072913	T	C	exonic	nonsynonymous SNV	ESRRA	0.44	0.03	9	25
11	64083328	rs79204587	C	T	exonic	nonsynonymous SNV	ESRRA	0.44	0.05	9	26.6
11	64083331	rs80310817	C	T	exonic	nonsynonymous SNV	ESRRA	0.44	0.05	9	25.2

Finally, our analysis identified the insertion rs754561621 (c.5765_5766insAT, p.S1923Ffs*63) in exon31 of the *Mucin 6* (*MUC6*) gene on chr11 that was found in 169 (84.5%) of OA samples with MAF = 0.42 in OA patients, while its global MAF was 0.097.

In the genome-wide case-only digenic interaction analysis, 369,853,788 tests were conducted on the WES data of 200 OA patients to investigate gene*gene interaction between 17,803 genes that included 38,731 variants (on average, 2 rare variants per gene) following the QC filtering. As a result, one exonic nonsynonymous SNV (rs56158521, c.G208A, p.D70N) in the *HLA Class II Histocompatibility Antigen, DR-1 Beta Chain* (*HLA-DRB1*) gene on chr6 was detected to be significantly ($P \leq 7.70* 10^{-5}$) interacting with aggregated rare variants in each of five genes, **Table 6.4**. The most significant gene*gene interaction ($P=9.45* 10^{-6}$, OR=13.75) was of rs56158521 SNV (MAF=0.139 in the European population) with aggregated rare variants in each of the *Cadherin 19* (*CDH19*), the *Suppressor of Glucose, Autophagy Associated 1* (*SOGA1*), and the *MORC Family CW-Type Zinc Finger 4* (*MORC4*) genes. This analysis revealed strong enrichment of rare heterozygous variants in the *CDH19, SOGA1,* and *MORC4* genes among 27 out of the 200 OA patients (13.5%), twelve of these 27 OA patients were also carriers of the rs56158521, **Figure 6.6**. As well as, rs56158521was significantly interacting with the aggregated rare heterozygous variants in the *Transmembrane O-Mannosyltransferase Targeting Cadherins 4* (*TMTC4*, $P=4.30* 10^{-5}$, OR=8.56) gene. We found that 28 of the 200 OA study participants (14%) having strong enrichment

Table 6.4: The digenic interaction results listing the significant interaction between the aggregated rare variants in each of the CDH19, SOGA1, MORC4, TMTC4, and ANK3 genes separately with the rs56158521 common missense variant in the HLA-DRB1 gene.

Chr	Position	Variant ID	OA MAF	Global MAF	BP change	AA change	Gene 1	Gene 2 variants	OR	P-value
6	32552048	rs56158521	0.12	0.14	c.208G>A	p.Asp70Asn	HLA-DRB1	CDH19	13.75	$9.45*10^{-6}$
6	32552048	rs56158521	0.12	0.14	c.208G>A	p.Asp70Asn	HLA-DRB1	SOGA1	13.75	$9.45*10^{-6}$
6	32552048	rs56158521	0.12	0.14	c.208G>A	p.Asp70Asn	HLA-DRB1	MORC4	13.75	$9.45*10^{-6}$
6	32552048	rs56158521	0.12	0.14	c.208G>A	p.Asp70Asn	HLA-DRB1	TMTC4	8.56	$4.30*10^{-5}$
6	32552048	rs56158521	0.12	0.14	c.208G>A	p.Asp70Asn	HLA-DRB1	ANK3	17.23	$7.70*10^{-5}$

Chr: chromosome number, OA MAF: minor allele frequency in the OA patients, Global MAF: minor allele frequency in the general population based on the data from the GnomAD database, BP change: change of the DNA base, AA change: change of the amino acid, Gene 1: the first gene containing the common variant, Gene 2 variants: aggregated rare variants in the second gene, OR: Odd Ratio.

of the rare heterozygous variants in the *TMTC4* gene, twelve of them also carried the rs56158521, **Figure 6.6**. Finally, rs56158521was found to significantly interacting ($P=7.70*10^{-5}$, OR=17.23) with the aggregated rare heterozygous variants in the *Ankyrin 3* (*ANK3*) gene. Strong enrichment of the rare heterozygous variants in the *ANK3* gene were identified in 77 out of 200 OA patients (38.5%), and 27 of them were carriers of the rs56158521 common SNV in the *HLA-DRB1* gene, **Figure 6.6**.

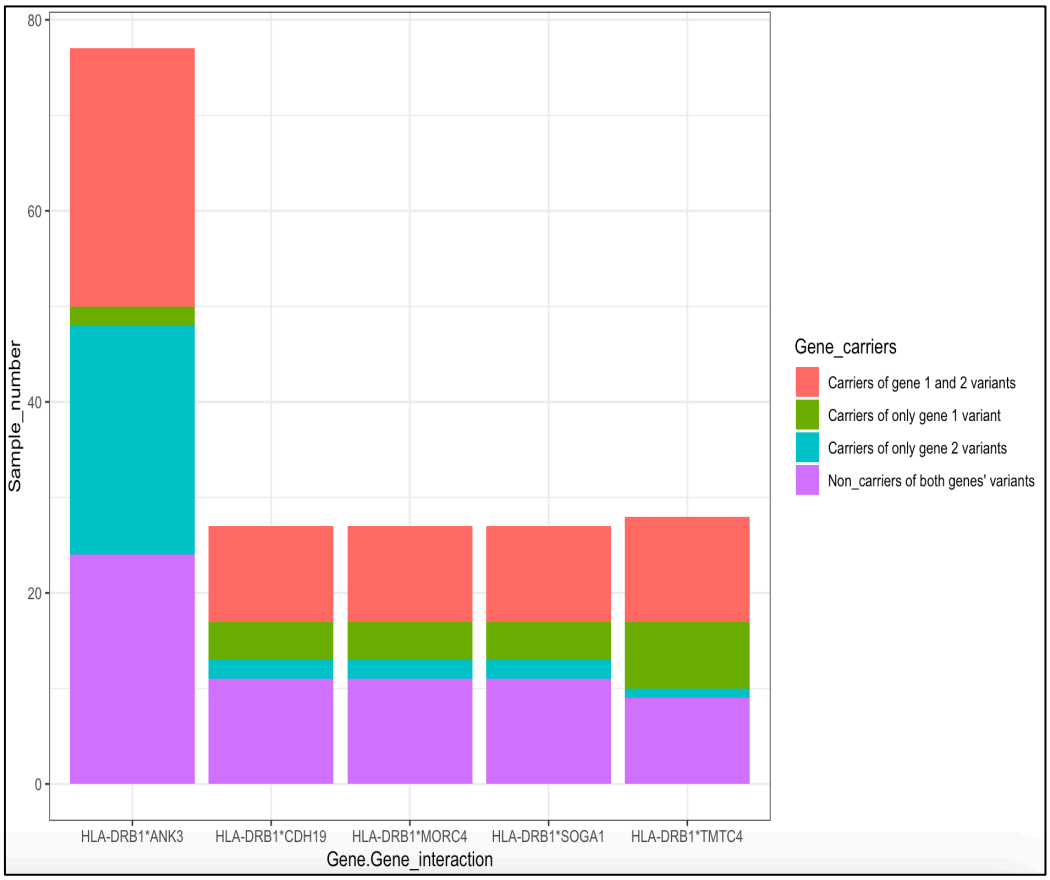


Figure 6.6: Bar chart illustrates the number of OA samples carrying the (rs56158521, c.G208A, p.D70N) variant in the HLA-DRB1gene on chr6 and its interaction with the aggregated variants in each of the CDH19, SOGA1, MORC4, TMTC4, and ANK3 genes.

6.6 Discussion:

To our knowledge, this study was the first study that used WES data for variant annotation analysis and the first ever study that investigated the genome-wide case-only digenic (gene*gene) interaction in OA. Interestingly, all of the identified genetic variations in our study are novel and have not been reported previously to be linked with OA development and progression.

Our variant annotation analysis using ANNOVAR software identified ten deleterious missense mutations to be associated with OA based on the pathogenicity prediction bioinformatics tools. SNV (rs149302457) in exon1 of the *HNRNPCL1* gene was detected in 98% of the OA patients with MAF = 0.49. *HNRNPCL1* gene is located on 1p36.21 and codes for a protein that may play a role in nucleosome assembly by neutralizing basic proteins (731). This protein also has a glycine rich arginine-glycine-glycine region called the (RGG) box which enables protein and RNA binding. It was estimated to affect many critical genes that are responsible for controlling metabolic pathways at the transcriptional, post-transcriptional, translation, and post-translation levels (731). Multiple studies reported the involvement of the *HNRNPCL1* gene up regulation in different cancer types and their lower prognosis (732–734). However, *HNRNPCL1* gene was not reported previously to play a role in musculoskeletal conditions, our findings suggested the implication of the *HNRNPCL1* gene in the OA development and progression through the regulation of the genes in the metabolic pathways.

Two missense damaging variants in the *IGSF3* gene including rs76151115 and rs76417519 were also identified in this study. These variants were found in ≥ 95.5% (MAF ≥ 0.47) of OA patients, but they were rare (MAF ≤ 0.04) in the global population. The *IGSF3* gene that is located on 1p13.1 is a member of the immunoglobulin (Ig) superfamily. The protein encoded by this gene is an immunoglobulin-like membrane protein containing several V-type Ig-like domains with a molecular background in immune synapse. IGSF3 belongs to a novel Ig subfamily containing a Glu-Trp- Ile (EWI) motif not seen in other Ig proteins. This Ig subfamily play a role in the T-cell activation and IL-17 production regulation in inflammatory bowel, rheumatoid diseases, and autoimmune disorders (735). Also, the IGSF3 has been known to regulate neuronal morphogenesis that might function through interactions with multiple partners including the tetraspanin TSPAN7 (736). We hypothesized that these damaging variants in the *IGSF3* gene are possibly implicated in the OA inflammatory pathway and pain mechanisms.

Moreover, rs140641854 (c.C2522A, p.P841H) in exon5 of the *ZNF717* gene on chr3 was common (91.5%) in the OA patients with MAF = 0.46, while it has MAF = 0.003 in the general population. The *ZNF717* gene encodes a Kruppel-associated box (KRAB) zinc-finger protein, which belongs to a large group of transcriptional regulators in mammals. These proteins bind nucleic acids and play important roles in various cellular functions, including cell proliferation, and differentiation (737). Recent studies identified a role of the *ZNF717* gene in osteogenic differentiation of mesenchymal stem cells (738). Also, mutations in the *ZNF717* gene were identified in patients with growth retardation and

intellectual disability (739). A study performed by Alkhateeb *et al.* (2021) identified mutations in the *ZNF717* gene to be implicated in the multiple sclerosis that is a chronic inflammatory autoimmune disease leading to neurodegenerative processes that cause neuron demyelination (740). Another study by Lamot *et al.* (2017) identified six pathogenic variants in the *ZNF717* gene that are implicated in the clavicular cortical hyperostosis rare inflammatory bone disorder (741). These studies suggested a possible critical role of the *ZNF717* gene in the inflammatory pathway and bone deformity. Thus, our findings suggested that the *ZNF717* gene abnormalities may lead to stimulate the pro-inflammatory and bone growth defect in OA patients.

Also, we identified the rs150930992 in exon5 of the *PRSS1* gene in 994.5% of the OA patients with MAF = 0.47 in the OA cohort and global MAF = 0.086. Interestingly, our results also identified the deletion rs773515866 (c.9_10del, p.E5Dfs*28) in exon2 of the of the *PRSS3* gene that was detected in 194 (97%) of the OA patients with MAF=0.46 in OA cohort, but MAF = 0.04 in the general population. The *PRSS1* gene is localized to the T cell receptor beta locus on chr7, while the *PRSS3* gene was translocated from chr7 and localized to the locus of the T cell receptor beta variable orphans on chr9 (742). The *PRSS1* and *PRSS3* genes encode trypsinogens, which are members of the trypsin family of serine proteases (743). The PRSS1 enzyme is secreted by the pancreas and cleaved to its active form in the small intestine. While the PRSS3 enzyme is expressed in the brain and pancreas and is resistant to common trypsin inhibitors. Both serine proteases are active on peptide linkages involving the carboxyl group of lysine or arginine. The *PRSS1* gene was reported to act in the metabolism of water-soluble vitamins and cofactors and

extracellular matrix organization pathways. Mutations in this gene are associated with hereditary pancreatitis (743). Moreover, a CNV study identified a heterozygous fusion of exons 1 and 2 from the *protease serine 2* (*PRSS2*) gene with exons 3-5 from the *PRSS1* gene that causes gout in a family from the French Caucasian ancestry (744). Also, the *PRSS1* gene has been found to play a role in the human chronic pain conditions (745,746). Moreover, a study performed by Bandesh *et al*. (2020) reported that mutations in the *PRSS3* gene were significantly associated ($p = 5.7*10^{-8}$) with type 2 diabetes (747). Another study detected the *PRSS3* gene to be implicated in autoimmune and mental disorders in the Danish population (748). Also, the expression of the *PRSS3* gene was reported to be 2.84 fold in RA patients compared to controls (749). Our results suggested that the *PRSS1* and *PRSS3* genes may be involved in OA pathogenesis through metabolic disorders and immunoinflammatory reactions, as well as implicated in the OA pain mechanisms.

The *AQP7* gene forms a channel that mediates water and glycerol transport across cell membranes at neutral pH (80). The channel is also permeable to urea (752). This gene plays an important role in body energy homeostasis under conditions that promote lipid catabolism, giving rise to glycerol and free fatty acids (750). The *AQP7* gene also mediates glycerol export from adipocytes. After release into the blood stream, glycerol is used for gluconeogenesis in the liver to maintain normal blood glucose levels and prevent fasting hypoglycemia (753,754). This gene was found previously to play a role in the hydrarthrosis, which is characterized by effusion of watery liquid into the cavity of a joint, and inflammatory synovitis (755). Studies in animal models investigated obesity-

induced inflammation in the pancreas using the *AQP7*-silenced rat β-cells stimulated by TNF-α and lipopolysaccharide (LPS). The *AQP7* gene as a main aquaglyceroporin in endocrine pancreas that is involved in insulin exocytosis was impaired by TNF-α along with an extreme reduction in insulin secretion, while it was upregulated by the LPS (756,755). These finding estimated the involvement of the *AQP7* gene in the immune cells' physiology and inflammation activation (756). Also, a study performed by Nagahara *et al.* (2010) in Japan revealed that the *AQP7* gene was expressed in the synovial tissues of 29 (91%) OA patients and nine (75%) RA patients of study subjects (754) suggesting the involvement of the *AQP7* gene in the inflammatory pathway activation and development of OA and RA (754). Moreover, Zhang *et al*. (2020) identified the *AQP7* gene among 118 up-regulated genes in the knee lateral tibial and medial tibial subchondral bone samples in OA patients from China (757). This study reported a role of the *AQP7* gene in the chemical homeostasis, intrinsic and integral component of plasma membrane, regulation of lipolysis in adipocytes, and the peroxisome proliferator-activated receptors signaling pathways (84). Interestingly, our variant annotation analysis findings came in agreement with these studies and detected rs139024279 and rs145516206 that are eight bps apart in exon6 of the *AQP7* gene on chr9 with high frequency ($\geq$ 91.5%, MAF $\geq$ 0.46) in study participants and global MAF $\leq$ 0.05. These findings may explain the implication of the *AQP7* gene in the cartilage and synovial fluid homeostasis, extracellular degradation, and inflammatory pathway activation in the OA patients from NFOAS.

Moreover, the *ESRRA* gene encodes a nuclear receptor that is most closely related to the estrogen receptor (758). This protein binds to an estrogen-related receptor-alpha response element containing a single consensus half-site, 5'-TNAAGGTCA-3'. Thus, it acts as a site-specific transcription factor and interacts with members of the peroxisome proliferator-activated receptor γ coactivator 1 alpha (PGC-1A) family of transcription cofactors and form the ERRalpha/PGC1A complex that regulates the expression of majority of cellular energy production genes in energy metabolism, as well as it acts in the process of mitochondrial biogenesis (759). It also induces the expression of the *PERM1* gene in the skeletal muscles (760). An animal study performed in 2021 identified systematic reduction of the mitochondrial functional genes including the *ESRRA* gene to be associated closely with skeletal muscle aging in mice. This analysis suggested that the *ESRRA* gene may play significant roles in the progression of skeletal muscle aging and serves as potential biomarker for future diagnosis and treatment (761). Another study reported that the *ESRRA* gene mutations inhibit the gene transcriptional activities and subsequently increase the risk of eating disorders (EDs) development including anorexia nervosa and bulimia nervosa (762). Also, a study performed by Tang *et al*. (2021) suggested an inflammatory cytokines, and growth factor's regulatory role of the *ESRRA* gene and thus exert significant effects on the occurrence and development of OA (763). Remarkably, our variant annotation analysis detected rs201072913 (c.T1154C, p.L385P), rs79204587 (c.C1162T, p.L388F), and rs80310817 (c.C1165T, p.R389C) in exon7 of the *ESRRA* gene on chr11 in 176 (88%) of the OA patients with MAF of 0.44 in the OA samples and ≤ 0.05 in the general population. Thus, we hypothesized that the identified variants in the *ESRRA* gene may play a critical role in the metabolic disorders, skeletal

muscles weakness, and pro-inflammatory pathway activation that are associated with development and progression of OA.

Also, this analysis detected two frameshift deletions and two frameshift insertions to be associated with OA. These frameshift variants included rs764426432 (c.35_36del, p.C12Sfs*17) and rs751581456 (c.30_31insTT, p.L11Ffs*29) in exon1 of the *TRBV73* gene on chr7. The deletion rs764426432 was identified in 187 (93.5%) of OA samples with MAF = 0.45 in OA patients, while its MAF in the general population was 0.012, insertion rs751581456 was detected in 189 (94.5%, MAF = 0.47) of OA samples and global MAF = 0.01. The *TRBV73* is a protein coding gene that is a part of the T cell receptor complex estimated to be involved in plasma membrane receptor signaling pathway. Alpha-beta T cell receptors (TRs) are antigen specific receptors which are essential to the immune response and are present on the cell surface of the T lymphocytes. Recognize peptide-major histocompatibility (pMH) complexes that are displayed by antigen presenting cells, a prerequisite for efficient T cell adaptive immunity against pathogens (764). Thus, three major signaling pathways including the calcium, the mitogen-activated protein kinase (MAPK), and the nuclear factor NF-kappa-B (NF-kB) pathways are recruited, leading to the mobilization of transcription factors that are critical for gene expression and essential for the T cell growth and differentiation (765). The T cell repertoire is generated in the thymus, by V-(D)-J rearrangement. This repertoire is then shaped by intrathymic selection events to generate a peripheral T cell pool of self-MH restricted, non-autoaggressive T cells. Post-thymic interaction of alpha-beta TR with the pMH complexes shapes TR structural and functional avidity (766).

Despite the critical role of the *TRBV73* gene in the immune response and gene transcription, it has not been reported previously to be implicated in autoimmune and musculoskeletal conditions. Our results suggested that due to consequences of aging in joints, abnormalities in the immune mechanisms cascade in the carriers of these identified frameshift variants in exon1 of the *TRBV73* gene may lead to the development and progression of OA.

Furthermore, the *MUC6* gene encodes a member of the mucin protein family. The *MUC6* gene encodes high-molecular-weight mucin that forms an insoluble mucous barrier of the protective mucous layer of the gastrointestinal tract (767). It is well known that the trefoil factor family (TFF) peptides and mucins are co-expressed in mucin-producing epithelia, especially of the gastrointestinal tract. For instance, the TFF3 interacts with MUC5AC in the human stomach and with MUC2 in the duodenum, while the TFF2 is co-produced with MUC6 in the human stomach and duodenum (768). The TFF1 is also bound to MUC5AC in human gastric mucosa (769). However, the molecular mechanisms of these interactions are still not fully understood, there are only a few studies concerning the expression of mucins in synovial tissue. For example, the TFF peptides were estimated to modify the viscosity of the synovial fluid in in mucin solutions and have been shown to support catabolic functions in OA (770). A study performed by Popp *et al.* (2019) reported that the TFF2 was significantly upregulated in RA and OA samples. While the TFF3 protein was significantly downregulated in OA patients in comparison to healthy individuals, and it was significantly upregulated in the RA compared to the OA patients (771). Moreover, other studies reported an upregulated MUC3 mucin expression, and the

expression of MUC5AC in synovial tissues of OA and RA patients compared to healthy tissues (772). Also, the MUC1 was identified in synovial membrane cells and mononuclear cells in RA synovial tissues, but not in OA, suggesting the implication of mucins in the RA and OA pathogenesis (773). Remarkably, Our analysis detected the insertion rs754561621 (c.5765_5766insAT, p.S1923Ffs*63) in exon31 of the *MUC6* gene on chr11 that was found in 169 (84.5%) of OA samples with MAF = 0.42 in OA patients, while its global MAF was 0.097. This finding suggests the possible involvement of the *MUC6* gene in the pathogenies of the synovial tissues and immunoinflammatory reactions in OA patients from NFOAS.

I also conducted in this project a genome-wide case-only digenic interaction in the OA patients using the WES data. Following the case-only digenic analysis pipeline developer recommendations, I used 200 OA samples in this analysis to optimize the study power (726). Although, the digenic (gene*gene) interaction of the OA has not been investigated previously based on the information from the DIgenic diseases Database (DIDA) (774), our analysis identified the common (MAF=0.14) missense rs56158521 variant in the *HLA-DRB1* gene to be significantly interacting ($p \leq 7.70*10^{-5}$) with aggregated rare variants in each of the *CDH19, SOGA1, MORC4, TMTC4*, and *ANK3* genes.

The *HLA-DRB1* gene belongs to the human leukocyte antigen (HLA) complex family of genes called histocompatibility complex class II (MHC class II) that provide instructions for making proteins on the surface of certain immune system cells (775). These proteins attach to polypeptides outside the cell that help the immune system to distinguish

between the body's own proteins and proteins made by foreign invaders, such as viral or bacterial proteins. If the immune system recognizes the peptides as foreign, it triggers a response to attack the invading viruses or bacteria (775). Each MHC class II gene has many possible variations, allowing the immune system to react to a wide range of foreign invaders. Different variations of the *HLA-DRB1* gene affect different amino acids in the beta chain. These changes occur near the antigen-recognizing binding groove, which is the part of the protein that binds to viral or bacterial peptides (776,777). This binding triggers the immune response that attacks foreign invaders.

Different transcripts of the *HLA-DRB1* gene have been associated with the increased susceptibility of type 1 diabetes (778) and many other autoimmune disorders, including pemphigus vulgaris (779), sarcoidosis (780), and RA (777). Multiple variations of the *HLA-DRB1* gene have been identified to elevate person's susceptibility to develop RA, and the variations of this gene are the most significant known genetic risk factor for the disease (781). Although the mechanism by which the *HLA-DRB1* variations increase the risk of RA is unclear, researchers suspect it to be related to changes in peptide binding that stimulate an abnormal immune response (782). Moreover, a GWAS conducted by Nakajima *et al.* (2010) in Japan identified two SNPs (rs7775228 and rs10947262) adjacent to the *HLA-DRB1* gene to be significantly associated with knee OA (301). Also, Moos *et al.* (2002) stated that due to up-regulation of the inflammatory cytokines including TNF-α, IL6, and/or IL-1β in chondrocytes of the OA patients, the *HLA-DRB1* gene is up-regulated in immunological response to the low-grade inflammation of OA (783). Furthermore, a study performed by Kooshkaki *et al.* (2020)

identified a significant association of the *HLA-DRB1*0101* gene variants with knee OA (784).

In this analysis two aggregated rare variants in the *CDH19* gene were identified to significantly interacting ($p=9.45*10^{-6}$) with the common rs56158521 in the *HLA-DRB1* gene in 27 (13.5%) of the OA patients. *CDH19* is one of three related type II cadherin genes situated in a cluster on chr18. This gene encodes a protein that is a calcium dependent cell-cell adhesion glycoprotein containing five extracellular cadherin repeats (785,786). Cadherins preferentially interact with themselves in a homophilic manner in connecting cells, and they may thus contribute to the sorting of heterogeneous cell types (785). The *CDH19* gene is also involved in vascular remodeling and plays a central role in the structural integrity of blood vessels (787). Recent studies have reported the involvement of classic cadherin in many complex processes including angiogenesis, morphogenesis, cellular communication, and cellular proliferation (788–790). Angiogenesis is fundamental to many physiological events including embryogenesis, growth, wound healing, and female reproductive cycle (791). A study performed by Niu *et al.* (2008) identified that the knockdown of *CDH12* and *CDH19* expression significantly inhibits monocyte chemotactic protein-1-induced protein (MCPIP) and suppresses the capillary-like tube formation in human umbilical vascular endothelial cells (HUVECs) (792). Inflammation and angiogenesis are closely associated in OA. Blood vessels from the subchondral bone invade the articular cartilage and stimulate the progression of forming osteophytes and development of OA (793). Angiogenesis also

associated with synovitis (794), it can impair chondrocyte function and homeostasis of the articular cartilage (795), in part by redistributing blood vessels away from the synovial surface and thus contributing towards articular hypoxia (796). Interestingly, Sun *et al.* (2014) reported a significant association between angiogenesis and downregulation of the *HLA-DRB1* gene in OA patients (797). Furthermore, a recent analysis conducted by Deng *et al.* (2022) identified the *CDH19* gene among the immune genes that were upregulated in the epicardial adipose tissue from patients with coronary artery disease (798).

Moreover, our analysis revealed that two aggregated rare variants in the *SOGA1* gene were significantly interacting ($p=9.45*10^{-6}$) with the common rs56158521 in the *HLA-DRB1* gene in 27 (13.5%) of the OA patients. The *SOGA1* gene acts in insulin receptor signaling pathway; negative regulation of gluconeogenesis; and regulation of autophagy by playing a role in the reduction of glucose production in an adiponectin- and insulin-dependent manner (799). Following to nutritional scarcity, autophagy stimulates the release of glucose from the liver by promoting the hydrolysis of proteins, glycogen, and triglycerides (800–802). Due to adiponectin-stimulated activation of the insulin signaling pathway in hepatocytes, expression of the *SOGA1* gene increases, and this increase contributes to the reduction of glucose production by inhibiting autophagy through not fully understood mechanism (803). Thus, SOGA1 may regulate glucose and glycogen metabolism by directly cooperating with glycogen synthase and the glycogen synthase-associated protein glycogenin (803). There is evidence that impaired glucose homeostasis

promotes release of the TNF-α and activate the Nf-kB, which cause inflammation and oxidative stress intracellularly and might promote cartilage degradation in the OA patients, in turn the immune response pathway is activated (804,805).

Additionally, our analysis revealed that two aggregated rare variants in the *MORC4* gene were significantly interacting ($p=9.45*10^{-6}$) with the common rs56158521 in the *HLA-DRB1* gene in 27 (13.5%) of the OA patients. The *MORC4* gene is a histone methylation reader binding to non-methylated (H3K4me0), monomethylated (H3K4me1), dimethylated (H3K4me2) and trimethylated (H3K4me3) 'Lys-4' on histone (806) that is located on the X-chromosome. This gene belongs to the microrchidia (MORC) family of CW zinc finger proteins. Compared with the other MORC members, MORC4 protein has an N-terminal ATPase-like ATP-binding region containing an HATPase-c domain, followed by a putative nuclear matrix-binding domain and a two-stranded coiled-coil motif near its C-terminus (807). In addition, MORC4 has 3 putative nuclear localization signals and 4 potential SUMOylation sites and is able to recruit targeting of C-terminal binding protein (CtBP) corepressors (807). MORC proteins participate in fundamental biologic processes ranging from DNA-based activities, such as transcription, replication, and recombination to ubiquitination and assembly of large protein complexes (806,807). They are widely expressed in testis formation and male germ cell development. Also, emerging evidence reveals an important role of the MORC members in cancer development and bone homeostasis (807). However, the *MORC4* gene was not reported previously to be associate with OA, previous studies reported the involvement of the *MORC4* gene in inflammatory diseases including Crohn's disease, ulcerative colitis

(808), tropical calcific pancreatitis (809), and alcohol or non-alcohol chronic pancreatitis. Differential expression of the *MORC4* gene was found in diffuse large B-cell lymphoma. The *MORC4* gene encodes OX-TES-4 antigen, which could elicit an antibody response in 50% of diffuse large B-cell lymphomas, suggesting a role for the *MORC4* as a potential lymphoma biomarker (807).

Also, we identified two aggregated rare variants in the *TMTC4* gene were significantly interacting ($p=4.30*10^{-5}$) with the common rs56158521 in the *HLA-DRB1* gene in 28 (14%) of the OA patients. The *TMTC4* gene encodes a transmembrane protein that belongs to a family of proteins containing an N-terminal transmembrane domain and a C-terminal tetratricopeptide repeat (TPR) domain. TPR domains mediate protein-protein interactions in various cellular processes, such as synaptic vesicle fusion, protein folding, protein translocation, and protein glycosylation pathways (810). The TMTC4 protein contains a potential N-terminal signal sequence, 10 and 12 hydrophobic segments and 11 and eight C-terminal TPR motifs, respectively possess five and three predicted N-linked glycosylation consensus sites (811). A transcriptome-wide association study identified an important role of the *TMTC4* gene in influencing BMD suggesting involvement of the *TMTC4* gene in the skeletal muscles satellite cell differentiation (812).

Finally, our analysis revealed eight aggregated rare variants in the *ANK3* gene to be significantly interacting ($p=7.70*10^{-5}$) with the same common rs56158521 in the *HLA-DRB1* gene in 77 (38.5%) of the OA patients. The *ANK3* is an immunologically distinct gene product from ankyrins 1 and 2 that participates in the maintenance and/or targeting

of ion channels and cell adhesion molecules at the nodes of Ranvier and axonal initial segments (813). However, the role of *ANK3* gene in inflammatory diseases, especially in arthritis, has not been studied thoroughly, a study performed by Liu *et al.* (2018) in China identified dramatical higher regulation of the *ANK3* gene in OA compared to the RA patients and healthy controls (814). Another study conducted by Long *et al.* (2019) reported a significantly higher expression of the *ANK3* gene in the OA compared to the RA patients and healthy controls from the Korean population (815). These findings suggested a role of the *ANK3* gene in the pro-inflammatory cytokines and immune response in the articular cartilage and synovial fluid of the OA patients.

Based on these results, we hypothesized that due to the impaired glucose homeostasis and age-related muscles weakness, immunoinflammatory reactions and angiogenesis are possibly stimulated in the pathogenesis of OA.

The main weakness of the functional variant annotation study using ANNOVAR software was the lack of OA-free controls from the NL population. Thus, we were not able to compare the MAF of the identified variants in ≥ 80% of the OA patients with their MAF in controls from the same population, and instead we compared the MAF of the identified variants in our OA patients with their MAF in the general (Caucasian) population from the single nucleotide polymorphism database (dbSNP), the 1000 genomes project, exome aggregation consortium (ExAC), and genome aggregation database (gnomAD).

It is worth mentioning that rare variant association analyses in complex diseases using sequencing data have several limitations (816,817). The limitation of these analyses include how to account for population structure, control for unbalanced case-control designs, and model different data types in a statistical test to address the challenges of inflated type I error and increase the statistical power of the rare variants test (816). However this case-only approach is based on the aggregation of rare variants within a gene as a unit of analysis to address the lack of power inherent to the studies of rare variants, reduce the number of tests required relative to testing at allele level, and potentially decrease the amount of computer time required (726). Moreover, case-only model has multiple advantages over classic case-control tests by avoiding recruitment of controls and its potential associated bias. Also, case-only method is more powerful than the corresponding case-control test for detecting digenic interactions in various population stratification scenarios when common variants are tested for interaction in the context of GWAS, and case-only test is a powerful and timely tool for detecting digenic interaction in WES data from patients (726).

While confirmation is required, the findings of our genome-wide case-only digenic interaction analysis provided new insights into better understanding of knee and hip OA pathogenesis and hold promising as druggable targets for developing OA therapies.

Chapter 7: Novel genes for poor outcome of the total joint replacement surgery in the osteoarthritis patients identified by a genome-wide association study

Salem Werdyani[1], Ming Liu[1], Andrew Furey[2], Zhiwei Gao[3], Proton Rahman[4], Guangju Zhai[1]

[1]Division of Biomedical Sciences (Genetics),
[2]Discipline of Surgery,
[3]Division of Community Health and Humanities,
[4]Discipline of Medicine,
Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada.

An abstract of this chapter is published in the osteoarthritis and cartilage 2022; 104:12.

7.1 Co-authorship statement:

GZ and SW were responsible for the study conception and design; statistical expertise; analysis, and interpretation of the results. AF, ML, PR, and GZ participated in specimens; data collecting; and assembly. AF, PR, and GZ were responsible for the provision of study materials or patients. GZ obtained the study funding. ML, PR, GZ provided the administrative, technical, and logistic support.

7.2 Abstract

Purpose: TJR is considered by far as the most effective treatment for end-stage OA patients. Majority of patients achieve symptomatic improvement following the TJR. However, about 22% either do not improve or deteriorate after the surgery. The aim of this study was to identify the genetic variants for the poor outcome of the TJR in primary OA patients by a GWAS analyses.

Methods: Study participants were total knee or hip replacement patients due to primary OA who were recruited to the NFOAS before 2017. The WOMAC-OA index was used to assess pain and functional impairment pre- and up until 3.99±1.38 years post-surgery. Participants with a change score less than 7/20 points for pain were considered as pain non-responders; and those with less than 22/68 points for function were classified as function non-responders based on the absolute WOMAC score change. Also, the WOMAC point score was transformed into 0-100 interval scale. Then, the Outcome Measures in Arthritis Clinical Trials and the Osteoarthritis Research Society International

(OMERACT-OARSI) criteria were implemented, and participants with an improvement <50% and absolute score change <20 for pain or function were considered as pain or function non-responders, respectively. Blood DNA samples were genotyped using the Illumina HumanOmni2.58 and Infinium Global Diversity 8 v1.0 genotyping microarrays. Then, pre-association QC filtering and population structure analyses were performed for the study participants based on the genotyping data, genome-wide imputation analysis was conducted, and the GWAS analysis was performed using the MLR implemented in the MilorGWAS package in R version 4.1.2 to investigate the association between the whole-genome genetic variants and non-responders to the TJR pain, function, both (pain and function) and either (pain or function) in the OA patients from the NFOAS.

Results: 441 primary OA patients with WOMAC data available were included in the current study. These OA patients included 315 (71.43%) knee and 126 (28.57%) hip OA patients. There was no significant difference (all $p \geq 0.35$) in the estimated prevalence of the TJR non-responders between the absolute WOMAC change score and the OMERACT-OARSI classification criteria. Our analysis identified four chromosomal regions on chr7, 9, 18, and 21 to be significantly associated with non-responders to the TJR. We found that rs563726736 in the *DGKB* gene and rs62456377 in the *VWC2* gene on chr7; rs2441639 in the *PGM5P2* gene, rs149992177 and rs9408586 adjacent to the *FOXD4L6* gene on chr9; rs4797006 and rs146474469 in the *MC5R* gene on chr18; in addition to rs376101925 that is close to the *TEKT4P2* Pseudogene and rs1907506 that is adjacent to the *TPTE* gene on chr21 were significantly associated with the TJR pain non-responders (all $p \leq 2.79*10^{-08}$). The GWAS analysis of the 315 knee OA patients revealed

that only rs4797006 and rs146474469 on chr18 remained significantly associated with the non-responding to TJR pain (OR= 0.15, $p=2.79*10^{-11}$), while none of the identified variants were associated with non-responding to TJR pain in the 126 hip OA patients. Furthermore, our association analysis identified two loci on chr21 to be significantly associated with non-responding to TJR function in all 441 OA samples. We found that rs373135624 in the *TEKT4P2* Pseudogene, rs28971219 and rs113445703 that are adjacent to the *TPTE* gene on chr21 were significantly associated with the non-responding to TJR function (all $p \leq 1.87*10^{-08}$). Following to the association analysis in the 315 knee OA patients, only rs28971219 on chr21 remained to be significantly associated with non-responding to TJR function (OR= 0.24, $p=1.78*10^{-8}$). Whereas none of the identified variants were associated with non-responding to TJR pain in the 126 hip OA patients.

Conclusion: Our results suggested that genes implicated in the pro-inflammatory response, prolonged OA pain, and pathogenesis of the Fibroblast-like synoviocytes (FLS) to be associated with the TJR poor outcome and could be novel targets for developing strategies to improve the outcome of the TJR.

7.3 Introduction:

Despite the high prevalence and societal burden of OA, there is no cure for it yet. When non-surgical treatments, such as medication and physiotherapy do not relieve pain and restore joint function of the OA joint, doctors recommend TJR surgery (367,818). TJR is

considered by far the most effective therapy and a commonly performed surgical intervention for the end-stage OA patients (819). More than 55,000 hip and 67,000 knee replacements were performed in Canada between 2015-2016, and these numbers increase annually by more than 15% (820,821). Most of the patients achieve symptomatic improvement in pain reduction, joint function improvement, and better health related quality of life following the TJR. However, a significant proportion of the patients either do not achieve improvement or deteriorate after surgery even within groups implanted with the same prosthesis in the same institution (822,823). Latest studies showed that up to 23% of THR patients and 34% of TKR patients reported an unfavorable long-term post-operative pain outcome (824). Therefore, it has been important to identify factors that are associated with poor outcome of these surgical procedures and develop strategies to improve post-operative therapies and interventions.

A number of potential non-genetic predictors for the TJR outcome have been investigated, which include (pre-operative pain and function score, joint replaced, implant type, anesthesia used, time in surgery, length of hospital stay, and experience of surgeons), demographics (age, sex, BMI, ethnicity), anthropological factors (socioeconomic status, marital status, level of education), comorbidities, medical history (radiographic OA grade, mental health, history of joint injury or surgery), and patient expectations. However, the results were either inconclusive or had a very limited predictive power (822,825,826). Also, the clinical risk prediction tool created by *Judge at al.* (2012) for the outcome of THR with predictors including pre-operative function, age, sex, BMI, previous hip injury, and radiographic OA severity can only predict up to 29.9%

of the THR improved patients (823). The discriminatory ability of a predictive model for non-satisfaction following TKR created with predictors including pre-operative pain and function, current smoking, treatment for anxiety and not having been treated with injected corticosteroids was very moderate as indicated by the corrected area AUC of 0.65 (equal or less than 0.5 indicates no discriminatory ability) and has a poor generalizability to other populations (827). Thus, there is an urgent need to identify novel predictors for the outcome of the TJR that improve our understanding of the potential mechanisms leading to a poor outcome in non-responders to the TJR.

Our initial analysis in 830 primary OA patients from the NFOAS showed that ~22% of patients did not reach the MCID in term of pain reduction or function improvement. Thus, these patients were classified as non-responders to the TJR operations. A previous study performed by Costello *et al* (2020) to investigate the association between the epidemiological factors and non-responders to the TJR from the NFOAS identified a significant association between the clinical depression, multisite musculoskeletal pain, and younger age with pain, function, and/or pain and function non-responders (828). For female-related variables, the study found age at menopause to be significantly associated with pain, function, and both (pain and function) non-responders, while it did not detect association between non-responders and BMI, comorbidities, and type of prosthesis (828). Therefore, we hypothesized that genetic factors may contribute to the TJR poor outcome and can be used to predict the outcome in non-responders to the TJR.

The GWAS remains one of the most suitable methods of identifying key genes and variants that are associated with many common human diseases (681,829). Thus, we conducted a GWAS analysis in a well-established NFOAS to identify the genetic variants for poor outcome of the TJR in primary OA patients.

7.4 Materials and methods:

7.4.1 Study participants:

This study was performed as part of the NFOAS, which was initiated in 2011 and aimed to identify novel genetic, epigenetic, and biochemical biomarkers for the OA (441). Study participants were THR or TKR patients due to primary OA who were recruited to the NFOAS (450) between November 2011 and September 2017 in St Clare's Mercy Hospital and Health Science Centre General Hospital in St John's, NL, Canada (441). Knee and hip OA diagnosis was made based on the ACR-OA clinical diagnostic criteria (71). Pathology reports on cartilage and osteophytic irregularities were investigated following the surgery to confirm the OA diagnose. Ethical approval for the study was obtained from the Health Research Ethics Authority of NL (reference number 11.311, **Appendix C**), and a written consent was received from all study participants, **Appendix D**.

7.4.2 Assessment of total joint replacement outcome:

The outcome of the TJR is assessed by the Patient-reported outcomes (PROs), which are accepted endpoints in the evaluation of patients' treatment. One of the common PROs is

the WOMAC score that is a self-reported standard questionnaire used to assess the TJR pain and function, **Appendix A**. The study participants completed the WOMAC Likert version 3.0 to assess their TJR outcome pre- and up to 3.99±1.38 years post-surgery by 2017 (575). The WOMAC Likert version 3.0 estimated the pain during walking, using stairs, in bed, sitting or lying, and standing. As well as it evaluated the physical function during stair use, rising from sitting, standing, bending, walking, getting in/out of a car, shopping, putting on/taking off socks, rising from bed, lying in bed, getting in/out of bath, sitting, getting on/off toilet, heavy and light household duties. Score of four points was given to each of these items. Where pain had 0-20 points, and function had 0-68 points, with 0 represents no pain or functional difficulties (64). Un-answered WOMAC questions were imputed with the average responses of that question across all patients. Patients with more than four missing questions in the pain section and more than eleven missing questions in the function section were considered missing for analysis on the respective subsection (828). Then, the MCID of patient-reported WOMAC was used to identify the smallest amount of patient's response to the surgery (576).

Classification of OA patients based on their response to the TJR was performed using two previously reported non-responder classification criteria (16,25–27) to control for variation in classification methods and allow for comparison of significant genetic variations under both sets of criteria. The first criteria used the MCID of the absolute WOMAC change score over the follow-up period. Participants with absolute WOMAC change score less than 7 of 20 points for pain relief were considered as pain non-

responders; and those patients with the WOMAC change score less than 22 of 68 points for function improvement were classified as function non-responders (830).

The second classification method used in our analysis was the Outcome Measures in Arthritis Clinical Trials and the OARSI (OMERACT-OARSI) criteria which used the transformed WOMAC score on a scale of 0-100 to estimate the high improvement (16,27). Transformed-pain score was calculated as (pain*100/20) points and transformed-function score was identified as (function*100/68) points. Study participants with pain or function transformed score ≤ 20 points and ≤ 50% improvement at the follow-up point compared with the baseline scores were considered as non-responder to TJR (831).

Based on the TJR non-responding categories resulted from each of the classification criteria used in this analysis, if the OA patients were considered as non-responders to TJR pain relief and non-responders to TJR function improvement, they were classified as non-responders to both TJR pain and function. While the study participants who were categorized as non-responders to TJR pain and responders to TJR function, or the participants who were categorized as responders to TJR pain and non-responders TJR function were considered as non-responders to either TJR pain or function.

7.4.3 Demographic information:

Self-administered questionnaire was used to collect the patients' demographic, occupation, medical history, and family history information including age, sex, height, and weight at surgery, **Appendix E**. Patient's age was calculated by subtracting the birth

date from the TJR surgery date, then divided by 365.25 to get the number of years. The BMI was calculated by dividing patient's weight in kilograms by the squared height in meters (441,513).

7.4.4 Genome-wide genotyping:

The genome-wide genotyping analysis was previously described in detail in section 5.4.4 of the materials and methods of chapter 5 of this thesis.

7.4.5 Pre-association QC and imputation analysis of the genotyping data:

Since pre-association QC filtering is one of the necessary steps in the GWAS analysis to eliminate low genotyping quality of the DNA samples and genetic variants, the QC check was performed separately for the Illumina HumanOmni2.58 and the Infinium GDA 8 v1.0 SNP genotyping datasets as described in detail in section 5.4.5 of the materials and methods in chapter 5 (583). Then, the genetic variations IDs were updated in both SNP genotyping datasets using PLINK (V2.0) (586) and were used in the genome-wide imputation analysis.

For this study, imputation analysis using ShapeIT4 (589) and Impute5 (590) programs was performed on the Compute Canada computing cluster and the Wellcome Trust Sanger Imputation Service for both SNP genotyping datasets as explained in chapter 5, section 5.4.6 (588).

Subsequently, both OA Illumina HumanOmni2.58 and the Infinium GDA 8 v1.0 SNP genotyping datasets were merged into one VCF file using BCFtools (591) and converted to PLINK binary BED file format with PLINK(V2.0) (586). Then, the post imputation QC filtering was performed with PLINK(V1.7) (584) to exclude duplicated and non-OA samples, as well as eliminate rare SNPs with MAF<0.01 that are deviated from the HWE ($p < 0.0001$).

7.4.6 Genome-wide association analysis:

The genome-wide association analysis was performed to investigate the association between the whole-genome genetic variants and non-responders to TJR pain, function, both (pain and function), and either (pain or function) in all OA, knee OA, and hip OA patients. This GWAS analysis was conducted using the mixed effects logistic regression (MLR) of the genome-wide association analysis implemented in the MilorGWAS package in R version 4.1.2 and included the top 10 PCs to completely correct for population structure (601) similar to the genome-wide association analysis conducted in section 5.4.7 of the materials and methods in chapter 5. Finally, the genome-wide significance level was defined at ($p < 5*10^{-8}$) to control for multiple testing.

7.5 Results:

The pre-association QC check of the Illumina HumanOmni2.58 SNP genotyping raw data led to the exclusion of 865,471 variants with MAF<1%; three individuals with discordance sex information; a total of 39,242 non-autosome variants; two participants

and 9,717 variants with genotyping call rate <95%; a number of 192 variants deviated from the HWE with $p < 0.0001$; and four non-Caucasian study participants were identified by the PCA as outliers. Therefore, the clean Illumina HumanOmni2.58 dataset contained 95 subjects and 1,466,378 autosomal genetic variants subsequence to the QC filtering that were used in the downstream imputation analyses.

Similarly, a total of 947,440 variants with MAF<1%; one study participant with disagreement of the sex information; 35,324 non-autosome variants; one sample and 18,787 variants with genotyping call rate <95%; a number of 61 variants out of HWE with $p < 0.0001$; two duplicated samples with IBD>0.5; and two non-Caucasian subjects were identified by the PCA as outliers were excluded from the Illumina Infinium GDA 8 v1.0 dataset following to the QC filtering. Thus, the remaining 482 OA patients and 823,710 genetic variants were used for the imputation analyses.

In-house genotype imputation analysis pipeline detected 81,181,531 and 81,126,288 autosomal variants in the Illumina HumanOmni2.58 and Illumina Infinium GDA 8 v1.0 datasets, respectively. Our imputation results came in concordance with the imputation findings from the Wellcome Trust Sanger Imputation Service. Subsequence to merging the imputed genetic variations, a total of 9,617,754 autosomal genetic variants from 577 samples passed the post-imputation QC check, 441 of them were primary OA patients with WOMAC data available and were included in the current study. Theses OA patients included 315 (71.43%) knee and 126 (28.57%) hip OA patients, **Table 7.1.**

Table 7.1: The characteristics of the 315 knee OA patients and 126 hip OA patient of this study.

	All OA patients (n=441)	Knee OA patients (n=315)	Hip OA patients (n=126)	Difference between Knee and Hip OA patients P-value
Sex: Female N (%)	251 (56.9)	186 (59.05)	65 (51.59)	0.19
Age (mean±SD) years	65.67±7.7	65.49±7.38	66.11±8.43	0.44
BMI (mean±SD) kg/m^2	33.93±7.03	35.09±6.93	31.01±6.42	0.0001

Values are mean ± SD for continuous variable and percentage for sex. P-values were obtained from Chi squared test for sex distribution and Student's t-test for age and BMI continuous variables.

Classification of the OA patients based on the absolute WOMAC change score identified 383 responders and 53 non-responders to the TJR pain; 369 responders and 68 non-responders to the TJR function; 356 responders and 42 non-responders to the TJR both (pain and function); and 357 responders and 79 non-responders to the TJR either (pain or function), **Table 7.2**. While the OMERACT-OARSI classification criteria detected 378 responders and 52 non-responders to the TJR pain, 364 responders and 69 non-responders to the TJR function, 349 responders and 42 non-responders to the TJR both (pain and function), and 349 responders and 79 non-responders to the TJR either (pain or function), **Table 7.2**. Moreover, **Table 7.2** illustrates the joint specific (knee and hips) classification of OA non-responders based on the absolute WOMAC change score and OMERACT-OARSI classification criteria.

Our analyses did not identify statistically significant differences (all $p \geq 0.35$) of the TJR responders and non-responders between the absolute WOMAC change score and OMERACT-OARSI classification criteria, and we estimated the prevalence of the OA non-responders to be about 20% of the TJR patients in the Newfoundland population, **Table 7.2**.

Moreover, we did not find significant difference (all $p \geq 0.06$) between the TJR responders and non-responders within each classification method, and there was no significant difference in the TJR non-responders between the absolute WOMAC change score and OMERACT-OARSI classification criteria except for BMI of THR pain ($p=0.029$) and either (pain or function, $p=0.001$) non-responders, **Table 7.3**.

Table 7.2: Classification of the OA patients of this study to responders and non-responders to the TJR surgery based on the absolute WOMAC change score and the OMERACT-OARSI classification criteria.

		All OA (n=441)			Knee OA (n=315)			Hip OA (n=126)		
		Absolute WOMAC Score	OMERACT-OARSI	**P-value**	Absolute WOMAC Score	OMERACT-OARSI	**P-value**	Absolute WOMAC Score	OMERACT-OARSI	**P-value**
Pain	Responders # (%)	383 (86.85)	378 (85.71)	0.53	272 (86.35)	269 (85.40)	0.53	111 (88.10)	109 (86.51)	0.52
	Non-responders # (%)	53 (12.02)	52 (11.79)		42 (13.33)	42 (13.33)		11 (8.73)	10 (7.94)	
Function	Responders # (%)	369 (83.67)	364 (82.54)	0.48	261 (82.86)	254 (80.63)	0.35	108 (85.71)	110 (87.30)	0.35
	Non-responders # (%)	68 (15.42)	69 (15.65)		52 (16.51)	56 (17.77)		16 (12.70)	13 (10.32)	
Both	Responders # (%)	356 (80.73)	349 (79.14)	0.51	252 (80)	245(77.78)	0.46	104 (82.54)	104 (82.54)	0.51
	Non-responders # (%)	42 (9.52)	42 (9.52)		33 (10.48)	34 (10.79)		9 (7.14)	8 (6.35)	
Either	Responders # (%)	357 (80.95)	349 (79.14)	0.48	253 (80.32)	245 (77.78)	0.38	104 (82.54)	104 (82.54)	0.38
	Non-responders # (%)	79 (17.91)	79 (17.91)		61 (19.37)	64 (20.32)		18 (14.29)	15 (11.90)	

Difference between percentages of responders and non-responders to TJR was calculated using the Chi-square test.

Table 7.3: Differences between responders and non-responders of the TJR patients in each classification method, and the differences of non-responders to the TJR between the absolute WOMAC change score and the OMERACT-OARSI classification criteria.

		The absolute WOMAC change score			**The OMERACT-OARSI classification criteria**			difference between TJR non-responders in both classification methods
		Responders	Non-responders	**P-value**	Responders	Non-responders	**P-value**	**P-value**
TJR pain	Sex (female #, %)	218 (56.92)	33 (62.26)	0.461	217 (57.41)	31 (59.62)	0.763	0.83
	Age (mean±SD)	64.09±7.78	64.03±6.83	0.109	65.72±7.78	64.08±6.73	0.148	0.969
	BMI (mean±SD)	33.93±7.16	34.15±6.07	0.829	34.08±7.16	33.85±5.88	0.823	0.798
TJR function	Sex (female #, %)	207 (56.10)	43 (63.24)	0.275	204 (56.04)	44 (63.77)	0.236	0.87
	Age (mean±SD)	65.90±7.75	64.29±7.29	0.116	65.90±7.75	64.29±7.29	0.079	1
	BMI (mean±SD)	33.90±7.13	34.29±6.45	0.673	33.90±7.13	34.29±6.45	0.257	1
TJR both (pain and function)	Sex (female #, %)	202 (56.74)	28 (66.67)	0.221	199 (57.02)	27 (64.29)	0.369	0.94
	Age (mean±SD)	65.95±7.74	64.48±6.85	0.242	65.81±7.76	63.90±6.42	0.128	0.689
	BMI (mean±SD)	33.89±7.12	33.98±5.56	0.931	34.00±7.17	34.44±5.84	0.700	0.713
TJR either (pain or function)	Sex (female #, %)	203 (56.86)	48 (60.76)	0.526	199 (57.02)	48 (60.76)	0.544	0.93
	Age (mean±SD)	65.98±7.75	64.02±7.22	0.041	65.81±7.76	64.18±7.16	0.09	0.889
	BMI (mean±SD)	33.87±7.12	34.36±6.64	0.575	34.00±7.17	34.44±6.30	0.612	0.936

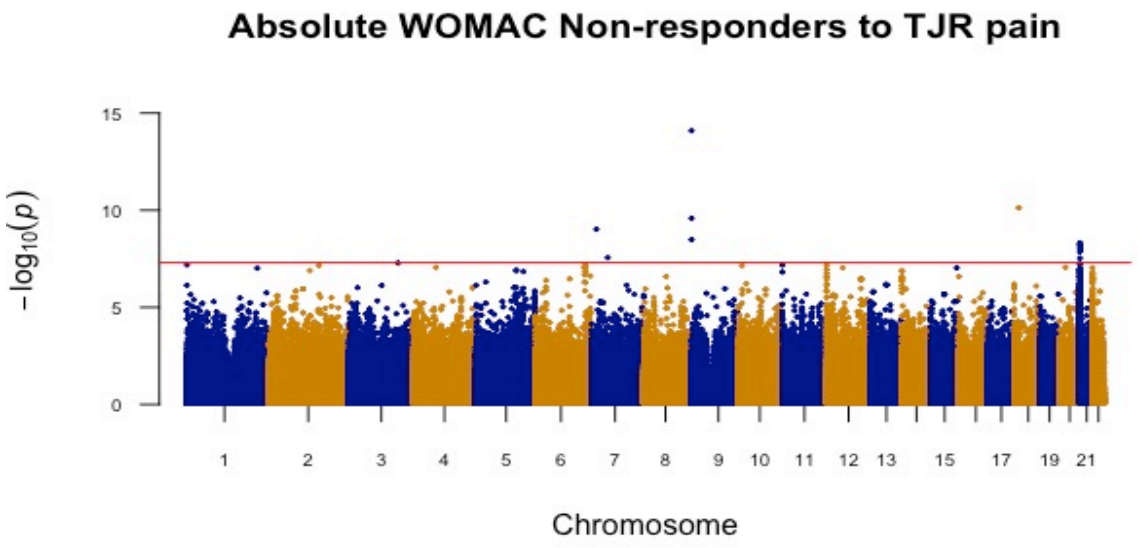
		The absolute WOMAC change score			**The OMERACT-OARSI classification criteria**			difference between TJR non-responders in both classification methods
		Responders	Non-responders	**P-value**	Responders	Non-responders	**P-value**	**P-value**
TKR pain	Sex (female #, %)	161 (59.19)	25 (59.52)	0.967	161 (59.85)	24 (57.14)	0.739	0.89
	Age (mean±SD)	65.75±7.46	63.96±6.67	0.144	65.71±7.50	63.59±6.10	0.084	0.7915
	BMI (mean±SD)	35.01±7.09	35.50±5.84	0.666	35.05±7.12	35.05±6.62	0.938	0.742
TKR function	Sex (female #, %)	152 (58.24)	33 (63.26)	0.485	149 (58.66)	35 (62.50)	0.597	0.77
	Age (mean±SD)	65.71±7.41	64.29±7.05	0.204	65.74±7.49	63.99±6.57	0.109	0.8194
	BMI (mean±SD)	35.03±7.07	35.43±6.22	0.703	35.02±7.12	35.74±6.03	0.484	0.7931
TKR both (pain and function)	Sex (female #, %)	149 (59.13)	22 (66.67)	0.407	146 (59.59)	21 (61.76)	0.81	0.94
	Age (mean±SD)	65.95±7.74	64.48±6.85	0.234	65.71±7.45	63.08±5.28	0.06	0.3514
	BMI (mean±SD)	33.89±7.12	33.98±5.56	0.860	35.13±7.15	35.68±5.48	0.668	0.212
TKR either (pain or function)	Sex (female #, %)	150 (59.29)	36 (59.02)	0.969	146 (59.59)	38 (59.37)	0.975	0.76
	Age (mean±SD)	65.85±7.42	64.11±7.05	0.102	65.71±7.45	64.21±6.86	0.148	0.9361
	BMI (mean±SD)	34.95±7.05	35.61±6.40	0.503	35.13±7.15	35.32±6.06	0.850	0.7951

		The absolute WOMAC change score			**The OMERACT-OARSI classification criteria**			difference between TJR non-responders in both classification methods
		Responders	Non-responders	**P-value**	Responders	Non-responders	**P-value**	**P-value**
THR pain	Sex (female #, %)	57 (51.35)	8 (72.73)	0.188	56 (51.38)	7 (70.00)	0.269	0.83
	Age (mean±SD)	66.07±8.50	64.32±7.42	0.512	65.75±8.42	66.12±8.58	0.895	0.612
	BMI (mean±SD)	28.99±3.69	34.15±6.07	0.26	31.48±6.56	28.81±3.99	0.208	**0.029**
THR function	Sex (female #, %)	55 (50.93)	10 (62.50)	0.389	55 (50.00)	9 (69.23)	0.198	0.83
	Age (mean±SD)	66.34±8.51	64.32±8.01	0.376	66.14±8.36	64.45±8.79	0.507	0.967
	BMI (mean±SD)	31.15±6.50	30.57±5.75	0.734	31.12±6.45	31.12±6.10	0.948	0.805
THR both (pain and function)	Sex (female #, %)	53 (50.96)	6 (66.67)	0.372	53 (50.96)	6 (75.00)	0.207	1
	Age (mean±SD)	66.30±8.48	65.58±7.65	0.806	66.04±8.43	67.38±9.12	0.666	0.664
	BMI (mean±SD)	31.24±6.58	29.53±3.76	0.441	31.33±6.49	29.19±4.15	0.358	0.807
THR either (pain or function)	Sex (female #, %)	53 (50.96)	12 (66.67)	0.223	53 (50.96)	10 (66.67)	0.26	0.69
	Age (mean±SD)	66.30±8.48	63.69±7.77	0.228	66.04±8.43	64.04±8.29	0.394	0.901
	BMI (mean±SD)	31.25±6.59	30.13±5.60	0.498	31.33±6.49	37.25±5.92	0.73	**0.001**

Values are numbers percentage for sex and mean±SD for age and BMI variables. P-values were obtained from Chi squared test for sex distribution and Student's t-test for continuous variables. TJR: total joint replacement, TKR: total knee replacement, THR: total hip replacement, BMI: body mass index.

Remarkably, the association analysis using the MLR in the R - MilorGWAS package between the genetic variations and the TJR poor outcome based on both TJR non-responders' classification methods identified similar associated variants. Our GWAS analysis in the entire 441 OA patient's cohort based on both classification criteria identified four chromosomal regions on chr7, 9, 18, and 21 to be significantly associated with non-responders to the TJR pain at the GWAS significance level as shown in the Manhattan plot in **Figure 7.1**. There was no evidence of population stratification in the association analysis between the genetic variations and non-responding to the TJR pain relief throughout the 441 OA patients based on the Absolute WOMAC Score classification criteria with inflation factor (λ)=0.97, and the OMERACT-OARSI classification method with inflation factor (λ)=0.98. Moreover, the Q-Q plot of the association test P-values did not visually show obvious deviation from what was expected under the assumption of no genetic association in both classification criteria, **Figure 7.2**.

A



B

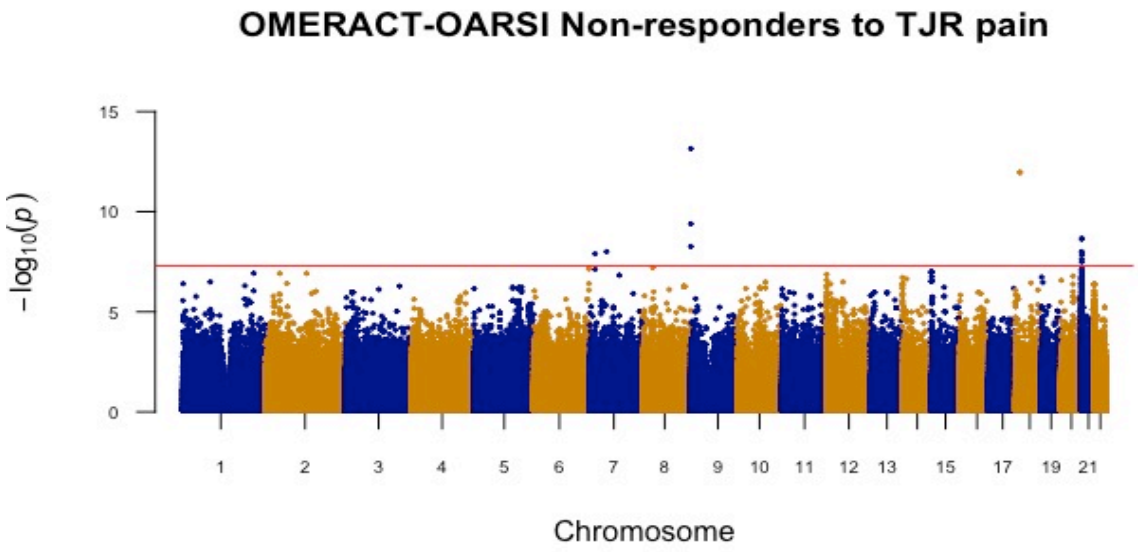
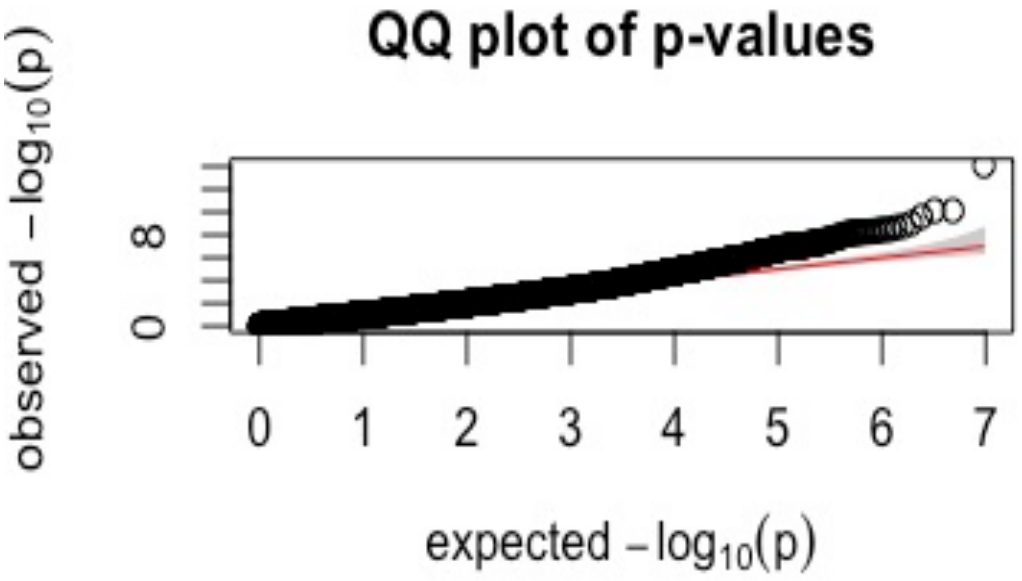


Figure 7.1: A Manhattan plot of the GWAS results showing the significantly associated genetic variations with non-responders to the TJR pain in 441 OA patients from the NFOAS based on **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods. The red line indicates the GWAS significance at $\alpha = 5*10^{-8}$.

A



B

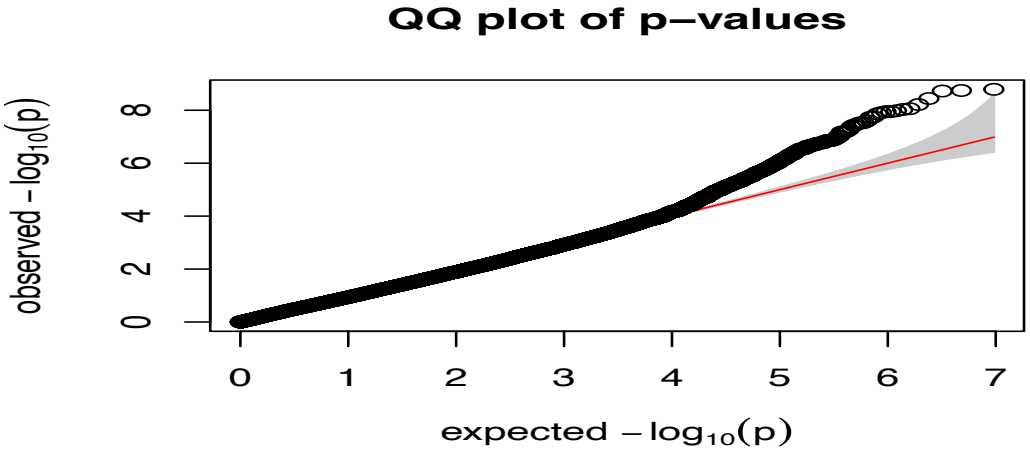


Figure 7.2: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the non-responders to the TJR pain in the 441 OA patients based on **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods.

Two regions on chr7 were detected to be significantly associated with the TJR pain non-responders in the whole 441 OA cohort. The INDEL rs563726736 was significantly associated with the TJR pain non-responders with OR of 12.8 ($p=9.64*10^{-10}$) and OR of 10.3 ($p=1.26*10^{-08}$) based on the Absolute WOMAC Score and the OMERACT-OARSI classification methods (**Table 7.4**), respectively. This INDEL is an intronic variant in intron one of the *diacylglycerol kinase betas* (*DGKB*) gene on chr7 as shown in the regional association plot in **Figure 7.3A**. The second variant that was detected to be significantly associated with the TJR poor pain outcome was rs62456377 with OR of 6.71 ($p=2.79*10^{-08}$) based on the Absolute WOMAC Score criteria and OR of 6.76 ($p=9.95*10^{-09}$) based on the OMERACT-OARSI classification criteria, **Table 7.4**. This SNP is intronic variant that falls in the intron two of the *von Willebrand factor C domain containing 2* (*VWC2*) gene on chr7, **Figure 7.3B**.

Moreover, we found that A allele of the rs2441639 SNP was significantly associated with the TJR pain non-responders with OR of 9.10 ($p=8.13*10^{-15}$) from the Absolute WOMAC Score classification method, and OR of 7.53 ($p=7.07*10^{-14}$) based on the OMERACT-OARSI classification criteria, **Table 7.4**. This SNP is in intron two of the *Phosphoglucomutase 5 Pseudogene 2* (*PGM5P2*) on chr9, **Figure 7.3C**. The other significantly associated variants with the TJR pain poor outcome on chr9 are adjacent to the rs2441639 SNP and included the INDEL rs149992177 (OR=5.98, $p=3.28*10^{-09}$) from the Absolute WOMAC Score classification method, and (OR=5.59, $p=5.47*10^{-09}$) based on the OMERACT-OARSI classification criteria, and SNP rs9408586 (OR=7.89, $p=2.57*10^{-10}$) from the Absolute WOMAC Score, and (OR=7.18, $p=4.03*10^{-10}$) from the OMERACT-OARSI classification method, **Table 7.4**. These two variants

are adjacent to the *Forkhead Box Protein D4-Like 6* (*FOXD4L6*) gene on chr9 as shown in the regional association plot in **Figure 7.3C**.

Table 7.4: The most significantly associated genetic variants with the TJR pain non-responding in the 441 OA samples.

CHR	position	Variant ID	Allele 1	The Absolute WOMAC Score				The OMERACT-OARSI			
				MAF_Aff	MAF_Un-aff	OR	P-value	MAF_Aff	MAF_Un-aff	OR	P-value
7	14983262	rs563726736	ACACACAC ACACACAC ACACACAC	0.11	0.01	12.8	$9.64*10^{-10}$	0.11	0.01	10.29	$1.26*10^{-08}$
7	49914144	rs62456377	G	0.13	0.02	6.76	$2.79*10^{-08}$	0.13	0.02	6.71	$9.95*10^{-09}$
9	52415	rs2441639	A	0.25	0.03	9.10	$8.13*10^{-15}$	0.23	0.03	7.53	$7.07*10^{-14}$
9	52625	rs149992177	CA	0.18	0.02	5.98	$3.28*10^{-09}$	0.17	0.03	5.59	$5.47*10^{-09}$
9	55099	rs9408586	C	0.18	0.02	7.89	$2.57*10^{-10}$	0.17	0.03	7.18	$4.03*10^{-10}$
18	13824807	rs4797006	G	0.15	0.02	8.96	$7.43*10^{-11}$	0.16	0.02	11.25	$1.09*10^{-12}$
18	13825028	rs146474469	C	0.15	0.02	8.96	$7.43*10^{-11}$	0.16	0.02	11.25	$1.09*10^{-12}$
21	9827857	rs71236670	C	0.12	0.46	0.19	$6.14*10^{-09}$	0.12	0.46	0.17	$2.34*10^{-09}$
21	9955015	rs189820349	A	0.18	0.52	0.24	$5.40*10^{-09}$	0.19	0.52	0.26	$2.38*10^{-08}$
21	9972507	rs376101925	A	0.20	0.54	0.26	$4.75*10^{-09}$	0.20	0.54	0.27	$9.96*10^{-09}$
21	9972611	rs4361469	A	0.20	0.53	0.27	$1.37*10^{-08}$	0.20	0.53	0.29	$3.18*10^{-08}$
21	10476925	rs1907506	A	0.18	0.51	0.25	$1.07*10^{-08}$	0.16	0.51	0.22	$2.11*10^{-09}$
21	10775663	rs111909407	A	0.09	0.42	0.16	$8.28*10^{-09}$	0.11	0.42	0.19	$1.48*10^{-08}$
21	10839945	rs113445703	CGAATG	0.15	0.48	0.21	$6.55*10^{-09}$	0.15	0.48	0.22	$1.07*10^{-08}$

CHR: chromosome, Allele 1: the risk allele, MAF_Aff: minor allele frequency of the risk allele in the TJR pain non-responders, MAF_un-Aff: minor allele frequency of the risk allele in the TJR pain responders, Allele 2: Major allele, OR: odd ratio.

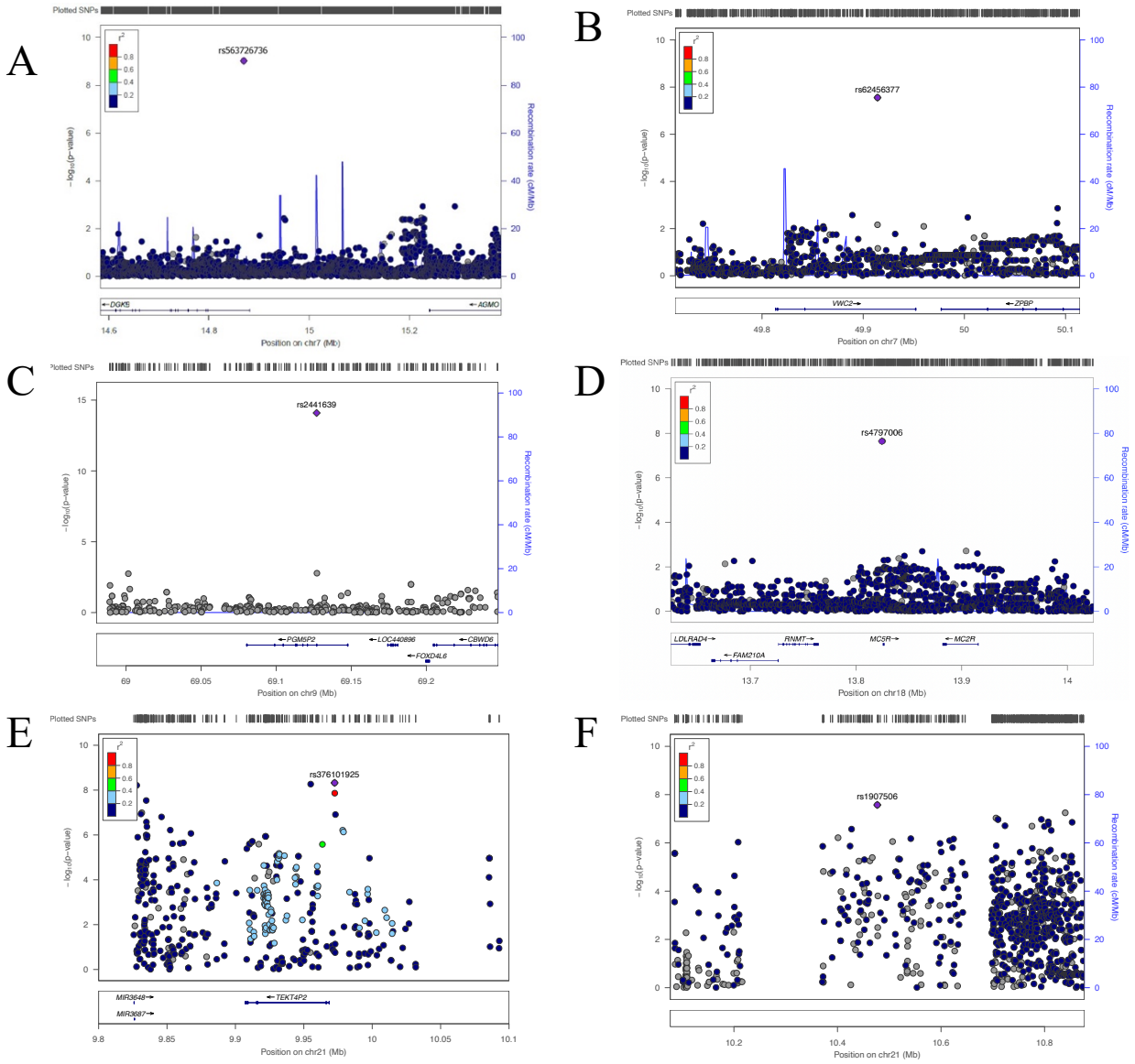


Figure 7.3: The regional association plot for the chromosomal region showing the most significantly associated genetic variants with the TJR pain non-responders: **A)** INDEL rs563726736 with $p \leq 1.26*10^{-08}$. This INDEL is an intronic variant in intron one of the *diacylglycerol kinase beta* (*DGKB*) gene on chr7; **B)** The second locus on chr7 included SNP rs62456377 with $p \leq 2.79*10^{-08}$. This SNP is intronic variant that falls in the intron two of the *von Willebrand factor C domain containing 2* (*VWC2*) gene on chr7; **C)** The top SNP on chr9 was rs2441639 with $p \leq 7.07*10^{-14}$. Variants on this locus are in intron two of the Phosphoglucomutase 5 Pseudogene 2 (PGM5P2) and adjacent to the *Forkhead Box Protein D4-Like 6* (*FOXD4L6*) gene on chr9; **D)** Variants rs4797006 and rs146474469 on chr18 ($p \leq 7.43*10^{-11}$). These variants are intronic variants in locate in intron one of the *melanocortin5 receptor* (*MC5R*) gene; **E)** The top SNP in the first associated chromosomal region on chr21 was rs376101925 with $p \leq 9.96*10^{-9}$. SNPs of this locus are adjacent to the Tektin 4 Pseudogene 2 (TEKT4P2) Pseudogene; **F)** the top variant of the second associated locus on chr21 was rs1907506 with $p \leq 1.07*10^{-08}$. Variants of this region fall ~ 400 KBPs upstream of the *Transmembrane Phosphatase With Tensin Homology* (*TPTE*) gene on chr21.

Furthermore, we found that SNP rs4797006 (T>G, MAF=0.16) and INDEL rs146474469 (CTTACTTAC>CTTAC, MAF=0.16) on chr18 were significantly associated with non-responders to the TJR pain (OR= 8.96, $p=7.43*10^{-11}$) based on the Absolute WOMAC Score, and (OR= 11.25, $p= 1.09 *10^{-12}$) based on the OMERACT-OARSI classification method in the 441 OA patients, **Table 7.4**. Pairwise linkage analysis revealed that these two variants are in LD, $r^2=1$ in intron one of the *melanocortin5 receptor* (*MC5R*) gene on chr18 as shown in the regional association plot in **Figure 7.3D**.

Additionally, two chromosomal regions on chr21 were found to be significantly associated with non-responders to the TJR pain. The top SNP on the first region was rs376101925 (OR= 0.26, $p=4.75*10^{-9}$) based on the Absolute WOMAC Score, and (OR= 0.27, $p= 9.96 *10^{-9}$) based on the OMERACT-OARSI classification method. This variant is adjacent to the *Tektin 4 Pseudogene 2* (*TEKT4P2*) Pseudogene, **Figure 7.3E**. Also, the rs1907506 was the top variant on the second associated locus on chr21 (OR=0.25, $p=1.07*10^{-8}$; and OR=0.22, $p=2.11*10^{-9}$ from the Absolute WOMAC Score and the OMERACT-OARSI classification methods, respectively). Although, this SNP is in gene desert, bioinformatics analysis showed that it is located ~ 400 KBPs upstream of the *Transmembrane Phosphatase With Tensin Homology* (*TPTE*) gene on chr21, **Figure 7.3F**.

The GWAS analysis of the 315 knee OA patients revealed that only variants rs4797006 and rs146474469 on chr18 remained significantly associated with the non-responding to the TJR pain (**Table 7.5)** based on the Absolute WOMAC Score (OR= 0.15, $p=1.77*10^{-11}$) and the OMERACT-OARSI classification methods (OR= 0.15, $p=2.79*10^{-11}$), **Figure 7.4**. While none

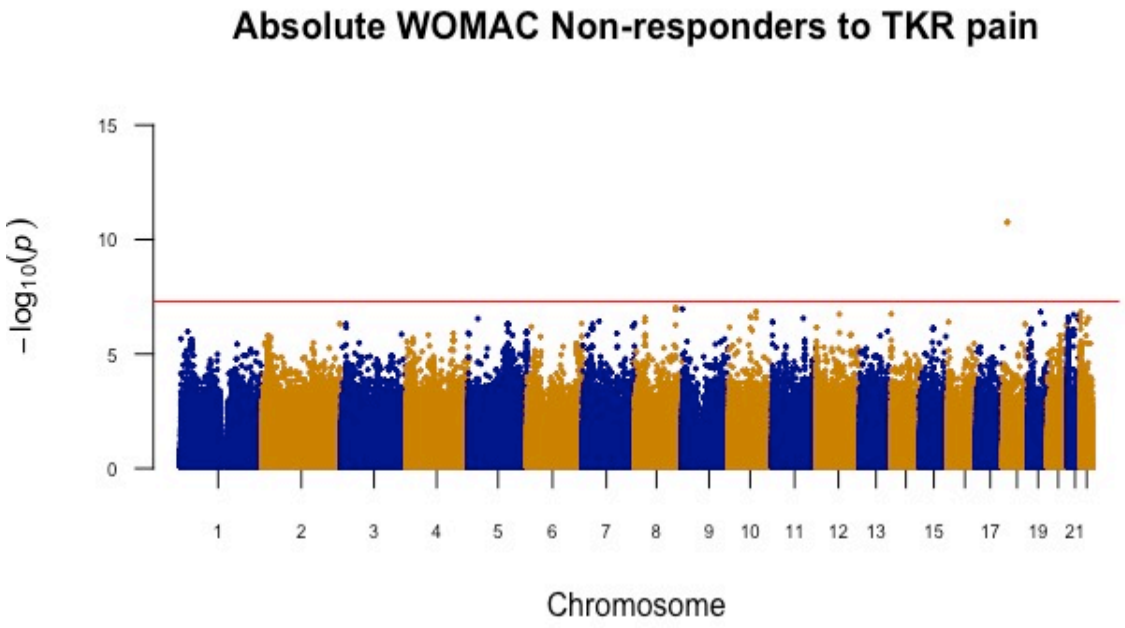
of the identified variants were associated with non-responding to the TJR pain in the 126 hip OA patients, **Figure 7.5**.

Table 7.5: The most significantly associated genetic variants with the TJR pain non-responding in the 315 knee OA samples.

CHR	position	Variant ID	Minor Allele	The Absolute WOMAC Score				The OMERACT-OARSI			
				MAF_Aff	MAF_Un-aff	OR	P-value	MAF_Aff	MAF_Un-aff	OR	P-value
18	13824807	rs4797006	G	0.15	0.012	14.42	$1.77*10^{-11}$	0.15	0.013	11.25	$2.79*10^{-11}$
18	13825028	rs146474469	C	0.15	0.012	14.42	$1.77*10^{-11}$	0.15	0.013	11.25	$2.79*10^{-11}$

CHR: chromosome, Allele 1: the risk allele, MAF_Aff: minor allele frequency of the risk allele in the TJR pain non-responders, MAF_un-Aff: minor allele frequency of the risk allele in the TJR pain responders, Allele 2: Major allele, OR: odd ratio.

A



B

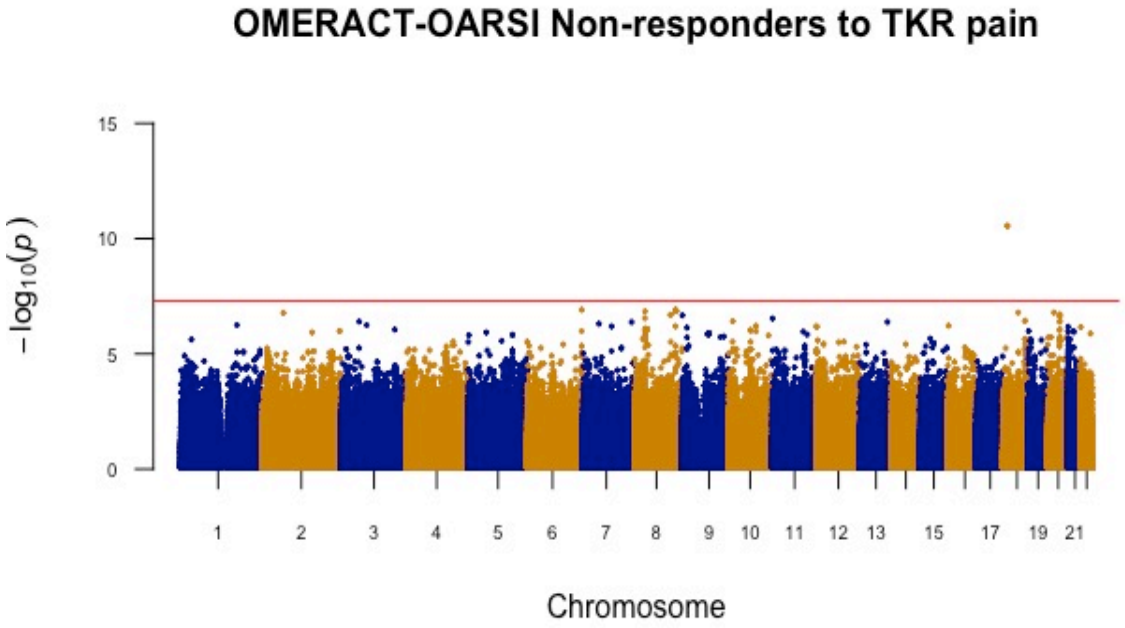
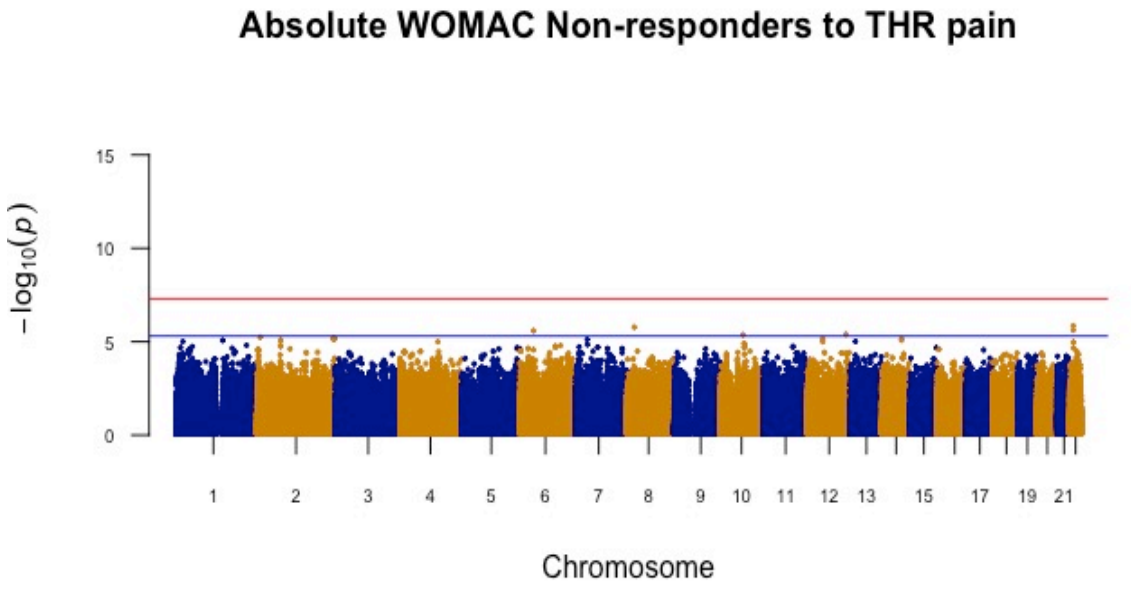


Figure 7.4: A Manhattan plot of the GWAS results showing the associated rs4797006 and rs146474469 on chr18 with the TJR pain non-responders in 315 knee OA patients based on **A)** the Absolute WOMAC Score (OR=14.42, p=1.77*10-11), **B)** the OMERACT-OARSI (OR=11.25, p=2.79*10^{-11}) classification methods.

A



B

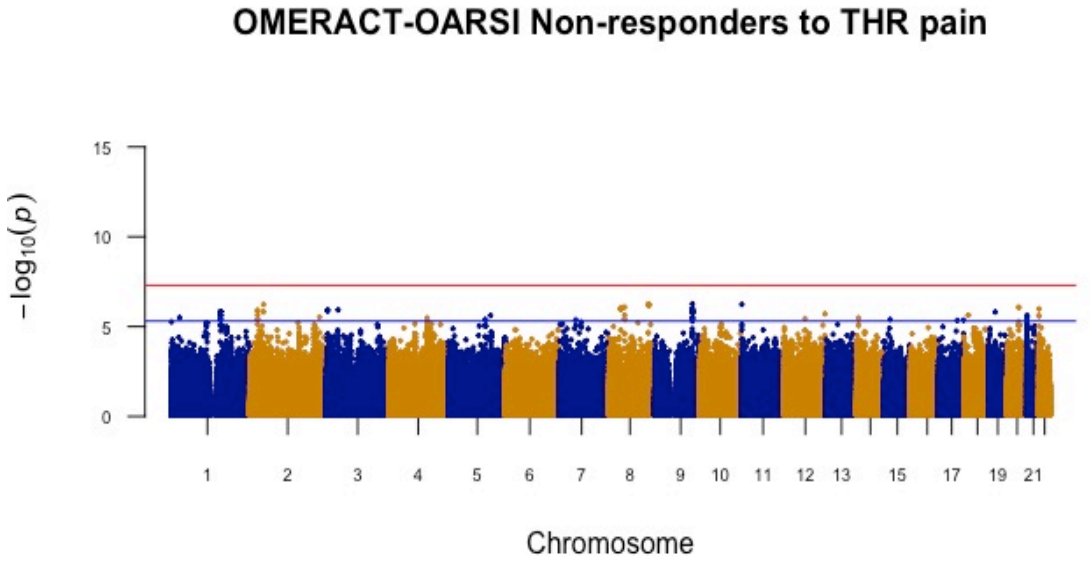
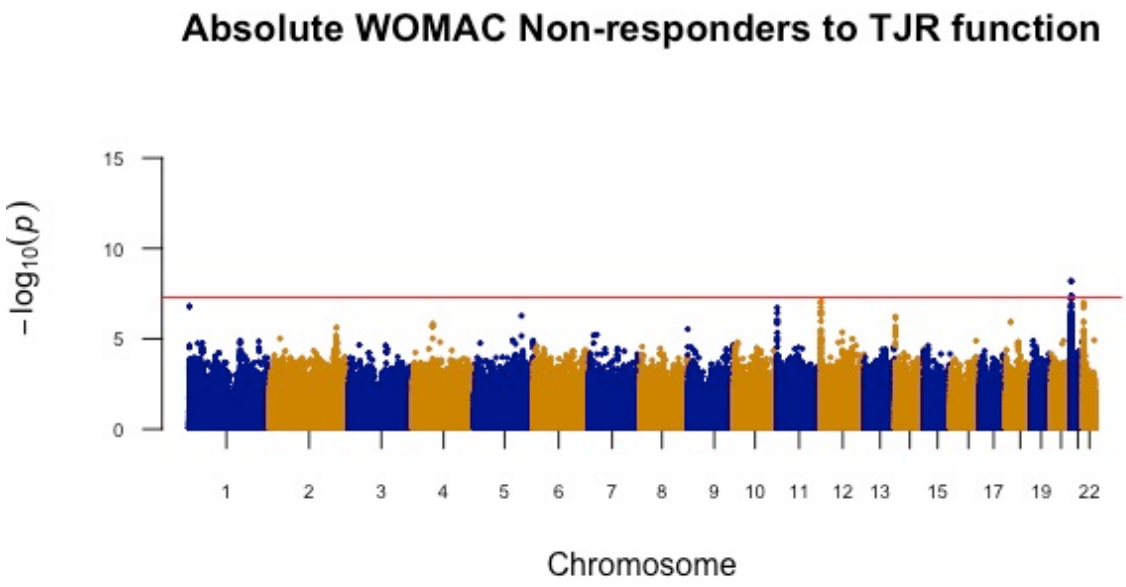


Figure 7.5: A Manhattan plot of the GWAS results showing no association of the genetic variations at $p<5*10^{-8}$ with the non-responders to TJR pain in 126 hip OA patients in our study based on: **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods. The red line indicates the GWAS significance at $\alpha = 5*10^{-8}$.

Furthermore, our association analysis identified two chromosomal regions on chr21 to be significantly associated with non-responding to the TJR function in all 441 OA samples based on the Absolute WOMAC Score (**Figure 7.6A**) and the OMERACT-OARSI (**Figure 7.6B**) classification methods. There was no evidence of population stratification in the Absolute WOMAC Score method with inflation factor (λ)=0.96 and the OMERACT-OARSI method with inflation factor (λ)=0.97 in the association analysis between the genetic variations and the TJR function non-responders in the 441 OA patients. Also, the Q-Q plot of the association test P-values did not visually show obvious deviation from what was expected under the assumption of no genetic association, **Figure 7.7**.

A



B

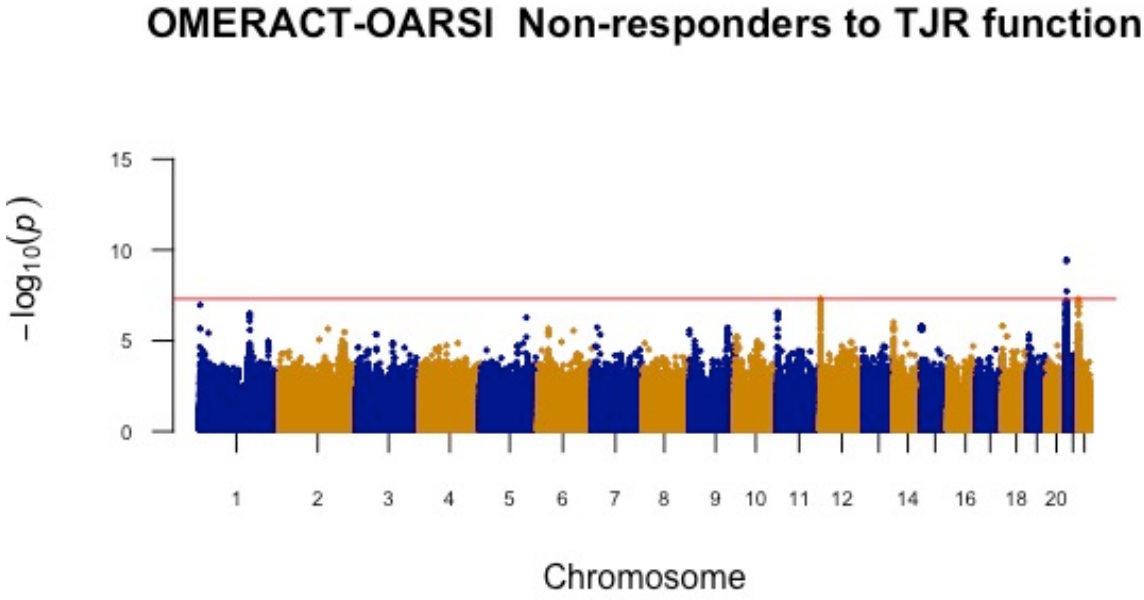
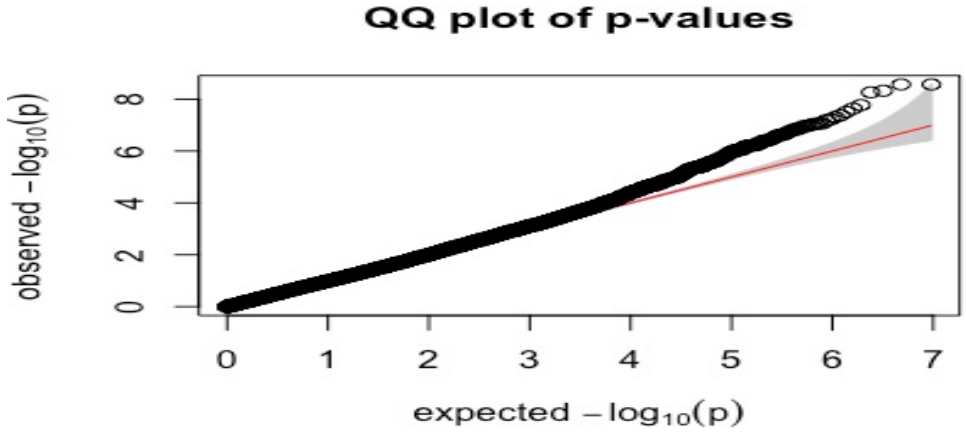


Figure 7.6: A Manhattan plot of the GWAS results showing the significantly associated genetic variations with non-responders to the TJR function in 441 OA patients from the NFOAS based on **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods. The red line indicates the GWAS significance at $\alpha = 5*10^{-8}$.

A



B

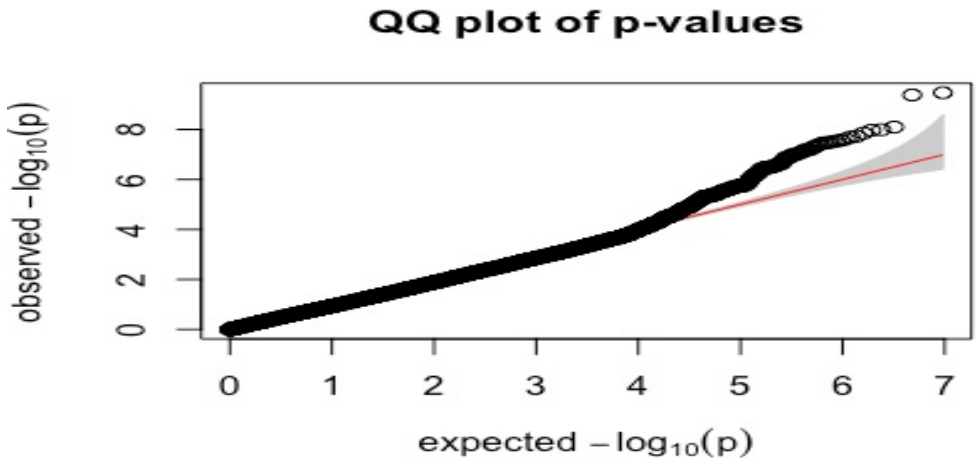


Figure 7.7: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the non-responders to the TJR function in the 441 OA patients based on **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods.

We found that allele G of SNP rs373135624 was significantly associated with non-responders to the TJR function based on the Absolute WOMAC Score (OR= 0.27, $p=5.97*10^{-9}$) and the OMERACT-OARSI (OR= 0.24, $p=4.15*10^{-10}$) classification methods as shown in the **Table 7.6.** This SNP is an intronic variant in the *TEKT4P2* Pseudogene on chr21 as shown in **Figure 7.8A.** Also, the intergenic SNP rs28971219 (C>A, MAF=0.20) on chr21 was significantly associated with the non-responders to the TJR function based on the Absolute WOMAC Score (OR= 0.30, $p=7.03*10^{-9}$), and the OMERACT-OARSI (OR= 0.26, $p=3.39*10^{-10}$) classification methods, **Table 7.6**. The adjacent genes to the rs28971219 include the *TPTE* gene, **Figure 7.8B**. Remarkably, this SNP is about 330 KBs downstream of the rs1907506 that was detected to be significantly associated with the non-responders to the TJR pain. However, LD analysis showed that these two SNPs are not linked ($r^2<0.2$). The third and last variant that was found to be significantly associated with the non-responding to the TJR function was the INDEL rs113445703 (**Table 7.6**) that is upstream of the rs28971219 and adjacent to the *TPTE* gene on chr21.

Table 7.6: The most significantly associated genetic variants with the TJR function non-responding in the 441 OA samples.

CHR	position	Variant ID	Allele 1	The Absolute WOMAC Score				The OMERACT-OARSI			
				MAF_Aff	MAF_Un-aff	OR	P-value	MAF_Aff	MAF_Un-aff	OR	P-value
21	9912348	rs373135624	G	0.16	0.46	0.27	$5.97*10^{-09}$	0.16	0.47	0.24	$4.15*10^{-10}$
21	10807264	rs28971219	A	0.22	0.51	0.30	$7.03*10^{-09}$	0.20	0.52	0.26	$3.39*10^{-10}$
21	10839945	rs113445703	CGAATG	0.21	0.48	0.30	$3.95*10^{-08}$	0.20	0.49	0.29	$1.87*10^{-08}$

CHR: chromosome, Allele 1: the risk allele, MAF_Aff: minor allele frequency of the risk allele in the TJR function non-responders, MAF_un-Aff: minor allele frequency of the risk allele in the TJR pain responders, Allele 2: Major allele, OR: odd ratio.

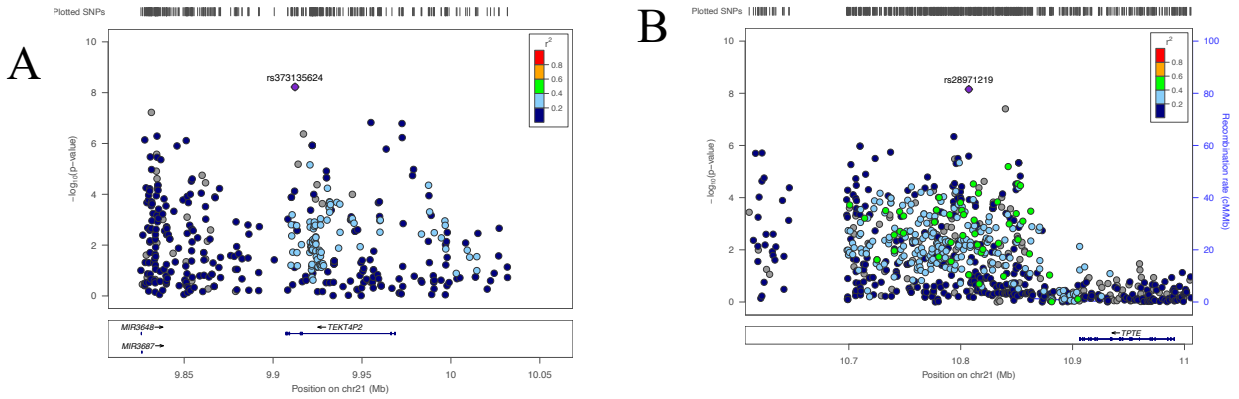
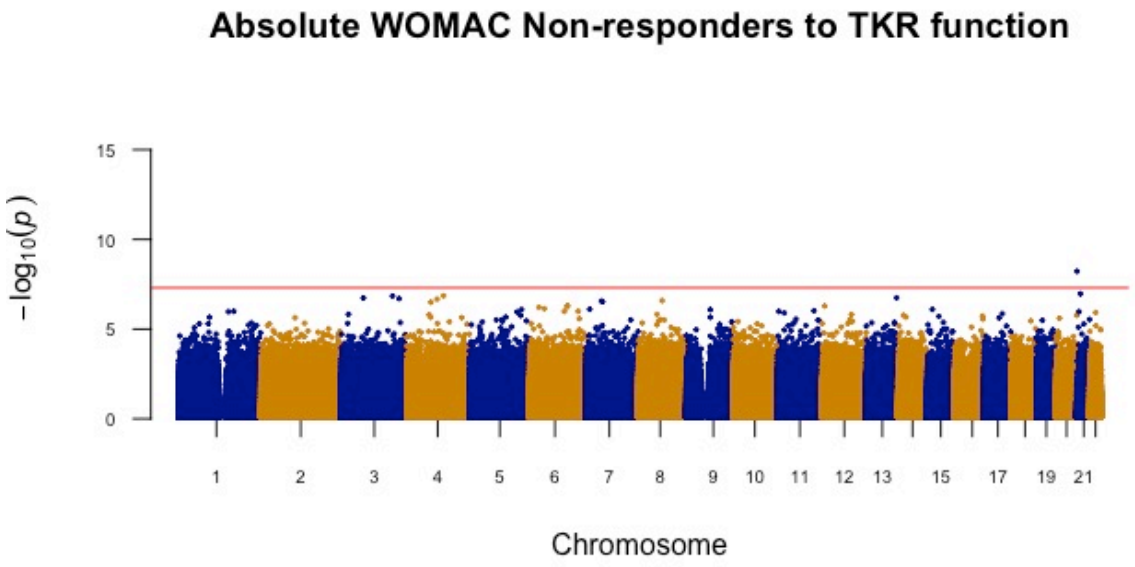


Figure 7.8: The regional association plot for the chromosomal region showing the most significantly associated variants on chr21 with the TJR function non-responders: **A)** SNP rs373135624 was significantly associated with non-responders to TJR function with $p \leq 5.97*10^{-9}$. This SNP is an intronic variant in the Tektin 4 Pseudogene 2 (TEKT4P2) Pseudogene on chr21, **B)** The top variant in the second associated locus was rs28971219 with $p \leq 7.03*10^{-09}$. Variants in this locus are intergenic that are adjacent to the *Transmembrane Phosphatase With Tensin Homology* (*TPTE*) gene.

Following the association analysis in the 315 knee OA patients, only rs28971219 on chr21 remained to be significantly associated with non-responding to the TJR function based on the Absolute WOMAC Score (OR= 0.29, $p=6.01*10^{-9}$, **Figure 7.9A**) and the OMERACT-OARSI (OR= 0.24, $p=1.78*10^{-8}$, **Figure 7.9B**) classification methods. Whereas none of the identified variants were associated with non-responding to the TJR function in the 126 hip OA patients, **Figure 7.10A and B**.

A



B

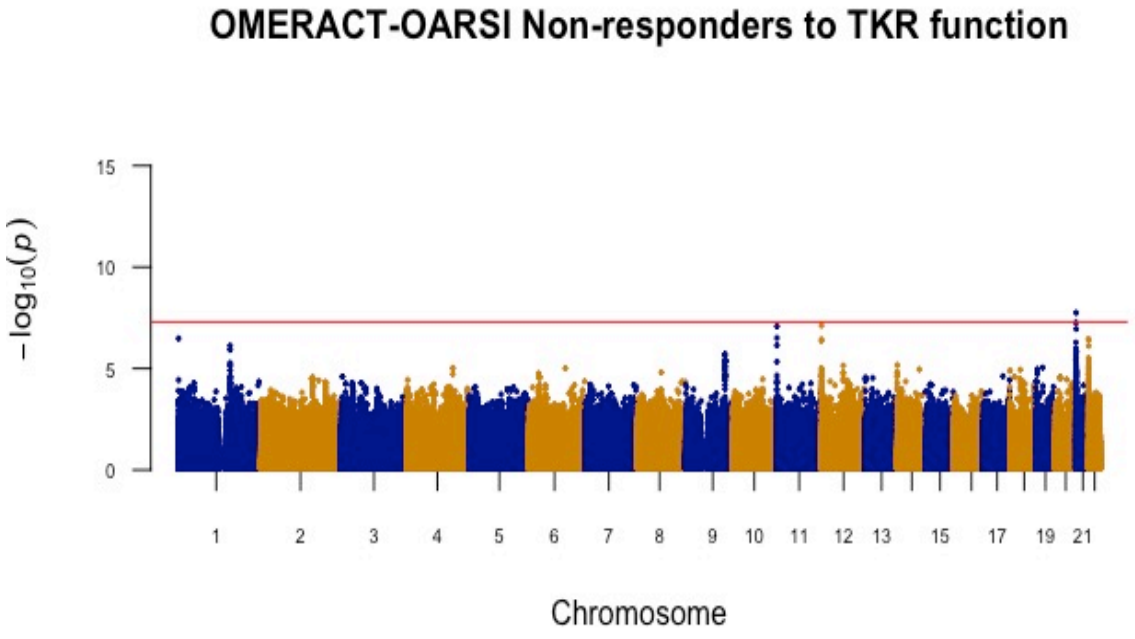
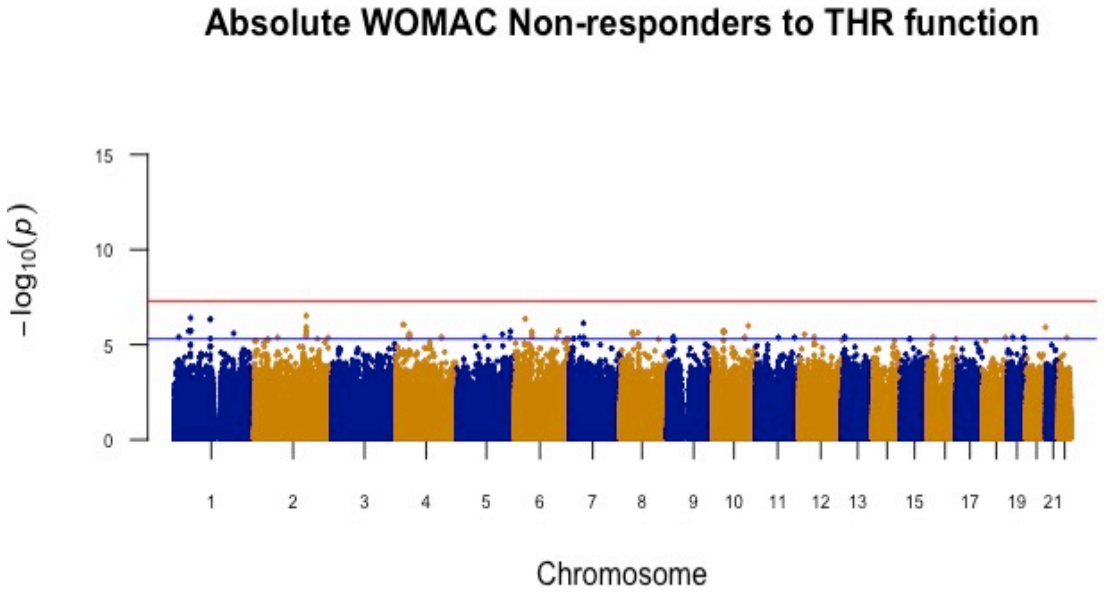


Figure 7.9: A Manhattan plot of the GWAS results showing the associated rs28971219 on chr21 with the TJR function non-responders in the 315 knee OA patients based on **A)** the Absolute WOMAC Score (OR=0.29, $p=6.01*10^{-9}$), and **B)** the OMERACT-OARSI (OR=0.24, $p=1.78*10^{-8}$) classification methods.

A



B

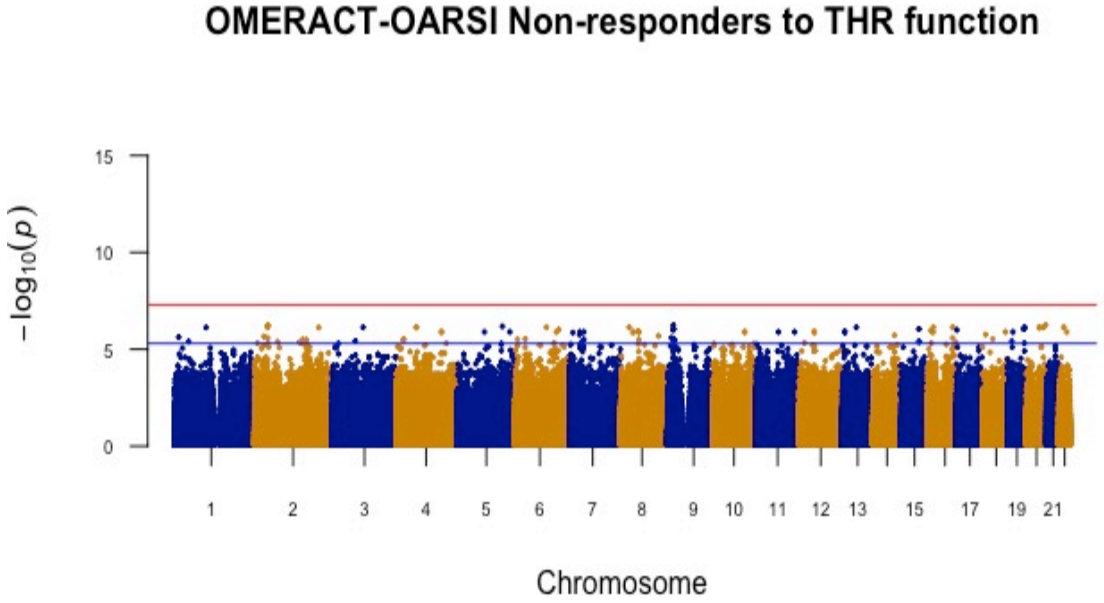
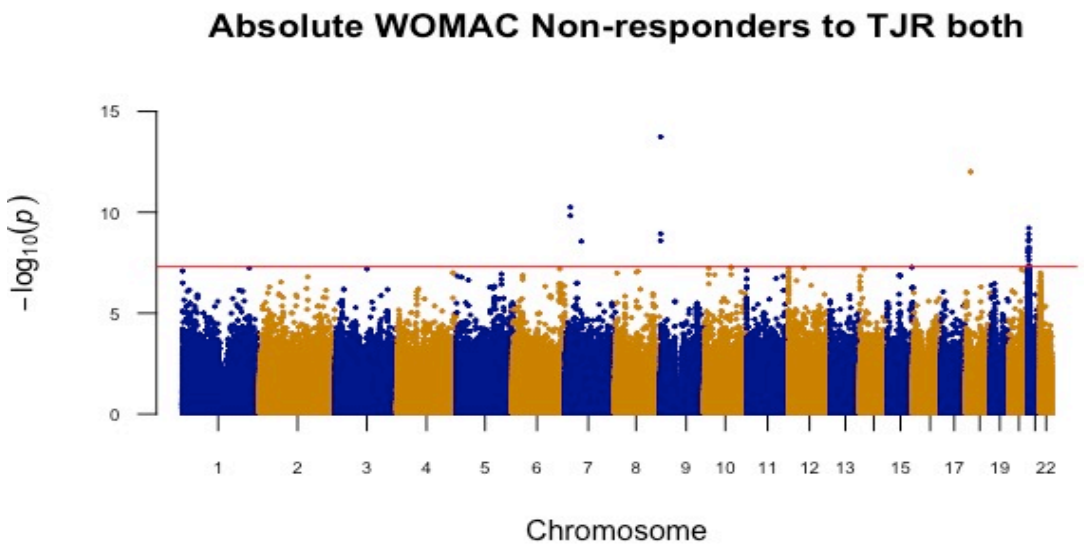


Figure 7.10: A Manhattan plot of the GWAS results showing no association of the genetic variations at $p<5*10^{-8}$ with the non-responders to the TJR function in the 126 hip OA patients in our study based on: **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods. The red line indicates the GWAS significance at $\alpha = 5*10^{-8}$.

Remarkably, the genetic variants on chr7, 9, 18, and 21 that were significantly associated with each of the TJR pain and TJR function non-responders separately were found to be significantly associated with non-responding to the TJR both (pain and function) in all 441 OA patients based on both classification criteria as shown in the Manhattan plot in **Figure 7.11**. There was no evidence of population stratification with inflation factor (λ)=0.98 based on the Absolute WOMAC Score and (λ)=0.99 based on the OMERACT-OARSI classification methods in the association analysis between the genetic variations and the TJR both (pain and function) non-responders in the 441 OA patients. Also, the Q-Q plot of the association test P-values did not visually show obvious deviation from what was expected under the assumption of no genetic association, **Figure 7.12**.

A



B

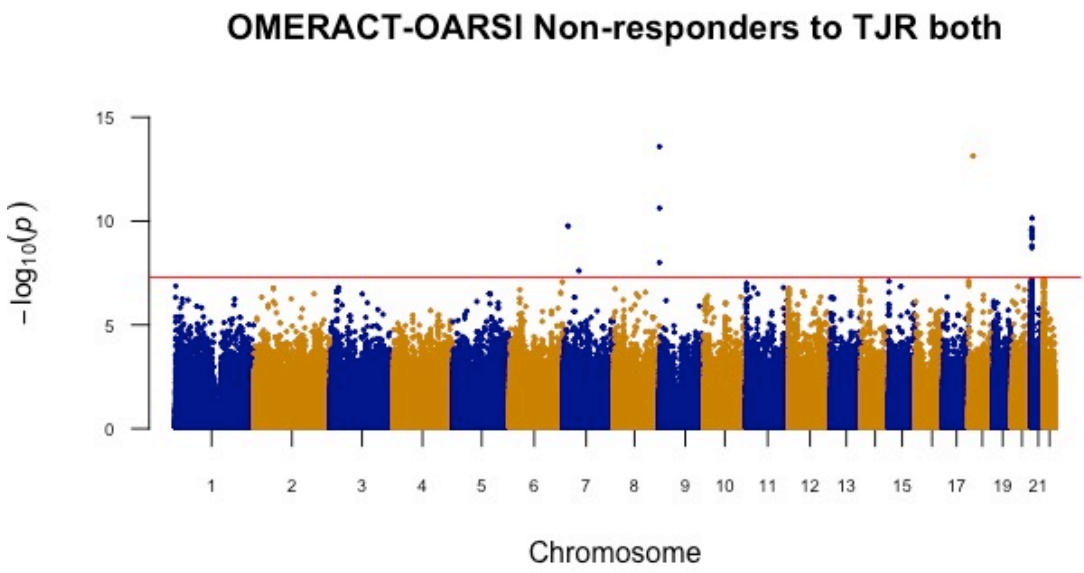
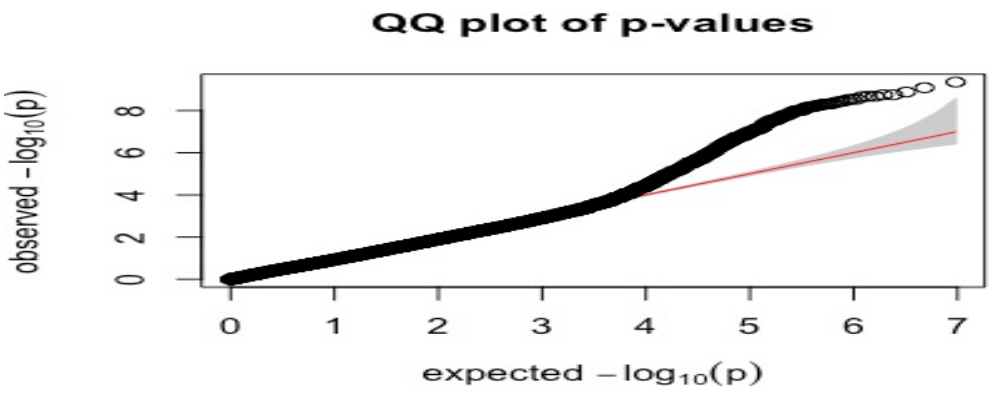


Figure 7.11: A Manhattan plot of the GWAS results showing the significantly associated genetic variations with non-responders to the TJR both (pain and function) in 441 OA patients from the NFOAS based on **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods. The red line indicates the GWAS significance at $\alpha = 5*10^{-8}$.

A



B

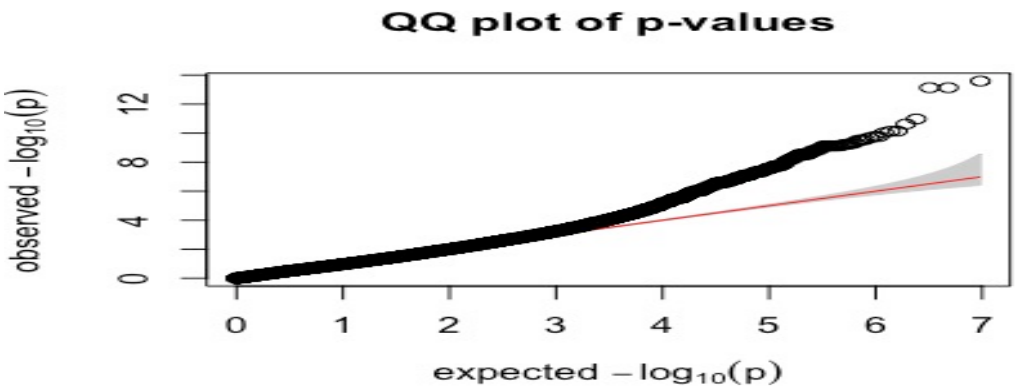


Figure 7.12: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the non-responders to both (pain and function) TJR in the 441 OA patients based on **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods.

We found that rs563726736, rs73058988, and rs62456377 on chr7 significantly associated with non-responders to the TJR both (pain and function) by both classification methods (all $p \leq 2.49*10^{-08}$) As shown in **Table 7.7**. Moreover, rs2441639, rs149992177, and rs9408586 on chr9 were significantly associated with non-responders to the TJR both (pain and function) at (all $p \leq 9.91*10^{-09}$) based on the Absolute WOMAC Score and the OMERACT-OARSI classification methods, **Table 7.7**. Furthermore. variants rs4797006 and rs146474469 on chr18 were found to be to most significantly associated genetic variations with non-responders to the TJR both (pain and function) from both classification method results (all $p \leq 9.83*10^{-13}$), **Table 7.7**. Likewise, two chromosomal regions on chr21 were detected to be significantly associated with non-responders to the TJR both (pain and function). The first locus included rs71236670, rs74843288, rs189820349, rs376101925, and rs4361469 variants were significantly associated with non-responders to the TJR both (pain and function) with all $p \leq 1.56*10^{-08}$, while the second region included variants rs2945173, rs1907506, rs4913173, rs28971219, and rs113445703 that were associated with non-responders to the TJR both (pain and function) based on both classification methods (all $p \leq 6.24*10^{-09}$), **Table 7.7**.

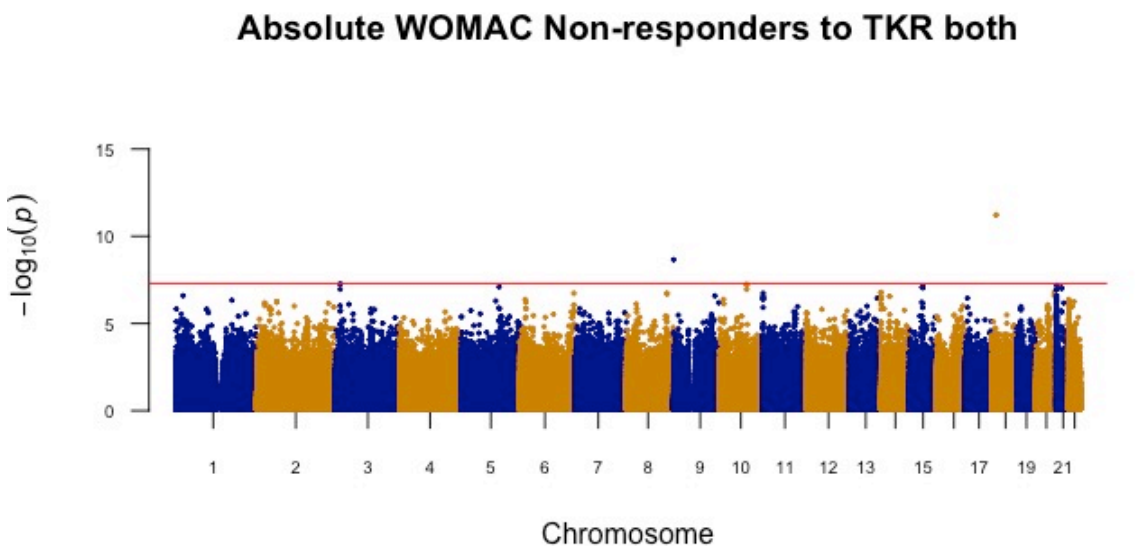
Table 7.7: The most significantly associated genetic variants with non-responding to the TJR both (pain and function) in the 441 OA samples.

CHR	position	Variant ID	Allele 1	The Absolute WOMAC Score				The OMERACT-OARSI			
				MAF_Aff	MAF_Un-aff	OR	P-value	MAF_Aff	MAF_Un-aff	OR	P-value
7	14983262	rs563726736	ACACACAC ACACACAC ACACACAC	0.13	0.01	15.80	$5.64*10^{-11}$	0.13	0.01	13.98	$1.74*10^{-10}$
7	14983276	rs73058988	C	0.13	0.01	13.53	$1.49*10^{-10}$	0.13	0.01	13.98	$1.74*10^{-10}$
7	49914144	rs62456377	G	0.15	0.02	9.09	$2.74*10^{-09}$	0.14	0.02	7.71	$2.49*10^{-08}$
9	52415	rs2441639	A	0.27	0.03	9.39	$1.82*10^{-14}$	0.25	0.03	9.21	$2.57*10^{-14}$
9	52625	rs149992177	CA	0.20	0.03	6.45	$2.57*10^{-09}$	0.18	0.02	6.11	$9.91*10^{-09}$
9	55099	rs9408586	C	0.19	0.02	8.16	$1.16*10^{-09}$	0.19	0.02	9.12	$2.34*10^{-11}$
18	13824807	rs4797006	G	0.18	0.02	12.01	$9.83*10^{-13}$	0.19	0.02	12.14	$7.35*10^{-14}$
18	13825028	rs146474469	C	0.18	0.02	12.01	$9.83*10^{-13}$	0.19	0.02	12.14	$7.35*10^{-14}$
21	9827857	rs71236670	C	0.10	0.46	0.15	$1.02*10^{-08}$	0.06	0.47	0.09	$2.76*10^{-10}$
21	9834803	rs74843288	C	0.14	0.50	0.19	$6.68*10^{-09}$	0.13	0.49	0.17	$1.57*10^{-09}$
21	9955015	rs189820349	A	0.15	0.53	0.20	$1.56*10^{-08}$	0.12	0.53	0.15	$2.10*10^{-10}$
21	9972507	rs376101925	A	0.15	0.54	0.20	$2.71*10^{-09}$	0.14	0.54	0.19	$4.49*10^{-10}$
21	9972611	rs4361469	A	0.15	0.53	0.21	$7.88*10^{-09}$	0.14	0.53	0.20	$1.55*10^{-09}$
21	10427188	rs2945173	C	0.83	0.45	5.05	$1.23*10^{-09}$	0.83	0.44	5.26	$3.03*10^{-10}$
21	10476925	rs1907506	A	0.14	0.51	0.18	$6.24*10^{-09}$	0.12	0.51	0.14	$2.95*10^{-10}$
21	10598148	rs4913173	T	0.10	0.47	0.15	$5.32*10^{-09}$	0.10	0.48	0.14	$1.98*10^{-09}$
21	10807264	rs28971219	A	0.13	0.52	0.17	$2.20*10^{-09}$	0.11	0.52	0.13	$7.27*10^{-11}$
21	10839945	rs113445703	CGAATG	0.10	0.49	0.13	$6.09*10^{-10}$	0.11	0.49	0.14	$6.79*10^{-10}$

CHR: chromosome, Allele 1: the risk allele, MAF_Aff: minor allele frequency of the risk allele in the TJR function non-responders, MAF_un-Aff: minor allele frequency of the risk allele in the TJR pain responders, Allele 2: Major allele, OR: odd ratio.

Variants rs2441639 on chr9; rs4797006, and rs146474469 on chr18 remained significantly associated (all $p \leq 2.20*10^{-09}$) with non-responding to the TJR both (pain and function) in the 315 knee OA patients from both classification methods analysis as shown in the Manhattan plot (**Figure 7.13)** and **Table 7.8**. While the GWAS analysis in the 126 hip OA patients did not detect these variants to be significantly associated with the non-responding to the TJR both (pain and function), **Figure 7.14**.

A



B

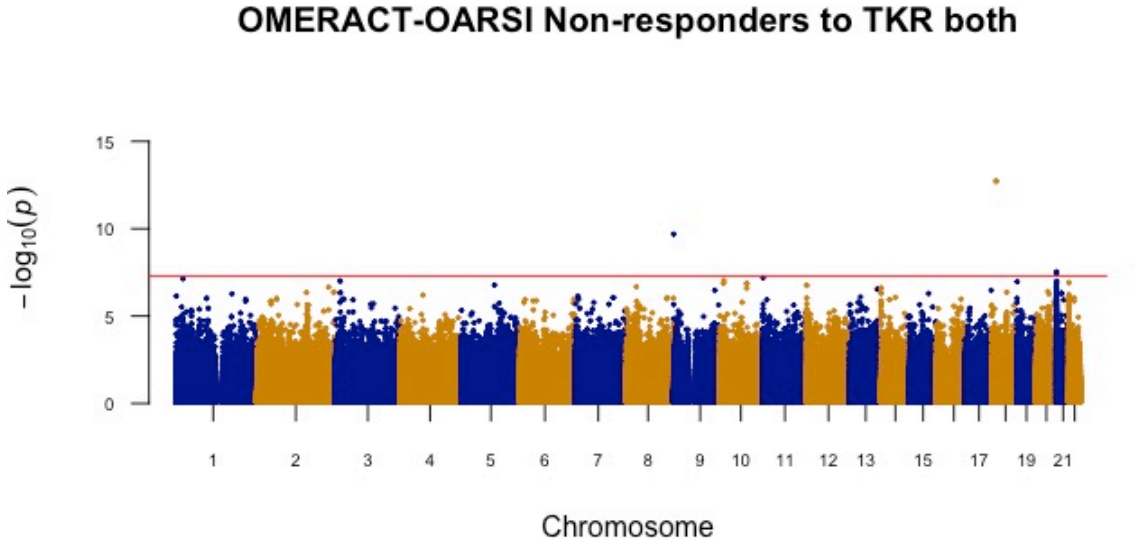


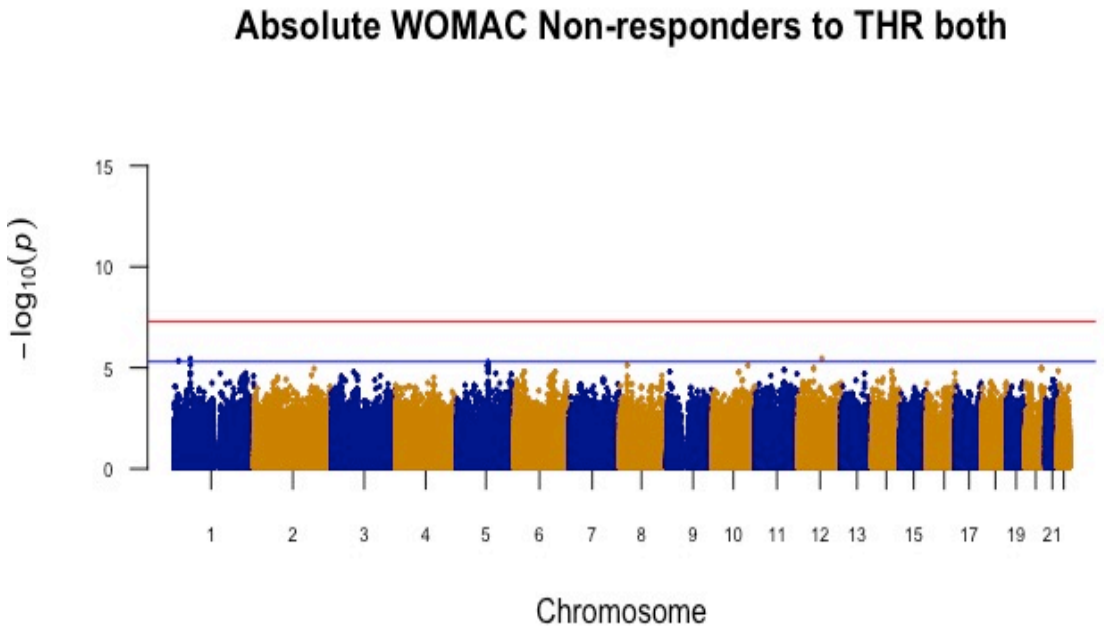
Figure 7.13: A Manhattan plot of the GWAS results showing the significantly associated genetic variants with the TJR both (pain and function) non-responders in the 315 knee OA patients based on **A)** the Absolute WOMAC Score, **B)** the OMERACT-OARSI classification methods.

Table 7.8: The most significantly associated genetic variants with non-responding to the TJR both (pain and function) in the 315 knee OA patients.

CHR	Position	Variant ID	Allele 1	The Absolute WOMAC Score				The OMERACT-OARSI			
				MAF_Aff	MAF_Un-aff	OR	P-value	MAF_Aff	MAF_Un-aff	OR	P-value
9	52415	rs2441639	A	0.23	0.03	7.70	$2.20*10^{-09}$	0.19	0.01	8.11	$2.05*10^{-10}$
18	13824807	rs4797006	G	0.18	0.01	17.18	$5.93*10^{-12}$	0.19	0.01	18.70	$1.91*10^{-13}$
18	13825028	rs146474469	C	0.18	0.01	17.18	$5.93*10^{-12}$	0.19	0.01	18.70	$1.91*10^{-13}$
21	9827857	rs71236670	C	0.09	0.47	0.13	$5.07*10^{-07}$	0.06	0.48	0.08	$3.44*10^{-08}$
21	10807264	rs28971219	A	0.15	0.53	0.20	$3.15*10^{-07}$	0.13	0.53	0.16	$2.95*10^{-08}$

CHR: chromosome, Allele 1: the risk allele, MAF_Aff: minor allele frequency of the risk allele in the TJR function non-responders, MAF_un-Aff: minor allele frequency of the risk allele in the TJR pain responders, Allele 2: Major allele, OR: odd ratio.

A



B

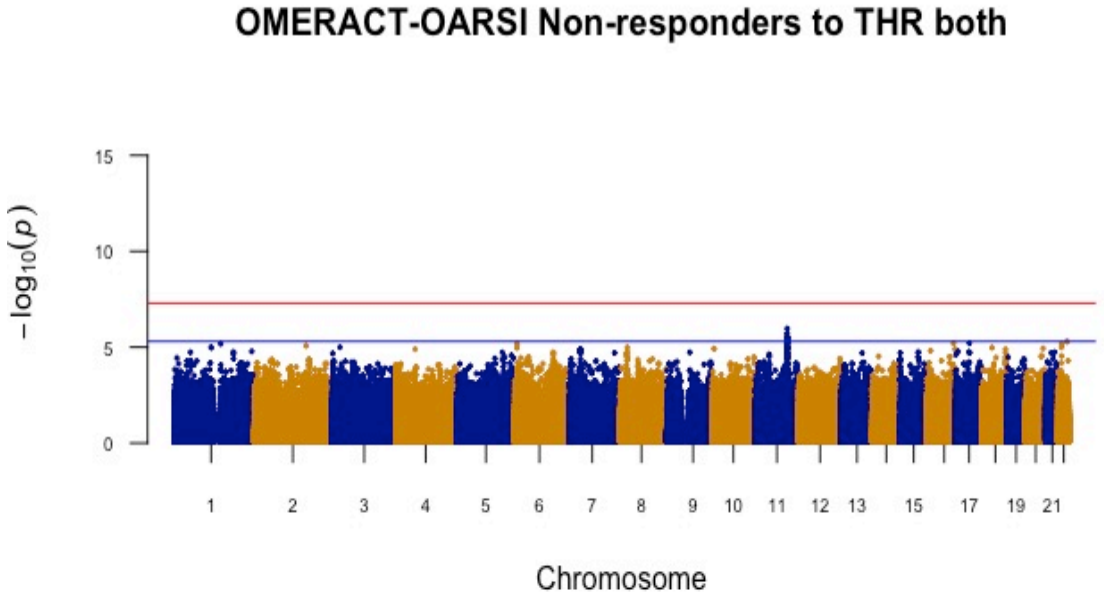
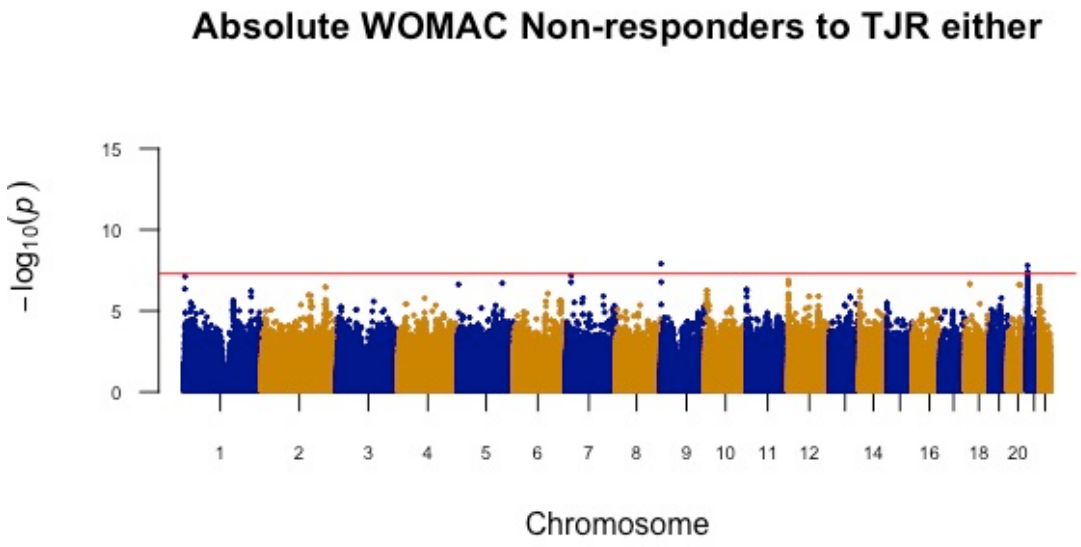


Figure 7.14: A Manhattan plot of the GWAS results showing no association of the genetic variations at $p<5*10^{-8}$ with the non-responders to the TJR both (pain and function) in the 126 hip OA patients in our study based on: **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods. The red line indicates the GWAS significance at $\alpha = 5*10^{-8}$.

Further, the GWAS analysis in all 441 OA patients identified one SNP on chr9 and two SNPs on chr21 from both classification methods to be significantly associated with the non-responding of the TJR to either (pain or function) as shown in the Manhattan plot in **Figure 7.15**. There was no evidence of population stratification with inflation factor (λ)=0.96 based on the Absolute WOMAC Score criteria and (λ)=0.97 based on the OMERACT-OARSI classification method. Also, the Q-Q plot of the association test P-values did not visually show obvious deviation from what was expected under the assumption of no genetic association, **Figure 7.16**.

A



B

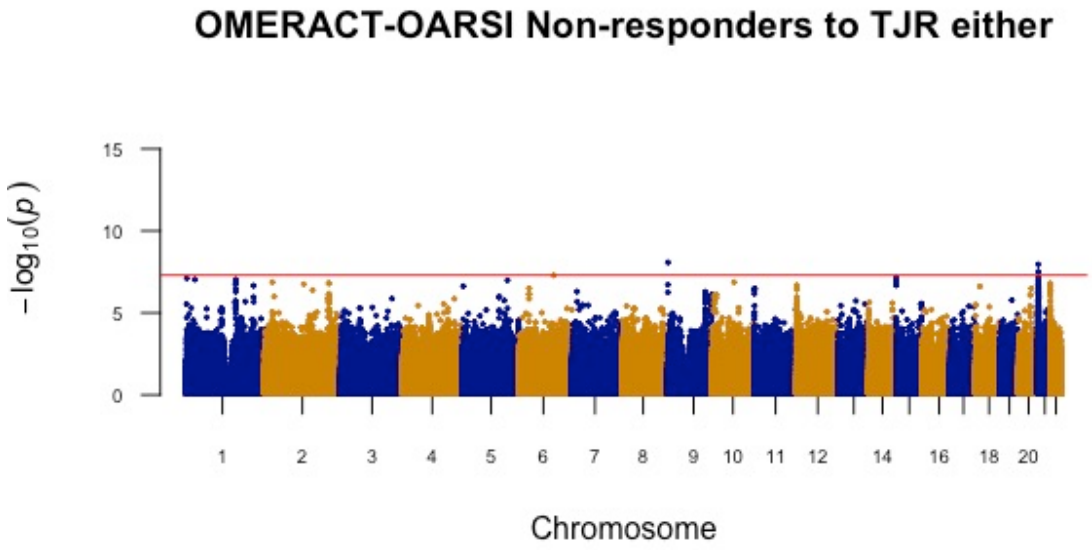
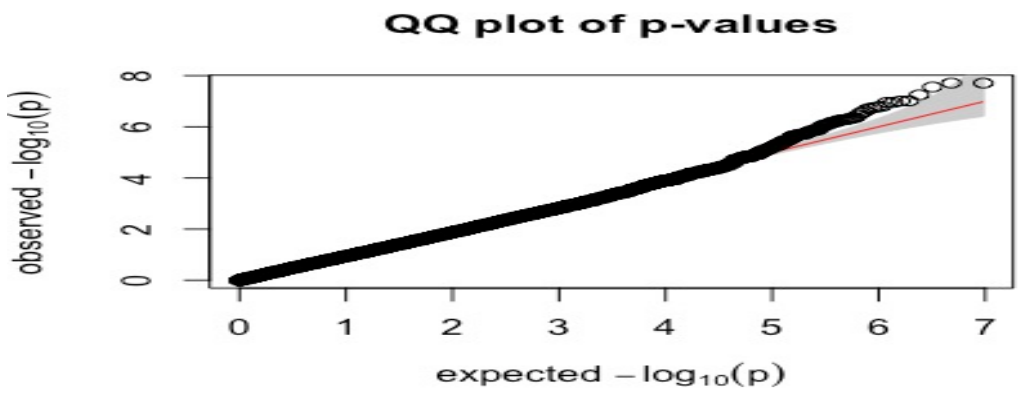


Figure 7.15: A Manhattan plot of the GWAS results showing the significantly associated genetic variations with non-responders to the TJR either (pain or function) in 441 OA patients from the NFOAS based on **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods. The red line indicates the GWAS significance at $\alpha = 5*10^{-8}$.

A



B

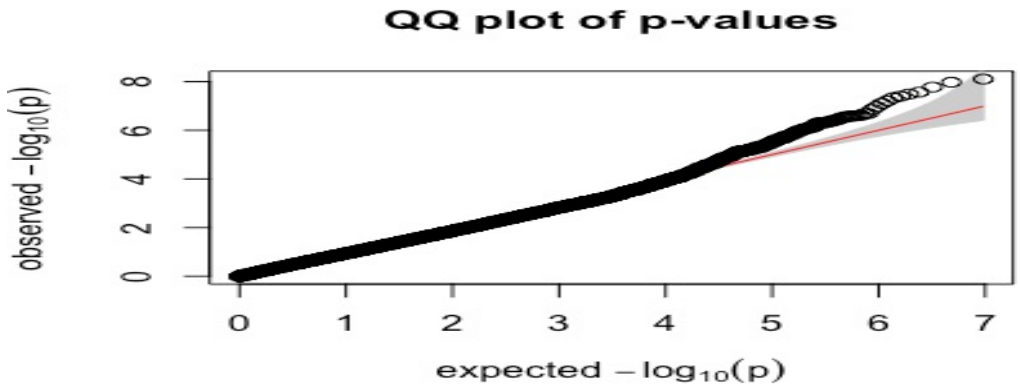


Figure 7.16: The Q-Q plot that shows the deviation of the observed P-value from the null hypothesis that there is no association between the genetic variants and the non-responders to either (pain or function) TJR in the 441 OA patients based on **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods.

SNP rs2441639 on chr9 was detected to be significantly associated with non-responding to the TJR either (pain or function) with OR=5.14 (all $p \leq 1.24*10^{-8}$) based on both of the classification criteria, **Table 7.9**. Also, SNPs rs377713075 and rs373135624 on chr21 were significantly associated with non-responding to the TJR either (pain or function) with OR~0.30 (all $p \leq 3.92*10^{-8}$) from the analyses of the Absolute WOMAC Score and the OMERACT-OARSI classification methods, **Table 7.9**.

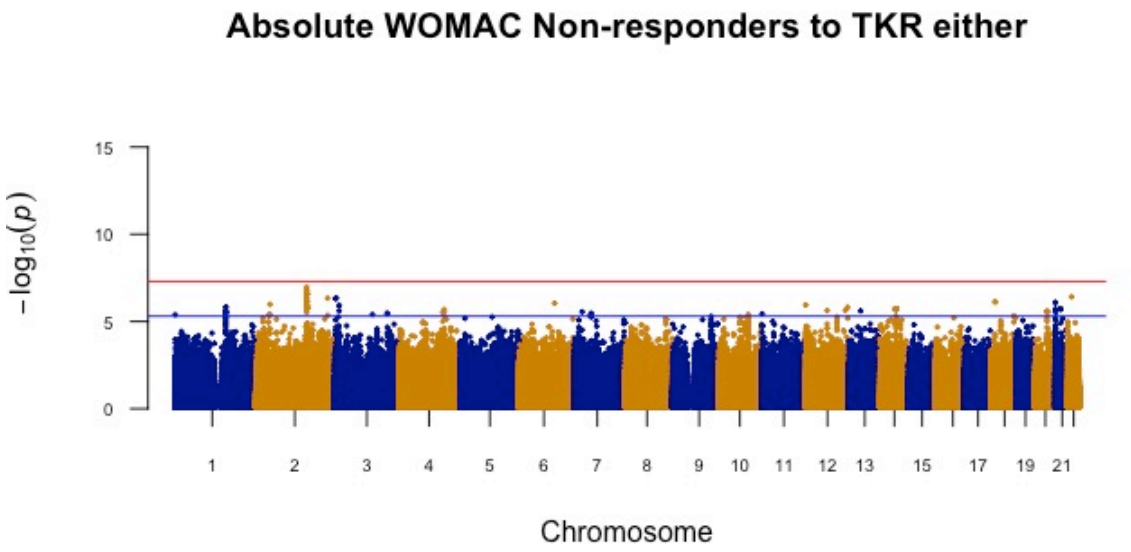
Table 7.9: The most significantly associated genetic variants with non-responding to either (pain or function) of the TJR in the 441 OA samples.

CHR	position	Variant ID	Allele 1	The Absolute WOMAC Score				The OMERACT-OARSI			
				MAF_Aff	MAF_Un-aff	OR	P-value	MAF_Aff	MAF_Un-aff	OR	P-value
9	52415	rs2441639	A	0.1646	0.02801	5.14	$1.24*10^{-08}$	0.1646	0.02865	5.14	$8.35*10^{-09}$
21	9831558	rs377713075	T	0.1646	0.4314	0.29	$1.58*10^{-08}$	0.1709	0.4341	0.30	$3.08*10^{-08}$
21	9912348	rs373135624	G	0.2089	0.465	0.33	$3.92*10^{-08}$	0.2025	0.4699	0.31	$1.11*10^{-08}$

CHR: chromosome, Allele 1: the risk allele, MAF_Aff: minor allele frequency of the risk allele in the TJR function non-responders, MAF_un-Aff: minor allele frequency of the risk allele in the TJR pain responders, Allele 2: Major allele, OR: odd ratio.

However, the association analysis in the 315 knee (**Figure 7.17)** and the 126 hip (**Figure 7.18)** OA patients did not identify genetic variation to be associated with the non-responding to the TJR either (pain or function) at the GWAS significance level of $p=5*10^{-8}$.

A



B

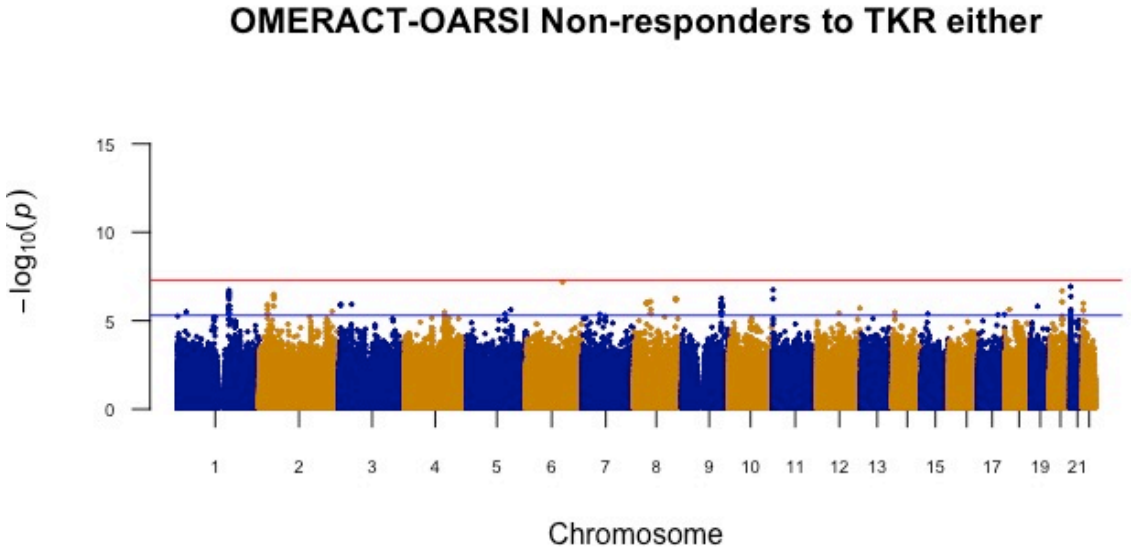
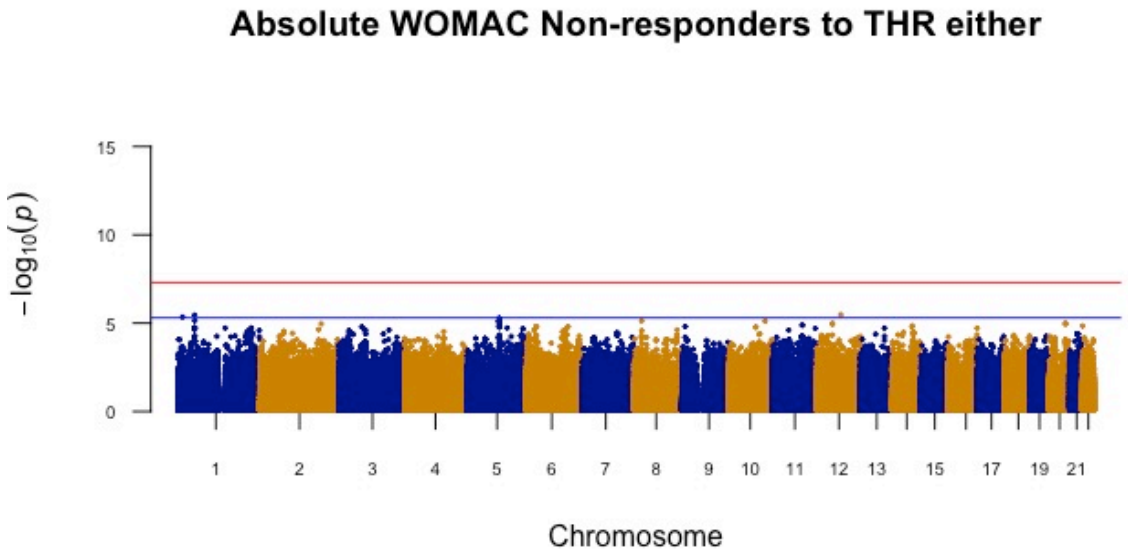


Figure 7.17: A Manhattan plot of the GWAS results showing the significantly associated genetic variants with the TJR either (pain or function) non-responders in the 315 knee OA patients based on **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods.

A



B

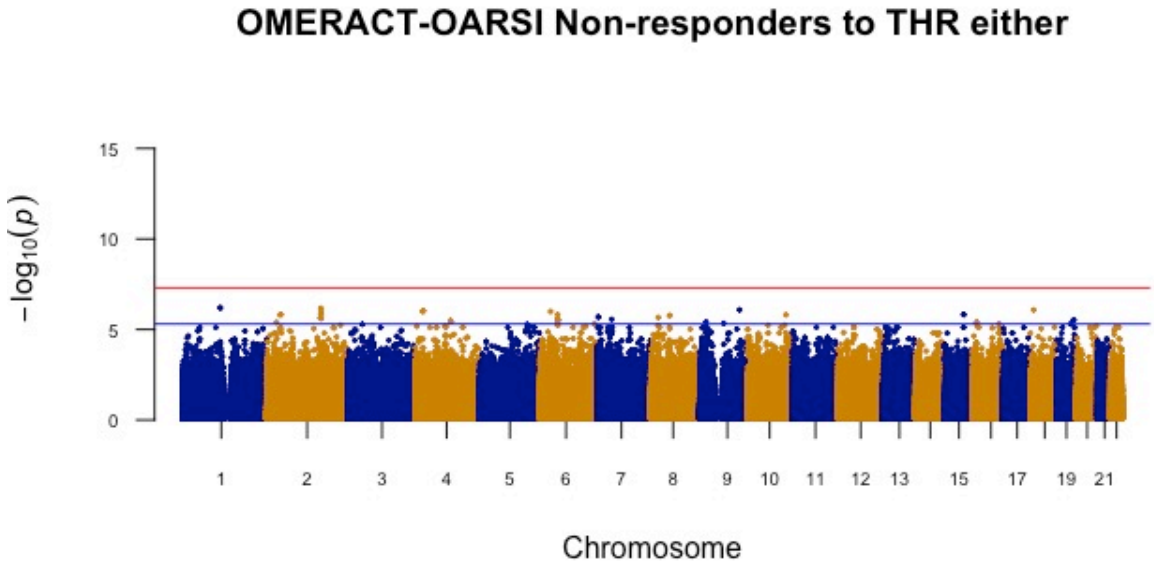


Figure 7.18: A Manhattan plot of the GWAS results showing no association of the genetic variations at $p<5*10^{-8}$ with the non-responders to the TJR either (pain or function) in the 126 hip OA patients in our study based on: **A)** the Absolute WOMAC Score, and **B)** the OMERACT-OARSI classification methods. The red line indicates the GWAS significance at $\alpha = 5*10^{-8}$.

7.6 Discussion:

This analysis was one of the few studies that investigated the genome-wide association between the genetic factors and the TJR outcome. Our study identified novel genetic variations to be associated with non-responding to the TJR pain relief and function improvement. The identified variants had a good discriminatory power to predict non-responders to the TJR therapy as indicated by the AUC ≥ 0.75 with an average sensitivity of 0.68 and specificity of 0.77. If these genetic variants were experimentally validated, they can be used in direct clinical applications as biomarkers to predict the TJR outcome. They can help clinicians to decide whether to undergo the TJR or not, and they can be used to educate OA patients about their TJR sugary expectations. Furthermore, these genetic variants can be applied to improve the OA management and rehabilitation strategies, as well as develop DMOADs. As a result, these findings can contribute to the reduction of the TJR healthcare service cost and improve the quality of life of people living with OA.

Our GWAS analysis showed that most of the identified signals for the non-responder to the TJR therapy in the entire OA cohort (n=441) analysis were due to the TKR non-responders (n=315), since none of the detected genetic variants were significantly associated with non-responders to the THR (n=126) at GWAS significance level likely due to the small hip OA sample size.

Our analysis identified two loci on chr7 to be significantly associated with the TJR poor outcome at the GWAS significance level of $p \leq 5*10^{-8}$. The INDEL rs563726736 was the top significantly associated variant with non-responding to the TJR pain ($p \leq 1.26*10^{-8}$) in the whole cohort of 441

OA samples. Interestingly, this INDEL was also significantly associated ($p \leq 1.74*10^{-10}$) with non-responding to the TJR both (pain and function) in the entire OA cohort. This INDEL is an intronic variant in intron one of the *diacylglycerol kinase beta* (*DGKB*) gene on chr7. The *DGKB* gene codes a member of ten diacylglycerol kinase enzymes that regulate the intracellular concentration of the second messenger diacylglycerol and thus play a key role in cellular processes. The encoded protein by the *DGKB* gene metabolizes 1,2,diacylglycerol to produce phosphatidic acid and regulates the respective levels of these two bioactive lipids (832). Thereby, it acts as a central switch between the signaling pathways activated by these second messengers with different cellular targets and probably opposite effects in numerous biological processes. Among its related pathways are GPCR downstream signalling and Response to elevated platelet cytosolic Ca2+. However, the *DGKB* gene has not been previously linked to OA, a recently published study by Lovšin *et al* (2023) found a novel copy number variant in the *DGKB* gene to be implicated in the BMD, fragility, and remodeling in OP patients from the European population (833). It has estimated that the inverse association of OA with OP can be explained by shared genetic components between OA and OP (168). It was reported that individuals with high BMD have a tendency for formation of multiple osteophytes, which elevates their susceptibility to OA (166). Furthermore, the studies of high BMD and OA in non-weightbearing joints in hands referred this association to osteophyte formation rather than cartilage loss and JSN (167). The second variant on chr7 that was detected to be significantly associated with the non-responders to the TJR pain in the 441 OA samples was rs62456377 ($p \leq 2.79*10^{-08}$). Moreover, this SNP was significantly associated ($p \leq 2.49*10^{-08}$) with non-responders to the TJR both (pain and function) in the whole OA cohort. This SNP is intronic variant that falls in the intron two of the *von Willebrand factor C domain containing 2* (*VWC2*) gene on chr7. The

VWC2 gene encodes a secreted bone morphogenic protein antagonist. The *VWC2* gene encoded protein is possibly involved in neural function and development and may have a role in cell adhesion. A study conducted by Almehmadi *et al* (2018) identified that VWC2 promotes bone formation by inhibiting Activin-Smad2 signaling pathway (834). Thus, our results suggested implication of rs62456377 in the bone remodeling and osteophytes formation, and it is association with non-responding to the TJR pain mechanisms in our study patients. The identified variants in the *DGKB* and the *VWC2* genes may represent targets for OA bone remodeling and osteophytes formation therapies in the non-responders to the TJR surgery.

Moreover, we found a locus on chr9 including rs2441639, rs149992177, and rs9408586 to be significantly associated with the TJR pain non-responders. The SNP (rs2441639) was the most significantly associated variant with non-responding to the TJR pain ($p \leq 7.07*10^{-14}$) in the whole cohort of 441 OA samples. Remarkably, rs2441639 was significantly associated with the TJR non-responders to the TJR both (pain and function) in the whole cohort of 441 OA samples ($p \leq 2.57*10^{-14}$) and remained significantly associated with the TJR non-responders to the TJR both (pain and function) in the 315 knee OA patients ($p \leq 2.20*10^{-9}$). Moreover, this SNP was significantly associated with the non-responders to the TJR either (pain or function) in the 441 OA samples with $p \leq 1.24*10^{-8}$. SNP rs2441639 is in the intron two of the *PGM5P2* pseudogene on chr9. A previous study conducted by Lin *et al.* (2020) detected a genetic variant in the *PGM5P2* pseudogene to be significantly associated with the familial short stature (FSS) in a group of Han Chinese ancestry in Taiwan (835). Another study performed by Zhang *et al.* (2016) reported that the dysregulation of the *PGM5P2* pseudogene contributes to the pathological processes of FLS in the RA patients (836). Synovitis in RA drives pain, cartilage degradation,

and pannus formation with subsequent erosions (837). It is increasingly recognized that synovitis is also observed both in early and late OA stages (838). Synovitis of OA may sometimes resemble that which is characteristic of RA. Subchondral inflammation might also contribute to increased bone turnover and joint damage in both OA and RA. Although, OA and RA differ in their aetiologies, with mechanical factors being key to knee OA progression and specific immunity driving RA, common mechanisms may contribute to joint damage and pain in both conditions (839). Thus, our findings suggested the implication of rs2441639 and the *PGM5P2* pseudogene in the OA synovitis and pain mechanisms in the TJR pain non-responders in our study, and the identified variant in the *PGM5P2* pseudogene can be a target for OA synovitis and pain mechanisms management strategies.

Our analysis showed that rs4797006 and rs146474469 on chr18 were significantly associated with non-responding to the TJR pain in the 441 OA samples ($p \leq 7.43*10^{-11}$) and in the 315 knee OA patients ($p \leq 1.77*10^{-11}$). Moreover, these variants were detected to be significantly associated with non-responding to the TJR both (pain and function) in the entire OA cohort ($p \leq 9.83*10^{-13}$) and in the 315 knee OA patients ($p \leq 5.93*10^{-12}$). Variants rs4797006 and rs146474469 are located 221 bps apart in full LD ($r^2=1$) in the intron one of the *MC5R* gene that encodes the melanocortin 5 receptor. The MC5R is a member of the G protein-coupled MCR family that includes (MC1R, MC2R, MC3R, MC4R, and MC5R) members. GPCRs comprise the largest and most diverse family of integral membrane proteins (602,840). All GPCRs are characterized by the presence of seven membrane-spanning α-helical segments separated by alternating intracellular and extracellular loop regions (840). GPCRs link extracellular stimuli to intracellular responses through signal transduction mechanisms and mediate most of our

physiological responses to hormones, neurotransmitters, heart contractility, immune responses, and environmental stimulants (841). Four out of six GPCRs families are present in humans, including rhodopsin-like (family A), secretin (family B), metabotropic glutamate-like (family C), and adhesion and frizzled/smoothened (family F) (842,843). Due to the involvement of the GPCRs in wide range of physiological and pathophysiological processes, they became exciting targets for drug discovery for metabolic diseases, cancers, and neurodegenerative diseases (610,611). Different types of the GPCRs were reported to play critical roles in modulating OA symptoms including cartilage degeneration, subchondral bone sclerosis and chronic pain (602). For instance, the MCR family share a common structure consisting of seven α-helical transmembrane domains, an extracellular N terminus and intracellular C tail, connected by three alternating extracellular and three intracellular loops. Sequence of MCRs revealed high homologies, ranging from 38% identity between MC2R and MC4R to 60% identity between MC4R and MC5R (844). The MCRs are distributed on multiple cell types throughout body tissues and are stimulated by the melanocortin peptides including ACTH, and MSH alpha, beta, and gamma (845). Activated MCRs by melanocortin peptides regulate diverse intracellular signal transduction mechanisms, containing cAMP and PK signaling pathways (846). Melanocortin system plays a key role in maintaining the homeostasis of our body via their neuro-immune-endocrine activities and regulates a diverse array of physiological functions, including melanogenesis, anti-inflammation, immunomodulation, fatty acid oxidation, adrenocortical steroidogenesis, hemodynamics, natriuresis, and energy homeostasis (844). They are believed to be involved in many disease states including pigmentary disorders, adrenal disorders, obesity, anorexia, prolonged and neuropathic pain, and inflammatory response (847). In particular, MC1R is involved in skin and hair pigmentation and immune cell regulation (848); MC2R binds

exclusively to ACTH melanocortin peptide and acts in the adrenal gland to stimulate glucocorticoid synthesis and release (849); MC3R regulates immune cell function; and both MC3R and MC4R contribute to energy homeostasis and regulate neuronal interactions with autonomic functions (850–852); where the MC5R is expressed in skin, skeletal muscles, bone marrow, lung, spleen, and brain cells. It has the ability to respond to almost all melanocortin peptides, except gamma-MSH (853). However, its function remains to be fully elucidated, the MC5R is reported to have anti-inflammatory effects including negative regulation of pro-inflammatory response to antigenic stimulus, interact with opioid system, affect the pain sensitivity, and regulate metabolic processes (854,855). Studies reported that the MC5R contributes to the immunomodulatory functions in B and T lymphocytes, and mast cells (856,857). Arthritis studies identified that melanocortin peptides and their receptors play a central role in the OA progression, and joint prolonged pain and disability. Specifically, activation of MC1R and MC3R was observed to downregulate IL-1β, IL-6 and IL-8 release, MMPs expression, and inhibit cell death in chondrocytes (858). Furthermore, a study performed by Lorenz *et al*. (2014) in animal model noted that mice with MC1R deficient developed extreme OA pathology of cartilage degradation (859). From another point of view, accumulating evidence have suggested that melanocortin peptides and their MC1R and MC4R receptors regulate pain. The MC1R is known to be expressed in the periaqueductal gray (PAG) of the brain where it plays a role in pain regulation (629). A study by Delaney *et al*. (2010) reported a reduced inflammatory pain response and an increased tolerance to noxious heat in female mice lacking the MC1R (860). Additionally, the MC4R was identified to play a critical role in the regulation of pain processing within a wide range of experimental models with painful conditions (861). Interestingly, our findings suggested that dysfunction and downregulation of

the *MC5R* gene by similarity with the *MC1R*, *MC3R*, and *MC4R* may lead to immune modulation deficiencies, suppression of the anti-inflammatory effect, and elevated inflammatory pain in non-responders to the TJR. If these results are experimentally validated, the identified genetic variants in the *MC5R* gene will become promising biomarkers to predict the TJR outcome and exciting potential target for anti-inflammatory drug discoveries in OA.

In addition, our association analysis identified two loci on chr21 to be significantly associated with poor outcome of the TJR. The first locus included variants rs71236670, rs189820349, rs376101925, rs4361469, rs373135624, rs74843288, and rs377713075 that were significantly associated (all $p \leq 3.92*10^{-8}$) with the TJR poor outcome in the whole OA cohort. Only rs71236670 remained significantly associated ($p=3.44*10^{-8}$) with non-responding to the TJR both (pain and function) in the 315 knee OA patients based on the OMERACT-OARSI classification method only. These variants are harbouring the *TEKT4P2* Pseudogene on chr21 that has unknown function. Furthermore, the top SNP in the second significantly associated locus with the TJR poor outcome on chr21was rs1907506 ($p \leq 1.07*10^{-8}$). This SNP is adjacent to the top associated SNP rs28971219 with non-responders to the TJR function on Chr21. Bioinformatics analysis showed that these genetic variants are close to the *TPTE* gene, but interestingly, pairwise linkage analysis revealed that none of the identified variants on chr21are linked (LD $r^2=0.2$). The *TPTE* gene encodes a phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase with a tensin-like domain, which negatively regulates intracellular levels of phosphatidylinositol-3,4,5-trisphosphate. The *TPTE* gene acts as a tumor suppressor gene by negatively regulating Akt/PKB signaling pathway. The PTEN opposes the action of PI3K that is needed for the activation of the Akt anti-apoptotic and tumor-promoting molecule (862). An

animal model study conducted by Li *et al*. (2019) identified an implication of the PTEN in negative regulation of pro-inflammatory cytokines and chemokines of the FLS in adjuvant-induced arthritis rats (863). Another study reported that the suppression of RA-FLS was mediated by the PTEN involving the *survivin* gene silencing (864). The FLS produce major components of the synovial fluid and participate in the joint homeostasis (865). Pathogenesis of the FLS is linked to the OA joint pain and disability (866,867). Our results supported these findings and suggested the identified genetic variants as genetic predictors of the TJR poor outcome; however, these findings warrant further investigations. If the identified variants are validated, they will represent potential target for the DMOADs.

To the best of our knowledge, this is the first study that investigated the genome-wide association between the genetic factors and the TJR non-responding. Our analysis assessed pain and disability in study participants before and up to 4 years following TJR therapy using the WONAC OA-index. This study used multiple non-responders to the TJR classification criteria. However, our study has a number of limitations, this study had a small sample size (n=441) for a GWAS analysis that may lower the power of the study and increase the type I error prediction, and all the study participants were from NL, which is a genetically/ethnically homogeneous population that may limit the generalizability of our results to other populations.

In conclusion, Our GWAS analysis identified novel genetic variants implicated in the pro-inflammatory response, prolonged OA pain, and pathogenesis of the FLS to be associated with non-responding to the TJR therapy. If these genetic variations are validated, they can predict the TJR poor outcome to educate patients about the expectation of their operation, help physicians to

decide whether to perform the TJR operation, contribute to the reduction of the healthcare service resources use, and improve the quality of life of people living with arthritis.

Chapter 8: General discussion and conclusion

8.1 General discussion:

OA is a heterogeneous group of overlapping distinct conditions that have different etiologies, but similar biological and clinical manifestations. Advances in the physiopathology and manifestations of OA revealed that variety of potential molecular and cellular changes can be involved in the joint destruction process leading to deteriorate all components of the synovial joints in hands, knees, hips, shoulders, and/or spine. OA occurs in different extents across patients and results in an array of possible structural OA manifestations including variable degrees of inflammation, meniscal lesions, and bone damage leading to joint pain, swelling, bone deformity, and muscle weakness (494). The OARSI defined OA as 'a disorder involving movable joints characterized by cell stress and extracellular matrix degradation initiated by micro- and macro-injury that activates maladaptive repair responses including pro-inflammatory pathways of innate immunity. The disease manifests first as a molecular derangement (abnormal joint tissue metabolism) followed by anatomic and/or physiologic derangement (characterized by cartilage degradation, bone remodeling, osteophyte formation, joint inflammation, and loss of normal joint function) that can culminate in illness' (10).

OA is the most common chronic progressive joint condition and one of the ten most disabling diseases in the developed countries. The worldwide prevalence and economic burden of OA has doubled since the fifties of the last century. It was estimated to affect about 10% of the world population aged 60 years and older (112,428). Pain is the strongest risk factor that forces OA patients to seek medical assistance (868). Moreover, about 80% of OA patients have limitations in movement, and 25% of them cannot perform their major daily activities with reduced health-related quality of life, and excess mortality (113).

An important aspect of OA is its extraordinary interpatient variability in clinical and structural manifestations (869,870). This heterogeneity may be one of the major factors associated with the complexity of OA and with the difficulties to identify general OA therapeutic strategies. However, efforts have been made to unravel OA heterogeneity, and several phenotypes and endotypes have been proposed, none of them has been sufficiently validated for clinical or research use yet, and the current treatment of OA is primary palliative focusing on symptom management, and no therapies other than the TJR have been proven to modify OA progression or to be highly effective for symptomatic relief to improve physical activity and health-related quality of life (871−873).

For instance, randomized controlled trials of the DMOADs have failed to show both significant pain relief and structural outcomes due to the heterogeneity of the disease, study cohort inclusion and exclusion criteria, and limited duration of studies. As well as definition of structural outcomes (874). Also, study findings may not be applicable in real-world conditions, since OA primarily affects older adults who often have multiple comorbidities, and the effect of these comorbidities on OA progression or treatment response remains unclear (872,874). Moreover, one of the challenges of OA clinical trial design is the difficulty of accurately selecting the suitable population that efficiently demonstrate the treatment effect (874), since it is unclear at present which patients would be most suitable for a specific therapy (873). For example, the failure of bisphosphonates to slow OA progression might have been due to enrolling any patient with symptomatic OA rather than selecting patients with greater subchondral bone turnover (873).

However, contemporary studies have focused on the identification of clinical phenotypes and endotypes based on functional and pathophysiologic mechanisms, including molecular pathways (872), overlap among clinical phenotypes and molecular endotypes further complicates the picture, since single patient may display a combination of clinical phenotypes; further, a single phenotype may have several underlying endotypes. This variability may account for the clinical heterogeneity seen in practice (872).

In order to address the heterogeneity of OA to improve clinical research and trials, a new model of understanding OA based on a phenotype-guided approach is urgently needed. The ACR/Arthritis Foundation and the OARSI recently published recommendations for OA management emphasizing coordination of pharmacologic and non-pharmacologic modalities and highlighting unmet needs for additional pharmacologic therapies to palliate pain and reduce structural progression including the DMOADs (875,876). Given the burden of OA, one hope is that precision medicine might yield opportunities for both OA prevention and improved treatments targeting OA pain and progression. Precision medicine typically leverages large amounts of data, to identify patient subgroups sharing specific relevant characteristics. The use of precision medicine in rheumatology remains limited; although in RA, several biomarkers have been identified that provide prognostic value regarding disease severity and progression (874,877). Hence, identifying specific OA phenotypes and endotypes focusing on OA patients at higher risk of progression; those more likely to benefit from a given existing treatment; and identify specific pathological processes including disease mechanisms representing specific endotypes for targeted treatment can inform both prognosis and guide therapeutic development for this prevalent disease, with the potential of positively impacting patient care (877).

By far, Dell'Isola *et al* (2016) identified six distinct phenotypes of knee OA based on review of 24 studies including chronic pain, metabolic syndrome, mechanical overload, inflammatory, bone and cartilage metabolism, and minimal joint disease phenotypes (878); subsequently, this group of researchers recognized the existence of a separate complex knee OA group that had features of more than one phenotype (872). Recently, Mobasheri *et al* (2019) expanded the number of potential phenotypes to 14 within four larger categories including age-related and systemic, intraarticular, extraarticular and secondary phenotypes (872). Mobasheri and colleagues focused their discussion on the 'synovitis-driven inflammatory phenotype' while noting that studies of anti-cytokine therapies, including IL-1 and TNF inhibitors, had been unsuccessful for both hand and knee OA (872). Additionally, Driban *et al* (2020) described a phenotype of accelerated knee OA that progresses from normal or minimal knee OA to advanced-disease within 4 years (879), and Hochberg *et al* (2016) reported a phenotype of rapidly progressive knee OA associated with JSN with or without subchondral bone destruction (880). These data suggest that while the identification of reliable phenotypes and endotypes could improve study design and outcomes, there is still much about individual endotypes and related phenotypes that remains poorly understood.

Metabolomics provides a snapshot of the entire physiology of the host and its response to the environment and genetics, which can be associated with the outcome phenotype and endotypes (881). Recently, application of metabolomics to OA research has identified several promising and potentially clinical actionable metabolic markers (480,882). Hence, I carried out a metabolomics approach to identify three distinct endotypes of OA patients in 615 primary OA patients from the well-established NFOAS (Chapter 3). Furthermore, this analysis identified several major contributing factors including C4, arginine, and a number of glycerophospholipids

to be used for the differentiation between the identified clusters. Cluster A of 66 OA patients were characterized by a significantly higher concentration of C4, but lower concentration of PC ae C40:3 than other two clusters and unaffected controls. The ratio of PC ae C40:3 to C4 had a great discriminatory power to classify OA patients in cluster A from controls. Cluster B of 200 OA patients had a significantly lower arginine concentration compared with other clusters and controls. Additionally, cluster C of 349 OA patients was distinguished by the remarkable different concentration of lysoPC a C16:0 and PC ae C38:2 from other two clusters and unaffected controls. The significant metabolite contributors to each of the three endotypes implied that the primary OA patients can be classified as muscle weakness, arginine deficient, and low inflammatory OA. These findings provide new insights into the pathogenesis of primary OA and have great potential for future development of personalized tools for OA management. However, this study used a commercially available metabolomics assay kit that has limited coverage of metabolites. Also, study participants included both knee and hip OA patients that have different aetiology. Though, we did not find a different distribution of knee and hip OA in the identified three clusters. Further studies of knee and hip OA cohorts with sufficient sample sizes are needed to confirm the findings. Lastly, our findings were detected in the OA patients from the NL population that may not be generalized to other populations.

Our endotypes of OA study revealed that muscle weakness might be responsible for a subset of OA patients. Hence, to better understand the potential mechanisms of the age-related muscle strength reduction I investigated the metabolomic signatures for the longitudinal reduction of muscle strength over 10 years in a community-based older adult cohort from Southern Tasmania, Australia (Chapter 4). The metabolites that were associated with the longitudinal reduction of

hand grip, knee extension, and leg muscle strength over 10 years were tested, and a GWAS analysis in 77 individuals from the NFOAS was performed to explore the potential mechanisms of the association between the identified metabolomic markers and the longitudinal reduction of muscle strength. The elevated level of ADMA was significantly associated with the reduction of average hand grip and knee extension strength over 10 years. GWAS analysis detected that SNP rs1125718 adjacent to the *WISP1*genes was significantly associated with increased ADMA levels. The *WISP1*genes plays a central role in the muscle stem cells regeneration following to muscle injury or pathogenesis. Additionally, we found that the increased concentration of uric acid was significantly associated with the decline of leg strength over 10 years but not with hand grip or knee extension strength. These results assumed that elevated serum levels of ADMA and uric acid increase the susceptibility of study participants to age-related muscle strength reduction and may elevate their risk to develop OA. This study was the first population-based study that investigated the relationship between the serum metabolome and the longitudinal reduction rates in hand grip, knee extension, and leg muscle strength over 10-years follow-up period in a large sample size of older adults that were randomly selected from a general population in Southern Tasmania, Australia. Despite the longitudinal nature of this study that allowed us to detect significant metabolite associations for muscle strength changes overtime within an individual, our study used a commercially available metabolomics assay kit that has a limited coverage of metabolome. Thus, we might miss some metabolites that may contribute to the longitudinal reduction of muscle strength. Further studies with multiple time points metabolomic profiling are needed. Also, our results may not be generalized to populations with different area-specific socioeconomic indexes and health provisions than that in Southern Tasmania, Australia. However, our findings provided new insights into the pathogenesis of age-related muscle

strength decline and novel targets for developing strategies to prevent muscle strength loss over time.

OA has a hereditary predisposition and many studies have sought genetic markers associated with OA with the hope of identifying high-risk groups for targeted prevention initiatives. By far, GWAS of OA have identified ~140 loci to be significantly associated with multiple OA phenotypes. However, these genetic variations can only account for a small fraction of OA heritability (334). While some of the identified genes are associated with mechanisms of action that support treatments currently in clinical trials for disease modification, the majority of the genes are implicated in traits associated with the risk of OA including higher BMD and BMI, and greater hip and waist circumference and body fat (196). Total trait variance including tendency to OA, accounted for by known genetic variants remains modest, suggesting major variants are yet undiscovered or that genotypic effects are outweighed by environmental factors. Thus, we took advantage of the NL founder population and the availability of the NFOAS and conducted an independent GWAS in 557 primary OA patients from the NFOAS to identify novel genetic variants associated with OA in the NL population that were not identified previously (Chapter 5). Our analysis identified 29 genetic variants in or adjacent to the *AGRN, PERM1*, *FAM20C*, *GPR123*, *ADAM8*, *TUBGCP2*, *MUC5B*, *TOLLIP*, *BRSK2*, *FADS2*, *TMEM132D*, *MMP17*, *CA5A*, *BANP*, *MED16*, *PLPPR3*, *AZU1*, *PRTN3*, *ELANE*, *DPF1*, *SIPA1L3*, *PPP1R14A*, *PSMD8*, *NTSR1*, *OGFR*, *COL9A3*, *DIDO1*, *GID8*, *RANBP1*, *RTN4R*, *DGCR6L*, and *TMEM191B* genes to be significantly associated with OA at the GWAS association threshold ($p<5*10^{-8}$). These genes play central roles in multiple OA pathophysiological mechanisms including fatty acid metabolism; cellular response to glucose stimulus; maintenance of protein

homeostasis by removing misfolded or damaged proteins; collagen chain trimerization; breakdown of the extracellular matrix and cartilage deterioration; inflammatory signaling; innate immune pathway; regulation of cell adhesion and migration; Abnormal bone growth and bone remodeling; panic-disorder and anxiety related behaviour; and pain mechanisms.

Among the 29 identified variants to be significantly associated with OA in the FNOAS, only rs141053204 on chr7 was reported in the recent largest GWAS-meta-analysis to be associated with OA at p=0.048 in patients underwent TKR therapy. These differences might be due to the differences of phenotype classification between our study and previous studies, possible high false positive calling and low replication rates, and the variations of functional loci and the LD patterns containing the identified SNPs through populations.

Also, one of this study drawbacks include the small sample size (n=557) for a GWAS analysis. If the identified genetic variants to be associated with OA in the NL population are experimentally validated, our findings provided new insights into the pathogenesis of OA and novel targets for developing OA drugs.

However, GWAS meta-analyses have so far identified about 140 OA genetic risk variants, the aetiology and pathogenesis of OA still not fully understood. Thus, I conducted a study to identify the possible putative genes for knee and hip OA by WES analysis using ANNOVAR variant annotation software (Chapter 6). Our analysis identified deleterious SNVs in the *IGSF3, ZNF717, PRSS1, AQP7,* and *ESRRA* genes in ≥ 80 of OA patients that have not been reported in previous OA GWAS studies. These genes act in metabolism of water-soluble vitamins and

cofactors, aquaporin-mediated transport, glucagon signaling in metabolic regulation, regulation of lipolysis in adipocytes, and the ECM degradation. While confirmation is required, these results provided new insights into better understanding of knee and hip OA pathogenesis and hold promise as druggable targets for developing OA therapies.

Since, the gene*gene interaction of the OA has not been tested previously, I also used the WES dataset of 200 OA patients from the NFOAS to perform the genome-wide case-only digenic interaction of the OA. I identified significant interactions between a common missense variant (rs56158521, c.G208A, p.D70N) in the *HLA-DRB1* gene and these aggregated rare variants in each of the *CDH19*, *SOGA1*, *MORC4*, *TMTC4*, and *ANK3* genes. Those identified genes play central roles in BMD, skeletal muscles' cells differentiation, angiogenesis, morphogenesis, cellular communication, cellular proliferation, oxidative stress, pro-inflammatory cytokines, and immune response in the articular cartilage and synovial fluid of OA patients. Thus, we suggested that the interaction of the aggregated rare variants in the *CDH19, SOGA1, MORC4, TMTC4*, and *ANK3* genes with the rs56158521 common variant in the *HLA-DRB1* gene in our study can be due to the impaired glucose homeostasis and oxidative stress, abnormal bone mineral density, and defects in the skeletal muscles differentiation that may stimulate the immunoinflammatory reactions and angiogenesis in the pathogenesis of OA in our study. While confirmation is required, these findings provided new insights into better understanding of OA pathogenesis and hold promise as druggable targets for developing OA therapies.

Despite the high prevalence and social burden of OA, it does not have a cure yet. If non-surgical treatments such as medication and physiotherapy did not relief pain and restore function, doctors

may recommend the TJR surgery for the OA patients. TJR is considered by far to be the most effective treatment for end-stage OA patients. The majority of the patients that underwent the TJR achieve joint pain reduction and function improvement following to the TJR therapy, however about 22% of them do not improve or they get worse after surgery. A number of potential non-genetic predictors for the outcome of the TJR have been investigated, but the results were either inconclusive or with very limited predictive power. Furthermore, none of the previously performed studies considered the genetic factors as a predictor for the TJR non responders. Therefor, I conducted a GWAS analysis using the MLR to identify the genetic variants associating with poor outcome of the TJR in 441 OA patients from the NFOAS (Chapter 7). Our analysis identified four chromosomal regions on chr7, 9, 18, and 21 to be significantly associated with non-responders to the TJR. We found that INDEL rs563726736 was significantly associated with the TJR pain non-responders with OR of 10.3 ($p=1.26*10^{-08}$). This INDEL is an intronic variant in *DGKB* on chr7. The second variant that was detected to be significantly associated with the TJR poor pain outcome was rs62456377 with OR of 6.71 ($p=2.79*10^{-08}$). This SNP is an intronic variant in the *VWC2* gene on chr7. Moreover, we found that minor A allele of the rs2441639 SNP was significantly associated with the TJR pain non-responders with OR of 7.53 ($p=7.07*10^{-14}$). This SNP is in intron two of the *PGM5P2* on chr9. The other significantly associated variants with TJR pain poor outcome on chr9 are adjacent to the rs2441639 SNP and included the INDEL rs149992177 (OR=5.59, $p=5.47*10^{-09}$, and SNP rs9408586 (OR=7.18, $p=4.03*10^{-10}$). These two variants are close to the *FOXD4L6* gene on chr9. Furthermore, we found that SNP rs4797006, and INDEL rs146474469 on chr18 were significantly associated with non-responders to the TJR pain (OR= 8.96, $p=7.43*10^{-11}$) in the entire OA cohort. These variants are in LD in intron one of the *MC5R* gene. Additionally, SNP

rs376101925 on chr21 was significantly associated (OR= 0.27, p= 9.96 *10^{-9}) with non-responders to the TJR pain. This variant is adjacent to the *TEKT4P2* Pseudogene. Also, the rs1907506 on chr21 was significantly associated (OR=0.25, p=1.07*10^{-8}) with non-responders to the TJR pain in the whole cohort of 441 OA patients. Although, this SNP is in a gene desert, bioinformatics analysis showed that it is located ~ 400 Kbps upstream of the *TPTE* gene on chr21. The GWAS analysis in the 315 knee OA patients revealed that only variants rs4797006 and rs146474469 on chr18 remained significantly associated with the non-responding to the TJR pain (OR= 0.15, p=2.79*10^{-11}), while none of the identified variants were associated with non-responding to TJR pain in the 126 hip OA patients.

Furthermore, our association analysis identified two loci on chr21 to be significantly associated with non-responding to the TJR function in all 441 OA samples. We found that allele G of SNP rs373135624 was significantly associated with non-responders to the TJR function (OR= 0.27, p=5.97*10^{-9}). This SNP is an intronic variant in the *TEKT4P2* Pseudogene on chr21. Also, the intergenic SNP rs28971219 (C>A) on chr21 was significantly associated (OR= 0.30, p=7.03*10^{-9}) with non-responders to the TJR function on chr21. The adjacent genes to rs28971219 include the *TPTE* gene. Remarkably, this SNP is ~330 KBs downstream of the rs1907506 that was detected to be significantly associated with non-responders to the TJR pain. The third and last variant that was found to be significantly associated with the non-responding to the TJR function was the INDEL rs113445703 that is upstream of the rs28971219 and adjacent to the *TPTE* gene on chr21. Although, these variants are in the same locus, they are not in LD (r^2<0.2). Following the association analysis in the 315 knee OA patients, only rs28971219 on chr21 remained to be significantly associated with non-responding to the TJR function (OR= 0.24, p=1.78*10^{-8}).

Whereas none of the identified variants were associated with non-responding to the TJR pain in the 126 hip OA patients.

This project was the first study that investigated the association between the genetic factors and poor TJR outcome. Our results demonstrated that genes related to immunoinflammatory reactions, pain mechanisms, cartilage degradation, and subchondral bone remodeling play a significant role in poor outcome of the TJR. However, this study has a small sample size (n=441) for a GWAS analysis, if these genetic variations are validated, they can be used as biomarkers to predict the TJR poor outcome and help physicians to decide whether to perform the TJR operation or not, educate patients about their TJR expectation, contribute to the reduction of the healthcare service resources use, and improve the quality of life of people living with arthritis.

8.2 Strengths:

I benefited from the advantage of the Newfoundland's unique population characteristics in order to identify novel disease-causing genes implicated in the OA pathogenicity. Also, I was fortunate to join the Zhai lab and benefit from the availability of the well-established NFOAS that aims to identify novel genomic, epigenomic, and metabolomic markers associated with the development and progression of OA, as well as the outcome of the TJR therapy.

Since, metabolites are the low molecular weight intermediates and downstream products of the genome, transcriptome, proteome, they provide snapshot of the entire physiology of the host and its response to the environment and genetics. As well as the association of OA with obesity, diabetes, systemic low-grade inflammation, and oxidative stress has been suggested the

involvement of metabolites in initiation and progression of OA. Accordingly, my thesis was one of the first studies that implemented genomic and metabolomic (multi-omic) approach to investigate the development and progression of OA.

Moreover, the longitudinal nature of the muscle strength reduction study allowed us to detect significant metabolite associations for muscle strength changes overtime within an individual that cannot be achieved in a cross-sectional analysis. The longitudinal muscle strength reduction also underscored the importance of the longer follow-up time with multiple time point measurements as it could minimize the effect of fluctuating variability on the measurements and provide more accurate estimate of changes over time. Also, our findings improved our understanding of the relationship between the age-related muscle strength weakness and the OA pathogenesis.

Furthermore, the GWAS analyses in my thesis were conducted using the MilorGWAS package in R that uses the MLR and includes the top 10 PCs as fixed effects to completely correct for population structure and type I error in the GWAS analysis.

Also, to the best of our knowledge, the genome-wide case-only analysis in the WES data of 200 OA patients was the first digenic analysis that investigated gene*gene interaction in the OA patients.

8.3 Limitations:

Study participants were from NL, which is a genetically/ethnically homogeneous population. Hence, our results may not be generalized to other populations. Also, giving that the NL population is a small, isolated population, assembling a large OA sample size for a GWAS analysis was not possible. Thus, the small sample size used for the GWAS analysis may increase the possibility of false positive rate inflation in our GWAS analysis. However, our association analysis was conducted using the MilorGWAS package in R that includes the top 10 PCs to correct type I error and completely correct for population structure which makes it less likely to be false positive association of the associated variants with OA.

Our study identified novel genetic variations to be potentially associated with OA that play critical role in the immune system, pro-inflammatory pathway, OA pain, cartilage degradation, subchondral bone remodeling, and skeletal muscle weakness. While confirmation is required, these findings provided new insights into the pathogenesis of OA and novel targets for developing OA drugs.

Also, the metabolic profiling of my thesis used commercially available metabolomics assay kits that offer limited coverage of metabolites. Thus, we might miss some metabolites that may contribute to the OA and muscle strength weakness. Moreover, the heterogeneity of OA contributes to the complex classification of the disease and may lead to bias in OA analysis. Likewise, the study participants included both knee and hip OA patients that share a number of risk factors, but the aetiology for knee and hip OA might be different. However, we did not find a different distribution of knee and hip OA throughout our studies. Other issue considers the

prevalence of OA, particularly knee OA that was different between men and women. Nonetheless, we did not find a significant sex difference in any of our studies. furthermore, all OA patients from the NFOAS are end stage patients who were underwent the TJR surgery, and the results might not represent the earlier stages of the disease.

Additionally, rare variants association analyses in complex diseases using sequencing data have several limitations including how to account for population structure, control for unbalanced case-control designs, and model different data types in a statistical test to address the challenges of inflated type I error and increase the statistical power of the rare variants test, however the case-only approach used in my thesis is based on the aggregation of rare variants within a gene as the unit of analysis to address the lack of power inherent to studies of rare variants, reduce the number of tests required relative to testing at allele level, and potentially decrease the amount of computer time required. Moreover, case-only method is more powerful than the corresponding case-control test for detecting digenic interactions in various population stratification scenarios when common variants are tested for interaction in the context of the GWAS, and case-only test is a powerful and timely tool for detecting digenic interaction in the WES data from patients. Thus, the only known limitation of this test is that it assumes independence in the general population of the variants tested.

8.4 Conclusion:

Our metabolomic approach analysis demonstrated that at least three distinct endotypes existed in primary OA, suggesting muscle weakness, arginine deficient, and low inflammatory OA subtypes that can be distinguished by specific blood metabolic markers. Moreover, the

longitudinal reduction of muscle strength analysis identified that baseline elevated serum concentrations of ADMA and uric acid were significantly associated with age-related muscle strength reduction. Also, our GWAS analysis demonstrated that variants harbouring genes that act in metabolism; collagen chain trimerization; cartilage deterioration; inflammatory signaling; innate immune pathway; regulation of cell adhesion and migration; abnormal bone growth and bone remodeling; panic-disorder and anxiety related behaviour; and pain mechanisms are associated with OA in the NL population. Likewise, our results demonstrated that genes related to immunoinflammatory reactions, pain mechanisms, and disability play a significant role in poor outcome of the TJR in the 441 OA patients from the NFOAS. Further variant annotation analysis using the WES data identified genes acting in metabolism of water-soluble vitamins and cofactors, aquaporin-mediated transport, glucagon signaling in metabolic regulation, regulation of lipolysis in adipocytes, and the extracellular matrix degradation. Finally, our genome-wide case-only analysis for detecting digenic interaction of OA suggested that the impaired glucose homeostasis and oxidative stress, abnormal bone mineral density, and defects in the skeletal muscles' differentiation may stimulate the immunoinflammatory reactions and angiogenesis in the pathogenesis of OA. While confirmation is required, these findings provide new insights into the understanding of OA pathogenesis and hold promise in developing personalized tools for OA management toward reduction of economic burden and better quality of life for OA patients.

8.5 Future directions:

Our analysis in this thesis is a step toward future investigations in the field of genomics and metabolomic of OA. In the endotypes of primary OA project, it will be beneficial to replicate the

analysis using a comprehensive metabolic profiling assay kit. Hence, we get all possible metabolites that may contribute to the endotypes of OA.

Regarding the GWAS analysis that identified significantly associated genetic variants with OA in the NL population (Chapter 4), and the GWAS that identified novel genetic variants to be significantly associated with the TJR poor outcome in the OA patients from the NFOAS (Chapter 5), it is important to replicate our GWAS analysis in larger sample size of the OA patients to increase the power of analysis. Also, our findings require an experimental validation in the whole NFOAS cohort using the allelic discrimination assays (TaqMan SNP genotyping) and replicate the GWAS analysis in the Longitudinal Evaluation in the Arthritis Program, Osteoarthritis Study (LEAP OA) cohort from Toronto Western Hospital, University Health Network (UHN) University of Toronto. Since replication of GWAS analysis helps to ensure that observed genotype-phenotype association is due to a true association and not a chance finding or an artifact because of uncontrolled biases. Furthermore, the significantly associated genetic variants with OA and the TJR poor outcome require functional analyses to determine their molecular mechanisms involvement in OA.

Recent progress in genetics has enabled the creation of polygenic predictors of complex human traits, including risk for many important complex diseases. Based on the heterogeneity of OA, many genetic variants can each confer a small effect on overall risk. Thus, polygenic risk score analysis is recommended in the OA studies that may give an estimate of how likely an individual is to have OA only based on genetics, without taking environmental factors into account.

If the identified novel genetic variants to be significantly associated with the TJR poor outcome in chapter 7 are experimentally validated and replicated in the LEAP OA cohort, the least absolute shrinkage and selection operator (LASSO) regression analysis can be utilized to predict the TJR poor outcome for the end-stage OA patients beforehand the TJR therapy. LASSO regression is a hypothesis-generating approach to data analysis rather than a hypothesis-testing approach where statistical methods are used to determine the most predictive variables and build the model. LASSO is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model. It is used to identify the prediction of an outcome and finding a parsimonious subset of variables that are associated with that outcome. In our analysis, LASSO regression can be used to select the subset of genetic markers with greatest discriminatory ability to minimize the prediction error and to examine the collective predictive power of all genetic and potential non-genetic factors, which include age, sex, BMI, comorbidities including diabetes, cardiovascular diseases, cancer, as well as blood lipid profiles, the patient's general health, physical, emotional and social function, motivation, self-efficacy, type of prosthesis, preoperative pain and function scores. Joint (knee and hip separately) and sex specificity will also be examined by the LASSO regression method with data on knee and hip joint as well as men and women separately. Finally, all the significant genetic markers and non-genetic factors selected by LASSO regression, in addition to the metabolic markers we recently found will be used in the logistic regression modeling to create a statistical prediction model for clinical utility and application. Also, further studies of knee and hip OA cohorts with sufficient sample sizes are needed to confirm these findings.

REFERENCES:

1. Paul C. The Burden of Musculoskeletal Conditions at the Start of the New Millennium. Report of a WHO Scientific Group. Geneva: WHO Technical Report Series, 919, 2003, pp. 218. ISBN: 92-4-120919-4. International Journal of Epidemiology. 2004 Aug 27;34(1):228–9.

2. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis. 2014 Jul;73(7):1323–30.

3. Cicuttini FM, Wluka AE. Is OA a mechanical or systemic disease? Nat Rev Rheumatol. 2014 Sep;10(9):515–6.

4. Zhang Y, Jordan JM. Epidemiology of Osteoarthritis. Clinics in Geriatric Medicine. 2010 Aug;26(3):355–69.

5. Prieto-Alhambra D, Judge A, Javaid MK, Cooper C, Diez-Perez A, Arden NK. Incidence and risk factors for clinically diagnosed knee, hip and hand osteoarthritis: influences of age, gender and osteoarthritis affecting other joints. Ann Rheum Dis. 2014 Sep;73(9):1659–64.

6. Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. Bull World Health Organ. 2003;81(9):646–56.

7. Glyn-Jones S, Palmer AJR, Agricola R, Price AJ, Vincent TL, Weinans H, et al. Osteoarthritis. The Lancet. 2015 Jul;386(9991):376–87.

8. Karsdal MA, Christiansen C, Ladel C, Henriksen K, Kraus VB, Bay-Jensen AC. Osteoarthritis – a case for personalized health care? Osteoarthritis and Cartilage. 2014 Jan;22(1):7–16.

9. Loeser RF. Age-Related Changes in the Musculoskeletal System and the Development of Osteoarthritis. Clinics in Geriatric Medicine. 2010 Aug;26(3):371–86.

10. Kraus VB, Blanco FJ, Englund M, Karsdal MA, Lohmander LS. Call for standardized definitions of osteoarthritis and risk stratification for clinical trials and clinical use. Osteoarthritis and Cartilage. 2015 Aug;23(8):1233–41.

11. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington, D.C.: National Academies Press; 2011.

12. Felson DT, Zhang Y. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. Arthritis & Rheumatism. 1998 Aug;41(8):1343–55.

13. Krishnan Y, Grodzinsky AJ. Cartilage diseases. Matrix Biology. 2018 Oct;71–72:51–69.

14. Donahue SW. Krogh's principle for musculoskeletal physiology and pathology. J Musculoskelet Neuronal Interact. 2018 Sep 1;18(3):284–91.

15. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: Classification of osteoarthritis of the knee. Arthritis & Rheumatism. 1986 Aug;29(8):1039–49.

16. Hutton CW. Osteoarthritis: the cause not result of joint failure? Annals of the Rheumatic Diseases. 1989 Nov 1;48(11):958–61.

17. Cubukcu D, Sarsan A, Alkan H. Relationships between Pain, Function and Radiographic Findings in Osteoarthritis of the Knee: A Cross-Sectional Study. Arthritis. 2012 Nov 19;2012:1–5.

18. Hochberg MC, Chang RW, Dwosh I, Lindsey S, Pincus T, Wolfe F. The american college of rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis: The American college of rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. Arthritis & Rheumatism. 1992 May;35(5):498–502.

19. Kellgren JH, Lawrence JS. Radiological Assessment of Osteo-Arthrosis. Annals of the Rheumatic Diseases. 1957 Dec 1;16(4):494–502.

20. A Treatise on Rheumatic Gout, or Chronic Rheumatic Arthritis of All the Joints. Br Foreign Med Chir Rev. 1858 Jul;22(43):41–6.

21. Cuperus N, Vliet Vlieland TPM, Mahler EAM, Kersten CC, Hoogeboom TJ, van den Ende CHM. The clinical burden of generalized osteoarthritis represented by self-reported health-related quality of life and activity limitations: a cross-sectional study. Rheumatol Int. 2015 May;35(5):871–7.

22. Bastick AN, Belo JN, Runhaar J, Bierma-Zeinstra SMA. What Are the Prognostic Factors for Radiographic Progression of Knee Osteoarthritis? A Meta-analysis. Clinical Orthopaedics & Related Research. 2015 Sep;473(9):2969–89.

23. Nelson AE, Smith MW, Golightly YM, Jordan JM. “Generalized osteoarthritis”: a systematic review. Semin Arthritis Rheum. 2014 Jun;43(6):713–20.

24. Silawal, Sandeep, et al. "Osteoarthritis and the complement cascade." *Clinical Medicine Insights:* Arthritis and Musculoskeletal Disorders 11 (2018): 1179544117751430.

25. Poole AR. Osteoarthritis as a Whole Joint Disease. HSS Jrnl. 2012 Feb;8(1):4–6.

26. Archer CW, Dowthwaite GP, Francis-West P. Development of synovial joints. Birth Defects Research Part C: Embryo Today: Reviews. 2003;69(2):144–55.

27. Khan IM, Redman SN, Williams R, Dowthwaite GP, Oldfield SF, Archer CW. The Development of Synovial Joints. In: Current Topics in Developmental Biology. Academic Press; 2007.

28. Kheir E, Shaw D. Hyaline articular cartilage. Orthopaedics and Trauma. 2009 Dec 1;23(6):450–5.

29. Muir H. The chondrocyte, architect of cartilage. Biomechanics, structure, function and molecular biology of cartilage matrix macromolecules. Bioessays. 1995 Dec;17(12):1039–48.

30. Buchanan WW, Kean WF. Osteoarthritis II: Pathology and pathogenesis. Inflammopharmacology. 2002 Feb;10(1–2):23–52.

31. Buckwalter JA. Articular Cartilage: Injuries and Potential for Healing. J Orthop Sports Phys Ther. 1998 Oct;28(4):192–202.

32. Heijink A, Gomoll AH, Madry H, Drobnič M, Filardo G, Espregueira-Mendes J, et al. Biomechanical considerations in the pathogenesis of osteoarthritis of the knee. Knee Surg Sports Traumatol Arthrosc. 2012 Mar;20(3):423–35.

33. Guerne PA, Sublet A, Lotz M. Growth factor responsiveness of human articular chondrocytes: Distinct profiles in primary chondrocytes, subcultured chondrocytes, and fibroblasts. J Cell Physiol. 1994 Mar;158(3):476–84.

34. Hayes DW, Brower RL, John KJ. Articular cartilage. Anatomy, injury, and repair. Clin Podiatr Med Surg. 2001 Jan;18(1):35–53.

35. Knecht S, Vanwanseele B, Stüssi E. A review on the mechanical quality of articular cartilage – Implications for the diagnosis of osteoarthritis. Clinical Biomechanics. 2006 Dec;21(10):999–1012.

36. Adams MA. The mechanical environment of chondrocytes in articular cartilage. Biorheology. 2006;43(3,4):537–45.

37. Kempson G, Muir H, Swanson S, Freeman M. Correlations between stiffness and the chemical constituents of cartilage on the human femoral head. Biochimica et Biophysica Acta (BBA) - General Subjects. 1970 Jul 21;215(1):70–7.

38. Jurvelin J, Säämänen AM, Arokoski J, Helminen HJ, Kiviranta I, Tammi M. Biomechanical Properties of the Canine Knee Articular Cartilage as Related to Matrix Proteoglycans and Collagen. Engineering in Medicine. 1988 Oct;17(4):157–62.

39. Juneja P, Munjal A, Hubbard JB. Anatomy, Joints. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2022.

40. Madry H, van Dijk CN, Mueller-Gerbl M. The basic science of the subchondral bone. Knee Surg Sports Traumatol Arthrosc. 2010 Apr;18(4):419–33.

41. Burr DB, Gallant MA. Bone remodelling in osteoarthritis. Nat Rev Rheumatol. 2012 Nov;8(11):665–73.

42. Goldring MB, Goldring SR. Articular cartilage and subchondral bone in the pathogenesis of osteoarthritis: Articular cartilage and subchondral bone. Annals of the New York Academy of Sciences. 2010 Apr;1192(1):230–7.

43. Holmdahl DE, Ingelmark BE. The contact between the articular cartilage and the medullary cavities of the bones. Cells Tissues Organs. 1951;12(4):341–9.

44. Castañeda S, Roman-Blas JA, Largo R, Herrero-Beaumont G. Subchondral bone as a key target for osteoarthritis treatment. Biochemical Pharmacology. 2012 Feb;83(3):315–23.

45. Suri S, Walsh DA. Osteochondral alterations in osteoarthritis. Bone. 2012 Aug;51(2):204–11.

46. Barland P, Novikoff AB, Hamerman D. electron microscopy of the human synovial membrane. Journal of Cell Biology. 1962 Aug 1;14(2):207–20.

47. Wang M, Peng Z, Vasilev K, Ketheesan N. Investigation of Wear Particles Generated in Human Knee Joints Using Atomic Force Microscopy. Tribol Lett. 2013 Jul;51(1):161–70.

48. Stannus O, Jones G, Cicuttini F, Parameswaran V, Quinn S, Burgess J, et al. Circulating levels of IL-6 and TNF-α are associated with knee radiographic osteoarthritis and knee cartilage loss in older adults. Osteoarthritis and Cartilage. 2010 Nov;18(11):1441–7.

49. Xia B, Di Chen, Zhang J, Hu S, Jin H, Tong P. Osteoarthritis Pathogenesis: A Review of Molecular Mechanisms. Calcif Tissue Int. 2014 Dec;95(6):495–505.

50. Buckwalter JA, Mankin HJ, Grodzinsky AJ. Articular cartilage and osteoarthritis. Instr Course Lect. 2005;54:465–80.

51. Sulzbacher I. Osteoarthritis: histology and pathogenesis. Wien Med Wochenschr. 2013 May;163(9–10):212–9.

52. Goldring MB, Goldring SR. Osteoarthritis. J Cell Physiol. 2007 Dec;213(3):626–34.

53. Zhang P, Li K, Kamali A, Ziadlou R, Ahmad P, Wang X, et al. Small molecules of herbal origin for osteoarthritis treatment: in vitro and in vivo evidence. Arthritis Res Ther. 2022 Dec;24(1):105.

54. Tanamas SK, Wluka AE, Pelletier JP, Martel-Pelletier J, Abram F, Wang Y, et al. The association between subchondral bone cysts and tibial cartilage volume and risk of joint replacement in people with knee osteoarthritis: a longitudinal study. Arthritis Res Ther. 2010;12(2):R58.

55. Kapoor M, Mahomed NN, editors. Osteoarthritis. Cham: Springer International Publishing; 2015.

56. Man GS, Mologhianu G. Osteoarthritis pathogenesis - a complex process that involves the entire joint. J Med Life. 2014 Mar 15;7(1):37–41.

57. Evans CH, Mears DC, Mcknight JL. A preliminary ferrographic survey of the wear particles in human synovial fluid. Arthritis & Rheumatism. 1981 Jul;24(7):912–8.

58. Loeuille D, Chary-Valckenaere I, Champigneulle J, Rat AC, Toussaint F, Pinzano-Watrin A, et al. Macroscopic and microscopic features of synovial membrane inflammation in the osteoarthritic knee: Correlating magnetic resonance imaging findings with disease severity. Arthritis Rheum. 2005 Nov;52(11):3492–501.

59. Yuan GH, Tanaka M, Masuko-Hongo K, Shibakawa A, Kato T, Nishioka K, et al. Characterization of cells from pannus-like tissue over articular cartilage of advanced osteoarthritis. Osteoarthritis and Cartilage. 2004 Jan;12(1):38–45.

60. Ashraf S, Walsh DA. Angiogenesis in osteoarthritis. Current Opinion in Rheumatology. 2008 Sep;20(5):573–80.

61. Martel-Pelletier J. Pathophysiology of osteoarthritis. Osteoarthritis and Cartilage. 2004;12:31–3.

62. Arden E, Arden N, Hunter D. Osteoarthritis. 6th ed. Oxford ; New York: Oxford University Press; 2008. 199 p. (Facts).

63. Klippel JH, Stone JH, Crofford LJ, White PH. Pocket Primer on the Rheumatic Diseases. 2nd ed. London: Springer London Springer e-books; 2010.

64. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988 Dec;15(12):1833–40.

65. Wright AA, Cook C, Abbott JH. Variables associated with the progression of hip osteoarthritis: A systematic review. Arthritis Rheum. 2009 Jul 15;61(7):925–36.

66. Johnston SA. Osteoarthritis. Veterinary Clinics of North America: Small Animal Practice. 1997 Jul;27(4):699–723.

67. Cooper C, Egger P, Coggon D, Hart DJ, Masud T, Cicuttini F, et al. Generalized osteoarthritis in women: pattern of joint involvement and approaches to definition for epidemiological studies. J Rheumatol. 1996 Nov;23(11):1938–42.

68. Sharma L. Osteoarthritis of the Knee. N Engl J Med. 2021 Jan 7;384(1):51–9.

69. Katz JN, Arant KR, Loeser RF. Diagnosis and Treatment of Hip and Knee Osteoarthritis: A Review. JAMA. 2021 Feb 9;325(6):568–78.

70. Punzi L, Frigato M, Frallonardo P, Ramonda R. Inflammatory osteoarthritis of the hand. Best Practice & Research Clinical Rheumatology. 2010 Jun;24(3):301–12.

71. Altman R, Alarcón G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum. 1991 May;34(5):505–14.

72. Sinusas K. Osteoarthritis: diagnosis and treatment. Am Fam Physician. 2012 Jan 1;85(1):49–56.

73. Abhishek A, Doherty M. Diagnosis and Clinical Presentation of Osteoarthritis. Rheumatic Disease Clinics of North America. 2013 Feb;39(1):45–66.

74. Wang X, Oo WM, Linklater JM. What is the role of imaging in the clinical diagnosis of osteoarthritis and disease management? Rheumatology. 2018 May 1;57(suppl_4):iv51–60.

75. Chinese Orthopaedic Association. Diagnosis and treatment of osteoarthritis. Orthop Surg. 2010 Feb;2(1):1–6.

76. Fukui N, Yamane S, Ishida S, Tanaka K, Masuda R, Tanaka N, et al. Relationship between radiographic changes and symptoms or physical examination findings in subjects with symptomatic medial knee osteoarthritis: a three-year prospective study. BMC Musculoskelet Disord. 2010 Dec;11(1):269.

77. Oo WM, Bo MT. Role of Ultrasonography in Knee Osteoarthritis. J Clin Rheumatol. 2016 Sep;22(6):324–9.

78. Kohn MD, Sassoon AA, Fernando ND. Classifications in Brief: Kellgren-Lawrence Classification of Osteoarthritis. Clin Orthop Relat Res. 2016 Aug;474(8):1886–93.

79. Altman RD, Hochberg M, Murphy WA, Wolfe F, Lequesne M. Atlas of individual radiographic features in osteoarthritis. Osteoarthritis Cartilage. 1995 Sep;3 Suppl A:3–70.

80. Bany Muhammad M, Yeasin M. Interpretable and parameter optimized ensemble model for knee osteoarthritis assessment using radiographs. Sci Rep. 2021 Jul 12;11(1):14348.

81. Rousseau JC, Garnero P. Biological markers in osteoarthritis. Bone. 2012 Aug;51(2):265–77.

82. Hoaglund FT, Maale G. Diagnosis of osteoarthritis: the value of laboratory tests. Orthop Clin North Am. 1979 Apr;10(2):299–305.

83. Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part II. Arthritis Rheum. 2008 Jan;58(1):26–35.

84. Guccione AA, Felson DT, Anderson JJ, Anthony JM, Zhang Y, Wilson PW, et al. The effects of specific medical conditions on the functional limitations of elders in the Framingham Study. Am J Public Health. 1994 Mar;84(3):351–8.

85. Allen KD, Thoma LM, Golightly YM. Epidemiology of osteoarthritis. Osteoarthritis and Cartilage. 2022 Feb;30(2):184–95.

86. Quicke JG, Conaghan PG, Corp N, Peat G. Osteoarthritis year in review 2021: epidemiology & therapy. Osteoarthritis and Cartilage. 2022 Feb;30(2):196–206.

87. Jin Z, Wang D, Zhang H, Liang J, Feng X, Zhao J, et al. Incidence trend of five common musculoskeletal disorders from 1990 to 2017 at the global, regional and national level: results from the global burden of disease study 2017. Ann Rheum Dis. 2020 Aug;79(8):1014–22.

88. Swain S, Sarmanova A, Mallen C, Kuo CF, Coupland C, Doherty M, et al. Trends in incidence and prevalence of osteoarthritis in the United Kingdom: findings from the Clinical Practice Research Datalink (CPRD). Osteoarthritis and Cartilage. 2020 Jun;28(6):792–801.

89. Canada PHA of. Osteoarthritis in Canada. 2020. Available from: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/osteoarthritis.html

90. Johnson VL, Hunter DJ. The epidemiology of osteoarthritis. Best Practice & Research Clinical Rheumatology. 2014 Feb;28(1):5–15.

91. Neogi T, Zhang Y. Epidemiology of Osteoarthritis. Rheumatic Disease Clinics of North America. 2013 Feb;39(1):1–19.

92. Felson DT, Naimark A, Anderson J, Kazis L, Castelli W, Meenan RF. The prevalence of knee osteoarthritis in the elderly. the framingham osteoarthritis study. Arthritis & Rheumatism. 1987 Aug;30(8):914–8.

93. Jordan JM, Helmick CG, Renner JB, Luta G, Dragomir AD, Woodard J, et al. Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis Project. J Rheumatol. 2007 Jan;34(1):172–80.

94. Nguyen USDT, Zhang Y, Zhu Y, Niu J, Zhang B, Felson DT. Increasing Prevalence of Knee Pain and Symptomatic Knee Osteoarthritis: Survey and Cohort Data. Ann Intern Med. 2011 Dec 6;155(11):725.

95. Li D, Li S, Chen Q, Xie X. The Prevalence of Symptomatic Knee Osteoarthritis in Relation to Age, Sex, Area, Region, and Body Mass Index in China: A Systematic Review and Meta-Analysis. Front Med (Lausanne). 2020;7:304.

96. Driban JB, Bannuru RR, Eaton CB, Spector TD, Hart DJ, McAlindon TE, et al. The incidence and characteristics of accelerated knee osteoarthritis among women: the Chingford cohort. BMC Musculoskelet Disord. 2020 Jan 31;21(1):60.

97. Losina E, Weinstein AM, Reichmann WM, Burbine SA, Solomon DH, Daigle ME, et al. Lifetime Risk and Age at Diagnosis of Symptomatic Knee Osteoarthritis in the US: Incidence and Lifetime Risk of Diagnosed Symptomatic Knee OA. Arthritis Care Res. 2013 May;65(5):703–11.

98. Murphy L, Schwartz TA, Helmick CG, Renner JB, Tudor G, Koch G, et al. Lifetime risk of symptomatic knee osteoarthritis. Arthritis Rheum. 2008 Sep 15;59(9):1207–13.

99. Hong JW, Noh JH, Kim DJ. The prevalence of and demographic factors associated with radiographic knee osteoarthritis in Korean adults aged ≥ 50 years: The 2010–2013 Korea National Health and Nutrition Examination Survey. Kwon YD, editor. PLoS ONE. 2020 Mar 20;15(3):e0230613.

100. Jordan JM, Helmick CG, Renner JB, Luta G, Dragomir AD, Woodard J, et al. Prevalence of Hip Symptoms and Radiographic and Symptomatic Hip Osteoarthritis in African Americans and Caucasians: The Johnston County Osteoarthritis Project. J Rheumatol. 2009 Apr;36(4):809–15.

101. Kim C, Linsenmeyer KD, Vlad SC, Guermazi A, Clancy MM, Niu J, et al. Prevalence of Radiographic and Symptomatic Hip Osteoarthritis in an Urban United States Community: The Framingham Osteoarthritis Study: Urban US Prevalence of Radiographic and Symptomatic Hip OA. Arthritis & Rheumatology. 2014 Nov;66(11):3013–7.

102. Iidaka T, Muraki S, Oka H, Horii C, Kawaguchi H, Nakamura K, et al. Incidence rate and risk factors for radiographic hip osteoarthritis in Japanese men and women: a 10-year follow-up of the ROAD study. Osteoarthritis and Cartilage. 2020 Feb;28(2):182–8.

103. Oliveria SA, Felson DT, Reed JI, Cirillo PA, Walker AM. Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintenance organization. Arthritis & Rheumatism. 1995 Aug;38(8):1134–41.

104. Peat G, Rathod-Mistry T, Paskins Z, Marshall M, Thomas MJ, Menz HB, et al. Relative prevalence and distribution of knee, hand and foot symptomatic osteoarthritis subtypes in an English population. Musculoskeletal Care. 2020 Jun;18(2):219–24.

105. Haugen IK, Englund M, Aliabadi P, Niu J, Clancy M, Kvien TK, et al. Prevalence, incidence and progression of hand osteoarthritis in the general population: the Framingham Osteoarthritis Study. Annals of the Rheumatic Diseases. 2011 Sep 1;70(9):1581–6.

106. Zhang Y. Prevalence of Symptomatic Hand Osteoarthritis and Its Impact on Functional Status among the Elderly: The Framingham Study. American Journal of Epidemiology. 2002 Dec 1;156(11):1021–7.

107. Snyder EA, Alvarez C, Golightly YM, Renner JB, Jordan JM, Nelson AE. Incidence and progression of hand osteoarthritis in a large community-based cohort: the Johnston County Osteoarthritis Project. Osteoarthritis and Cartilage. 2020 Apr;28(4):446–52.

108. Qin J, Barbour KE, Murphy LB, Nelson AE, Schwartz TA, Helmick CG, et al. Lifetime Risk of Symptomatic Hand Osteoarthritis: The Johnston County Osteoarthritis Project. Arthritis & Rheumatology. 2017 Jun;69(6):1204–12.

109. Dillon CF, Hirsch R, Rasch EK, Gu Q. Symptomatic Hand Osteoarthritis in the United States: Prevalence and Functional Impairment Estimates from the Third U.S. National Health and Nutrition Examination Survey, 1991–1994. American Journal of Physical Medicine & Rehabilitation. 2007 Jan;86(1):12–21.

110. Callahan LF, Cleveland RJ, Allen KD, Golightly Y. Racial/Ethnic, Socioeconomic, and Geographic Disparities in the Epidemiology of Knee and Hip Osteoarthritis. Rheumatic Disease Clinics of North America. 2021 Feb;47(1):1–20.

111. Gellhorn AC, Katz JN, Suri P. Osteoarthritis of the spine: the facet joints. Nat Rev Rheumatol. 2013 Apr;9(4):216–24.

112. WHO Scientific Group on Rheumatic Diseases, editor. Rheumatic diseases: report of a WHO scientific group. Geneva: World Health Organization; 1992. 59 p. (WHO technical report series).

113. Hunter DJ, Schofield D, Callander E. The individual and socioeconomic impact of osteoarthritis. Nat Rev Rheumatol. 2014 Jul;10(7):437–41.

114. Palazzo C, Nguyen C, Lefevre-Colau MM, Rannou F, Poiraudeau S. Risk factors and burden of osteoarthritis. Annals of Physical and Rehabilitation Medicine. 2016 Jun;59(3):134–8.

115. Nuesch E, Dieppe P, Reichenbach S, Williams S, Iff S, Juni P. All cause and disease specific mortality in patients with knee or hip osteoarthritis: population based cohort study. BMJ. 2011 Mar 8;342(mar08 2):d1165–d1165.

116. Hawker GA, Croxford R, Bierman AS, Harvey PJ, Ravi B, Stanaitis I, et al. All-Cause Mortality and Serious Cardiovascular Events in People with Hip and Knee Osteoarthritis: A Population Based Cohort Study. Sim RB, editor. PLoS ONE. 2014 Mar 7;9(3):e91286.

117. Barbour KE, Lui LY, Nevitt MC, Murphy LB, Helmick CG, Theis KA, et al. Hip Osteoarthritis and the Risk of All-Cause and Disease-Specific Mortality in Older Women: A Population-Based Cohort Study: HIP OA AND Mortality in Older Women. Arthritis & Rheumatology. 2015 Jul;67(7):1798–805.

118. Postler A, Luque Ramos A, Goronzy J, Günther KP, Lange T, Schmitt J, et al. Prevalence and treatment of hip and knee osteoarthritis in people aged 60 years or older in Germany: an analysis based on health insurance claims data. CIA. 2018 Nov;Volume 13:2339–49.

119. Deshpande BR, Katz JN, Solomon DH, Yelin EH, Hunter DJ, Messier SP, et al. Number of Persons With Symptomatic Knee Osteoarthritis in the US: Impact of Race and Ethnicity, Age, Sex, and Obesity: Symptomatic Knee OA in the US. Arthritis Care & Research. 2016 Dec;68(12):1743–50.

120. Dieppe, Paul A., and L. Stefan Lohmander. "Pathogenesis and management of pain in osteoarthritis." *The Lancet* 365.9463 (2005): 965-973.

121. Sharif B, Kopec J, Bansback N, Rahman MM, Flanagan WM, Wong H, et al. Projecting the direct cost burden of osteoarthritis in Canada using a microsimulation model. Osteoarthritis and Cartilage. 2015 Oct;23(10):1654–63.

122. Bitton R. The economic burden of osteoarthritis. Am J Manag Care. 2009 Sep;15(8 Suppl):S230-235.

123. Aspden RM. Osteoarthritis: a problem of growth not decay? Rheumatology. 2008 Aug 13;47(10):1452–60.

124. Silverwood V, Blagojevic-Bucknall M, Jinks C, Jordan JL, Protheroe J, Jordan KP. Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. Osteoarthritis and Cartilage. 2015 Apr;23(4):507–15.

125. Felson DT, Zhang Y, Hannan MT, Naimark A, Weissman BN, Aliabadi P, et al. The incidence and natural history of knee osteoarthritis in the elderly, the framingham osteoarthritis study. Arthritis & Rheumatism. 1995 Oct;38(10):1500–5.

126. Vina ER, Kwoh CK. Epidemiology of osteoarthritis: literature update. Current Opinion in Rheumatology. 2018 Mar;30(2):160–7.

127. Nelson AE. Osteoarthritis year in review 2017: clinical. Osteoarthritis and Cartilage. 2018 Mar;26(3):319–25.

128. Litwic A, Edwards MH, Dennison EM, Cooper C. Epidemiology and burden of osteoarthritis. British Medical Bulletin. 2013 Mar 1;105(1):185–99.

129. Srikanth VK, Fryer JL, Zhai G, Winzenberg TM, Hosmer D, Jones G. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. Osteoarthritis and Cartilage. 2005 Sep;13(9):769–81.

130. Allen KD, Golightly YM. State of the evidence. Curr Opin Rheumatol. 2015 May;27(3):276–83.

131. Kalichman L, Li L, Kim DH, Guermazi A, Berkin V, O'Donnell CJ, et al. Facet joint osteoarthritis and low back pain in the community-based population. Spine (Phila Pa 1976). 2008 Nov 1;33(23):2560–5.

132. Maillefert JF. Sex differences in hip osteoarthritis: results of a longitudinal study in 508 patients. Annals of the Rheumatic Diseases. 2003 Oct 1;62(10):931–4.

133. Spector TD, Nandra D, Hart DJ, Doyle DV. Is hormone replacement therapy protective for hand and knee osteoarthritis in women?: The Chingford study. Annals of the Rheumatic Diseases. 1997 Jul 1;56(7):432–4.

134. Cirillo DJ, Wallace RB, Wu L, Yood RA. Effect of hormone therapy on risk of hip and knee joint replacement in the women's health initiative. Arthritis Rheum. 2006 Oct;54(10):3194–204.

135. de Klerk BM, Schiphof D, Groeneveld FPMJ, Koes BW, van Osch GJVM, van Meurs JBJ, et al. No clear association between female hormonal aspects and osteoarthritis of the hand, hip and knee: a systematic review. Rheumatology. 2009 Sep;48(9):1160–5.

136. Warner SC, Valdes AM. Genetic association studies in osteoarthritis: is it fairytale? Curr Opin Rheumatol. 2017 Jan;29(1):103–9.

137. Guilak F, Fermor B, Keefe FJ, Kraus VB, Olson SA, Pisetsky DS, et al. The Role of Biomechanics and Inflammation in Cartilage Injury and Repair. Clinical Orthopaedics & Related Research. 2004 Jun;423:17–26.

138. Pottie P, Presle N, Terlain B, Netter P, Mainard D, Berenbaum F. Obesity and osteoarthritis: more complex than predicted! Annals of the Rheumatic Diseases. 2006 Nov 1;65(11):1403–5.

139. Thakur M, Dawes JM, McMahon SB. Genomics of pain in osteoarthritis. Osteoarthritis Cartilage. 2013 Sep;21(9):1374–82.

140. Wilkinson JM, Zeggini E. The Genetic Epidemiology of Joint Shape and the Development of Osteoarthritis. Calcif Tissue Int. 2021 Sep;109(3):257–76.

141. Evangelou E, Valdes AM, Kerkhof HJM, Styrkarsdottir U, Zhu Y, Meulenbelt I, et al. Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. Ann Rheum Dis. 2011 Feb;70(2):349–55.

142. Baird DA, Evans DS, Kamanu FK, Gregory JS, Saunders FR, Giuraniuc CV, et al. Identification of Novel Loci Associated With Hip Shape: A Meta-Analysis of Genomewide Association Studies. J Bone Miner Res. 2019 Feb;34(2):241–51.

143. Richette P, Poitou C, Garnero P, Vicaut E, Bouillot JL, Lacorte JM, et al. Benefits of massive weight loss on symptoms, systemic inflammation and cartilage turnover in obese patients with knee osteoarthritis. Annals of the Rheumatic Diseases. 2011 Jan 1;70(1):139–44.

144. Christensen R, Bartels EM, Astrup A, Bliddal H. Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: a systematic review and meta-analysis. Annals of the Rheumatic Diseases. 2006 Sep 19;66(4):433–9.

145. Hong JQ, Wang YX, Li SH, Jiang GY, Hu B, Yang YT, et al. Association between SMAD3 gene polymorphisms and osteoarthritis risk: a systematic review and meta-analysis. J Orthop Surg Res. 2018 Sep 12;13(1):232.

146. Huo JZ, Ji XH, Su ZY, Shang P, Gao F. Association of ADAMTS4 and ADAMTS5 polymorphisms with musculoskeletal degenerative diseases: a systematic review and meta-analysis. Biosci Rep. 2018 Dec 21;38(6):BSR20181619.

147. Grotle M, Hagen KB, Natvig B, Dahl FA, Kvien TK. Obesity and osteoarthritis in knee, hip and/or hand: An epidemiological study in the general population with 10 years follow-up. BMC Musculoskelet Disord. 2008 Dec;9(1):132.

148. Kluzek S, Newton JL, Arden NK. Is osteoarthritis a metabolic disorder? Br Med Bull. 2015 Sep;115(1):111–21.

149. Hart DJ, Doyle DV, Spector TD. Association between metabolic factors and knee osteoarthritis in women: the Chingford Study. J Rheumatol. 1995 Jun;22(6):1118–23.

150. Abella V, Scotece M, Conde J, López V, Lazzaro V, Pino J, et al. Adipokines, Metabolic Syndrome and Rheumatic Diseases. Journal of Immunology Research. 2014;2014:1–14.

151. Li B, Aspden RM. Mechanical and material properties of the subchondral bone plate from the femoral head of patients with osteoarthritis or osteoporosis. Annals of the Rheumatic Diseases. 1997 Apr 1;56(4):247–54.

152. Spector TD, MacGregor AJ. Risk factors for osteoarthritis: genetics. Osteoarthritis Cartilage. 2004;12 Suppl A:S39-44.

153. Lo IKY, Griffin S, Kirkley A. The development of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: The Western Ontario Osteoarthritis of the Shoulder (WOOS) index. Osteoarthritis and Cartilage. 2001 Nov;9(8):771–8.

154. Puenpatom RA, Victor TW. Increased Prevalence of Metabolic Syndrome in Individuals with Osteoarthritis: An Analysis of NHANES III Data. Postgraduate Medicine. 2009 Nov 1;121(6):9–20.

155. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). The Upper Extremity Collaborative Group (UECG). Am J Ind Med. 1996 Jun;29(6):602–8.

156. Guermazi A, Hayashi D, Roemer FW, Felson DT. Osteoarthritis. Rheumatic Disease Clinics of North America. 2013 Aug;39(3):567–91.

157. Lo K, Au M, Ni J, Wen C. Association between hypertension and osteoarthritis: A systematic review and meta-analysis of observational studies. J Orthop Translat. 2022 Jan;32:12–20.

158. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg. 1994 Nov;3(6):347–52.

159. Frey MI, Barrett-Connor E, Sledge PA, Schneider DL, Weisman MH. The effect of noninsulin dependent diabetes mellitus on the prevalence of clinical osteoarthritis. A population based study. J Rheumatol. 1996 Apr;23(4):716–22.

160. Louati K, Vidal C, Berenbaum F, Sellam J. Association between diabetes mellitus and osteoarthritis: systematic literature review and meta-analysis. RMD Open. 2015 Jun 2;1(1):e000077–e000077.

161. Piva SR, Susko AM, Khoja SS, Josbeno DA, Fitzgerald GK, Toledo FGS. Links Between Osteoarthritis and Diabetes. Clinics in Geriatric Medicine. 2015 Feb;31(1):67–87.

162. McDonald Blumer MH. Bone mineral content versus bone density in a population with osteoarthritis: a new twist to the controversy? J Rheumatol. 2005 Oct;32(10):1868–9.

163. Deng Z han, Zeng C, Li Y sheng, Yang T, Li H, Wei J, et al. Relation between phalangeal bone mineral density and radiographic knee osteoarthritis: a cross-sectional study. BMC Musculoskelet Disord. 2016 Dec;17(1):71.

164. Mezhov V, Ciccutini FM, Hanna FS, Brennan SL, Wang YY, Urquhart DM, et al. Does obesity affect knee cartilage? A systematic review of magnetic resonance imaging data. Obes Rev. 2014 Feb;15(2):143–57.

165. Blagojevic M, Jinks C, Jeffery A, Jordan KP. Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. Osteoarthritis and Cartilage. 2010 Jan;18(1):24–33.

166. Hardcastle SA, Dieppe P, Gregson CL, Hunter D, Thomas GER, Arden NK, et al. Prevalence of radiographic hip osteoarthritis is increased in high bone mass. Osteoarthritis Cartilage. 2014 Aug;22(8):1120–8.

167. King LK, March L, Anandacoomarasamy A. Obesity & osteoarthritis. Indian J Med Res. 2013;138:185–93.

168. Reynard LN, Loughlin J. Insights from human genetic studies into the pathways involved in osteoarthritis. Nat Rev Rheumatol. 2013 Oct;9(10):573–83.

169. Hunter DJ, Gerstenfeld L, Bishop G, Davis AD, Mason ZD, Einhorn TA, et al. Bone marrow lesions from osteoarthritis knees are characterized by sclerotic bone that is less well mineralized. Arthritis Res Ther. 2009;11(1):R11.

170. Frost HM. Wolff's Law and bone's structural adaptations to mechanical usage: an overview for clinicians. Angle Orthod. 1994;64(3):175–88.

171. Reyes C, Garcia-Gil M, Elorza JM, Mendez-Boo L, Hermosilla E, Javaid MK, et al. Socio-economic status and the risk of developing hand, hip or knee osteoarthritis: a region-wide ecological study. Osteoarthritis and Cartilage. 2015 Aug;23(8):1323–9.

172. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, Boirie Y, Cederholm T, Landi F, et al. Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. Age and Ageing. 2010 Jul 1;39(4):412–23.

173. Mobasheri A, Rayman MP, Gualillo O, Sellam J, van der Kraan P, Fearon U. The role of metabolism in the pathogenesis of osteoarthritis. Nat Rev Rheumatol. 2017 May;13(5):302–11.

174. Brandt, Kenneth D. "Putting some muscle into osteoarthritis." *Annals of Internal Medicine* 127.2 (1997): 154-156.

175. Hortobágyi T, Garry J, Holbert D, Devita P. Aberrations in the control of quadriceps muscle force in patients with knee osteoarthritis. Arthritis Care & Research. 2004;51(4):562–9.

176. Hortobágyi T, Westerkamp L, Beam S, Moody J, Garry J, Holbert D, et al. Altered hamstring-quadriceps muscle balance in patients with knee osteoarthritis. Clinical Biomechanics. 2005 Jan 1;20(1):97–104.

177. Scott D, Blizzard L, Fell J, Jones G. Prospective study of self-reported pain, radiographic osteoarthritis, sarcopenia progression, and falls risk in community-dwelling older adults. Arthritis Care Res (Hoboken). 2012 Jan;64(1):30-7. doi: 10.1002/acr.20545. PMID: 21739619.

178. Johannesdottir F, Aspelund T, Siggeirsdottir K, Jonsson BY, Mogensen B, Sigurdsson S, et al. Mid-thigh cortical bone structural parameters, muscle mass and strength, and association with lower limb fractures in older men and women (AGES-Reykjavik Study). Calcif Tissue Int. 2012 May;90(5):354–64.

179. Abellan van Kan G. Epidemiology and consequences of sarcopenia. J Nutr Health Aging. 2009 Oct;13(8):708-12. doi: 10.1007/s12603-009-0201-z. PMID: 19657554.

180. Hurley, Michael V. "The role of muscle weakness in the pathogenesis of osteoarthritis." *Rheumatic disease clinics of North America* 25.2 (1999): 283-298.

181. Brandt KD. Putting some muscle into osteoarthritis. Ann Intern Med. 1997 Jul 15;127(2):154–6.

182. Wollheim FA, Burmester GR. [Osteoarthritis concerns us all]. Z Rheumatol. 2003 Aug;62(4):406–8.

183. Hurley MV. Muscle dysfunction and effective rehabilitation of knee osteoarthritis: what we know and what we need to find out. Arthritis Rheum. 2003 Jun 15;49(3):444–52.

184. Cao Y long, Pang J, Zhan H sheng, Shi Y yu. [Clinical advances of muscle status in osteoarthritis]. Zhongguo Gu Shang. 2008 Jun;21(6):476–9.

185. Kim JR, Yoo JJ, Kim HA. Therapeutics in Osteoarthritis Based on an Understanding of Its Molecular Pathogenesis. Int J Mol Sci. 2018 Feb 27;19(3).

186. Tanamas S, Hanna FS, Cicuttini FM, Wluka AE, Berry P, Urquhart DM. Does knee malalignment increase the risk of development and progression of knee osteoarthritis? A systematic review. Arthritis Rheum. 2009 Apr 15;61(4):459–67.

187. Slemenda C, Brandt KD, Heilman DK, Mazzuca S, Braunstein EM, Katz BP, et al. Quadriceps weakness and osteoarthritis of the knee. Ann Intern Med. 1997 Jul 15;127(2):97–104.

188. Ajuied A, Wong F, Smith C, Norris M, Earnshaw P, Back D, et al. Anterior cruciate ligament injury and radiologic progression of knee osteoarthritis: a systematic review and meta-analysis. Am J Sports Med. 2014 Sep;42(9):2242–52.

189. Cerejo R, Dunlop DD, Cahue S, Channin D, Song J, Sharma L. The influence of alignment on risk of knee osteoarthritis progression according to baseline stage of disease. Arthritis Rheum. 2002 Oct;46(10):2632–6.

190. van Buuren MMA, Arden NK, Bierma-Zeinstra SMA, Bramer WM, Casartelli NC, Felson DT, et al. Statistical shape modeling of the hip and the association with hip osteoarthritis: a systematic review. Osteoarthritis Cartilage. 2021 May;29(5):607–18.

191. Slauterbeck JR, Kousa P, Clifton BC, Naud S, Tourville TW, Johnson RJ, et al. Geographic mapping of meniscus and cartilage lesions associated with anterior cruciate ligament injuries. J Bone Joint Surg Am. 2009 Sep;91(9):2094–103.

192. Øiestad BE, Engebretsen L, Storheim K, Risberg MA. Winner of the 2008 Systematic Review Competition: Knee Osteoarthritis after Anterior Cruciate Ligament Injury. Am J Sports Med. 2009 Jul;37(7):1434–43.

193. Turkiewicz A, Petersson IF, Björk J, Hawker G, Dahlberg LE, Lohmander LS, et al. Current and future impact of osteoarthritis on health care: a population-based study with projections to year 2032. Osteoarthritis Cartilage. 2014 Nov;22(11):1826–32.

194. Cooper C, Inskip H, Croft P, Campbell L, Smith G, McLaren M, et al. Individual risk factors for hip osteoarthritis: obesity, hip injury, and physical activity. Am J Epidemiol. 1998 Mar 15;147(6):516–22.

195. Juhakoski R, Heliövaara M, Impivaara O, Kröger H, Knekt P, Lauren H, et al. Risk factors for the development of hip osteoarthritis: a population-based prospective study. Rheumatology (Oxford). 2009 Jan;48(1):83–7.

196. Tachmazidou I, Hatzikotoulas K, Southam L, Esparza-Gordillo J, Haberland V, Zheng J, et al. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nat Genet. 2019 Feb;51(2):230–6.

197. Muraki S, Akune T, Oka H, Mabuchi A, En-Yo Y, Yoshida M, et al. Association of occupational activity with radiographic knee osteoarthritis and lumbar spondylosis in elderly patients of population-based cohorts: a large-scale population-based study. Arthritis Rheum. 2009 Jun 15;61(6):779–86.

198. Croft P, Coggon D, Cruddas M, Cooper C. Osteoarthritis of the hip: an occupational disease in farmers. BMJ. 1992 May 16;304(6837):1269–72.

199. Oiestad BE, Holm I, Aune AK, Gunderson R, Myklebust G, Engebretsen L, et al. Knee function and prevalence of knee osteoarthritis after anterior cruciate ligament reconstruction: a prospective study with 10 to 15 years of follow-up. Am J Sports Med. 2010 Nov;38(11):2201–10.

200. Snoeker B, Turkiewicz A, Magnusson K, Frobell R, Yu D, Peat G, et al. Risk of knee osteoarthritis after different types of knee injuries in young adults: a population-based cohort study. Br J Sports Med. 2020 Jun;54(12):725–30.

201. Lie MM, Risberg MA, Storheim K, Engebretsen L, Øiestad BE. What's the rate of knee osteoarthritis 10 years after anterior cruciate ligament injury? An updated systematic review. Br J Sports Med. 2019 Sep;53(18):1162–7.

202. Thomas AC, Hubbard-Turner T, Wikstrom EA, Palmieri-Smith RM. Epidemiology of Posttraumatic Osteoarthritis. J Athl Train. 2017 Jun 2;52(6):491–6.

203. Hadler NM, Gillings DB, Imbus HR, Levitin PM, Makuc D, Utsinger PD, et al. Hand structure and function in an industrial setting. Arthritis & Rheumatism. 1978 Mar;21(2):210–20.

204. Jiang L, Tian W, Wang Y, Rong J, Bao C, Liu Y, et al. Body mass index and susceptibility to knee osteoarthritis: a systematic review and meta-analysis. Joint Bone Spine. 2012 May;79(3):291–7.

205. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012 Dec 15;380(9859):2163–96.

206. Buckwalter JA, Lane NE. Athletics and osteoarthritis. Am J Sports Med. 1997 Dec;25(6):873–81.

207. Atukorala I, Makovey J, Lawler L, Messier SP, Bennell K, Hunter DJ. Is There a Dose-Response Relationship Between Weight Loss and Symptom Improvement in Persons With Knee Osteoarthritis? Arthritis Care Res (Hoboken). 2016 Aug;68(8):1106–14.

208. Spector TD, Harris PA, Hart DJ, Cicuttini FM, Nandra D, Etherington J, et al. Risk of osteoarthritis associated with long-term weight-bearing sports: a radiologic survey of the hips and knees in female ex-athletes and population controls. Arthritis Rheum. 1996 Jun;39(6):988–95.

209. Gersing AS, Schwaiger BJ, Nevitt MC, Joseph GB, Chanchek N, Guimaraes JB, et al. Is Weight Loss Associated with Less Progression of Changes in Knee Articular Cartilage among Obese and Overweight Patients as Assessed with MR Imaging over 48 Months? Data from the Osteoarthritis Initiative. Radiology. 2017 Aug;284(2):508–20.

210. Turkiewicz A, Kiadaliri AA, Englund M. Cause-specific mortality in osteoarthritis of peripheral joints. Osteoarthritis Cartilage. 2019 Jun;27(6):848–54.

211. Yoshimura N. [Epidemiology of osteoarthritis in Japan : the ROAD study]. Clin Calcium. 2011 Jun;21(6):821–5.

212. Ho-Pham LT, Lai TQ, Mai LD, Doan MC, Pham HN, Nguyen TV. Prevalence of radiographic osteoarthritis of the knee and its relationship to self-reported pain. PLoS One. 2014;9(4):e94563.

213. Nelson AE, Braga L, Renner JB, Atashili J, Woodard J, Hochberg MC, et al. Characterization of individual radiographic features of hip osteoarthritis in African American and White women and men: the Johnston County Osteoarthritis Project. Arthritis Care Res (Hoboken). 2010 Feb;62(2):190–7.

214. Plumb MS, Aspden RM. High levels of fat and (n-6) fatty acids in cancellous bone in osteoarthritis. Lipids Health Dis. 2004 Jun 18;3:12.

215. Masuko K, Murata M, Suematsu N, Okamoto K, Yudoh K, Nakamura H, et al. A metabolic aspect of osteoarthritis: lipid as a possible contributor to the pathogenesis of cartilage degradation. Clin Exp Rheumatol. 2009 Apr;27(2):347–53.

216. Calder PC. n-3 polyunsaturated fatty acids, inflammation, and inflammatory diseases. Am J Clin Nutr. 2006 Jun;83(6 Suppl):1505S-1519S.

217. Baker KR, Matthan NR, Lichtenstein AH, Niu J, Guermazi A, Roemer F, et al. Association of plasma n-6 and n-3 polyunsaturated fatty acids with synovitis in the knee: the MOST study. Osteoarthritis Cartilage. 2012 May;20(5):382–7.

218. Thomas S, Browne H, Mobasheri A, Rayman MP. What is the evidence for a role for diet and nutrition in osteoarthritis? Rheumatology (Oxford). 2018 May 1;57(suppl_4):iv61–74.

219. Simopoulos AP. The importance of the omega-6/omega-3 fatty acid ratio in cardiovascular disease and other chronic diseases. Exp Biol Med (Maywood). 2008 Jun;233(6):674–88.

220. Wang Y, Wluka AE, Hodge AM, English DR, Giles GG, O'Sullivan R, et al. Effect of fatty acids on bone marrow lesions and knee cartilage in healthy, middle-aged subjects without clinical knee osteoarthritis. Osteoarthritis Cartilage. 2008 May;16(5):579–83.

221. Stürmer T, Sun Y, Sauerland S, Zeissig I, Günther KP, Puhl W, et al. Serum cholesterol and osteoarthritis. The baseline examination of the Ulm Osteoarthritis Study. J Rheumatol. 1998 Sep;25(9):1827–32.

222. Tabas I. Consequences of cellular cholesterol accumulation: basic concepts and physiological implications. J Clin Invest. 2002 Oct;110(7):905–11.

223. Prasad K, Lee P. Suppression of oxidative stress as a mechanism of reduction of hypercholesterolemic atherosclerosis by aspirin. J Cardiovasc Pharmacol Ther. 2003 Mar;8(1):61–9.

224. Jenkins DJ, Kendall CW, Vuksan V. Viscous fibers, health claims, and strategies to reduce cardiovascular disease risk. Am J Clin Nutr. 2000 Feb;71(2):401–2.

225. Grover AK, Samson SE. Benefits of antioxidant supplements for knee osteoarthritis: rationale and reality. Nutr J. 2016 Jan 5;15:1.

226. Henrotin Y, Kurz B. Antioxidant to treat osteoarthritis: dream or reality? Curr Drug Targets. 2007 Feb;8(2):347–57.

227. Felson DT, Niu J, Clancy M, Aliabadi P, Sack B, Guermazi A, et al. Low levels of vitamin D and worsening of knee osteoarthritis: Results of two longitudinal studies. Arthritis Rheum. 2007 Jan;56(1):129–36.

228. Mabey T, Honsawek S. Role of Vitamin D in Osteoarthritis: Molecular, Cellular, and Clinical Perspectives. Int J Endocrinol. 2015;2015:383918.

229. Shen M, Luo Y, Niu Y, Chen L, Yuan X, Goltzman D, et al. 1,25(OH)2D deficiency induces temporomandibular joint osteoarthritis via secretion of senescence-associated inflammatory cytokines. Bone. 2013 Aug;55(2):400–9.

230. Beaudart C, Buckinx F, Rabenda V, Gillain S, Cavalier E, Slomian J, et al. The effects of vitamin D on skeletal muscle strength, muscle mass, and muscle power: a systematic review and meta-analysis of randomized controlled trials. J Clin Endocrinol Metab. 2014 Nov;99(11):4336–45.

231. Berkner KL. Vitamin K-dependent carboxylation. Vitam Horm. 2008;78:131–56.

232. Bliddal H, Leeds AR, Christensen R. Osteoarthritis, obesity and weight loss: evidence, hypotheses and horizons - a scoping review. Obes Rev. 2014 Jul;15(7):578–86.

233. Muthuri SG, Hui M, Doherty M, Zhang W. What if we prevent obesity? Risk reduction in knee osteoarthritis estimated through a meta-analysis of observational studies. Arthritis Care Res (Hoboken). 2011 Jul;63(7):982–90.

234. Hui M, Doherty M, Zhang W. Does smoking protect against osteoarthritis? Meta-analysis of observational studies. Ann Rheum Dis. 2011 Jul;70(7):1231–7.

235. Valdes AM, Spector TD. The Contribution of Genes to Osteoarthritis. Medical Clinics of North America. 2009 Jan;93(1):45–66.

236. Gullo TR, Golightly YM, Cleveland RJ, Renner JB, Callahan LF, Jordan JM, et al. Defining multiple joint osteoarthritis, its frequency and impact in a community-based cohort. Semin Arthritis Rheum. 2019 Jun;48(6):950–7.

237. Wang T, Liang Y, Li H, Li H, He Q, Xue Y, et al. Single Nucleotide Polymorphisms and Osteoarthritis: An Overview and a Meta-Analysis. Medicine (Baltimore). 2016 Feb;95(7):e2811.

238. Guo SW. Sibling Recurrence Risk Ratio as a Measure of Genetic Effect: Caveat Emptor! The American Journal of Human Genetics. 2002 Mar;70(3):818–9.

239. Charcot JM. Clinical lectures on senile and chronic diseases. New York: Arno Press; 1979. 307 p. (Aging and old age).

240. Duckworth D. A treatise on gout. 71st ed. London: Blackiston, Son and Company; 1889.

241. Stecher RM. HEBERDEN'S NODES: Heredity in Hypertrophic Arthritis of the Finger Joints. The American Journal of the Medical Sciences. 1941 Jun;201(6):801–9.

242. Kellgren JH, Lawrence JS, Bier F. Genetic Factors in Generalized Osteo-Arthrosis. Annals of the Rheumatic Diseases. 1963 Jul 1;22(4):237–55.

243. Felson DT, Couropmitree NN, Chaisson CE, Hannan MT, Zhang Y, McAlindon TE, et al. Evidence for a Mendelian gene in a segregation analysis of generalized radiographic osteoarthritis: the Framingham Study. Arthritis Rheum. 1998 Jun;41(6):1064–71.

244. Valdes AM, Spector TD. Genetic epidemiology of hip and knee osteoarthritis. Nat Rev Rheumatol. 2011 Jan;7(1):23–32.

245. Yoo K, Origitano TC. Familial cervical spondylosis. Case report. J Neurosurg. 1998 Jul;89(1):139–41.

246. Matsui H, Terahata N, Tsuji H, Hirano N, Naruse Y. Familial predisposition and clustering for juvenile lumbar disc herniation. Spine (Phila Pa 1976). 1992 Nov;17(11):1323–8.

247. Varlotta GP, Brown MD, Kelsey JL, Golden AL. Familial predisposition for herniation of a lumbar disc in patients who are less than twenty-one years old. J Bone Joint Surg Am. 1991 Jan;73(1):124–8.

248. Postacchini F, Lami R, Pugliese O. Familial predisposition to discogenic low-back pain. An epidemiologic and immunogenetic study. Spine (Phila Pa 1976). 1988 Dec;13(12):1403–6.

249. Scapinelli R. Lumbar disc herniation in eight siblings with a positive family history for disc disease. Acta Orthop Belg. 1993;59(4):371–6.

250. Sahu M, Prasuna JG. Twin Studies: A Unique Epidemiological Tool. Indian J Community Med. 2016 Sep;41(3):177–82.

251. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet. 2002 Nov;3(11):872–82.

252. Spector TD, MacGregor AJ. The St. Thomas' UK Adult Twin Registry. twin res. 2002 Oct 1;5(5):440–3.

253. McDonnell SM, Sinsheimer J, Price AJ, Carr AJ. Genetic influences in the aetiology of anteromedial osteoarthritis of the knee. J Bone Joint Surg Br. 2007 Jul;89(7):901–3.

254. Lanyon P, Muir K, Doherty S, Doherty M. Assessment of a genetic contribution to osteoarthritis of the hip: sibling study. BMJ. 2000 Nov 11;321(7270):1179–83.

255. Felson DT. Developments in the clinical understanding of osteoarthritis. Arthritis Res Ther. 2009;11(1):203.

256. Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet. 2015 May;16(5):275–84.

257. Lee YH, Rho YH, Choi SJ, Ji JD, Song GG. Osteoarthritis susceptibility loci defined by genome scan meta-analysis. Rheumatol Int. 2006 Sep;26(11):996–1000.

258. Loughlin J. Polymorphism in signal transduction is a major route through which osteoarthritis susceptibility is acting. Curr Opin Rheumatol. 2005 Sep;17(5):629–33.

259. Dawn Teare M, Barrett JH. Genetic linkage studies. Lancet. 2005 Sep 17;366(9490):1036–44.

260. Patnala R, Clements J, Batra J. Candidate gene association studies: a comprehensive guide to useful in silicotools. BMC Genet. 2013 Dec;14(1):39.

261. Zhu M, Zhao S. Candidate gene identification approach: progress and challenges. Int J Biol Sci. 2007 Oct 25;3(7):420–7.

262. Rodriguez-Fontenla C, Calaza M, Evangelou E, Valdes AM, Arden N, Blanco FJ, et al. Assessment of Osteoarthritis Candidate Genes in a Meta-Analysis of Nine Genome-Wide Association Studies. Arthritis & Rheumatology. 2014 Apr;66(4):940–9.

263. Bratus A, Aeschlimann A, Russo G, Sprott H. Candidate gene approach in genetic epidemiological studies of osteoarthritis-related pain. Pain. 2014 Feb;155(2):217–21.

264. Buxton P, Edwards C, Archer CW, Francis-West P. Growth/differentiation factor-5 (GDF-5) and skeletal development. J Bone Joint Surg Am. 2001;83-A Suppl 1(Pt 1):S23-30.

265. Edwards CJ, Francis-West PH. Bone morphogenetic proteins in the development and healing of synovial joints. Seminars in Arthritis and Rheumatism. 2001 Aug;31(1):33–42.

266. Miyamoto Y, Mabuchi A, Shi D, Kubo T, Takatori Y, Saito S, et al. A functional polymorphism in the 5′ UTR of GDF5 is associated with susceptibility to osteoarthritis. Nat Genet. 2007 Apr;39(4):529–33.

267. Sun K, Guo J, Yao X, Guo Z, Guo F. Growth differentiation factor 5 in cartilage and osteoarthritis: A possible therapeutic candidate. Cell Prolif. 2021 Mar.

268. Chapman K, Takahashi A, Meulenbelt I, Watson C, Rodriguez-Lopez J, Egli R, et al. A meta-analysis of European and Asian cohorts reveals a global role of a functional SNP in the 5' UTR of GDF5 with osteoarthritis susceptibility. Hum Mol Genet. 2008 May 15;17(10):1497–504.

269. Egli RJ, Southam L, Wilkins JM, Lorenzen I, Pombo-Suarez M, Gonzalez A, et al. Functional analysis of the osteoarthritis susceptibility-associated GDF5 regulatory polymorphism. Arthritis Rheum. 2009 Jul;60(7):2055–64.

270. Southam L, Rodriguez-Lopez J, Wilkins JM, Pombo-Suarez M, Snelling S, Gomez-Reino JJ, et al. An SNP in the 5'-UTR of GDF5 is associated with osteoarthritis susceptibility in Europeans and with in vivo differences in allelic expression in articular cartilage. Hum Mol Genet. 2007 Sep 15;16(18):2226–32.

271. Evangelou E, Chapman K, Meulenbelt I, Karassa FB, Loughlin J, Carr A, et al. Large-scale analysis of association between GDF5 and FRZB variants and osteoarthritis of the hip, knee, and hand. Arthritis Rheum. 2009 Jun;60(6):1710–21.

272. Valdes AM, Evangelou E, Kerkhof HJM, Tamm A, Doherty SA, Kisand K, et al. The GDF5 rs143383 polymorphism is associated with osteoarthritis of the knee with genome-wide statistical significance. Ann Rheum Dis. 2011 May;70(5):873–5.

273. Hao SW, Jin QH. Association between the +104T/C polymorphism in the 5'UTR of GDF5 and susceptibility to knee osteoarthritis: a meta-analysis. Mol Med Rep. 2013 Feb;7(2):485–8.

274. Valdes AM, Doherty S, Muir KR, Zhang W, Maciewicz RA, Wheeler M, et al. Genetic contribution to radiographic severity in osteoarthritis of the knee. Ann Rheum Dis. 2012 Sep;71(9):1537–40.

275. Minafra L, Bravatà V, Saporito M, Cammarata FP, Forte GI, Caldarella S, et al. Genetic, clinical and radiographic signs in knee osteoarthritis susceptibility. Arthritis Res Ther. 2014 Apr 9;16(2):R91.

276. Reynard LN, Bui C, Syddall CM, Loughlin J. CpG methylation regulates allelic expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. Hum Genet. 2014 Aug;133(8):1059–73.

277. Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, et al. Identification of ten loci associated with height highlights new biological pathways in human growth. Nat Genet. 2008 May;40(5):584–91.

278. Sanna S, Jackson AU, Nagaraja R, Willer CJ, Chen WM, Bonnycastle LL, et al. Common variants in the GDF5-UQCC region are associated with variation in human height. Nat Genet. 2008 Feb;40(2):198–203.

279. Masuya H, Nishida K, Furuichi T, Toki H, Nishimura G, Kawabata H, et al. A novel dominant-negative mutation in Gdf5 generated by ENU mutagenesis impairs joint formation and causes osteoarthritis in mice. Hum Mol Genet. 2007 Oct 1;16(19):2366–75.

280. Harada M, Takahara M, Zhe P, Otsuji M, Iuchi Y, Takagi M, et al. Developmental failure of the intra-articular ligaments in mice with absence of growth differentiation factor 5. Osteoarthritis Cartilage. 2007 Apr;15(4):468–74.

281. Daans M, Luyten FP, Lories RJU. GDF5 deficiency in mice is associated with instability-driven joint damage, gait and subchondral bone changes. Ann Rheum Dis. 2011 Jan;70(1):208–13.

282. Marian AJ. Molecular genetic studies of complex phenotypes. Transl Res. 2012 Feb;159(2):64–79.

283. Siontis KCM, Patsopoulos NA, Ioannidis JPA. Replication of past candidate loci for common diseases and phenotypes in 100 genome-wide association studies. Eur J Hum Genet. 2010 Jul;18(7):832–7.

284. Need AC, Goldstein DB. Whole genome association studies in complex diseases: where do we stand? Dialogues Clin Neurosci. 2010;12(1):37–46.

285. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. Nat Protoc. 2011 Feb;6(2):121–33.

286. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet. 2009;10:387–406.

287. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010 Jul;11(7):499–511.

288. Abraham G, Inouye M. Fast Principal Component Analysis of Large-Scale Genome-Wide Data. Zhang Y, editor. PLoS ONE. 2014 Apr 9;9(4):e93766.

289. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. Eur J Hum Genet. 2016 Aug;24(8):1202–5.

290. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014 Jan;42(Database issue):D1001-1006.

291. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. Lewitter F, Kann M, editors. PLoS Comput Biol. 2012 Dec 27;8(12):e1002822.

292. Burton PR, Hansell AL, Fortier I, Manolio TA, Khoury MJ, Little J, et al. Size matters: just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. Int J Epidemiol. 2009 Feb;38(1):263–73.

293. Panoutsopoulou K, Zeggini E. Advances in osteoarthritis genetics. J Med Genet. 2013 Nov;50(11):715–24.

294. Zeggini E, Ioannidis JPA. Meta-analysis in genome-wide association studies. Pharmacogenomics. 2009 Feb;10(2):191–201.

295. Panagiotou OA, Willer CJ, Hirschhorn JN, Ioannidis JPA. The power of meta-analysis in genome-wide association studies. Annu Rev Genomics Hum Genet. 2013;14:441–65.

296. Miyamoto Y, Shi D, Nakajima M, Ozaki K, Sudo A, Kotani A, et al. Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. Nat Genet. 2008 Aug;40(8):994–8.

297. Nakajima M, Miyamoto Y, Ikegawa S. Cloning and characterization of the osteoarthritis-associated gene DVWA. J Bone Miner Metab. 2011 May;29(3):300–8.

298. Meulenbelt I, Chapman K, Dieguez-Gonzalez R, Shi D, Tsezou A, Dai J, et al. Large replication study and meta-analyses of DVWA as an osteoarthritis susceptibility locus in European and Asian populations. Human Molecular Genetics. 2009 Jan 27;18(8):1518–23.

299. Lee SJ, Kim MJ, Kee SJ, Song SK, Kweon SS, Shin MH, et al. Association study of the candidate gene for knee osteoarthritis in Koreans. Rheumatol Int. 2013 Mar;33(3):783–6.

300. Valdes AM, Spector TD, Doherty S, Wheeler M, Hart DJ, Doherty M. Association of the DVWA and GDF5 polymorphisms with osteoarthritis in UK populations. Ann Rheum Dis. 2009 Dec;68(12):1916–20.

301. Nakajima M, Takahashi A, Kou I, Rodriguez-Fontenla C, Gomez-Reino JJ, Furuichi T, et al. New Sequence Variants in HLA Class II/III Region Associated with Susceptibility to Knee Osteoarthritis Identified by Genome-Wide Association Study. Agarwal S, editor. PLoS ONE. 2010 Mar 18;5(3):e9723.

302. Shi D, Zheng Q, Chen D, Zhu L, Qin A, Fan J, et al. Association of single-nucleotide polymorphisms in HLA class II/III region with knee osteoarthritis. Osteoarthritis Cartilage. 2010 Nov;18(11):1454–7.

303. Valdes AM, Styrkarsdottir U, Doherty M, Morris DL, Mangino M, Tamm A, et al. Large scale replication study of the association between HLA class II/BTNL2 variants and osteoarthritis of the knee in European-descent populations. PLoS One. 2011;6(8):e23371.

304. Kerkhof HJM, Lories RJ, Meulenbelt I, Jonsdottir I, Valdes AM, Arp P, et al. A genome-wide association study identifies an osteoarthritis susceptibility locus on chromosome 7q22. Arthritis Rheum. 2010 Feb;62(2):499–510.

305. Raine EVA, Wreglesworth N, Dodd AW, Reynard LN, Loughlin J. Gene expression analysis reveals HBP1 as a key target for the osteoarthritis susceptibility locus that maps to chromosome 7q22. Ann Rheum Dis. 2012 Dec;71(12):2020–7.

306. Castaño Betancourt MC, Cailotto F, Kerkhof HJ, Cornelis FMF, Doherty SA, Hart DJ, et al. Genome-wide association and functional studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. Proc Natl Acad Sci U S A. 2012 May 22;109(21):8218–23.

307. Sovio U, Bennett AJ, Millwood IY, Molitor J, O'Reilly PF, Timpson NJ, et al. Genetic Determinants of Height Growth Assessed Longitudinally from Infancy to Adulthood in the Northern Finland Birth Cohort 1966. Gibson G, editor. PLoS Genet. 2009 Mar 6;5(3):e1000409.

308. Evangelou E, Valdes AM, Castano-Betancourt MC, Doherty M, Doherty S, Esko T, et al. The DOT1L rs12982744 polymorphism is associated with osteoarthritis of the hip with genome-wide statistical significance in males. Ann Rheum Dis. 2013 Jul;72(7):1264–5.

309. Evangelou E, Kerkhof HJ, Styrkarsdottir U, Ntzani EE, Bos SD, Esko T, et al. A meta-analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. Ann Rheum Dis. 2014 Dec;73(12):2130–6.

310. Xu J, Liao L, Ning G, Yoshida-Komiya H, Deng C, O'Malley BW. The steroid receptor coactivator SRC-3 (p/CIP/RAC3/AIB1/ACTR/TRAM-1) is required for normal growth, puberty, female reproductive function, and mammary gland development. Proc Natl Acad Sci U S A. 2000 Jun 6;97(12):6379–84.

311. Nicholls JJ, Brassill MJ, Williams GR, Bassett JHD. The skeletal consequences of thyrotoxicosis. J Endocrinol. 2012 Jun;213(3):209–21.

312. Millward-Sadler SJ, Khan NS, Bracher MG, Wright MO, Salter DM. Roles for the interleukin-4 receptor and associated JAK/STAT proteins in human articular chondrocyte mechanotransduction. Osteoarthritis Cartilage. 2006 Oct;14(10):991–1001.

313. Panoutsopoulou K, Southam L, Elliott KS, Wrayner N, Zhai G, Beazley C, et al. Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. Ann Rheum Dis. 2011 May;70(5):864–7.

314. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature. 2010 Oct 28;467(7319):1061–73.

315. Day-Williams AG, Southam L, Panoutsopoulou K, Rayner NW, Esko T, Estrada K, et al. A variant in MCF2L is associated with osteoarthritis. Am J Hum Genet. 2011 Sep 9;89(3):446–50.

316. Mitchell RE, Huitema LFA, Skinner REH, Brunt LH, Severn C, Schulte-Merker S, et al. New tools for studying osteoarthritis genetics in zebrafish. Osteoarthritis Cartilage. 2013 Feb;21(2):269–78.

317. arcOGEN Consortium, arcOGEN Collaborators, Zeggini E, Panoutsopoulou K, Southam L, Rayner NW, et al. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. Lancet. 2012 Sep 1;380(9844):815–23.

318. Lionel AC, Tammimies K, Vaags AK, Rosenfeld JA, Ahn JW, Merico D, et al. Disruption of the ASTN2/TRIM32 locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. Hum Mol Genet. 2014 May 15;23(10):2752–68.

319. Freilinger T, Anttila V, de Vries B, Malik R, Kallela M, Terwindt GM, et al. Genome-wide association analysis identifies susceptibility loci for migraine without aura. Nat Genet. 2012 Jun 10;44(7):777–82.

320. Wilson PM, Fryer RH, Fang Y, Hatten ME. Astn2, A Novel Member of the Astrotactin Gene Family, Regulates the Trafficking of ASTN1 during Glial-Guided Neuronal Migration. Journal of Neuroscience. 2010 Jun 23;30(25):8529–40.

321. Klüppel M, Wight TN, Chan C, Hinek A, Wrana JL. Maintenance of chondroitin sulfation balance by chondroitin-4-sulfotransferase 1 is required for chondrocyte development and growth factor signaling during cartilage morphogenesis. Development. 2005 Sep;132(17):3989–4003.

322. Karlsson C, Dehne T, Lindahl A, Brittberg M, Pruss A, Sittinger M, et al. Genome-wide expression profiling reveals new candidate genes associated with osteoarthritis. Osteoarthritis Cartilage. 2010 Apr;18(4):581–92.

323. Izu Y, Sun M, Zwolanek D, Veit G, Williams V, Cha B, et al. Type XII collagen regulates osteoblast polarity and communication during bone formation. Journal of Cell Biology. 2011 Jun 13;193(6):1115–30.

324. Lanske B, Amling M, Neff L, Guiducci J, Baron R, Kronenberg HM. Ablation of the PTHrP gene or the PTH/PTHrP receptor gene leads to distinct abnormalities in bone development. J Clin Invest. 1999 Aug;104(4):399–407.

325. Yuan Q, Sato T, Densmore M, Saito H, Schüler C, Erben RG, et al. Deletion of PTH Rescues Skeletal Abnormalities and High Osteopontin Levels in Klotho−/− Mice. White KE, editor. PLoS Genet. 2012 May 17;8(5):e1002726.

326. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet. 2009 Jan;41(1):18–24.

327. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007 May 11;316(5826):889–94.

328. Panoutsopoulou K, Metrustry S, Doherty SA, Laslett LL, Maciewicz RA, Hart DJ, et al. The effect of FTO variation on increased osteoarthritis risk is mediated through body mass index: a Mendelian randomisation study. Ann Rheum Dis. 2014 Dec;73(12):2082–6.

329. Liu F, van der Lijn F, Schurmann C, Zhu G, Chakravarty MM, Hysi PG, et al. A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans. Gibson G, editor. PLoS Genet. 2012 Sep 13;8(9):e1002932.

330. Komori T. Regulation of bone development and extracellular matrix protein genes by RUNX2. Cell Tissue Res. 2010 Jan;339(1):189–95.

331. Zengini E, Hatzikotoulas K, Tachmazidou I, Steinberg J, Hartwig FP, Southam L, et al. Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. Nat Genet. 2018 Apr;50(4):549–58.

332. Zhang R, Oyajobi BO, Harris SE, Chen D, Tsao C, Deng HW, et al. Wnt/β-catenin signaling activates bone morphogenetic protein 2 expression in osteoblasts. Bone. 2013 Jan;52(1):145–56.

333. Usmani SE, Pest MA, Kim G, Ohora SN, Qin L, Beier F. Transforming growth factor alpha controls the transition from hypertrophic cartilage to bone during endochondral bone growth. Bone. 2012 Jul;51(1):131–41.

334. Boer CG, Hatzikotoulas K, Southam L, Stefánsdóttir L, Zhang Y, Coutinho de Almeida R, et al. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell. 2021 Sep;184(18):4784-4818.e17.

335. Sedaghati-Khayat B, Boer CG, Runhaar J, Bierma-Zeinstra SMA, Broer L, Ikram MA, et al. Risk Assessment for Hip and Knee Osteoarthritis Using Polygenic Risk Scores. Arthritis & Rheumatology. 2022 Sep;74(9):1488–96.

336. Neogi T, Zhang Y. Osteoarthritis prevention. Current Opinion in Rheumatology. 2011 Mar;23(2):185–91.

337. Bijlsma JWJ, Knahr K. Strategies for the prevention and management of osteoarthritis of the hip and knee. Best Practice & Research Clinical Rheumatology. 2007 Feb;21(1):59–76.

338. Kundu S, Ashinsky BG, Bouhrara M, Dam EB, Demehri S, Shifat-E-Rabbi M, et al. Enabling early detection of osteoarthritis from presymptomatic cartilage texture maps via transport-based learning. Proc Natl Acad Sci USA. 2020 Oct 6;117(40):24709–19.

339. Hunter, David J. "Lower extremity osteoarthritis management needs a paradigm shift." *British Journal of Sports Medicine* 45.4 (2011): 283-288.

340. Felson DT, Lawrence RC, Hochberg MC, McAlindon T, Dieppe PA, Minor MA, et al. Osteoarthritis: new insights. Part 2: treatment approaches. Ann Intern Med. 2000 Nov 7;133(9):726–37.

341. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage. 2014 Mar;22(3):363–88.

342. Puett DW, Griffin MR. Published trials of nonmedicinal and noninvasive therapies for hip and knee osteoarthritis. Ann Intern Med. 1994 Jul 15;121(2):133–40.

343. Minor MA, Hewett JE, Webel RR, Anderson SK, Kay DR. Efficacy of physical conditioning exercise in patients with rheumatoid arthritis and osteoarthritis. Arthritis Rheum. 1989 Nov;32(11):1396–405.

344. Hunter DJ. Lower extremity osteoarthritis management needs a paradigm shift. Br J Sports Med. 2011 Apr;45(4):283–8.

345. Bennell KL, Hinman RS. A review of the clinical evidence for exercise in osteoarthritis of the hip and knee. J Sci Med Sport. 2011 Jan;14(1):4–9.

346. Messier SP, Mihalko SL, Legault C, Miller GD, Nicklas BJ, DeVita P, et al. Effects of intensive diet and exercise on knee joint loads, inflammation, and clinical outcomes among overweight and obese adults with knee osteoarthritis: the IDEA randomized clinical trial. JAMA. 2013 Sep 25;310(12):1263–73.

347. Jones A, Silva PG, Silva AC, Colucci M, Tuffanin A, Jardim JR, et al. Impact of cane use on pain, function, general health and energy expenditure during gait in patients with knee osteoarthritis: a randomised controlled trial. Ann Rheum Dis. 2012 Feb;71(2):172–9.

348. Raja K, Dewan N. Efficacy of knee braces and foot orthoses in conservative management of knee osteoarthritis: a systematic review. Am J Phys Med Rehabil. 2011 Mar;90(3):247–62.

349. van Raaij TM, Reijman M, Brouwer RW, Bierma-Zeinstra SMA, Verhaar JAN. Medial knee osteoarthritis treated by insoles or braces: a randomized trial. Clin Orthop Relat Res. 2010 Jul;468(7):1926–32.

350. Fokmare PS, Phansopkar P. A Review on Osteoarthritis Knee Management via Contrast Bath Therapy and Physical Therapy. Cureus. 2022 Jul 27.

351. Pincus T, Koch GG, Sokka T, Lefkowith J, Wolfe F, Jordan JM, et al. A randomized, double-blind, crossover clinical trial of diclofenac plus misoprostol versus acetaminophen in patients with osteoarthritis of the hip or knee. Arthritis Rheum. 2001 Jul;44(7):1587–98.

352. Pincus T, Swearingen C, Cummins P, Callahan LF. Preference for nonsteroidal antiinflammatory drugs versus acetaminophen and concomitant use of both types of drugs in patients with osteoarthritis. J Rheumatol. 2000 Apr;27(4):1020–7.

353. Towheed TE, Maxwell L, Judd MG, Catton M, Hochberg MC, Wells G. Acetaminophen for osteoarthritis. Cochrane Database Syst Rev. 2006 Jan 25;(1):CD004257.

354. Chan AT, Manson JE, Albert CM, Chae CU, Rexrode KM, Curhan GC, et al. Nonsteroidal antiinflammatory drugs, acetaminophen, and the risk of cardiovascular events. Circulation. 2006 Mar 28;113(12):1578–87.

355. Rahme E, Barkun A, Nedjar H, Gaugris S, Watson D. Hospitalizations for upper and lower GI events associated with traditional NSAIDs and acetaminophen among the elderly in Quebec, Canada. Am J Gastroenterol. 2008 Apr;103(4):872–82.

356. Miceli-Richard C, Le Bars M, Schmidely N, Dougados M. Paracetamol in osteoarthritis of the knee. Ann Rheum Dis. 2004 Aug;63(8):923–30.

357. Brown GA. AAOS clinical practice guideline: treatment of osteoarthritis of the knee: evidence-based guideline, 2nd edition. J Am Acad Orthop Surg. 2013 Sep;21(9):577–9.

358. Simon LS, Grierson LM, Naseer Z, Bookman AAM, Shainhouse ZJ. Efficacy and safety of topical diclofenac containing dimethyl sulfoxide (DMSO) compared with those of topical placebo, DMSO vehicle and oral diclofenac for knee osteoarthritis. Pain. 2009 Jun;143(3):238–45.

359. Tugwell PS, Wells GA, Shainhouse JZ. Equivalence study of a topical diclofenac solution (pennsaid) compared with oral diclofenac in symptomatic treatment of osteoarthritis of the knee: a randomized controlled trial. J Rheumatol. 2004 Oct;31(10):2002–12.

360. Chou R, McDonagh MS, Nakamoto E, Griffin J. Analgesics for Osteoarthritis: An Update of the 2006 Comparative Effectiveness Review. Rockville (MD): Agency for Healthcare Research and Quality (US); 2011.

361. Mason L, Moore RA, Derry S, Edwards JE, McQuay HJ. Systematic review of topical capsaicin for the treatment of chronic pain. BMJ. 2004 Apr 24;328(7446):991.

362. Bannuru RR, Natov NS, Obadan IE, Price LL, Schmid CH, McAlindon TE. Therapeutic trajectory of hyaluronic acid versus corticosteroids in the treatment of knee osteoarthritis: a systematic review and meta-analysis. Arthritis Rheum. 2009 Dec 15;61(12):1704–11.

363. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, et al. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. American College of Rheumatology. Arthritis Rheum. 1995 Nov;38(11):1541–6.

364. Raynauld JP, Buckland-Wright C, Ward R, Choquette D, Haraoui B, Martel-Pelletier J, et al. Safety and efficacy of long-term intraarticular steroid injections in osteoarthritis of the knee: a randomized, double-blind, placebo-controlled trial. Arthritis Rheum. 2003 Feb;48(2):370–7.

365. Nüesch E, Rutjes AW, Husni E, Welch V, Jüni P. Oral or transdermal opioids for osteoarthritis of the knee or hip. Cochrane Database Syst Rev. 2009 Oct 7;(4):CD003115.

366. Solomon DH, Rassen JA, Glynn RJ, Lee J, Levin R, Schneeweiss S. The comparative safety of analgesics in older adults with arthritis. Arch Intern Med. 2010 Dec 13;170(22):1968–76.

367. Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage. 2010 Apr;18(4):476–99.

368. Reichenbach S, Sterchi R, Scherer M, Trelle S, Bürgi E, Bürgi U, et al. Meta-analysis: chondroitin for osteoarthritis of the knee or hip. Ann Intern Med. 2007 Apr 17;146(8):580–90.

369. Clegg DO, Reda DJ, Harris CL, Klein MA, O'Dell JR, Hooper MM, et al. Glucosamine, chondroitin sulfate, and the two in combination for painful knee osteoarthritis. N Engl J Med. 2006 Feb 23;354(8):795–808.

370. Hochberg MC, Zhan M, Langenberg P. The rate of decline of joint space width in patients with osteoarthritis of the knee: a systematic review and meta-analysis of randomized placebo-controlled trials of chondroitin sulfate. Curr Med Res Opin. 2008 Nov;24(11):3029–35.

371. Fransen M, Agaliotis M, Nairn L, Votrubec M, Bridgett L, Su S, et al. Glucosamine and chondroitin for knee osteoarthritis: a double-blind randomised placebo-controlled clinical trial evaluating single and combination regimens. Ann Rheum Dis. 2015 May;74(5):851–8.

372. Arden NK, Cro S, Sheard S, Doré CJ, Bara A, Tebbs SA, et al. The effect of vitamin D supplementation on knee osteoarthritis, the VIDEO study: a randomised controlled trial. Osteoarthritis and Cartilage. 2016 Nov;24(11):1858–66.

373. Fidelix TSA, Macedo CR, Maxwell LJ, Fernandes Moça Trevisani V. Diacerein for osteoarthritis. Cochrane Musculoskeletal Group, editor. Cochrane Database of Systematic Reviews. 2014 Feb 10.

374. Christensen R, Bartels EM, Astrup A, Bliddal H. Symptomatic efficacy of avocado-soybean unsaponifiables (ASU) in osteoarthritis (OA) patients: a meta-analysis of randomized controlled trials. Osteoarthritis Cartilage. 2008 Apr;16(4):399–408.

375. Cleland LG, James MJ. Osteoarthritis. Omega-3 fatty acids and synovitis in osteoarthritic knees. Nat Rev Rheumatol. 2012 Apr 17;8(6):314–5.

376. Lopez HL. Nutritional interventions to prevent and treat osteoarthritis. Part I: focus on fatty acids and macronutrients. PM R. 2012 May;4(5 Suppl):S145-154.

377. 2013 Annual Meeting Abstract Supplement. Arthritis & Rheumatism, 65: S1-S1331. https://doi.org/10.1002/art.38216.

378. Li J, Li YX, Luo LJ, Ye J, Zhong DL, Xiao QW, et al. The effectiveness and safety of acupuncture for knee osteoarthritis: An overview of systematic reviews. Medicine. 2019 Jul;98(28):e16301.

379. Hou PW, Fu PK, Hsu HC, Hsieh CL. Traditional Chinese medicine in patients with osteoarthritis of the knee. J Tradit Complement Med. 2015 Oct;5(4):182–96.

380. Lo PC, Lin FC, Tsai YC, Lin SK. Traditional Chinese medicine therapy reduces the risk of total knee replacement in patients with knee osteoarthritis. Medicine. 2019 Jun;98(23):e15964.

381. Chen LX, Zhou ZR, Li YL, Ning GZ, Li Y, Wang XB, et al. Transcutaneous Electrical Nerve Stimulation in Patients With Knee Osteoarthritis: Evidence From Randomized-controlled Trials. The Clinical Journal of Pain. 2016 Feb;32(2):146–54.

382. Rodriguez-Merchan, E. Carlos. "The current role of disease-modifying osteoarthritis drugs." *Archives of Bone and Joint Surgery* 11.1 (2023): 11.

383. Song GG, Lee YH. Relative efficacy and tolerability of 2.5, 5, and 10 mg tanezumab for the treatment of osteoarthritis: A Bayesian network meta-analysis of randomized controlled trials based on patient withdrawal. CP. 2021 Feb 1;59(02):147–55.

384. Schnitzer TJ, Easton R, Pang S, Levinson DJ, Pixton G, Viktrup L, et al. Effect of Tanezumab on Joint Pain, Physical Function, and Patient Global Assessment of Osteoarthritis Among Patients With Osteoarthritis of the Hip or Knee: A Randomized Clinical Trial. JAMA. 2019 Jul 2;322(1):37.

385. Berenbaum F, Blanco FJ, Guermazi A, Miki K, Yamabe T, Viktrup L, et al. Subcutaneous tanezumab for osteoarthritis of the hip or knee: efficacy and safety results from a 24-week randomised phase III study with a 24-week follow-up period. Ann Rheum Dis. 2020 Jun;79(6):800–10.

386. Lane NE, Schnitzer TJ, Birbara CA, Mokhtarani M, Shelton DL, Smith MD, et al. Tanezumab for the Treatment of Pain from Osteoarthritis of the Knee. N Engl J Med. 2010 Oct 14;363(16):1521–31.

387. Schnitzer TJ, Lane NE, Birbara C, Smith MD, Simpson SL, Brown MT. Long-term open-label study of tanezumab for moderate to severe osteoarthritic knee pain. Osteoarthritis and Cartilage. 2011 Jun;19(6):639–46.

388. Birbara C, Dabezies E, Burr A, Fountaine R, Smith M, Brown M, et al. Safety and efficacy of subcutaneous tanezumab in patients with knee or hip osteoarthritis. JPR. 2018 Jan;Volume 11:151–64.

389. Cohen SB, Proudman S, Kivitz AJ, Burch FX, Donohue JP, Burstein D, et al. A randomized, double-blind study of AMG 108 (a fully human monoclonal antibody to IL-1R1) in patients with osteoarthritis of the knee. Arthritis Res Ther. 2011;13(4):R125.

390. Chevalier X, Goupille P, Beaulieu AD, Burch FX, Bensen WG, Conrozier T, et al. Intraarticular injection of anakinra in osteoarthritis of the knee: A multicenter, randomized, double-blind, placebo-controlled study. Arthritis Rheum. 2009 Mar 15;61(3):344–52.

391. Wang J. Efficacy and safety of adalimumab by intra-articular injection for moderate to severe knee osteoarthritis: An open-label randomized controlled trial. J Int Med Res. 2018 Jan;46(1):326–34.

392. Kloppenburg M, Ramonda R, Bobacz K, Kwok WY, Elewaut D, Huizinga TWJ, et al. Etanercept in patients with inflammatory hand osteoarthritis (EHOA): a multicentre, randomised, double-blind, placebo-controlled trial. Ann Rheum Dis. 2018 Dec;77(12):1757–64.

393. Li S, Cao P, Chen T, Ding C. Latest insights in disease-modifying osteoarthritis drugs development. Therapeutic Advances in Musculoskeletal. 2023 Jan;15:1759720X2311698.

394. Siebuhr AS, Werkmann D, Bay-Jensen AC, Thudium CS, Karsdal MA, Serruys B, et al. The Anti-ADAMTS-5 Nanobody® M6495 Protects Cartilage Degradation Ex Vivo. IJMS. 2020 Aug 20;21(17):5992.

395. Snijders GF, Van Den Ende CH, Van Riel PL, Van Den Hoogen FH, Den Broeder AA, on behalf of NOAC study group. The effects of doxycycline on reducing symptoms in knee osteoarthritis: results from a triple-blinded randomised controlled trial. Annals of the Rheumatic Diseases. 2011 Jul 1;70(7):1191–6.

396. Zhang X, Deng XH, Song Z, Croen B, Carballo CB, Album Z, et al. Matrix Metalloproteinase Inhibition With Doxycycline Affects the Progression of Posttraumatic Osteoarthritis After Anterior Cruciate Ligament Rupture: Evaluation in a New Nonsurgical Murine ACL Rupture Model. Am J Sports Med. 2020 Jan;48(1):143–52.

397. Hellio Le Graverand MP, Clemmer RS, Redifer P, Brunell RM, Hayes CW, Brandt KD, et al. A 2-year randomised, double-blind, placebo-controlled, multicentre study of oral selective iNOS inhibitor, cindunistat (SD-6010), in patients with symptomatic osteoarthritis of the knee. Ann Rheum Dis. 2013 Feb;72(2):187–95.

398. Krzeski P, Buckland-Wright C, Bálint G, Cline GA, Stoner K, Lyon R, et al. Development of musculoskeletal toxicity without clear benefit after administration of PG-116800, a matrix metalloproteinase inhibitor, to patients with knee osteoarthritis: a randomized, 12-month, double-blind, placebo-controlled study. Arthritis Res Ther. 2007;9(5):R109.

399. Hunter DJ, Pike MC, Jonas BL, Kissin E, Krop J, McAlindon T. Phase 1 safety and tolerability study of BMP-7 in symptomatic knee osteoarthritis. BMC Musculoskelet Disord. 2010 Dec;11(1):232.

400. Müller S, Lindemann S, Gigout A. Effects of Sprifermin, IGF1, IGF2, BMP7, or CNP on Bovine Chondrocytes in Monolayer and 3D Culture. J Orthop Res. 2020 Mar;38(3):653–62.

401. Lohmander LS, Hellot S, Dreher D, Krantz EFW, Kruger DS, Guermazi A, et al. Intraarticular Sprifermin (Recombinant Human Fibroblast Growth Factor 18) in Knee Osteoarthritis: A Randomized, Double-Blind, Placebo-Controlled Trial: Sprifermin Effects in Knee Osteoarthritis. Arthritis & Rheumatology. 2014 Jul;66(7):1820–31.

402. Li J, Wang X, Ruan G, Zhu Z, Ding C. Sprifermin: a recombinant human fibroblast growth factor 18 for the treatment of knee osteoarthritis. Expert Opinion on Investigational Drugs. 2021 Sep 2;30(9):923–30.

403. Madry H, Cucchiarini M. Gene therapy for human osteoarthritis: principles and clinical translation. Expert Opinion on Biological Therapy. 2016 Mar 3;16(3):331–46.

404. Ham O, Song BW, Lee SY, Choi E, Cha MJ, Lee CY, et al. The role of microRNA-23b in the differentiation of MSC into chondrocyte by targeting protein kinase A signaling. Biomaterials. 2012 Jun;33(18):4500–7.

405. Matsukawa T, Sakai T, Yonezawa T, Hiraiwa H, Hamada T, Nakashima M, et al. MicroRNA-125b regulates the expression of aggrecanase-1 (ADAMTS-4) in human osteoarthritic chondrocytes. Arthritis Res Ther. 2013;15(1):R28.

406. Meng F, Zhang Z, Chen W, Huang G, He A, Hou C, et al. MicroRNA-320 regulates matrix metalloproteinase-13 expression in chondrogenesis and interleukin-1β-induced chondrocyte responses. Osteoarthritis and Cartilage. 2016 May;24(5):932–41.

407. Tuddenham L, Wheeler G, Ntounia-Fousara S, Waters J, Hajihosseini MK, Clark I, et al. The cartilage specific microRNA-140 targets histone deacetylase 4 in mouse cells. FEBS Letters. 2006 Jul 24;580(17):4214–7.

408. Miyaki S, Sato T, Inoue A, Otsuki S, Ito Y, Yokoyama S, et al. MicroRNA-140 plays dual roles in both cartilage development and homeostasis. Genes Dev. 2010 Jun 1;24(11):1173–85.

409. El Hajjaji H, Williams JM, Devogelaer JP, Lenz ME, Thonar EJMA, Manicourt DH. Treatment with calcitonin prevents the net loss of collagen, hyaluronan and proteoglycan aggregates from cartilage in the early stages of canine experimental osteoarthritis. Osteoarthritis and Cartilage. 2004 Nov;12(11):904–11.

410. Karsdal MA, Byrjalsen I, Henriksen K, Riis BJ, Lau EM, Arnold M, et al. The effect of oral salmon calcitonin delivered with 5-CNAC on bone and cartilage degradation in osteoarthritic patients: a 14-day randomized study. Osteoarthritis and Cartilage. 2010 Feb;18(2):150–9.

411. Goldring MB, Berenbaum F. Emerging targets in osteoarthritis therapy. Current Opinion in Pharmacology. 2015 Jun;22:51–63.

412. Wang Y, Fan X, Xing L, Tian F. Wnt signaling: a promising target for osteoarthritis therapy. Cell Commun Signal. 2019 Dec;17(1):97.

413. Childs BG, Gluscevic M, Baker DJ, Laberge RM, Marquess D, Dananberg J, et al. Senescent cells: an emerging target for diseases of ageing. Nat Rev Drug Discov. 2017 Oct;16(10):718–35.

414. Jeon OH, Kim C, Laberge RM, Demaria M, Rathod S, Vasserot AP, et al. Local clearance of senescent cells attenuates the development of post-traumatic osteoarthritis and creates a pro-regenerative environment. Nat Med. 2017 Jun;23(6):775–81.

415. Van Deursen JM. Senolytic therapies for healthy longevity. Science. 2019 May 17;364(6441):636–7.

416. Atobe M. Activation of Transient Receptor Potential Vanilloid (TRPV) 4 as a Therapeutic Strategy in Osteoarthritis. CTMC. 2019 Nov 18;19(24):2254–67.

417. Li C, Zheng Z, Ha P, Jiang W, Berthiaume EA, Lee S, et al. Neural EGFL like 1 as a potential pro-chondrogenic, anti-inflammatory dual-functional disease-modifying osteoarthritis drug. Biomaterials. 2020 Jan;226:119541.

418. Kjelgaard-Petersen CF, Sharma N, Kayed A, Karsdal MA, Mobasheri A, Hägglund P, et al. Tofacitinib and TPCA-1 exert chondroprotective effects on extracellular matrix turnover in bovine articular cartilage ex vivo. Biochemical Pharmacology. 2019 Jul;165:91–8.

419. Sabha M, Siaton BC, Hochberg MC. Lorecivivint, an intra-articular potential disease-modifying osteoarthritis drug. Expert Opinion on Investigational Drugs. 2020 Dec 1;29(12):1339–46.

420. Yazici Y, McAlindon TE, Gibofsky A, Lane NE, Lattermann C, Skrepnik N, et al. A Phase 2b randomized trial of lorecivivint, a novel intra-articular CLK2/DYRK1A inhibitor and Wnt pathway modulator for knee osteoarthritis. Osteoarthritis and Cartilage. 2021 May;29(5):654–66.

421. Guo H, Yin W, Zou Z, Zhang C, Sun M, Min L, et al. Quercitrin alleviates cartilage extracellular matrix degradation and delays ACLT rat osteoarthritis development: An in vivo and in vitro study. Journal of Advanced Research. 2021 Feb;28:255–67.

422. Kirkley A, Birmingham TB, Litchfield RB, Giffin JR, Willits KR, Wong CJ, et al. A randomized trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med. 2008 Sep 11;359(11):1097–107.

423. Thorlund JB, Juhl CB, Roos EM, Lohmander LS. Arthroscopic surgery for degenerative knee: systematic review and meta-analysis of benefits and harms. BMJ. 2015 Jun 16;350(jun16 3):h2747–h2747.

424. Katz JN, Brophy RH, Chaisson CE, de Chaves L, Cole BJ, Dahm DL, et al. Surgery versus physical therapy for a meniscal tear and osteoarthritis. N Engl J Med. 2013 May 2;368(18):1675–84.

425. Buckwalter JA, Lohmander S. Operative treatment of osteoarthrosis. Current practice and future development. J Bone Joint Surg Am. 1994 Sep;76(9):1405–18.

426. Hiligsmann M, Cooper C, Arden N, Boers M, Branco JC, Luisa Brandi M, et al. Health economics in the field of osteoarthritis: An Expert's consensus paper from the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). Seminars in Arthritis and Rheumatism. 2013 Dec;43(3):303–13.

427. Hiligsmann M, Cooper C, Guillemin F, Hochberg MC, Tugwell P, Arden N, et al. A reference case for economic evaluations in osteoarthritis: An expert consensus article from the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). Seminars in Arthritis and Rheumatism. 2014 Dec;44(3):271–82.

428. WHO Scientific Group on the Burden of Musculoskeletal Conditions at the Start of the New Millennium. The burden of musculoskeletal conditions at the start of the new millennium. World Health Organ Tech Rep Ser. 2003;919:i–x, 1–218, back cover.

429. MacDonald KV, Sanmartin C, Langlois K, Marshall DA. Symptom onset, diagnosis and management of osteoarthritis. Health Rep. 2014 Sep;25(9):10–7.

430. Johnston SA. Osteoarthritis. Joint anatomy, physiology, and pathobiology. Vet Clin North Am Small Anim Pract. 1997 Jul;27(4):699–723.

431. Dieppe PA, Lohmander LS. Pathogenesis and management of pain in osteoarthritis. The Lancet. 2005 Mar;365(9463):965–73.

432. Cicuttini FM, Wluka AE. Osteoarthritis: Is OA a mechanical or systemic disease? Nat Rev Rheumatol. 2014 Sep;10(9):515–6.

433. Kean WF, Kean R, Buchanan WW. Osteoarthritis: symptoms, signs and source of pain. Inflammopharmacology. 2004 Feb;12(1):3–31.

434. Herrero-Beaumont G, Roman-Blas JA, Castañeda S, Jimenez SA. Primary Osteoarthritis No Longer Primary: Three Subsets with Distinct Etiological, Clinical, and Therapeutic Characteristics. Seminars in Arthritis and Rheumatism. 2009 Oct;39(2):71–80.

435. McGonagle D, Tan AL, Carey J, Benjamin M. The anatomical basis for a novel classification of osteoarthritis and allied disorders. JAnat. 2010 Mar;216(3):279–91.

436. Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with relevance for clinical practice. The Lancet. 2011 Jun;377(9783):2115–26.

437. Scanzello CR, Umoh E, Pessler F, Diaz-Torne C, Miles T, DiCarlo E, et al. Local cytokine profiles in knee osteoarthritis: elevated synovial fluid interleukin-15 differentiates early from end-stage disease. Osteoarthritis and Cartilage. 2009 Aug;17(8):1040–8.

438. Solovieva S, Kämäräinen OP, Hirvonen A, Hämäläinen S, Laitala M, Vehmas T, et al. Association Between Interleukin 1 Gene Cluster Polymorphisms and Bilateral Distal Interphalangeal Osteoarthritis. J Rheumatol. 2009 Sep;36(9):1977–86.

439. MacGregor AJ, Li Q, Spector TD, Williams FMK. The genetic influence on radiographic osteoarthritis is site specific at the hand, hip and knee. Rheumatology. 2008 Oct 17;48(3):277–80.

440. Haynes MK, Hume EL, Smith JB. Phenotypic characterization of inflammatory cells from osteoarthritic synovium and synovial fluids. Clin Immunol. 2002 Dec;105(3):315–25.

441. Zhang W, Likhodii S, Zhang Y, Aref-Eshghi E, Harper PE, Randell E, et al. Classification of osteoarthritis phenotypes by metabolomics analysis. BMJ Open. 2014 Nov 19;4(11):e006286.

442. Zhai G, Randell EW, Rahman P. Metabolomics of osteoarthritis: emerging novel markers and their potential clinical utility. Rheumatology (Oxford). 2018 Dec 1;57(12):2087–95.

443. Li Z, Zhang Y, Hu T, Likhodii S, Sun G, Zhai G, et al. Differential metabolomics analysis allows characterization of diversity of metabolite networks between males and females. PLoS One. 2018;13(11):e0207775.

444. Katz JD, Agrawal S, Velasquez M. Getting to the heart of the matter: osteoarthritis takes its place as part of the metabolic syndrome. Current Opinion in Rheumatology. 2010 Sep;22(5):512–9.

445. Zhuo Q, Yang W, Chen J, Wang Y. Metabolic syndrome meets osteoarthritis. Nat Rev Rheumatol. 2012 Dec;8(12):729–37.

446. Gowda GN, Zhang S, Gu H, Asiago V, Shanaiah N, Raftery D. Metabolomics-based methods for early disease diagnostics. Expert Review of Molecular Diagnostics. 2008 Sep;8(5):617–33.

447. Berenbaum F. Osteoarthritis as an inflammatory disease (osteoarthritis is not osteoarthrosis!). Osteoarthritis and Cartilage. 2013 Jan;21(1):16–21.

448. Courties A, Gualillo O, Berenbaum F, Sellam J. Metabolic stress-induced joint inflammation and osteoarthritis. Osteoarthritis and Cartilage. 2015 Nov;23(11):1955–65.

449. Lepetsos P, Papavassiliou AG. ROS/oxidative stress signaling in osteoarthritis. Biochim Biophys Acta. 2016 Apr;1862(4):576–91.

450. Zhai Guangju, Aref-Eshghi Erfan, Rahman Proton. Attempt to replicate the published osteoarthritis-associated genetic variants in the newfoundland & labrador population. Journal of Orthopedics & Rheumatology. April 2014 Vol.:2, Issue:2.

451. Fontaine-Bisson B, Thorburn J, Gregory A, Zhang H, Sun G. Melanin-concentrating hormone receptor 1 polymorphisms are associated with components of energy balance in

the Complex Diseases in the Newfoundland Population: Environment and Genetics (CODING) study. The American Journal of Clinical Nutrition. 2014 Feb 1;99(2):384–91.

452. Sun G, French CR, Martin GR, Younghusband B, Green RC, Xie Y gang, et al. Comparison of multifrequency bioelectrical impedance analysis with dual-energy X-ray absorptiometry for assessment of percentage body fat in a large, healthy population. The American Journal of Clinical Nutrition. 2005 Jan;81(1):74–8.

453. Zhang W, Sun G, Aitken D, Likhodii S, Liu M, Martin G, et al. Lysophosphatidylcholines to phosphatidylcholines ratio predicts advanced knee osteoarthritis. Rheumatology (Oxford). 2016 Sep;55(9):1566–74.

454. Zhang W, Likhodii S, Aref-Eshghi E, Zhang Y, Harper PE, Randell E, et al. Relationship Between Blood Plasma and Synovial Fluid Metabolite Concentrations in Patients with Osteoarthritis. J Rheumatol. 2015 May;42(5):859–65.

455. Zhai G, Wang-Sattler R, Hart DJ, Arden NK, Hakim AJ, Illig T, et al. Serum branched-chain amino acid to histidine ratio: a novel metabolomic biomarker of knee osteoarthritis. Ann Rheum Dis. 2010 Jun;69(6):1227–31.

456. Mittelstrass K, Ried JS, Yu Z, Krumsiek J, Gieger C, Prehn C, et al. Discovery of Sexual Dimorphisms in Metabolic and Genetic Biomarkers. McCarthy MI, editor. PLoS Genet. 2011 Aug 11;7(8):e1002215.

457. Williams B, Onsman A, Brown T. Exploratory factor analysis: A five-step guide for novices. Australasian Journal of Paramedicine. 2010 Aug 2.

458. Macrosson WDK. Scree Plots, Data Structure, and Random Variance. Psychol Rep. 1999 Apr;84(2):533–40.

459. Buja A, Eyuboglu N. Remarks on Parallel Analysis. Multivariate Behavioral Research. 1992 Oct;27(4):509–40.

460. Thompson B. Factor analysis. The Blackwell Encyclopedia of Sociology. In 2007.

461. Dinning PG. Factor analysis identifies subgroups of constipation. WJG. 2011;17(11):1468.

462. Lawson RG, Jurs PC. New index for clustering tendency and its application to chemical problems. JChemInfComputSci. 1990;30(1):36–41.

463. Ng A. Clustering with the k-means algorithm. Mach Learning. 2012;

464. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. 1987 Nov;20:53–65.

465. Brock G, Pihur V, Datta S, Datta S. **clValid** : An *R* Package for Cluster Validation. J Stat Soft. 2008.

466. Shutaywi M, Kachouie NN. Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering. Entropy. 2021 Jun 16;23(6):759.

467. Petersen AK, Krumsiek J, Wägele B, Theis FJ, Wichmann HE, Gieger C, et al. On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. BMC Bioinformatics. 2012 Jun 6;13:120.

468. Williams CK, Engelhardt A, Cooper T, Mayer Z, Ziem A, Scrucca L, et al. Package 'caret'. [Internet]. 2019. Available from: : https://cran.r-project.org/web/packages/caret/caret.pdf.

469. Zhang W, Likhodii S, Zhang Y, Aref-Eshghi E, Harper PE, Randell E, et al. Classification of osteoarthritis phenotypes by metabolomics analysis. BMJ Open. 2014 Nov 19;4(11):e006286-2014–006286.

470. Mels CMC, Jansen van Rensburg P, van der Westhuizen FH, Pretorius PJ, Erasmus E. Increased Excretion of C4-Carnitine Species after a Therapeutic Acetylsalicylic Acid Dose: Evidence for an Inhibitory Effect on Short-Chain Fatty Acid Metabolism. ISRN Pharmacology. 2011 Feb 22;2011:1–8.

471. Nagan N, Kruckeberg KE, Tauscher AL, Bailey KS, Rinaldo P, Matern D. The frequency of short-chain acyl-CoA dehydrogenase gene variants in the US population and correlation with the C(4)-acylcarnitine concentration in newborn blood spots. Mol Genet Metab. 2003 Apr;78(4):239–46.

472. Pedersen CB, Kølvraa S, Kølvraa A, Stenbroen V, Kjeldsen M, Ensenauer R, et al. The ACADS gene variation spectrum in 114 patients with short-chain acyl-CoA dehydrogenase (SCAD) deficiency is dominated by missense variations leading to protein misfolding at the cellular level. Hum Genet. 2008 Aug;124(1):43–56.

473. O'Reilly S, Jones A, Doherty M. Muscle weakness in osteoarthritis. Curr Opin Rheumatol. 1997 May;9(3):259–62.

474. Villarreal-Pérez JZ, Villarreal-Martínez JZ, Lavalle-González FJ, Torres-Sepúlveda MDR, Ruiz-Herrera C, Cerda-Flores RM, et al. Plasma and urine metabolic profiles are reflective of altered beta-oxidation in non-diabetic obese subjects and patients with type 2 diabetes mellitus. Diabetol Metab Syndr. 2014;6:129.

475. Makrecka-Kuka M, Sevostjanovs E, Vilks K, Volska K, Antone U, Kuka J, et al. Plasma acylcarnitine concentrations reflect the acylcarnitine profile in cardiac tissues. Sci Rep. 2017 Dec 13;7(1):17528.

476. Zhang W, Sun G, Likhodii S, Aref-Eshghi E, Harper PE, Randell E, et al. Metabolomic analysis of human synovial fluid and plasma reveals that phosphatidylcholine metabolism is associated with both osteoarthritis and diabetes mellitus. Metabolomics. 2016 Feb;12(2):24.

477. Zhang W, Sun G, Likhodii S, Liu M, Aref-Eshghi E, Harper PE, et al. Metabolomic analysis of human plasma reveals that arginine is depleted in knee osteoarthritis patients. Osteoarthritis Cartilage. 2016 May;24(5):827–34.

478. Pascale V, Pascale W, Lavanga V, Sansone V, Ferrario P, De Gennaro Colonna V. L-arginine, asymmetric dimethylarginine, and symmetric dimethylarginine in plasma and synovial fluid of patients with knee osteoarthritis. Med Sci Monit. 2013;19:1057–62.

479. Ohnishi A, Osaki T, Matahira Y, Tsuka T, Imagawa T, Okamoto Y, et al. Correlation of plasma amino acid concentrations and chondroprotective effects of glucosamine and fish collagen peptide on the development of osteoarthritis. J Vet Med Sci. 2013 May 2;75(4):497–502.

480. Zhai G. Alteration of Metabolic Pathways in Osteoarthritis. Metabolites. 2019 Jan 9;9(1):11.

481. Tousoulis D, Antoniades C, Tentolouris C, Goumas G, Stefanadis C, Toutouzas P. L-arginine in cardiovascular disease: dream or reality? Vasc Med. 2002 Aug;7(3):203–11.

482. Gornik HL, Creager MA. Arginine and endothelial and vascular health. J Nutr. 2004 Oct;134(10 Suppl):2880S-2887S; discussion 2895S.

483. Sudar-Milovanovic E, Obradovic M, Jovanovic A, Zaric B, Zafirovic S, Panic A, et al. Benefits of L-Arginine on Cardiovascular System. Mini Rev Med Chem. 2016;16(2):94–103.

484. Zhang W, Sun G, Likhodii S, Aref-Eshghi E, Harper PE, Randell E, et al. Metabolomic analysis of human synovial fluid and plasma reveals that phosphatidylcholine metabolism is associated with both osteoarthritis and diabetes mellitus. Metabolomics. 2016;12(2):24.

485. Goetzl EJ. Pleiotypic mechanisms of cellular responses to biologically active lysophospholipids. Prostaglandins Other Lipid Mediat. 2001 Apr;64(1–4):11–20.

486. Kabarowski JHS, Xu Y, Witte ON. Lysophosphatidylcholine as a ligand for immunoregulation. Biochem Pharmacol. 2002 Jul 15;64(2):161–7.

487. Hsieh CC, Yen MH, Liu HW, Lau YT. Lysophosphatidylcholine induces apoptotic and non-apoptotic death in vascular smooth muscle cells: in comparison with oxidized LDL. Atherosclerosis. 2000 Aug;151(2):481–91.

488. Leistad L, Feuerherm AJ, Faxvaag A, Johansen B. Multiple phospholipase A2 enzymes participate in the inflammatory process in osteoarthritic cartilage. Scand J Rheumatol. 2011;40(4):308–16.

489. Hilbert N, Schiller J, Arnhold J, Arnold K. Cartilage degradation by stimulated human neutrophils: elastase is mainly responsible for cartilage damage. Bioorg Chem. 2002 Apr;30(2):119–32.

490. Zhai G, Pelletier JP, Liu M, Aitken D, Randell E, Rahman P, et al. Activation of The Phosphatidylcholine to Lysophosphatidylcholine Pathway Is Associated with Osteoarthritis Knee Cartilage Volume Loss Over Time. Sci Rep. 2019 Jul 4;9(1):9648.

491. Zhai G, Pelletier JP, Liu M, Randell EW, Rahman P, Martel-Pelletier J. Serum lysophosphatidylcholines to phosphatidylcholines ratio is associated with symptomatic responders to symptomatic drugs in knee osteoarthritis patients. Arthritis Res Ther. 2019 Nov 6;21(1):224.

492. Fuchs B, Schiller J, Wagner U, Häntzschel H, Arnold K. The phosphatidylcholine/lysophosphatidylcholine ratio in human plasma is an indicator of the severity of rheumatoid arthritis: investigations by 31P NMR and MALDI-TOF MS. ClinBiochem. 2005;38(10):925–33.

493. Bultink IEM, Lems WF. Osteoarthritis and osteoporosis: what is the overlap? Curr Rheumatol Rep. 2013 May;15(5):328.

494. Loeser RF, Goldring SR, Scanzello CR, Goldring MB. Osteoarthritis: A disease of the joint as an organ. Arthritis & Rheumatism. 2012 Jun;64(6):1697–707.

495. Shorter E, Sannicandro AJ, Poulet B, Goljanek-Whysall K. Skeletal Muscle Wasting and Its Relationship With Osteoarthritis: a Mini-Review of Mechanisms and Current Interventions. Curr Rheumatol Rep. 2019 Aug;21(8):40.

496. Papa EV, Dong X, Hassan M. Skeletal Muscle Function Deficits in the Elderly: Current Perspectives on Resistance Training. J Nat Sci. 2017 Jan;3(1):e272.

497. Bordoni B, Varacallo M. Anatomy, Tendons. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022. Available from: http://www.ncbi.nlm.nih.gov/books/NBK513237/

498. Bordoni B, Varacallo M. Anatomy, Head and Neck, Scalenus Muscle. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022.

499. Narici MV, Maffulli N. Sarcopenia: characteristics, mechanisms and functional significance. British Medical Bulletin. 2010 Sep 1;95(1):139–59.

500. Silva JM de S, Alabarse PVG, Teixeira V de ON, Freitas EC, de Oliveira FH, Chakr RM da S, et al. Muscle wasting in osteoarthritis model induced by anterior cruciate ligament transection. Williams JL, editor. PLoS ONE. 2018 Apr 30;13(4):e0196682.

501. Larsson L, Degens H, Li M, Salviati L, Lee Y il, Thompson W, et al. Sarcopenia: Aging-Related Loss of Muscle Mass and Function. Physiological Reviews. 2019 Jan 1;99(1):427–511.

502. McCormick R, Vasilaki A. Age-related changes in skeletal muscle: changes to life-style as a therapy. Biogerontology. 2018 Dec;19(6):519–36.

503. Holloszy JO. The Biology of Aging. Mayo Clinic Proceedings. 2000 Jan;75(1):S3–9.

504. Volpi E, Nazemi R, Fujita S. Muscle tissue changes with aging: Current Opinion in Clinical Nutrition and Metabolic Care. 2004 Jul;7(4):405–10.

505. Valderrabano V, Steiger C. Vol. 2011, Journal of Aging Research. Hindawi; 2010. p. e374653 Treatment and Prevention of Osteoarthritis through Exercise and Sports. Available from: https://www.hindawi.com/journals/jar/2011/374653/

506. Kalyani RR, Corriere M, Ferrucci L. Age-related and disease-related muscle loss: the effect of diabetes, obesity, and other diseases. The Lancet Diabetes & Endocrinology. 2014 Oct;2(10):819–29.

507. Legrand D, Vaes B, Matheï C, Adriaensen W, Van Pottelbergh G, Degryse JM. Muscle strength and physical performance as predictors of mortality, hospitalization, and disability in the oldest old. J Am Geriatr Soc. 2014 Jun;62(6):1030–8.

508. Abellan Van Kan G. Epidemiology and consequences of sarcopenia. J Nutr Health Aging. 2009 Oct;13(8):708–12.

509. Brandt KD. Putting Some Muscle into Osteoarthritis. Ann Intern Med. 1997 Jul 15;127(2):154.

510. Hortobágyi T, Westerkamp L, Beam S, Moody J, Garry J, Holbert D, et al. Altered hamstring-quadriceps muscle balance in patients with knee osteoarthritis. Clinical Biomechanics. 2005 Jan;20(1):97–104.

511. Hortobágyi T, Garry J, Holbert D, Devita P. Aberrations in the control of quadriceps muscle force in patients with knee osteoarthritis: Quadriceps Control in Knee OA. Arthritis & Rheumatism. 2004 Aug 15;51(4):562–9.

512. Cicuttini FM, Teichtahl AJ, Wluka AE, Davis S, Strauss BJG, Ebeling PR. The relationship between body composition and knee cartilage volume in healthy, middle-aged subjects. Arthritis Rheum. 2005 Feb;52(2):461–7.

513. Werdyani S, Liu M, Zhang H, Sun G, Furey A, Randell EW, et al. Endotypes of primary osteoarthritis identified by plasma metabolomics analysis. Rheumatology. 2021 Jun 18;60(6):2735–44.

514. Gallois A, Mefford J, Ko A, Vaysse A, Julienne H, Ala-Korpela M, et al. A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. Nat Commun. 2019 Oct 21;10(1):4788.

515. Assfalg M, Bertini I, Colangiuli D, Luchinat C, Schäfer H, Schütz B, et al. Evidence of different metabolic phenotypes in humans. Proc Natl Acad Sci U S A. 2008 Feb 5;105(5):1420–4.

516. Zhai G, Cicuttini F, Srikanth V, Cooley H, Ding C, Jones G. Factors associated with hip cartilage volume measured by magnetic resonance imaging: The Tasmanian Older Adult Cohort Study. Arthritis & Rheumatism. 2005;52(4):1069–76.

517. Fess, Elaine, Moran, Christine. American Society of Hand Therapists Clinical Assessment Recommendations. 1981.

518. Cervo MM, Shivappa N, Hebert JR, Oddy WH, Winzenberg T, Balogun S, et al. Longitudinal associations between dietary inflammatory index and musculoskeletal health in community-dwelling older adults. Clin Nutr. 2020 Feb;39(2):516–23.

519. El-gohary TM, Abd Elkader SM, Al-shenqiti AM, Ibrahim MI. Assessment of hand-grip and key-pinch strength at three arm positions among healthy college students: Dominant versus non-dominant hand. Journal of Taibah University Medical Sciences. 2019 Dec;14(6):566–71.

520. Martín-San Agustín R, Medina-Mirapeix F, Casaña-Granell J, García-Vidal JA, Lillo-Navarro C, Benítez-Martínez JC. Tensiomyographical responsiveness to peripheral fatigue in quadriceps femoris. PeerJ. 2020 Feb 28;8:e8674.

521. Scott D, Blizzard L, Fell J, Jones G. Statin therapy, muscle function and falls risk in community-dwelling older adults. QJM. 2009 Sep 1;102(9):625–33.

522. Pinheiro J, Bates D, DebRoy S, Sarkar D. Linear and Nonlinear Mixed Effects Models. 2012;R pakage version 3.1-152.

523. G. Zhai, M. Liu, P. Rahman, A. Furey. Genetic predictors for non-responders to total joint replacement in osteoarthritis patients. Orthopaedic Proceedings. 2020 Jul 1.

524. Bassey EJ, Harries UJ. Normal values for handgrip strength in 920 men and women aged over 65 years, and longitudinal changes over 4 years in 620 survivors. Clin Sci (Lond). 1993 Mar;84(3):331–7.

525. Martin JA, Ramsay J, Hughes C, Peters DM, Edwards MG. Age and grip strength predict hand dexterity in adults. PLoS One. 2015;10(2):e0117598.

526. Skelton DA, Greig CA, Davies JM, Young A. Strength, power and related functional ability of healthy people aged 65-89 years. Age Ageing. 1994 Sep;23(5):371–7.

527. Hughes VA, Frontera WR, Wood M, Evans WJ, Dallal GE, Roubenoff R, et al. Longitudinal muscle strength changes in older adults: influence of muscle mass, physical activity, and health. J Gerontol A Biol Sci Med Sci. 2001 May;56(5):B209-217.

528. Hägg GM, Milerad E. Forearm extensor and flexor muscle exertion during simulated gripping work -- an electromyographic study. Clin Biomech (Bristol, Avon). 1997 Jan;12(1):39–43.

529. Li ZM. The influence of wrist position on individual finger forces during forceful grip. J Hand Surg Am. 2002 Sep;27(5):886–96.

530. Li ZM, Zatsiorsky VM, Latash ML. The effect of finger extensor mechanism on the flexor force during isometric tasks. J Biomech. 2001 Aug;34(8):1097–102.

531. Schieber MH, Santello M. Hand function: peripheral and central constraints on performance. J Appl Physiol (1985). 2004 Jun;96(6):2293–300.

532. Jarque-Bou NJ, Sancho-Bru JL, Vergara M. A Systematic Review of EMG Applications for the Characterization of Forearm and Hand Muscle Activity during Activities of Daily Living: Results, Challenges, and Open Issues. Sensors (Basel). 2021 Apr 26;21(9):3035.

533. Bergquist R, Weber M, Schwenk M, Ulseth S, Helbostad JL, Vereijken B, et al. Performance-based clinical tests of balance and muscle strength used in young seniors: a systematic literature review. BMC Geriatr. 2019 Jan 9;19(1):9.

534. Obayashi K, Saeki K, Maegawa T, Sakai T, Kitagawa M, Otaki N, et al. Association of Serum Asymmetric Dimethylarginine With Muscle Strength and Gait Speed: A Cross-Sectional Study of the HEIJO-KYO Cohort. J Bone Miner Res. 2016 May;31(5):1107–13.

535. Kunz HE, Dorschner JM, Berent TE, Meyer T, Wang X, Jatoi A, et al. Methylarginine metabolites are associated with attenuated muscle protein synthesis in cancer-associated muscle wasting. J Biol Chem. 2020 Dec 18;295(51):17441–59.

536. Rawal N, Rajpurohit R, Lischwe MA, Williams KR, Paik WK, Kim S. Structural specificity of substrate for S-adenosylmethionine:protein arginine N-methyltransferases. Biochim Biophys Acta. 1995 Apr 5;1248(1):11–8.

537. Bode-Böger SM, Scalera F, Ignarro LJ. The L-arginine paradox: Importance of the L-arginine/asymmetrical dimethylarginine ratio. Pharmacol Ther. 2007 Jun;114(3):295–306.

538. Lukjanenko L, Karaz S, Stuelsatz P, Gurriaran-Rodriguez U, Michaud J, Dammone G, et al. Aging Disrupts Muscle Stem Cell Function by Impairing Matricellular WISP1 Secretion from Fibro-Adipogenic Progenitors. Cell Stem Cell. 2019 Mar 7;24(3):433-446.e7.

539. Wang H, Zhang R, Wen S, McCafferty DM, Beck PL, MacNaughton WK. Nitric oxide increases Wnt-induced secreted protein-1 (WISP-1/CCN4) expression and function in colitis. J Mol Med (Berl). 2009 Apr;87(4):435–45.

540. Vallance P, Leone A, Calver A, Collier J, Moncada S. Accumulation of an endogenous inhibitor of nitric oxide synthesis in chronic renal failure. Lancet. 1992 Mar 7;339(8793):572–5.

541. Biferali B, Proietti D, Mozzetta C, Madaro L. Fibro-Adipogenic Progenitors Cross-Talk in Skeletal Muscle: The Social Network. Front Physiol. 2019;10:1074.

542. Almada AE, Wagers AJ. Molecular circuitry of stem cell fate in skeletal muscle regeneration, ageing and disease. Nat Rev Mol Cell Biol. 2016 May;17(5):267–79.

543. Blau HM, Cosgrove BD, Ho ATV. The central role of muscle stem cells in regenerative failure with aging. Nat Med. 2015 Aug;21(8):854–62.

544. Lukjanenko L, Jung MJ, Hegde N, Perruisseau-Carrier C, Migliavacca E, Rozo M, et al. Loss of fibronectin from the aged stem cell niche affects the regenerative capacity of skeletal muscle in mice. Nat Med. 2016 Aug;22(8):897–905.

545. Xu ZR, Zhang Q, Chen LF, Xu KY, Xia JY, Li SM, et al. Characteristics of hyperuricemia in older adults in China and possible associations with sarcopenia. Aging Med (Milton). 2018 Jun;1(1):23–34.

546. Macchi C, Molino-Lova R, Polcaro P, Guarducci L, Lauretani F, Cecchi F, et al. Higher circulating levels of uric acid are prospectively associated with better muscle function in older persons. Mech Ageing Dev. 2008 Sep;129(9):522–7.

547. Kanellis J, Kang DH. Uric acid as a mediator of endothelial dysfunction, inflammation, and vascular disease. Semin Nephrol. 2005 Jan;25(1):39–42.

548. Beavers KM, Beavers DP, Serra MC, Bowden RG, Wilson RL. Low relative skeletal muscle mass indicative of sarcopenia is associated with elevations in serum uric acid levels: findings from NHANES III. J Nutr Health Aging. 2009 Mar;13(3):177–82.

549. Mannion JJ, Memorial University of Newfoundland, editors. The Peopling of Newfoundland: essays in historical geography. St. John's, Nfld.: Institute of Social and Economic Research, Memorial University of Newfoundland; 1977. 289 p. (Social and economic papers).

550. Statistics Canada 2019–20 Departmental Results Report, Catalogue no:11-628-X, ISSN: 2368-1160.

551. Rahman P, Jones A, Curtis J, Bartlett S, Peddle L, Fernandez BA, et al. The Newfoundland population: a unique resource for genetic investigation of complex diseases. Human Molecular Genetics. 2003 Oct 15;12(suppl 2):R167–72.

552. Bear JC, Nemec TF, Kennedy JC, Marshall WH, Power AA, Kolonel VM, et al. Inbreeding in outport Newfoundland. Am J Med Genet. 1988 Mar;29(3):649–60.

553. Bear JC, Nemec TF, Kennedy JC, Marshall WH, Power AA, Kolonel VM, et al. Persistent genetic isolation in outport Newfoundland. Am J Med Genet. 1987 Aug;27(4):807–30.

554. Zhai G, Zhou J, Woods MO, Green JS, Parfrey P, Rahman P, et al. Genetic structure of the Newfoundland and Labrador population: founder effects modulate variability. Eur J Hum Genet. 2016 Jul;24(7):1063–70.

555. de Vries Y, Lwiwski N, Levitus M, Kuyt B, Israels SJ, Arwert F, et al. A Dutch Fanconi Anemia *FANCC* Founder Mutation in Canadian Manitoba Mennonites. Anemia. 2012;2012:1–6.

556. Strachan T, Read AP, Strachan T. Human molecular genetics. 4th ed. New York: Garland Science; 2011. 781 p.

557. Kristiansson K, Naukkarinen J, Peltonen L. Isolated populations and complex disease gene identification. Genome Biol. 2008;9(8):109.

558. Olufemi SE, Green JS, Manickam P, Guru SC, Agarwal SK, Kester MB, et al. Common ancestral mutation in the MEN1 gene is likely responsible for the prolactinoma variant of MEN1 (MEN1Burin) in four kindreds from Newfoundland. Hum Mutat. 1998;11(4):264–9.

559. Spirio L, Green J, Robertson J, Robertson M, Otterud B, Sheldon J, et al. The identical 5' splice-site acceptor mutation in five attenuated APC families from Newfoundland demonstrates a founder effect. Hum Genet. 1999 Nov;105(5):388–98.

560. Kopciuk KA, Choi YH, Parkhomenko E, Parfrey P, McLaughlin J, Green J, et al. Penetrance of HNPCC-related cancers in a retrolective cohort of 12 large Newfoundland families carrying a MSH2 founder mutation: an evaluation using modified segregation models. Hered Cancer Clin Pract. 2009 Oct 28;7(1):16.

561. Froggatt NJ, Green J, Brassett C, Evans DG, Bishop DT, Kolodner R, et al. A common MSH2 mutation in English and North American HNPCC families: origin, phenotypic expression, and sex specific differences in colorectal cancer. J Med Genet. 1999 Feb;36(2):97–102.

562. Grewal KK, Stefanelli MG, Meijer IA, Hand CK, Rouleau GA, Ives EJ. A founder effect in three large Newfoundland families with a novel clinically variable spastic ataxia and supranuclear gaze palsy. Am J Med Genet A. 2004 Dec 15;131(3):249–54.

563. Meijer IA, Hand CK, Grewal KK, Stefanelli MG, Ives EJ, Rouleau GA. A locus for autosomal dominant hereditary spastic ataxia, SAX1, maps to chromosome 12p13. Am J Hum Genet. 2002 Mar;70(3):763–9.

564. Kaurah P, MacMillan A, Boyd N, Senz J, De Luca A, Chun N, et al. Founder and recurrent CDH1 mutations in families with hereditary diffuse gastric cancer. JAMA. 2007 Jun 6;297(21):2360–72.

565. Xie YG, Zheng H, Leggo J, Scully MF, Lillicrap D. A founder factor VIII mutation, valine 2016 to alanine, in a population with an extraordinarily high prevalence of mild hemophilia A. Thromb Haemost. 2002 Jan;87(1):178–9.

566. Young TL, Ives E, Lynch E, Person R, Snook S, MacLaren L, et al. Non-syndromic progressive hearing loss DFNA38 is caused by heterozygous missense mutation in the Wolfram syndrome gene WFS1. Hum Mol Genet. 2001 Oct 15;10(22):2509–14.

567. Shekarabi M, Girard N, Rivière JB, Dion P, Houle M, Toulouse A, et al. Mutations in the nervous system--specific HSN2 exon of WNK1 cause hereditary sensory neuropathy type II. J Clin Invest. 2008 Jul;118(7):2496–505.

568. Lafreniere RG, MacDonald MLE, Dube MP, MacFarlane J, O'Driscoll M, Brais B, et al. Identification of a novel gene (HSN2) causing hereditary sensory and autonomic neuropathy type II through the Study of Canadian Genetic Isolates. Am J Hum Genet. 2004 May;74(5):1064–73.

569. Aksentijevich I, Masters SL, Ferguson PJ, Dancey P, Frenkel J, van Royen-Kerkhoff A, et al. An autoinflammatory disease with deficiency of the interleukin-1-receptor antagonist. N Engl J Med. 2009 Jun 4;360(23):2426–37.

570. Merner ND, Hodgkinson KA, Haywood AFM, Connors S, French VM, Drenckhahn JD, et al. Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene. Am J Hum Genet. 2008 Apr;82(4):809–21.

571. Leppert M, Baird L, Anderson KL, Otterud B, Lupski JR, Lewis RA. Bardet-Biedl syndrome is linked to DNA markers on chromosome 11q and is genetically heterogeneous. Nat Genet. 1994 May;7(1):108–12.

572. Aref-Eshghi E, Zhang Y, Hart D, Valdes AM, Furey A, Martin G, et al. SMAD3 is associated with the total burden of radiographic osteoarthritis: the Chingford study. PLoS One. 2014;9(5):e97786.

573. Aref-Eshghi E, Liu M, Razavi-Lopez SB, Hirasawa K, Harper PE, Martin G, et al. SMAD3 Is Upregulated in Human Osteoarthritic Cartilage Independent of the Promoter DNA Methylation. J Rheumatol. 2016 Feb;43(2):388–94.

574. Aref-Eshghi E, Liu M, Harper PE, Doré J, Martin G, Furey A, et al. Overexpression of MMP13 in human osteoarthritic cartilage is associated with the SMAD-independent TGF-β signalling pathway. Arthritis Res Ther. 2015 Sep 23;17:264.

575. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis & Rheumatism. 2001 Oct;45(5):453–61.

576. Davis AM, Perruccio AV, Lohmander LS. Minimally clinically important improvement: all non-responders are not really non-responders an illustration from total knee replacement. Osteoarthritis and Cartilage. 2012 May;20(5):364–7.

577. Zhai G, Aref-Eshghi E, Rahman P, Zhang H, Martin G, Furey A, et al. Attempt to replicate the published osteoarthritis-associated genetic variants in the Newfoundland & Labrador Population. Journal of Orthopedics & Rheumatology. 2014 Apr 23.

578. Zhang W, Likhodii S, Zhang Y, Aref-Eshghi E, Harper PE, Randell E, et al. Classification of osteoarthritis phenotypes by metabolomics analysis. BMJ Open. 2014 Nov;4(11):e006286.

579. Altman R, Alarcón G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis & Rheumatism. 1991 May;34(5):505–14.

580. Baecke J, Burema J, Frijters J. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. The American Journal of Clinical Nutrition. 1982 Nov;36(5):936–42.

581. Infinium Omni2.5-8 Kit. https://www.illumina.com/products/by-type/microarray-kits/infinium-omni25-8.html

582. infinium-global-diversity-array-data-sheet-m-gl-00153.pdf. https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/infinium-global-diversity-array-data-sheet-m-gl-00153/infinium-global-diversity-array-data-sheet-m-gl-00153.pdf

583. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. International Journal of Methods in Psychiatric Research. 2018;27(2):e1608.

584. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics. 2007 Sep;81(3):559–75.

585. Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT, et al. Quality Control Procedures for Genome-Wide Association Studies. Current Protocols in Human Genetics. 2011 Jan.

586. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaSci. 2015 Dec;4(1):7.

587. Marino AD, Mahmoud AA, Bose M, Bircan KO, Terpolovsky A, Bamunusinghe V, et al. A comparative analysis of current phasing and imputation software. Genomics; 2021 Nov.

588. Baldwin S. Compute Canada: Advancing Computational Research. J Phys: Conf Ser. 2012 Feb 9;341:012001.

589. Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable and integrative haplotype estimation. Nat Commun. 2019 Dec;10(1):5436.

590. Rubinacci S, Delaneau O, Marchini J. Genotype imputation using the Positional Burrows Wheeler Transform. PLOS Genetics. 2020 Nov 16;16(11):e1009049.

591. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021 Feb 16;10(2):giab008.

592. The 1000 Genomes Project Consortium, Corresponding authors, Auton A, Abecasis GR, Steering committee, Altshuler DM, et al. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68–74.

593. Delaneau O, Marchini J, The 1000 Genomes Project Consortium, McVean GA, Donnelly P, Lunter G, et al. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat Commun. 2014 Jun 13;5(1):3934.

594. Li N, Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics. 2003 Dec;165(4):2213–33.

595. Durbin R. Efficient haplotype matching and storage using the positional Burrows–Wheeler transform (PBWT). Bioinformatics. 2014 May 1;30(9):1266–72.

596. Browning BL, Browning SR. Improving the Accuracy and Efficiency of Identity-by-Descent Detection in Population Data. Genetics. 2013 Jun 1;194(2):459–71.

597. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016 Oct;48(10):1284–7.

598. Milet J, Courtin D, Garcia A, Perdry H. Mixed logistic regression in genome-wide association studies. BMC Bioinformatics. 2020 Dec;21(1):536.

599. Chen Z, Huang H, Ng HKT. Testing for association in case-control genome-wide association studies with shared controls. Stat Methods Med Res. 2016 Apr;25(2):954–67.

600. Devlin B, Roeder K. Genomic Control for Association Studies. Biometrics. 1999;55(4):997–1004.

601. Milet J, Courtin D, Garcia A, Perdry H. Mixed logistic regression in genome-wide association studies. BMC Bioinformatics. 2020 Nov 23;21(1):536.

602. Wang F, Liu M, Wang N, Luo J. G Protein-Coupled Receptors in Osteoarthritis. Front Endocrinol. 2022 Jan 28;12:808835.

603. Liu Y, Wang G, Cui T, Lv L. Adhesion GPR123 is an indicator for recurrence and prognosis in bladder cancer. Genes Genomics. 2021 Nov;43(11):1317–25.

604. Monfoulet LE, Philippe C, Mercier S, Coxam V, Wittrant Y. Featured Article: Deficiency of G-protein coupled receptor 40, a lipid-activated receptor, heightens *in vitro*- and *in vivo*-induced murine osteoarthritis. Exp Biol Med (Maywood). 2015 Jul;240(7):854–66.

605. Chen Y, Zhang D, Ho KW, Lin S, Suen WCW, Zhang H, et al. GPR120 is an important inflammatory regulator in the development of osteoarthritis. Arthritis Res Ther. 2018 Dec;20(1):163.

606. Gu J, Lin H, Zhang Y, Xu T, Wang T, Xue X, et al. Activation of GPR40 Suppresses AGE-Induced Reduction of Type II Collagen and Aggrecan in Human SW1353 Chondrocytes. DDDT. 2020 Jun;Volume 14:2371–9.

607. Xu Z, Ke T, Zhang Y, Fu C, He W. Agonism of GPR120 prevented IL-1β-induced reduction of extracellular matrix through SOX-9. Aging. 2020 Jun 24;12(12):12074–85.

608. Wang F, Ma L, Ding Y, He L, Chang M, Shan Y, et al. Fatty acid sensing GPCR (GPR84) signaling safeguards cartilage homeostasis and protects against osteoarthritis. Pharmacological Research. 2021 Feb;164:105406.

609. Pirozzi C, Francisco V, Guida FD, Gómez R, Lago F, Pino J, et al. Butyrate Modulates Inflammation in Chondrocytes via GPR43 Receptor. Cell Physiol Biochem. 2018;51(1):228–43.

610. Davenport AP, Scully CCG, de Graaf C, Brown AJH, Maguire JJ. Advances in therapeutic peptides targeting G protein-coupled receptors. Nat Rev Drug Discov. 2020 Jun;19(6):389–413.

611. Catania A, Gatti S, Colombo G, Lipton JM. Targeting Melanocortin Receptors as a Novel Strategy to Control Inflammation. Pharmacol Rev. 2004 Mar 1;56(1):1–29.

612. Koller G, Schlomann U, Golfi P, Ferdous T, Naus S, Bartsch JW. ADAM8/MS2/CD156, an emerging drug target in the treatment of inflammatory and invasive pathologies. Curr Pharm Des. 2009;15(20):2272–81.

613. Zack MD, Malfait AM, Skepner AP, Yates MP, Griggs DW, Hall T, et al. ADAM-8 isolated from human osteoarthritic chondrocytes cleaves fibronectin at Ala ^{271}. Arthritis Rheum. 2009 Sep;60(9):2704–13.

614. Weber S, Saftig P. Ectodomain shedding and ADAMs in development. Development. 2012 Oct;139(20):3693–709.

615. Duan B, Liu Y, Hu H, Shi FG, Liu YL, Xue H, et al. Notch1-ADAM8 positive feed-back loop regulates the degradation of chondrogenic extracellular matrix and osteoarthritis progression. Cell Commun Signal. 2019 Dec;17(1):134.

616. Mandelin J, Li TF, Hukkanen MVJ, Liljeström M, Chen ZK, Santavirta S, et al. Increased expression of a novel osteoclast-stimulating factor, ADAM8, in interface tissue around loosened hip prostheses. J Rheumatol. 2003 Sep;30(9):2033–8.

617. Monteagudo S, Lories RJ. A Notch in the joint that exacerbates osteoarthritis. Nat Rev Rheumatol. 2018 Oct;14(10):563–4.

618. Saito T, Tanaka S. Molecular mechanisms underlying osteoarthritis development: Notch and NF-κB. Arthritis Res Ther. 2017 Dec;19(1):94.

619. Kuyinu EL, Narayanan G, Nair LS, Laurencin CT. Animal models of osteoarthritis: classification, update, and measurement of outcomes. J Orthop Surg Res. 2016 Feb 2;11:19.

620. Hayashi N, Yokoyama N, Seki T, Azuma Y, Ohba T, Nishimoto T. RanBP1, a Ras-like nuclear G protein binding to Ran/TC4, inhibits RCC1 via Ran/TC4. Mol Gen Genet. 1995 Jun 25;247(6):661–9.

621. Fournier AE, GrandPre T, Strittmatter SM. Identification of a receptor mediating Nogo-66 inhibition of axonal regeneration. Nature. 2001 Jan 18;409(6818):341–6.

622. Li X, Ke Q, Li Y, Liu F, Zhu G, Li F. DGCR6L, a novel PAK4 interaction protein, regulates PAK4-mediated migration of human gastric cancer cell via LIMK1. Int J Biochem Cell Biol. 2010 Jan;42(1):70–9.

623. Dunham I, Shimizu N, Roe BA, Chissoe S, Hunt AR, Collins JE, et al. The DNA sequence of human chromosome 22. Nature. 1999 Dec 2;402(6761):489–95.

624. Huzé C, Bauché S, Richard P, Chevessier F, Goillot E, Gaudon K, et al. Identification of an agrin mutation that causes congenital myasthenia and affects synapse function. Am J Hum Genet. 2009 Aug;85(2):155–67.

625. Nicole S, Chaouch A, Torbergsen T, Bauché S, de Bruyckere E, Fontenille MJ, et al. Agrin mutations lead to a congenital myasthenic syndrome with distal muscle weakness and atrophy. Brain. 2014 Sep;137(Pt 9):2429–43.

626. Kim N, Stiegler AL, Cameron TO, Hallock PT, Gomez AM, Huang JH, et al. Lrp4 is a receptor for Agrin and forms a complex with MuSK. Cell. 2008 Oct 17;135(2):334–42.

627. Gee SH, Montanaro F, Lindenbaum MH, Carbonetto S. Dystroglycan-alpha, a dystrophin-associated glycoprotein, is a functional agrin receptor. Cell. 1994 Jun 3;77(5):675–86.

628. Gautam M, Noakes PG, Moscoso L, Rupp F, Scheller RH, Merlie JP, et al. Defective neuromuscular synaptogenesis in agrin-deficient mutant mice. Cell. 1996 May 17;85(4):525–35.

629. Caruso V, Lagerström MC, Olszewski PK, Fredriksson R, Schiöth HB. Synaptic changes induced by melanocortin signalling. Nat Rev Neurosci. 2014 Feb;15(2):98–110.

630. Engel AG. Congenital Myasthenic Syndromes in 2018. Curr Neurol Neurosci Rep. 2018 Jun 12;18(8):46.

631. Simpson MA, Hsu R, Keir LS, Hao J, Sivapalan G, Ernst LM, et al. Mutations in FAM20C are associated with lethal osteosclerotic bone dysplasia (Raine syndrome), highlighting a crucial molecule in bone development. Am J Hum Genet. 2007 Nov;81(5):906–12.

632. Li X, Han L, Nookaew I, Mannen E, Silva MJ, Almeida M, et al. Stimulation of Piezo1 by mechanical signals promotes bone anabolism. Elife. 2019 Oct 7;8:e49631.

633. Sun W, Chi S, Li Y, Ling S, Tan Y, Xu Y, et al. The mechanosensitive Piezo1 channel is required for bone formation. Elife. 2019 Jul 10;8:e47454.

634. Zhou T, Gao B, Fan Y, Liu Y, Feng S, Cong Q, et al. Piezo1/2 mediate mechanotransduction essential for bone formation through concerted activation of NFAT-YAP1-ß-catenin. Elife. 2020 Mar 18;9:e52779.

635. Wang L, You X, Lotinun S, Zhang L, Wu N, Zou W. Mechanical sensing protein PIEZO1 regulates bone homeostasis via osteoblast-osteoclast crosstalk. Nat Commun. 2020 Jan 15;11(1):282.

636. Hendrickx G, Fischer V, Liedert A, von Kroge S, Haffner-Luntzer M, Brylka L, et al. Piezo1 Inactivation in Chondrocytes Impairs Trabecular Bone Formation. J Bone Miner Res. 2021 Feb;36(2):369–84.

637. Dzamukova M, Brunner TM, Miotla-Zarebska J, Heinrich F, Brylka L, Mashreghi MF, et al. Mechanical forces couple bone matrix mineralization with inhibition of angiogenesis to limit adolescent bone growth. Nat Commun. 2022 Jun 1;13(1):3059.

638. Desseyn JL, Buisine MP, Porchet N, Aubert JP, Laine A. Genomic organization of the human mucin gene MUC5B. cDNA and genomic sequences upstream of the large central exon. J Biol Chem. 1998 Nov 13;273(46):30157–64.

639. Juge PA, Lee JS, Ebstein E, Furukawa H, Dobrinskikh E, Gazal S, et al. MUC5B Promoter Variant and Rheumatoid Arthritis with Interstitial Lung Disease. N Engl J Med. 2018 Dec 6;379(23):2209–19.

640. Dai Y, Wang W, Yu Y, Hu S. Rheumatoid arthritis-associated interstitial lung disease: an overview of epidemiology, pathogenesis and management. Clin Rheumatol. 2021 Apr;40(4):1211–20.

641. Juge PA, Solomon JJ, van Moorsel CHM, Garofoli R, Lee JS, Louis-Sydney F, et al. MUC5B promoter variant rs35705950 and rheumatoid arthritis associated interstitial lung disease survival and progression. Semin Arthritis Rheum. 2021 Oct;51(5):996–1004.

642. Palomäki A, FinnGen Rheumatology Clinical Expert Group, Palotie A, Koskela J, Eklund KK, Pirinen M, et al. Lifetime risk of rheumatoid arthritis-associated interstitial lung disease in MUC5B mutation carriers. Ann Rheum Dis. 2021 Dec;80(12):1530–6.

643. Burns K, Clatworthy J, Martin L, Martinon F, Plumpton C, Maschera B, et al. Tollip, a new component of the IL-1RI pathway, links IRAK to the IL-1 receptor. Nat Cell Biol. 2000 Jun;2(6):346–51.

644. Ye J, Xu B, Fan B, Zhang J, Yuan F, Chen Y, et al. Discovery of Selenocysteine as a Potential Nanomedicine Promotes Cartilage Regeneration With Enhanced Immune Response by Text Mining and Biomedical Databases. Front Pharmacol. 2020 Jul 24;11:1138.

645. Miura K, Miyoshi O, Yun K, Inazawa J, Miyamoto T, Hayashi H, et al. Repeat-directed isolation of a novel gene preferentially expressed from the maternal allele in human placenta. J Hum Genet. 1999;44(1):1–9.

646. Glaser C, Heinrich J, Koletzko B. Role of FADS1 and FADS2 polymorphisms in polyunsaturated fatty acid metabolism. Metabolism. 2010 Jul;59(7):993–9.

647. Ishikawa K, Nagase T, Suyama M, Miyajima N, Tanaka A, Kotani H, et al. Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. DNA Res. 1998 Jun 30;5(3):169–76.

648. Demirkan A, van Duijn CM, Ugocsai P, Isaacs A, Pramstaller PP, Liebisch G, et al. Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. PLoS Genet. 2012;8(2):e1002490.

649. Naik RR, Sotnikov SV, Diepold RP, Iurato S, Markt PO, Bultmann A, et al. Polymorphism in Tmem132d regulates expression and anxiety-related behavior through binding of RNA polymerase II complex. Transl Psychiatry. 2018 Jan 10;8(1):1.

650. Wang X, Jiang W, Luo S, Yang X, Wang C, Wang B, et al. The C. elegans homolog of human panic-disorder risk gene TMEM132D orchestrates neuronal morphogenesis through the WAVE-regulatory complex. Mol Brain. 2021 Mar 16;14(1):54.

651. Erhardt A, Czibere L, Roeske D, Lucae S, Unschuld PG, Ripke S, et al. TMEM132D, a new candidate for anxiety phenotypes: evidence from human and mouse studies. Mol Psychiatry. 2011 Jun;16(6):647–63.

652. Itoh Y, Kajita M, Kinoh H, Mori H, Okada A, Seiki M. Membrane type 4 matrix metalloproteinase (MT4-MMP, MMP-17) is a glycosylphosphatidylinositol-anchored proteinase. J Biol Chem. 1999 Nov 26;274(48):34260–6.

653. Puente XS, Pendás AM, Llano E, Velasco G, López-Otín C. Molecular cloning of a novel membrane-type matrix metalloproteinase from a human breast carcinoma. Cancer Res. 1996 Mar 1;56(5):944–9.

654. Davidson RK, Waters JG, Kevorkian L, Darrah C, Cooper A, Donell ST, et al. Expression profiling of metalloproteinases and their inhibitors in synovium and cartilage. Arthritis Res Ther. 2006;8(4):R124.

655. Nagao Y, Batanian JR, Clemente MF, Sly WS. Genomic organization of the human gene (CA5) and pseudogene for mitochondrial carbonic anhydrase V and their localization to chromosomes 16q and 16p. Genomics. 1995 Aug 10;28(3):477–84.

656. Birot A, Duret L, Bartholin L, Santalucia B, Tigaud I, Magaud J, et al. Identification and molecular analysis of BANP. Gene. 2000 Aug 8;253(2):189–96.

657. Zhang X, Krutchinsky A, Fukuda A, Chen W, Yamamura S, Chait BT, et al. MED1/TRAP220 exists predominantly in a TRAP/ Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription. Mol Cell. 2005 Jul 1;19(1):89–100.

658. Bräuer AU, Savaskan NE, Kühn H, Prehn S, Ninnemann O, Nitsch R. A new phospholipid phosphatase, PRG-1, is involved in axon growth and regenerative sprouting. Nat Neurosci. 2003 Jun;6(6):572–8.

659. Iversen LF, Kastrup JS, Bjørn SE, Rasmussen PB, Wiberg FC, Flodgaard HJ, et al. Structure of HBP, a multifunctional protein with a serine proteinase fold. Nat Struct Biol. 1997 Apr;4(4):265–8.

660. Wang X, Lei D, Ding J, Liu S, Tao L, Zhang F, et al. A DNA-Methylated Sight on Autoimmune Inflammation Network across RA, pSS, and SLE. J Immunol Res. 2018;2018:4390789.

661. Campanelli D, Melchior M, Fu Y, Nakata M, Shuman H, Nathan C, et al. Cloning of cDNA for proteinase 3: a serine protease, antibiotic, and autoantigen from human neutrophils. J Exp Med. 1990 Dec 1;172(6):1709–15.

662. Germeshausen M, Deerberg S, Peter Y, Reimer C, Kratz CP, Ballmaier M. The spectrum of ELANE mutations and their implications in severe congenital and cyclic neutropenia. Hum Mutat. 2013 Jun;34(6):905–14.

663. Zhang Y, Yang Y, Wang C, Wan S, Yao Z, Zhang Y, et al. Identification of Diagnostic Biomarkers of Osteoarthritis Based on Multi-Chip Integrated Analysis and Machine Learning. DNA and Cell Biology. 2020 Dec 1;39(12):2245–56.

664. Chestkov AV, Baka ID, Kost MV, Georgiev GP, Buchman VL. The d4 gene family in the human genome. Genomics. 1996 Aug 15;36(1):174–7.

665. Evers C, Paramasivam N, Hinderhofer K, Fischer C, Granzow M, Schmidt-Bacher A, et al. SIPA1L3 identified by linkage analysis and whole-exome sequencing as a novel gene for autosomal recessive congenital cataract. Eur J Hum Genet. 2015 Dec;23(12):1627–33.

666. Yamawaki K, Ito M, Machida H, Moriki N, Okamoto R, Isaka N, et al. Identification of human CPI-17, an inhibitory phosphoprotein for myosin phosphatase. Biochem Biophys Res Commun. 2001 Jul 27;285(4):1040–5.

667. Kominami K, DeMartino GN, Moomaw CR, Slaughter CA, Shimbara N, Fujimuro M, et al. Nin1p, a regulatory subunit of the 26S proteasome, is necessary for activation of Cdc28p kinase of Saccharomyces cerevisiae. EMBO J. 1995 Jul 3;14(13):3105–15.

668. Yuan WH, Xie QQ, Wang KP, Shen W, Feng XF, Liu Z, et al. Screening of osteoarthritis diagnostic markers based on immune-related genes and immune infiltration. Sci Rep. 2021 Mar 29;11(1):7032.

669. Li J, Chen C, Lei X, Wang Y, Chen C, He Q, et al. The NTSR1 gene modulates the association between hippocampal structure and working memory performance. Neuroimage. 2013 Jul 15;75:79–86.

670. Zagon IS, Verderame MF, Allen SS, McLaughlin PJ. Cloning, sequencing, chromosomal location, and function of cDNAs encoding an opioid growth factor receptor (OGFR) in humans. Brain Res. 2000 Feb 21;856(1–2):75–83.

671. Brewton RG, Wood BM, Ren ZX, Gong Y, Tiller GE, Warman ML, et al. Molecular cloning of the alpha 3 chain of human type IX collagen: linkage of the gene COL9A3 to chromosome 20q13.3. Genomics. 1995 Nov 20;30(2):329–36.

672. Ikeda T, Mabuchi A, Fukuda A, Kawakami A, Ryo Y, Yamamoto S, et al. Association analysis of single nucleotide polymorphisms in cartilage-specific collagen genes with knee and hip osteoarthritis in the Japanese population. J Bone Miner Res. 2002 Jul;17(7):1290–6.

673. Jakkula E, Melkoniemi M, Kiviranta I, Lohiniva J, Räinä SS, Perälä M, et al. The role of sequence variations within the genes encoding collagen II, IX and XI in non-syndromic, early-onset osteoarthritis. Osteoarthritis Cartilage. 2005 Jun;13(6):497–507.

674. Solovieva S, Lohiniva J, Leino-Arjas P, Raininko R, Luoma K, Ala-Kokko L, et al. COL9A3 gene polymorphism and obesity in intervertebral disc degeneration of the lumbar spine: evidence of gene-environment interaction. Spine (Phila Pa 1976). 2002 Dec 1;27(23):2691–6.

675. Solovieva S, Lohiniva J, Leino-Arjas P, Raininko R, Luoma K, Ala-Kokko L, et al. Intervertebral disc degeneration in relation to the COL9A3 and the IL-1ss gene polymorphisms. Eur Spine J. 2006 May;15(5):613–9.

676. Forghanifard MM, Naeimi Khorasanizadeh P, Abbaszadegan MR, Javdani Mallak A, Moghbeli M. Role of DIDO1 in Progression of Esophageal Squamous Cell Carcinoma. J Gastrointest Cancer. 2020 Mar;51(1):83–7.

677. Lu Y, Xie S, Zhang W, Zhang C, Gao C, Sun Q, et al. Twa1/Gid8 is a β-catenin nuclear retention factor in Wnt signaling and colorectal tumorigenesis. Cell Res. 2017 Dec;27(12):1422–40.

678. Flint J. GWAS. Curr Biol. 2013 Apr 8;23(7):R265-6. doi: 10.1016/j.cub.2013.01.040. PMID: 23578868.

679. Osterman MD, Kinzy TG, Bailey JNC. Polygenic Risk Scores. Curr Protoc. 2021 May;1(5):e126.

680. International HapMap Consortium. A haplotype map of the human genome. Nature. 2005 Oct 27;437(7063):1299–320.

681. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nat Rev Methods Primers. 2021 Dec;1(1):59.

682. Ji SG, Juran BD, Mucha S, Folseraas T, Jostins L, Melum E, et al. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. Nat Genet. 2017 Feb;49(2):269–73.

683. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet. 2015 Apr;47(4):373–80.

684. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747–53.

685. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012 Jan 13;90(1):7–24.

686. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. J Hum Genet. 2014 Jan;59(1):5–15.

687. Sanger F, Brownlee GG, Barrell BG. A two-dimensional fractionation procedure for radioactive nucleotides. J Mol Biol. 1965 Sep;13(2):373–98.

688. Min Jou W, Haegeman G, Ysebaert M, Fiers W. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. Nature. 1972 May 12;237(5350):82–8.

689. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977 Dec;74(12):5463–7.

690. Behjati S, Tarpey PS. What is next generation sequencing? Arch Dis Child Educ Pract Ed. 2013 Dec;98(6):236–8.

691. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. Hum Immunol. 2021 Nov;82(11):801–11.

692. Sun Y, Ruivenkamp CAL, Hoffer MJV, Vrijenhoek T, Kriek M, van Asperen CJ, et al. Next-generation diagnostics: gene panel, exome, or whole genome? Hum Mutat. 2015 Jun;36(6):648–55.

693. Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A, et al. Large-scale whole-genome sequencing of the Icelandic population. Nat Genet. 2015 May;47(5):435–44.

694. Styrkarsdottir U, Lund SH, Thorleifsson G, Zink F, Stefansson OA, Sigurdsson JK, et al. Meta-analysis of Icelandic and UK data sets identifies missense variants in SMO, IL11, COL11A1 and 13 more new loci associated with osteoarthritis. Nat Genet. 2018 Dec;50(12):1681–7.

695. Nakazato T, Ohta T, Bono H. Experimental Design-Based Functional Mining and Characterization of High-Throughput Sequencing Data in the Sequence Read Archive. Aziz RK, editor. PLoS ONE. 2013 Oct 22;8(10):e77910.

696. Auwera G van der, O'Connor BD. Genomics in the cloud: using Docker, GATK, and WDL in Terra. First edition. Sebastopol, CA: O'Reilly Media; 2020. 467 p.

697. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. Genomics; 2017 Nov.

698. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep;38(16):e164.

699. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016 Jan 4;44(D1):D733-745.

700. Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2021 update. Nucleic Acids Res. 2021 Jan 8;49(D1):D1046–57.

701. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. Nucleic Acids Res. 2022 Jan 7;50(D1):D988–95.

702. Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, et al. GENCODE 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D916–23.

703. Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biol. 2006;7 Suppl 1:S12.1-14.

704. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001 Jan 1;29(1):308–11.

705. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. Nucleic Acids Res. 2017 Jan 4;45(D1):D840–5.

706. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020 May 28;581(7809):434–43.

707. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Research. 2012 Jul 1;40(W1):W452–7.

708. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248–9.

709. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res. 2009 Sep;19(9):1553–61.

710. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010 Aug;7(8):575–6.

711. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Research. 2011 Sep;39(17):e118–e118.

712. Shihab HA, Gough J, Cooper DN, Day INM, Gaunt TR. Predicting the functional consequences of cancer-associated amino acid substitutions. Bioinformatics. 2013 Jun 15;29(12):1504–10.

713. Cooper GM, Stone EA, Asimenos G, NISC Comparative Sequencing Program, Green ED, Batzoglou S, et al. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005 Jul;15(7):901–13.

714. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010 Jan;20(1):110–21.

715. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol. 2010 Dec 2;6(12):e1001025.

716. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014 Mar;46(3):310–5.

717. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Research. 2019 Jan 8;47(D1):D886–94.

718. Wong AK, Sealfon RSG, Theesfeld CL, Troyanskaya OG. Decoding disease: from genomes to networks to phenotypes. Nat Rev Genet. 2021 Dec;22(12):774–90.

719. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Research. 2018 Jan 4;46(D1):D794–801.

720. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010 Oct;28(10):1045–8.

721. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc. 2015 Oct;10(10):1556–66.

722. Deltas C. Digenic inheritance and genetic modifiers. Clin Genet. 2018 Mar;93(3):429–38.

723. Fan R, Lo SH. A robust model-free approach for rare variants association studies incorporating gene-gene and gene-environmental interactions. PLoS One. 2013;8(12):e83057.

724. Kwon M, Leem S, Yoon J, Park T. GxGrare: gene-gene interaction analysis method for rare variants from high-throughput sequencing data. BMC Syst Biol. 2018 Mar 19;12(Suppl 2):19.

725. Zhao J, Zhu Y, Xiong M. Genome-wide gene-gene interaction analysis for next-generation sequencing. Eur J Hum Genet. 2016 Mar;24(3):421–8.

726. Kerner G, Bouaziz M, Cobat A, Bigio B, Timberlake AT, Bustamante J, et al. A genome-wide case-only test for the detection of digenic inheritance in human exomes. Proc Natl Acad Sci USA. 2020 Aug 11;117(32):19367–75.

727. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012; Available from: https://arxiv.org/abs/1207.3907

728. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w ^{1118} ; iso-2; iso-3. Fly. 2012 Apr;6(2):80–92.

729. Zheng X, Gogarten SM, Lawrence M, Stilp A, Conomos MP, Weir BS, et al. SeqArray—a storage-efficient high-performance data format for WGS variant calls. Birol I, editor. Bioinformatics. 2017 Aug 1;33(15):2251–7.

730. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012 Dec 1;28(24):3326–8.

731. Yang M, Li L, Wang J, Gao T, Sun Y, Li H, et al. Heterogeneous nuclear ribonucleoproteins (hnRNPs) and human transformer-2-beta1 (hTra2-beta1)-regulated estrogen receptor-alpha improves prognosis of endometrial cancer. Eur J Gynaecol Oncol. 2014;35(6):701–7.

732. Chen WC, Wang CY, Hung YH, Weng TY, Yen MC, Lai MD. Systematic Analysis of Gene Expression Alterations and Clinical Outcomes for Long-Chain Acyl-Coenzyme A Synthetase Family in Cancer. PLoS One. 2016;11(5):e0155660.

733. Liu M, Zhou X, Liu J, Lu C, Zhang G, Zhang J, et al. Predictive Biomarkers of Dicycloplatin Resistance or Susceptibility in Prostate Cancer. Front Genet. 2021;12:669605.

734. Gao Y, Zhang X, Wang T, Zhang Y, Wang Q, Hu Y. HNRNPCL1, PRAMEF1, CFAP74, and DFFB: Common Potential Biomarkers for Sporadic and Suspected Lynch Syndrome Endometrial Cancer. Cancer Manag Res. 2020;12:11231–41.

735. in the frame of GSF (Groupe Sarcoïdose France), Calender A, Rollat Farnier PA, Buisson A, Pinson S, Bentaher A, et al. Whole exome sequencing in three families segregating a pediatric case of sarcoidosis. BMC Med Genomics. 2018 Dec;11(1):23.

736. Usardi A, Iyer K, Sigoillot SM, Dusonchet A, Selimi F. The immunoglobulin-like superfamily member IGSF3 is a developmentally regulated protein that controls neuronal morphogenesis. Dev Neurobiol. 2017 Jan;77(1):75–92.

737. Coulombe P, Nassar J, Peiffer I, Stanojcic S, Sterkers Y, Delamarre A, et al. The ORC ubiquitin ligase OBI1 promotes DNA replication origin firing. Nat Commun. 2019 Dec;10(1):2426.

738. Querques F, D'Agostino A, Cozzolino C, Cozzuto L, Lombardo B, Leggiero E, et al. Identification of a Novel Transcription Factor Required for Osteogenic Differentiation of Mesenchymal Stem Cells. Stem Cells and Development. 2019 Mar 15;28(6):370–83.

739. Diao H, Zhu P, Dai Y, Chen W. Identification of 11 potentially relevant gene mutations involved in growth retardation, intellectual disability, joint contracture, and hepatopathy. Medicine (Baltimore). 2018 Nov;97(46):e13117.

740. Alkhateeb AM, Salman DS, Al-Hayk KA. Exome Sequencing Analysis of Familial Cases of Multiple Sclerosis and a Monozygotic Discordant Twin. Arab J Sci Eng. 2021 Jun;46(6):5421–7.

741. Proceedings of the 24th Paediatric Rheumatology European Society Congress: Part one: Athens, Greece. 14-17 September 2017. Pediatr Rheumatol. 2017 Sep;15(S2):64, s12969-017-0185–x.

742. Rowen L, Koop BF, Hood L. The complete 685-kilobase DNA sequence of the human beta T cell receptor locus. Science. 1996 Jun 21;272(5269):1755–62.

743. Liu J, Zhang H xin. A comprehensive study indicates PRSS1 gene is significantly associated with pancreatitis. Int J Med Sci. 2013;10(8):981–7.

744. Zheng Dong, , Jiucun Wang. Copy number variation and gout: the next frontier. Gout and Hyperuricemia. 2015;2(2):42–9.

745. Zorina-Lichtenwalter K, Meloto CB, Khoury S, Diatchenko L. Genetic predictors of human chronic pain conditions. Neuroscience. 2016 Dec;338:36–62.

746. Mogil JS. Pain genetics: past, present and future. Trends in Genetics. 2012 Jun;28(6):258–66.

747. Bandesh K, Bharadwaj D. Genetic variants entail type 2 diabetes as an innate immune disorder. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 2020 Sep;1868(9):140458.

748. Liu X, Nudel R, Thompson WK, Appadurai V, Schork AJ, Buil A, et al. Genetic factors underlying the bidirectional relationship between autoimmune and mental disorders – Findings from a Danish population-based study. Brain, Behavior, and Immunity. 2021 Jan;91:10–23.

749. Wang JG, Xu WD, Zhai WT, Li Y, Hu JW, Hu B, et al. Disorders in angiogenesis and redox pathways are main factors contributing to the progression of rheumatoid arthritis: A comparative proteomics study. Arthritis & Rheumatism. 2012 Apr;64(4):993–1004.

750. Kondo H, Shimomura I, Kishida K, Kuriyama H, Makino Y, Nishizawa H, et al. Human aquaporin adipose (AQPap) gene. Genomic structure, promoter analysis and functional mutation. Eur J Biochem. 2002 Apr;269(7):1814–26.

751. Mósca AF, de Almeida A, Wragg D, Martins AP, Sabir F, Leoni S, et al. Molecular Basis of Aquaporin-7 Permeability Regulation by pH. Cells. 2018 Nov 10;7(11):207.

752. Kuriyama H, Kawamoto S, Ishida N, Ohno I, Mita S, Matsuzawa Y, et al. Molecular cloning and expression of a novel human aquaporin from adipose tissue with glycerol permeability. Biochem Biophys Res Commun. 1997 Dec 8;241(1):53–8.

753. Funahashi T, Nagasawa A, Hibuse T, Maeda N. Impact of glycerol gateway molecule in adipocytes. Cell Mol Biol (Noisy-le-grand). 2006 Oct 30;52(7):40–5.

754. Nagahara M, Waguri-Nagaya Y, Yamagami T, Aoyama M, Tada T, Inoue K, et al. TNF-α-induced aquaporin 9 in synoviocytes from patients with OA and RA. Rheumatology. 2010 May;49(5):898–906.

755. Rodríguez A, Catalán V, Gómez-Ambrosi J, Frühbeck G. Aquaglyceroporins serve as metabolic gateways in adiposity and insulin resistance control. Cell Cycle. 2011 May 15;10(10):1548–56.

756. da Silva IV, Cardoso C, Méndez-Giménez L, Camoes SP, Frühbeck G, Rodríguez A, et al. Aquaporin-7 and aquaporin-12 modulate the inflammatory phenotype of endocrine pancreatic beta-cells. Arch Biochem Biophys. 2020 Sep 30;691:108481.

757. Zhang Y, Yu R, Zhang J, Feng E, Wang H, Li Z, et al. Identification of Key Genes and Pathways Associated with Differences of Subchondral Bone in Osteoarthritis. In Review; 2020 Aug.

758. Giguère V. To ERR in the estrogen pathway. Trends Endocrinol Metab. 2002 Jul;13(5):220–5.

759. Cho Y, Hazen BC, Russell AP, Kralli A. Peroxisome proliferator-activated receptor γ coactivator 1 (PGC-1)- and estrogen-related receptor (ERR)-induced regulator in muscle 1 (Perm1) is a tissue-specific regulator of oxidative capacity in skeletal muscle cells. J Biol Chem. 2013 Aug 30;288(35):25207–18.

760. Cho Y, Hazen BC, Gandra PG, Ward SR, Schenk S, Russell AP, et al. Perm1 enhances mitochondrial biogenesis, oxidative capacity, and fatigue resistance in adult skeletal muscle. FASEB J. 2016 Feb;30(2):674–87.

761. Kan J, Hu Y, Ge Y, Zhang W, Lu S, Zhao C, et al. Declined expressions of vast mitochondria-related genes represented by CYCS and transcription factor ESRRA in skeletal muscle aging. Bioengineered. 2021 Jan 1;12(1):3485–502.

762. Cui H, Moore J, Ashimi SS, Mason BL, Drawbridge JN, Han S, et al. Eating disorder predisposition is associated with ESRRA and HDAC4 mutations. J Clin Invest. 2013 Nov 1;123(11):4706–13.

763. Tang J, Liu T, Wen X, Zhou Z, Yan J, Gao J, et al. Estrogen-related receptors: novel potential regulators of osteoarthritis pathogenesis. Mol Med. 2021 Jan 15;27(1):5.

764. Rossjohn J, Gras S, Miles JJ, Turner SJ, Godfrey DI, McCluskey J. T cell antigen receptor recognition of antigen-presenting molecules. Annu Rev Immunol. 2015;33:169–200.

765. Brownlie RJ, Zamoyska R. T cell receptor signalling networks: branched, diversified and bounded. Nat Rev Immunol. 2013 Apr;13(4):257–69.

766. Nikolich-Zugich J, Slifka MK, Messaoudi I. The many important facets of T-cell repertoire diversity. Nat Rev Immunol. 2004 Feb;4(2):123–32.

767. Sakai H, Jinawath A, Yamaoka S, Yuasa Y. Upregulation of MUC6 mucin gene expression by NFkappaB and Sp factors. Biochem Biophys Res Commun. 2005 Aug 12;333(4):1254–60.

768. Hoffmann W, Jagla W. Cell type specific expression of secretory TFF peptides: colocalization with mucins and synthesis in the brain. Int Rev Cytol. 2002;213:147–81.

769. Ruchaud-Sparagano MH, Westley BR, May FEB. The trefoil protein TFF1 is bound to MUC5AC in human gastric mucosa. Cell Mol Life Sci. 2004 Aug;61(15):1946–54.

770. Thim L, Madsen F, Poulsen SS. Effect of trefoil factors on the viscoelastic properties of mucus gels. Eur J Clin Invest. 2002 Jul;32(7):519–27.

771. Popp J, Schicht M, Garreis F, Klinger P, Gelse K, Sesselmann S, et al. Human Synovia Contains Trefoil Factor Family (TFF) Peptides 1-3 Although Synovial Membrane Only Produces TFF3: Implications in Osteoarthritis and Rheumatoid Arthritis. Int J Mol Sci. 2019 Dec 3;20(23):6105.

772. Volin MV, Shahrara S, Haines GK, Woods JM, Koch AE. Expression of mucin 3 and mucin 5AC in arthritic synovial tissue. Arthritis Rheum. 2008 Jan;58(1):46–52.

773. Hamaguchi M, Kawahito Y, Ishino H, Takeuchi N, Tokunaga D, Hojo T, et al. Mucin from rheumatoid arthritis synovial fluid enhances interleukin-6 production by human peripheral blood mononuclear cells. Hum Immunol. 2011 Mar;72(3):241–8.

774. Gazzo AM, Daneels D, Cilia E, Bonduelle M, Abramowicz M, Van Dooren S, et al. DIDA: A curated and annotated digenic diseases database. Nucleic Acids Res. 2016 Jan 4;44(D1):D900-907.

775. Bondinas GP, Moustakas AK, Papadopoulos GK. The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. Immunogenetics. 2007 Jul;59(7):539–53.

776. Weyand CM. The Influence of HLA-DRB1 Genes on Disease Severity in Rheumatoid Arthritis. Ann Intern Med. 1992 Nov 15;117(10):801.

777. Kerlan-Candon S, Combe B, Vincent R, Clot J, Pinet V, Eliaou JF. HLA-DRB1 gene transcripts in rheumatoid arthritis. Clinical and Experimental Immunology. 2002 Jan 12;124(1):142–9.

778. Noble JA, Valdes AM. Genetics of the HLA region in the prediction of type 1 diabetes. Curr Diab Rep. 2011 Dec;11(6):533–42.

779. Yan L, Wang J, Zeng K. Association between HLA-DRB1 polymorphisms and pemphigus vulgaris: a meta-analysis. British Journal of Dermatology. 2012 Oct;167(4):768–77.

780. Spagnolo P, du Bois RM. Genetics of sarcoidosis. Clinics in Dermatology. 2007 May;25(3):242–9.

781. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet. 2012 Jan 29;44(3):291–6.

782. Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. Nat Rev Rheumatol. 2013 Mar;9(3):141–53.

783. Moos V, Menard J, Sieper J, Sparmann M, Müller B. Association of HLA-DRB1*02 with osteoarthritis in a cohort of 106 patients. Rheumatology (Oxford). 2002 Jun;41(6):666–9.

784. Kooshkaki O, Atabati E, Shayesteh M, Salmani F, Sarab GA. The Association Between Knee Osteoarthritis and HLA-DRB1*0101 in the East of Iran. Curr Rheumatol Rev. 2020;16(2):134–8.

785. Kools P, Van Imschoot G, van Roy F. Characterization of three novel human cadherin genes (CDH7, CDH19, and CDH20) clustered on chromosome 18q22-q23 and with high homology to chicken cadherin-7. Genomics. 2000 Sep 15;68(3):283–95.

786. Tariq M, Chishti MS, Ali G, Ahmad W. A novel locus for ectodermal dysplasia of hairs, nails and teeth type maps to chromosome 18q22.1-22.3. Ann Hum Genet. 2008 Jan;72(Pt 1):19–25.

787. Blaschuk OW, Rowlands TM. Cadherins as modulators of angiogenesis and the structural integrity of blood vessels. Cancer Metastasis Rev. 2000;19(1–2):1–5.

788. Li F, Wan B, Li XQ. Expression Profile and Prognostic Values of CDH Family Members in Lung Adenocarcinoma. Dis Markers. 2022;2022:9644466.

789. Saito M, Tucker DK, Kohlhorst D, Niessen CM, Kowalczyk AP. Classical and desmosomal cadherins at a glance. J Cell Sci. 2012 Jun 1;125(Pt 11):2547–52.

790. Yulis M, Kusters DHM, Nusrat A. Cadherins: cellular adhesive molecules serving as signalling mediators. J Physiol. 2018 Sep;596(17):3883–98.

791. Walsh DA. Angiogenesis and arthritis. Rheumatology (Oxford). 1999 Feb;38(2):103–12.

792. Niu J, Azfer A, Zhelyabovska O, Fatma S, Kolattukudy PE. Monocyte chemotactic protein (MCP)-1 promotes angiogenesis via a novel transcription factor, MCP-1-induced protein (MCPIP). J Biol Chem. 2008 May 23;283(21):14542–51.

793. Imhof H, Breitenseher M, Kainberger F, Trattnig S. Degenerative joint disease: cartilage or vascular disease? Skeletal Radiol. 1997 Jul;26(7):398–403.

794. Haywood L, McWilliams DF, Pearson CI, Gill SE, Ganesan A, Wilson D, et al. Inflammation and angiogenesis in osteoarthritis. Arthritis Rheum. 2003 Aug;48(8):2173–7.

795. Walsh DA, Bonnet CS, Turner EL, Wilson D, Situ M, McWilliams DF. Angiogenesis in the synovium and at the osteochondral junction in osteoarthritis. Osteoarthritis Cartilage. 2007 Jul;15(7):743–51.

796. Mapp PI, Avery PS, McWilliams DF, Bowyer J, Day C, Moores S, et al. Angiogenesis in two animal models of osteoarthritis. Osteoarthritis Cartilage. 2008 Jan;16(1):61–9.

797. Sun Y, Mauerhan DR, Steuerwald NM, Ingram J, Kneisl JS, Hanley EN. Expression of Phosphocitrate-Targeted Genes in Osteoarthritis Menisci. BioMed Research International. 2014;2014:1–17.

798. Deng Y, Wang X, Liu Z, Lv X, Ma B, Nie Q, et al. In Silico Identification of Key Genes and Immune Infiltration Characteristics in Epicardial Adipose Tissue from Patients with Coronary Artery Disease. Aihara K ichi, editor. BioMed Research International. 2022 Oct 29;2022:1–13.

799. Kruse R, Krantz J, Barker N, Coletta RL, Rafikov R, Luo M, et al. Characterization of the CLASP2 Protein Interaction Network Identifies SOGA1 as a Microtubule-Associated Protein. Mol Cell Proteomics. 2017 Oct;16(10):1718–35.

800. Kotoulas OB, Kalamidas SA, Kondomerkos DJ. Glycogen autophagy in glucose homeostasis. Pathol Res Pract. 2006;202(9):631–8.

801. Mortimore GE, Pösö AR. Intracellular protein catabolism and its control during nutrient deprivation and supply. Annu Rev Nutr. 1987;7:539–64.

802. Singh R, Kaushik S, Wang Y, Xiang Y, Novak I, Komatsu M, et al. Autophagy regulates lipid metabolism. Nature. 2009 Apr 30;458(7242):1131–5.

803. Cowherd RB, Asmar MM, Alderman JM, Alderman EA, Garland AL, Busby WH, et al. Adiponectin lowers glucose production by increasing SOGA. Am J Pathol. 2010 Oct;177(4):1936–45.

804. Yoshimura N, Muraki S, Oka H, Tanaka S, Kawaguchi H, Nakamura K, et al. Accumulation of metabolic risk factors such as overweight, hypertension, dyslipidaemia, and impaired glucose tolerance raises the risk of occurrence and progression of knee osteoarthritis: a 3-year follow-up of the ROAD study. Osteoarthritis Cartilage. 2012 Nov;20(11):1217–26.

805. Engström G, Gerhardsson de Verdier M, Rollof J, Nilsson PM, Lohmander LS. C-reactive protein, metabolic syndrome and incidence of severe hip and knee osteoarthritis. A population-based cohort study. Osteoarthritis Cartilage. 2009 Feb;17(2):168–73.

806. Perry J, Zhao Y. The CW domain, a structural module shared amongst vertebrates, vertebrate-infecting parasites and higher plants. Trends Biochem Sci. 2003 Nov;28(11):576–80.

807. Liggins AP, Cooper CDO, Lawrie CH, Brown PJ, Collins GP, Hatton CS, et al. MORC4, a novel member of the MORC family, is highly expressed in a subset of diffuse large B-cell lymphomas. Br J Haematol. 2007 Aug;138(4):479–86.

808. Söderman J, Norén E, Christiansson M, Bragde H, Thiébaut R, Hugot JP, et al. Analysis of single nucleotide polymorphisms in the region of CLDN2-MORC4 in relation to inflammatory bowel disease. World J Gastroenterol. 2013 Aug 14;19(30):4935–43.

809. Paliwal S, Bhaskar S, Nageshwar Reddy D, Rao GV, Thomas V, Singh SP, et al. Association Analysis of PRSS1-PRSS2 and CLDN2-MORC4 Variants in Nonalcoholic Chronic Pancreatitis Using Tropical Calcific Pancreatitis as Model. Pancreas. 2016 Sep;45(8):1153–7.

810. Li J, Akil O, Rouse SL, McLaughlin CW, Matthews IR, Lustig LR, et al. Deletion of Tmtc4 activates the unfolded protein response and causes postnatal hearing loss. J Clin Invest. 2018 Nov 1;128(11):5150–62.

811. Graham JB, Sunryd JC, Mathavan K, Weir E, Larsen ISB, Halim A, et al. Endoplasmic reticulum transmembrane protein TMTC3 contributes to O-mannosylation of E-cadherin, cellular adherence, and embryonic gastrulation. Mol Biol Cell. 2020 Feb 1;31(3):167–83.

812. Ma M, Huang DG, Liang X, Zhang L, Cheng S, Cheng B, et al. Integrating transcriptome-wide association study and mRNA expression profiling identifies novel genes associated with bone mineral density. Osteoporos Int. 2019 Jul;30(7):1521–8.

813. Kordeli E, Lambert S, Bennett V. Ankyrin. Journal of Biological Chemistry. 1995 Feb;270(5):2352–9.

814. Liu F. Analysis of differentially expressed genes in rheumatoid arthritis and osteoarthritis by integrated microarray analysis. J of Cellular Biochemistry. 2019 Aug;120(8):12653–64.

815. Long NP, Park S, Anh NH, Min JE, Yoon SJ, Kim HM, et al. Efficacy of Integrating a Novel 16-Gene Biomarker Panel and Intelligence Classifiers for Differential Diagnosis of Rheumatoid Arthritis and Osteoarthritis. J Clin Med. 2019 Jan 6;8(1):50.

816. Momozawa Y, Mizukami K. Unique roles of rare variants in the genetics of complex diseases in humans. J Hum Genet. 2021 Jan;66(1):11–23.

817. Chen W, Coombes BJ, Larson NB. Recent advances and challenges of rare variant association analysis in the biobank sequencing era. Front Genet. 2022 Oct 6;13:1014947.

818. Hamel MB, Toth M, Legedza A, Rosen MP. Joint replacement surgery in elderly patients with severe osteoarthritis of the hip or knee: decision making, postoperative recovery, and clinical outcomes. Arch Intern Med. 2008 Jul 14;168(13):1430–40.

819. Wood AM, Brock TM, Heil K, Holmes R, Weusten A. A Review on the Management of Hip and Knee Osteoarthritis. Int J Chronic Dis. 2013;2013:845015.

820. Canadian Institute for Health Information. "Hip and knee replacements in Canada, 2016–2017: Canadian joint replacement registry annual report." (2018).

821. Lovse L, Poitras S, Dobransky J, Huang A, Beaulé PE. Should the Pareto Principle Be Applied as a Cost Savings Method in Hip and Knee Arthroplasty? The Journal of Arthroplasty. 2019 Dec;34(12):2841–5.

822. Nilsdotter AK. Predictors of patient relevant outcome after total hip replacement for osteoarthritis: a prospective study. Annals of the Rheumatic Diseases. 2003 Oct 1;62(10):923–30.

823. Judge A, Javaid MK, Arden NK, Cushnaghan J, Reading I, Croft P, et al. Clinical tool to identify patients who are most likely to achieve long-term improvement in physical function after total hip arthroplasty. Arthritis Care Res. 2012 Jun;64(6):881–9.

824. Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. BMJ Open. 2012;2(1):e000435.

825. Weber M, Craiovan B, Woerner ML, Schwarz T, Grifka J, Renkawitz TF. Predictors of Outcome After Primary Total Joint Replacement. The Journal of Arthroplasty. 2018 Feb;33(2):431–5.

826. Jiang Y, Sanchez-Santos MT, Judge AD, Murray DW, Arden NK. Predictors of Patient-Reported Pain and Functional Outcomes Over 10 Years After Primary Total Knee Arthroplasty: A Prospective Cohort Study. The Journal of Arthroplasty. 2017 Jan;32(1):92-100.e2.

827. Garriga C, Sanchez-Santos MT, Judge A, Perneger T, Hannouche D, Lübbeke A, et al. Development of a model predicting non-satisfaction 1 year after primary total knee replacement in the UK and transportation to Switzerland. Sci Rep. 2018 Dec;8(1):3380.

828. Costello CA, Liu M, Furey A, Rahman P, Randell EW, Zhai G. Association Between Epidemiological Factors and Nonresponders to Total Joint Replacement Surgery in Primary Osteoarthritis Patients. The Journal of Arthroplasty. 2021 May;36(5):1502-1510.e5.

829. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007 Jun;447(7145):661–78.

830. Chesworth BM, Mahomed NN, Bourne RB, Davis AM. Willingness to go through surgery again validated the WOMAC clinically important difference from THR/TKR surgery. Journal of Clinical Epidemiology. 2008 Sep;61(9):907–18.

831. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI Initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis and Cartilage. 2004 May;12(5):389–99.

832. Caricasole A, Bettini E, Sala C, Roncarati R, Kobayashi N, Caldara F, et al. Molecular cloning and characterization of the human diacylglycerol kinase beta (DGKbeta) gene: alternative splicing generates DGKbeta isotypes with different properties. J Biol Chem. 2002 Feb 15;277(7):4790–6.

833. Lovšin N. Copy Number Variation and Osteoporosis. Curr Osteoporos Rep. 2023 Apr;21(2):167–72.

834. Almehmadi A, Ohyama Y, Kaku M, Alamoudi A, Husein D, Katafuchi M, et al. VWC2 Increases Bone Formation Through Inhibiting Activin Signaling. Calcif Tissue Int. 2018 Dec;103(6):663–74.

835. Lin YJ, Cheng CF, Wang CH, Liang WM, Tang CH, Tsai LP, et al. Genetic Architecture Associated With Familial Short Stature. The Journal of Clinical Endocrinology & Metabolism. 2020 Jun 1;105(6):1801–13.

836. Zhang Y, Xu YZ, Sun N, Liu JH, Chen FF, Guan XL, et al. Long noncoding RNA expression profile in fibroblast-like synoviocytes from patients with rheumatoid arthritis. Arthritis Res Ther. 2016 Oct 6;18(1):227.

837. Woolley DE, Tetlow LC. Observations on the microenvironmental nature of cartilage degradation in rheumatoid arthritis. Ann Rheum Dis. 1997 Mar;56(3):151–61.

838. Benito MJ, Veale DJ, FitzGerald O, van den Berg WB, Bresnihan B. Synovial tissue inflammation in early and late osteoarthritis. Ann Rheum Dis. 2005 Sep;64(9):1263–7.

839. Walsh DA, McWilliams DF, Turley MJ, Dixon MR, Fransès RE, Mapp PI, et al. Angiogenesis and nerve growth factor at the osteochondral junction in rheumatoid arthritis and osteoarthritis. Rheumatology (Oxford). 2010 Oct;49(10):1852–61.

840. Rosenbaum DM, Rasmussen SGF, Kobilka BK. The structure and function of G-protein-coupled receptors. Nature. 2009 May;459(7245):356–63.

841. Foord SM, Bonner TI, Neubig RR, Rosser EM, Pin JP, Davenport AP, et al. International Union of Pharmacology. XLVI. G Protein-Coupled Receptor List. Pharmacol Rev. 2005 Jun;57(2):279–88.

842. Lagerström MC, Schiöth HB. Structural diversity of G protein-coupled receptors and significance for drug discovery. Nat Rev Drug Discov. 2008 Apr;7(4):339–57.

843. Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB, Gloriam DE. Trends in GPCR drug discovery: new agents, targets and indications. Nat Rev Drug Discov. 2017 Dec;16(12):829–42.

844. Wang W, Guo DY, Lin YJ, Tao YX. Melanocortin Regulation of Inflammation. Front Endocrinol (Lausanne). 2019 Oct 9;10:683.

845. Mountjoy KG, Robbins LS, Mortrud MT, Cone RD. The Cloning of a Family of Genes That Encode the Melanocortin Receptors. Science. 1992 Aug 28;257(5074):1248–51.

846. Loram LC, Culp ME, Connolly-Strong EC, Sturgill-Koszycki S. Melanocortin peptides: potential targets in systemic lupus erythematosus. Inflammation. 2015 Feb;38(1):260–71.

847. Victor J. Hruby, Minying Cai, James P. Cain, Alexander V. Mayorov, Matthew M. Dedek, Devendra Trivedi. Design, Synthesis and Biological Evaluation of Ligands Selective for the Melanocortin-3 Receptor. curr top med chem. 2007 Jun 1;7(11):1107–19.

848. Abdel-Malek Z, Scott MC, Suzuki I, Tada A, Im S, Lamoreux L, et al. The Melanocortin-1 Receptor is a Key Regulator of Human Cutaneous Pigmentation: The Melanocortin-1 Receptor and Human Cutaneous Pigmentation. Pigment Cell Research. 2000 Jun;13:156–62.

849. Chan LF, Metherell LA, Clark AJL. Effects of melanocortins on adrenal gland physiology. European Journal of Pharmacology. 2011 Jun;660(1):171–80.

850. Magenis RE, Smith L, Nadeau JH, Johnson KR, Mountjoy KG, Cone RD. Mapping of the ACTH, MSH, and neural (MC3 and MC4) melanocortin receptors in the mouse and human. Mammalian Genome. 1994 Aug;5(8):503–8.

851. Mountjoy KG, Mortrud MT, Low MJ, Simerly RB, Cone RD. Localization of the melanocortin-4 receptor (MC4-R) in neuroendocrine and autonomic control circuits in the brain. Molecular Endocrinology. 1994 Oct;8(10):1298–308.

852. Lasaga M, Debeljuk L, Durand D, Scimonelli TN, Caruso C. Role of α-melanocyte stimulating hormone and melanocortin 4 receptor in brain inflammation. Peptides. 2008 Oct;29(10):1825–35.

853. Xu Y, Guan X, Zhou R, Gong R. Melanocortin 5 receptor signaling pathway in health and disease. Cell Mol Life Sci. 2020 Oct;77(19):3831–40.

854. Levitskaia NG, Kamenskiĭ AA. [Melanocortin system]. Usp Fiziol Nauk. 2009 Jan-Mar;40(1):44-65. Russian. PMID: 19326848.

855. Chowdhary BP, Gustavsson I, Wikberg JES, Chhajlani V. Localization of the human melanocortin-5 receptor gene (MC5R) to chromosome band 18p11.2 by fluorescence in situ hybridization. Cytogenet Genome Res. 1995;68(1–2):79–81.

856. Buggy JJ. Binding of alpha-melanocyte-stimulating hormone to its G-protein-coupled receptor on B-lymphocytes activates the Jak/STAT pathway. Biochem J. 1998 Apr 1;331 (Pt 1):211–6.

857. Taylor A, Namba K. In vitro induction of CD25+ CD4+ regulatory T cells by the neuropeptide alpha-melanocyte stimulating hormone (alpha-MSH). Immunol Cell Biol. 2001 Aug;79(4):358–67.

858. Kaneva MK, Kerrigan MJ, Grieco P, Curley GP, Locke IC, Getting SJ. Chondroprotective and anti-inflammatory role of melanocortin peptides in TNF-α activated human C-20/A4 chondrocytes: Melanocortins and chondrocytes. British Journal of Pharmacology. 2012 Sep;167(1):67–79.

859. Lorenz J, Seebach E, Hackmayer G, Greth C, Bauer RJ, Kleinschmidt K, et al. Melanocortin 1 Receptor-Signaling Deficiency Results in an Articular Cartilage Phenotype and Accelerates Pathogenesis of Surgically Induced Murine Osteoarthritis. Malaval L, editor. PLoS ONE. 2014 Sep 5;9(9):e105858.

860. Delaney A, Keighren M, Fleetwood-Walker SM, Jackson IJ. Involvement of the Melanocortin-1 Receptor in Acute Pain and Pain of Inflammatory but Not Neuropathic Origin. Baccei ML, editor. PLoS ONE. 2010 Sep 13;5(9):e12498.

861. Li ZX, Liu BW, He ZG, Xiang HB. Melanocortin-4 receptor regulation of pain. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2017 Oct;1863(10):2515–22.

862. Das UN. Molecular biochemical aspects of cancer. New York, NY: Humana Press; 2020. 254 p.

863. Li XF, Chen X, Bao J, Xu L, Zhang L, Huang C, et al. PTEN negatively regulates the expression of pro-inflammatory cytokines and chemokines of fibroblast-like synoviocytes in adjuvant-induced arthritis. Artificial Cells, Nanomedicine, and Biotechnology. 2019 Dec 4;47(1):3687–96.

864. Chen D, Liu D, Liu D, He M, Peng A, Xu J, et al. Rheumatoid Arthritis Fibroblast-like Synoviocyte Suppression Mediated by PTEN Involves Survivin Gene Silencing. Sci Rep. 2017 Mar 23;7(1):367.

865. Longobardi L, Li T, Tagliafierro L, Temple JD, Willcockson HH, Ye P, et al. Synovial Joints: from Development to Homeostasis. Curr Osteoporos Rep. 2015 Feb;13(1):41–51.

866. Mathiessen A, Conaghan PG. Synovitis in osteoarthritis: current understanding with therapeutic implications. Arthritis Research & Therapy. 2017 Feb 2;19(1):18.

867. Silverstein AM, Stefani RM, Sobczak E, Tong EL, Attur MG, Shah RP, et al. Toward understanding the role of cartilage particulates in synovial inflammation. Osteoarthritis Cartilage. 2017 Aug;25(8):1353–61.

868. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage. 2013 Sep;21(9):1145–53.

869. Bierma-Zeinstra SMA, Verhagen AP. Osteoarthritis subpopulations and implications for clinical trial design. Arthritis Res Ther. 2011 Apr 5;13(2):213.

870. Felson DT. Identifying different osteoarthritis phenotypes through epidemiology. Osteoarthritis Cartilage. 2010 May;18(5):601–4.

871. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. The Lancet. 2019 Apr;393(10182):1745–59.

872. Mobasheri A, Saarakkala S, Finnilä M, Karsdal MA, Bay-Jensen AC, Van Spil WE. Recent advances in understanding the phenotypes of osteoarthritis. F1000Res. 2019 Dec 12;8:2091.

873. Tiulpin A, Klein S, Bierma-Zeinstra SMA, Thevenot J, Rahtu E, Meurs J van, et al. Multimodal Machine Learning-based Knee Osteoarthritis Progression Prediction from Plain Radiographs and Clinical Data. Sci Rep. 2019 Dec 27;9(1):20038.

874. Hyrich KL. Real world data in rheumatology. Semin Arthritis Rheum. 2019 Dec;49(3S):S22–4.

875. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, Hip, and Knee. Arthritis Care Res. 2020 Feb;72(2):149–62.

876. Bannuru RR, Osani MC, Vaysbrot EE, Arden NK, Bennell K, Bierma-Zeinstra SMA, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage. 2019 Nov;27(11):1578–89.

877. Aletaha D. Precision medicine and management of rheumatoid arthritis. J Autoimmun. 2020 Jun;110:102405.

878. Dell'Isola A, Allan R, Smith SL, Marreiros SSP, Steultjens M. Identification of clinical phenotypes in knee osteoarthritis: a systematic review of the literature. BMC Musculoskelet Disord. 2016 Oct 12;17(1):425.

879. Driban JB, Harkey MS, Barbe MF, Ward RJ, MacKay JW, Davis JE, et al. Risk factors and the natural history of accelerated knee osteoarthritis: a narrative review. BMC Musculoskelet Disord. 2020 May 29;21(1):332.

880. Hochberg MC, Tive LA, Abramson SB, Vignon E, Verburg KM, West CR, et al. When Is Osteonecrosis Not Osteonecrosis?: Adjudication of Reported Serious Adverse Joint Events in the Tanezumab Clinical Development Program. Arthritis Rheumatol. 2016 Feb;68(2):382–91.

881. Idle JR, Gonzalez FJ. Metabolomics. Cell Metab. 2007 Nov;6(5):348–51.

882. Zhai G, Randell EW, Rahman P. Metabolomics of osteoarthritis: emerging novel markers and their potential clinical utility. Rheumatology. 2018;57(12):2087–95.

APPENDIX

Appendix A: A copy of the Western Ontario and McMasters Universities Osteoarthritis Index (WOMAC).

PATIENT NAME	DOB

WESTERN ONTARIO AND MCMASTER OSTEOARTHRITIS INDEX (WOMAC)

Please circle the appropriate rating for each item.

RATE YOUR PAIN WHEN...	NONE	SLIGHT	MODERATE	SEVERE	EXTREME	HOSPITAL USE ONLY
Walking	0	1	2	3	4	
Climbing stairs	0	1	2	3	4	
Sleeping at night	0	1	2	3	4	
Resting	0	1	2	3	4	
Standing	0	1	2	3	4	TOTAL
RATE YOUR STIFFNESS IN THE...	**NONE**	**SLIGHT**	**MODERATE**	**SEVERE**	**EXTREME**	**HOSPITAL USE ONLY**
Morning	0	1	2	3	4	
Evening	0	1	2	3	4	TOTAL
RATE YOUR DIFFICULTY WHEN...	**NONE**	**SLIGHT**	**MODERATE**	**SEVERE**	**EXTREME**	**HOSPITAL USE ONLY**
Descending stairs	0	1	2	3	4	
Ascending stairs	0	1	2	3	4	
Rising from sitting	0	1	2	3	4	
Standing	0	1	2	3	4	
Bending to floor	0	1	2	3	4	
Walking on even floor	0	1	2	3	4	
Getting in/out of car	0	1	2	3	4	
Going shopping	0	1	2	3	4	
Putting on socks	0	1	2	3	4	
Rising from bed	0	1	2	3	4	
Taking off socks	0	1	2	3	4	
Lying in bed	0	1	2	3	4	
Getting in/out of bath	0	1	2	3	4	
Sitting	0	1	2	3	4	
Getting on/off toilet	0	1	2	3	4	
Doing light domestic duties (cooking, dusting)	0	1	2	3	4	
Doing heavy domestic duties (moving furniture)	0	1	2	3	4	TOTAL

PATIENT SIGNATURE	DATE

Appendix B: A list of genetic variations that were found by far to be significantly associated with OA in 11 OA phenotypes. This table was adapted from the Boer *et al.* (2021).

Novel SNV	Position	Risk Allele	EAF	OR	95% CI	p-value	Annotation	Gene	OA phenotype
rs11588154	1:55301936	T	0.17	0.83	0.79–0.88	$6.08*10^{-10}$	intron	C1orf177	FingerOA
rs4411121	1:118757034	T	0.31	1.07	1.05–1.09	$2.16*10^{-11}$	intergenic	SPAG17	HipOA
rs1327123	1:184014593	C	0.35	0.91	0.89–0.93	$2.44*10^{-16}$	intergenic	TSEN15	THR
rs11588850	1:227927242	A	0.82	0.87	0.84–0.91	$3.53*10^{-10}$	intron	SNAP47	ThumbOA
rs74676797	2:633063	A	0.82	1.05	1.03–1.07	$6.39*10^{-10}$	intergenic	TMEM18	KneeHipOA
rs66989638	2:106689736	A	0.13	1.12	1.08–1.15	$3.31*10^{-11}$	intron	C2orf40	THR
rs2276749	3:11643465	T	0.05	0.86	0.82–0.90	$3.34*10^{-9}$	missense	VGLL4	THR
rs62242105	3:20630395	A	0.33	0.97	0.96–0.98	$2.93*10^{-9}$	intergenic	RNU6-815P	AllOA
rs781661531	3:187051013	T	0.999	0.11	0.05-0.21	$8.36*10^{-11}$	intergenic	RTP4	HipOA
rs747952496	3:188311659	A	$4*10^{-4}$	7.02	3.93–12.55	$4.91*10^{-11}$	intron	LPP	HipOA
rs9835230	3:189735461	A	0.24	1.07	1.04–1.09	$1.34*10^{-9}$	intron	LEPREL1	HipOA
rs201194999	4:66666895	T	0.3	0.88	0.85–0.92	$3.05*10^{-9}$	intergenic	RNU2-40P	AllOA
rs11729628	4:121584282	T	0.24	0.97	0.96–0.98	$4.74*10^{-9}$	intergenic	RP11-501E14.1	AllOA
rs75686861	4:145621328	A	0.09	1.12	1.08–1.16	$3.04*10^{-9}$	intron	HHIP	THR
rs2066928	5:30843787	A	0.48	0.96	0.95–0.97	$1.20*10^{-8}$	intergenic	RPL19P11	KneeOA
rs56132153	5:67825133	A	0.61	1.07	1.05–1.09	$3.80*10^{-9}$	intron	CTC-537E7.1	THR
rs1560080	5:115338732	A	0.83	0.91	0.88–0.94	$9.61*10^{-9}$	intron	AQPEP	HandOA
rs17615906	5:128018413	T	0.84	0.95	0.93–0.96	$3.76*10^{-11}$	intron	SLC27A6	KneeHipOA
rs10062749	5:141805088	T	0.27	1.08	1.6–1.11	$2.04*10^{-9}$	intron	AC005592.2	HandOA
rs9396861	6:18404133	A	0.61	1.13	1.09–1.17	$9.35*10^{-11}$	intron	RNF144B	FingerOA
rs2038740	6:35114542	T	0.72	0.94	0.93–0.96	$6.20*10^{-10}$	intron	TCP11	TJR
rs116934101	7:101775597	A	0.27	1.06	1.04–1.08	$7.12*10^{-9}$	intron	CUX1	TJR
rs12667224	7:114024316	A	0.52	0.97	0.96–0.98	$1.66*10^{-9}$	intron	FOXP2	AllOA
rs571734653	7:137143697	A	$3*10^{-4}$	6.03	3.30–11.03	$5.56*10^{-9}$	intron	DGKI	KneeHipOA
rs7787744	7:150521096	A	0.67	1.08	1.05–1.11	$1.29*10^{-9}$	upstream_gene	AOC1	TKR
rs76340814	9:98321412	A	0.05	0.89	0.86–0.92	$1.87*10^{-9}$	intergenic	RP11-332M4.1	TJR
rs79895530	9:110416422	T	0.13	0.88	0.85–0.91	$3.86*10^{-14}$	intergenic	RNU6-996P	THR

rs7862601	9:118343026	A	0.62	0.94	0.92–0.96	$6.19*10^{-9}$	intergenic	RP11-284G10.1	HipOA
rs10983775	9:120521100	T	0.54	0.95	0.93–0.97	$4.65*10^{-9}$	intergenic	RP11-281A20.2	HipOA
rs10465114	9:129917824	A	0.22	1.06	1.04–1.09	$9.04*10^{-9}$	intron	RALGPS1	HipOA
rs3740129	10:73767859	A	0.46	1.08	1.05–1.10	$1.70*10^{-11}$	Missense	CHST3	THR
rs10824456	10:78615458	C	0.58	0.95	0.94–0.97	$1.16*10^{-8}$	intergenic	KCNMA1	TJR
rs3993110	11:12794530	A	0.61	1.09	1.06–1.11	$3.75*10^{-11}$	intron	TEAD1	HandOA
rs1631174	11:47974373	A	0.34	1.04	1.03–1.05	$7.28*10^{-9}$	regulatory_region	PTPRJ	KneeHipOA
rs72979233	11:74355523	A	0.75	0.92	0.89–0.95	$2.52*10^{-9}$	intron	POLD3	TKR
rs10831475	11:95796907	A	0.81	1.08	1.05–1.10	$5.89*10^{-12}$	intron	MAML2	TJR
rs10842226	12:23959589	A	0.42	1.04	1.03–1.06	$4.68*10^{-10}$	intron	SOX5	KneeHipOA
rs7967762	12:48420214	T	0.16	1.11	1.07–1.15	$4.41*10^{-10}$	upstream_gene	RP1-228P16.4	TKR
rs1426371	12:108629780	A	0.27	0.95	0.93–0.97	$8.86*10^{-10}$	intron	WSCD2	KneeOA
rs58973023	13:42959133	A	0.49	1.06	1.04–1.08	$4.72*10^{-10}$	intergenic	FABP3P2	KneeOA
rs28929474	14:94844947	T	0.02	0.81	0.76–0.86	$1.06*10^{-10}$	Missense	SERPINA1	TJR
rs746239049	15:63067433	D	0.21	0.9	0.87–0.93	$8.19*10^{-12}$	intron	TLN2	THR
rs12914479	15:99174828	C	0.66	1.04	1.03–1.06	$7.12*10^{-9}$	intergenic	RP11-35O15.1	KneeOA
rs6500609	16:4515334	C	0.11	0.94	0.91–0.96	$5.16*10^{-9}$	intron	NMRAL1	KneeOA
rs227732	17:54769890	T	0.3	1.06	1.04–1.09	$1.61*10^{-9}$	intergenic	NOG	TJR
rs9908159	17:54841961	T	0.51	1.04	1.03–1.05	$4.44*10^{-11}$	intergenic	C17orf67	KneeHipOA
rs1039257158	18:77950448	T	$6*10^{-4}$	3.62	2.35–5.60	$6.56*10^{-9}$	intron	PARD6G	AllOA
rs551471509	19:9943264	T	0.999	0.18	0.10–0.32	$1.15*10^{-8}$	upstream_gene	CTD-623N2.11	KneeHipOA
rs8112559	19:46390455	C	0.89	1.13	1.09–1.18	$7.32*10^{-11}$	upstream_gene	IRF2BP1	HandOA
rs9981884	21:40585633	A	0.49	0.95	0.94–0.97	$7.93*10^{-9}$	intron	BRWD1	TJR
rs11705555	22:28206912	A	0.76	1.05	1.03–1.07	$2.99*10^{-9}$	regulatory_region	MN1	KneeOA
rs12160491	22:38195796	A	0.71	0.93	0.90–0.95	$1.28*10^{-10}$	intergenic	H1F0	THR
Previously reported									
rs11164653	1:103464210	T	0.41	0.92	0.91–0.94	$2.77*10^{-18}$	intron	COL11A1	HipOA
1:150214028	1:150214028	D	0.38	1.04	1.02–1.05	$8.58*10^{-10}$	intergenic	RNU2-17P/ANP32E	AllOA
rs10797923	1:183901966	T	0.69	1.05	1.04–1.07	$6.20*10^{-9}$	intron	COLGALT2	TJR
rs2605100	1:219644224	A	0.32	1.07	1.05–1.09	$4.49*10^{-15}$	intergenic	RP11-95P13.1	TJR
rs7581446	2:33423801	T	0.48	0.95	0.94–0.97	$4.87*10^{-11}$	intron	LTBP1	KneeHipOA
rs3771501	2:70717653	A	0.47	1.04	1.03–1.05	$4.05*10^{-15}$	intron	TGFA	AllOA

rs62182810	2:204387482	A	0.54	1.03	1.02–1.04	$3.82*10^{-9}$	intron	RAPH1	AllOA
rs3774354	3:52817675	A	0.37	1.1	1.07–1.12	$1.40*10^{-16}$	intron	ITIH1	THR
rs1530586	4:1760927	T	0.8	1.09	1.06–1.11	$3.34*10^{-14}$	regulatory_ region	TACC3	TJR
rs1913707	4:13039440	A	0.6	1.09	1.06–1.11	$1.23*10^{-13}$	intergenic	RNU6-962P	THR
rs13107325	4:103188709	T	0.07	1.08	1.06–1.10	$3.25*10^{-17}$	missense	SLC39A8	AllOA
rs3884606	5:170871074	A	0.52	0.96	0.95–0.97	$8.96*10^{-10}$	intron	FGF18	KneeHipOA
rs79220007	6:26098474	T	0.93	0.9	0.87–0.93	$2.22*10^{-9}$	3_prime_UTR	HFE	HipOA
rs2856821	6:33046742	T	0.79	1.05	1.03–1.06	$5.71*10^{-9}$	intron	HLA-DPA1	KneeHipOA
rs17288390	6:45384018	T	0.65	0.92	0.90–0.94	$9.16*10^{-13}$	intron	RUNX2	THR
rs9475400	6:55638258	T	0.1	1.15	1.10–1.19	$1.73*10^{-13}$	intron	BMP5	THR
rs12209223	6:76164589	A	0.11	1.22	1.18–1.26	$1.92*10^{-29}$	intron	FILIP1	THR
rs111844273	7:18436337	A	0.02	1.26	1.18–1.34	$1.05*10^{-12}$	intron	HDAC9	HipOA
rs143083812	7:128843410	T	$1.1*10^{-3}$	3.3	2.34–4.66	$1.11*10^{-11}$	missense	SMO	THR
rs11984666	8:130730280	A	0.2	0.9	0.87–0.92	$1.69*10^{-15}$	intergenic	RP11-274M4.1	THR
rs10974438	9:4291928	A	0.65	1.04	1.03–1.06	$7.39*10^{-11}$	intron	GLIS3	KneeHipOA
rs72760655	9:116916214	A	0.33	1.05	1.03–1.06	$5.97*10^{-13}$	upstream_gene	COL27A1	KneeHipOA
rs1330349	9:117840742	C	0.59	1.1	1.07–1.12	$6.47*10^{-17}$	intron	TNC	THR
rs1321917	9:119324929	C	0.41	1.1	1.08–1.13	$9.87*10^{-19}$	intron	ASTN2	THR
rs62578126	9:129375338	T	0.37	0.92	0.90–0.94	$1.39*10^{-12}$	intron	RP11-123K19.1	THR
rs1517572	11:28829882	A	0.41	1.04	1.03–1.05	$6.79*10^{-10}$	intron	RP11-115J23.1	KneeHipOA
rs67924081	11:65342981	A	0.74	1.1	1.07–1.12	$2.14*10^{-13}$	upstream_gene	EHBP1L1	THR
rs34560402	11:66872320	T	0.06	0.86	0.82–0.90	$2.64*10^{-10}$	intergenic	KDM2A	THR
rs1149620	11:76506572	A	0.44	0.96	0.95–0.97	$2.87*10^{-9}$	intron	TSKU	KneeHipOA
rs7294636	12:15054016	A	0.37	1.16	1.12–1.20	$2.99*10^{-16}$	intron	C12orf60	FingerOA
rs10843013	12:28025196	A	0.78	0.86	0.84–0.88	$2.53*10^{-30}$	intergenic	RP11-993B23.1	THR
rs17120227	12:59289349	T	0.07	1.17	1.12–1.22	$7.21*10^{-13}$	intron	LRIG3	THR
rs7953280	12:94136009	C	0.5	1.04	1.03–1.06	$4.84*10^{-12}$	intron	CRADD	KneeHipOA
rs753350451	12:123732769	D	0.2	0.93	0.91–0.95	$3.36*10^{-10}$	intron	C12orf65	KneeOA
rs1809889	12:124801226	T	0.28	1.07	1.05–1.09	$5.70*10^{-14}$	downstream_ gene	FAM101A	TJR
rs4380013	15:50759428	A	0.19	1.06	1.04–1.08	$8.73*10^{-10}$	intron	USP8	KneeOA
rs11071366	15:58334244	A	0.61	0.9	0.88–0.92	$4.88*10^{-17}$	intron	ALDH1A2	HandOA
rs12908498	15:67366488	C	0.54	1.08	1.06–1.10	$1.85*10^{-16}$	intron	SMAD3	HipOA

rs9940278	16:53800200	T	0.43	1.06	1.04–1.07	$1.45*10^{-18}$	intron	FTO	KneeHipOA
rs34195470	16:69955690	A	0.45	0.95	0.94–0.96	$3.13*10^{-13}$	intron	WWP2	KneeOA
rs216175	17:2167690	A	0.83	1.04	1.03–1.06	$2.74*10^{-12}$	intron	SMG6	AllOA
rs7212908	17:59654593	A	0.8	0.91	0.89–0.94	$1.95*10^{-11}$	intergenic	NACA2	THR
rs2716212	17:67503653	A	0.62	0.93	0.91–0.95	$3.56*10^{-10}$	intron	MAP2K6	THR
rs10405617	19:10752968	A	0.32	1.03	1.02–1.04	$9.33*10^{-11}$	intron	SLC44A2	AllOA
rs75621460	19:41833784	A	0.03	1.21	1.14–1.28	$2.72*10^{-10}$	intron	TGFB1	TJR
rs4252548	19:55879672	T	0.02	1.39	1.29–1.49	$2.49*10^{-19}$	Missense	IL11	THR
rs143384	20:34025756	A	0.59	1.07	1.06–1.09	$1.01*10^{-23}$	5_prime_UTR	GDF5	KneeOA
rs9981408	21:40017446	T	0.23	1.1	1.07–1.12	$2.21*10^{-12}$	intron	ERG	THR
Female-specific									
rs116112221	2:59439973	T	$6.1*10^{-3}$	1.95	1.58–2.41	$4.61*10^{-10}$	upstream_gene	FANCL	THR
rs10282983	8:69590554	T	0.22	1.15	1.11–1.19	$2.21*10^{-14}$	intron	C8orf34	THR
rs10453201	9:34050345	T	0.22	1.05	1.03–1.06	$1.05*10^{-8}$	upstream_gene	UBAP2	AllOA
Early-onset									
rs148693048	8:24598320	T	0.001	6.26	3.26–12.00	$3.37*10^{-8}$	intron	NEFM	AllOA

Appendix C: A copy of the updated study ethical approval that was acquired from the Health Research Ethics Authority of NL with the reference number 11.311.

From: "do-not-reply-mun@researchservicesoffice.com" <do-not-reply-mun@researchservicesoffice.com>
Date: Tuesday, June 13, 2023 at 3:41 PM
To: "Zhai, Guangju" <Guangju.Zhai@med.mun.ca>
Cc: Hreaadministrator <administrator@hrea.ca>
Subject: HREB - Approval of Ethics Renewal 20161599

Researcher Portal File #: 20161599

Dear Dr. Guangju Zhai:

This e-mail serves as notification that your ethics renewal for study HREB # 2011.311 – Genetics Newfoundland Osteoarthritis Study (NFOA) – has been **approved**. Please log in to the Researcher Portal to view the approved event.

Ethics approval for this project has been granted for a period of twelve months effective from August 3, 2023 to August 3, 2024.

Please note, it is the responsibility of the Principal Investigator (PI) to ensure that the Ethics Renewal form is submitted prior to the renewal date each year. Though the Research Ethics Office makes every effort to remind the PI of this responsibility, the PI may not receive a reminder. The Ethics Renewal form can be found on the Researcher Portal as an "Event".

The ethics renewal was reviewed by the Health Research Ethics Board at their meeting dated June 13, 2023.

Thank you,

Research Ethics Office
Health Research Ethics Authority
760 Topsail Road
Mount Pearl, NL A1N 3J5
(e) info@hrea.ca
(t) 709-864-8871
(f) 709-864-8870
(w) www.hrea.ca

Appendix D: A copy of the consent taken from the NFOAS participants to take a part in the research study.

MEMORIAL UNIVERSITY
Faculty of Medicine
Discipline of Genetics
Health Sciences Centre
St. John's, NL Canada A1B 3V6
Tel: 709 777-6807 Fax:709 777-7497
www.med.mun.ca

Consent to Take Part in Research

TITLE: Newfoundland Osteoarthritis Study (NFOAS)

INVESTIGATOR(S): Drs. G. Zhai, A. Furey, G. Martin, P. Rahman, R. Green, and G. Sun.

You have been invited to take part in a research study. It is up to you to decide whether to be in the study or not. Before you decide, you need to understand what the study is for, what risks you might take and what benefits you might receive. This consent form explains the study.

The researchers will:

discuss the study with you
answer your questions
keep confidential any information which could identify you personally
be available during the study to deal with problems and answer questions

If you decide not to take part or to leave the study, your normal treatment will not be affected.

Introduction/Background:
You have a form of arthritis called osteoarthritis. Many people in Newfoundland and other places have arthritis, but we still don't know much about what causes it. We know that what you inherit from your parents plays a part because if you have a close relative with arthritis, you are more likely to get the disease yourself. You inherit genes (made up of DNA) from both your parents. As part of this study we will be examining DNA from osteoarthritis patients and comparing it with DNA from those who do not have arthritis.

2. Purpose of study:
We will recruit about 1000 people who have osteoarthritis and 1000 people who don't have, and be looking to see how the genes you inherit from your parents contribute to the cause of arthritis. We will also look to see whether small changes to these genes within the joints themselves can make a difference.

3. Description of the study procedures and tests:
You are going to have a knee or hip joint replaced because of arthritis in the joint. You will get exactly the same treatment whether you decide to join our research study or not.

If you join the study we will ask you for some extra tubes of blood (up to 4 extra tubes) when you have your routine blood collection. From this blood we will prepare your DNA (the material that makes up your genes) that we will store in a freezer. We will test the DNA from hundreds of arthritis patients to find changes that could be involved in causing arthritis.

We will ask you to complete a questionnaire about your general medical history plus a short form just about your arthritis. We will help to complete the questionnaires, if you would like.

During your joint surgery we will be keeping small samples of the cartilage, subchondral bone, synovial membrane, and of the fluid inside the joint, which are normally thrown away. This material may also be tested for its genetic content.

We also ask your consent to access your medical records, both now and as long as this study continues (at least 5 years). This lets us compare the symptoms you have or may develop, and when you get them, to any genetic changes we may find.

4. Length of time:
Completing the questionnaires should take about 30 minutes.
There is nothing else for you to do.

5. Possible risks and discomforts:
Bruising/discomfort after blood sample.

6. Benefits:
It is unlikely that this study will benefit you personally. We hope that our findings may help in preventing arthritis in the future or in finding better treatments.

7. Liability statement:
Signing this form gives us your consent to be in this study. It tells us that you understand the information about the research study. When you sign this form, you do not give up your legal rights. Researchers or agencies involved in this research study still have their legal and professional responsibilities.

8. What about my privacy and confidentiality?
Protecting your privacy is an important part of this study. Every effort to protect your privacy will be made. However it cannot be guaranteed. For example we may be required by law to allow access to research records.
When you sign this consent form you give us permission to

Collect information from you
Collect information from your health record
Share information with the people conducting the study
Share information with the people responsible for protecting your safety

Access to records
The members of the research team will see health and study records that identify you by name. Other people may need to look at the study records that identify you by name. This might include the research ethics board. You may ask to see the list of these people. They can look at your records only when one of the research team is present.

Use of records
The research team will collect and use only the information they need for this research study.

This information will include your
date of birth
sex
medical conditions
medications
the results of tests and procedures you had before and during the study
information from study interviews and questionnaires

Your name and contact information will be kept secure by the research team in Newfoundland and Labrador. It will not be shared with others without your permission. Your name will never appear in any report or article published as a result of this study.

Information collected for this study will be kept for at least five years after the end of the study.

If you decide to withdraw from the study, the information collected up to that time will continue to be used by the research team. It may not be removed. This information will only be used for the purposes of this study. At any time, you may ask that your DNA and any other specimens be destroyed.

After your part in this study ends, we may continue to review your health records to check that the information we collected is correct.

Information collected and used by the research team will be stored by Dr. Zhai who is the person responsible for keeping it secure.

Your access to records
You may ask the study doctor or researcher to see the information that has been collected about you.

9. Questions:
If you have any questions about taking part in this study, you can meet with the investigator, Dr. Guangju Zhai, 709-864-6683 who is in charge of the study at this institution.

Or you can talk to someone who is not involved with the study, but can advise you on your rights as a participant in a research study. This person can be reached through: Health Research Ethics Authority (HREA) at 709-777-6974 or email**: info**@hrea.ca

Future use of DNA or other samples:

In order to preserve a valuable resource, your DNA and other samples may be stored at the end of this research project. It is possible that these samples may be used in a future research project. **Any future research would first have to be approved by a Research Ethics Board (REB).**

Please tick **one** of the following two options:

	I agree that my samples can be used for an approved research project without contacting me again, but **only if my name* cannot be linked, in any way, to the samples.**
	Under no circumstances may my samples be used for future research. **My samples must be destroyed at the end of this present project**.

*Includes name, MCP number or any other identifying information.

The DNA and other samples from this study will be stored in St. John's, NL for an indefinite period of time.

At any time, you may ask that your DNA and any other specimens be destroyed. To do this please call Dr. Zhai at (709) 864-6683, contact anyone else on the research team, or contact the Office of the HREA at 709-777-6974 or email: info@hrea.ca

To protect your samples, we will ask that you confirm your request in writing before we destroy them.

After signing this consent you will be given a copy.

Appendix E: The self-administered questionnaire that was used in the NFOAS to collect the OA patients' demographic and medical information including age, sex, BMI, and comorbidities.

MEMORIAL UNIVERSITY
Faculty of Medicine

Discipline of Genetics
Health Sciences Centre
St. John's, NL Canada A1B 3V6
Tel: 709 777-6807 Fax:709 777-7497
www.med.mun.ca

Genetic Study of Osteoarthritis in the Newfoundland Population

General Questionnaire

Date form completed: (dd/mm/yyyy) ☐☐ / ☐☐ / ☐☐☐☐

Instruction for completing the questionnaire:

Please answer all questions to the best of your ability (leave blank if unknown).

Please write in block letters using the boxes where provided.

Use a black/blue pen.

Cross out any mistakes & write correct answers just below the relevant boxes.

Indicate your response by filling in the box next to the most appropriate answer or by writing clearly in the boxes or space provided.

Your answers will be completely confidential.

Self administered: ☐

Research assistant administered: ☐

Name and address

Surname

Given name

Title

Maiden Name (if applicable)

Address

Province

Postal code

Date of Birth (dd/mm/yyyy)
/ /

Place of Birth
City/Town

Province/Country

Gender: Male ☐ Female ☐

MCP number:

Section 1: Demographics

1. Ethnic: White ☐ Black ☐ Other ☐, please specify ☐

2. Height: ☐☐☐ cm

3. Weight: ☐☐☐ kg

4a. Smoker: yes (current) ☐ no ☐ ex-smoker ☐

4b. If Yes, how many cigarettes do/did you smoke a day? ☐☐☐

5a. In the past 4 weeks approximately how many units of alcohol did you drink per week? (1 unit = 1 glass of wine/½ pint of beer /1 shot of sprit)? ☐☐☐

5b. Do you think your drinking habits in the last 4 weeks reflect your typical drinking habit? Yes ☐ no, less than usual ☐ no, more than usual ☐

6a. How heavy were you when you were born?

☐☐☐ grams or ☐☐ lbs ☐☐ ozs

6b. If weight unknown, were you Light ☐ Average ☐ Heavy ☐

6c. Were you born prematurely (more than 1 week early) Yes ☐ No ☐

7a. How heavy were you at age 20 yrs? ☐☐☐ kg

7b. How heavy were you at age 50 yrs? ☐☐☐ kg

For women only:

8. At what age did your period start? ☐☐

9. At what age did your period stop? ☐☐

Section 1: Demographics (continued)

10a. Have you had a hysterectomy (removal of the womb)? Yes ☐ No ☐

10b. If Yes, how old were you? ☐☐

10c. Did the hysterectomy include removal of the ovaries?

Yes ☐ No ☐ or Unknown ☐

11. Have you ever taken an oral contraceptive pill? Yes ☐ No ☐

12a. Have you ever taken hormone replacement therapy? Yes ☐ No ☐

12b. If Yes, how long in total did you take it for?

Less than 3 months ☐

3 to 12 months ☐

1 to 5 years ☐

Longer than 5 years ☐

13. How many live births have you had? ☐☐

Section 2 - Occupation

14a. What was your current/last occupation (job title)? ☐

14b. In what industry did you carry out this occupation (eg farming, shipyard, car factory, shoe shop, hospital, insurance office)? ☐

14c. Number of years in job: ☐☐

15a. What was the main occupation that you held for the longest period of time (job title)? ☐

15b. In what industry did you carry out this occupation (eg farming, shipyard, car factory, shoe shop, hospital, insurance office)? ☐

15c. Number of years in job: ☐☐

For your main occupation in an average working day, did you:

16. Sit for more than two hours in total? Yes ☐ No ☐ Don't know ☐

17. Stand or walk for more than two hours in total? Yes ☐ No ☐ Don't know ☐

18. Kneel for more than one hour in total? Yes ☐ No ☐ Don't know ☐

19. Squat for more than one hour in total? Yes ☐ No ☐ Don't know ☐

20. Drive for more than 4 hours in total? Yes ☐ No ☐ Don't know ☐

21. Walk more than 2 miles in total? Yes ☐ No ☐ Don't know ☐

Section 2 – Occupation (continued)

22. In the course of your work how often on average did you lift or carry weights of 10 kg or more?

Never ☐

Less than once per week ☐

1 to 10 times per week ☐

More than 10 times per week ☐

23. In the course of your work how often on average did you lift or carry weights of 25kg or more (Equivalent to half a bag of cement)

Never ☐

Less than once per week ☐

1 to 10 times per week ☐

More than 10 times per week ☐

Section 3 – Medical history (1)

Please list in the box below all medication that the patient is currently taking:

Have you **EVER** been told by a Doctor or other health professional that you have **ANY** of the following conditions (please tick all that apply to you):

Cardiology

24. Congenital Heart Disease ☐
25. Coronary Heart Disease ☐
26. Heart Attack ☐
27. Hypertension (high blood pressure) ☐
28. High Blood Pressure in Pregnancy ☐
29. Angina ☐
30. High Cholesterol ☐
31. Deep Vein Thrombosis ☐
32. Varicose Veins ☐
33. Pulmonary Embolism ☐

Immunology/Chest Medicine

34. Asthma ☐
35. Hayfever ☐
36. Eczema ☐
37. Sinusitis ☐

Gastroenterology/Endocronology

38. Heartburn ☐
39. Irritable Bowel Syndrome ☐
40. Crohn's ☐
41. Diabetes ☐

Neurology/Psychiatry

42. Dyslexia ☐
43. Clinical Depression ☐
44. Anxiety/Stress Disorder ☐
45. Epilepsy ☐
46. Stroke ☐
47. Motion Sickness ☐
48. Migraine ☐

Section 3 – Medical history (2)

Have you **EVER** been told by a Doctor or other health professional that you have **ANY** of the following conditions (please tick all that apply to you):

Oncology/Cancers

49. Breast Cancer ☐

50. Colon Cancer ☐

51a. Skin Cancer ☐

if yes, was it:

51b. Melanoma ☐

51c. Basal Cell Carcinoma ☐

51d. Squamous Cell Carcinoma ☐

Rheumatology

52. Gout ☐

53. Paget's Disease ☐

54. Bunions ☐

55. Frozen Shoulder ☐

56. Osteoporosis ☐

57. Carpal Tunnel ☐

58. Tennis Elbow ☐

59. Golfer's Elbow ☐

Dermatology/Skin

60. Acne (that caused scarring) ☐

61. Viral Warts ☐

62. Cold Sores ☐

Hearing

63. Hearing Loss ☐

64. Tinnitus (ringing in ears) ☐

Opthalmology/Eyes

65. Glaucoma ☐

66. Cataract ☐

67. Myopia (short sightedness) ☐

68. Age-related Macular Degeneration (AMD) ☐

Urology

69. Incontinence (leak urine) ☐

70. Polycystic ovary syndrome ☐

Section 3 – Medical history (3)

Please answer the following questions by ticking the appropriate box:

71a. Have you ever lost the use of an arm, leg, vision, or ability to speak?

Yes ☐ No ☐

71b. If Yes, how long for : less than 24 hours ☐ or more than 24 hours ☐

72a. Do you usually bring up phlegm from your chest in winter? Yes ☐ No ☐

72b. Do you usually bring up phlegm on most days for at least 3 months a year?

Yes ☐ No ☐

73a. Have you had heartburn or acid regurgitation in the last year? Yes ☐ No ☐

73b. If Yes, how many times have you had heartburn/acid regurgitation in the last year?

Less than once a month ☐

About once a month ☐

Once a week or more ☐

74a. Have you been bothered by recurrent headaches? Yes ☐ No ☐

74b. If Yes, do you still have recurrent headaches? Yes ☐ No ☐

74c. If Yes, are your most troubling headaches

One sided ☐

Accompanied by sensitivity to light/noise ☐

4 to 72 hours in duration if untreated ☐

Section 3 – Medical history (4)

Please answer the following questions by ticking the appropriate box:

75. Since turning 16 have you ever fractured or broken a bone? Yes ☐ No ☐

If Yes, please tick which of the following bones you have fractured or broken

Wrist ☐ Arm ☐ Ribs ☐ Hip ☐ Ankle ☐ Vertebra ☐ Other ☐

76. In the past 3 months have you had pain in your back on most days?

Yes ☐ No ☐

If Yes, does this pain typically radiate to either leg? Yes ☐ No ☐

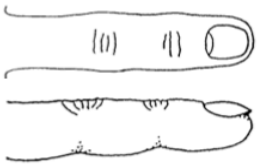
77. In the past 3 months have you had any pain in any part of your body lasting at least 24 hours? Yes ☐ No ☐

Section 4 – Nodal status

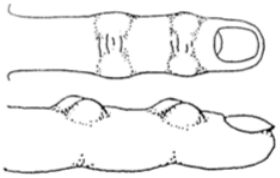
We are interested in knowing whether you have any finger nodes. These sometimes relate to arthritis at the hand and other joints. A finger node is a firm, bobbly swelling on the back of the finger joint.

For example:

A finger **without** nodes:



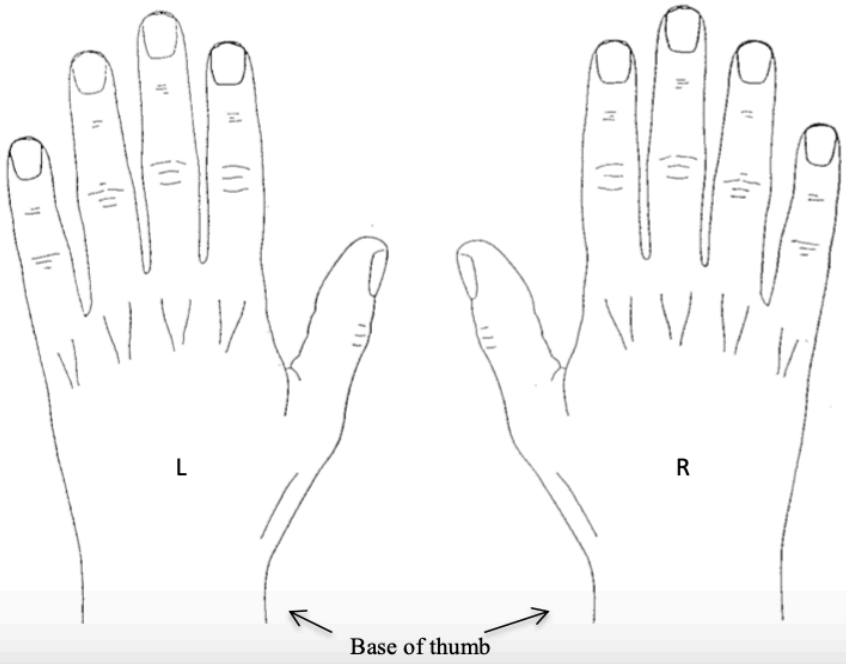
A finger **with** nodes:



When you meet with the research assistant, please look at your hands and then answer the following questions:

78a. Do you think you have any nodes/swellings on your hands? Yes ☐ No ☐

If Yes, for each hand please circle the finger joint(s) where you have these nodes. (You may circle several joints).



Section 4 – Nodal status (continued)

78b. If Yes, at what age did the nodes first develop? ☐☐

78c. How many nodes do you have on the: left hand ☐☐

right hand ☐☐

79. Which hand do you write with? Left ☐

Right ☐

80a. Have you suffered from pain in the fingers for most days for at least one month?

Yes ☐ No ☐

80b. If Yes, at what age did you first develop 'significant' pain in your fingers? ☐☐

81. Do you have pain in the base of your thumb (as arrow on drawing)?

Yes ☐ No ☐

Section 5 – Family History of Osteoarthritis

82. Does/did your mother suffer from osteoarthritis of the knee/hip?

Yes ☐ No ☐ Don't know ☐

If Yes, has/did your mother had/have a total joint replacement of the knee/hip?

Yes ☐ No ☐ Don't know ☐

83. Does/did your father suffer from osteoarthritis of the knee/hip?

Yes ☐ No ☐ Don't know ☐

If Yes, has/did your father had/have a total joint replacement of the knee/hip?

Yes ☐ No ☐ Don't know ☐

84. Does/did your brothers/sisters suffer from osteoarthritis of the knee/hip?

Yes ☐ No ☐ Don't know ☐

If Yes, has/did your brothers/sisters had/have a total joint replacement of the knee/hip? Yes ☐ No ☐ Don't know ☐

Appendix F: A copy of the TASOAC study participants' interview questionnaire.

3039191927

Individual study No.

TASOAC

Interview form

Scales Calibrated ○

Blood Pressure Machine Calibrated ○

Appointment Date / /

Age at Appointment

Name and Address

Surname

Maiden Name (if applicable)

Given Names

Address

Suburb

State

Post Code

Home Phone Number

Business Phone Number

Mobile Phone Number

Gender Male ○1 Female ○2

Date of Birth / /

5466191923

Research Assistant Initials ☐☐☐

Station 1

1.1 Spectrophotometer readings

Spectrophotometer Number ☐

Time Seated ☐☐:☐☐

Time last reading taken ☐☐:☐☐

Time elapsed ☐☐ minutes

Reading **Upper Inner Arm**

1.1.1 Flash Number ☐☐☐ 400nm ☐☐.☐☐ 420nm ☐☐.☐☐

1.1.2 Flash Number ☐☐☐ 400nm ☐☐.☐☐ 420nm ☐☐.☐☐

1.1.3 Flash Number ☐☐☐ 400nm ☐☐.☐☐ 420nm ☐☐.☐☐

1.2 Photo of Hands

Camera Type

Canon Digital G2 ○ Number 1

Other ○ Please specify ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

Date Taken

☐☐ / ☐☐ / ☐☐☐☐

Time Taken ☐☐:☐☐

0246191924

Research Assistant Initials ☐☐☐

1.3 Blood Pressure Reading

Room Temperature ☐☐.☐ °C

Machine Type

Omron M4 ○ Number ☐

Other ○ Please specify ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

Arm Circumference ☐☐.☐ cm

Arm used Right ○ 1
Left ○ 2

Cuff size used Small (17 - 22 cm) ○ 1
Regular (22 - 32 cm) ○ 2
Large (32 - 42 cm) ○ 3

If *Left* arm was used please specify why

Right arm injured ○ 1
Right arm amputated ○ 2
Recent surgery to right arm ○ 3
Other ○ 4

☐

Time first reading taken ☐☐:☐☐

Subject must have been seated for at least 5 minutes prior to taking blood pressure readings

	Systolic	Diastolic
1.3.1 Reading 1	☐☐☐ mmHg	☐☐☐ mmHg
1.3.2 Reading 2	☐☐☐ mmHg	☐☐☐ mmHg
1.3.3 Reading 3	☐☐☐ mmHg	☐☐☐ mmHg

1163191923

1.4 Blood taken Yes ○ 1 No ○ 2 Date Taken ☐☐ / ☐☐ / ☐☐☐☐

Time Taken ☐☐ : ☐☐ Hrs

Fasted Yes ○ No ○

Time Fasted (food & fluids) from ☐☐ : ☐☐ Hrs

Period Fasted (food & fluids) for ☐☐ : ☐☐ Hrs

Buccal Swab taken Yes ○ 1 No ○ 2 Date Taken ☐☐ / ☐☐ / ☐☐☐☐

Time Taken ☐☐ : ☐☐ Hrs

Reason for Incomplete blood Sample

- Refusal ○ 1
- Patient felt faint ○ 2
- Difficulty finding vein ○ 3
- Vein Collapsed ○ 4

Blood Followup

- New Appointment Made ○ 1
- Arranged to have blood taken elsewhere ○ 2

Comment

☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

1.5 Height ☐☐☐ . ☐ cm

Stadiometer Type Leicester ○ 1

Other ○ 2

If other please specify ☐☐☐☐☐☐☐☐☐☐☐☐

Stadiometer Number ☐

7212191924

1.6 Weight ☐☐☐.☐ kg

Scales Type Heine ○ 1

Other ○ 2

If other please specify ☐☐☐☐☐☐☐☐☐☐☐☐

Scales Number ☐

1.7 Girth Measurement

Was the waist measurement taken:

over skin ○ 1

over clothes ○ 2

Was the hip measurement taken:

over skin ○ 1

over clothes ○ 2

	Waist	Hip
Reading 1	☐☐☐.☐ cm	☐☐☐.☐ cm
Reading 2	☐☐☐.☐ cm	☐☐☐.☐ cm

If there is a difference of 2cm or more between readings 1 and 2 then a third reading should be taken.

Reading 3	☐☐☐.☐ cm	☐☐☐.☐ cm

8142191927

Research Assistant initials | | | |

Station 2

2.1 Short Form Physiological

2.1.1 Edge contrast Sensitivity (MET)

Score | | | **(1)**

2.1.2 Reaction Time - Hand

	Trial 1	Trial 2	
Score 1			(360 - 244)
Score 2			(360 - 244)
Score 3			(360 - 244)
Score 4			(360 - 244)
Score 5			(360 - 244)
Score 6			(360 - 244)
Score 7			(360 - 244)
Score 8			(360 - 244)
Score 9			(360 - 244)
Score 10			(360 - 244)

2.1.3 Proprioception

Score 1 ☐☐.☐ (6 - 1.5)

Score 2 ☐☐.☐ (6 - 1.5)

Score 3 ☐☐.☐ (6 - 1.5)

Score 4 ☐☐.☐ (6 - 1.5)

Score 5 ☐☐.☐ (6 - 1.5)

2.1.4 Knee extension (quads)

Dominant Leg ☐☐ (kg) (5 - 20.1)

If dominant leg unsuitable use the **non Dominant Leg** ☐☐ (kg) (5 - 20.1)

2.1.5 Balance

(a) Sway on foam, eyes open.

Anterior/Posterior ☐☐☐ mm (0 -50)

Lateral ☐☐☐ mm (0 - 50)

(b) Sway on foam, eyes closed.

Anterior/Posterior ☐☐☐ mm (0 -50)

Lateral ☐☐☐ mm (0 - 50)

5830191928

2.1.6 Falls risk calculation Z score ☐☐.☐☐

2.2 Hand Grip Strength

Right Hand 2.2.1a ☐☐.☐☐ lbs/sq in 2.2.1b ☐☐.☐☐ lbs/sq in

Left Hand 2.2.2a ☐☐.☐☐ lbs/sq in 2.2.2b ☐☐.☐☐ lbs/sq in

2.3 Leg Strength

2.3.1 Reading ☐☐☐ kg 2.3.2 Reading ☐☐☐ kg

2.4 Baldness Scale (Men only)

Scale Type Norwood ○ 1

Other ○ 2

If other please specify ☐☐☐☐☐☐☐☐☐☐☐☐

Type of Baldness ☐☐☐☐☐

If totally bald at what age did this occur? ☐☐ Years

9723191923

STATION 3

3.1 Xrays taken

3.1.1 ***Hands*** date ☐☐ / ☐☐ / ☐☐☐☐

3.1.1 ***Hips*** date ☐☐ / ☐☐ / ☐☐☐☐

3.1.1 ***Knees*** date ☐☐ / ☐☐ / ☐☐☐☐

If no XRays taken state reason

Refusal ○ 1
Physical restrictions ○ 2
Other ○ 3

☐☐☐☐☐☐☐☐☐☐

3.2 MRI

Yes ○ 1
No ○ 2

Date taken ☐☐ / ☐☐ / ☐☐

If no MRI taken state reason

Refusal ○ 1
Physical restrictions ○ 2
Claustrophobic ○ 3
Does not require an MRI ○ 4
Other ○ 5

☐☐☐☐☐☐☐☐☐☐

8618191925

Research Assistant Initials

STATION 4

4.1 Bone Density Test completed

Date / /

If no bone density test taken please state reason

Refusal ○ 1
Physical restrictions ○ 2
Severe breathing ○ 3
> 130 kg ○ 4
other ○ 5

Results given to participant on day of test Yes ○ 1

Results given to participant Date / /

Results given to participant to pass on to GP Yes ○ 1

Date sent to GP / /

4001191927

Research Assistant Initials

STATION 5

5.1 Dietary Questionnaire completed Yes ○ 1 No ○ 2

5.2 General Questionnaire completed Yes ○ 1 No ○ 2

5.3 Physical Activity Questionnaire completed Yes ○ 1 No ○ 2

5.4 Interview form completed Yes ○ 1 No ○ 2

5.5 Aus AQOL Questionnaire completed Yes ○ 1 No ○ 2

5.6 Urine bottle received to Menzies / / Date

5.7 Pedometer

Research Assistant Initials

Number of Pedometer

Date given / /

Sensitivity number

5.8 Pedometer Calibrated

Yes ○ 1 No ○ 2

If participant unable to use Pedometer state reason why

Refusal ○
Physical restrictions ○
Unable to comprehend instructions ○
other ○

Signature of Participant .....

5.9 Total mini-mental status exam test score (Computed maximum 30 points)

5.10 General Comments

Appendix G: Copyright approval to adapt and use the leg muscle strength test using a dynamometer figure from Scott et al. (2009).

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Mar 14, 2023

This Agreement between Salem Werdyani ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	5507940676323
License date	Mar 14, 2023
Licensed content publisher	Oxford University Press
Licensed content publication	QJM: An International Journal of Medicine
Licensed content title	Statin therapy, muscle function and falls risk in community-dwelling older adults
Licensed content author	Scott, D.; Blizzard, L.
Licensed content date	Jul 24, 2009
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Genomic and metabolomic studies for better understanding osteoarthritis pathogenesis
Publisher of your work	Memorial University of Newfoundland
Expected publication date	Sep 2023
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD

Appendix H: A list of the significantly associated variants with OA in the recent largest GWAS-meta-analysis *(Boer et al. Cell 2021)* that were identified to be associated with OA in all 557 OA patients, knee OA (n=386) patients, and hip OA (n=171) patients from the NFOAS with p≤0.05.

Significantly associated genetic variants with OA from the recent largest GWAS-meta-analysis *(Boer et al. Cell 2021).*								**Associated SNPs (p≤0.05) with OA in the 557 OA patients from NFOAS**			
Osteoarthritis phenotype	**SNP_ID**	**Chr:position**	**EA**	**EAF**	**OR**	**95% CI**	**p**	**EA freq**	**beta**	**SD**	**p**
TJR	rs2038740	6:35114542	T	0.72	0.94	0.93–0.96	$6.20*10^{-10}$	0.70	-0.34	0.17	0.04
								Associated SNPs (p≤0.05) with OA in the 386 knee OA patients from NFOAS			
KneeOA	rs6500609	16:4515334	C	0.11	0.94	0.91–0.96	$5.16*10^{-9}$	0.11	-0.55	0.24	0.02
								Associated SNPs (p≤0.05) with OA in the 171 hip OA patients from NFOAS			
Finger OA, Hand OA	rs9396861	6:18404133	A	0.61	1.13	1.09–1.17	$9.35*10^{-11}$	0.60	0.43	0.18	0.02
Hip OA, TJR, THR	rs1321917	9:119324929	C	0.41	1.1	1.08–1.13	$9.87*10^{-19}$	0.42	0.35	0.17	0.04

EA: effect allele, **EAF:** effect allele frequency, **OR:** Odd Ratio, **95% CI:** 95% confidence interval, **p:** p-value, **EA freq:** effect allele frequency in the NFOAS, **SD:** standard variation.