Genetic Screening Algorithm for Inflammatory Back Pain

By:

©Rebecca Power

Submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Master of Science in Medicine

Discipline of Genetics

Memorial University of Newfoundland

May 2018

St. John's, Newfoundland and Labrador

Abstract

Objective: To develop a single nucleotide polymorphism (SNP) based genetic-based algorithm among patients with low back pain to screen for axial spondyloarthritis (SpA).

Methods: An 18-plex genetic assay was designed using a MassARRAY, consisting of SNPs associated with ankylosing spondylitis (AS), psoriasis, inflammatory bowel disease (IBD) and uveitis. 1172 AS cases and 848 controls have been analyzed over two cohorts. A machine learning algorithm was created using a J48/C4.5 decision tree model; the first decision was *human leukocyte antigen B 27* (*HLA-B*27)* status. The initial algorithm was validated in an independent cohort. The discovery and validation cohorts were then combined and the final genetic-based screening algorithm was weighted.

Results: The SNP based algorithm that included *HLA-B*27* positivity had a precision, specificity and sensitivity of; 0.83, 0.83, and 0.80, respectively which is higher than the current *HLA-B*27* based Assessment of Spondyloarthritis International Society (ASAS) classification criteria. The SNP based algorithm that included *HLA-B*27* negativity had a precision, specificity and sensitivity of, 0.58, 0.32, and 0.69, respectively.

Conclusions: This genetic screening algorithm is inexpensive, out performs the clinical arm of the current ASAS classification criteria and can potentially lead to earlier detection of axial SpA.

Acknowledgements

I would first like to acknowledge Dr. Quan Li. Thank you for all your work creating the machine learning algorithm. Secondly, I would like to thank my supervisors Dr. Proton Rahman and Dr. Darren O'Rielly. Thank you to Nadine Burry, Amanda Dohey and Dianne Codner for your assistance and teaching in the lab. Thank you to my committee member Dr. Dicks and Dr. Gao for helping me with my thesis writing.

Thank you to Dr. Bob Gendron, Dr. Helene Paradis, and Dr. Vina Broderick for convincing me to do my Master's in Science. Thank you to my family, my boyfriend, friends and the entire Genetics Discipline for helping me through my project and showing me that I am no longer a dunder. Special thank you to my Dad, George Power who edited my thesis.

Thank you to the University of Toronto and University of Alberta for assisting with this project. Finally, thank you to the donors of the scholarships I was awarded over the past two years. The Noah Curtis Godwin Llyod Student Award, the Faculty of Medicine Scholarship, Sail Canada's Instructor of the Year and the Newfoundland and Labrador Neurotruama Initiative Award.

Table of Contents

List of Tables

Table of Figures

Abbreviations List

Abbreviations	Name
A	Adenine
AAU	Anterior Acute Uveitis
ADS	Assay Design Suite
ASAS	Assessment of Spondyloarthritis International Society
AVG	Average
Axial SpA	Axial Spondyloarthritis
BF	Bayesian Factor
C	Cytosine
CARD9	Caspase Recruitment Domain-Containing Protein 9
CCP	Cyclic Citrullinated Peptide
CD8	Cluster of Differentiation 8
Chr	Chromosome
CNV	Copy Number Variant
CRP	C-reactive protein
DMARDs	Disease-Modifying Antirheumatic Drugs
DNA	Deoxyribose nucleic acid
dNTP	Deoxynucleotide
ER	Endoplasmic Reticulum
ERAP1	Endoplasmic Reticulum Aminopeptidase 1
ERAP2	Endoplasmic Reticulum Aminopeptidase 2
ERK	Extracellular Signal-Regulated Kinase
EXT	Extend
FBXL19	F-box and Leucine-Rich Repeat Protein 19
G	Guanine
GWA	Genome-wide association
Hetero	Heterozygote
HLA	Human Leukcyte Antigen
HLA-B	Human Leukcyte Antigen B
HLA-B*08	Human Leukcyte Antigen B 08
HLA-B*13	Human Leukcyte Antigen B 13
HLA-B*27	Human Leukcyte Antigen B 27
HLA-B*38	Human Leukcyte Antigen B 38
HLA-B*39	Human Leukcyte Antigen B 39
HLA-B*40	Human Leukcyte Antigen B 40
HLA-B*44	Human Leukcyte Antigen B 44
HLA-B*47	Human Leukcyte Antigen B 47
HLA-B*51	Human Leukcyte Antigen B 51
HLA-C	Human Leukcyte Antigen C
HLA-Cw6	Human Leukcyte Antigen C w 6
HLA-DRB1	Human Leukcyte Antigen class II, DR beta 1
Homo	Homozygote

HPLC	High Performance Liquid Chromatography
IBD	Inflammatory Bowel Disease
IL-12	Interleukin 12
IL-17	Interleukin 17
IL-23	Interleukin 23
IL-23R	Interleukin 23 receptor
IL12B	Interleukin 12B
INF	Interferon
LBP	Lower Back Pain
ITM2A	Integral Membrane Protein 2A
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
MAP	Mitogen-Activated Protein Kinase
MAX	Maximum
MHC	Major Histocompatiability Complex
MIN	Minimum
MRI	Magnetic Resonance Imaging
MRPS11	Mitochondrial Ribosomal Protein S11
NCBI	National Center for Biotechnology Information
ncRNA	Non-Coding RNA
NFkB	Nuclear Factor Kappa-B
nL	Nano Litre
NOD2	Nucleotide Binding Oligomerization Domain Containing 2
NPEPPS	Puromycin-Sensitive Aminopeptidase
nr-axial SpA	Non-Radiographic Axial SpA
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
OR	Odds Ratio
PsA	Psoriatic Arthritis
RNA	Ribose Nucleic Acid
SAP	Shrimp Alkaline Phosphatase
SD	Standard Deviation
sec	Seconds
SEC16A	Sec16 homolog A
SNP	Single Nucleotide Polymorphism
SpA	Spondyloarthritis
T	Thymine
Temp	Tempature
Th-1	T helper cells 1
Th-17	T helper cells 17
TNF	Tumor necrosis factor
TNFα	Tumor necrosis factor alpha
TNFAIP3	TNF-Induced Protein 3
TNFΔARE mice	TNF Overexpression Mouse Model
TRAF3IP2	TNF Receptor-Associated Factor Interacting Protein 2
TRAF6	TNF Receptor-Associated Factor 6

UCSC	University of California Santa Cruz
UGT2B17	UDP-Glucrosnosyltransferase 2 Polypeptide B17
WHO	World Health Organization
WT	Wild-Type
ΔPH	Difference in Peak Height
μL	Micro Litre
μM	Micro Molar

Chapter 1 Introduction

1.1 SpA in brief

Inflammatory back pain is primarily associated with the disease spondyloarthritis (SpA). SpA is a group of inflammatory rheumatic diseases that encompasses axial spondyloarthritis (axial SpA) and peripheral spondyloarthritis [which includes psoriatic arthritis (PsA), reactive arthritis (reiter's syndrome) and arthritis associated with inflammatory bowel disease (IBD)] (Figure 1.1). These disorders are interrelated and share overlapping pathophysiological pathways, clinical features, treatments and genetic variants (1). A characteristic feature of SpA is enthesitis which refers to inflammation on the insertion point of the tendons and ligaments on to the bone. SpA causes pain and stiffness of the axial spine, sacroiliac joints and occasionally axial SpA. The inflammation not only leads to pain and stiffness, but can also result in new bone formation, such as syndesmophytes causing bridging of the vertebral bodies (ankylosis) (2). The ankylosis cannot be reversed; therefore, it is very important to diagnose this disease early in order to reduce inflammation and hopefully prevent disease progression. The broad term for this disorder is axial SpA. In some patients the scaroiliitis will progress to ankylosis and will become present on radiographic imagery such as an X-Ray. When the disease is radiographic it can be referred to as ankylosing spondylitis (AS).

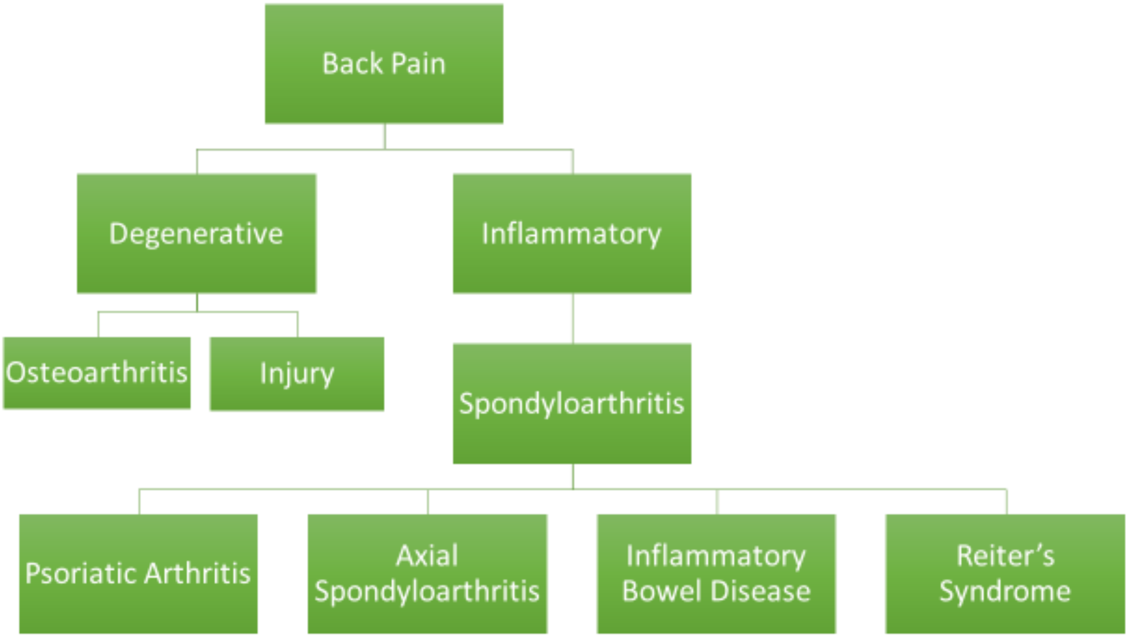


Figure 1.0.1 Classification of back pain.

This figure focuses on spondyloarthritis and its subtypes.

1.2 Pathogenesis of axial SpA

Axial SpA is a seronegative arthritis, meaning it lacks circulating rheumatoid factor and anti- cyclic citrullinated peptide (CCP). SpA has characteristics of both an auto-inflammatory and an auto-immune disease (3). Through many different types of investigations, many immune specific pathways have been attributed to SpA. The following will be discussed in this section; immune related pathways (antigen presentation, innate and adaptive response), microbial and synovial entheseal complex pathogenesis that have been associated to axial SpA.

1.2.1 Immune related pathogenesis

Human leukocyte antigen B 27 (*HLA-B*27)* has one of the strongest genetic associations of any complex disease (4-7). The main function of the *HLA-B*27* molecules is to bind peptides and present them to the surface of cells to be recognized by CD8+ T cells (8).

*HLA-B*27* is a very polymorphic locus and, dysfunction in this loci have been implicated in AS and axial SpA pathogenesis (8). The details of the pathogenesis and genetic variation of *HLA-B*27* is in section 1.3.1 below.

Interaction with *HLA-B*27* has been studied and has shown association to AS pathogenesis. Particularly, three genes from the M1 family of zinc metallopeptidases have been associated with AS: encoding endoplasmic reticulum aminopeptidase (*ERAP1*), encoding endoplasmic reticulum aminopeptidase 2 (*ERAP2*) and encoding puromycin-sensitive aminopeptidase (*NPEPPS*) (9). These aminopeptidases act as molecular scissors and trim peptides presented by *HLA-B*27*, variation at these loci can cause unusual antigen processing which either increase or reduce the availability of these peptides (8). This altered concentration may have variable effects on antigen presentation, through differing *HLA-B*27* heavy chain expression, enzymatic activity of the aminopeptidases and activation of T helper 17 (Th-17) cells (10). Direct *HLA-B*27* and *ERAP1* interaction is still not clear, however a reduction of *ERAP1* activity causes a decreased stability of *HLA-B*27* (11). Two studies have suggested that reduced *ERAP1* leads to increased amount of *HLA-B*27* dimers and longer peptide lengths, whereas, another study attributed reduced ERAP1 activity to a decrease in *HLA-B*27* heavy chain expression (10, 12, 13).

It has been shown that when a person has a variant in *ERAP1* with the combination of a variant in *HLA-B*27,* the risk of developing AS increases significantly (discussion in section 1.3) (9). Investigations have shown that *ERAP1* and *ERAP2* may have functional differences. *ERAP1* and *ERAP2* can form heterodimers which result in increased peptide trimming efficiency (14). However, a variant in *ERAP2* is not sufficient for altered expression of *HLA-B*27.* The study of these modifier proteins and variants has uncovered some evidence that antigen presentation and processing as well as the interaction of proteins with *HLA-B*27* maybe involved in AS and axial SpA pathogenesis.

An axial SpA multigenerational familial exome sequencing study showed that a rare variant in the gene SEC16 homolog A (*SEC16A)* was associated with disease (15). SEC16A is involved in the formation of the coat protein complex II and is required for the transport of proteins from the endoplasmic reticulum (ER) to the Golgi apparatus. In this multiplex family nine members had variants in both *HLA-B*27* and *SEC16A*, seven of the nine had axial SpA (15). These variants were proven to have functional consequences and there is a suggestion of a gene-gene interaction between *HLA-B*27* and *SEC16A*. Studying families with axial SpA could give further insight into the pathogenic pathways into axial SpA as rare variants are exposed to be pathogenic.

An initial hypothesis suggests that the early activation of the innate immune response associated with injury to enthesis may stimulate the development of axial SpA (16, 17). This potential prolonged auto-inflammation can stimulate dendritic cells in the synovial fluid to releases important immune cytokines such as interferon (INF) and tumor necrosis factor (TNF) (16, 18).

INF interacts with immune receptors to stimulate production of pro-inflammatory cytokines such as TNFα and interleukin-1 (IL-1). Disruptions in this pathways, causes the accumulation of pro-inflammatory factors which lead to further inflammation and contribute to SpA pathogenesis. The INF signalling pathway is also required for the activation of the nuclear factor kappa-B (NFkB) signalling pathway, which is another key innate immunity pathway. NFkB forms a protein complex that regulates transcription and cytokine production (TNFα, IL-1, IL-17) which can lead to an inflammatory response. Multiple studies have shown through genetic and functional dysregulation of the NFkB complex has contributed to SpA pathogenesis (2, 9, 19).

The adaptive immune response is also linked to SpA pathogenesis. Traditionally this has been restricted to T helper 1 (Th-1) cells, given that the differentiation of these cells are complete by the upstream immune mediators TNFα, INFγ, INFα and IL-1β (16). When genetic variation disrupts this pathway Th-1 cells display a phenotype that contributes to SpA disease pathogenesis.

In the last decade published work has outlined that Th-17 cells play an important role in pathophysiological role in SpA. Similar to Th-1 cells, TNFα, INFγ, INFα and IL-1β induce release of IL-12/IL-23 by dendritic cells, which in turn causes Th-17 cells to differentiate (20). Mature Th-17 cells produce and release IL-17. IL-17 induces cascades of pro-inflammatory cytokines and angiogenic factors (21), which causes naïve T cell determination to the Th-17 cell lineage – this is a positive feedback loop (20). Genetic variation can cause disruption of this pathway thereby contributing to SpA pathogenesis. From GWA studies we know that multiple variants found to be significantly associated with SpA are involved with this pathway (2, 9, 19).

Both the innate immunity response (through interferon and NFkB signalling) and the adaptive immunity response (through Th-1 and Th-17 signalling pathways) have links to disease pathogenesis. In addition many of the strongest genetic associations to axial SpA such as; *HLA-B*, *HLA-C* and *ERAP1* are involved in immune related pathways (2, 5, 6, 9, 19). These significant genetic associations provide evidence that innate immune pathways play a pivotal role in SpA pathogenesis.

1.2.2 Microbial pathogenesis

Recently the Microbiome has developed a widening appreciation in the study of complex disease. Many studies have shown that the trillions of microbes living within our bodies can interact and effect complex disease. Various protective mechanisms have evolved to prevent these interactions such as physiological barriers (mucosal layers, proteins with antimicrobial properties), tight junctions in epithelial cells and the lamina propria. All of these mechanisms prevent the immune system activation and influence homeostasis. However, the guts interactions have been proven to disrupt homeostasis and initiate an inflammatory response via the adaptive and innate immune systems (22). Specifically, the interactions between the microbiome and the inflammatory response have been implicated in IBD, which as discussed above is a comorbidity of axial SpA. Furthermore it has been shown that AS patients and their first degree relatives have increased gut permeability, which facilitates gut and immune interaction (23). In addition, the bacterium *Dialister* is positively correlated with AS disease activity scores (24). In *HLA-B*27* Transgenic rat models, when the rats were exposed to a germ-free environment disease was prevented; however, when common gut bacterium were transferred to the animals, they developed disease (25, 26). Increasing evidence is showing that the microbiome plays a crucial role in disease pathogenesis, however as these findings are promising, research in this area has only just begun.

1.2.3 Synovial Entheseal Complex

Mechanical stress both internal and external can promote inflammation. Since SpA is attributed to inflammation it is very possible that mechanical stress plays a role in disease pathogenesis. Axial SpA characterises itself from other forms of inflammatory arthritis by its defining symptoms of enthesitis of large weight barring joints and the axial skeleton. One possible reason for this distinct symptom could be from evidence that entheseal resident T cells (CD3+, CD4-, CD8-, ROR- γt+) cells are present at these effected joints (27). Another perspective is that since the entheses are subjected to repetitive mechanical stressing forces, this could lead to further inflammation within the effected joint. Jacques et al., investigated this hypothesis with a TNF overexpression mouse model (TNF$^{\Delta ARE}$ mice). They suspended the hind limbs of mice so that significant stress was relieved from their load bearing joints (28). With the reduced stress there was a decrease in the development of enthesitis and osteoproliferation in the mice (28). The authors suggested that a possible explanation for these results are that mechanoreceptors in weight bearing joints trigger the MAP kinase extracellular signal-regulated kinase (ERK) 1/2 pathway which then stimulates inflammatory pathways such as the TNF pathway (28) causing inflammation. Other studies also suggest a similar pathological mechanism; that mechanical stress stimulates prostaglandin which induces ERK and EP4 receptors to inhibit sclerostin, therefore activating osteoblasts (29, 30). These hypothesises have been supported through GWA studies, as a variant that is involved in this pathway, prostaglandin receptor 4 (*PTGER4*) has shown significance in its association with AS (19). Biomechanical stress seems to play a role in disease pathogenesis, however more work still needs to be completed to understand this concept further.

1.3 Genetics

A significant amount of work has been completed studying the genetics of AS and axial SpA. The precise etiology of AS remains uncertain, however genetic, environmental, mechanical and immunological factors have provided evidence that contribute to disease susceptibility. Research projects studying the strong genetic bias of AS have used heritability, GWA studies, copy-number variation (CNV) and transcriptome analysis to try and provide evidence to the genetic component of AS.

AS and axial SpA have the highest heritability of any immune-mediated complex disease with a calculated heritability of >90% (31, 32). This disease is also highly familial with a sibling recurrence risk ratio of ≥52 (33, 34). There is substantial evidence through twin studies that show that there is a large genetic susceptibility component to AS and axial SpA (31, 32).

1.3.1 HLA B*27

There have been multiple genetic associations to AS, particularly the genetic variant human leukocyte antigen B 27 (*HLA-B*27)* which has one of the strongest genetic associations of any complex disease (4-7), with an odds ratio of 40-90 and a p-value of $<10^{-200}$ (2, 9, 19). *HLA-B*27* has a likelihood ratio of 9.0 (35) and is incorporated in the current axial SpA diagnostic evaluation used globally. However, only 1-5% *HLA-B*27* carriers actually develop axial SpA suggesting that there are other genetic factors influencing this disease (9).

*HLA-B*27* is a member of the major histocompatibility complex (MHC), and is a class I surface antigen encoded by the MHC B locus. The *HLA-B*27* gene is located at the cytogenetic location; 6p21.33 and codes for 105 identified protein subtypes (36). The most common *HLA-B*27* subtypes that are associated with AS are; *HLA-B*27:05*, *HLA-B*27:02*, and *HLA-B*27:04* (36). *HLA-B*27:05* is the most abundant protein variant in the European Caucasian population, this variant consists of approximately 90% of the

*HLA-B*27* variants within this population. *HLA-B*27:02* comprises approximately the remaining 10% of the variation in the Western European Caucasian population. Finally, *HLA-B*27:04* is the primary associated variant subtype for East Asian populations. The remaining identified subtypes are considered rare and are restricted mainly to familial disease inheritance (36).

Currently there are three major hypotheses for why *HLA-B*27* plays such an important role in AS. The first hypothesis suggests that the *HLA-B*27* heavy chain homodimerizes or misfolds, while in the ER (37). The mechanism to this misfolding is thought to be due to disulfide bonds formation because of the presence of Cys67 residue in the α1 domain (38). These disulfide bonds cause the protein to form a heavy chain homodimer, which elicits the ER pro-inflammatory unfolded protein response (39). The second hypothesis is based on evidence that in diseased patients there is the presence of cell-surface expression of homodimers (39, 40). These cell-surface homodimers are thought not to come from misfolding of the protein in the ER but rather arise from endosomal recycling in normally folded *HLA-B*2*7 protein (41). These cell-surface homodimers are recognized by natural killer cells leading to an inflammatory response (41, 42). The arthritogenic peptide theory is the last hypothesis and discusses the ability of *HLA-B*27* to bind to unique peptides or arthritogenic peptides. These peptides are then recognized by $CD8^+$ T cell receptors, thereby initiating a cytotoxic T cell autoimmune response (43).

1.3.2 Additional MHC Genes

The MHC region is one of the densest, and polymorphic regions of the genome, and contains a large amount of genes related to immune function. Most immune-related disorders have genetic associations to this region of the genome. Thus it is not a surprise that many MHC variants have prominent genetic associations to AS and axial SpA. Furthermore, *HLA-B*27* is not the only MHC variant association to AS. These additional variants include *HLA-B*13:02*, *HLA-B*40:01*, *HLA-B*40:02*, *HLA-B*47:01* and *HLA-B*51:01* (44).

Other variants within the MHC locus have gained significance in axial SpA and its related phenotypes. Specifically, to AS and axial SpA there have been studies showing that *HLA-B*40:01* is an important genetic variant to pathogenesis (45). *HLA-B*40* is composed of the HLA-B serotypes B60 and B61. This variant has been linked to patients that are *HLA-B*27* negative.

Other HLA-B genetic variants have been linked to SpA with genome wide significance. Particularly with psoriasis and PsA, *HLA-B*44, HLA-B*39, HLA-B*38* and *HLA-B*08* have reached significance. *HLA-B*39* and *HLA-B*38* have been associated with PsA (46). These variants have been associated with peripheral polyarthritis ion PSA (46) .

The most prominent psoriasis risk allele known is *HLA-Cw6* (47, 48). Similar to HLA-B, HLA-C is an MHC class I antigen. The HLA-C association to psoriasis also includes an early age onset of disease (47, 48). However, with the surge of information provided by genome-wide association (GWA) studies there is suggestion that non-MHC genetic variants also have an important role in disease pathogenesis.

1.3.3 GWA Studies and non-MHC variants

GWA Studies are used to identify what a genetic variants effect size is to a particular disease. These studies are designed in a case/control format and for a particular genetic variant to be deemed associated to the disease of study it must have a p-value of < 10^{-8}.

There have been a total of four GWA studies completed in AS (2, 9, 17, 19). Three of the GWA studies have been completed in the European population (2, 9, 19) and one has been completed in the Han Chinese population (17). These studies have un-covered uncovered many immune specific pathways that are associated with this disease along with genetic variants outside the MHC region (2, 9, 17, 19). The latest total of variants that meet genome-wide association is 48 genetic loci, including many non-MHC variants that reached genome-wide significance (9). These non-MHC significant variants have given insight into the specific pathogenesis of axial SpA, such that the pathogenesis has been attributed to auto-immune pathways (3) specifically, the up-regulation of the inflammatory response (3, 49). In particular genetic variation in the IL-12/23 and IL-17 inflammatory pathway axis that promote the activation of Th-1 cells and Th-17 cells have proven to be strongly associated with axial SpA (49). It is hypothesized that transcript variation and differential expression in this area could also contribute to the susceptibility and heritability of axial SpA.

Promotion and up-regulation of tumour necrosis factor α (TNFα) has been associated with an increased inflammatory response in axial SpA and other inflammatory diseases (33, 49). Similarly, to the IL-12/23 and IL-17 inflammatory pathways, genetic variants in the TNFα pathway have genome-wide significance (2, 9, 19). In addition to these immune specific pathways, modifier proteins and enzymes have been associated with axial SpA pathogenesis. Three aminopeptidases; *ERAP1*, *ERAP2* and *NPEPPS* have genome wide significance in AS (9). ERAP1 and ERAP2 act as molecular peptide trimmers prior to presentation to MHC proteins. Genetic variants in ERAP1 have been associated with *HLA-B*27* and *HLA-B*60/B*40* positive variants (9, 44). This gene-gene interaction has proven to increase the odds ratio (OR) of an individual developing disease from 40 to over 100 (9).

These example provide information that other factors outside the MHC region may contribute to disease pathogenesis and progression. However, even with strong genome-wide significant associations reported via GWAS, these findings have only accounted for 25% of axial SpA heritability (9, 16).

This gap in heritability can be attributed to the limitations of GWA studies as it primarily assesses one type of genetic variant (single nucleotide polymorphisms (SNPs)) and searches only common variants. In order to fully comprehend the pathogenesis of axial SpA alternative investigations using other techniques need to be completed (50).

1.3.4 CNVs

Copy number variants (CNV) are structural variants that are caused by duplications or deletions in a particular gene. CNVs have gained importance within the study of genetics in AS and axial SpA, however there have been few studies completed investigating CNVs (50). The first study to use a genome-wide CNV microarray analysis in the Korean population found 227 CNV regions associated with AS. Of the examined CNVs 9 were independently replicated (51). Of the 9 CNVs, 5 were a deletion-type of CNV and were associated with an increased risk of AS - the remaining 4 CNVs were considered protective. The CNVs found to be significantly associated were physically related to genetic variants already implicated in AS pathogenesis (51). Another study used a genome-wide CNV microarray to examine a multiplex AS family. This study found that the CNV UDP-glucrosnosyltransferase 2 polypeptide B17 (*UGT2B17*) segregated within the affected family members (52). Another study investigating CNVs in the Chinese population found that low copy numbers of the genes *FCGR3A* and *FCGR3B* were significantly associated to AS (53). These studies show that CNVs may contribute to AS pathogenesis.

1.3.5 Gene expression profiling

Gene expression profiling has been investigated through microarray analysis in AS. These studies have shown that cases can be differentiated from controls based on an individual's transcriptome profile (50). Many of the differentially expressed genes are related to immune specific pathways such as the TNFα pathway, NFkB signalling, B-cell receptor signalling, T-cell receptor signalling and IFN genes (54-57)

Although findings from these studies are interesting, these studies have many limitations. The majority of these studies investigate using peripheral blood instead of synovial fluid from affected joints, as well as have small cohort sizes. In addition, there are some inconsistencies between the gene expression profiles reported by these studies. This problem can be partly mended by meta-analysis, and thus far two transcriptome meta-analyses have been completed in AS. One meta-analysis reported that 423 genes were downregulated and 482 genes upregulated; many of these differentially expressed genes were related to the antigen processing and presentation pathway (58). Another meta-analysis identified 65 differentially expressed genes, 23 upregulated and 42 downregulated (59). The upregulated gene with the largest effect size was *integral membrane protein 2A (ITM2A)* which is related to T-cell activation and the downregulated gene was *mitochondrial ribosomal protein S11 (MRPS11)* which is related to mitochondrial translation (59). Transcriptome analysis can explain missing portion of the heritability of AS, however there are still many study limitations in the published literature.

1.3.6 Linkage Analysis

Linkage is the tendency of genetic information to be inherited together based on location of genetic loci. This is studied in multiplex families or sibling pairs as particular loci segregate together. Linkage analysis is a way of capturing the segregation of this genetic information. These studies can be helpful in identifying genetic loci that have large effect sizes.

Thus far there have been three genome-wide linkage studies of importance using the AS population. All of these studies reported a strong linkage to MHC chromosomal regions. One study used an AS population from the United Kingdom; this study found that there was significant linkage in chromosome 16q and inferred linkage at chromosome 2q, 9q,10q, and 19q (60). Another study that used a sibling pair from North America found that there was suggestive linkage at chromosome 6q and 11q (61). The last study was completed using a French AS population, this study found that there was suggestive linkage at the locus 5q, 9q, 13q, and 17q. A meta-analysis of the combined data from these three studies found that there was suggestive linkage at the 10q and 16q locus sites and nominal linkage in chromosomal regions 1q, 3q, 5q, 6q, 17q, and 19q (62). At the time these studies showed great importance and progress in AS genomics. However, with the advent of hundred of thousands of SNP markers arrays, the ability to detect genetic variant is superior in GWA studies rather than linkage analysis.

1.3.7 Selection of Non-MHC Genetic variants for Panel

The type of genetic variants that were selected for this genetic screening algorithm were SNPs. This is a SNP based algorithm. A full literature review was completed analyzing the latest GWAS studies in AS, psoriasis, IBD, and acute anterior uveitis. SNP's were selected based on clinical significance to identify extra-articular features, genome-wide significance and a minor allele frequency of between 5-45%. 20 different SNPs were chosen and the decision was finalized by Dr. Proton Rahman. See Table 2.1 in Methods.

1.3.7a Antigen Presentation

ERAP1

As discussed above *ERAP1* is an aminopeptidase that plays an integral to the MHC Class 1 presentation pathway. *ERAP1* acts on peptides that have been processed by proteasomes and have been transported from the cytoplasm into the ER. *ERAP1* then trims any N-terminally extended peptides to 9 amino-acids in length (63, 64). This is the optimal length for MHC Class 1 loading and presentation. *ERAP1* has been reported to be overexpressed in the dendritic cells of AS patients, resulting in a loss of function protein phenotype (9, 19). Interestingly, *ERAP1* is only associated with AS in patients with a positive *HLA-B*27* variant. As discussed above this gene-gene interaction dramatically increases the OR of a person developing the disease (9, 19). Other gene-gene interactions have been reported with other MHC proteins in other diseases, such as; *ERAP1* with *HLA-B*51* in Bechet`s disease and *HLA-Cw6* in psoriasis. These interactions with HLA genes suggest that *ERAP1* is a very important regulator in complex disease pathogenesis (33).

ERAP2

As discussed above *ERAP2* has a very similar function to *ERAP1*. *ERAP2* shares 49% sequence homology to *ERAP1* and these two proteins have been known to form minor heterodimers with *ERAP1* (65). *ERAP2* is in strong linkage disequilibrium with *ERAP1*, therefore is not possible to determine if *ERAP2* is associated with AS in *HLA-B*27* positive patients (4, 9, 19). However, it has been determined that *ERAP2* is associated with AS in patients negative to *HLA-B*27* (9, 19). *ERAP2* is also associated with other inflammatory diseases such as psoriasis and IBD (66, 67).

1.3.7.b Th-17 Signalling

IL-23R

Interleukin 23 receptor (*IL-23R*) encodes a subunit of the IL-23 receptor. This receptor is crucial for the IL-23/IL-12 pathway axis, as IL-23 signals the transduction of this pathway through the receptor. Once this pathway is initiated, IL-23 stimulates the differentiation of CD4+ T cells into Th-17 cells, which secretes IL-17 a pro-inflammatory cytokine. Dysfunction of IL23R causes an upregulation of this pathway and animal model studies have shown that IL-23 overexpression is sufficient to cause SpA (68). *IL23R* particularly the SNP rs11209026 has been significantly associated with AS, psoriasis, uveitis and IBD (2, 9, 19, 66, 67, 69-72). Thereby showing the importance of the function of the IL-23/IL-12 axis of the pathogenesis of these inflammatory diseases.

IL-12B

Interleukin 12B (*IL-12B*) encodes the subunit of IL-23 and IL-12 cytokines, IL-12 p40 (4). When the IL-12 p40 subunit combines with IL-12 p35 subunit it forms the IL-12 cytokine and when the IL-12 p40 subunit is combined with the IL-23 p19 cytokine it forms IL-23. These protein products bind IL-23R and initiate the IL-12/23 pathway axis as described in the IL-23R section. IL-12B upregulated expression has also been reported in psoriatic skin lesions (73). The *IL-12B* variant is significantly genetically associated with AS, PsA, uveitis and IBD (2, 9, 19, 66, 67, 69-72).

TRAF3IP2

TNF receptor-associated factor interacting protein 2 (TRAF3IP2) is responsible for regulating cytokines related to Th-17 cellular inflammatory response and the NFkB signalling pathway (74). *TRAF3IP2* is mainly associated with PsA (75). This particular variant has shown that it can no longer interact with TRAF6 (TNF receptor-associated factor 6) (75). Showing that *TRAF3IP2* is an important link between the innate and adaptive IL-17 immunity (16).

1.3.7.c NFkB signalling genes

TNFA1P3

TNFAIP3 encodes TNF-induced protein 3 (TNFAIP3), this genes expression is induced by TNF (76). The ubiquitination of this protein occurs when the NFkB complex is activated in order to prevent additional NFkB complex activation (77). The protein is also involved with cytokine immune and inflammatory responses. *TNFAIP3* is significantly genetically associated with psoriasis (66, 69, 78, 79).

FBXL19

F-box and leucine-rich repeat protein 19 (FBXL19) is significantly genetically associated with psoriasis (66, 69, 78, 79). *FBXL19* protein product reversibly inhibits the NFkB signalling. This protein is found to be significantly elevated in psoriatic skin as compared to normal skin (80).

CARD9

CARD9 encodes for caspase recruitment domain-containing protein 9 (hCARD9), and is responsible for stimulating production of TNF, IL-6 and IL-23. hCARD9 uses the NFkB signalling pathway to induce differentiations of Th-17 cells, which secret IL-17 and IL-23 (81). The *CARD9* variant is significantly associated with AS (2, 9, 19) and was also associated with PsA and IBD.

1.3.7.d Autophagy

CARD15/NOD2

Nucleotide binding oligomerization domain containing 2 (NOD2) has many critical functions related to the immune system. *NOD2* is very significantly associated to Crohn's Disease (67, 82) but not AS itself. The NOD2 protein is active in monocytes, macrophages, and dendritic cells (83). It is also active in epithelial cells in the intestine recognizing certain bacteria and stimulating the immune system via the NFkB signalling pathway (83). NOD2 also plays an important role in autophagy and dendritic cells with a NOD2 variants show impaired autophagy (84). It is suggested that the *NOD2* variant cannot recognize bacteria allowing chronic inflammation to accumulate (83).

ATG16L1

Autophagy related 16 like 1 (ATG16L1) is a protein that is required for the autophagy process (85). Autophagy is related to the inflammatory response and assists the immune system destroy harmful microbes (85). *ATG16L1* is one of the most associated genetic variants to Crohn`s disease (67). Dendritic cells extracted from Crohn`s patients with variants in ATG16L1 show deficient autophagy processes, bacteria tracking and antigen presentation, resulting in chronic inflammation (84). This provides evidence that the autophagy process is key to chronic inflammation and Crohn`s disease pathogenesis

1.3.7.e Intergenic 2p15

The intergenic region 2p15 has no translated gene product encoded. RNA sequencing data has revealed that long non-coding RNA (ncRNA) transcripts are present (2). This gives insight that an unknown possible germline regulation mechanism is associated with this variant. Interestingly, this variant is associated with AS and uveitis (2, 9, 19, 72).

1.4 Clinical

SpA is a group of rheumatic diseases that consists of axial SpA, psoriatic arthritis, reactive arthritis, undifferentiated SpA and SpA associated with IBD. There is increasing support that SpA disease group is a heterogenous disease with various phenotypic manifestations (86-88). SpA is considered axial or peripheral based on its dominant clinical feature displayed. Axial SpA involves predominant sacorilitis and spondylitis and is mainly attributed to chronic inflammatory back pain. Inflammatory back pain is diagnosed and defined as an age of onset younger than 40 years, chronic back pain of greater than 3 months, insidious onset, morning stiffness and improvement with exercise (89).

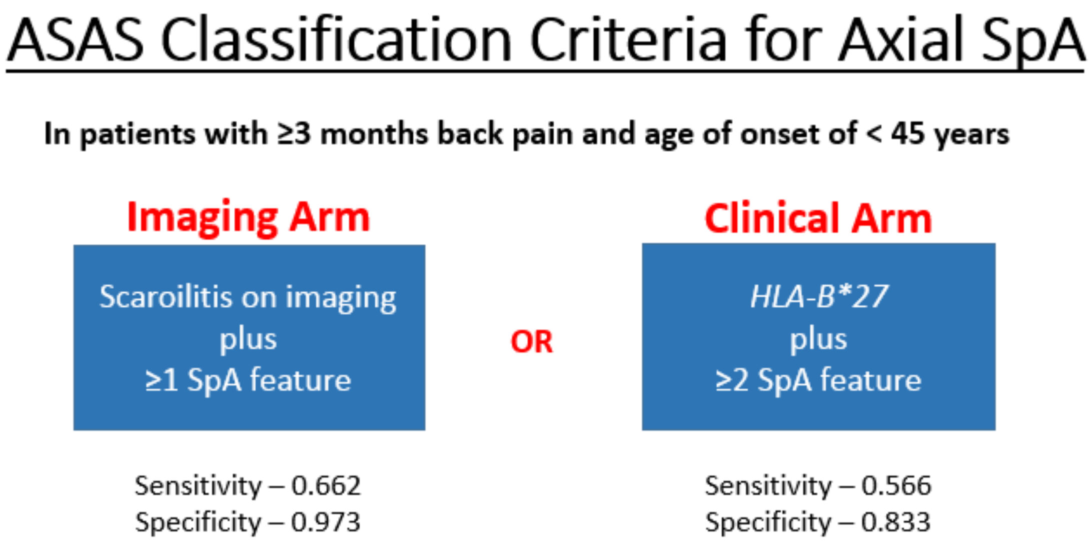
Axial SpA predominantly affects males, commonly manifesting before the age of 40 (3), with a population prevalence of 0.55% in European and 0.23% in Han Chinese (9, 17). Axial SpA is an auto-immune disease; the increase inflammation causes ankylosis which leads to decreased spinal mobility. The most characteristic phenotypic trait is bamboo spine, which occurs when the vertebrae have fused resulting in a concave thoracic spine and thereby leads to the decreased mobility. As axial SpA is a part of the SpA family many individuals suffering from the disease have peripheral musculoskeletal manifestations, such as enthesitis and dactylitis. Enthesitis usually manifests in the Achilles tendon, costochondral and costovertebral joints. Whereas, dactylitis occurs in the fingers and toes due to inflammation of the tendon sheath. In addition, as discussed earlier in this chapter axial SpA has many extra-articular manifestations which includes AAU, psoriasis and IBD. All of these diseases are discussed in detailed and are included in the following section. Currently there is no cure for axial SpA; however, if treated the disease progression may be halted and symptoms can be minimized. This is why it is very important to diagnosis this disease early in order to prevent vertebral fusion and overall disease progression.

1.4.1 Diagnosis

In 2009 the Assessment of Spondyloarthritis International Society (ASAS) changed the diagnostic evaluation and criteria of SpA (90, 91). The concept of axial SpA was introduced, in order to capture earlier and broader spectrum of the disease. This new criteria captures AS which was classified by the modified New York Criteria (92) and non-radiographic axial SpA (nr-axial SpA) (90, 91).

Radiographic scaroiliitis is the hallmark symptom of AS, however it can take up to 7-10 years to occur after the onset of inflammatory back pain. This is one of the main reasons of the diagnosis delay in AS. Therefore, with the ASAS criteria the goal was to improve patient care by encompassing nr-axial SpA with AS in order to speed up diagnosis and may prevent disease progression in patients, thereby introducing the new term axial SpA.

The diagnostic evaluations has an initial criteria of chronic back pain for greater than 3 months and an age of onset of less than 45 years (90) (Figure 1.2). From that initial step the diagnostic evaluation splits into two arms: the imaging arm and the clinical arm (90). The imaging arm criteria is based off the presence of scaroiliitis on imaging. Scaroiliitis on imaging meaning either; "active inflammation on magnetic resonance imaging (MRI) highly suggestive of scaroiliitis associated with SpA and or definitive radiographic scaroiliitis according to the modified New York Criteria" (90). In order to fulfill the imaging arm the patient must also have at least one SpA feature, SpA features are listed in Table 1.1 (90). The clinical arm criteria is based off a positive *HLA-B*27* test and greater than 2 SpA features listed in Table 1.1 (90).



ASAS Classification Criteria for Axial SpA

In patients with ≥3 months back pain and age of onset of < 45 years

Imaging Arm

Scaroilitis on imaging
plus
≥1 SpA feature

Sensitivity – 0.662
Specificity – 0.973

OR

Clinical Arm

*HLA-B*27*
plus
≥2 SpA feature

Sensitivity – 0.566
Specificity – 0.833

Figure 1.0.2: ASAS Classification Criteria for Axial SpA.
Illustration of criteria's imaging and clinical arms. Please refer to Box 1, for list of SpA features. Scaroiliitis on imaging refers to active inflammation on MRI suggestive of SpA characteristic scaroiliitis. As well as, radiographic scaroiliitis as defined by the New York Criteria. This figure was adaptive from the ASAS study by Rudwaleit et al., (90).

Table 1.1: SpA Features that are incorporated in the ASAS Diagnostic evaluation

SpA Features
• **Inflammatory Back Pain**
• **Arthritis**
• **Enthesitis (heel)**
• **Uveitis**
• **Dactylitis**
• **Psoriasis**
• **Crohn’s Disease/Ulcerative colitis**
• **Family History of SpA**
• **HLA-B27**
• **Elevated CRP**

The overall sensitivity and specificity of the ASAS criteria was reported at 82.9 and 84.4%, for the imaging arm the sensitivity and specificity was reported at 66.2 and 97.3% (90) and for the clinical arm it was reported at 56.6 and 83.3% (93). These sensitivities and specificities are low, and show the need for improvement, especially in the clinical arm. The ASAS criteria mainly relies on the imaging arm of the criteria, which is expensive on the health care system and time-consuming.

Magnetic resonance imaging (MRI) is the gold standard for the imaging arm, it is able to detect active inflammation as well as structural abnormalities (1, 94, 95). It is of high diagnostic value, however, inflammation can still be picked up in healthy individuals so it is still important to consider the additional clinical features of the disease (94, 95). Therefore, it is important that an MRI is not used for the diagnosis unless the patient is highly suspected to have axial SpA. In addition, MRIs are very costly, and currently in Canada wait times for MRIs are quite long. Therefore, an improved clinical arm criteria would potentially alleviate the need for unnecessary MRIs.

1.4.2 Management

There is currently no cure for axial SpA, however if the disease is treated early it may prevent disease progression. There are varied forms of management for axial SpA, these include both pharmacological treatments and non-pharmacological treatments such as physiotherapy.

All patients diagnosed with axial SpA are recommended to visit a physiotherapist and participate in joint-directed therapeutic exercises (96). The effects of physiotherapy on management has been reviewed by a systemic literature review and all reviewed studies suggested that physiotherapy relieved symptoms such as pain, physical function, spinal mobility and patients over well-being as compared to no intervention and home-based exercise programs (97). Other forms of successful non-pharmacological are spa therapy and balneotherapy (97). One study suggested that balneotherapy had the same effect on pain as non-steroidal anti-inflammatory drugs (NSAIDs), however, this study was limited by its participant size (98).

Initially axial SpA is treated with anti-inflammatory medications, specifically (NSAIDs). NSAIDs are very important to the treatment of axial SpA, as 70 to 80% of patients with axial SpA report a significant benefit and a good NSAIDs response is a part of the current diagnostic evaluation (96). The NSAIDs commonly used are naproxen, ibuprofen, meloxicam and or indomethacin. No one NSAID is superior over another (96). For patients with localized joint swelling corticosteroids may be injected into joints or tendon sheaths (99). These medications allow for effective pain relief in a localized area.

Some patients do not respond to the above forms of treatment. These patients are commonly prescribed disease-modifying antirheumatic drugs (DMARDs). Generally

DMARDs are not prescribed to treat axial inflammation, however these medications can be quite useful for patients with peripheral symptoms (96, 99). The most common DMARDs prescribed are sulfasalazine and methotrexate (99).

Patients with more severe forms of the disease and have failed all of the above lines of treatment are treated with biologic agents. Biologic agents target patient's immune system to supress the inflammatory response in axial SpA patients. Anti-TNF agents have been available for patients since 2003, these agents work by targeting and inhibiting the TNFα pathway. Anti-TNF agents significantly reduce axial inflammation, enhanced mobility and improved quality of life (96, 97). These agents may halt disease progression in some patients. Currently there are five types of anti-TNF agents available for patients; etanercept, adalimumab, infliximab, golimumab, and certolizumab (99).

Recently another biologic agent, secukinumab, has been approved that inhibits IL-17 activity (99). Secukinumab is an anti-IL-17A monoclonal antibody that inhibits the effector function of IL-17A (100). Anti-IL17 agents have shown promising success, especially for patients that have not responded to anti-TNF therapies. Biologic agents are expensive and sometimes result in adverse reactions in select patients, as well as can have serious side effects such as serious infections and malignancy. Therefore, it is important that they are managed properly and only used when necessary.

In some cases surgical intervention is helpful, these include total joint replacements – the most common being a total hip replacement (97). Spinal surgery is rarely used, except for in extreme cases when there are traumatic factures (96).

There are many different forms of axial SpA management. Some therapies have risk but it is important that proper treatments are put in place for patients with axial SpA in order to potentially prevent disease progression.

1.4.2 Comorbid diseases of axial SpA

When focusing on inflammatory back pain, the subset of SpA that is mainly responsible for this symptom is axial SpA. Axial SpA has extra-axial manifestations and is associated with other inflammatory diseases such as uveitis, psoriasis, and inflammatory bowel disease (IBD) (33, 87, 88). Epidemiological studies have reported the co-existence of these features with axial SpA. (33, 87, 88). Vander Cruyssen et al., (2007) published a study with a cohort of 847 patients stating that 42% of patients had one of these additional inflammatory diseases (101). Of this 42% of patients with an extra-axial manifestation, 50% had uveitis, 20% had psoriasis, 19% had IBD and 10% had a combination of the extra-axial manifestations (See Figure 1.2) (101). In addition these inflammatory diseases share common genetic loci, responses to treatment and etiological pathways (33, 87, 88).

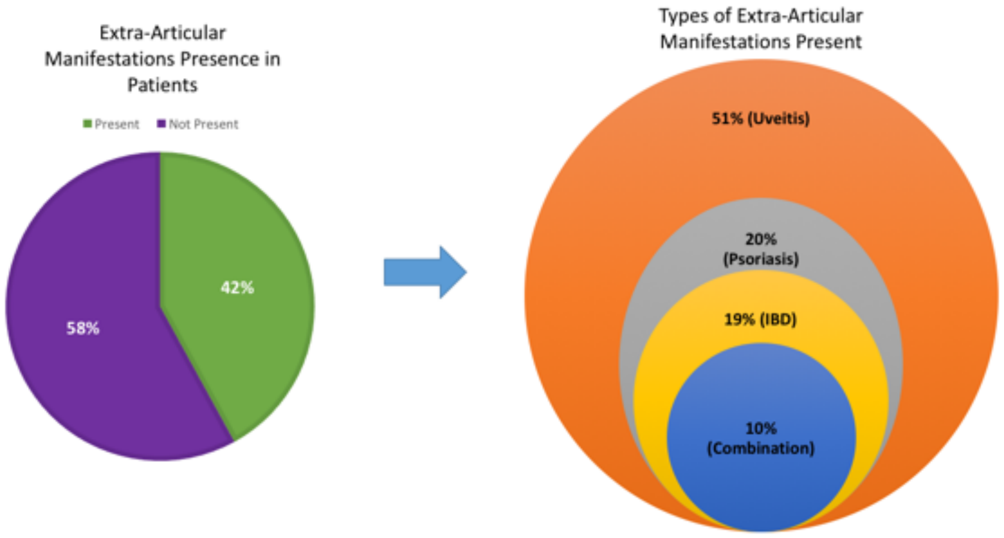


Figure 1.0.3: Extra-Axial Manifestations in AS Patients.

This has been supported with results from GWA studies that have showed notable genetic overlap in immune-mediated diseases (87, 102). Specifically to SpA, previous GWA studies have identified genetic associations to spondyloarthrophies that are shared amongst this disease group (33). In addition robust cohorts have demonstrated that these shared genetic loci are interrelated on wide-spread network analyses (87, 88).

Evidence of the shared inflammatory response prove that there is a common pathophysiology pathway and follow the same pathophysiological axis's (33, 87, 88). Particular pathways related to the IL-17 pathway, IL-23/IL-12 pathway, activation of nuclear factor kappa B (NFkB), amino acid trimming for MHC antigen presentation and the TNFα pathways (87, 88). Since these diseases share the same pathophysiological response, the same treatments are used to treat these diseases. These drugs specifically target these common inflammatory pathophysiological pathways and include non-steroidal anti-inflammatory drugs (NSAIDs) and biologic therapies that repress the TNFα, IL-12/IL-23 and IL-17 pathways. To conclude, there is increasing evidence that these inflammatory diseases are the same diseases just with different phenotypic displays.

1.4.2.a Psoriasis

Psoriasis is hyperproliferative auto-immune skin disorder that affects up to 2% of the North American population. This disease is characterized by demarcated, papular, scaly erythematous skin lesions which can vary in size and severity. These lesions usually occur on extensor surfaces on the elbows and knees.

Approximately 30% of psoriasis patients develop psoriatic arthritis (PsA) which is in an inflammatory musculoskeletal disease. PsA affects both men and women at equal portions and onsets around middle-aged. As PsA is a form of SpA it is caused by enthesitis and can affects the spine (spondylitis) and peripheral joints. PsA also causes inflammatory of the digits (dactylitis) and nails.

PsA has a strong genetic basis, its recurrence risk ratio among first-degree relatives is second only to AS with ranges from 30 to 55. This is substantially higher than the recurrence risk that is seen in psoriasis patients.

Similar to axial SpA, there are strong genetic associations to MHC class 1 proteins (103). The major effect area of the MHC region is *HLA-CW*0602* (48). Psoriasis patients with *HLA-CW*0602* variant have earlier age of onset and higher likelihood of familial psoriasis (47). *HLA-CW*0602* is associated with PsA, however the association is not as strong and PsA patients positive for the HLA-CW*0602 variant have a delayed onset of symptoms (62). Other HLA variants have been associated with PsA; these include *HLA-B13, HLA-B27, HLA-B38/B39, HLA-B57* and *HLA-DRB1*04* (46, 62).

There have multiple large scale GWA studies that have been completed in psoriasis in the European Population. These studies have identified 36 genetic variants that have reached genome-wide significance. These variants account for 22% of heritability associated with psoriasis and are also comprised of non-MHC variants. Four GWA studies have been published in PsA, three of which were relatively small (70, 75, 78, 104). These studies have collectively identified 13 genetic variants associated with PsA. The largest and latest study identified 5 genetic variants at genome-wide significance, these were all known risk variants, *HLA-B, IL12B, TRAF3IP2, TNIP1*, and *TYK2* (78).

1.4.2.b Inflammatory Bowel Disease

IBD is a group of conditions that are caused by chronic inflammation of the colon and small intestine. IBD encompasses both Crohn's Disease and Ulcerative colitis. Crohn's disease can affect the entire gastrointestinal tract, whereas ulcerative colitis only affects the large intestine and rectum. Symptoms of IBD include intense abdominal pain, vomiting, diarrhea, rectal bleeding, weight loss and inflammatory back pain. IBD has a disease onset of 20 to 30 years (105), and incidence rate is 1 per 1000 individuals (106, 107).

Key pathological features of Crohn's disease includes the aggregation of macrophages that form non-caseting granulomas, these aggregations may be segmental and inflammation can be transmural (105). Ulcerative colitis commonly has diffuse mucosal inflammation that can produce a complex mixture of immune mediators such as significant numbers of neutrophils that aggregate in the lamina propria and crypts (105).

Disease heritability has long been recognized in IBD by the aggregation of this disease within families. First degree relative risk ratio of fivefold or greater (105), inherited component is stronger in Crohn's disease as compared to UC (105). Multiple large scale GWA studies have been completed in IBD, mostly within the European population (67, 106). The latest GWA studies included approximately 100,000 IBD patients and had approximately 10,000 non-European ancestry patients within the case cohort (67). This studied showed that although most risk variants clustered within specific populations that some of these variants transcend all populations (67). To date these GWA studies have identified 200 genetic loci that are significantly associated with IBD (67, 106). Of the 200 risk loci 138 showed overlap with other complex diseases (87, 88, 106).

Certain variants have been pointed out as of importance with very strong p-values and have given insight into the pathogenesis of IBD. NOD2 was first found through hypothesis driven studies. *NOD2* is primarily associated with increased risk of Crohn's disease and resequencing and fine-mapping studies have confirmed this variant to be causal (108). NOD2 plays an important role in innate immunity system through the NFkB pathway by the risk variant failing to activate the autophagy pathway (105). Hypothesis suggest the *NOD2* variants relation to IBD pathogenesis is related to epithelial cell lesions and downregulation of toll like receptors (106).

Resequencing and fine mapping studies have also stated that additional variants are causal for IBD (108, 109). *ATG16L1*, is strongly associated with risk of Crohn's disease. It encodes essential proteins for autophagy and several studies have identified this protein in inflammatory signalling. Interleukin 23 receptor (IL23R) has also been identified as causal variant for IBD. IL23R is also significantly associated with axial SpA, as well as psoriasis and uveitis. IL23R is a key molecule in the IL-23/12 and IL-17 pathways and is curial for T-helper 17 (Th17) signalling (106). Another important variant of interest is *CARD9* which is significantly associated with IBD and axial SpA. CARD9 encodes an important protein for innate immunity, which provides protection against infections and also assists in the activation of NFkB signalling (106).

IBD is a chronic inflammation and recent investigations into the genomic architecture of this disease has proven that has shared pathogenesis and genetic loci with other forms of SpA.

1.4.2.c Anterior Acute Uveitis

Uveitis is a disease characterized by the inflammation of the uvea, which is the middle layer of the eye and contains most of the eye's blood vessels. Uveitis can slightly decrease vision or cause severe vision impairment, it can also lead to other visual impairments such as glaucoma, the development of cataracts and complete vision loss (72). This disease primarily affects people between the ages of 20 and 60 years old and has an incidence rate of 0.2% in the European population (110), which accounts for 10% of blindness in Americans. Uveitis can be acute or chronic, specifically anterior acute uveitis (AAU) is an extra-articular manifestation of SpA, it occurs in 30-40% of patients with AS (4). Acute uveitis is sudden in onset and these episodes last for an average of 6-8 weeks (111). It also increases in frequency with AS disease progression, with 60 % of > 50 years AS patients developing AAU (112).

Previous studies have reported that AAU has a large genetic component and similarly to axial SpA, AAU is strongly associated with *HLA-B*27* (113). Approximately half of all patients with AAU are positive for *HLA-B*27* (72). This corresponds with the first degree relative recurrence risk of AAU, which is much higher when the patient is *HLA-B*27* positive (111).

Recent genetic studies have made genome-wide significant findings for AAU outside the MHC region. These variants include; *ERAP1, IL23R, GPR25-KIF21B*, and the intergenic regions 2p15 and 21q22 (72, 114). All of these variants are associated with AS (72, 114), suggesting a common etiology. Specifically, the strong association with *ERAP1*, shows that aminopeptidases and MHC antigen presentation is critical portion of AAU disease pathogenesis. Similarly, IL23R suggest that the IL-12/23 and IL-17 pathways are also important factors for the pathogenesis of this disease.

1.5 Rationale for the Study

1.5.1 Back Pain – A world health problem

Back pain is a very common health problem that most people will have to deal with at some point in their life. It is the largest cause of inactivity globally and causes a large financial burden on patients, families, communities and governments (115-119). Back pain has a high worldwide disease burden, once thought of as primarily a Western nation health problem, back pain is also a major problem in the developing world (115, 120).

Low back pain is the most common and dominant kind of back pain. The Global Burden of Disease Study by the World Health Organization (WHO) reported that low back pain is the leading cause of years lived with a disability (120) and the lifetime prevalence of low back pain worldwide is estimated at 60-70% (121). In Canada, the reported back pain point of prevalence is 15%-25%, and a lifetime prevalence of 60%-80% (122-124). Multiple large-scale research studies have demonstrated that low back pain is a major health problem worldwide.

In Canada the overall annual socioeconomic burden of musculoskeletal disease was estimated to be $22.3 billion, in which low back pain was considered to be the dominant factor and contributor to this estimate (125). According to an estimate in the United States, low back pain cause an approximate yearly direct and indirect health care expenditure of $100-200 billion dollars (121, 126, 127).

Low back pain can be acute, sub-acute and or chronic (121). Most define chronic back pain as greater than 3 months of consistent pain (90, 91). Chronic low back pain has many risk factors such as: obesity, posture and age; however, the cause of the onset remains mainly unclear and makes diagnosing back pain very difficult for physicians (121). To add to these difficulties, back pain can be primarily categorized into two categories: degenerative and inflammatory (See Figure 1.1). Both types of back pain present with very similar symptoms; however, the causes of the two classes are very different. Degenerative back pain can be attributed mainly to degeneration or physical injury to the musculoskeletal system or joints. Whereas, inflammatory back pain can be mainly attributed to chronic inflammation of the joints, which is primarily due to auto-immune responses. Inflammatory back pain is primarily associated with the disease, spondyloarthritis (SpA).

1.5.2 The challenge of diagnosing back pain and axial SpA

The current primary care physician evaluation of patients with musculoskeletal pain, particularly inflammatory low back pain is very unreliable. When most general practitioners are confronted with axial SpA, symptoms overlap with mechanical low back pain – making it very difficult to distinguish. These difficulties can be also be attributed to the volume of inflammatory back pain that general practitioners see in their clinics. A combination of these problems is reflected by a delay of 7 to 10 years from the onset of symptoms to diagnosis (128). Additionally, the costs associated with inappropriate diagnostic evaluations such as x-rays, bone scans, CT scans, and MRI are significant (16).

As discussed earlier *HLA-B*27* is one the strongest genetic associations to any complex disease. But, *HLA-B*27* has low positive predictive value when used to diagnosis axial SpA as only approximately 1-5% of carriers actually develop disease (9).

However, with the surge of GWA studies in many of these inflammatory diseases (discussed above) have yielded large numbers of significant genetic associations. These individual variants are not very discriminative at predicting disease outcome individually. However, geneticists have shown that combining multiple significant loci into a global genetic risk model, can increase prediction accuracy for some complex diseases (129-134). Therefore, given that the current evaluation is both expensive and time consuming - the development of a genetically enhanced screening algorithm will represent a major advance in the early detection of axial SpA.

1.5.3 Genetic Risk Models

A genetic risk model was published by a group studying psoriasis. These researchers used 10 SNP variants to create two types of genetic risk models (129). They created an additive genetic risk model, where the risk variants were added together in order to equate combined risk. They also created a weighted genetic risk model where each SNP was weighted via each individual SNPs odds ratio. The conclusion of this study was that both genetic risk models were significantly better at predicting disease risk as compared to any individual SNP. When both genetic risk models were compared the weighted genetic risk model significantly out preformed the additive genetic risk model at predicting risk of psoriasis (129). In addition, utilized MassARRAY technology, which can effectively and efficiently multi-plex multiple genetic variants into the same assay. This technology makes SNP-based testing easy and cost effective.

These models can aid in diagnosis; however, there is still a need for improvement. Two research groups studying heart disease created very similar additive genetic risk models to the psoriasis research group`s. They determined that although the creation of the genetic risk models were beneficial for risk prediction in two separate cohorts, it did not significantly improve the current risk prediction methods (133, 134). Genetic risk models can increase the prediction accuracy of complex disease diagnostic evaluations; however, the use of more sophisticated statistical programming techniques such as machine learning may improve risk prediction in complex diseases (135).

1.5.4 Machine Learning in health care

Machine learning is a type of artificial intelligence programming that provides the program the ability to learn without being explicitly programmed to do so (136). Machine learning algorithms can improve with experience, meaning the more data the program acquires the more sophisticated decisions the program can make. These programs can be as simple as a data sorter to as complex as making financial decisions on the New York Stock Exchange.

When visualising how a machine learning algorithm works it is best to use an everyday example of complex decision making. For example, if you went shopping for oranges, your friend told you that the oranges with the brightest colour are the tastiest oranges. Therefore, you go to the shop and buy the brightest oranges. However, you learn when you try the oranges that this is not always the case and you notice that the oranges that are also squishier are tastier. Then the next time you go to the shop you buy the squishy and bright oranges. Every time you go to the shop you learn a different way to identify which oranges are the tastiest – this is how machine learning works. The more data or experience the program has the better the program is at say buying delicious oranges or making a complex prediction.

A machine learning algorithm works by first using a portion of a large data set as a training set. This training set gives the algorithm the ability to "train" and to understand data that it has been given (136). There are two main types of machine learning training methods: supervised and unsupervised learning (136). Supervised learning methods train on a dataset that have labels and make predictions on unlabelled examples, whereas unsupervised learning methods look for structures in data sets without using labels (136). For the purpose of this study I will describe the supervised learning method. This learning method uses training data set will have binary labels such as case and control. The algorithm then can produce a model from the training data set that can be used with a prediction algorithm to assign predicted labels (such as case or control) to an unlabelled testing data set. Every time the dataset grows the algorithm has the ability to "train" in order to increase the accuracy of the prediction algorithm (136).

This type of programming can and will be very helpful in medicine. With the advent of electronic health records, large data sets can be more easily acquired (135). This gives the potential of making interesting predictions from machine learning algorithm and complex data structure. Machine learning could easily change the way diseases are diagnosis and pick up on risk factors that have not been researched in the medical field (135). The potential of this field within medicine is huge, that is why it is important to start designing studies using machine learning in genomics.

1.6 Research Objectives

To develop a genetic screening algorithm using MassARRAY technology and a genetic risk model via machine learning to aid in the diagnosis of inflammatory back pain.

Chapter 2 Methods

This study was approved by Memorial Universities ethics board. The study number is 1999.172. Ethics board in University of Alberta and Toronto had approved DNA collection for genetics studies led by Drs. Walter Makysmowych and Robert Inman. An overview of this section is provided in the Figure 2.1.

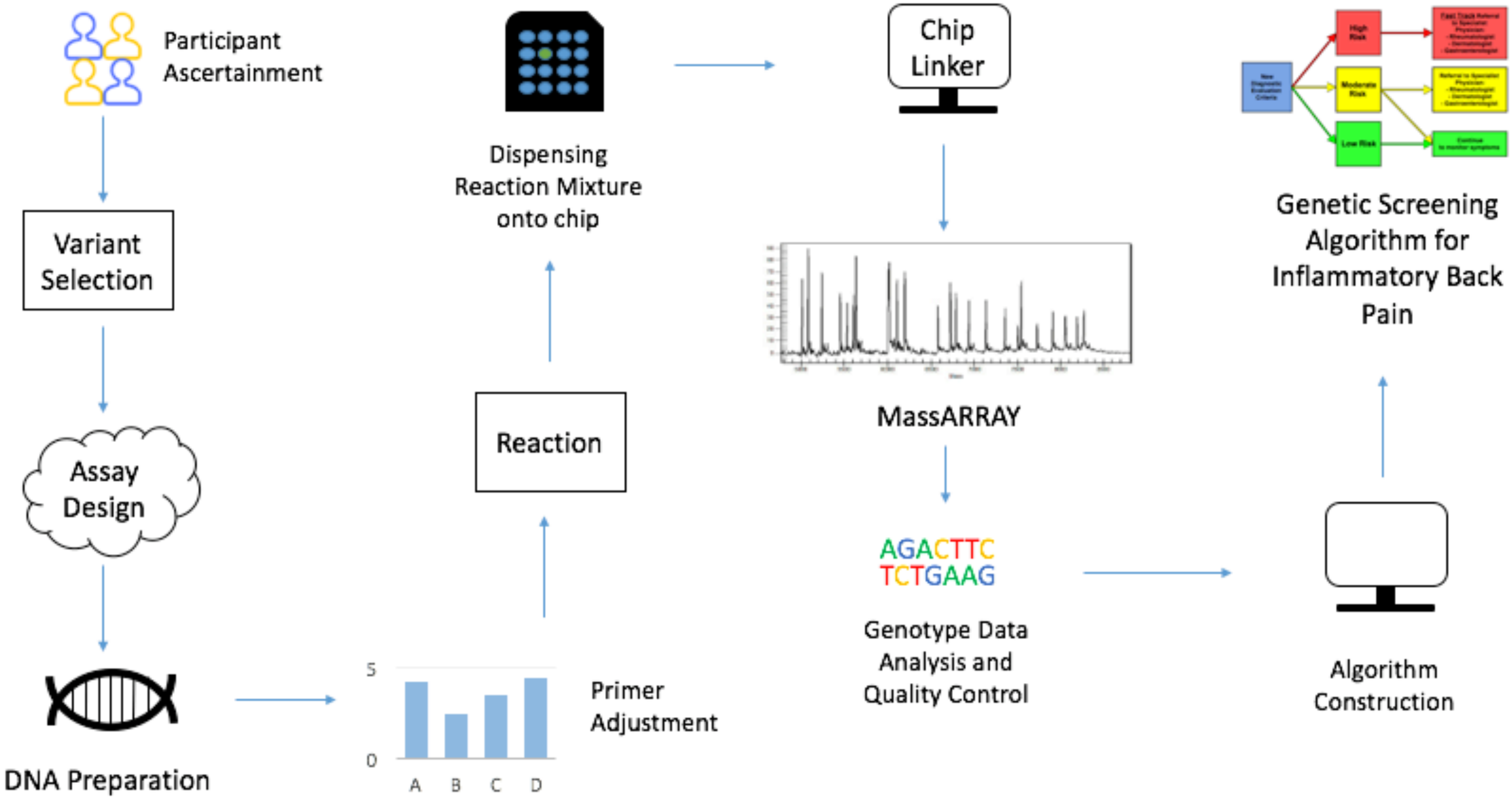

Participant Ascertainment
Variant Selection
Assay Design
DNA Preparation
5
0
A
B
C
D
Primer Adjustment
Reaction
Dispensing Reaction Mixture onto chip
Chip Linker
MassARRAY
AGACTTC
TCTGAAG
Genotype Data Analysis and Quality Control
Algorithm Construction
New Diagnostic Evaluation Criteria
High Risk
Moderate Risk
Low Risk
Fast Track Referral to Specialist Physician
- Rheumatologist
- Dermatologist
- Gastroenterologist
Referral to Specialist Physician
- Rheumatologist
- Dermatologist
- Gastroenterologist
Continue to monitor symptoms
Genetic Screening Algorithm for Inflammatory Back Pain

Figure 2.1: Overview of the Methods for Project.

2.1 Participants

Study case participants were recruited and ascertained by rheumatologist physicians at three sites; Memorial University, St. John's, NL, University of Toronto, ON, and University of Alberta, Edmonton, AB. All participants were of European decent. All case participants were diagnosed by a rheumatologist with AS or non radiographic axial SpA at the time of collection. Patients with non-radiographic axial SpA had MRI changes compatible with scaroiliitis. This is because DNA collection started prior to the ASAS classification. Now that we have a new classification criteria, patients with AS (defined by New York criteria) and non-radiographic axial SpA were now classified as imaging arm of axial SpA. Axial SpA is a broader clinical classification and therefore contains patients with both radiographic and nr-axial SpA. Axial SpA is diagnosed by the ASAS diagnostic evaluation.

In this study all case participants met the current ASAS diagnostic evaluation. Clinical information for all participants was collected and blood was collected for DNA extraction. Controls participants were accessed from the existing control database from Memorial University, University of Toronto and University of Alberta. Control participants came from previous case control studies, where they acted as controls, in osteoarthritis and type 2 diabetes. The majority of the controls were assessed clinically by an internist and did not have an obvious autoimmune disease. Participant's numbers that were genotyped for each set are listed in the results section (Table 3.1).

2.2 Genetic Variant Selection

The type of genetic variants that were selected for this genetic screening algorithm were SNPs. This is a SNP based algorithm. A full literature review was completed analyzing the latest GWAS studies in ankylosing spondylitis, psoriasis, IBD, and acute anterior uveitis. SNP's were selected based on genome-wide significance, a minor allele frequency of between 5-45%, gene-gene interaction and clinical significance. Eighteen different SNPs were chosen and the decision was finalized by Dr. Proton Rahman. See Table 2.1 below of SNPs (gene, MAF, rs number, clinical information).

Table 2.1: List of Genetic Variants included in Discovery Cohort**.**

MAF is taken from the 1000 Genomes Project. Information that is missing from p-value, quoted paper did not supply value.

SNP ID	Gene	p-value	Odds ratio	Minor Allele Frequency (MAF)	Associations with Uveitis, IBD, Crohn's, AS	Source
rs116488202	***HLA-B*2705***	< 1E-200	40.8	T=0.0136/68	AS/ Uveitis / IBD	(9)
rs1265163	***LD with HLA-B60***		1.8	G=0.1825/914	AS	(137)
rs10456057	***LD with HLA-CW6***	4.06E-214	4.66	G=0.1116/559	PS	(69)
rs3132528	***LD with HLA-B44***			C=0.2308/1156	PsA/ PS	(138)
rs10781500	***CARD9***	1.10E-06	1.1	T=0.3670/1838	AS/IBD	(19)
rs11209026	***IL23R***	8.12E-161	2.013	A=0.0228/114	AS/ Uveitis / IBD	(102)
rs2032890	***ERAP1***	2.11E-16	1.51	C=0.1619/811	AS/ Uveitis / IBD	(139)
rs2066844	***CARD15/NOD2***	9.19E-214	2.13	T=0.0144/72	IBD	(67)
rs582757	***TNFAIP3***	2.65E−16	1.6	C=0.2584/1294	PsA	(75)
rs6738490	***ATG16L1***	4.26E-78		C=0.3952/1979	IBD	(67)
rs33980500	***TRAF3IP2***	2.65E−16	1.6	T=0.0837/419	PsA	(75)
rs6759298	***intergenic 2p15***	4.90E-47	1.29	G=0.4097/2052	AS	(9)
rs6871626	***IL12B***	3.10E-08	1.1	A (HapMap)	AS/IBD	(9)
rs10782001	***FBXL19***	4.00E-08	1.26	A=0.4932/2470	PsA	(69)
rs2910686	***ERAP2***	4.50E-17	1.17	C=0.4177/2092	AS	(9)
rs2853931	***LD with HLA-B*3906***	<0.0001	3.74	T=0.2861/1433	PsA	(138)
rs3129944	***HLA-B*3801***	<0.0001	9.32	G=0.3313/1659	PsA	(138)
rs4349859	***HLA-B*2705***	1.00E-200	40.8	A=0.0136/68	AS/ Uveitis / IBD	(19)
rs6457374	***HLA-B*0801***			C=0.1310/656	PS	(138)
rs887466	***HLA-C*0602***	4.06E-214	4.66	A=0.4283/2145	PS	(138)

2.3 Assay Design

2.3.1 Assay Design Suite (ADS)

The web-based software offered by Agena Biosciences, Assay Design Suite (ADS) was used to design a 15-plex assay for the Sequenom MassARRAY. The version used was ADS v2.0.

2.3.1.a Running ADS – Overview

A user log in was created for ADS through agenacx.com, and ADS was accessed through this same website. The genetic variants with corresponding rs numbers were uploaded to ADS via the "Edit Text Input" tab by copy and pasting the rs numbers into the window. The order of the rs numbers was of importance when importing the SNPs as the SNPs imported first would be given priority when incorporated into the design. Due to this variable in the design stage, rs numbers deemed a higher priority for the design, such as the HLA variants were imported first.

The organism for the design was selected as "Human", the database selected was Feb.2009(GRCh37/hg19) and the chemistry selected was "iPLEX". The multiplex level selected for the design was 15, as that was the amount of rs numbers imported. The software then runs through the following five steps: 1) Retrieve and Format Sequences, 2) Find Proximal SNPs, 3) Identify Optimal Primer Areas 4) Design Assays 5) Validate.

Step 1 – Retrieve and Format Sequences. This step transfers the inputted sequence into a SNP Group file format and then displays the sequence corresponding to the Flank Size specified in the advanced settings.

Step 2 – Find Proximal SNPs. This step aligns the SNP sequence with the genome selected, this is complete to specify if other proximal SNPs lie within the SNP sequences flanking region. Proximal SNPs can cause primer design problems and can prevent primers from being designed for the specific SNP sequence.

Step 3 – Identify Optimal Primer Areas. The purpose of this step is to design specific PCR amplification primers for the SNP sequence. Once the primers are designed, they are referenced back to the genome selected in order to ensure that the PCR primers created are specific for the particular SNP sequence.

Steph 4 – Design Assays. This step is the multiplexing step, where the extend primers sequence, directionality and position are created. This step creates a file containing the multiplexed design to use for ordering PCR and extend primers.

Step 5 – Validate. The purpose of this step is to validate the primers created to ensure that they will accurately genotype the SNP sequences. This step also ensures that there are no unintended amplification causing false positives during the mass spectra analysis phase.

These five steps must be completed in order for the design to be complete. After each step is completed you are able to view the steps results and summaries. At each step, you can export the results to Excel or Text Tab Delimited files or can view the results using the ADS browser.

2.3.1.b Advanced Settings

Advanced Settings are offered by ADS to make the design criteria more or less stringent. These settings can maximize ADS's ability to create a design and can also be used to eliminate errors from the design. The settings are divided via which step of the ADS process they affect. The advanced settings that were used that differed from the default settings are as follows:

Step 1 - Retrieve and Format Sequence

Flank Size – default is 100, changed to 300. This is the size of the flanking region on either side of the polymorphism. Increasing this area gives the software more area to create a suitable primer for PCR amplification, however, it does not limit the designing capacities of the design.. Please refer to Table 2.2.

Table 2.2: Advanced Settings for Step 1 - Retrieve and Format Sequence

Step 1 - Retrieve and Format Sequence			
Setting	**Default setting**	**New setting**	**Rationale**
Flank size	100	300	Increasing this area gives the software more area to create a suitable primer for PCR amplification.

Step 2 - Annotate Proximal SNPs

In this step, the advanced settings are divided under two tabs, Matching Constraints and Filtering. Please refer to Table 2.3 for details of changes. The Matching Constraints tab was used to determine how much the SNP sequence has to be aligned and matched to the genome sequence. The Filtering tab allows filtering on the basis of frequency, population and/or validation status.

Table 2.3: Advanced Settings for Step 2 - Annotate Proximal SNPs

Step 2 - Annotate Proximal SNPs			
Setting	**Default setting**	**New setting**	**Rationale**
Matching Constraints tab			
SNP Flanking Sequence	100	300	This is the same as Flank Size in Step 1 Retrieve and Format Sequence. It is very important to keep these values the same.
Filtering tab			
Validation Status	unchecked	checked	This excludes proximal SNPs with statuses not validated.
Filter proximal SNPs based on population	unchecked	checked	This filters the SNPs that are in the flanking regions of specific rs numbers by populations. For the case of this assay, it is only for a European population, therefore SNPs that are in the European population were selected to be included.
Exclude proximal SNPS with frequency below cut off – 0.01	unchecked	checked	This is used to eliminate SNPs from the software that have low population minor allele frequencies. This was checked in this design because there were problems with proximal SNPs that were making it difficult for primers to design.
Exclude proximal SNPs with no population information	unchecked	checked	This is used to eliminate SNPs from the software that do not have any information. This was checked in my design because I had problems with proximal SNPs

Step 3 - Identify Optimal Primer Areas

Please refer to Table 2.4 for changes to this section of the design.

Table 2.4: Advanced Settings for Step 3 - Identify Optimal Primer Areas

Step 3 - Identify Optimal Primer Areas			
Setting	**Default setting**	**New setting**	**Rationale**
Amplicon Length			
Minimum	80	80	This is a very important setting. This increases the size of the possible amplicon (PCR product). Increasing this can increase the area in which the PCR primers can bind, thus making it easier for a primer to be designed. Changing this setting is the first recommended action to eliminate errors caused by proximal SNPs
Optimum	100	100	
Maximum	120	300	

Step 4 - Design Assay

In this step, there are 5 tabs that have corresponding advanced settings that can be adjusted. These tabs are: Quick Fix, General, Amplicon, Extend Primer, and Multiplex. When the Quick Fix tab settings are adjusted, it will change their parameters in the further sections of the advanced settings (See Table 2.5).

Table 2.5: Advanced Settings for Step 4 - Design Assay, Quick Fix Tab

Step 4 - Design Assay			
Setting	**Default setting**	**New setting**	**Rationale**
Quick Fix Tab			
Amplicon Primer Potential			
False Priming	1	100	This number is a penalty for the primer designs that have the potential to hybridize to alternative target sites. A higher value means a better primer – as it will only map to the specific site in the genome
Hairpin/Dimer Extension	1	100	This number is a penalty for primer design so the primer will not extend against themselves and a result in a hairpin or dimer substructure. A higher value means a stricter primer design.
Extend Primer Potential			
Hairpin/Dimer Extension	1	0.9	This number is a penalty for primer design so the primer will not extend against themselves and result in a hairpin or dimer substructure
Multiplex Evaluation Potential			
False Priming	1	0.8	This number is a penalty for the primer designs that have the potential to hybridize to alternative target sites. A higher value means a better primer – as it will only map to the specific site on the specific amplicon. This is not as crucial as the Amplicon primer stage as there are less binding sites on other amplicons for extend primers. This change was suggested in the Agena Protocol as well as by Agena employees.
Hairpin/Dimer Extension	1	0.8	This number is a penalty for primer design so the primer will not extend against themselves as a result of a hairpin or dimer substructure. A higher value means a stricter primer design. Cross-primer dimer potential is the main limiting factor to multiplexing efficiency. This change was suggested in the Agena Protocol as well as by Agena employees.

The Amplicon tab is the section strictly used for design settings of PCR primers and the production of the amplicon (See Table 2.6).

Table 2.6: Advanced Settings for Step 4 - Design Assay, Amplicon Tab

Step 4 - Design Assay			
Setting	**Default setting**	**New setting**	**Rationale**
Amplicon tab			
Minimum	80	80	This is the same as above in Step 3 - Identify Optimal Primer Areas, you must change both to the same values. By expanding the maximum amplicon length it gives more options for the creation of the extension primer.
Optimum	100	100	
Maximum	120	300	

The Extend Primer tab section is strictly used for the design settings of the extend primers (See Table 2.7).

Table 2.7: Advanced Settings for Step 4 - Design Assay, Extend Primer Tab

Step 4 - Design Assay			
Setting	**Default setting**	**New setting**	**Rationale**
Extend Primer Tab			
Minimum	15	17	This for the length of the extend primer. A longer extend primer would hopefully follow a more specific and strict design and would avoid possible false amplifications.
Maximum	30	30	

The Multiplex tab section is used to change the design settings relating to the multiplexing of the assay. Making it possible for multiple SNPs to be in the same plex and not interact with each other (See Table 2.8).

Table 2.8: Advanced Settings for Step 4 - Design Assay, Multiplex Tab

Step 4 - Design Assay			
Setting	**Default setting**	**New setting**	**Rationale**
Multiplex Tab			
Design Iterations			
Number of Iterations	1	10	This is the number of times the software will design the assay. The higher the number the more likely you will get the best design. 10 is the maximum setting available.
Best Iteration Selection Criteria	Highest Average Multiplex	Fewest Rejects by Low Plex	I chose this option as I was more concerned that my inputted SNPs would not be rejected than having a large number of SNPs in one well

Once these advanced settings were adjusted the settings were saved and "Begin Run" was selected in order to start an ADS run.

2.3.1.c ADS Output Design Reports

When ADS has completed running through the five steps, there are a series of reports produced to show how "well" the design will theoretically work. These can be viewed online using the software or can be export as a zip folder containing them (these are using in excel). There are many reports outputted; however, the reports that will be described are the reports relevant to the design stage.

Step 4 Report - Design Summary

The Design Summary gives information of the design parameters, the designs overall statistics and a text summary report of the designed assay. This information from the software was accessed by, selecting the "View Results" tab next to Step 4 Design Assay then select the “Design Summary” tab on the bottom bar. The Design Summary Report is attached in the Appendix 1.A and details of what is included in the report are below (Table 2.9).

Table 2.9: Design Summary Report Contents

Design Summary Report contents
Number and which rs numbers are in each well
Uniplex confidence percentages (this is to determine the probability of how “well” the rs number will work in the reaction)
PCR Primers
Extend Primers
Primer lengths
Primer direction
Primer extension nucleotides
Masses of un-extended primers and extended primers.
A Spectrum of the Assay results of the MassArray
Warnings for possible problems for the rs number in the reaction need to describe some warnings. o D – primer dimer potential between primers of multiplexed assays o H – primer hairpin or self-dimer potential o Dh – both primer dimer potential and hairpin potential.

Step 5 Report - Validation Hits Report

Once you have checked the Design Summary it is very important to check the Validation Hits. This report gives details of each uniplex genomic alignment, how the primers created will map with the genome and the potential for false positives and negatives of each rs number. The Validation Hits report is below in Table 2.10.

Table 2.10: Validation Hits Report from the Summary Reports in ADS.

See Description below in text for further details.

Gene	SNP ID	True Hits	False Hits	Null Hits	PCR2 Hits	PCR1 Hits
HLA-CW6	rs10456057	1	0	0	80	58
CARD9	rs10781500	1	0	0	12	11
FBXL19	rs10782001	1	0	0	71	46
IL23R	rs11209026	1	0	0	102	12
*HLA-B*2705*	rs116488202	2	0	0	61	100
HLA-B60	rs1265163	1	0	0	15	83
ERAP1	rs2032890	1	0	0	112	23
CARD15/NOD2	rs2066844	1	0	0	133	7
ERAP2	rs2910686	1	0	0	81	49
HLA-B44	rs3132528	1	0	0	82	174
TRAF3IP2	rs33980500	1	0	0	73	26
TNFAIP3	rs582757	1	0	0	42	22
ATG16L1	rs6738490	1	0	0	61	29
Intergenic 2p15	rs6759298	1	0	0	33	105
IL12B	rs6871626	1	0	0	65	43

The following is a description of the values generated by the validation report and the importance of each descriptor. True Hits – This is the number of amplicons produced by the PCR primer pair. Therefore, it is important that this value at 1, as a value of more than one will result in more than one amplicon produced. The SNP rs116488202, gave a True Hit of 2, this was checked with Agena and they confirmed that it would be OK, as well as this variant passed the secondary validation that this research group did in addition to using this software. False Hits – This is the number of amplicons produced that contain an invalid target for the extend primer of the same rs number. This is very important to keep at 0, as we need the extend primer to be very specific. Null Hits – This is the number of amplicons produced by the primer pair that do not contain a target sequence for the extend primer to bind. This means that the extend primer will not bind therefore there will be no genotype generated for the rs number. It is very important to keep this value at 0. To summarize it is important to have the True Hits at 1, False Hits at 0, and Null Hits at 0.

Other values that are reported that are of importance are the PCR1 Hits and PCR2 Hits. These are the number of matches to the genome each PCR forward (PCR1) and reverse (PCR2) primer has to the genome. These numbers are usually quite high (20-150) but on secondary validation through NCBI's BLAST and UCSC's Blat, these values mean incomplete matches to the genome. This will be discussed at length in the Secondary Validation section of the Methods.

In the ADS browser under the validation report, there is additional information which is not exported to excel Validation report. By selecting each rs number ADS will give you more detail about each rs number reaction. The information provided is such as the amplicon and its length and SNP position in the genome. This information was double checked in the Secondary Validation to ensure that these positions correspond to the correct genomic position of each SNP of interest.

Step 5 Report - Cross Assay Hits

The Cross Assay Hits report provides information of the amount and types of amplicons that could result from a cross-hybridization reactions from primers in the designed assay. It is important that there are no results for this section. Results will only appear if there is a problem.

2.3.1.d Error Codes and Design Problems related to ADS

If ADS encounters a problem it result in a step error. The error will be recorded as a "Reject", as the error or "Reject" will exclude the problem SNP from the design. If the "Reject" is selected - the software will indicate why the SNP was rejected and will give information that can indicate what will need to be changed in the design for the specific rs number to be incorporated into the specific design.

When designing the assay, a problem was encountered relating to error codes appearing in Step 3, Identify Optimal Primer Areas of the ADS process. This resulted in many SNPs being rejected from the design.

The particular error code that was encountered was "Error 190 – Multiple Extend hits for scanned primer triplet". This error indicates that a PCR amplicon containing a unique site for the extension primer cannot be generated. This is due to proximal SNPs in the binding regions of the primer or could be caused by highly repetitive regions surrounding the SNP of interest. These polymorphisms and repetitive regions make it unable to design primers for the SNP of interest.

This error code was encountered for the all of the HLA variants selected for the design. These errors could be attributed to the HLA region as it is a highly polymorphic and highly repetitive region – making it very difficult to design primers. It was very important for the stringency of the assay design to not be reduced and it was necessary for these variants to be incorporated into the design. Therefore, alternative measures were explored in order to incorporate these variants into the design.

In the advanced settings, the PCR amplicon length can be adjusted to eliminate the problem. Adjusting the amplicon to 1000 (a less stringent setting) did not improve the problem and there was the error code 190 still appeared for the HLA variants. Therefore, the PCR amplicon length was adjusted to 300 and a focus was put on finding a SNP in linkage disequilibrium (LD) with the SNP of interest that could be incorporated into the design.

The SNPs in LD were found using the program tool HaploReg v4 issued by the Broad Institute and MIT (140). HaploReg is a tool for discovering LD variants on haplotype blocks, such as candidate regulatory SNPs at disease-associated loci (140). The LD information is provided by the latest 1000 Genomes Project data. When HaploReg was used the default settings were selected as they were advised by the published paper on this software (140). The following settings were used: an LD threshold, r^2: 0.8. The 1000 Genome Phase 1 population for LD calculation: EUR (European) and the source for epigenomes: Chromm (Core 15-state model). The Mammalian conservation algorithm: SiPhy-omega and the position was relative to: GENCODE genes.

By entering the rs number corresponding to the SNP of interest into the Query box, HaploReg provided a list of SNPs in LD with the SNP of interest. It provided r^2 and D' values, genomic position of the rs number, and additional information of each rs number listed. The HaploReg information was exported into excel and ranked the LD SNP by r^2 and D' values.

A new design in ADS was created, using the same advanced settings as listed above, and each LD SNP was run through the ADS program one at a time. This was repeated with a different LD SNP until an LD SNP did not get rejected and the validation report resulted with True Hits of 1, False Hits and Null Hits of 0. This was a very lengthy process as many of the LD SNPs were rejected by the Error code 190 and the program takes 15 minutes to run for each iteration. Once an LD SNP passed the criteria above, the LD information was recorded and incorporated the design with the other SNPs.

The LD SNP is the best method in my opinion as there was no reduction of the stringency of the assay. The following is a table of the LD SNPs corresponding to the tag SNP that was selected in the original SNP selection. All LD SNPs selected had an $r^2 > 0.95$ and a D' > 0.98 (see Table 2.11).

Table 2.11: LD SNPs that were incorporated into the design**.**

Corresponding r^2 and D' values, MAF and Chromosomal location are listed.

Gene	**SNP in assay**	**Chr. Location**	**Originally Associated tag SNP**	**r²**	**D'**	**Minor Allele Frequency (MAF)**	**Source**
HLA-B60	rs1265163	6p21.3	rs1265110	0.98	1	G=0.1825/914	(137)
HLA-CW6	rs10456057	6p21.33	rs10484554	0.95	0.98	G=0.1116/559	(69)
HLA-B44	rs3132528	6p21.3	rs3130501	0.97	0.98	C=0.2308/1156	(138)
HLA-B39	rs2853931	6p21.4	rs2844603	1	1	T=0.2861/1433	(138)

LD SNP Validation Study

All samples used in the discovery phase of this study were also used in the most recent AS GWAS. A validation was completed in order to determine the concordance of the LD SNPs genotype and the originally selected tag SNP. Please see Results section 3.5 for discussion of LD SNP Validation Study.

2.3.2 Typer 4 – Assay Designer

2.3.2.a Typer 4 Overview

Five HLA variant rs numbers could not be incorporated into an assay design using ADS. All five variants including these variants LD SNPs received the same error code, which was error code 190. These five variants were necessary for the design of the assay. Therefore, these five variants were designed using a different program – which was Typer 4.

Typer 4 is a software program which is accessed offline and was purchased from Agena Biosciences. This program is similar to ADS; however, it has less advanced settings than ADS and computes the assay design much quicker than ADS. Typer 4 can generate a design in less than 10 seconds while it takes ADS approximately 15 minutes. Typer 4 is located on the this research's group shared computer. Typer 4 software offers four different tools, particularly for assay design the program Assay Designer will be discussed and was used for this portion of the study.

When Typer 4 was opened, the "Assay Designer" tab on the Typer home page was selected in order to access the Typer 4 Assay Designer Program. The first step is to import the genetic variants into Typer 4. This step differs from ADS as the variants need to be submitted as a SNP group file which are in a text tab delimited file. The SNP group file used for the Typer 4 analysis was the SNP group file that was generated from ADS and accessed from Step 2 of the ADS process. This SNP group file had a flanking region of 300 nucleotides on each side of the SNP.

Similar to ADS, Typer 4 offers design parameters that can be adjusted in order to maximize the assay design. The "Assay Type" was selected as iPLEX from the dropdown menu and the “Max. Multiplex Level” was 5. Another similarity to ADS, Typer 4 also has advanced settings. The advanced settings are separated into 4 tabs; 1) General, 2) Amplicons, 3) Probes and 4) Multiplexing. There was an effort to ensure that the settings were kept very similar to the advanced settings adjusted using ADS.

For the General tab, nothing was changed from the default settings. In the Amplicon tab, which is dedicated to the design of the PCR primers and PCR product amplicon, there were changes from the default settings. The amplicon length was changed to the following: minimum – 80 (default), optimum – 100 (default), maximum – 300 (default is 120). In the Amplicon tab, there is a tab that can be selected in order to specify the settings further, this tab is “Expert Settings”. In the Expert Settings, the False Priming Potential was adjusted to 100 from the default of 1 and the Hairpin/Dimer Extension Potential was adjusted to 100 from the default of 1. These adjusted values in the Expert Settings are penalties similar to the ones described in the ADS advanced settings section.

In the Probes tab of Typer 4, the tab is dedicated to the design of the extend primers. The oligo length was adjusted; the minimum to 17 from a default of 15 and the maximum to 30 from a default of 28. In this tab, the Expert Settings were also adjusted. The Hairpin/Dimer Extension Potential was adjusted to 0.9 from a default of 1 in the ME Primer Design section of the Expert Settings. This adjusted value is a penalty.

The last tab (Multiplex tab) focused on the design of the multiplexing of the assay. The Expert Settings were adjusted specifically within the Multiplex Evaluation section of this tab. The False Priming Potential was adjusted to 0.8 from a default of 1, and the Primer-Dimer Potential was adjusted to 0.8 from a default of 1, similarly these adjusted values were penalties. These settings were saved by selecting close on each tab. By selecting the

"Reset" button, the advanced settings will return to default. Once all of the above was inputted into Typer 4, the tab "Run" was selected.

2.3.3 Design Report

When the program is finished running in order to view the assay design result, "Design Report" was selected. Typer 4 generates reports, one as a text tab delimited file with all the details including the advanced settings and information regarding the confidence of the assay to perform, as well as the primer details. The other report is exported into excel, this report gives more detailed information about the primers that can be used to import information into the MassARRAY system (Appendix 2.A)

From a design stand point for Typer 4, the most important values to analyze are the multiplex confidence percentage and the uniplex confidence percentage. As well as the warnings detailed in the "Design Report" in the text tab delimited file. For the Assay which was designed in Typer 4, the multiplex confidence was quite high at 93.0% and there were only 2 warnings.

2.3.4 Secondary Validation

It is very important to do a Secondary Validation of the design through online software programs to ensure that ADS and Typer 4 produced a viable design. This in part due to troubles experienced with using that software. This was recommended to us by another group (correspondence Children's Hospital of Eastern Ontario, Ottawa, ON).

2.3.3.a PCR Primer Secondary Validation

Four online software programs were used to validate the PCR primers created by ADS and Typer 4 were specific and sensitive to each rs number. The programs used were National

Center for Biotechnology Information's (NCBI) BLAST, University of California Santa Cruz's (UCSC) Blat, UCSC's In-Silco PCR, and SNPCheck. All of the results from the Secondary Validation are below in the Tables 2.12 and 2.13.

BLAST

NCBI's Basic Local Alignment Search Tool (BLAST) specifically genomic BLAST is a bioinformatic tool that can be used to compare DNA sequences to a reference genome. BLAST can give information regarding primers to see if they are specific to the rs number, it also gives information with how many matches a primer has to the genome and the genomic position of each match.

The PCR forward and reverse primers were copied and pasted into the query sequence box separately from the primer order form outputted by ADS. The human database, and the "Genome (all assemblies top-level)" database was selected – which were the latest databases available. The megablast optimization was selected – this selection ensures that only highly specific sequences to the primers were outputted in the results. Once all of the above the options were selected the tab "BLAST" was selected. The BLAST results were very similar to the PCR1 and PCR2 Hits. However, on closer inspection BLAST reports matches to the genome with sequence mismatches, i.e, the reported result could be a match of 18 out of 20 nucleotides from the primer. The BLAST results also output a genomic location, this genomic location for the complete match was cross referenced with the genomic location of the primer and rs number genomic location outputted by ADS. All PCR primers genomic location were concordant.

Blat

UCSC's Blat is another bioinformatics genomic alignment tool offered online very similar to BLAST. The PCR forward and reverse primers were copied and pasted into the query sequence box. The genome selected was "Human" and the assembly was "Dec.2013 (GRCh38/hg38), and the query type was "DNA". Once all of the above was selected the "submit" tab was selected. Similarly, to BLAST the database used by Blat is more update than the one used for designing primers in ADS. Blat only reported 100% matches to the genome, unlike BLAST. Blat also output's a genomic location for the primer entered into the program. This genomic location was cross-referenced with the genomic location of the primer and rs number outputted by ADS and there was 100% concordance.

In-Silico PCR

UCSC's In-Silico PCR is a computational software tool that is used to theoretically calculate the amplification products of forward and reverse primers. The genome selected was "Human" and the assembly was "Dec.2013 (GRCh38/hg38), and the target was "genome assembly". The forward and reverse primers were copied and pasted into their specific query sequence boxes. When the above was completed the "submit" tab was selected. In-Silico PCR outputted the theoretical amplification product and the genomic location of that product. This result was then cross referenced with the genomic location and the sequence of the amplicon produced by ADS. The number of unique PCR amplification products should be and was the same value as the value of the True Hits for each rs number. Ideally, there should be only one PCR amplification product. All the results were 100% concordant with the results from ADS.

SNPCheck

SNPCheck is a software tool used to check if there are proximal SNPs in PCR primer binding sites. The forward and reverse primers were copied and pasted into the sequence box along with the rs number and the chromosome which the rs number is located on. This secondary validation was important as discussed above in the error code proximal SNP problems when designing the assay.

Table 2.12: Secondary PCR Primer Validation for LBP_1W well.

BLAT, BLAST, In-Sillico PCR Match, and SNP Checker descrptions are all above in section 2.3.3.a Chr denotes chromosome

SNP ID	Gene	Forward Primer Sequence	BLAT	BLAST	Reverse Primer Sequence	BLAT	BLAST	In-Silico PCR - Match	SNP Checker (SNPS found)
rs6759298	*Intergenic 2p15*	ACGTTGGATGAGTTGCAGGCTATTGGTGTC	1	68	ACGTTGGATGCTTTGTGGTGGTTCTGTAGG	1	138	1	1
rs10782001	*FBXL19*	ACGTTGGATGTGTTCCCCTCATAGAGCAAG	1	26	ACGTTGGATGACACAGTTATCTGCTCCCAC	1	46	1	1
rs6871626	*IL12B*	ACGTTGGATGCATTATGGGCTAAGTGGGTG	1	122	ACGTTGGATGGCAGAGAAAGTTACCTGTCC	1	161	1	1
rs1265163	*HLA-B60*	ACGTTGGATGAGAAACTGGCACATCCAAGG	8	248*	ACGTTGGATGTAACCTGACAGGTGTTCTCG	1	54	1	2
rs2032890	*ERAP1*	ACGTTGGATGTAAAGACCCAGTGGTGGGAG	2	101	ACGTTGGATGCATCCTGGCGAAACTCCTTG	1	32	1	0
rs2910686	*ERAP2*	ACGTTGGATGAACTTAAATCCCAGCTCACC	1	210	ACGTTGGATGACAAGTGACCACAATGTGGC	1	72	1	1
rs11209026	*IL23R*	ACGTTGGATGGGGAATGATCGTCTTTGCTG	1	85	ACGTTGGATGGAAATTCTGCAAAAACCTAC	2	255	1	3
rs33980500	*TRAF3IP2*	ACGTTGGATGCTGGGATTGGTTTCAGCAAC	2	177	ACGTTGGATGTGAACCGAAGCATTCCTGTG	1	45	1	1
rs582757	*TNFAIP3*	ACGTTGGATGTAGCCTCATGTGGAATAAGC	1	108	ACGTTGGATGATAAGGCTACCAAGGCCTAC	1	60	1	0
rs10781500	*CARD9*	ACGTTGGATGTCTCTAACCATATCGGAAGC	1	13	ACGTTGGATGATCTGTGGGTTATTTAGCGG	1	72	1	0
rs3132528	*HLA-B44*	ACGTTGGATGAGCCTTATCTTGACCTGTTC	6	63*	ACGTTGGATGCCATTTTAAAAACTTGGGCTC	0	174 **	1	1

rs6738490	*ATG16L1*	ACGTTGGATGGTAAACCTGACGACTTTCTC	1	47	ACGTTGGATGGAGAACTACTGATTTTGCAC	2	177	1	0
rs116488202	*HLA-B*2705*	ACGTTGGATGCCCGCACCAAATTCAGTACA	1	45	ACGTTGGATGACCAAGCCTCAGACCATGC	0	58 **	2	5
rs2066844	*CARD15/NOD2*	ACGTTGGATGAGTGCCAGACATCTGAGAAG	1	48	ACGTTGGATGATGGAGTGGAAGTGCTTGCG	1	72	1	2
rs10456057	*HLA-CW6*	ACGTTGGATGTGTTTTCAGAGGTTCTGGAC	3	241*	ACGTTGGATGGGCACTGCAATATTGAGTTC	8	87*	1	3

* Different results in Blast and BLAT, there was only 1 100% match in BLAST

** Blast showed a 100% match

Table 2.13: Summary of PCR Secondary Validation for HLA well.

Same Description as Table 2.13

SNP ID	Gene	Chr.	Forward Primer Sequence	BLAT	BLAST	Reverse Primer Sequence	BLAT	BLAST	In-Silico PCR - Match	SNP Checker (SNPS found)
rs887466	HLA-C*0602	6	TCCGCACCTATCACACCTAC	6	76 *	AATCCTTCCTGACCTAGAGC	7	205*	7 (2)**	1
rs2853931	LD with HLA-B*3906	6	ACGCTCTTTTCAGGACGATG	7	114*	GCATAGAATATCATGCTGCAC	7	55*	7 (2)**	3
rs6457374	HLA-B*0801	6	TTTCAAACCTCCTGCATCTG	7	264*	CCTAACAGTATGACACTCG	0	120*	5 (3) **	3
rs3129944	HLA-B*3801	6	CTGTGGAGAACAAGGAAGAG	8	547*	TGTGCTTATAAGGTACCCAC	8	114*	8 (2) **	2
rs4349859	HLA-B*2705	6	AAGCAGCCTAATCCCCTTAC	5	92*	AGAGAGCAGTCCTACAAATG	8	218*	5(1) **	0

* Different results in Blast and BLAT, there was only 1 100% match in BLAST
** Blast showed a 100% match

2.3.3.b Extend Primer Secondary Validation

There was no program online to validate the primers against a PCR amplification product, therefore a method was created using Microsoft Excel 2011 and an online software tool reverse complement primer. Extend primer sequences and direction and the PCR product amplicon was obtained from the Design Summary excel file outputted by ADS. Depending on the direction of the Extend primer. Forward primers were entered into the program reverse complement for generation of the complement sequence of the primer. This forward complement sequence was then validated to be within the amplicon by using the control + F function in Microsoft Excel 2011.

For reverse primers, the primer was entered into the program reverse complement for the generation of the reverse complement sequence of the primer. This reverse primer reverse complement sequence was then validated to be within the amplicon by using the control + F function in excel.

2.3.5 Final Design

The final design consisted of a 2-well assay. One well was design using ADS comprising of a 15-plex assay and the other well was designed using Typer 4 and comprised of a 5-plex assay. Below are Tables (2.6.1, 2.6.2, 2.6.3, and 2.6.4) of the PCR Primers for both designs which were ordered and created by Integrated DNA Technologies (IDT) technologies in December 2016 at the concentration of 25 nano moles (nm), 100 micro molar (μM) in 100 micro litres (μL) Below is also a Table of extend primers that were ordered and created by IDT technologies in December 2015 at a concentration of 250nm, 500µM in 40µL, And in March 2016 and Oct 2016 at a concentration of 250nm, 500µM in 80µL.

2.4 Typer 4 – Plate Editor

2.4.1 Importing Design Files into Typer 4

In order to import the designs into the Typer 4, the program “Assay Editor” was opened and a new assay project was created using the Database Browser. This was completed by selecting the “Project Administrator” and adding a new “Assay Project”, for each design the name that was assigned was “LBP_1wdesign_v12” and “LBP_HLAwell”. The design files used for importing was the Design Summary Report from ADS and the Design Summary excel report from the Typer 4 Assay Designer.

2.4.2 Creating a Plate File in Typer 4

The program Typer 4 was selected and on the home screen of the program “Plate Editor” was selected. Under the tab “Plate” the appropriate folder was selected (ROR lab -> LBP). Under that folder LBP was right clicked and then create a new plate was selected and the plate was given a name corresponding to the sample plate name and the assay design which was being run. Under the “Assay” tab, wells were highlighted corresponding to the reaction plate. The appropriate assay design was added to the plate by finding the appropriate folder then selecting the design “LBP_1wdesign_v12” or “LBP_HLAwell” depending on which assay design was run. Under the “Sample” tab, the sample group was uploaded first by right clicking on the folder which is needs to be in and then selecting import sample group. This sample group was given a specific name. This file must have the samples in vertical order, in a single column and in a text tab delimited file. In Typer 4 the wells which will receive the samples must be highlighted. On the right side of the window under “Sample

Tab" settings the following must be specified; Apply sample direction – Vertical, 4 (96) – 1(384) Mode should be "False", keep in selected region should be "False", No Sample ID – CROSS and No Sample ID Color should be 150; 150; 150. Once these settings are selected import the sample group onto the plate/chip by right clicking the sample group file and selecting "Apply Samples from Group…". Once this is all complete the sample plate is ready to be linked to the MassARRAY machine.

2.5 DNA Preparation

DNA was extracted from peripheral blood (performed by a research assistant). This research assistant was trained by Memorial University, and/or University of Toronto and/or University of Alberta specifically to handle human specimens and extract DNA. This DNA extraction was extracted at three different sites; Memorial University, St. John's, NL, University of Toronto, ON, and University of Alberta, Edmonton, AB. The stock DNA concentration was determined using the NanoDrop 1000 (Thermo Fischer Scientific) apparatus and the stock DNA was diluted to ~5-10ng/µL

2.6 Primer Adjustment

Primer mixes were made manually for both PCR and extend primers.

2.6.1 PCR Primer Mix

All the PCR Primers are ordered at the concentration 100µM. Both Forward and Reverse were included into the same mix. In order to make 1 mL of PCR primer mix, 5µL of each primer were added to the mix. The remainder of the 1 mL is HPLC molecular-grade water. The calculations for the PCR Primer Mix are below.

2.6.1.a HLA well

The HLA well is a 5-plex reaction with 5 Forward PCR primers and 5 Reverse PCR primers, for a total of 10 primers. The required primer concentration is 0.5µM. 50 µL of Primers and 950 µL of molecular-grade water was added to the PCR mix. Once this mix was completed it was vortexed spun and aliquoted. See List of PCR Primers for this well below (Table 2.14).

Table 2.14: List of PCR Primers for HLA well

F denotes forward primer, R denotes reverse primer.

SNP Primer ID	Primer Sequence
rs2853931_F	ACGTTGGATGGCATAGAATATCATGCTGCAC
rs2853931_R	ACGTTGGATGACGCTCTTTTCAGGACGATG
rs3129944_F	ACGTTGGATGTGTGCTTATAAGGTACCCAC
rs3129944_R	ACGTTGGATGCTGTGGAGAACAAGGAAGAG
rs4349859_F	ACGTTGGATGAGAGAGCAGTCCTACAAATG
rs4349859_R	ACGTTGGATGAAGCAGCCTAATCCCCTTAC
rs6457374_F	ACGTTGGATGCCTAACAGTATGACACTCG
rs6457374_R	ACGTTGGATGTTTCAAACCTCCTGCATCTG
rs887466_F	ACGTTGGATGAATCCTTCCTGACCTAGAGC
rs887466_R	ACGTTGGATGTCCGCACCTATCACACCTAC

2.6.1.b LBP_1W (15-plex) well

The LBP_1W well is a 15-plex reaction with 15 Forward PCR primers and 15 Reverse PCR primers, for a total of 30 primers. The required primer concentration is 0.5µM. 150 µL of Primers and 850 µL of molecular-grade water was added to the PCR mix. Once this mix was completed it was vortexed spun and aliquoted. See list of PCR Primers below (Table 2.15).

Table 2.15: LBP_1W well PCR Primer Sequences.

F denotes forward primer; R denotes reverse primer.

SNP Primer ID	Primer Sequence
rs10456057_F	ACGTTGGATGTGTTTTCAGAGGTTCTGGAC
rs10456057_R	ACGTTGGATGGGCACTGCAATATTGAGTTC
rs10781500_F	ACGTTGGATGTCTCTAACCATATCGGAAGC
rs10781500_R	ACGTTGGATGATCTGTGGGTTATTTAGCGG
rs10782001_F	ACGTTGGATGTGTTCCCCTCATAGAGCAAG
rs10782001_R	ACGTTGGATGACACAGTTATCTGCTCCCAC
rs11209026_F	ACGTTGGATGGGGAATGATCGTCTTTGCTG
rs11209026_R	ACGTTGGATGGAAATTCTGCAAAAACCTAC
rs116488202_F	ACGTTGGATGCCCGCACCAAATTCAGTACA
rs116488202_R	ACGTTGGATGACCAAGCCTCAGACCATGC
rs1265163_F	ACGTTGGATGAGAAACTGGCACATCCAAGG
rs1265163_R	ACGTTGGATGTAACCTGACAGGTGTTCTCG
rs2032890_F	ACGTTGGATGTAAAGACCCAGTGGTGGGAG
rs2032890_R	ACGTTGGATGCATCCTGGCGAAACTCCTTG
rs2066844_F	ACGTTGGATGAGTGCCAGACATCTGAGAAG
rs2066844_R	ACGTTGGATGATGGAGTGGAAGTGCTTGCG
rs2910686_F	ACGTTGGATGAACTTAAATCCCAGCTCACC
rs2910686_R	ACGTTGGATGACAAGTGACCACAATGTGGC
rs3132528_F	ACGTTGGATGAGCCTTATCTTGACCTGTTC
rs3132528_R	ACGTTGGATGCCATTTTAAAAACTTGGGCTC
rs33980500_F	ACGTTGGATGCTGGGATTGGTTTCAGCAAC
rs33980500_R	ACGTTGGATGTGAACCGAAGCATTCCTGTG
rs582757_F	ACGTTGGATGTAGCCTCATGTGGAATAAGC
rs582757_R	ACGTTGGATGATAAGGCTACCAAGGCCTAC
rs6738490_F	ACGTTGGATGGTAAACCTGACGACTTTCTC
rs6738490_R	ACGTTGGATGGAGAACTACTGATTTTGCAC
rs6759298_F	ACGTTGGATGAGTTGCAGGCTATTGGTGTC
rs6759298_R	ACGTTGGATGCTTTGTGGTGGTTCTGTAGG
rs6871626_F	ACGTTGGATGCATTATGGGCTAAGTGGGTG
rs6871626_R	ACGTTGGATGGCAGAGAAAGTTACCTGTCC

2.6.2 Extension Primer Mix

Extend Primers are ordered at a concentration of 500µM. There is an inverse relationship between peak intensity and analyte mass; therefore, extension primers must be adjusted by concentration in order for low mass SNPs to perform well. Extension primers with a higher mass need a higher concentration in the primer mix then the lower mass extension primers. From the design file SNP's primers extension primers are sorted by weight/analyte mass from lowest weight to highest weight. This design file was obtained from ADS. See tables of extend primers (Table 2.16 and 2.17).

2.6.2.a LBP_1W (15-plex) well

For the LBP_1W the three-tier method was used, which meant that the extend primers were separated into three groups by analyte mass. For the low mass extension primers, the concentration of the primer for the extend primer mix was 5µM, for the medium mass primers the concentration was 10µM and for the high mass the concentration was 15µM. For calculating the volume of each SNP's primer that was added to the initial extend primer mix see the calculations below. Therefore, there was 5 primers in the low mass group, 5 in the medium mass group and 5 in the high mass group. Total water that was added to the mix = total volume – (25µL + 50µL + 75µL) = 500µL – 150µL = 350µL. The extend mix was then vortexed and spun quickly.

Table 2.16: LBP_1W well Extend Primer Sequences and Masses

SNP ID	Sequence	Primer Mass
rs2910686	AATCCCAGCTCACCATTTAC	5050.3
rs116488202	TCAGACCATGCCCAGCCTAGCTTACT	5106.3
rs6871626	CTGTCCTTCATCACTTGG	5127.3
rs2032890	GAGAAACCTGATCCGGTAT	5194.4
rs10782001	ATGAAGGCTTGTCAACA	5531.6
rs6759298	TCTTCCAACACAGTGCC	5683.7
rs6738490	ACTGATTTTGCACAATCAGAATGC	5786.8
rs33980500	TGGGTATGGTTCTGATTCAT	5864.8
rs11209026	CTGCAAAAACCTACCCAGTT	6030
rs1265163	TCTCTTTCTGTCCTTTCAC	6240.1
rs582757	CTGCATTTTTATCCTTTTAGCA	6352.1
rs2066844	GCCAGACATCTGAGAAGGCCCTGCTC	6391.2
rs3132528	CCTGTTCTATTAAAACCTGCCACA	6653.4
rs10781500	GCTAAAAATCGGTAACAGATAT	6775.5
rs10456057	CTGCAATATTGAGTTCATATAACAAG	7383.8

2.6.2.b HLA Well

For the HLA well, since it was a 5-plex reaction all of the extension primers were added at the same concentration of 10μM. For calculating the volume of each SNP's primer that was added to the initial extend primer mix see the calculations below. Total water to be added = total volume – (50μL) = 500μL – 50μL = 450μL. This extend mix was then vortexed and spun quickly. See below the HLA Sequences

Table 2.17: HLA well Extend Primer Sequences

SNP ID	Sequence	Primer Mass
rs887466	TCTACCCTCTCCGGAAA	5090.3
rs2853931	CTGCACATGAAGAAATAGG	5869.9
rs6457374	ACCAGATAGGTTTAGTGGTG	6212.1
rs3129944	AGTCAATAGACACTCAATAAAA	6728.4
rs4349859	TCTTACATGTCTTTGTACTTACT	6945.5

2.6.2.c Quality Control of Extension Primers

From prepared extend mixes 1μL was added to 49μL of molecular-grade water. This was repeated three times and dispensed into three wells of a plate (Axygen PCR Microplate 96-FLT-C). This was done for both designs. The primers were mixed by vortexing and quickly spin plate. Then the film from the plate was removed and place into the Agena Nanodispenser MTP1 position, well A1 forward. The quality control chip was placed into the Nanodispenser. The Nanodispenser was switched from the 6-pin format to the 1-pin format. Follow the spotting and running protocols that are discussed in Nanodispensing, Typer Plate setup, ChipLinker and SpectroAcquire sections of methods for the remaining time frame for the protocol.

Once the running of the chip is completed, Typer 4 was opened and under "File" then "Reports", "Primer Adjustment Report" was selected. The "Primer Adjustment Report" includes an excel file and a coloured histogram file. Using the excel file the average of the three wells using the value of "percent to add" is calculated. Primers with a value of >50 average of percent to add were highlighted. The highlighted average was then multiplied by the original value added, for example., low mass primer will have the average percent to add multiplied by 5μL, a medium mass primer multiplied by 10μL and the high mass primers by 15μL. The amount needed to be added was then added to the mix. The extend mix was then vortexed and spun quickly and the quality control spotting was run again using the same method as discussed above. Once the second run was completed, the Typer 4 "Primer Adjustment Report" was exported and the average percent to add was calculated again. If no average percent to add values were > 50 then the extend mix was complete. However, if there was a primer with an average percent to add was > 50 then the amount to add was calculated again and the mix was respotted and run using the same method as discussed above until there were no primers with >50 average percent to add.

2.7 Reaction

Before initiating Agena MassARRAY reaction bench top and pipets are cleaned with 10% bleach and 70% ethanol. Clean filtered pipet tips are used at every step. All regents were aliquoted and stored at -20°C, these reagents were thawed on ice and mixed and spun quickly before use. All reagents and plates are kept on ice when in use. All lot numbers and expiry dates were recorded.

2.7.1 Amplification/PCR Stage

The samples genomic DNA must be amplified containing the SNP of interest, for each rs number via PCR. The PCR amplification product produces an amplicon containing the SNP of interest and binding sites for the extend primers to bind to. First a PCR master mix was made and contained: High Performance Liquid Chromatography (HPLC) molecular-grade water, 10X PCR Buffer, 2mM $MgCl_a$, 25mM dNTP, 5u/μL Agena PCR Enzyme and 0.5uM PCR Primer Mix. These reagents have been optimized to ensure that proper DNA amplification will occur. The PCR master mix was then mixed and spun quickly. See Table 2.18 below with ratio volumes and concentrations. An overhang for the reaction of 15-20% was used to prevent pipetting errors. PCR Master Mix was dispensed per well at 3 µL on the plate used for the reaction. This was done by using a P20 multichannel pipet and the PCR master mix was divided by 8 and equally divided into each tube on the strip. The plate was visually inspected to ensure that all PCR Master Mix was dispensed appropriately into the plate. Next 2 µL of sample DNA was dispensed per well, at a concentration of ~5-10ng/µL. A new filtered pipet tip was used for each well to prevent cross-contamination.

Table 2.18: PCR Master Mix combination ratio per well.

1X indicates the ratio per well reaction

Reagent	**1x**
HPLC grade water	0.80
10X PCR Buffer	0.50
25 mM MgCl2	0.40
25 mM dNTP	0.10
5 u/µl Sequenom PCR Enzyme	0.20
PCR Primer (0.5uM)	1.00
Total Master Mix per well	3.00
DNA per well	2.00

Once all the PCR master mix and DNA is dispensed into the appropriate wells, the plate was tightly covered with film. The plate was then gently vortexed and spun quickly. The plate was then placed in a ThermoFischer Verti Thermocycler. The program used for the PCR stage was required by Agena Biosciences; 95°C for 2 min, (95°C for 30ssec, 56°C for 30 sec, 72°C for 1 min, for 45 cycles), 72°C for 5 min, then held for 4°C until collected from the thermolcycler. Table 2.19 thermocycler protocol. Once the thermocycler program ended the reaction plate could be held at 4°C or stored at -20°C for up to two weeks.

Table 2.19: Thermocycler program and cycling

Abbreviations: Temperature (Temp), Seconds (sec)

Temp ˚C	**Time (sec)**	
95	**2 min**	
95	**30 sec**	**45 cycles**
56	**30 sec**	
72	**1 min**	
72	**5 min**	
4	∞	

2.7.2 Shrimp Alkaline Phosphatase (SAP) Stage

The SAP stage is a cleaning step used to remove any unincorporated nucleotides.

The SAP master mix was prepared in a 1.5 ml micro-centrifuge tube and contained: HPLC molecular-grade water, 10X SAP Buffer, 5u/µL Agena SAP Enzyme. The SAP master mix was kept on ice and mixed and spun quickly.

See Table 2.20 below with reaction volumes and concentrations. An overhang of 15% was used to ensure there was enough master mix if there was a pipetting error.

Table 2.20: SAP Master Mix Combination and Ratio

1X indicates the ratio per well reaction

Reagent	**1x**
HPLC grade water	1.53
SAP Buffer	0.17
SAP Enzyme	0.30
Total per well	2.00

The PCR-treated reaction plate was removed from the thermocycler at 4°C and was quickly spun. The film on the reaction plate was removed very carefully using a special technique. The special technique went as follows: Hold plate firmly and pull film from side with no samples on the plate, peal the film carefully and slowly off one column at a time and pullet a 180° angle. It is very important to do this carefully to ensure no reaction mixture splatters and cross-contaminates another well. This technique was used each time the film was removed.

Then 2 μL of the SAP master mix was dispensed per well onto the reaction plate. The plate was visually inspected to ensure that all SAP Master Mix was dispensed appropriately into the plate. A new film was then tightly sealed onto the reaction plate and the plate was then mixed gently and spun quickly. The reaction plate was then placed into the ThermoFishcer Verti thermocycler. The program used for the SAP stage was; 37°C for 40 min, 85°C for 5 min, then held for 4°C until collected from the thermocycler. Table 2.21 thermocycler protocol.

Table 2.21: SAP thermocycler program and cycling

Abbreviations: same as 2.19

Temp ˚C	Time
37	40
85	5
4	∞

Once the thermocycler program ended the reaction plate could be held at 4°C or stored at -20°C for up to two weeks.

2.7.3 Extension Stage

The extend master mix was prepared and contained: HPLC molecular-grade water, iPLEX 10X Buffer Plus, iPLEX 10X Termination Mix, Extend Primer Mix, iPLEX Pro Enzyme 32 u/µL. The extend master mix was kept on ice and mixed and spun quickly. See Table 2.22 below with reaction volumes and concentrations. An overhang of 15% was used to ensure there was enough master mix if there was a pipetting error. The SAP-treated reaction plate was removed from the thermocycler at 4°C or thawed from -20°C and was quickly spun. The film on the reaction plate was removed very carefully using the special technique described above. Then, 2 µL of the extend master mix was dispensed per well onto the reaction plate using a clean pipet tip each time. The plate was visually inspected to ensure that all extend master mix was dispensed appropriately into the SAP-treated reaction plate. A new film was then tightly sealed onto the plate and was gently mixed and spun quickly. The reaction plate was then placed into the ThermoFischer Verti thermocycler. The program used for the extend stage was required by Agena Bioscience; 95°C for 30 sec, [94°C for 5 secs, (52°C for 5 secs, 80°C for 5 sec for 5 cycles), for 40 cycles] 72°C for 3 mins and then held for 4°C until collected from the thermocycler. Table 2.23 thermocycler protocol. Once the thermocycler program ended the reaction plate could be held at 4°C or stored at -20°C for up to two weeks.

Table 2.22: Extension Master Mix Combination and ratio

1X indicates the ratio per well reaction

Reagent	**1x**
HPLC grade water	0.62
iPLEX-PRO Buffer	0.20
iPLEX Termination Mix	0.20
iPLEX PRO Enzyme	0.04
Extend Primer	0.94
Total per well	2.00

Table 2.23: Extension thermocycler program and cycling.

Abbreviations: same as 2.19

Temp ˚C	**Time (sec)**	**Cycling**	
94	**30**		
94	**5**		**40 cycles**
52	**5**	**5 cycles**	
80	**5**		
72	**180**		
4	∞		

2.7.4 Resin Stage

The resin Stage is completed to deionize the reaction contents so that the reaction can be performed on the MassARRAY system. The reaction plate was centrifuged for a quick spin. Clean resin was carefully spooned out (~3 spoonful's) onto a clean, dry dimple plate (that is mirrored to the 96 well reaction plate). Starting at one end a scraper was used to spread the Clean Resin out along the dimple plate. It was made sure that all necessary wells (depending on the sample number) were filled with the Clean Resin. Excess Clean Resin was scraped off and was returned to the stock container. The resin plate was dried for 10-12 minutes at room temperature. While the resin was drying, 41 μL of HPLC molecular-grade water was added to each well of the reaction plate and was centrifuged at 2500 g for 1 min. When the resin plate is dried, the reaction plate was gently inverted. The reaction plate was aligned with the resin plate to ensure resin would be in each reaction well. By tightly pressings the reaction and resin plates together, both plates were inverted so the resin dropped out of the dimple plate and into the reaction plate. If the resin did not come out of the resin plate, the plate was tapped until it fell out. The resin plate was dimple plate was removed and cleaned. The reaction plate was sealed firmly and then rotated for 15-45 mins at room temperature. The rotator rotated the reaction plate along a 360° axis. Once the rotation was complete, the reaction plate was centrifuged for 5 mins at 2000g.

2.8 Nanodispensing iPLEX Assay samples onto chips

In this step of the procedure the reaction mix was dispensed onto the MassARRAY chip. The supply and waste tanks and ultrasonic wash supply bottle was checked to confirm that the Agena Nanodispenser was ready for use. From the Nanodispenser computer home screen the "Tools" icon. "Sonicator Drain" was selected, which drains the 100% ethanol the pins were soaking in. “Sonicator Fill” was then selected, this moves the pins out of the a=way to get the container. The container was then removed and filled with 50% ethanol. The container is then inverted and put back into place. The pins are then cleaned 5 times with the 50% ethanol that was just placed in the container in the machine. Once this was completed “soak” was selected. The “Tools” icon was selected again and the 6-pin format was selected.

2.8.1 Mapping

The mapping step was used to make sure that the Nanodispenser dispenses the reaction mix in the appropriate place on the chip. The “Mapping” icon was selected and a new mapping method was created and was used for the all runs completed. In the mapping methods. “Tuning” was selected. The 96 well plate (96 MTP) to 96 spectochip-1 was selected using a 6-pin format.

2.8.2 Method

The method stage purpose was to ensure that the proper methods are used to ensure the Nanodispenser is working optimally as well as in a standard procedure. From the home screen, the “Method” icon was selected and a new method was created and used for the remaining runs. Under the first tab “setup”, sample tracking was not selected, auto-tuning was enabled and a target volume was selected at 14 nL, and the Volume check was enabled with a lower limit of 8 nL and an upper limit of 18 nL. Under the second tab which was “cleaning set-up” the following was selected; pre-rinse, wash, post-rinse, dry (pre), dry (wash), and dry (post) which a pre-transfer cleaning cycles of 5. The rinse time was 3.5, the wash 5 and the 0.2 for dry. Under the third tab “aspiration/dispensing” in the operation section: analyte and calibrant was selected, the spotting was selected as wet, the calibrant section was adjusted to a dispense of 1 and a speed of 150. The aspirate settings were adjusted to a time of 5, offset of 6.75 and speed of 60 and the dispense setting were set at a time of 0.2 and an offset of 1.0 and a speed of 100 – it was very important that these settings were kept at these values to ensure optimal volumes dispensed onto the chip.

2.8.3 Transfer

At the transfer stage the Nanodispenser transferred the reaction mix onto the chip. The chip's package was open carefully lifted out of the package and placed in the upper left slot with the barcode facing forward. The calibrant was brought to room temperature and 60μL was added to calibrant holder and placed in the Agena Nanodispenser. The reaction plate's film was removed using the special technique discussed above and placed in the plate rack with the title A1, which was on the left side. It was confirmed that the reaction plate was securely placed in the plate rack. The created "Method" file was then opened and then "Run" was selected. It was then confirmed that the Agena Nanodispenser's rinse station was operating. Once this was completed the run was initiated. Volumes and speed volumes were monitored during the spotting by selecting the tuning and volume tabs.

2.9 Chip Linker

The purpose of the Chip Linker step was to ensure that the Nanodispenser Chip layout is imported onto the MassARRAY, so the software can connect the Genotyping result with the Sample ID. The plate that was going to be run on the MassARRAY was selected on the left-hand side of the screen. Selected the Terminator chemistry as "iPLEX", the process was selected as "Genotype + Area", the dispenser was selected as 96-96 and the experiment name corresponded to the project which was "LBP". The Chip Barcode for Chip Linker was the barcode provided on the chip. This was saved. Chip Linker connects the plate information to the MassARRAY.

2.10 SpectroAcquire

At the SpectroAcquire step the MassARRAY genotyped the samples. The MassARRAY machine was opened and the stage was taken out of the machine. The chip that needed to be run was placed into the stage onto slot 1, if only one chip was being run a blank chip was placed onto slot 2. The stage was then placed back into the machine and then sent into the MassARRAY machine. The SpectroAcquire program was then selected on the desktop and the tab "Automatic run set-up" was selected. A "Barcode Report" then confirmed that the machine had the right information. The status of the machine was checked in order to ensure that the pressure within the machine had stabilized (this usually elapsed to be 10 mins). Then "Autorun" was selected and the MassARRAY initiated the run.

2.11 Plate Data Analysis

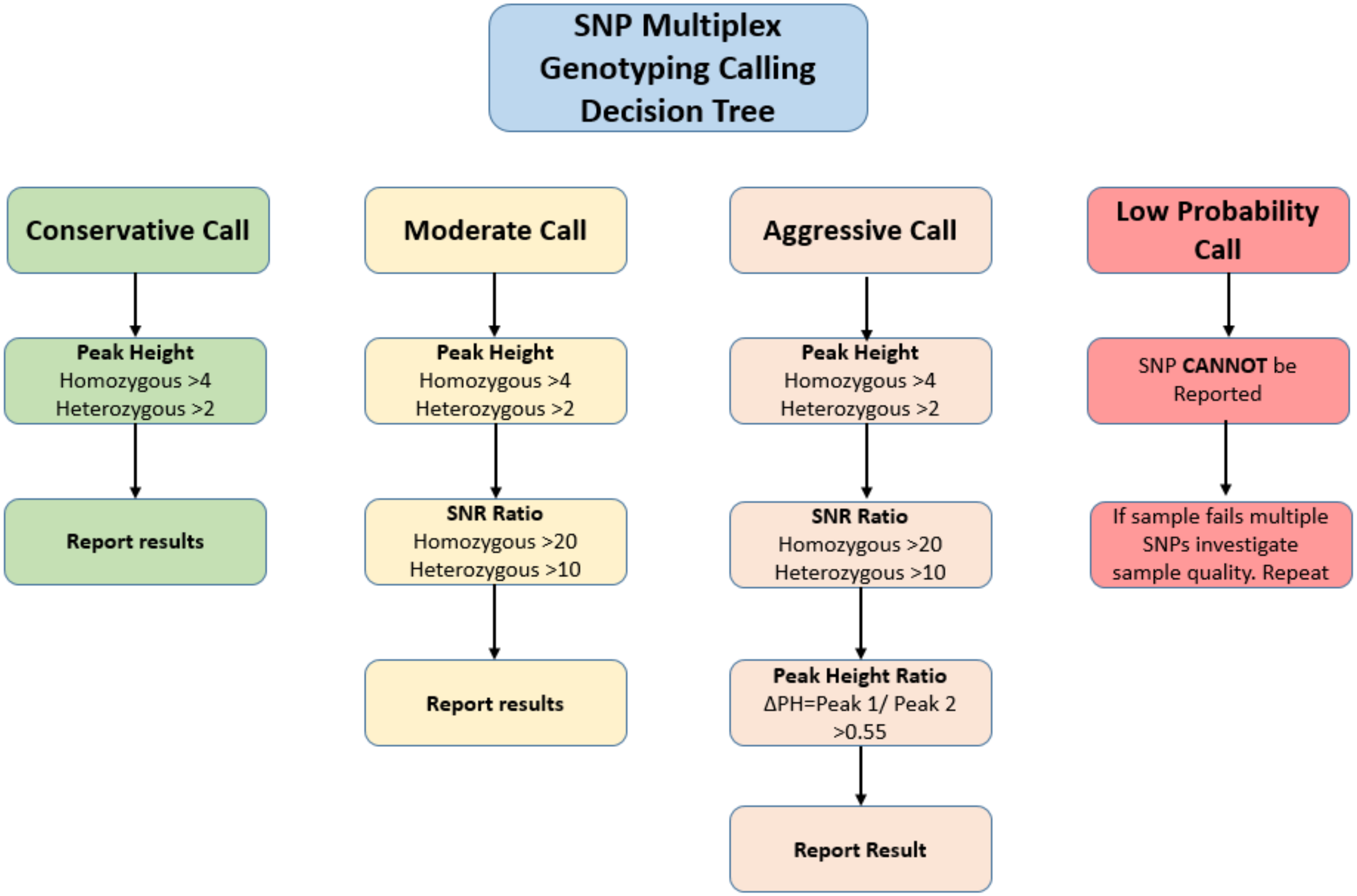
2.11.1 Typer 4 – Typer Analyzer

First pass data analysis was performed using the Typer 4 software specifically the "Typer Analyzer" program. Once Typer Analyzer was opened, the Assay of interest was selected under the "Chip List" section on the right-hand side of the program. Using the "Traffic Light pane" NTC's were selected to determine if there were any genotyping calls in these controls. Then each SNP's assay was examined in the "Post Processing Clusters" specifically the "Call Cluster Plot" to determine how the assay was clustering, the height of the peaks and the SNR of each assay. This was first completed by analyzing the high mass peak height vs. low mass peak height (log axes) and then yield vs. skew plot was analyzed to determine if the assay had sufficient yield and was performing adequately.

While analyze this Call Cluster Plot, the Detail pane was analyzed for the samples selected showing peak height, SNR and call probability. Calls that clustered far out of cluster from the other calls of that specific genotype and samples that had low SNR and low peak heights were manually changed to a “No Call” distinction. All changes were recorded in a log book and the software changed the call description to a “User Call”. See next section for detailed calling algorithm.

2.11.2 Genotype Calling Decision Tree

A variety of Quality Control metrics were applied to the genotyping calls for the samples used in this project. This Genotyping calling decision tree metrics were determined by reviewing trends and investigating this projects assay, as well as 4 other assays performance on the MassARRAY platform. This was done as a collaboration by Memorial University and Eastern Health’s Medical Genetics Laboratory. It was important to determine a calling algorithm that would give trusted results. Figures outlining this calling decision tree are below (Fig 2.2, 2.3, 2.4). Agena Biosciences was contacted and specific metrics of the Typer 4 software was outlined and based on this information the decision tree quality metrics. The first decision in this genotyping calling decision tree was “Call Description”. This metric encompasses many different quality control metrics such as peak height, Signal to Noise Ratio (SNR), Call Probability and Distribution to give a qualitative standard. The next standard that had to meet was peak height, followed by SNR, then Peak Height Ratio (ΔPH). 15 out of the 18 assays followed the genotyping decision tree in Figure 2.2. After reviewing the data extensively and comparing these metrics three assays required different genotyping calling decision trees, this was based on assay performance and genotyping clustering. One SNP (rs6871626) followed Figure 2.3, while two SNPs, (rs1265163 and rs6457374) followed Figure 2.4.


SNP Multiplex Genotyping Calling Decision Tree

Conservative Call

Peak Height
Homozygous >4
Heterozygous >2

Report results

Moderate Call

Peak Height
Homozygous >4
Heterozygous >2

SNR Ratio
Homozygous >20
Heterozygous >10

Report results

Aggressive Call

Peak Height
Homozygous >4
Heterozygous >2

SNR Ratio
Homozygous >20
Heterozygous >10

Peak Height Ratio
ΔPH=Peak 1/ Peak 2
>0.55

Report Result

Low Probability Call

SNP **CANNOT** be Reported

If sample fails multiple SNPs investigate sample quality. Repeat

Figure 2.2: SNP Genotyping Calling Decision Tree for 15 out of 18 SNPs in assays

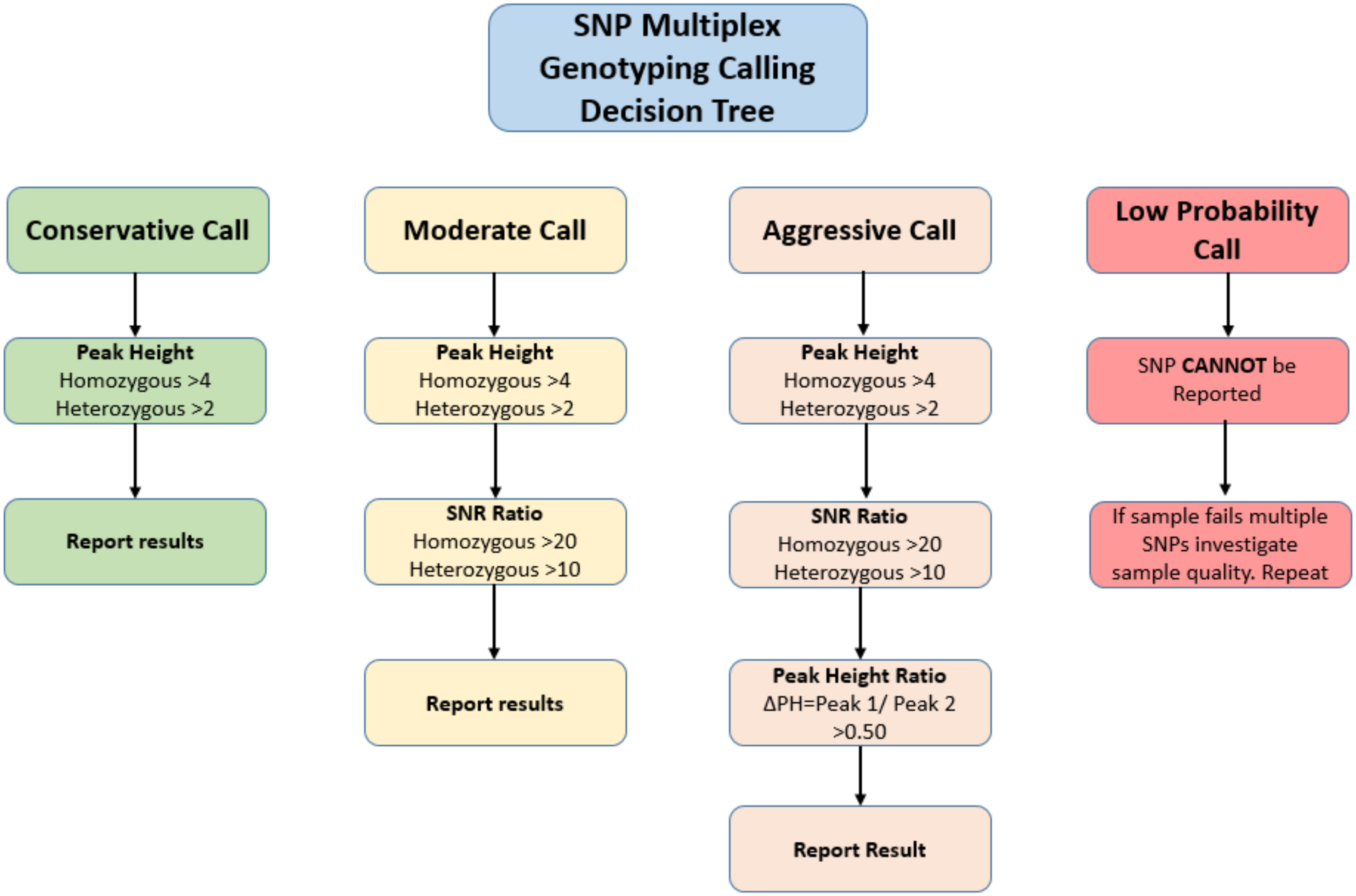

SNP Multiplex Genotyping Calling Decision Tree
Conservative Call
Peak Height Homozygous >4 Heterozygous >2
Report results
Moderate Call
Peak Height Homozygous >4 Heterozygous >2
SNR Ratio Homozygous >20 Heterozygous >10
Report results
Aggressive Call
Peak Height Homozygous >4 Heterozygous >2
SNR Ratio Homozygous >20 Heterozygous >10
Peak Height Ratio ΔPH=Peak 1/ Peak 2 >0.50
Report Result
Low Probability Call
SNP CANNOT be Reported
If sample fails multiple SNPs investigate sample quality. Repeat

Figure 2.3: SNP Genotyping Calling Decision Tree for rs6871626.

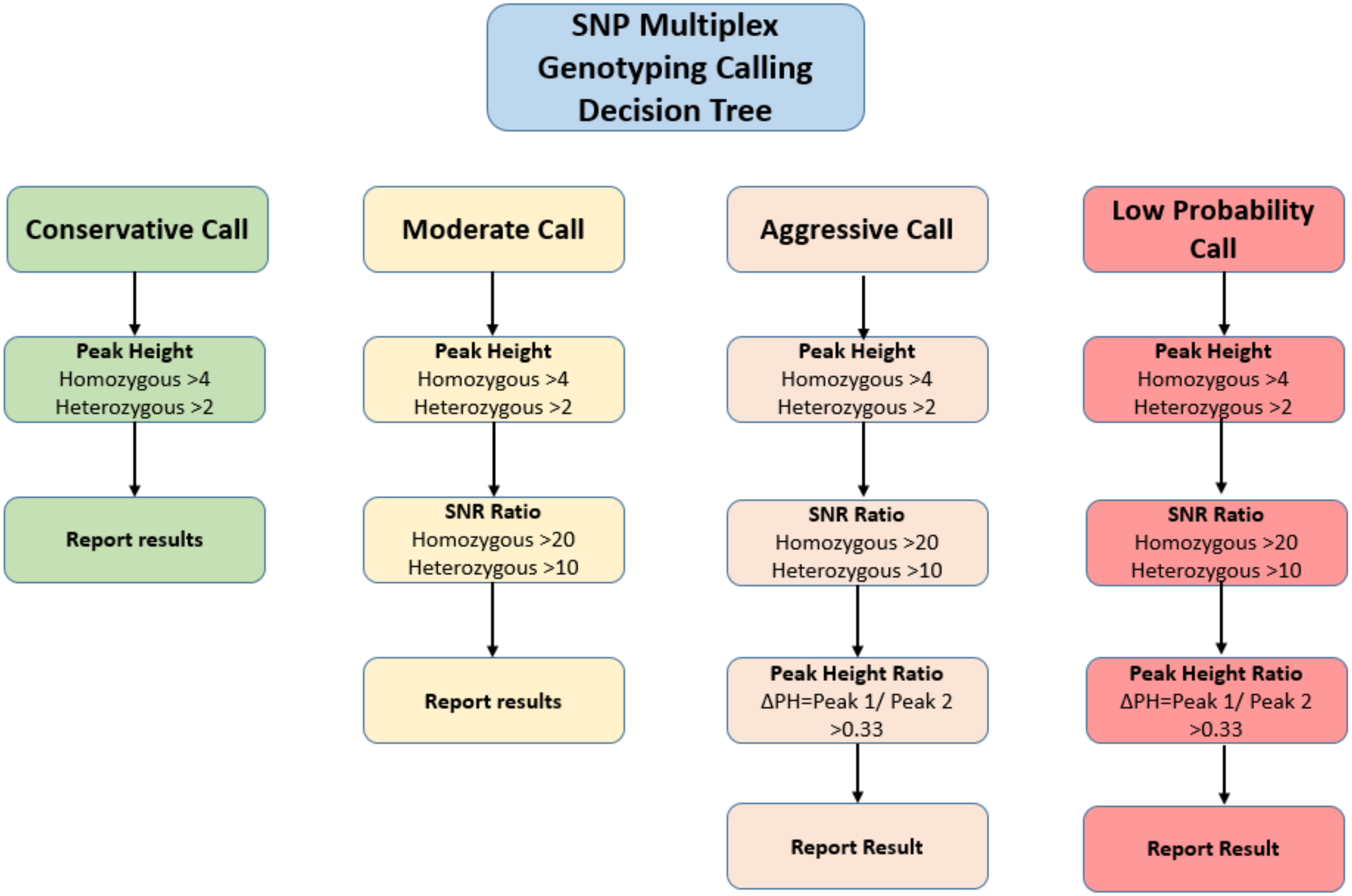


Figure 2.4: SNP Genotyping Calling Decision Tree for rs1265163 and rs6457374.

2.11.3 Plate Data Pane formatting

The Plate Data Pane was accessed for each experimental plate processed. This was done by accessing Typer 4 and opening the Plate Data Pane on the display browser. Once the Plate Data Pane was opened it was copied and pasted into an excel document and saved individually. All Plate Data Panes for every experiment were then combined on the basis of 4 categories; Discovery LBP_1W well, Discovery Cohort HLA well, Validation Cohort LBP_1W well and Validation Cohort HLA well. Once the Plate Data Panes were combined into these categories they were processed on the basis of the above quality control metrics. The data was first filtered, then each rs number was separated onto its own sheet within each category document. Once the quality control metrics were applied all un-necessary information was removed. The only information that was left was rs number, genotyping call and sample ID. All samples with less than 95% genotyping information were removed from the analysis.

2.12 Assay Optimization

2.12.1 Discovery Optimization

Before the Discovery Cohort was initialized the assay was optimization. The quality control metric that was used for optimization was peak height. A cohort of case samples was used for this optimization. These samples were all run twice by two different users (Amanda Dohey and Rebecca Power). A summary peak height table was constructed which was separated by each rs number and each call (Homozygous wild-type, homozygous mutant and heterozygous). Under each rs number heading the peak heights were summarized via average, standard deviation, minimum and maximum (See results section

3.3). Each rs number was examined and it was determined that three rs numbers did not meet quality control metrics. These rs numbers were rs1265163 (*HLA-B*60*), and rs582757 (*TNFAIP3*), rs6871626 (*IL12B)*. rs1265163 had a low yield, meaning a high proportion of un-extended primer was leftover in the reaction. From a literature review and reviewing the Agena standard protocol this was indicative of a low PCR efficiency reaction. Therefore, 50% more PCR primer was added to a 1mL PCR primer mix for rs1265163 (2.5uL for each forward and reverse primer). Both rs582757 and rs6871626 had low peak heights, from experience and review of the Agena standard protocol 50% more extend primers were added for each rs number. Rs6871626 was a low mass extend primer and rs5872757 was a medium mass extend primer, therefore there was a different volume added to the extend primer mix. These adjustments were made and then tested on the same cohort of samples and there was a marked improvement in the performance of all three rs numbers. This analysis was completed in Microsoft Excel and Typer 4.

At the end of the discovery section genotyping two SNPs were removed from the panel as they were removed. Two SNPs were included for both *HLA-B*27* and *HLA-Cw6* variant. It was only necessary to include one SNP for each variant therefore the best performing SNP for each was chosen (See results section). Rs4349859 was chosen for *HLA-B*27* as it performed excellently with high yield and signal. Rs887466 was chosen for *HLA-CW6* as it was the tag SNP not an LD one like rs10456057, and it performed excellently with high yield and signal. These changes made the LBP_1W into a 13-plex reaction from a 15-plex reaction.

2.12.2 Validation optimization

Once the discovery cohort genotyping was completed, summary tables were constructed (see tables in results section 3.3). Similar to the discovery optimization, peak heights were determined to be the quality control metric for optimization.

2.12.2.a HLA-B60 optimization

Rs1265163 performed poorly in the discovery cohort genotyping. Many samples fell below quality control metrics and had to be repeated. Agena was consulted about the issue. Agena's response was that the extend primer for rs1265163 had low ionization efficiency. This extend primer had an enrichment of thymine's, this enrichment of thymine's gives the primer a positive charge causing difficulties with the ionization ability. Furthermore, causing the primer not to fly as high on the MassARRAY causing low peak heights. This function cannot be controlled for on either of Agena's assay design software platforms, ADS or Typer 4 Assay Designer. Agena recommended to redesign the extend primer stating a combination of a poor PCR reaction and low ionization efficiency of the extend primer caused a poor reaction. Since the original extend primer was designed in the reverse direction, Agena suggested to design a primer in the forward direction. This change can be made by entering the original design in ADS, selecting "Current input: View", then selecting "SNPs" next to the SNP of choice (rs1265163). A window will appear and option for the direction of the extend primer will be available, by selecting forward, the primer will only be designed in the forward direction. The subsequent design, also designed new PCR primers for this new extend primer (see Table 2.24 below). The new primer set was named rs1265163F (F for forward) and the old primer set was named rs1265163R (R for reverse). These primers followed the same quality control metrics as previously outlined in methods section 2.3.4.

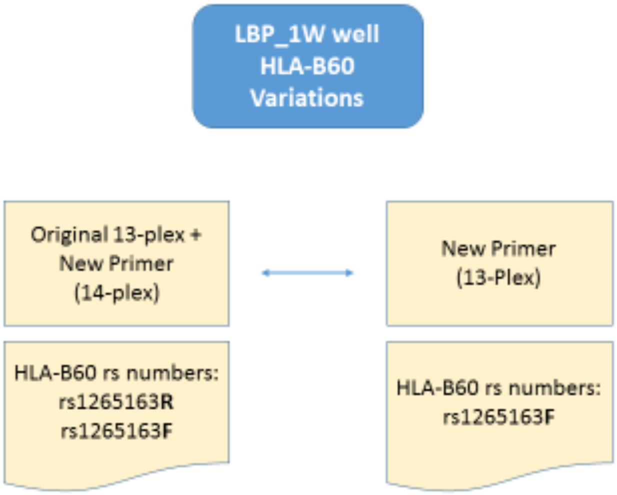
Table 2.24: Replication HLA-B*60 Optimization PCR Primers

rs1265163 Primer Direction	**Amplicon length**	**Forward PCR Primer**	**Reverse PCR Primer**
Reverse	103	ACGTTGGATGAGAAACTGGCACATCCAAGG	ACGTTGGATGTAACCTGACAGGTGTTCTCG
Forward	117	ACGTTGGATGTAACCTGACAGGTGTTCTCG	ACGTTGGATGACTACTCTTCCCCCAGAAAC

In order to compare the performances of the new primers we made 2 experimental groups in order to determine the effectiveness of the new primers. Two groups to determine the performances in the LBP_1W well.

For the LBP_1W well, the original reaction was a 13-plex reaction. The first group had both rs1265163 primer sets in it (forward rs1265163 and reverse rs1265163). The group was titled “Original 13-plex + New Primer (14-plex)”. The second group only had the new rs1265163F primer set, it was titled “New Primer 1W (13-plex)”. (See Figure 2.25)

Table 2.25: LBP_1W *HLA-B*60* Variations experimental group's comparison.


LBP_1W well HLA-B60 Variations
Original 13-plex + New Primer (14-plex)
New Primer (13-Plex)
HLA-B60 rs numbers: rs1265163R rs1265163F
HLA-B60 rs numbers: rs1265163F

PCR primer mixes were made and a new extend primer adjustment was completed for all experimental groups. The first round of analysis on the optimization was performed on 22 case samples that had genotyping information from the discovery cohort. Two separate runs on different dates of the four groups were run. 11 samples were replicated for each experimental group for each separate run (See Tables 2.26, 2.27, 2.28).

Table 2.26: Peak Height Comparisons of LBP_1W well Optimization Experimental Groups for Homozygous Wild-Type Call

See description in text above

		Original 13plex + New Primer (14-plex)		New Primer (13 plex)
Sample Id	**Call**	**rs1265163 F**	**rs1265163 R**	**rs1265163 F**
NF_AS_001_0001	C	10.6	11.1	9.4
NF_AS_002_0001	C	10.9	18.9	14.9
NF_AS_003_0001	C	13.8	23.4	6.8
NF_AS_004_0001	C	*n/a*	*n/a*	12.2
NF_AS_006_0001	C	16.6	24.3	*n/a*
NF_AS_005_0001	C	11.7	21.7	13.4
NF_AS_009_0001	C	10.1	17.7	17.6
NF_AS_010_0001	C	13.8	21.3	9.0
NF_AS_012_0001	C	10.1	20.7	15.4
AS_NF_122-0001	C	7.7	9.6	7.5
AS_NF_132-0001	C	9.2	8.9	8.9
	AVG	11.5	17.8	11.5
	SD	2.5	5.5	3.5
	MIN	7.7	8.9	6.8
	MAX	16.6	24.3	17.6

Table 2.27: Peak Height Comparison on LBP_1W well Optimization Experimental Groups for Homozygous Mutant Comparisons.

See description above in text.

		Original 13plex + New Primer (14-plex)		New Primer (13 plex)
Sample Id	**Call**	**rs1265163F**	**rs1265163R**	**rs1265163F**
NF_AS_030_0001	G	11.6	18.3	10.9
1-0340	G	*n/a*	*n/a*	*n/a*
1-0436	G	4.5	11.6	6.3
	AVG	8.0	15.0	8.6
	SD	3.5	3.4	2.3
	MIN	4.5	11.6	6.3
	MAX	11.6	18.3	10.9

Table 2.28: Peak Height Comparison on LBP_1W well Optimization Experimental Groups for Heterozygotes Comparison

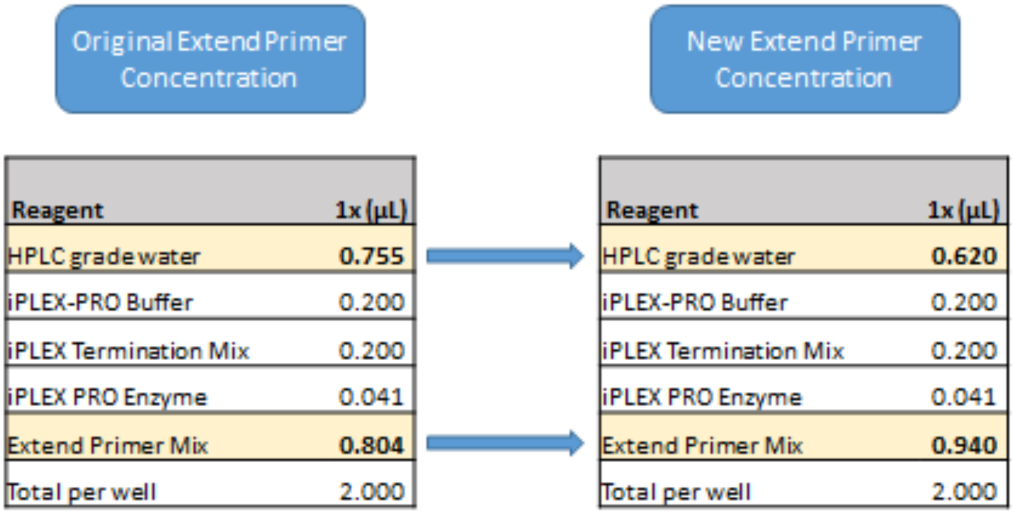
See description above in text.

		Original 13plex + New Primer (14-plex)				New Primer (13 plex)	
Sample Id	**Call**	**rs1265163F**		**rs1265163R**		**rs1265163F**	
NF_AS_007_0001	CG	4.1	5.5	4.1	5.8	4.5	4.6
AS_NF_127-0001	CG	3.4	6.0	7.8	8.6	2.2	4.1
AS_NF_138-0001	CG	2.2	3.6	*n/a*	*n/a*	5.1	3.0
AS_NF_144-0001	CG	2.0	2.8	3.5	4.7	3.5	3.0
AS_NF_148-0001	CG	*n/a*	*n/a*	4.1	3.3	*n/a*	*n/a*
AS_NF_165-0001	CG	4.1	5.6	5.6	7.0	*n/a*	*n/a*
AS_NF_191-0001	CG	2.6	4.7	5.4	4.6	3.2	4.8
AS_NF_203-0001	CG	2.5	3.4	2.0	3.0	3.7	4.3
	AVG	3.0	4.5	4.6	5.3	3.7	4.0
	SD	0.8	1.2	1.7	1.9	0.9	0.7
	MIN	2.0	2.8	2.0	3.0	2.2	3.0
	MAX	4.1	6.0	7.8	8.6	5.1	4.8

The conclusion from this comparison between the “Original 13plex + New Primer (14-plex)” group and the “New Primer (13-plex)” group, was that the primer set rs1265163F consistently performed poorly compared to the original primer set rs1265163R. Therefore, the option to have the primer set rs1265163F was eliminated. It was noticed that rs1265163R performed significantly better when in a well with rs1265163F’s primer set. From this observation, it was determined to keep rs1265163F’s PCR primers as these PCR primers significantly boosted rs1265163R’s performance. To conclude the final combination of primers for the LBP_1W well was rs1265163R and rs1265163F’s PCR Primers and only rs1265163R’s extend primer.

2.12.2.b Extend Primer concentration changes

Once optimization of the *HLA-B*60* primers was completed there was still other variants that needed to be optimized. A new extension mix calculation was released from Agena when the Discovery cohort was complete. This extension mix had a higher concentration of extend primer in the mix and a lower amount of water. This change is illustrated in the Figure 2.5 below; this was an 18% increase in the extend primer mix in the total extension mix.



Original Extend Primer Concentration

Reagent	1x (µL)
HPLC grade water	**0.755**
iPLEX-PRO Buffer	0.200
iPLEX Termination Mix	0.200
iPLEX PRO Enzyme	0.041
Extend Primer Mix	**0.804**
Total per well	2.000

New Extend Primer Concentration

Reagent	1x (µL)
HPLC grade water	**0.620**
iPLEX-PRO Buffer	0.200
iPLEX Termination Mix	0.200
iPLEX PRO Enzyme	0.041
Extend Primer Mix	**0.940**
Total per well	2.000

Figure 2.5: Changes in Extension Mix Calculation.

Abbreviations same as Table 2.22

In order to test that this change was beneficial for the assay. Both wells were tested using 22 samples from the discovery cohort (these were different samples than were used in the replication optimization). Both wells were tested in the original extend primer concentration and the new extend primer concentration. For the LBP_1W well there was increased peak heights using the new extend primers concentration. This was measured by comparing the same samples with the old extend primer concentrations versus the new concentration (Table 2.29). For the HLA well there was no observed improvement or benefit (Table 2.30). A decision was made to use the increased extend primer concentration for the Replication phase. A decision was made to also use the increased extend primer concentration, as to be consistent with the other well.

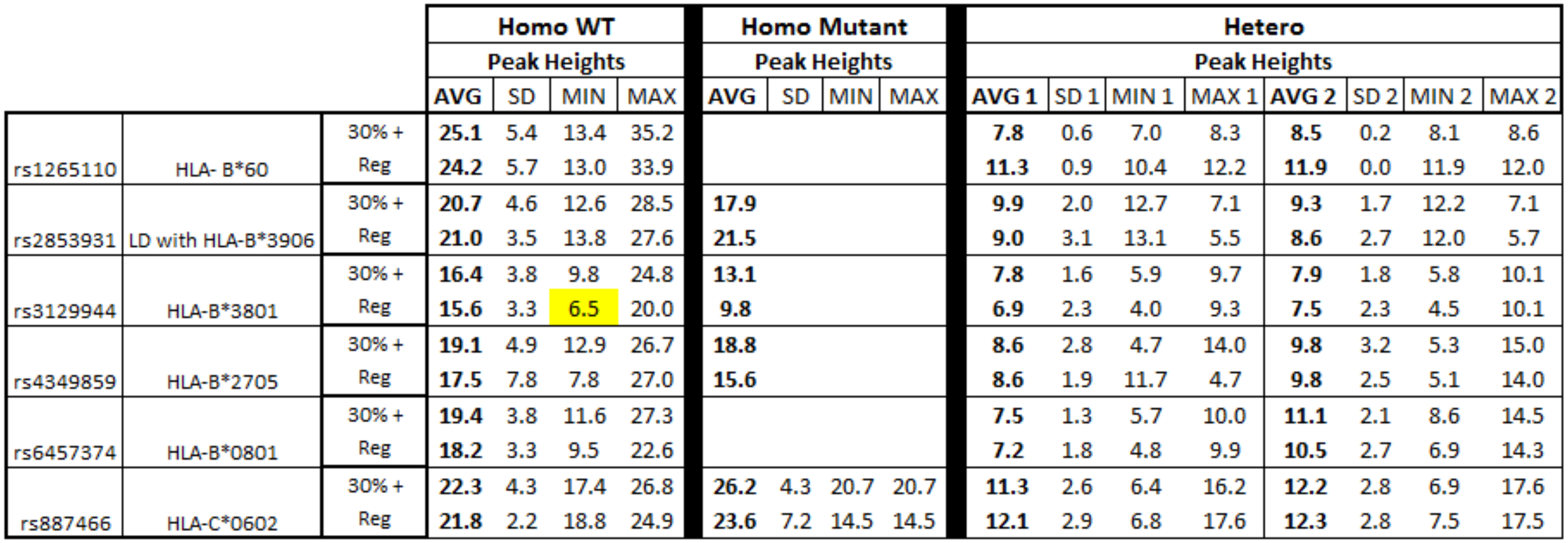
Table 2.29: Peak Height Comparison Table of New Extend Mix Primer Concentrations vs. Original Primer Mix Concentrations for LBP_1W.

Row with Reg (Regular) means original extend primer concentration, row with 18% + means extend primer concentration increased by 18%. Less than 7 and than 3.5 means number of samples that feel below this metric. Highlighted in yellow were minimums that fell below these metrics and Highlighted in blue are Averages that were below 10 for Homozygotes and 5 for Heterozygotes. Count meant number of samples per genotype.

			Homo WT						Homo Mutant						Hetero									
			Peak Heights						Peak Heights						Peak Heights									
			AVG	SD	MIN	MAX	Less than 7	Count	AVG	SD	MIN	MAX	Less than 7	Count	AVG 1	SD 1	MIN 1	MAX 1	AVG 2	SD 2	MIN 2	MAX 2	than 3.5	Count
rs10781500	CARD9	18% +	21.4	4.9	13.6	26.9	0.0	4.0	19.7	2.8	16.9	23.6	0.0	3.0	10.7	2.2	7.6	14.4	9.1	2.1	6.3	12.6	0.0	14.0
		Reg	13.6	1.1	12.3	15.0	0.0	4.0	16.6	6.4	11.0	25.6	0.0	3.0	8.8	3.6	2.7	15.5	7.2	3.0	2.4	12.2	2.0	16.0
rs10782001	FBXL19	18% +	8.5	1.5	6.6	11.7	1.0	8.0	10.0	1.7	8.2	11.7	0.0	2.0	5.0	1.1	3.5	7.0	4.9	1.1	3.3	6.4	1.0	10.0
		Reg	7.0	2.8	2.6	11.5	5.0	9.0	10.3	4.1	6.2	14.3	1.0	2.0	3.7	1.4	1.2	6.8	3.6	1.5	1.2	7.0	7.0	12.0
rs11209026	IL23R	18% +	19.7	4.9	12.7	30.0	0.0	18.0							5.8	0.1	5.6	5.9	6.7	0.3	6.4	7.0	0.0	2.0
		Reg	14.0	5.9	5.1	26.2	2.0	20.0							8.3	1.6	6.3	10.3	8.5	1.5	6.6	10.2	0.0	3.0
rs1265163 R	LD with B60	18% +	11.9	3.2	7.0	17.8	0.0	18.0							4.3	1.8	2.4	6.1	5.0	1.9	3.1	6.9	1.0	2.0
		Reg	9.4	3.8	3.0	17.3	7.0	21.0							3.6	1.7	1.9	5.3	4.0	2.3	1.7	6.3	1.0	2.0
rs2032890	ERAP1	18% +	11.0	2.7	7.4	15.6	0.0	9.0	16.5	3.7	10.6	20.7	0.0	4.0	5.4	0.7	4.4	6.7	5.6	0.7	4.6	6.9	0.0	7.0
		Reg	9.0	4.9	3.3	18.8	5.0	10.0	10.9	2.3	8.8	15.1	0.0	6.0	5.3	2.0	3.5	9.5	5.2	1.9	3.8	9.6	0.0	7.0
rs2066844	CARD15/NOD2	18% +	12.1	2.6	8.0	17.1	0.0	16.0							6.9	1.6	4.9	8.5	4.7	1.0	3.7	5.7	0.0	4.0
		Reg	9.7	3.8	3.3	17.6	3.0	18.0	7.0						4.7	1.1	3.5	6.5	3.3	0.8	2.5	4.6	3.0	4.0
rs2910686	ERAP2	18% +	18.8	4.0	15.4	26.9	0.0	6.0	20.1	3.7	15.7	24.8	0.0	3.0	7.9	2.1	5.0	12.5	8.1	2.0	5.7	12.3	0.0	11.0
		Reg	14.1	4.3	7.8	19.5	0.0	7.0	14.6	3.7	10.1	20.3	0.0	4.0	5.6	2.7	2.2	11.4	6.0	2.7	2.5	11.7	0.0	12.0
rs3132528	LD with B44	18% +	14.1	3.2	10.4	19.6	0.0	12.0	16.8	2.9	12.8	21.1	0.0	8.0										
		Reg	12.2	3.6	5.2	18.0	3.0	12.0	10.7	4.5	4.0	18.1	1.0	10.0	11.4				5.9					
rs33980500	TRAF3IP2	18% +	11.5	2.9	7.4	16.1	0.0	17.0							4.9	0.9	3.7	5.9	5.8	0.7	5.1	6.7	0.0	3.0
		Reg	8.5	2.8	3.1	14.7	4.0	19.0							4.5	1.7	1.7	6.1	4.7	1.7	1.9	6.6	1.0	4.0
rs582757	TNFAIP3	18% +	17.4	4.5	11.6	24.4	0.0	11.0	14.1						9.3	2.1	6.3	13.3	9.4	2.2	5.5	13.0	0.0	8.0
		Reg	13.1	4.3	4.3	21.2	1.0	14.0	27.0						5.8	1.9	2.7	8.7	5.7	1.8	2.8	8.2	2.0	8.0
rs6738490	ATG16L1	18% +	13.5	2.8	9.7	18.2	0.0	8.0	15.0	2.7	12.0	19.0	0.0	4.0	6.4	1.4	5.0	8.9	6.4	1.5	4.7	8.6	0.0	8.0
		Reg	11.6	3.5	6.6	18.6	1.0	11.0	11.3	4.8	4.6	17.4	1.0	4.0	4.0	1.0	2.0	5.0	3.9	1.1	1.9	5.4	2.0	8.0
rs6759298	intergenic 2p15	18% +	14.5	3.1	10.3	18.4	0.0	7.0	11.9	2.2	9.6	16.3	0.0	4.0	5.8	1.5	4.2	8.9	6.1	1.7	4.4	9.4	0.0	9.0
		Reg	8.6	3.1	3.4	13.9	1.0	7.0	10.3	3.5	8.1	16.3	0.0	4.0	4.6	1.7	1.8	7.9	4.9	1.9	1.7	8.0	2.0	12.0
rs6871626	IL12B	18% +	14.4	2.4	11.1	17.6	0.0	6.0	12.6	4.1	8.5	16.7	0.0	2.0	8.5	2.0	5.7	11.5	5.1	1.3	3.3	7.5	3.0	12.0
		Reg	12.3	5.0	4.2	18.8	1.0	6.0	7.1	1.2	5.4	8.1	1.0	3.0	6.5	2.3	2.6	11.8	3.9	1.3	1.9	7.4	1.0	14.0

Table 2.30: Peak Height Comparison Table of New Extend Mix Primer Concentrations vs. Original Primer Mix Concentrations for HLA well.

See description of Table 2.30



			Homo WT				Homo Mutant				Hetero							
			Peak Heights				Peak Heights				Peak Heights							
			AVG	SD	MIN	MAX	**AVG**	SD	MIN	MAX	**AVG 1**	SD 1	MIN 1	MAX 1	**AVG 2**	SD 2	MIN 2	MAX 2
rs1265110	HLA- B*60	30% +	**25.1**	5.4	13.4	35.2					**7.8**	0.6	7.0	8.3	**8.5**	0.2	8.1	8.6
		Reg	**24.2**	5.7	13.0	33.9					**11.3**	0.9	10.4	12.2	**11.9**	0.0	11.9	12.0
rs2853931	LD with HLA-B*3906	30% +	**20.7**	4.6	12.6	28.5	**17.9**				**9.9**	2.0	12.7	7.1	**9.3**	1.7	12.2	7.1
		Reg	**21.0**	3.5	13.8	27.6	**21.5**				**9.0**	3.1	13.1	5.5	**8.6**	2.7	12.0	5.7
rs3129944	HLA-B*3801	30% +	**16.4**	3.8	9.8	24.8	**13.1**				**7.8**	1.6	5.9	9.7	**7.9**	1.8	5.8	10.1
		Reg	**15.6**	3.3	6.5	20.0	**9.8**				**6.9**	2.3	4.0	9.3	**7.5**	2.3	4.5	10.1
rs4349859	HLA-B*2705	30% +	**19.1**	4.9	12.9	26.7	**18.8**				**8.6**	2.8	4.7	14.0	**9.8**	3.2	5.3	15.0
		Reg	**17.5**	7.8	7.8	27.0	**15.6**				**8.6**	1.9	11.7	4.7	**9.8**	2.5	5.1	14.0
rs6457374	HLA-B*0801	30% +	**19.4**	3.8	11.6	27.3					**7.5**	1.3	5.7	10.0	**11.1**	2.1	8.6	14.5
		Reg	**18.2**	3.3	9.5	22.6					**7.2**	1.8	4.8	9.9	**10.5**	2.7	6.9	14.3
rs887466	HLA-C*0602	30% +	**22.3**	4.3	17.4	26.8	**26.2**	4.3	20.7	20.7	**11.3**	2.6	6.4	16.2	**12.2**	2.8	6.9	17.6
		Reg	**21.8**	2.2	18.8	24.9	**23.6**	7.2	14.5	14.5	**12.1**	2.9	6.8	17.6	**12.3**	2.8	7.5	17.5

2.13 Algorithm construction

The genetic-based screening algorithm, statistical analysis and machine learning were programmed by Dr. Quan Li using the IBM cluster at the Centre for Health Informatics and Analytics (CHIA) at Memorial University Faculty of Medicine.

2.13.1 F-Score Analysis

Each maker included in the algorithm was interrogated for how discriminative it was at predicting disease individually. This was done using an F-Score Analysis for the discrimination between the affected and un-affected samples. Given training vectors X_k, k = 1, ..., n, if the number of affected and un-affected samples are n_+ and n_-, respectively, then the F-score of the i_{th} marker is defined as:

$$F(i) \equiv \frac{\left(\bar{\boldsymbol{x}}_i^{(+)} - \bar{\boldsymbol{x}}_i\right)^2 + \left(\bar{\boldsymbol{x}}_i^{(-)} - \bar{\boldsymbol{x}}_i\right)^2}{\frac{1}{n_+-1}\sum_{k=1}^{n_+}\left(x_{k,i}^{(+)} - \bar{\boldsymbol{x}}_i^{(+)}\right)^2 + \frac{1}{n_--1}\sum_{k=1}^{n_-}\left(x_{k,i}^{(-)} - \bar{\boldsymbol{x}}_i^{(-)}\right)^2},$$

Where $\bar{\boldsymbol{x}}_i, \bar{\boldsymbol{x}}_i^{(+)}, \bar{\boldsymbol{x}}_i^{(-)}$ are the average of the $\boldsymbol{i}_{th}$ marker of the whole, affected, and un-affected samples, respectively. The larger the F-score means the marker is more discriminative.

2.13.2 Machine Learning

Machine Learning methods are algorithms that can learn over time and make intelligent decisions that they were not explicitly programmed to do so. For this genetic-based screening algorithm a decision tree model was chosen. In machine learning, decision trees can easily show the process of interpretation and the structure of the decisions within the tree, making this model a good candidate for a strategic screening process.

The decision tree model was programmed in C/R using the J48/C5.0 decision tree model using a supervised learning technique. The first branch in the tree was *HLA-B*27* status. It was decided that this would be the first decision as *HLA-B*27* is the most discriminative and significant genetic variant to SpA. This was coded in a dominance inheritance form, with A/G or A/A as positive and G/G as negative. Once the model programming was complete, reduced-error based global pruning was applied to prevent overfitting the model.

First, a machine learning decision tree algorithm was performed for the discovery cohort. For the programming the cohort was subdivided into a training set (80%) and a testing set (20%) during 5-fold cross validation. Then another machine learning decision tree algorithm was separately programmed for the validation cohort. A similar subsetting of the cohort was also applied. Once both the discovery and validation cohort's training were completed they were compared the performance with F-measure and precision and it was determined that both of the training precisions and F-measure were equally comparable and high enough (both > 0.6) for the two cohorts to be combined.

In machine learning, the sample complexity is concerned with the performance of learning. In some cases, small sets of samples could not cover most of the domain of the data knowledge. Commonly, larger sample sizes are better than smaller one, as performance tends to increases with the size of the cohort. This was the main reason for the cohort combining in this project. For the cohort combining, first, the discovery cohort was used as the training set and the validation cohort was used as the testing set. For this considerable performance was obtained. Then we combined both the discovery cohort and validation cohort for the final tree model training. For the performance of combing cohort, we also randomly chose the 80% samples as training set and 20% as test set. For the combined cohort algorithm, the programming went one step further and markers voting weights were applied. The voting weights were based on the confidence its confidence value, the highest total vote is chosen as the final prediction.

When the final genetic-based screening algorithm was programmed the sensitivity (true positive (TP) rate), specificity (1.0 - false positive (FP) rate), precision (positive predictive value, TP/(TP+FP)), F-Measure and ROC were measured. Widely accepted performance measures can be derived from the following quantities;

(1) ***TP,*** the number of correctly classified as affected
(2) ***TN*** (True Negative), the number of correctly classified un-affected
(3) ***FP*,** the number of incorrectly classified as affected
(4) ***FN*** (False Negative), the number of incorrectly classified un-affected

$$Sensitivity = Q_{obs} = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

The F-measure (highest value of 1 and lowest value of 0) was defined as the harmonic mean of precision and recall:

$$2 \cdot \frac{\mathbf{precision} \cdot \mathbf{recall}}{\mathbf{precision} + \mathbf{recall}}.$$

Also by varying the threshold, a receiver operating characteristic (ROC) curve can be obtained by plotting true positive rate against false positive rate. The area under the ROC curve (AUC) can be used as a reliable, threshold-independent performance measure. The ROC curve of a random predictor had an AUC of 0.5 and of a perfect predictor has an AUC of 1.0.

Chapter 3 Results

3.1 Participant Info

All case participants in this study had been previously diagnosed by a Rheumatologist (Please see Methods Section). Participants that were included had 100% genotyping information for the 18 SNPs. In the discovery cohort 1075 samples were used (593 cases, 482 controls). In the validation cohort 943 samples were included (578 cases, 365 controls) (Table 3.1).

Table 3.1: Participant Ascertainment and Population Size.

The following table illustrates the number of participants from each ascertainment site that were included in the genetic-based algorithm. In addition, for the case participants the percent of males is included as well as the age of diagnosis.

	Control	**Case**		
Participant ascertainment	**N**	**N**	**Percent Male (%)**	**Mean Age at Diagnosis**
Newfoundland & Labrador	388	99	0.75	35
Ontario	95	495	0.72	31
Alberta	365	578	0.73	24
Total	848	1172	0.73	28

3.1.2 Case Participant Clinical Info

All case participants in this study had been diagnosed by a Rheumatologist (Please see Methods Section). Case samples that were ascertained from the University of Toronto have detailed clinical info. Specifically, these samples have *HLA-B27* status, C-reactive protein levels and info on if the patient has an extra-articular manifestations such as;

Iritis, Psoriasis, Cardiac Disease, Ulcerative Colitis, and or Crohn's Disease. Further details noted in Table 3.2.

Table 3.2: Clinical Information of Case Study Patients.

This info was ascertained by the Rheumatology Clinic at the University of Toronto. n/a means not applicable to certain category in table.

Clinical Information	**Count**	**Positives**	**Frequency in cohort (%)**	**AVG**
HLA-B27 Status	440	334	75.9	n/a
Iritis	462	142	30.7	n/a
Psoriasis	460	45	9.8	n/a
Cardiac disease	461	128	27.8	n/a
Ulcerative Colitis	462	27	5.8	n/a
Crohn's Disease	462	35	7.6	n/a
CRP Level	450	n/a	n/a	13.8

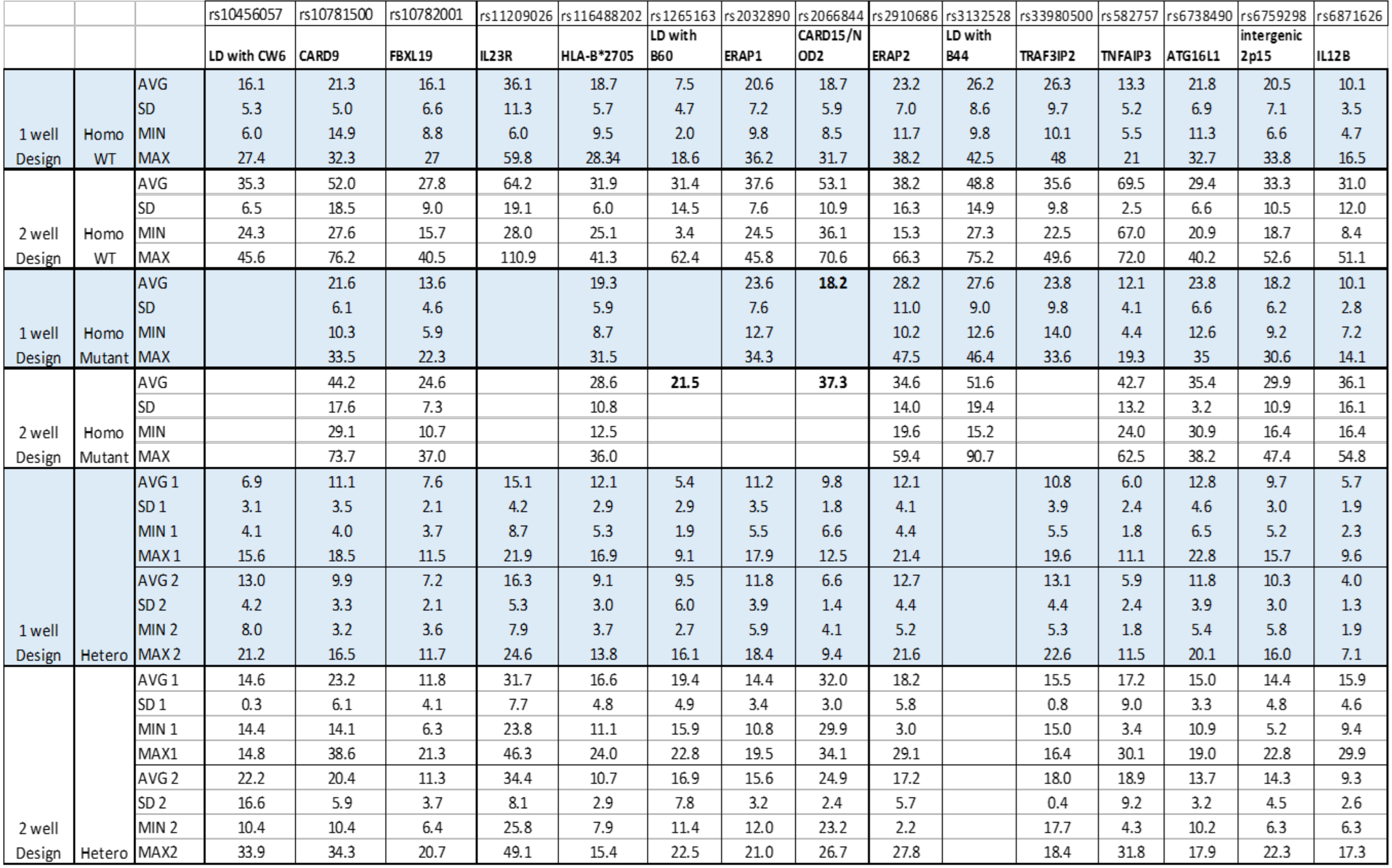
3.2 Assay Optimization

3.2.1 15-plex assay – 1 well vs. 2 well design

Across all assays for every rs number in the 15-plex reaction, the 2 well design had higher peak heights (Table 3.3). However, the 1 well design with the assay optimization met the quality metrics for peak heights outlined in the methods section. In addition, upon further inspection using Typer IV Analyzer, one rs3132528 (*HLA-B*44*) had abnormal clustering on the "Call Cluster Plot" of the "Post Processing Clusters" (Figure 3.1). This abnormal clustering occurred in homozygote samples with the genotype CC, these samples clustered near the heterozygote axis of the plot. This clustering made it difficult to determine the genotypes of certain samples. For an illustration of the abnormal cluster please refer to Figure 3.1 of the 2 well design which is the Typer IV Call Cluster Plot. Refer to Figure 3.2 for the Call Cluster Plot of the same assay and samples in the 1 well design. Another reason why the 1 well design was chosen over the 2 well design was that the 1 well design had a significant reduction of cost and labour. For these reasons the 1 well design was chosen over the 2 well design for the 15-plex reaction and the name of the well was then labelled LBP_1W.

Table 3.3: Peak Height Comparison of 1 well design vs. 2 well design.

Abbreviations; Average (AVG), Standard deviation (SD), Minimum (MIN) and Maximum (MAX). Cells of table Highlighted in Blue are metrics for the 1 well design assay.



			rs10456057	rs10781500	rs10782001	rs11209026	rs116488202	rs1265163	rs2032890	rs2066844	rs2910686	rs3132528	rs33980500	rs582757	rs6738490	rs6759298	rs6871626
			LD with CW6	CARD9	FBXL19	IL23R	HLA-B*2705	LD with B60	ERAP1	CARD15/NOD2	ERAP2	LD with B44	TRAF3IP2	TNFAIP3	ATG16L1	intergenic 2p15	IL12B
1 well Design	Homo WT	AVG	16.1	21.3	16.1	36.1	18.7	7.5	20.6	18.7	23.2	26.2	26.3	13.3	21.8	20.5	10.1
		SD	5.3	5.0	6.6	11.3	5.7	4.7	7.2	5.9	7.0	8.6	9.7	5.2	6.9	7.1	3.5
		MIN	6.0	14.9	8.8	6.0	9.5	2.0	9.8	8.5	11.7	9.8	10.1	5.5	11.3	6.6	4.7
		MAX	27.4	32.3	27	59.8	28.34	18.6	36.2	31.7	38.2	42.5	48	21	32.7	33.8	16.5
2 well Design	Homo WT	AVG	35.3	52.0	27.8	64.2	31.9	31.4	37.6	53.1	38.2	48.8	35.6	69.5	29.4	33.3	31.0
		SD	6.5	18.5	9.0	19.1	6.0	14.5	7.6	10.9	16.3	14.9	9.8	2.5	6.6	10.5	12.0
		MIN	24.3	27.6	15.7	28.0	25.1	3.4	24.5	36.1	15.3	27.3	22.5	67.0	20.9	18.7	8.4
		MAX	45.6	76.2	40.5	110.9	41.3	62.4	45.8	70.6	66.3	75.2	49.6	72.0	40.2	52.6	51.1
1 well Design	Homo Mutant	AVG		21.6	13.6		19.3		23.6	**18.2**	28.2	27.6	23.8	12.1	23.8	18.2	10.1
		SD		6.1	4.6		5.9		7.6		11.0	9.0	9.8	4.1	6.6	6.2	2.8
		MIN		10.3	5.9		8.7		12.7		10.2	12.6	14.0	4.4	12.6	9.2	7.2
		MAX		33.5	22.3		31.5		34.3		47.5	46.4	33.6	19.3	35	30.6	14.1
2 well Design	Homo Mutant	AVG		44.2	24.6		28.6	**21.5**		**37.3**	34.6	51.6		42.7	35.4	29.9	36.1
		SD		17.6	7.3		10.8				14.0	19.4		13.2	3.2	10.9	16.1
		MIN		29.1	10.7		12.5				19.6	15.2		24.0	30.9	16.4	16.4
		MAX		73.7	37.0		36.0				59.4	90.7		62.5	38.2	47.4	54.8
1 well Design	Hetero	AVG 1	6.9	11.1	7.6	15.1	12.1	5.4	11.2	9.8	12.1		10.8	6.0	12.8	9.7	5.7
		SD 1	3.1	3.5	2.1	4.2	2.9	2.9	3.5	1.8	4.1		3.9	2.4	4.6	3.0	1.9
		MIN 1	4.1	4.0	3.7	8.7	5.3	1.9	5.5	6.6	4.4		5.5	1.8	6.5	5.2	2.3
		MAX 1	15.6	18.5	11.5	21.9	16.9	9.1	17.9	12.5	21.4		19.6	11.1	22.8	15.7	9.6
		AVG 2	13.0	9.9	7.2	16.3	9.1	9.5	11.8	6.6	12.7		13.1	5.9	11.8	10.3	4.0
		SD 2	4.2	3.3	2.1	5.3	3.0	6.0	3.9	1.4	4.4		4.4	2.4	3.9	3.0	1.3
		MIN 2	8.0	3.2	3.6	7.9	3.7	2.7	5.9	4.1	5.2		5.3	1.8	5.4	5.8	1.9
		MAX 2	21.2	16.5	11.7	24.6	13.8	16.1	18.4	9.4	21.6		22.6	11.5	20.1	16.0	7.1
2 well Design	Hetero	AVG 1	14.6	23.2	11.8	31.7	16.6	19.4	14.4	32.0	18.2		15.5	17.2	15.0	14.4	15.9
		SD 1	0.3	6.1	4.1	7.7	4.8	4.9	3.4	3.0	5.8		0.8	9.0	3.3	4.8	4.6
		MIN 1	14.4	14.1	6.3	23.8	11.1	15.9	10.8	29.9	3.0		15.0	3.4	10.9	5.2	9.4
		MAX1	14.8	38.6	21.3	46.3	24.0	22.8	19.5	34.1	29.1		16.4	30.1	19.0	22.8	29.9
		AVG 2	22.2	20.4	11.3	34.4	10.7	16.9	15.6	24.9	17.2		18.0	18.9	13.7	14.3	9.3
		SD 2	16.6	5.9	3.7	8.1	2.9	7.8	3.2	2.4	5.7		0.4	9.2	3.2	4.5	2.6
		MIN 2	10.4	10.4	6.4	25.8	7.9	11.4	12.0	23.2	2.2		17.7	4.3	10.2	6.3	6.3
		MAX2	33.9	34.3	20.7	49.1	15.4	22.5	21.0	26.7	27.8		18.4	31.8	17.9	22.3	17.3

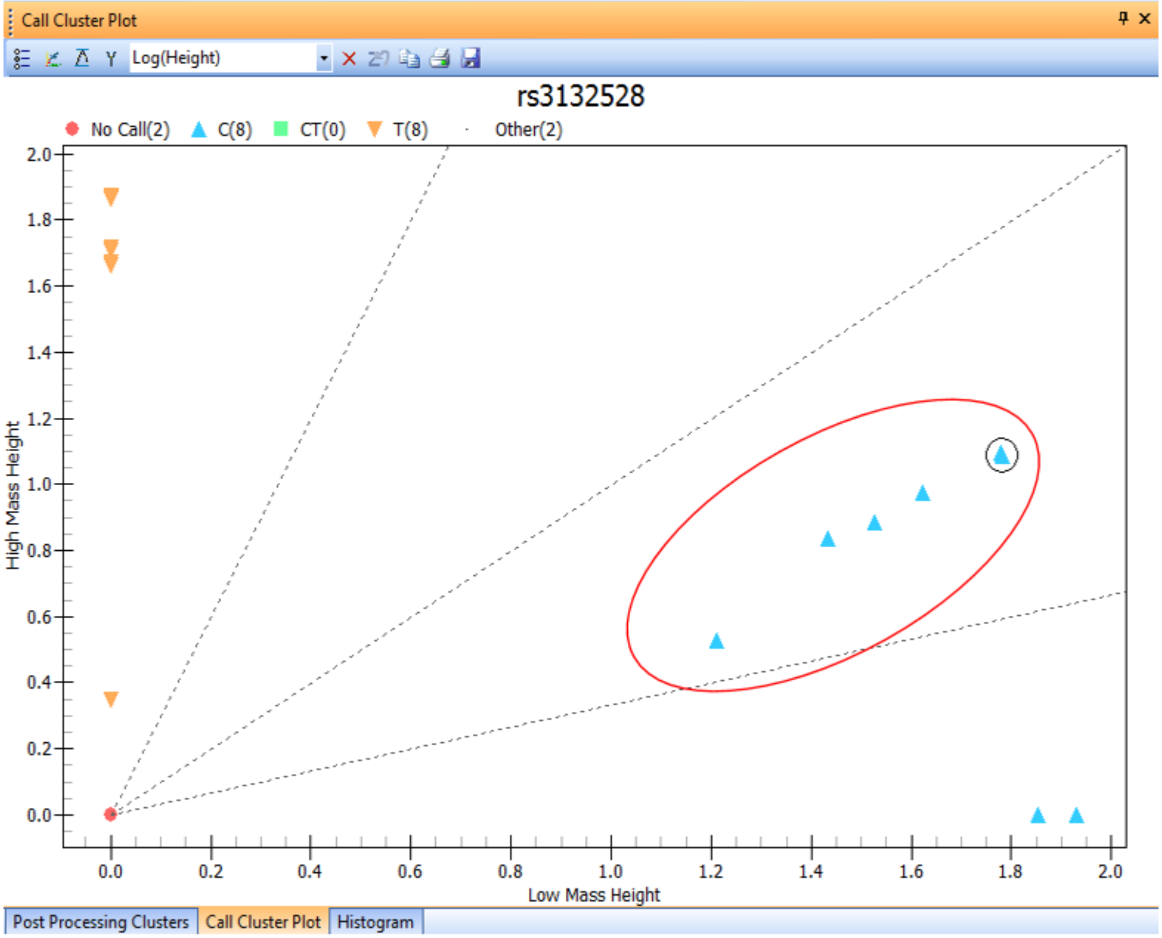


Figure 3.1: Typer IV Analyzer Call Cluster Plot for rs3132528 (*HLA-B*44*) for 2 well design.

Illustration of the abnormal clustering of the assay, as genotypes clustering near heterozygote zone.

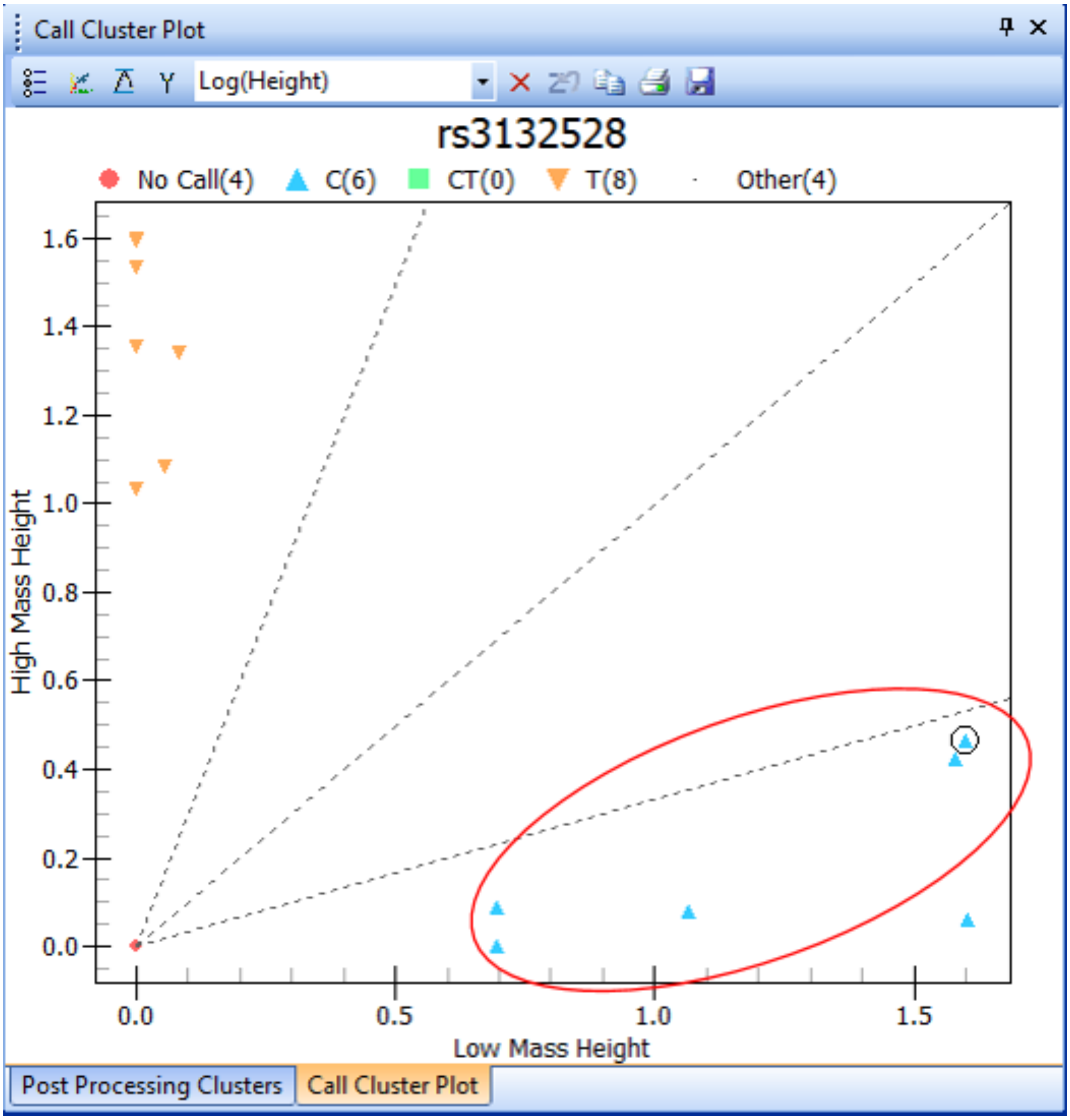


Figure 3.2: Typer IV Analyzer Call Cluster Plot for rs3132528 (*HLA-B*44*) for 1 well design. Illustration shows clustering of genotypes near the homozygous axis.

3.2.2 Gold vs. Pro iPLEX Regeant Kit

The Gold iPLEX regent kit is the more economical version of the regent set offered by Agena Biosciences. Using the Gold Regent kit offers a significant cost reduction than using the PRO regent kit. The difference between the kits is the extension enzyme (iPLEX Enzyme) used in the extend phase of the reaction. Agena Biosceicences states that these two product kits perform the same task.

Both kits had sufficient peak heights for the criteria established (Table 3.4 and 3.5). The iPLEX Pro Reagent kit had a higher average peak height in the majority of assays. A major problem was encountered with the Gold regent kit it was found that there were genotyping inconsistency. See Tables 3.6 and 3.7 below for disconcordant of Genotyping errors with Gold Kit. There were two errors which occurred in rs2853931 in the HLA well (Table 3.6) and rs2066844 in the 15-plex well (Table 3.7). All samples compared were run using both Gold and Pro and had been previously genotyped by the most recent AS GWAS study. These samples were then cross referenced with the ASGWAS study results, and the pro results. The PRO results were concordant with the AS GWAS microarray results.

Agena Biosciences was contacted and the company stated that in the past the Gold regent kit has yielded incorrect results and that the PRO kit is superior. Since there were major genotyping errors, the PRO Kit was chosen for genotyping this project.

Table 3.4: Peak Height Comparison of Gold iPLEX Reagent Kit vs. PRO iPLEX Reagent Kit for HLA well.

Only 4 assays were compared in the HLA well, as rs2853931 had 3 discrepant calls between the kits. Highlighted cells in yellow represent the kit that had the higher peak height average. The PRO iPLEX Reagent Kit had higher peak height averages in 12 out of the 14 comparisons.

			rs2853931	**rs3129944**	**rs4349859**	**rs6457374**	**rs887466**
			LD with HLA-B*3906	HLA-B*3801	HLA-B*2705	HLA-B*0801	HLA-C*0602
GOLD	**Homo WT**	**AVG**	**45.8**	**35.0**	**34.3**	**26.2**	**66.1**
		MIN	22.4	11.8	3.9	12.7	
PRO	**Homo WT**	**AVG**	**52.8**	**42.0**	**57.2**	**30.5**	**34.5**
		SD	28.2	19.9	24.8	18.5	
GOLD	**Homo Mutant**	**AVG**			**67.8**		**42.0**
		MIN					21.6
PRO	**Homo Mutant**	**AVG**			**45.6**		**52.3**
		SD					43.7
GOLD	**Hetero**	**AVG 1**	**19.9**	**20.1**	**19.1**	**10.9**	**22.2**
		MIN 1	7.5	9.1	2.4	6.5	9.2
		AVG 2	**19.0**	**14.7**	**15.0**	**10.5**	**20.5**
		MIN 2	8.0	7.7	2.2	5.1	7.5
PRO	**Hetero**	**AVG 1**	**26.4**	**24.2**	**26.0**	**12.4**	**27.3**
		MIN 1	19.1	16.4	15.1	10.7	18.1
		AVG 2	**25.5**	**25.4**	**27.2**	**21.5**	**30.2**
		MIN 2	18.7	14.7	8.1	17.5	18.8

Table 3.5: Peak Height Comparison of Gold iPLEX Reagent Kit and PRO iPLEX Reagent Kit for LBP_1W well.

Only 14 out of the 15 assays were compared in the 15-plex well, as there were discrepant genotyping calls between kits. Peak height averages for this Cells with red font are highlighted because in this assay one or more fell below the quality control metric set at the time of experimentation. Abbreviations same as Table 3.3 and Description for highlighting same as Table 3.4.

			rs10456057	rs10781500	rs10782001	rs11209026	rs116488202	rs1265163	rs2032890	rs2066844	rs2910686	rs3132528	rs33980500	rs582757	rs6738490	rs6759298	rs6871626
			LD with CW6	CARD9	FBXL19	IL23R	HLA-B*2705	LD with B60	ERAP1	CARD15/NOD2	ERAP2	LD with B44	TRAF3IP2	TNFAIP3	ATG16L1	intergenic 2p15	IL12B
GOLD	Homo WT	AVG	19.8	27.0	9.9	35.3	12.9	11.2	8.3	16.3	25.2	26.8	33.0	11.8	26.0	20.4	11.3
		MIN	13.0	19.9	9.6	22.0	8.1	6.6	5.6	11.2	18.4	18.5	22.2		19.8	12.5	5.2
PRO	Homo WT	AVG	17.6	24.3	14.8	37.4	20.6	7.5	23.9	18.4	27.2	27.8	28.4	11.5	18.4	21.8	9.6
		MIN	8.8	15.4	5.9	14.8	16.1	2.5	11.3	8.5	18.1	16.0	16.1		16.8	6.6	6.0
GOLD	Homo Mutant	AVG		18.4	5.1		21.6		25.1		32.8	34.5		12.3	26.8	21.0	7.7
		MIN		14.4	3.4		15.9		19.2		19.9	21.9		3.9	17.8	17.6	5.5
PRO	Homo Mutant	AVG		20.4	15.0		19.3		28.2		35.8	28.9		14.1	25.1	23.7	13.0
		MIN		15.9	14.5		8.7		22.1		23.7	12.5		8.4	17.2	16.8	10.6
GOLD	Hetero	AVG 1	6.6	14.0	4.3	23.7	13.2		7.7	14.5	17.9		12.1	8.6	14.0	10.5	6.3
		MIN 1	4.7	9.1	1.5		9.4		4.9		8.6			4.1	8.2	7.4	4.0
		AVG 2	14.5	9.6	3.3	16.9	10.4		5.1	9.7	7.2		8.7	6.3	13.9	10.4	5.0
		MIN 2	11.1	5.8	1.7		8.2		3.2		3.6			2.8	9.9	6.7	3.3
PRO	Hetero	AVG 1	6.0	11.5	8.2	13.5	12.8		11.1		12.1		6.7	8.0	14.1	9.4	5.9
		MIN 1	4.7	5.3	4.9		8.8		8.1		4.4			5.1	6.5	6.8	3.0
		AVG 2	12.3	10.1	8.1	13.7	10.3		11.1		13.0		8.0	8.2	13.1	10.5	4.5
		MIN 2	9.9	5.0	5.8		7.0		6.5		5.2			5.3	5.4	7.5	2.6

Table 3.6: Discrepant iPLEX Gold Reagent Kit Genotyping for HLA well rs2853931 (*HLA-B*39*).

Comparison between Gold Reagent Kit, Pro Reagent Kit and Microarray Genotyping calls. Microarray and iPLEX Pro Reagent Genotyping was concordant.

		Gold Reagent Kit			PRO Reagent Kit		
Sample ID	**Microarray Genotype**	**Call**	**Peak Height - T**	**Peak Height - C**	**Call**	**Peak Height - T**	**Peak Height - C**
AS_NF_017-0001	TT	TC	27.4	24.2	TT	50.4	1.5
AS_NF_036-0001	TT	TC	22.0	17.6	TT	77.8	2.2
AS_NF_040-0001	TT	TC	29.1	24.9	TT	37.6	0.8

Table 3.7: Discrepant iPLEX Gold Reagent Kit Genotyping for HLA well rs2066844 (*CARD15/NOD2*).

Comparison between Gold Reagent Kit, Pro Reagent Kit and Microarray Genotyping calls. Microarray and iPLEX Pro Reagent Genotyping was concordant.

		Gold Reagent Kit			PRO Reagent Kit		
Sample ID	**Microarray Genotype**	**Call**	**Peak Height - C**	**Peak Height - T**	**Call**	**Peak Height - C**	**Peak Height - T**
P-789	CC	CT	5.2	3.7	CC	13.5	0.0
P-792	CC	CT	4.6	3.3	CC	20.1	0.0

3.3 Quality Control Metric Analysis

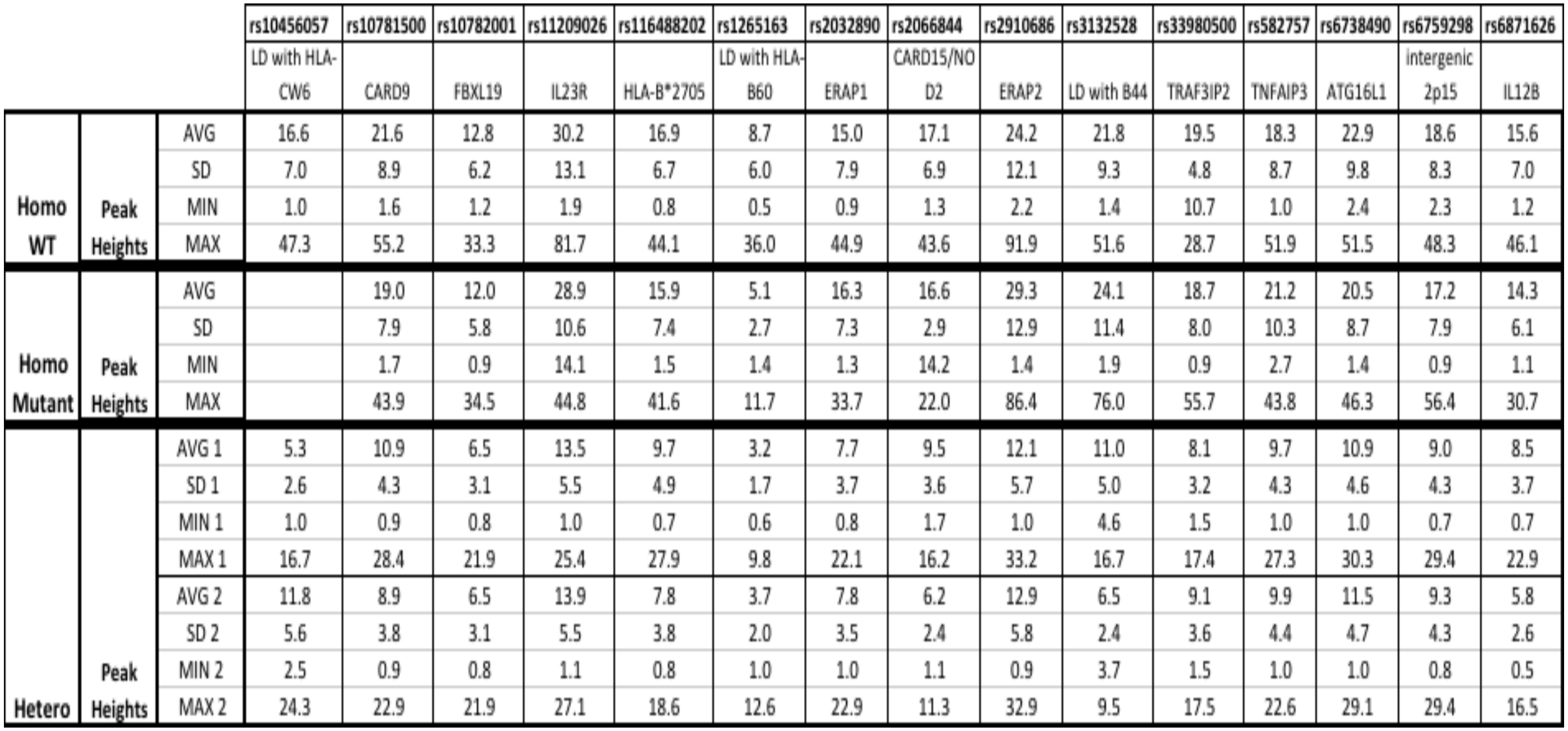
A series of quality control metrics were applied for genotype calling results as described in Methods section 2.11.

3.3.1 Peak Height Summaries

In the Tables below illustrate the peak heights for each rs number in both the LBP_1W well and the HLA well (Tables 3.8, 3.9, 3.10, and 3.11) for bot cases and controls combined by cohort. These tables have the peak height means, standard deviations, and range (minimum and maximum). These values were collected in order to monitor the quality of genotyping produced in each cohort.

Table 3.8: LBP_1W well Discovery Peak Heights.

Same abbreviations as above Table 3.3



			rs10456057	rs10781500	rs10782001	rs11209026	rs116488202	rs1265163	rs2032890	rs2066844	rs2910686	rs3132528	rs33980500	rs582757	rs6738490	rs6759298	rs6871626
			LD with HLA-CW6	CARD9	FBXL19	IL23R	HLA-B*2705	LD with HLA-B60	ERAP1	CARD15/NOD2	ERAP2	LD with B44	TRAF3IP2	TNFAIP3	ATG16L1	intergenic 2p15	IL12B
Homo WT	Peak Heights	AVG	16.6	21.6	12.8	30.2	16.9	8.7	15.0	17.1	24.2	21.8	19.5	18.3	22.9	18.6	15.6
		SD	7.0	8.9	6.2	13.1	6.7	6.0	7.9	6.9	12.1	9.3	4.8	8.7	9.8	8.3	7.0
		MIN	1.0	1.6	1.2	1.9	0.8	0.5	0.9	1.3	2.2	1.4	10.7	1.0	2.4	2.3	1.2
		MAX	47.3	55.2	33.3	81.7	44.1	36.0	44.9	43.6	91.9	51.6	28.7	51.9	51.5	48.3	46.1
Homo Mutant	Peak Heights	AVG		19.0	12.0	28.9	15.9	5.1	16.3	16.6	29.3	24.1	18.7	21.2	20.5	17.2	14.3
		SD		7.9	5.8	10.6	7.4	2.7	7.3	2.9	12.9	11.4	8.0	10.3	8.7	7.9	6.1
		MIN		1.7	0.9	14.1	1.5	1.4	1.3	14.2	1.4	1.9	0.9	2.7	1.4	0.9	1.1
		MAX		43.9	34.5	44.8	41.6	11.7	33.7	22.0	86.4	76.0	55.7	43.8	46.3	56.4	30.7
Hetero	Peak Heights	AVG 1	5.3	10.9	6.5	13.5	9.7	3.2	7.7	9.5	12.1	11.0	8.1	9.7	10.9	9.0	8.5
		SD 1	2.6	4.3	3.1	5.5	4.9	1.7	3.7	3.6	5.7	5.0	3.2	4.3	4.6	4.3	3.7
		MIN 1	1.0	0.9	0.8	1.0	0.7	0.6	0.8	1.7	1.0	4.6	1.5	1.0	1.0	0.7	0.7
		MAX 1	16.7	28.4	21.9	25.4	27.9	9.8	22.1	16.2	33.2	16.7	17.4	27.3	30.3	29.4	22.9
		AVG 2	11.8	8.9	6.5	13.9	7.8	3.7	7.8	6.2	12.9	6.5	9.1	9.9	11.5	9.3	5.8
		SD 2	5.6	3.8	3.1	5.5	3.8	2.0	3.5	2.4	5.8	2.4	3.6	4.4	4.7	4.3	2.6
		MIN 2	2.5	0.9	0.8	1.1	0.8	1.0	1.0	1.1	0.9	3.7	1.5	1.0	1.0	0.8	0.5
		MAX 2	24.3	22.9	21.9	27.1	18.6	12.6	22.9	11.3	32.9	9.5	17.5	22.6	29.1	29.4	16.5

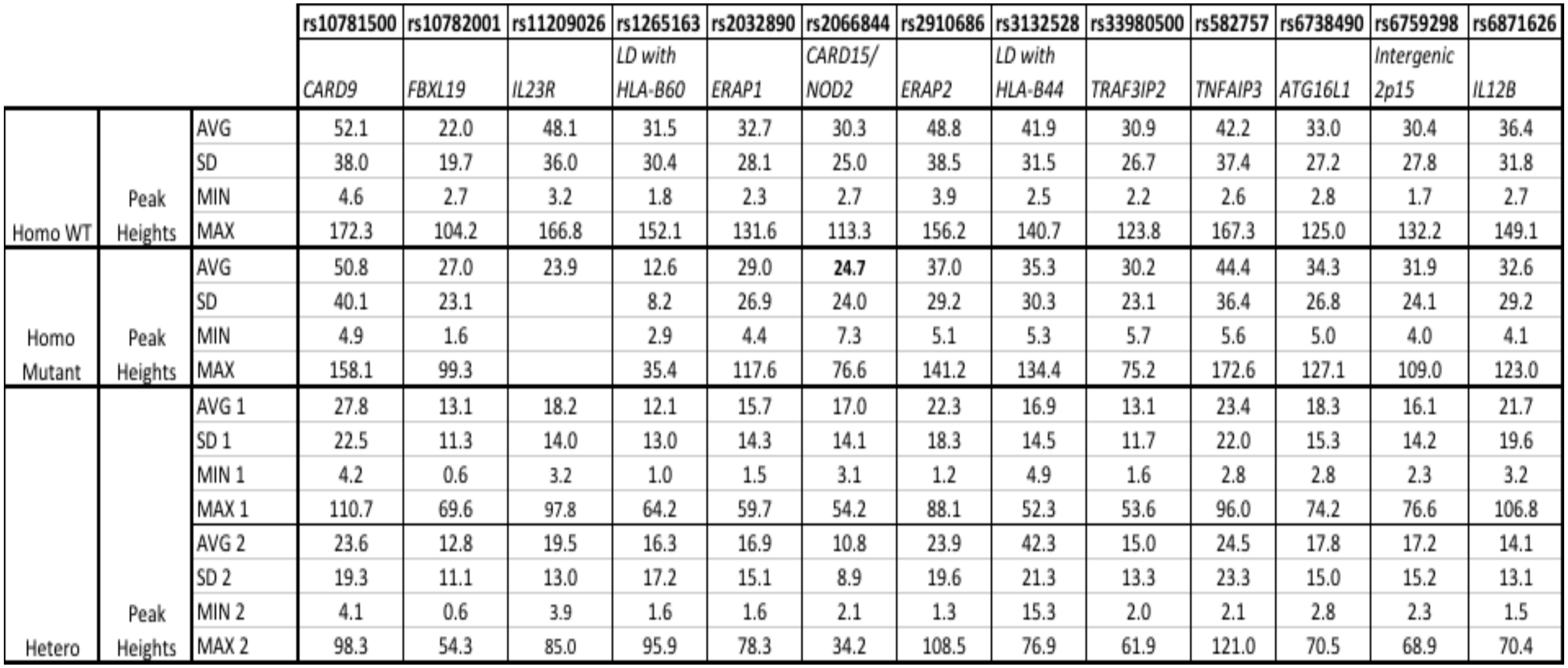
Table 3.9: HLA well Discovery Peak Heights.

Same abbreviations as above Table 3.3

			rs2853931	**rs3129944**	**rs4349859**	**rs6457374**	**rs887466**
			LD with HLA-B*3906	_HLA-B*3801_	_HLA-B*2705_	_HLA-B*0801_	_HLA-C*0602_
Homo WT	**Peak Heights**	AVG	33.7	30.0	29.0	28.8	38.7
		SD	14.6	13.2	14.5	16.9	16.4
		MIN	2.4	1.8	1.2	2.0	3.0
		MAX	92.7	74.7	72.1	64.6	81.9
Homo Mutant	**Peak Heights**	AVG	31.0	29.6	30.6	27.1	41.4
		SD	13.1	12.5	12.1	13.9	18.9
		MIN	4.3	2.5	12.2	2.0	3.2
		MAX	59.5	60.1	55.1	75.2	108.2
Hetero	**Peak Heights**	AVG 1	15.4	14.3	13.9	11.2	18.1
		SD 1	7.3	6.5	6.8	6.5	8.0
		MIN 1	1.2	1.1	0.9	1.0	1.3
		MAX 1	37.5	35.1	51.9	40.0	48.6
		AVG 2	14.4	14.6	13.9	14.8	20.4
		SD 2	6.8	6.6	6.6	8.2	9.1
		MIN 2	1.2	1.0	0.7	1.1	1.2
		MAX 2	35.66	42.3	36.6	45.1	54.0

Table 3.10: LBP_1W well Replication Peak Heights.

Same abbreviations as above Table 3.3. rs11209026 only had one genotype call homozygous mutant for this cohort.



			rs10781500	**rs10782001**	**rs11209026**	**rs1265163**	**rs2032890**	**rs2066844**	**rs2910686**	**rs3132528**	**rs33980500**	**rs582757**	**rs6738490**	**rs6759298**	**rs6871626**
			CARD9	*FBXL19*	*IL23R*	*LD with HLA-B60*	*ERAP1*	*CARD15/ NOD2*	*ERAP2*	*LD with HLA-B44*	*TRAF3IP2*	*TNFAIP3*	*ATG16L1*	*Intergenic 2p15*	*IL12B*
Homo WT	Peak Heights	AVG	52.1	22.0	48.1	31.5	32.7	30.3	48.8	41.9	30.9	42.2	33.0	30.4	36.4
		SD	38.0	19.7	36.0	30.4	28.1	25.0	38.5	31.5	26.7	37.4	27.2	27.8	31.8
		MIN	4.6	2.7	3.2	1.8	2.3	2.7	3.9	2.5	2.2	2.6	2.8	1.7	2.7
		MAX	172.3	104.2	166.8	152.1	131.6	113.3	156.2	140.7	123.8	167.3	125.0	132.2	149.1
Homo Mutant	Peak Heights	AVG	50.8	27.0	23.9	12.6	29.0	**24.7**	37.0	35.3	30.2	44.4	34.3	31.9	32.6
		SD	40.1	23.1		8.2	26.9	24.0	29.2	30.3	23.1	36.4	26.8	24.1	29.2
		MIN	4.9	1.6		2.9	4.4	7.3	5.1	5.3	5.7	5.6	5.0	4.0	4.1
		MAX	158.1	99.3		35.4	117.6	76.6	141.2	134.4	75.2	172.6	127.1	109.0	123.0
Hetero	Peak Heights	AVG 1	27.8	13.1	18.2	12.1	15.7	17.0	22.3	16.9	13.1	23.4	18.3	16.1	21.7
		SD 1	22.5	11.3	14.0	13.0	14.3	14.1	18.3	14.5	11.7	22.0	15.3	14.2	19.6
		MIN 1	4.2	0.6	3.2	1.0	1.5	3.1	1.2	4.9	1.6	2.8	2.8	2.3	3.2
		MAX 1	110.7	69.6	97.8	64.2	59.7	54.2	88.1	52.3	53.6	96.0	74.2	76.6	106.8
		AVG 2	23.6	12.8	19.5	16.3	16.9	10.8	23.9	42.3	15.0	24.5	17.8	17.2	14.1
		SD 2	19.3	11.1	13.0	17.2	15.1	8.9	19.6	21.3	13.3	23.3	15.0	15.2	13.1
		MIN 2	4.1	0.6	3.9	1.6	1.6	2.1	1.3	15.3	2.0	2.1	2.8	2.3	1.5
		MAX 2	98.3	54.3	85.0	95.9	78.3	34.2	108.5	76.9	61.9	121.0	70.5	68.9	70.4

Table 3.11: HLA well Replication Peak Heights.

Same abbreviations as above Table 3.3

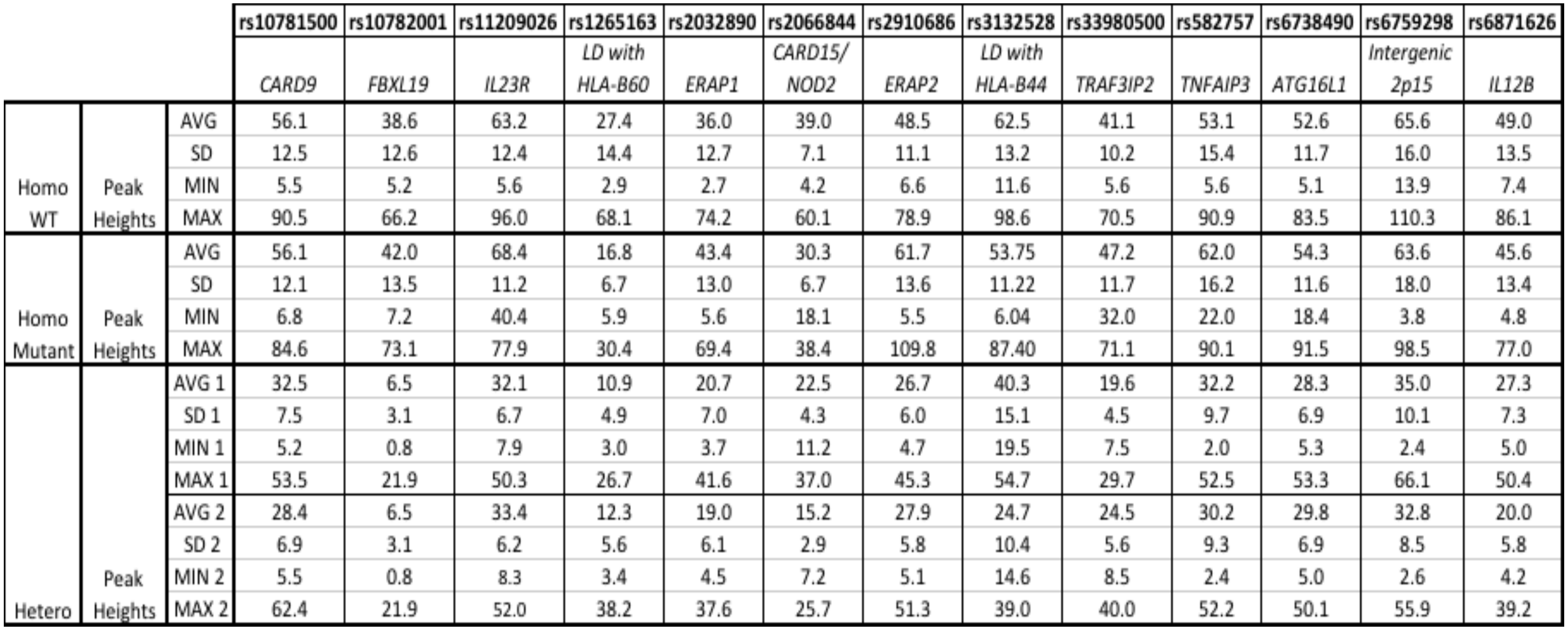
			rs2853931	rs3129944	rs4349859	rs6457374	rs887466
			*LD with HLA-B*3906*	*HLA-B*3801*	*HLA-B*2705*	*HLA-B*0801*	*HLA-C*0602*
Homo WT	**Peak Heights**	AVG	40.4	30.9	29.7	34.9	49.5
		SD	25.8	21.4	22.2	22.4	31.9
		MIN	5.8	5.6	5.1	5.3	6.6
		MAX	130.7	102.4	139.1	108.2	153.8
Homo Mutant	**Peak Heights**	AVG	30.6	29.4	45.9	31.6	43.6
		SD	19.5	22.7	33.4	20.1	26.7
		MIN	7.0	8.0	8.7	10.1	8.1
		MAX	89.7	100.0	122.0	89.4	153.9
Hetero	**Peak Heights**	AVG 1	18.9	14.3	21.2	12.6	23.8
		SD 1	13.0	10.6	16.8	8.3	16.5
		MIN 1	3.4	2.9	2.7	2.8	4.2
		MAX 1	66.9	56.6	90.6	55.6	83.9
		AVG 2	17.3	14.7	22.3	20.4	27.0
		SD 2	11.5	10.7	17.8	14.2	19.6
		MIN 2	3.3	3.0	3.6	2.7	4.9
		MAX 2	61.8	55.2	89.7	74.0	95.4

3.3.2 Signal to Noise Ratio Summaries

The following illustrate the Signal to Noise Ratio (SNR) for each rs number in both the LBP_1W well and the HLA well (Tables 3.12, 3.13, 3.14, and 3.15) for both cases and controls by cohort. These tables have the SNR means, standard deviations, and range (minimum and maximum). These values were collected in order to monitor the quality of genotyping produced in each cohort

Table 3.12: LBP_1W well Discovery Signal to Noise Ratio.

Same abbreviations as above Table 3.3



			rs10781500	rs10782001	rs11209026	rs1265163	rs2032890	rs2066844	rs2910686	rs3132528	rs33980500	rs582757	rs6738490	rs6759298	rs6871626
			CARD9	*FBXL19*	*IL23R*	*LD with HLA-B60*	*ERAP1*	*CARD15/ NOD2*	*ERAP2*	*LD with HLA-B44*	*TRAF3IP2*	*TNFAIP3*	*ATG16L1*	*Intergenic 2p15*	*IL12B*
Homo WT	Peak Heights	AVG	56.1	38.6	63.2	27.4	36.0	39.0	48.5	62.5	41.1	53.1	52.6	65.6	49.0
		SD	12.5	12.6	12.4	14.4	12.7	7.1	11.1	13.2	10.2	15.4	11.7	16.0	13.5
		MIN	5.5	5.2	5.6	2.9	2.7	4.2	6.6	11.6	5.6	5.6	5.1	13.9	7.4
		MAX	90.5	66.2	96.0	68.1	74.2	60.1	78.9	98.6	70.5	90.9	83.5	110.3	86.1
Homo Mutant	Peak Heights	AVG	56.1	42.0	68.4	16.8	43.4	30.3	61.7	53.75	47.2	62.0	54.3	63.6	45.6
		SD	12.1	13.5	11.2	6.7	13.0	6.7	13.6	11.22	11.7	16.2	11.6	18.0	13.4
		MIN	6.8	7.2	40.4	5.9	5.6	18.1	5.5	6.04	32.0	22.0	18.4	3.8	4.8
		MAX	84.6	73.1	77.9	30.4	69.4	38.4	109.8	87.40	71.1	90.1	91.5	98.5	77.0
Hetero	Peak Heights	AVG 1	32.5	6.5	32.1	10.9	20.7	22.5	26.7	40.3	19.6	32.2	28.3	35.0	27.3
		SD 1	7.5	3.1	6.7	4.9	7.0	4.3	6.0	15.1	4.5	9.7	6.9	10.1	7.3
		MIN 1	5.2	0.8	7.9	3.0	3.7	11.2	4.7	19.5	7.5	2.0	5.3	2.4	5.0
		MAX 1	53.5	21.9	50.3	26.7	41.6	37.0	45.3	54.7	29.7	52.5	53.3	66.1	50.4
		AVG 2	28.4	6.5	33.4	12.3	19.0	15.2	27.9	24.7	24.5	30.2	29.8	32.8	20.0
		SD 2	6.9	3.1	6.2	5.6	6.1	2.9	5.8	10.4	5.6	9.3	6.9	8.5	5.8
		MIN 2	5.5	0.8	8.3	3.4	4.5	7.2	5.1	14.6	8.5	2.4	5.0	2.6	4.2
		MAX 2	62.4	21.9	52.0	38.2	37.6	25.7	51.3	39.0	40.0	52.2	50.1	55.9	39.2

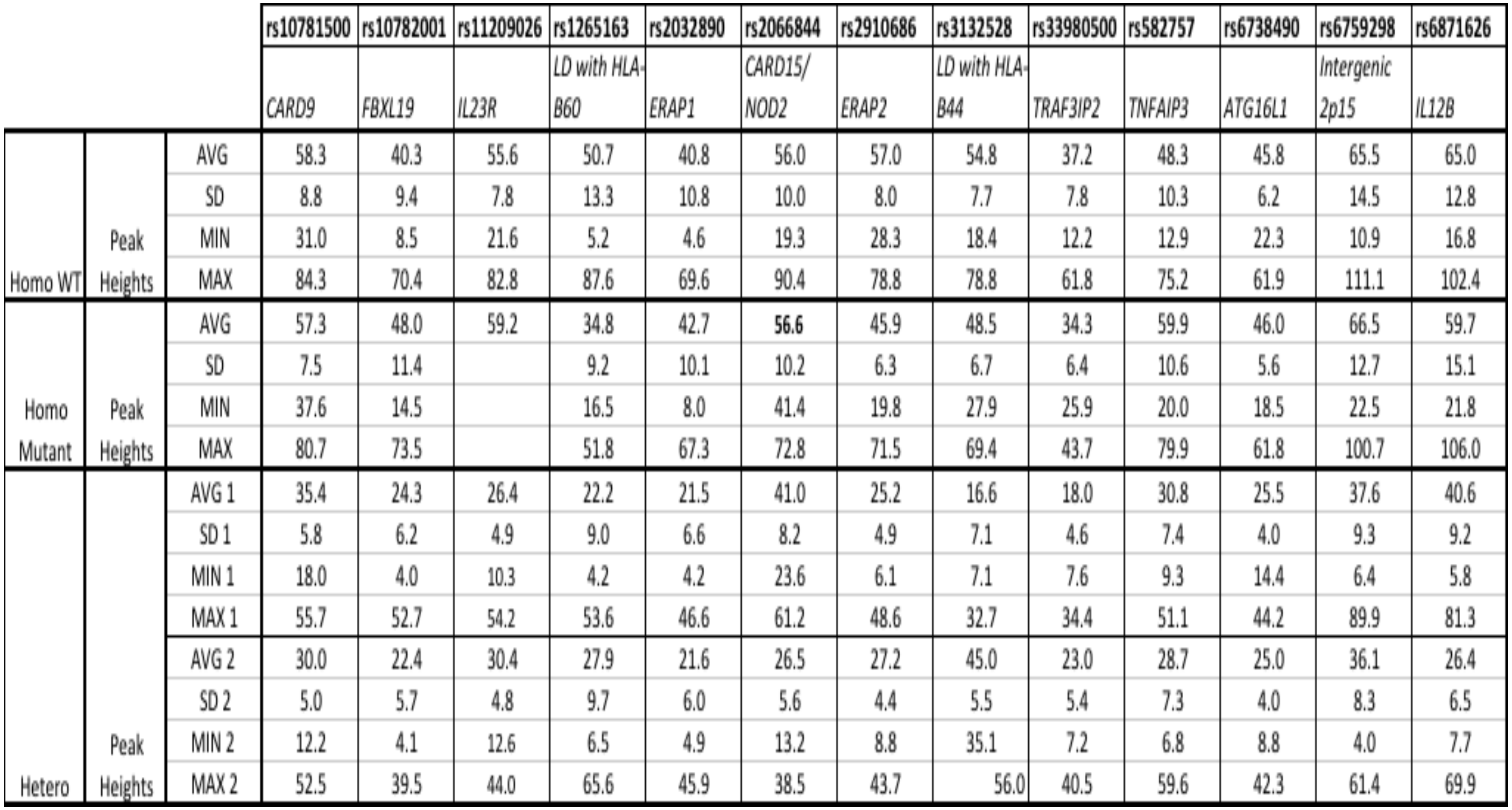
Table 3.13: HLA Well Discovery cohort Signal to Noise Ratio.

Same abbreviations as above Table 3.3

			rs2853931	**rs3129944**	**rs4349859**	**rs6457374**	**rs887466**
			*LD with HLA-B*3906*	*HLA-B*3801*	*HLA-B*2705*	*HLA-B*0801*	*HLA-C*0602*
Homo WT	**Peak Heights**	AVG	94.5	72.8	64.4	71.3	95.6
		SD	21.4	18.6	17.0	19.5	21.0
		MIN	18.1	6.0	4.4	6.1	15.8
		MAX	142.1	120.1	110.1	123.1	149.9
Homo Mutant	**Peak Heights**	AVG	89.1	77.2	75.2	72.4	107.3
		SD	18.4	19.7	9.7	22.6	25.9
		MIN	26.2	4.9	55.8	17.0	26.3
		MAX	132.7	113.8	90.8	119.0	177.1
Hetero	**Peak Heights**	AVG 1	55.5	44.61	33.5	38.1	63.6
		SD 1	15.0	11.92	10.1	15.4	17.2
		MIN 1	8.6	2.79	3.6	8.0	7.0
		MAX 1	90.1	74.18	72.8	91.6	107.1
		AVG 2	50.6	41.0	38.7	49.0	64.6
		SD 2	13.4	11.0	11.1	17.5	16.1
		MIN 2	8.5	3.1	3.4	7.4	6.8
		MAX 2	79.2	70.3	66.9	91.4	109.5

Table 3.14: LBP_1W well Replication Cohort Signal to Noise Ratio.

Same abbreviations as above Table 3.3



			rs10781500	rs10782001	rs11209026	rs1265163	rs2032890	rs2066844	rs2910686	rs3132528	rs33980500	rs582757	rs6738490	rs6759298	rs6871626
			CARD9	*FBXL19*	*IL23R*	*LD with HLA-B60*	*ERAP1*	*CARD15/ NOD2*	*ERAP2*	*LD with HLA-B44*	*TRAF3IP2*	*TNFAIP3*	*ATG16L1*	*Intergenic 2p15*	*IL12B*
Homo WT	Peak Heights	AVG	58.3	40.3	55.6	50.7	40.8	56.0	57.0	54.8	37.2	48.3	45.8	65.5	65.0
		SD	8.8	9.4	7.8	13.3	10.8	10.0	8.0	7.7	7.8	10.3	6.2	14.5	12.8
		MIN	31.0	8.5	21.6	5.2	4.6	19.3	28.3	18.4	12.2	12.9	22.3	10.9	16.8
		MAX	84.3	70.4	82.8	87.6	69.6	90.4	78.8	78.8	61.8	75.2	61.9	111.1	102.4
Homo Mutant	Peak Heights	AVG	57.3	48.0	59.2	34.8	42.7	**56.6**	45.9	48.5	34.3	59.9	46.0	66.5	59.7
		SD	7.5	11.4		9.2	10.1	10.2	6.3	6.7	6.4	10.6	5.6	12.7	15.1
		MIN	37.6	14.5		16.5	8.0	41.4	19.8	27.9	25.9	20.0	18.5	22.5	21.8
		MAX	80.7	73.5		51.8	67.3	72.8	71.5	69.4	43.7	79.9	61.8	100.7	106.0
Hetero	Peak Heights	AVG 1	35.4	24.3	26.4	22.2	21.5	41.0	25.2	16.6	18.0	30.8	25.5	37.6	40.6
		SD 1	5.8	6.2	4.9	9.0	6.6	8.2	4.9	7.1	4.6	7.4	4.0	9.3	9.2
		MIN 1	18.0	4.0	10.3	4.2	4.2	23.6	6.1	7.1	7.6	9.3	14.4	6.4	5.8
		MAX 1	55.7	52.7	54.2	53.6	46.6	61.2	48.6	32.7	34.4	51.1	44.2	89.9	81.3
		AVG 2	30.0	22.4	30.4	27.9	21.6	26.5	27.2	45.0	23.0	28.7	25.0	36.1	26.4
		SD 2	5.0	5.7	4.8	9.7	6.0	5.6	4.4	5.5	5.4	7.3	4.0	8.3	6.5
		MIN 2	12.2	4.1	12.6	6.5	4.9	13.2	8.8	35.1	7.2	6.8	8.8	4.0	7.7
		MAX 2	52.5	39.5	44.0	65.6	45.9	38.5	43.7	56.0	40.5	59.6	42.3	61.4	69.9

Table 3.15: LBP_1W well Replication Signal to Noise Ratio.

Same abbreviations as above Table 3.3.

			rs2853931	**rs3129944**	**rs4349859**	**rs6457374**	**rs887466**
			LD with HLA-B*3906	HLA-B*3801	HLA-B*2705	HLA-B*0801	HLA-C*0602
Homo WT	**Peak Heights**	AVG	84.0	58.3	59.0	66.9	82.1
		SD	16.8	17.3	10.3	12.8	15.5
		MIN	19.7	8.2	25.2	19.0	32.8
		MAX	127.2	104.4	93.0	101.4	129.8
Homo Mutant	**Peak Heights**	AVG	76.0	66.6	67.2	70.5	84.6
		SD	15.1	13.5	15.9	10.6	14.3
		MIN	21.0	34.9	32.5	35.9	28.7
		MAX	114.5	98.7	92.9	94.2	126.5
Hetero	**Peak Heights**	AVG 1	50.7	35.5	33.4	33.7	47.0
		SD 1	12.8	10.4	8.7	10.5	10.3
		MIN 1	9.9	4.3	9.3	12.1	14.9
		MAX 1	81.7	64.3	56.9	78.5	84.5
		AVG 2	45.8	32.8	40.1	52.0	52.0
		SD 2	11.1	9.9	9.9	12.8	10.4
		MIN 2	9.7	4.1	10.8	13.9	22.7
		MAX 2	71.7	60.1	66.9	84.0	96.1

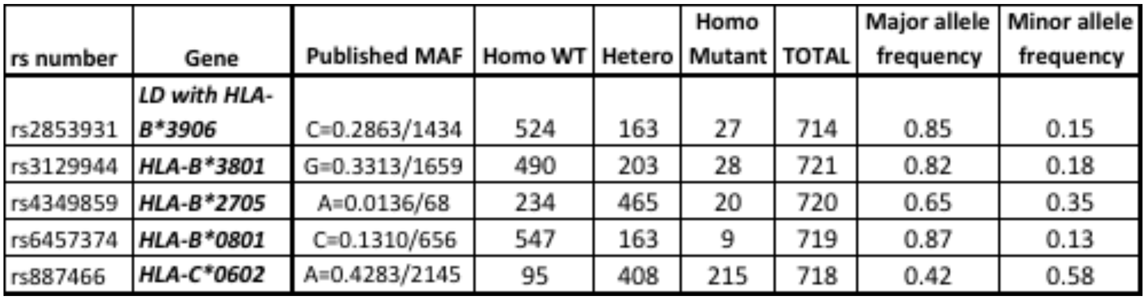
3.4 Genotype Frequency

Allele frequencies were determined for the discovery cohort, the following Tables (Table 3.16 And 3.17) show the allele frequencies for each variant in the discovery cohort.

Table 3.16 A and B: Case and Control Genotyping Frequencies HLA well for Discovery Cohort.

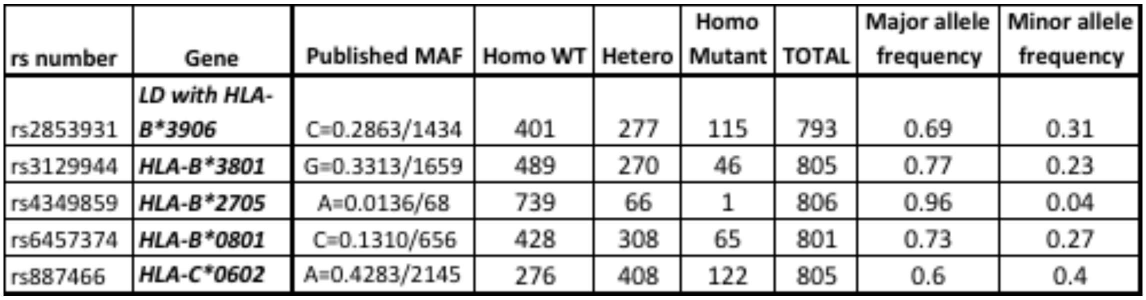
A) Case genotypes for HLA well. **B)** Control Genotypes for the HLA well. Published MAF was accessed through the 1000 genomes project. Abbreviations: Homo WT: Homozygous wild-type genotype, Hetero: Heterozygous genotype, Homo Mutant: Homozygous mutant genotype.

A



rs number	Gene	Published MAF	Homo WT	Hetero	Homo Mutant	TOTAL	Major allele frequency	Minor allele frequency
rs2853931	***LD with HLA-B*3906***	C=0.2863/1434	524	163	27	714	0.85	0.15
rs3129944	***HLA-B*3801***	G=0.3313/1659	490	203	28	721	0.82	0.18
rs4349859	***HLA-B*2705***	A=0.0136/68	234	465	20	720	0.65	0.35
rs6457374	***HLA-B*0801***	C=0.1310/656	547	163	9	719	0.87	0.13
rs887466	***HLA-C*0602***	A=0.4283/2145	95	408	215	718	0.42	0.58

B

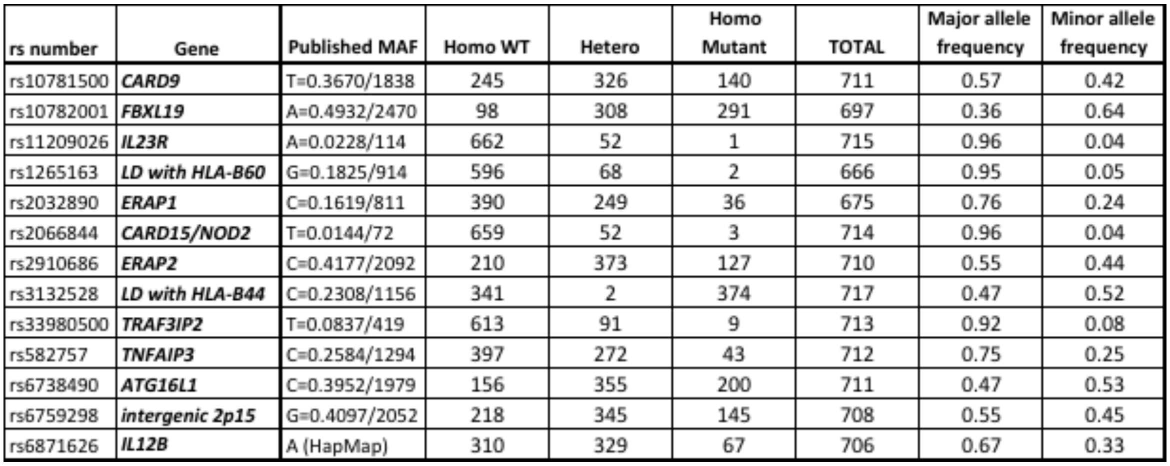


rs number	Gene	Published MAF	Homo WT	Hetero	Homo Mutant	TOTAL	Major allele frequency	Minor allele frequency
rs2853931	***LD with HLA-B*3906***	C=0.2863/1434	401	277	115	793	0.69	0.31
rs3129944	***HLA-B*3801***	G=0.3313/1659	489	270	46	805	0.77	0.23
rs4349859	***HLA-B*2705***	A=0.0136/68	739	66	1	806	0.96	0.04
rs6457374	***HLA-B*0801***	C=0.1310/656	428	308	65	801	0.73	0.27
rs887466	***HLA-C*0602***	A=0.4283/2145	276	408	122	805	0.6	0.4

Table 3.17 A and B: Case and Control Genotyping Frequencies LBP_1W well for Discovery cohort.

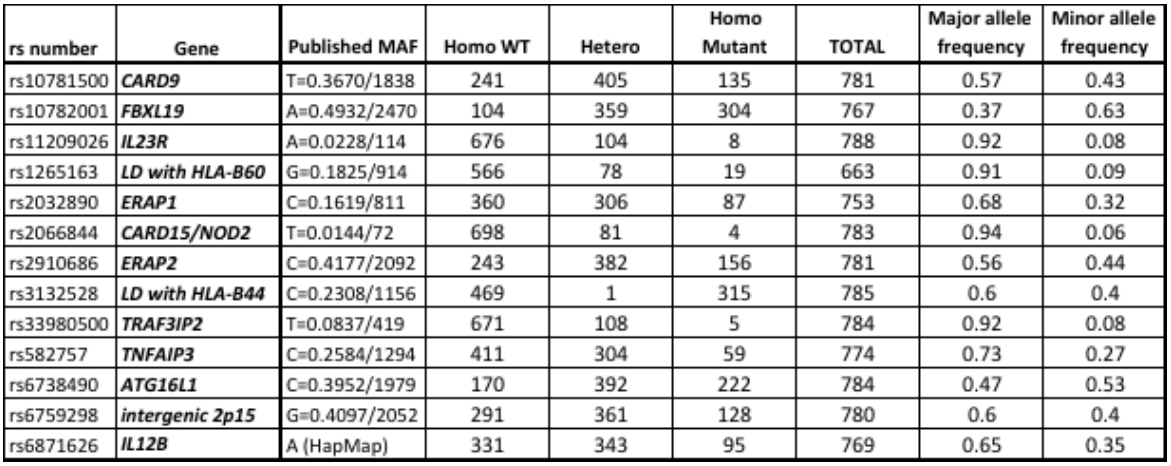
A) Case genotypes for LBP_1W well. **B)** Control Genotypes for the LBP_1W well. The abbreviations are the same as the Table 3.16

A



rs number	Gene	Published MAF	Homo WT	Hetero	Homo Mutant	TOTAL	Major allele frequency	Minor allele frequency
rs10781500	***CARD9***	T=0.3670/1838	245	326	140	711	0.57	0.42
rs10782001	***FBXL19***	A=0.4932/2470	98	308	291	697	0.36	0.64
rs11209026	***IL23R***	A=0.0228/114	662	52	1	715	0.96	0.04
rs1265163	***LD with HLA-B60***	G=0.1825/914	596	68	2	666	0.95	0.05
rs2032890	***ERAP1***	C=0.1619/811	390	249	36	675	0.76	0.24
rs2066844	***CARD15/NOD2***	T=0.0144/72	659	52	3	714	0.96	0.04
rs2910686	***ERAP2***	C=0.4177/2092	210	373	127	710	0.55	0.44
rs3132528	***LD with HLA-B44***	C=0.2308/1156	341	2	374	717	0.47	0.52
rs33980500	***TRAF3IP2***	T=0.0837/419	613	91	9	713	0.92	0.08
rs582757	***TNFAIP3***	C=0.2584/1294	397	272	43	712	0.75	0.25
rs6738490	***ATG16L1***	C=0.3952/1979	156	355	200	711	0.47	0.53
rs6759298	***intergenic 2p15***	G=0.4097/2052	218	345	145	708	0.55	0.45
rs6871626	***IL12B***	A (HapMap)	310	329	67	706	0.67	0.33

B



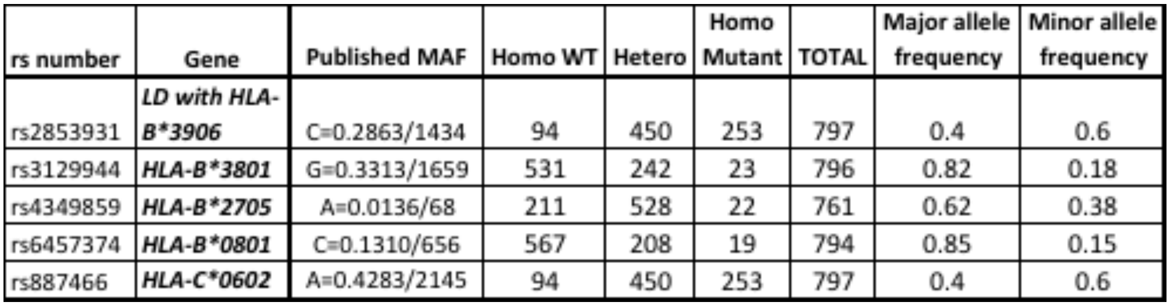
rs number	Gene	Published MAF	Homo WT	Hetero	Homo Mutant	TOTAL	Major allele frequency	Minor allele frequency
rs10781500	***CARD9***	T=0.3670/1838	241	405	135	781	0.57	0.43
rs10782001	***FBXL19***	A=0.4932/2470	104	359	304	767	0.37	0.63
rs11209026	***IL23R***	A=0.0228/114	676	104	8	788	0.92	0.08
rs1265163	***LD with HLA-B60***	G=0.1825/914	566	78	19	663	0.91	0.09
rs2032890	***ERAP1***	C=0.1619/811	360	306	87	753	0.68	0.32
rs2066844	***CARD15/NOD2***	T=0.0144/72	698	81	4	783	0.94	0.06
rs2910686	***ERAP2***	C=0.4177/2092	243	382	156	781	0.56	0.44
rs3132528	***LD with HLA-B44***	C=0.2308/1156	469	1	315	785	0.6	0.4
rs33980500	***TRAF3IP2***	T=0.0837/419	671	108	5	784	0.92	0.08
rs582757	***TNFAIP3***	C=0.2584/1294	411	304	59	774	0.73	0.27
rs6738490	***ATG16L1***	C=0.3952/1979	170	392	222	784	0.47	0.53
rs6759298	***intergenic 2p15***	G=0.4097/2052	291	361	128	780	0.6	0.4
rs6871626	***IL12B***	A (HapMap)	331	343	95	769	0.65	0.35

Allele frequencies were determined for the validation cohort, the following Tables (Table 3.18 And 3.19) show the allele frequencies for each variant in the discovery cohort. Hardy Weinberg was calculated for each variant.

Table 3.18 A and B: Case and Control Genotyping Frequencies for HLA well in Replication Cohort.

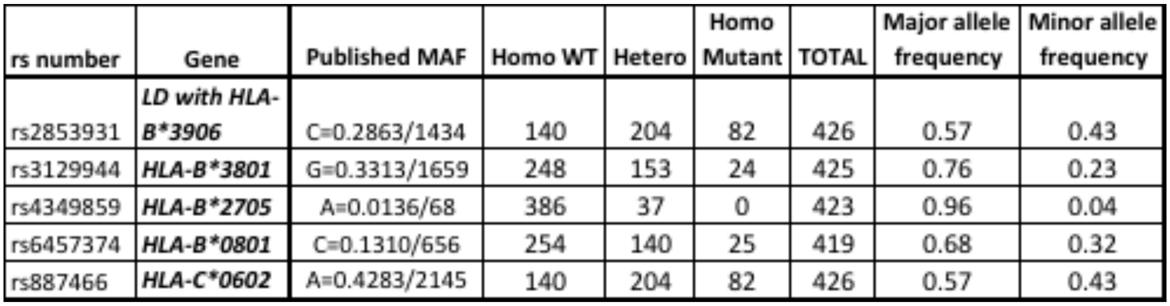
A) Case genotypes for HLA well. **B)** Control Genotypes for the HLA well. The abbreviations are the same as the Table 3.16

A



rs number	Gene	Published MAF	Homo WT	Hetero	Homo Mutant	TOTAL	Major allele frequency	Minor allele frequency
rs2853931	***LD with HLA-B*3906***	C=0.2863/1434	94	450	253	797	0.4	0.6
rs3129944	***HLA-B*3801***	G=0.3313/1659	531	242	23	796	0.82	0.18
rs4349859	***HLA-B*2705***	A=0.0136/68	211	528	22	761	0.62	0.38
rs6457374	***HLA-B*0801***	C=0.1310/656	567	208	19	794	0.85	0.15
rs887466	***HLA-C*0602***	A=0.4283/2145	94	450	253	797	0.4	0.6

B

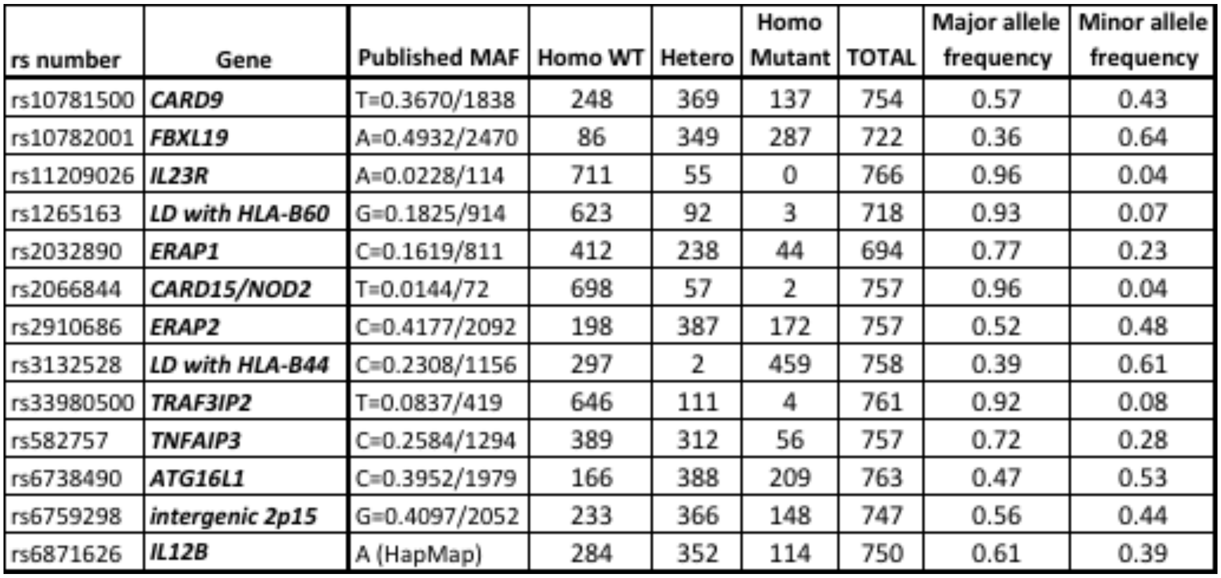


rs number	Gene	Published MAF	Homo WT	Hetero	Homo Mutant	TOTAL	Major allele frequency	Minor allele frequency
rs2853931	***LD with HLA-B*3906***	C=0.2863/1434	140	204	82	426	0.57	0.43
rs3129944	***HLA-B*3801***	G=0.3313/1659	248	153	24	425	0.76	0.23
rs4349859	***HLA-B*2705***	A=0.0136/68	386	37	0	423	0.96	0.04
rs6457374	***HLA-B*0801***	C=0.1310/656	254	140	25	419	0.68	0.32
rs887466	***HLA-C*0602***	A=0.4283/2145	140	204	82	426	0.57	0.43

Table 3.19 A and B: Case and Control Genotyping Frequencies for LBP_1W well for Replication Cohort.

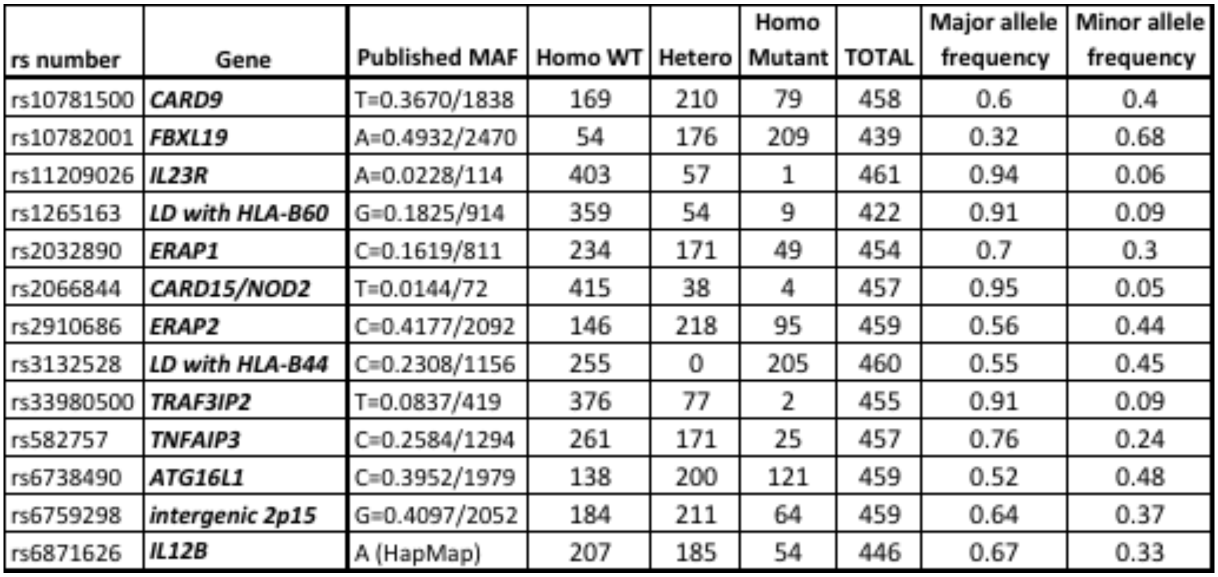
A) Case genotypes for LBP_1W well. **B)** Control Genotypes for the LBP 1W well. The abbreviations are the same as the Table 3.16

A



rs number	Gene	Published MAF	Homo WT	Hetero	Homo Mutant	TOTAL	Major allele frequency	Minor allele frequency
rs10781500	***CARD9***	T=0.3670/1838	248	369	137	754	0.57	0.43
rs10782001	***FBXL19***	A=0.4932/2470	86	349	287	722	0.36	0.64
rs11209026	***IL23R***	A=0.0228/114	711	55	0	766	0.96	0.04
rs1265163	***LD with HLA-B60***	G=0.1825/914	623	92	3	718	0.93	0.07
rs2032890	***ERAP1***	C=0.1619/811	412	238	44	694	0.77	0.23
rs2066844	***CARD15/NOD2***	T=0.0144/72	698	57	2	757	0.96	0.04
rs2910686	***ERAP2***	C=0.4177/2092	198	387	172	757	0.52	0.48
rs3132528	***LD with HLA-B44***	C=0.2308/1156	297	2	459	758	0.39	0.61
rs33980500	***TRAF3IP2***	T=0.0837/419	646	111	4	761	0.92	0.08
rs582757	***TNFAIP3***	C=0.2584/1294	389	312	56	757	0.72	0.28
rs6738490	***ATG16L1***	C=0.3952/1979	166	388	209	763	0.47	0.53
rs6759298	***intergenic 2p15***	G=0.4097/2052	233	366	148	747	0.56	0.44
rs6871626	***IL12B***	A (HapMap)	284	352	114	750	0.61	0.39

B



rs number	Gene	Published MAF	Homo WT	Hetero	Homo Mutant	TOTAL	Major allele frequency	Minor allele frequency
rs10781500	***CARD9***	T=0.3670/1838	169	210	79	458	0.6	0.4
rs10782001	***FBXL19***	A=0.4932/2470	54	176	209	439	0.32	0.68
rs11209026	***IL23R***	A=0.0228/114	403	57	1	461	0.94	0.06
rs1265163	***LD with HLA-B60***	G=0.1825/914	359	54	9	422	0.91	0.09
rs2032890	***ERAP1***	C=0.1619/811	234	171	49	454	0.7	0.3
rs2066844	***CARD15/NOD2***	T=0.0144/72	415	38	4	457	0.95	0.05
rs2910686	***ERAP2***	C=0.4177/2092	146	218	95	459	0.56	0.44
rs3132528	***LD with HLA-B44***	C=0.2308/1156	255	0	205	460	0.55	0.45
rs33980500	***TRAF3IP2***	T=0.0837/419	376	77	2	455	0.91	0.09
rs582757	***TNFAIP3***	C=0.2584/1294	261	171	25	457	0.76	0.24
rs6738490	***ATG16L1***	C=0.3952/1979	138	200	121	459	0.52	0.48
rs6759298	***intergenic 2p15***	G=0.4097/2052	184	211	64	459	0.64	0.37
rs6871626	***IL12B***	A (HapMap)	207	185	54	446	0.67	0.33

3.5 Concordance Analysis

3.5.1 Genotyping concordance between Microarray genotyping and iPLEX Pro Genotyping

A concordance analysis was completed using the results from the genotype results from this study and the genotype results from microarray genotyping information that was acquired from the previous AS GWA study. The rs numbers that were in both studies had 100 percent genotyping concordance see Table 3.20 below.

Table 3.20: Discovery Genotyping Concordance Analysis.

Comparison of Case Genotyping information from Microarray genotyping and iPLEX PRO genotyping.

rs number	Gene	Cohort Concordance (%)
rs10781500	*CARD9*	100
rs11209026	*IL23R*	100
rs2066844	*CARD15/NOD2*	100
rs887466	*HLA-C*0602*	100

3.5.2 Linkage Disequilibrium Validation Study

The rs numbers that were in LD with rs numbers that were significantly associated with the axial SpA were compared. The LD rs numbers genotype data came from this study and the significantly associated rs numbers came from the microarray genotyping information from the previous AS GWA study. Please see the Table below for concordance percentages. (Table 3.21)

Table 3.21: Cohort Concordance between assays that were in LD in the assay design. The LD SNPs have a SNP that is associated at a genome-wide significance.

rs number	Genetic variant	LD rs number	r^2	D'	Cohort Concordance %
rs2844603	*HLA-B*3906*	rs2853931	1	1	100
rs1265110	*HLA-B60*	rs1265163	0.98	1	98.9
rs3130501	*HLA-B44*	rs3132528	0.97	0.98	100

3.5.3 Concordance between assays which were replicated in the design

When the Discovery Cohort was genotyped there was two genetic variants that were replicated with two separate rs numbers. These genetic variants were *HLA-B27* and *HLA-CW6*. Both of these variants are very significantly associated with SpA, therefore it was important to capture the superior variant for the assay.

For the *HLA-CW6* the two rs numbers are not linked. The SNPs were compared to see if the results were concordant by comparing rs887466 and rs10456057. These genotypes were not concordant and were not in linkage disequilibrium. This is further illustrated by Table 3.22, if you observe the count for each genotype there are large differences in the allele frequency. Using the allele peak heights as a quality metric, rs887466 had higher average peak heights as compared to rs10456057. For these reasons the decision was made to only include rs887466 in the final panel.

Table 3.22: HLA-Cw6 Redundant Variants Peak Height and Allele Count Comparison.

Less than 7 (%) and Less than 3 (%); indicates the percentage of samples that feel below this quality control metric.

			rs10456057	rs887466
			LD with HLA-CW6	**HLA-C*0602**
Homo WT	**Peak Heights**	AVG	16.6	38.7
		SD	7.0	16.4
		MIN	1.0	3.0
		MAX	47.3	81.9
		Less than 7 (%)	8.9	3.5
	Count		1103	375
Homo Mutant	**Peak Heights**	AVG		41.4
		SD		18.9
		MIN		3.2
		MAX		108.2
		Less than 7 (%)		2
	Count			346
Hetero	**Peak Heights**	AVG 1	5.3	18.1
		SD 1	2.6	8.0
		MIN 1	1.0	1.3
		MAX 1	16.7	48.6
		Less than 3.5 (%)	21.8	2.0
		AVG 2	11.8	20.4
		SD 2	5.6	9.1
		MIN 2	2.5	1.2
		MAX 2	24.3	54.0
		Less than 3.5 (%)	4.1	2.1
	Count		243	814

For the *HLA-B*27* the two rs numbers were in linkage disequilibrium with r^{2} – 0.6 and D`- 0.77. These genotypes were somewhat concordant. Using the allele peak heights as a quality metric (Table 3.23), the rs4349859 outperformed rs116488202. Rs4349859 had a higher average of peak heights. In addition, after the Discovery cohort was complete it was observed that some of the samples were clustering off axis. It was then found that some samples had a proximal SNP in the primer binding area causing this strange clustering. For these reasons the decision was made to only include rs4349859 in the final panel.

Table 3.23: HLA-B*27 Redundant Variants Peak Height and Allele Count Comparison.

Same description as Table 3.23

			rs116488202	rs4349859
			HLA-B*2705	HLA-B*2705
Homo WT	**Peak Heights**	AVG	16.9	29.0
		SD	6.7	14.5
		MIN	0.8	1.2
		MAX	44.1	72.1
		Less than 7 (%)	7.6	5.6
	Count		896	972
Homo Mutant	**Peak Heights**	AVG	15.9	30.6
		SD	7.4	12.1
		MIN	1.5	12.2
		MAX	41.6	55.1
		Less than 7 (%)	9.9	0
	Count		252	23
Hetero	**Peak Heights**	AVG 1	9.7	13.9
		SD 1	4.9	6.8
		MIN 1	0.7	0.9
		MAX 1	27.9	51.9
		Less than 3.5 (%)	8.4	4.3
		AVG 2	7.8	13.9
		SD 2	3.8	6.6
		MIN 2	0.8	0.7
		MAX 2	18.6	36.6
		Less than 3.5 (%)	9.8	4.3
	Count		143.0	541

3.6 F-score (Discrimination Testing)

The F-score was used for the feature/markers selection to indicate the discrimination between the cases and controls. The larger of the F-score, the more likely this marker is more discriminative. This was completed using the Discovery Cohort. From the Figure 3.3 of F-score, we can find the *HLA-B*2705* and other HLA alleles are the most informative markers for individually predicting AS and axial SpA. This is the only figure of F-score analysis that will be published in this thesis as the remaining results and individual F-scores for this testing can not be published as it is apart of a patent application that is currently in progress.

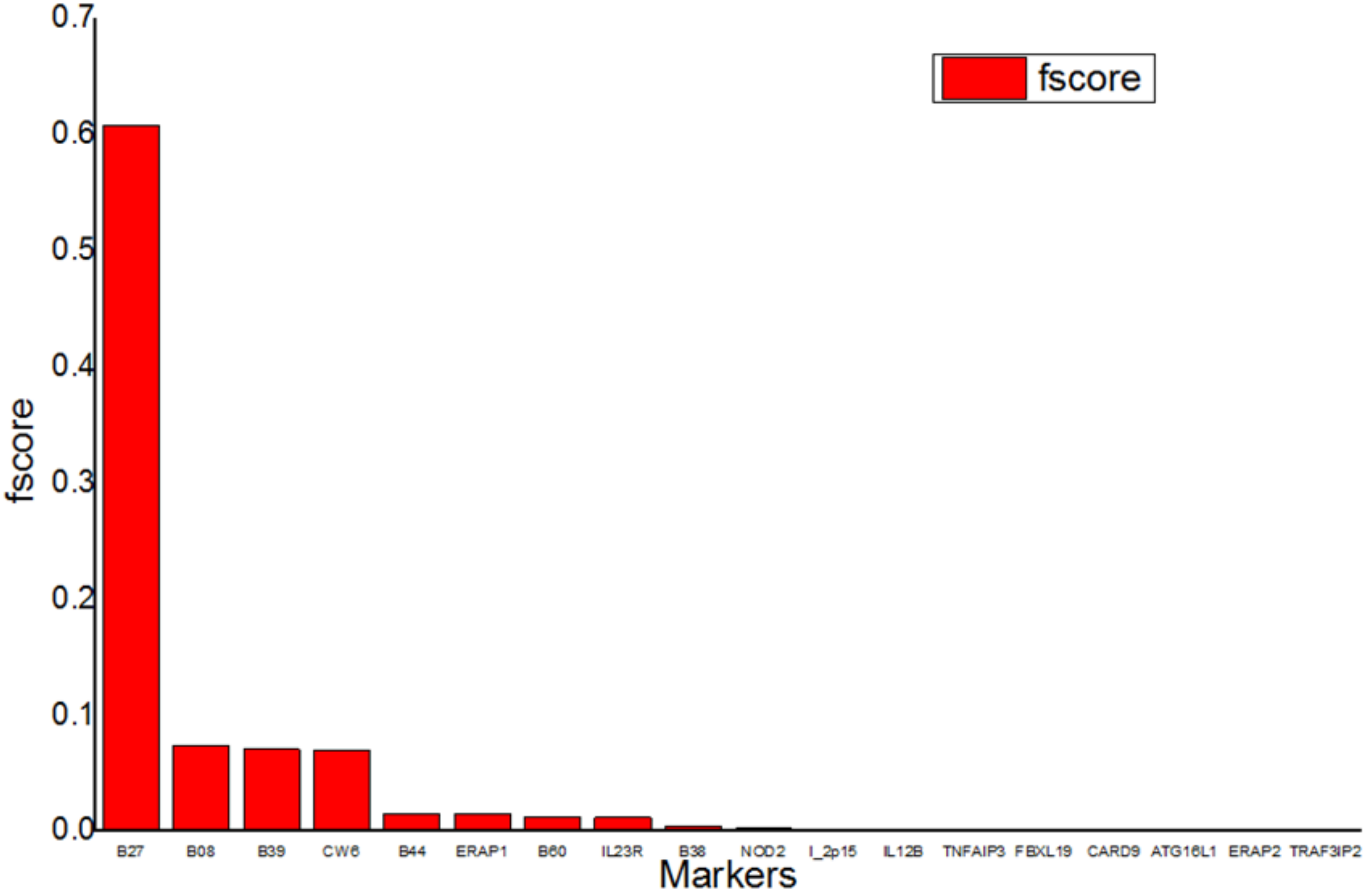


Figure 3.3: F-score Analysis of Discovery Cohort.

HLA alleles are the most discriminatory markers at predicting AS individually. All 18 markers listed on the x-axis.

3.7 Machine Learning Algorithm

A J48/C5.0. Decision tree model was applied independently on both the discovery and validation cohort samples to construct this tree. The decision tree uses each variant has a classifier. At each classifier a decision is made based on genotype and as each combination, each patient genetic information falls into a leaf which can give a risk score for how likely the person is to develop disease. The tree is spilt into two portions the *HLA-B*27* positive tree and the *HLA-B*27* negative tree. The results will be presented in tables based on these two trees and via cohort (Discovery Tables 3.24, 3.25 and Validation Tables 3.26, 3.27). Figure 3.4 shows a visualization of decision tree model for the discovery cohort.

Table 3.24: Discovery *HLA-B*27* Positive Decision Tree Results

Disease Class	Sensitivity	Specificity	ROC	F-Measure	Precision
Case	0.74	0.84	0.8	0.79	0.85
Control	0.84	0.74	0.8	0.78	0.72
Total	0.78	0.8	0.8	0.78	0.79

Table 3.25: Discovery *HLA-B*27* Negative Decision Tree Results

Disease Class	Sensitivity	Specificity	ROC	F-Measure	Precision
Case	0.21	0.91	0.58	0.30	0.50
Control	0.91	0.21	0.58	0.81	0.73
Total	0.70	0.43	0.58	0.65	0.66

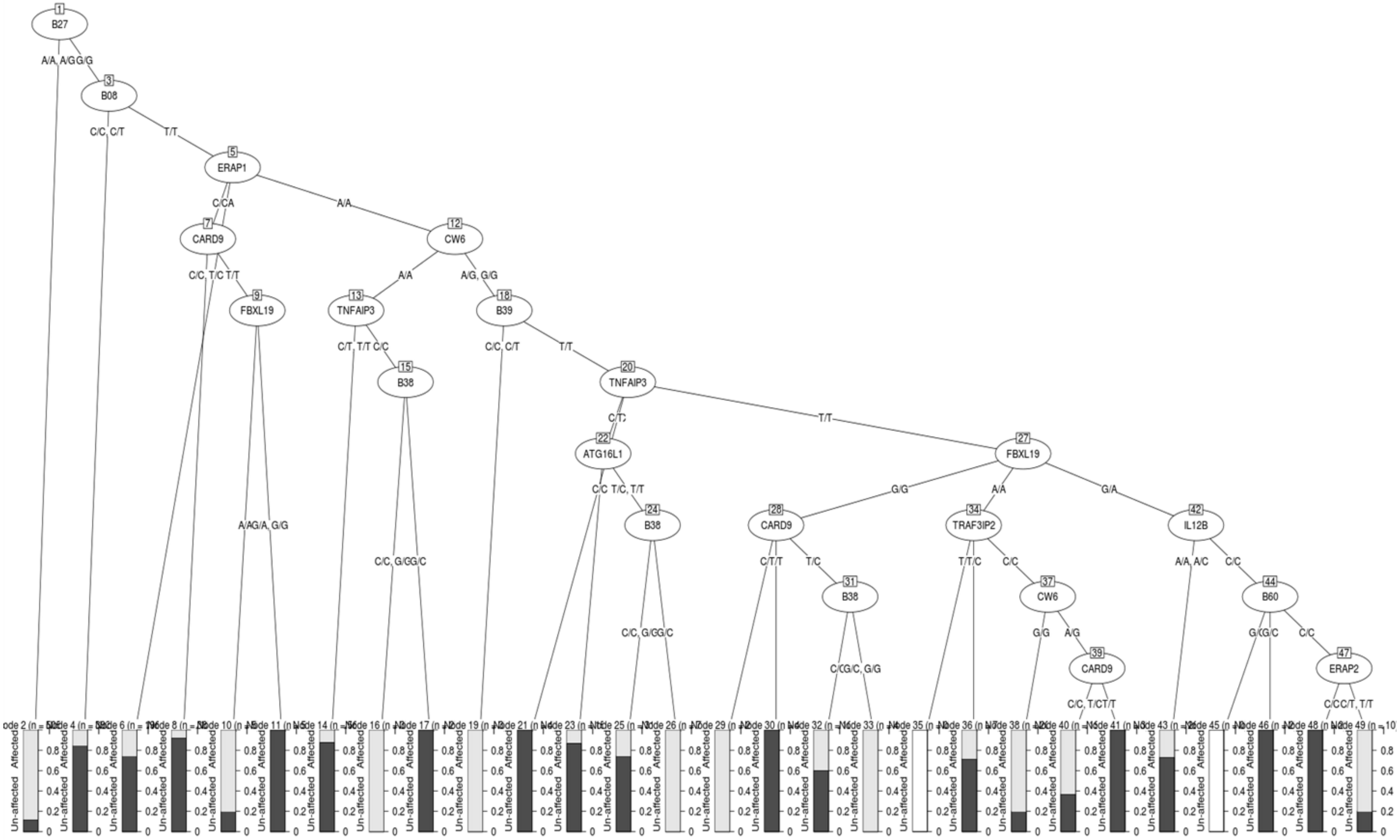


Figure 3.4: Discovery Cohort Machine Learning Decision Tree Model.

Grey leaves indicate cases and black leaves indicate controls. Nodes are Genes and each branch is the type of genetic variation of the node (the gene).

Table 3.26: Replication Cohort *HLA-B*27* Positive Decision Tree Results

Disease Class	Sensitivity	Specificity	ROC	F-Measure	Precision
Case	0.76	0.86	0.82	0.82	0.90
Control	0.86	0.76	0.82	0.77	0.70
Total	0.80	0.82	0.82	0.80	0.82

Table 3.27: Replication Cohort *HLA-B*27* Negative Decision Tree Results

Disease Class	Sensitivity	Specificity	ROC	F-Measure	Precision
Case	0.08	0.93	0.53	0.14	0.34
Control	0.93	0.09	0.53	0.80	0.70
Total	0.68	0.34	0.53	0.60	0.60

Both cohorts were then combined into one cohort. In order to make sure both of these cohorts were able to be combined multi-dimensional scaling was performed. This analysis showed that both cohorts were similarly distributed. The results for the combined cohort are displayed in Tables 3.28 and 3.29.

Table 3.28: Combined Cohort *HLA-B*27* Positive Decision Tree Results

Disease Class	Sensitivity	Specificity	ROC	F-Measure	Precision
Case	0.71	0.92	0.80	0.80	0.92
Control	0.92	0.71	0.80	0.79	0.70
Total	0.80	0.83	0.80	0.80	0.83

Table 3.29: Combined Cohort *HLA-B*27* Negative Decision Tree Results

Disease Class	Sensitivity	Specificity	ROC	F-Measure	Precision
Case	0.03	0.97	0.50	0.05	0.31
Control	0.97	0.03	0.50	0.81	0.70
Total	0.69	0.32	0.50	0.58	0.58

3.7.1. Additional Observations from Machine Learning Algorithm Discovery Cohort

Upon further investigation of the machine learning decision tree algorithm trends within the data appeared showing high predictive values for certain clusters of genotypes. *HLA-B*2705* positive patients with genetic variants in *HLA-B60*, *ERAP1*, *IL23R*, and *CARD9* occurs in 9% of the cohort population and resulted in specificity of 100%. For *HLA-B*2705* negative patients, with a combination of *HLA-B08*, *CARD9* and *ATG16L1* occurs in 10% of the cohort population and resulted in specificity of 96%.

In addition, to the machine learning algorithm an logistic regression algorithm was performed. This model used *HLA-B*2705* and all 18 variants reached the best performance of accuracy 0.871 with specificity 0.905 and sensitivity 0.716, Area under the curve (AUC) of 0.87, Matthew's Correlation Coefficient (MCC) of 0.636. These results were disregarded as it did not have the visual component as well as *HLA-B*27* was not given the most important weight.

3.8 Comparison of Genetic Screening Algorithm to the current Diagnostic Evaluation

A comparison was done using results from the clinical trial of the current diagnostic evaluation and this genetic-based screening algorithm (Table 3.30).

Table 3.30: Comparison of HLA-B*27 negative and positive decision tree to current ASAS Clinical Diagnostic Evaluation.

B27 represents *HLA-B*27,* ASAS Clinical represents the sensitivities and specificities of the clinical arm

Type of Sample	Sensitivity	Specificity
B27 Positive	0.80	0.83
B27 Negative	0.67	0.32
ASAS Clinical (B27 Positive)	0.57	0.83

Chapter 4 Discussion

Back pain is very prevalent, most people will experience back pain at some point in their life and it is the largest cause of disability worldwide (120). Back pain can be categorized into two main categories; non-inflammatory (degenerative or mechanical) and inflammatory. Presently the evaluation of musculoskeletal pain in the primary care setting is unreliable, particularly as it relates to inflammatory low back pain as it is difficult to visualize the axial spine, unlike peripheral joint pain. The management plan and overall prognosis of degenerative and inflammatory back pain is quite different, particularly with the advent of biologic therapy that has revolutionized the management of inflammatory back pain. Thus, there is a need a reliable and relatively inexpensive method to differentiate between these two fundamentally different causes of low back pain.

The most common disorder for inflammatory low back pain is axial SpA. Inflammatory back pain, particularly AS, is a highly heritable disease. AS exhibits the largest effect size of any genetic variant to a complex disease. More strategically utilizing the genetic basis of this disease may lead to a better screening process to discriminate inflammatory from degenerative low back pain. Although the utility of *HLA-B*27* is highly debated for the screening of low back pain (141), it is recommended to use this as a screening tool, as per the ASAS suggestions. *HLA-B*27* is very sensitive for AS and moderately sensitive for axial SpA at predicting disease; however, it is not very specific. *HLA-B*27* has a high MAF at around 7- 10%, meaning depending on the population, one in ten in the population carry this variant. However, only 5% of people with this variant actually develop disease. This causes problems when screening for inflammatory back pain and SpA, as the positive predictive value is poor.

In 2009, the ASAS developed a new diagnostic evaluation for axial SpA, with hopes to diagnose the disease earlier. The current ASAS diagnostic evaluation consists of two arms; the imaging arm that relies on scaroiliitis imaging and the clinical arm that relies on a positive *HLA-B*27* status and clinical features (90, 93). The imaging arm of the criteria has superior sensitivity and specificity as the MRI can detect inflammation and structural abnormalities (1, 90, 94, 95). Presently, to satisfy the classification criteria for axial SpA, a patient must have radiographic or MRI evidence to satisfy the criteria. Access to MRI may be challenging for some jurisdictions due to cost and so the clinical arm provides an avenue for identifying patients with axial SpA. The clinical arm capitalizes on *HLA-B*27* statuses along with some extra-articular manifestations or inflammatory features. However, the clinical arm is not as sensitive and there is sufficient need for improvement. This leads to the goal of this project which was to improve the clinical arm of axial SpA. Thereby, developing a genetically enhanced screening algorithm that will potentially represent a major advance in the early detection of SpA.

4.1 SNP based testing

The type of genetic variation selected in this genetic-based screening algorithm were SNPs. Previous work in our research group has shown that SNP-based testing can effectively replace *HLA-B*27* locus testing with high concordance and accuracy. Lehr et al., (2017) reported that in 1000 consecutive patients receiving *HLA-B*27* locus testing was also sequenced for the rs116488202 and rs4349859 *HLA-B*27* SNPs via real-time PCR (142). They found that both these SNPs had an analytical sensitivity and specificity of 97.6% and 99.9%. These results showed that SNP-based testing is a sufficient replacement for full *HLA-B*27* locus testing and SNPs can potentially accurately determine genetic risk. In addition, *HLA-B*27* locus testing costs $64.91 per patient and the replacement SNP based *HLA-B*27* testing costs $4.61 per patient. Therefore, SNP-based testing can offer the health care system large savings (142). Another example of SNP-based testing is from macular degeneration where the SNPs HTRA serine peptidase 1, complement factor H, and apolipoprotein E predict disease outcome.

Variant selection for this panel was very important. As discussed earlier in the methods section of this thesis, all of the chosen SNP variants for the gene panel came from large robust GWA studies. These robust studies have been completed using SNP microarrays to determine association, particularly using the Illumina Immunochip (200,000 SNP array). However, SNPs are not the only type of genetic variation and other variants such as CNVs, insertions and deletions could be a more predictive disease markers of SpA. It is hard to know which other variants would have been most effective as genome-wide studies on CNVs have been limited and the identification of exonic insertion and deletion variants have only been limited to next-generation sequencing studies in families. As more and more genome-wide studies are performed on other variant types, a more informative variant may come to light. At the time, these variants were closely considered and determined to be the most important to be included in this genetic panel, this process and selection may not be perfect. For example, there may have been a variant that was missed that was released in the literature recently that has a high predictive power. This is a limitation of this study that must be considered; however, the variants were selected very strategically in order to potential maximize predictive power. Other genetic prediction approaches will be discussed in a later section when other genetic screening algorithms for axial SpA are discussed and compared.

4.1.1 MassARRAY

For this project, it was determined that the most time efficient and cost-effective technology platform to use was the Agena Biosciences MassARRAY. The technology enables multiplexing variants together within one assay, it is cost-effective (approximately $25 a sample) and less-time consuming when compared to other genotyping platforms. The protocol for the MassARRAY is very automated and requires limited reaction volumes and produces accurate genotyping calls. Most importantly for the purpose of this project is that these multiplexed assays were custom designed. With the use of the ADS and Typer Software packages, various variants were able to be multiplexed over 2 assays. In addition, this was especially helpful as the HLA region is difficult to sequence, due to its repetitive and polymorphic nature. With this custom designed assay, appropriate custom changes were able to be made in order to ensure accurate genotyping results were produced.

4.1.2 Rationale for each SNP

Eighteen (18) SNPs were selected for this panel based on their GWA significance, clinical significance, MAF and odds ratio. These SNPs were also chosen on the basis of their association with AS, psoriasis, IBD and uveitis. The genetic variants *HLA-B*27*, *HLA-B60*, *CARD9*, *IL23R*, *ERAP1*, intergenic region 2p15, *IL12B*, and *ERAP2* were associated with axial SpA and AS. The genetic variants *HLA-B*27, ERAP1* and intergenic region 2p15 were associated with uveitis. The genetic variants *HLA-CW6, HLA-B44, HLA-B38, HLA-B39, HLA-B08, TNFAIP3, TRAF3IP2,* and *FBXL19* were associated with psoriasis. Finally, the genetic variants *CARD15/NOD2*, *HLA-B*27*, *CARD9*, *ATG16L1* and *IL12B* were associated with IBD. The pathological significance of each SNP was discussed in detail in the introduction section of this thesis. All of these SNPs play important roles in the pathophysiology of SpA and related diseases. By targeting multiple diseases associated genes, the panel was able to increase the sensitivity of the screening algorithm by predicting risk for related extra-articular manifestations, as well as increase specificity by including non-MHC associated genes and expanding the scope of the genetic screening in SpA from just *HLA-B*27*.

4.2 Positive HLA-B*27 Arm

The ASAS's clinical arm of the diagnostic evaluation is the most applicable to be compared to the *HLA-B*27* positive portion of the decision tree algorithm. Mainly, because in order for a physician to utilize the clinical arm, the patient must first be positive for *HLA-B*27* and secondly, as one of the primary goals of this project was to improve the genetic component of this classification criteria. When the machine learning algorithm was compared to the clinical arm, there was an improvement in the sensitivity and a similar specificity. This shows that an enhanced genetic based risk algorithm can give a more informative diagnostic screen than the current diagnostic evaluation. As well as utilizing SNP based technology, cost will be less than current PCR based methods of HLA-B*27 locus testing. Thus, the *HLA-B*27* positive portion of the screening algorithm has a higher sensitivity, the same specificity and lower costs than the clinical arm of the ASAS criteria.

Another interesting finding was that the results were comparable to the ASAS's overall diagnostic evaluation (the overall evaluation includes both the imaging and the clinical arm). The current overall ASAS's sensitivity and specificity are slightly greater than the sensitivities and specificities of this genetic-based screening algorithm, however, these values are only slightly higher. Thus, this shows that a genetic-based screening algorithm may not be as specific as scaroiliitis imaging, it can potentially attain similar prediction power needed for a diagnosis of axial SpA.

Although *HLA-B*27* is a fair indicator of for screening axial SpA, as a whole the ASAS diagnostic evaluation is not very sensitive, with low values in both arms of the criteria. A possible explanation for the increase of the sensitivity of this genetic screening algorithm, could be that related extra-articular features genetic variants were included. These comorbid diseases are already included as a SpA feature in the ASAS diagnostic evaluation (90, 91). Thereby, including their genetic contribution is both targetable and could have increased the sensitivity result from this genetic screening algorithm.

These related phenotypes are all seronegative immune-driven diseases and the current literature has shown that these chronic inflammatory diseases have substantial comorbidity. These extra-articular manifestations are commonly found in patients with inflammatory back pain (33, 87, 88). Vander Cruyssen et al., (2007) published a study with a cohort of 847 AS patients, they found that 42% of patients had one of these additional inflammatory diseases (101). Of this 42% of patients with an extra-axial manifestation, 50% had uveitis, 20% had psoriasis, 19% had IBD and 10% had a combination of the extra-axial manifestations (101). Robust consortium studies using pathway network analysis have shown that these auto-immune diseases share common genetic loci, responses to treatment and etiological pathways (33, 87, 88). In addition, all of these auto-immune diseases have large genetic associations. Psoriasis has a known genetic contribution of approximately one-third, and IBD of about 40%. These extra-articular feature phenotypes tend to worsen with disease progression and by incorporating these genetic variants the risk can potentially be established even before the onset of these extra-articular features manifest. This offers an opportunity as the possibility of having one of these diseases is both targetable and does not change over time. Thus, including the genetics of these diseases could have caused the screening algorithm to be more sensitive. Adding the genetic risk of these comorbid diseases also expands the predictive power of the algorithm to these other auto-immune diseases.

Possible explanation for why there was a slight increase in specificity in the algorithms of this project is that strictly testing for *HLA-B*27* is sensitive; however, it is not very specific. As stated above *HLA-B*27* has a MAF of about 10%; however, 10% of the population does not develop axial SpA – only approximately 5% of people with a positive *HLA-B*27* genetic variant develop disease. Most importantly, non-inflammatory back pain is very common, but inflammatory back pain is rare. The incidence rate of AS is 0.1 -0.3% and axial SpA is 0.3-1%. This incidence rate combined with the MAF of *HLA-B*27* creates a major diagnosing problem. As a person who is *HLA-B*27* positive is more likely to have chronic low back pain then have axial SpA. This makes testing for *HLA-B*27* genetic variant alone not very specific.

Previous GWA studies and Immunochip data have reported multiple SNP associations that have reached GWAS significance for AS as well as other related phenotypes; psoriasis, IBD and uveitis (33, 50). There have been 31 non-MHC genetic variants reaching genome-wide significance from previous GWA studies (9, 19, 50). These identified non-MHC genetic variants have uncovered new insight into pathways of pathogenesis of axial SpA and could be the answer to genetic specifically axial SpA diagnosis (4, 16). For example*,* *ERAP1* is one of the non-MHC loci that has a strong genetic associations to axial SpA and functionally this gene appears to be very important to our current understanding of AS (2, 9, 19). The *ERAP1* variant is correlated with axial SpA patients that are also positive for *HLA-B*27* (19). In patients that are positive for both *ERAP1* and *HLA-B*27* the odd ratio of disease increases to over 200 (19). There is strong evidence that *ERAP1* interacts with *HLA-B*27* (19)*,* as *ERAP1* is a modifier protein and when it is mutated there is less efficient trimming of *HLA-B*27* (3). By combining these loci on the genetic algorithm, the specificity will increase.

Many other non-MHC genetic variants are incorporated into the genetic screening algorithm. These other variants will capture the broader population and make the panel more specific. Even if the effect sizes of these non-MHC variants are low individually, by combining these loci together the genotypes of each individual will be more specifically related to SpA. Inclusion of the extra- articular features variants may further increase the sensitivity of the genetic-based screening algorithm and the addition of additional markers reaching GWAS significance and with MAF over 1% may help in improving the specificity. Thus, a combination of these two approaches were used to develop our screening algorithm.

4.3 Negative *HLA-B*27* Arm

Currently, there are no genetic screening tools or tests for *HLA-B*27* negative patients. Thus, this screening algorithm could be the first screening tool to attempt to screen *HLA-B*27* negative patients and therefore the *HLA-B*27* negative side of the tree cannot be compared to any other screening tool. Unfortunately, the *HLA-B*27* negative side of the tree did not perform very well and it can be inferred from the results that it is likely not going to be a very predictive tool for picking up negative *HLA-B*27* cases. However, from analysing the decision tree it was observed that certain patients with axial SpA that were *HLA-B*27* negative with genetic variants in *HLA-B08, CARD9, and ATG16L1*, which occurred in 10% of the population had a specificity of 96%. This shows that although the overall positive predictive value of the *HLA-B*27* negative tree is poor, a small subset of patients can be screened using this algorithm, if certain pattern of SNPs are found.

In addition, as the goal is for many screening tools, it is not to find everyone, but eliminate and catch as many people as possible. That being said the *HLA-B*27* negative side of the tree has a high negative predictive value. Meaning that it can effectively identify patients that do not have SpA. For example, if a patient is *HLA-B*27* negative and is also negative for a couple other variants within the tree, it can confidently be concluded that this person likely does not have or will not develop SpA. For a screening tool, this is a very good benefit as a physician can be told that a patient has a very high likelihood of never developing SpA; therefore can stop testing for SpA and continue on a track to the proper diagnosis.

4.4 The two cohorts: Discovery and Validation

This study had two separate cohorts: the discovery and the validation cohort. This was a valuable approach as there was two genetic-based screening algorithms created in separate independent cohorts. Each cohort had a different population make-up. The discovery cohort was comprised of participants from Newfoundland and Ontario, whom were diagnosed with both axial SpA and AS. While the validation cohort was comprised from participants from Alberta who were diagnosed with AS. Results from the discovery cohort were replicated in the validation cohort. Furthermore, from using this approach it was determined that both cohorts gave similar results, and were similarly distributed. Since both cohorts were similarly distributed, it was decided to combine the cohorts as they had similar precision values and machine learning are more robust the larger the cohort.

4.5 Genetic Screening Algorithms

4.5.1 Genetic risk score

In the current literature, genetic epidemiologists have concluded that that combining multiple significant loci into a global genetic risk model, can increase prediction accuracy for some complex diseases (129-134). Most of published genetic risk models rely on additive genetic risk, meaning equating risk to how many minor or risk allele an individual possesses. A genetic risk model was published by a group studying psoriasis. These researchers used 10 SNP variants to create two types of genetic risk models (129). They created an additive genetic and a weighted genetic risk model. The authors concluded that both genetic risk models were able to significantly better predict disease risk as compared to any individual SNP could (129). Thus, certain groups have shown that a genetic risk model can have a higher predictive power as compared to the predictive power of an individual genetic variant (129). This approach is useful as it includes additional variants rather than focusing on the monogenic approach which is currently in use in axial SpA. However, it neglects to include how discriminative each marker is individually as well as gene-gene interactions. This approach has just been used in axial SpA and will be discussed in a later section. It should be noted that when applying this approach to SpA, we have *HLA* genes that are very sensitive at predicting disease, compared to other non-MHC variants that should have different weightings.

4.5.2 Machine learning

Machine learning can capitalize on large complex data sets in order to make intelligent complex decisions. This type of programming is being implemented everywhere around us and particularly in health care machine learning is proving to be very useful in improving health care outcomes. Recently, Google has used machine learning to demonstrate its predictive power for diagnostics and screening. They used a deep learning algorithm to detect metastasis in cancer patients (143). Metastasis is hard to diagnosis by pathologists even with very extensive training. There are large number of slides to analyze and it is easy to miss something. Google programmed an algorithm that could detect metastasis in patients from tumour cell images from biopsy's and CT scans. This same dataset was then analysed by pathologists. The algorithm detected 89% of the metastasis in the dataset, while the pathologists given no time constraint only detected 73% (143). This study shows how from large amounts of data machine learning algorithms can assist highly trained specialists in diagnosing complex diseases and this approach can be applied in many other fields of medicine.

There are many different types of machine learning approaches. These range from deep learning and simple vector machine models to decision tree models. In this project, we decided to choose a decision tree model as it is the easiest machine learning model to visualize. Decision trees are unique as users can easily understand why the classifier or node makes a decision. They are one of the few models where the results are interpretable. This can be very useful in practice as users can determine exactly why the classier makes a decision. This is very helpful especially in settings where un-trained individuals are receiving information from the algorithm. As physicians with no training in machine learning algorithms and limited training in genetics will be potentially receiving the information from this algorithm, the decision tree model best fits with this consideration. It was important to make sure that the information from the algorithm was easy to understand and a decision tree algorithm can be assist in the ease of transfer of information from the result. The decision tree algorithm makes the result from the algorithm easy to understand and also assists in making the communication between the physician more effective and educational.

When programming the decision tree, the first decision made in the algorithm was chosen as *HLA-B*27* status (Figure 4.5.1). *HLA-B*27* is the most descriptive and informative genetic variant by far. In a study in 2005, Reudewalit et al determined that *HLA-B*27* status had a likelihood ratio of 9 of developing disease (144). In addition, a positive HLA-B*27 status is required for a patient to be diagnosed and screened under the clinical arm of the ASAS criteria. Therefore, from the literature review of both clinical and genetic practice, *HLA-B*27* was by far the most important variant and therefore the algorithm selected it as the first decision. When this algorithm splits it represents two separate tests an improved *HLA-B*27* screening tool that is more specific and the first ever negative *HLA-B*27* screening tool.

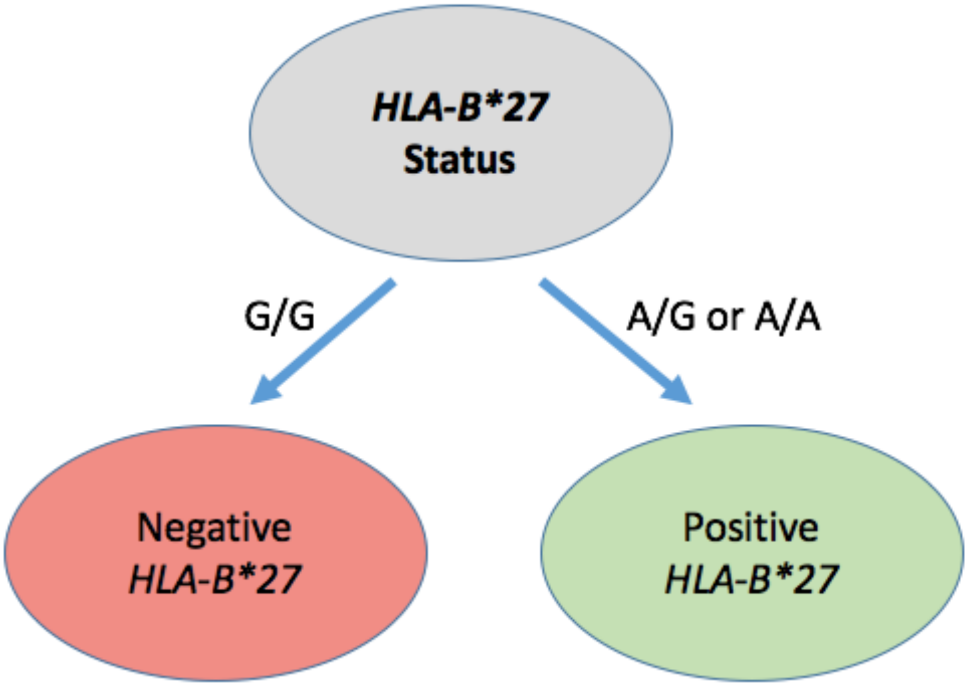


Figure 4.1: Machine Learning Decision Tree Model First Decision.

This decision was coded in a dominant form of inheritance.

4.6 Other Genetic Risk Algorithm studies in axial SpA

4.6.1 Immunochip Based Study

A study out of Australia was just published using whole genome SNP microarrays to profile axial SpA (145). This group used the Immunochip 200,000 SNP array used in previous GWA studies in AS. This study had a large cohort of 9,638 controls and 4,428 AS cases diagnosed via the previously used New York criteria and the current ASAS criteria (145). The genetic risk model for this study used was an ROC curve programmed in R. In this study, when patients that meet the ASAS imaging arm of the criteria were compared to the controls the genetic risk model gave an AUC of 0.83 (145). Within this study, patients negative for the ASAS imaging arm were compared to patients positive for the ASAS imaging arm, the AUC for this comparison was 0.65. This study showed that there is a significant difference in the genetics between axial SpA patients and controls (145). This is a promising step forward for utilizing multiple genetic loci to diagnosis; however, even with a 200,000 SNP array, this genetic-based screening algorithm's performance markers were comparable to the results of this study. This suggests that having additionally variants to diagnosis or predict risk of SpA may not be necessary. As well when you compare the costs of a SNP array to the Agena MassARRAY, the MassARRAY is significantly cheaper to implement and use. Making the MassARRAY more accessible (See discussion on economics).

4.6.2 Selected SNPs and CNV's study

Thus far, one other study has been published that utilized a genetic risk score to assist diagnosing AS. This study was done in the Korean population and consisted of 5 CNVs, 7 SNPs and *HLA-B*27* (146). The genotyping was done using a TaqMan assay, making this approach a much more time-consuming and costly approach to genotyping 13 variants. The genetic risk score was derived using a multi-variant logistic regression analysis. This genetic risk score was constructed similar to this project in a two-cohort approach with both a discovery and validation cohort. This study reported that their genetic risk score had an excellent AUC of 0.95 (146). The group also reported that their model performed superiorly over the current *HLA-B*27* model currently in use. In addition, the variants chosen by this study were primarily from the same gene locus, 6 out of the 7 SNPs were different SNPs in the *ERAP1* locus (146). This is a restricted approach as compared to this project where variants from various genetic loci were chosen to consist the genetic panel. Furthermore, the CNVs chosen for the Korean study may not be generalizable as there is not very robust evidence in the literature outlining the pathogenic features and associations of these CNVs at predicting AS. Particularly, the CNVs which have never been associated in the larger European ancestry studies, making it likely that this genetic risk score constructed would not be able to be translated across other populations.

4.6.3 HLA-B*27/B*60 Markers Study

Another study based out of the Netherlands was attempted but never made it to completion. This study's aim was to make a negative *HLA-B*27* test, by utilizing the genetic variant *HLA-B60*. As discussed in the introduction section of this thesis, patients with axial SpA that are negative *HLA-B*27* have been proven to have the variant *HLA-B*60*. However, when this group investigated the predictive power of *HLA-B*60* in a diagnostic setting, they determined that the results were inconclusive (Personal Communication with Dr. Rahman, from the International Congress of SpA 2016, Ghent, Belgium). When a comparison is made between this project and the studies and techniques discussed above it can be inferred that it's a powerful approach to use multiple unique variants as well as limiting the number of SNPs included. However, most importantly all of these other studies have only used regression-based analysis to make predictions of disease. As discussed in this thesis, machine learning is by far a more comprehensive and more powerful technique in order to establish risk and prediction in a population.

4.7 Limitations

4.7.1 Case definition: Radiographic versus non-radiographic spondylitis.

The case cohort used in this study, consisted of participants whom were mostly diagnosed with AS by experienced rheumatologist in three Canadian centres. All AS patients satisfied the modified New York criteria for AS. Therefore, they had at least bilateral grade 2 scaroiliitis or at least unilateral grade 3 scaroiliitis. A smaller subset of patients had non-radiographic axial SpA and were diagnosed by the new ASAS diagnostic evaluation. However, all patients satisfied the ASAS diagnostic evaluation. However, now that radiographic and non-radiographic AS has fallen into the broader category of axial SpA, this is why we have used this term throughout our discussion. Thus essentially our SNP based test was determining the performance of our test with a population that primarily had radiographic AS. The validation cohort primarily had individuals diagnosed with AS in the study. This may have impact on the stratification of the two cohorts, as well as the development of the algorithm. This will be tested eventually when this research is tested in a clinical trial.

4.7.2 Comparison to gold standard

As discussed earlier in this thesis, the MRI is the gold-standard for the diagnosis of SpA. The current costs for this diagnostic technology is quite high; however, these prices are decreasing. In addition, if the MRI becomes more accessible or if the price of the MRI significantly decreases, this may affect the utility of this project and research. Thus, the main limitation of this research is that it has not been tested in a clinical setting. Therefore, it is difficult to make conclusions if this genetic-based screening tool will be helpful in the current health-care system if the MRI is routinely available.

4.7.3 Control Cohort

This project's control cohort was selected from previous studies examining the genetics of complex disease. The controls selected were from studies in Type 2 Diabetes Mellitus, Obesity and Osteoarthritis. The controls from Newfoundland were ascertained on the basis that they did not have autoimmune disease and were Caucasians of North European ancestry. A small subset of the controls were from the Newfoundland and Labrador Colorectal Cancer Study. No information was acquired from these Cancer controls, however since the prevalence of auto-immune disease is very low, it very unlikely that these controls would possess SpA. Another limitation with this control population is the potential for population stratification. Most of the cases are from Alberta and Ontario, while most of the controls are from Newfoundland and Labrador. This cases the potential for population stratification, however population stratification was controlled and adjusted for when the machine learning algorithm was constructed.

4.7.4 LD within the MHC region

Including SNPs that were in linkage disequilibrium with GWA study-associated SNPs was not an ideal scenario. The HLA region of the genome is an extremely difficult area of the genome to sequence, as it is very polymorphic and repetitive. Therefore, these SNPs could either not be imported into the design or would only import in with drastically reducing the stringency of the assay; thereby causing problems for other variants within in the assay. This modification is a limitation but it was necessary in order to include these important SNPs (*HLA-B*39*, *HLA-B60* and *HLA-B44*). The r^2 and D' values were all greater than 0.97 and 0.98 respectively, making the linkage between both the LD SNP and the GWA study SNP very strong. In addition, this project is lucky to be a part of large consortium (SPARCC), so we had microarray data available to compare both genotypes. Upon comparison, we saw an increase in concordance in this data set as compared to the r^2 and D' values published.

4.7.5 Machine Learning - Overfitting

As discussed above machine learning algorithms gains power as the sample size increases. Therefore, a limitation of this study is the population. If there was a larger population used in the study, it may have increased the power and prediction value of the study. Another limitation to this projects machine learning is the potential of overfitting the model. Overfitting can easily happen in a machine-learning decision tree model, as the model gets more specific to the data set the algorithm gets more specific or "fitted" to the data set used to create it. If the model is over-fitted it will not be able to be applied to external data sets. This project attempted to address this issue by using the separate data sets to ensure that the model would not be over-fitted as well as this project used a global pruning technique to prevent this from occurring. It will not be able to be officially determined if this model is over-fitted until the screening algorithm is applied in a broader setting.

4.7.6 Generalizability

Another limitation surrounding this projects study cohort is the ethnicities of the participants. The genetic-based screening algorithm is most suited for the Caucasian population. Currently, the algorithm was created and replicated with participants from European ancestry only. The genetic panel was created using SNPs that had genome significance in the European population, not SNPs from other ancestries. Therefore, it cannot be determined how predictive this genetic-based screening algorithm will be at predicting disease risk in other populations.

4.7.7 Clinical Utility

We have developed a screening algorithm to fast track patients with high likelihood of axial SpA and suggest conservative management for patients that screen negative for axial SpA. For patients that screen positive, we may have overestimated the prevalence as our cohort was enriched for cases (approximately 50% cases and 50% controls) as compared only 5 to 10% of patients with inflammatory low back pain having axial SpA (which is seen in the current clinical setting). For patients that screen negative we likely underestimated the true negative predictive value. A better estimate of the clinical utility will be determined when consecutive patients presenting to primary care physicians are evaluated. The addition of more SNPs could have been helpful to increase the power of the genetic-based screening tool. This makes the number of SNPs included in the panel a potential limitation.

However, it is difficult to judge how helpful this approach would have been toward the performance of the screening algorithm. Another factor to consider is that as the number of variants increase, the algorithm and decision tree becomes more detailed and difficult to understand/visualise.

4.7.8 Implementation of genetic-based screening algorithm

As non-experts will be eventually receiving the information from this screening algorithm, it is important to ensure that the information being presented is easily understood and communicated. This illustrates that it is important to have a balance between the number variants on the panel and potential power from significant amounts of variants. The addition of multiple variants may have made the panel a more informative predictor of disease. It is hard to quantify how many variants of importance are needed to make the most predictive screening algorithm. It is also difficult to understand how primary-care physicians will adapt to this genetic-based screening algorithm. As many primary-care physicians will be unfamiliar with genetic risk and machine learning this may be a potential limitation in the future. That is why it will be important to ensure that physicians are well educated about this genetic screening process and the associated risks.

4.8 Future directions

4.8.1 Economics

Musculoskeletal disease has a large socioeconomic burden in Canada, in 2014 it was estimated to cost $22.3 billion, representing approximately 3% of Canada's gross domestic product and 5.7% of health expenditures (125). Low back pain is the primary expenditure of these estimates and the assessment and treatment of SpA is approximately $7 billion.

A study by Kobelt et al., (2006) estimated the average annual cost of treating and diagnosing axial SpA in Canada. This study estimated that it costed an average of $9,008 yearly to treat/diagnosis a patient with axial SpA (147). The total Canadian cost of treating and diagnosing axial SpA was estimated at $2.2 billion. This study also estimated that patient's out-of-pocket costs represented 33.1% and lost work capacity was 38% of this total figure (147). In addition, a trend was demonstrated that as physical function of the patient decreased the cost of management per year went from $4,000 to $30,000 annually (147).

When focusing on in on replacing the current *HLA-B*27* typing testing with this project's genetic-based screening algorithm, there are large cost savings. These costs savings will be the results of 1) the lower cost of the genetic-based screening algorithm versus the current *HLA-B*27* typing. 2) Fewer false positive results, thus avoiding unnecessary expensive investigations and consultations to confirm results. 3) Fewer false negatives, thus avoiding delayed diagnosis and treatments of affected patients. 4) A reduction of the need of advanced scaroiliitis imaging, such as bone scans, CT scans and MRIs and multiple consultations of specialist physicians.

From this project, a preliminary cost savings was estimated. This is illustrated in Figure 4.2, which is an economic decision tree used to estimate cost savings from this study. It was determined if this screening tool was implemented, there will be significant cost savings for the Canadian health-care system. Currently more than 50,000 patients need to be or are tested annually for *HLA-B*27*. This estimate was determined by the number of actual *HLA-B*27* tests that were ordered in Newfoundland and Labrador over a one-year period and then was extrapolated to match the Canadian population. Currently *HLA-B*27* gene locus typing costs $50 per test in Canada, the estimate cost for the genetic-based screening algorithm is $20 per test. Therefore, it is estimated that the genetic-based screening test would save an estimated $1.5 million dollars annually in direct genetic testing.

As discussed above the genetic-based screening algorithm has superior accuracy. If we analysed the cost estimates per 100 tests. By introducing the genetic-based screening algorithm we can estimate that there will be 10 fewer false positives and 10 fewer false negatives per 100 tests. The current direct health-care expenditure for a false positive is $800; this would include the cost of a CT scan and/or a MRI, which would be necessary to obtain the correct result. The current expenditure for a false negative is approximately $1000; which would be based on delayed treatment and unnecessary imaging. From these estimates, we can estimate that by reducing the false positive and negative rate we can have a total savings of $18,000 per 100 patients tested. When this analysis goes a step further and accounts for eliminating the need for scaroiliitis imaging for an estimate of 50 patients who test as true positives, the costs are reduced $800 per true positive (similarly cost estimate to the false positives). This estimate suggests a potential for a savings of $40,000 per 100 patients tested.

When an estimate of the total savings per 100 test, it is estimated a reduction of $30,000 can be accounted for with the replacement of *HLA-B*27* typing with the genetic-based screening algorithm. Thus, the total cost savings is shows a savings of $88,000 per 100 tests ordered. These estimated are based on direct health-care expenditure making these estimates a conservative. It is difficult to quantify potential savings and costs of fewer physician visits and costs associated with work loss from SpA.

From an economic model evaluation using these numbers above the total direct health-care expenditure for diagnosing inflammatory back pain costs approximately $45.5 million.

From this economic model, an estimated $27 million dollars of direct health-care expenditure can be saved by utilizing this genetic-based screening algorithm. This is a very conservative estimate; however, as this estimate does not account for savings for fewer physician visits and associated absences from work. While it costs to offer patients a new screening test, the capital added to initiating the technology would be easily offset by substantial long term cost savings. These long-term cost savings would include less scaroiliitis imaging and time-consuming laboratory testing, reduction of primary-care physician visits, more appropriate treatments and less work absences.

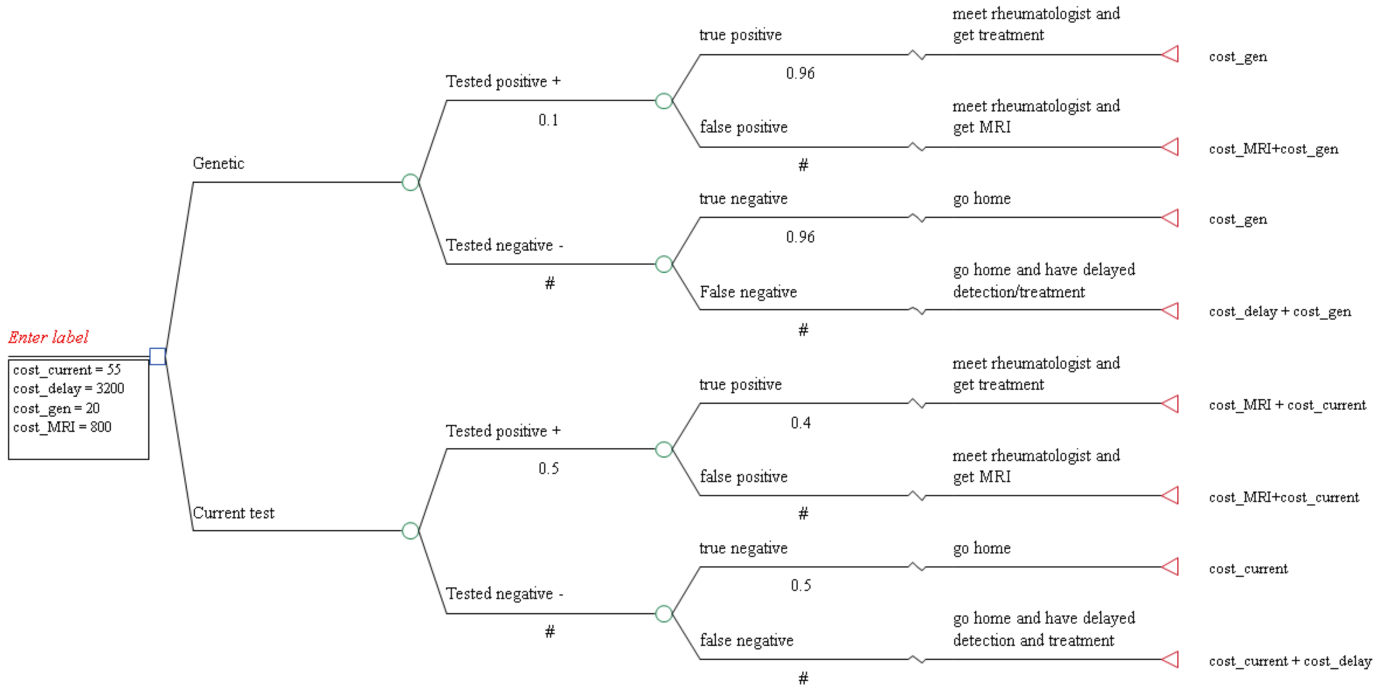


Figure 4.2: Economic Decision Tree

This figure was created by Dr. Hai Van Nguyen (an health economist at Memorial University)

4.8.2 Ethics

Over the last decade primary care physicians have been surveyed about genomic technologies and these surveys have resulted with positive attitudes towards the benefits of these technologies. However, they have low confidence concerning implementations of these technologies into their clinical practices, mainly because of a lack of credible evidence surrounding the clinical utility of this technology (148). There may be many barriers to these feelings and most importantly there is a need to address the knowledge deficit in this area of technology as well as knowledge management (handling and understanding the sheer volume of evidence across the breadth of a clinician's practice).

For the sake of this project this genetic-based algorithm is a screening tool not a diagnostic test. Meaning that the results of this algorithm can give a probability or risk of disease not a strict diagnosis. It is still important that a rheumatologist finishes the diagnosis. Therefore, it is very important that with genetic screening there has to be caution. Genetics is a complex discipline that not many in the general public fully understand and as genetic testing becomes more and more common it is important that people are properly informed for what information they are receiving means. A patient might fall into what is known as the "worried –well".

A "worried-well" patient is a patient that seeks medical treatment continually, in order to be reassured by their physical or emotional support. When a patient hears that they have a relative risk for developing a disease it is very important that this is communicated appropriately. This point further reiterates that importance of effective communication between the primary-care physician and the patient as well as to ensure that the primary-care physicians understands what it meant by the relative risk. This can be accomplished by ensuring that physicians are well educated on genetic relative risk and this genetic-based screening algorithm.

4.8.3 Prospective Clinical Trial

As compared to the current studies released by the Australian and Korean groups, no one has tested their genetic algorithm in a real-life clinical setting. This is the true test to ensure that a genetic-based screening algorithm has high clinical utility. Therefore, the future direction of this project is that it will be implemented into a prospective clinical trial to assess the ability of the algorithm to perform in a clinical setting (Figure 4.3). This will be completed at SPARCC rheumatology clinics. Patients that meet the criteria of younger than 40 years and have been suffering from chronic back pain for greater than 3 months will be eligible to enter the study. The aim is to recruit and enrol 1,200 patients. Patients enrolled in the prospective study will follow the current management and ASAS diagnosis guidelines, as well as DNA will be collected and patient's genotyping information will be assessed via the genetic-based algorithm.

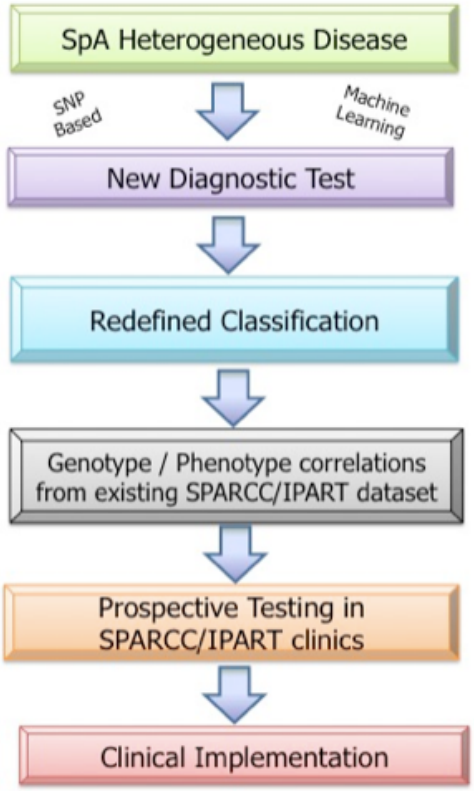
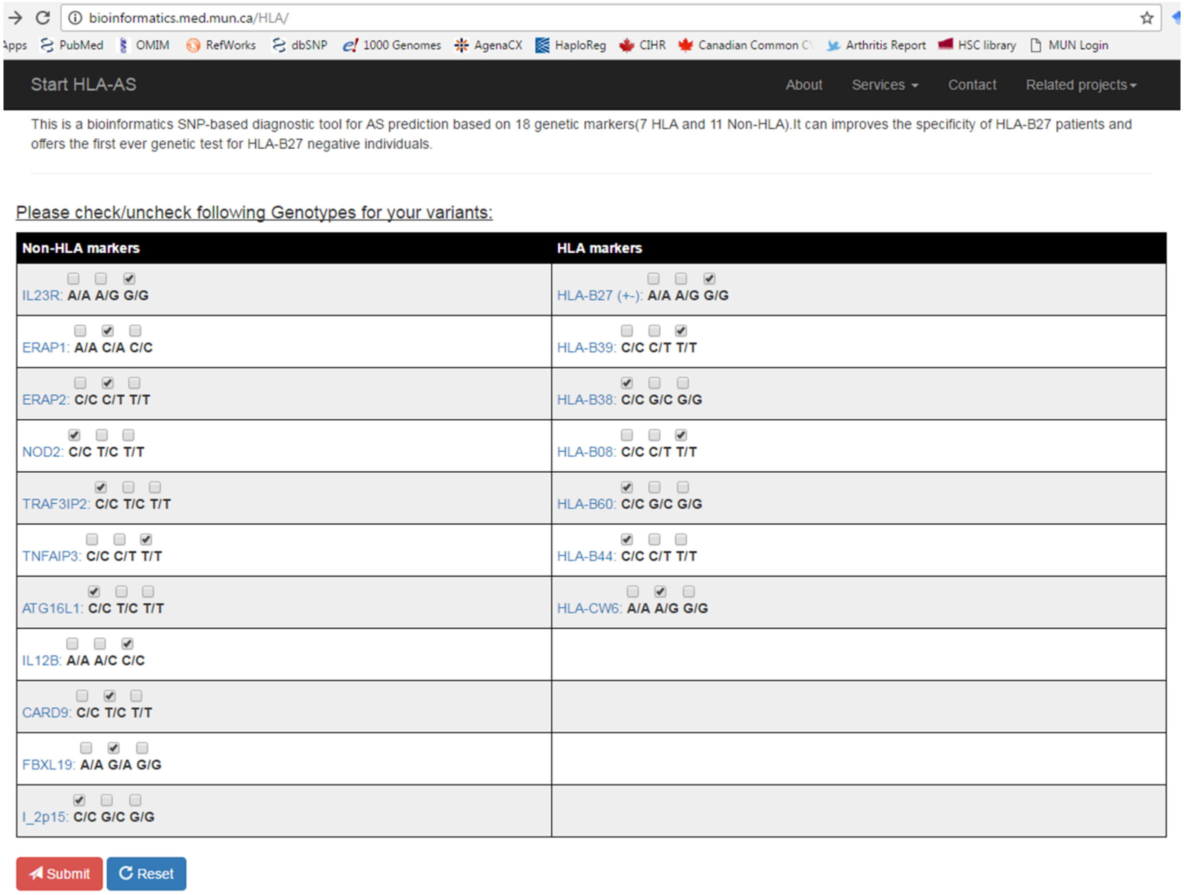

SpA Heterogeneous Disease
SNP Based
Machine Learning
New Diagnostic Test
Redefined Classification
Genotype / Phenotype correlations from existing SPARCC/IPART dataset
Prospective Testing in SPARCC/IPART clinics
Clinical Implementation

Figure 4.3: Flow Chart of Genetic-Based Screnning Algorithm Developement and Prospective Clinical Trial

A common bootstrap web server has been developed and can be accessed at the website; http://bioinformatics.med.mun.ca/HLA/. Once the individual's genotyping is completed, they can be submitted based on the web based tool (Figure 4.4). From there the physician would receive an easy to understand, animated decision-making graph outlining the results (Figure 4.5). The exact subsets will eventually be determined using a health technology assessment for each possible permutation. From this step, a patient will be given a genetic risk factor for disease (Figure 4.6). Patients that are given a high risk will be given a fast track referral to a rheumatologist, patients that are given an indeterminate risk will follow the current management guidelines and the ASAS diagnostic evaluation. Patients that are given a low risk of developing SpA will be referred back to a primary care physician and allied health care professionals.



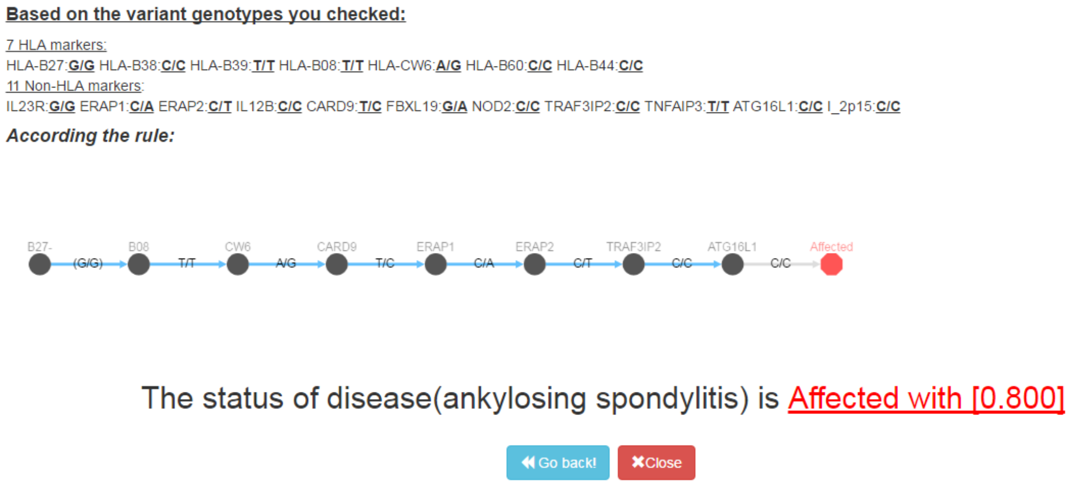
This is a bioinformatics SNP-based diagnostic tool for AS prediction based on 18 genetic markers(7 HLA and 11 Non-HLA).It can improves the specificity of HLA-B27 patients and offers the first ever genetic test for HLA-B27 negative individuals.

Please check/uncheck following Genotypes for your variants:

Non-HLA markers	HLA markers
IL23R: ☐ A/A ☐ A/G ☑ G/G	HLA-B27 (+-): ☐ A/A ☐ A/G ☑ G/G
ERAP1: ☐ A/A ☑ C/A ☐ C/C	HLA-B39: ☐ C/C ☐ C/T ☑ T/T
ERAP2: ☐ C/C ☑ C/T ☐ T/T	HLA-B38: ☑ C/C ☐ G/C ☐ G/G
NOD2: ☑ C/C ☐ T/C ☐ T/T	HLA-B08: ☐ C/C ☐ C/T ☑ T/T
TRAF3IP2: ☑ C/C ☐ T/C ☐ T/T	HLA-B60: ☑ C/C ☐ G/C ☐ G/G
TNFAIP3: ☐ C/C ☐ C/T ☑ T/T	HLA-B44: ☑ C/C ☐ C/T ☐ T/T
ATG16L1: ☑ C/C ☐ T/C ☐ T/T	HLA-CW6: ☐ A/A ☑ A/G ☐ G/G
IL12B: ☐ A/A ☐ A/C ☑ C/C	
CARD9: ☐ C/C ☑ T/C ☐ T/T	
FBXL19: ☐ A/A ☑ G/A ☐ G/G	
I_2p15: ☑ C/C ☐ G/C ☐ G/G	

Figure 4.4: Genetic-Based Screening Algorithm Website.

Access to the algorithm will be through a web portal. http://bioinformatics.med.mun.ca/HLA/



Based on the variant genotypes you checked:

7 HLA markers:
HLA-B27:**G/G** HLA-B38:**C/C** HLA-B39:**T/T** HLA-B08:**T/T** HLA-CW6:**A/G** HLA-B60:**C/C** HLA-B44:**C/C**
11 Non-HLA markers:
IL23R:**G/G** ERAP1:**C/A** ERAP2:**C/T** IL12B:**C/C** CARD9:**T/C** FBXL19:**G/A** NOD2:**C/C** TRAF3IP2:**C/C** TNFAIP3:**T/T** ATG16L1:**C/C** I_2p15:**C/C**

According the rule:

The status of disease(ankylosing spondylitis) is Affected with [0.800]

Figure 4.5: Web-based tool Prediction Screening Algorithm Output

Access to the algorithm will be through a web portal. http://bioinformatics.med.mun.ca/HLA/

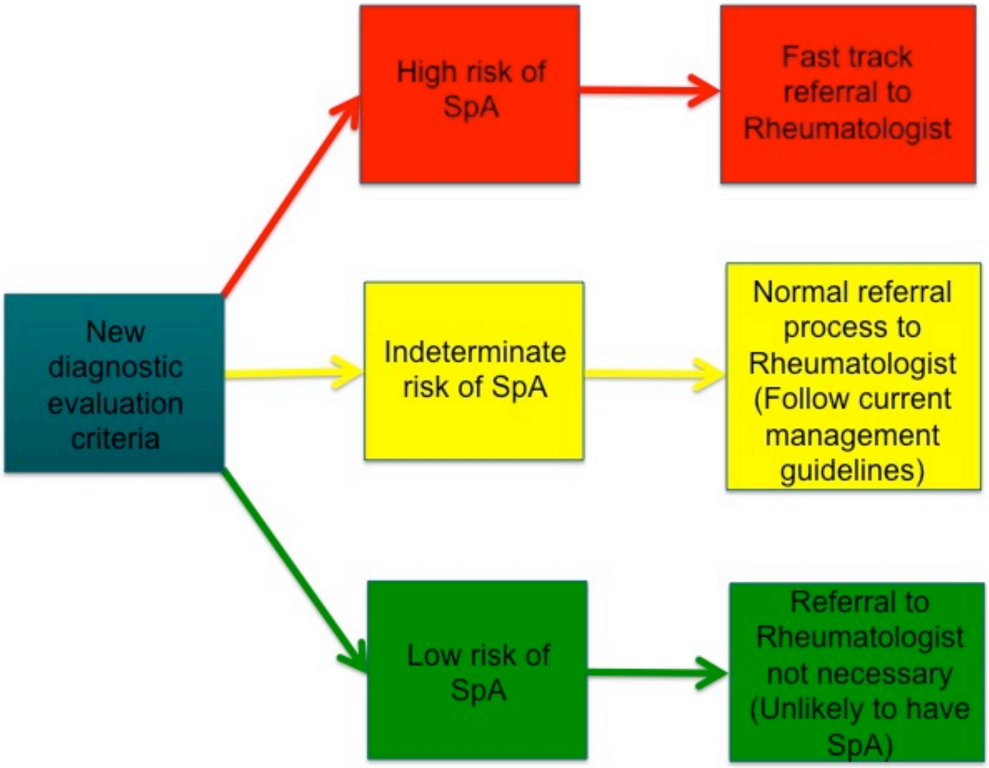

New diagnostic evaluation criteria
High risk of SpA
Fast track referral to Rheumatologist
Indeterminate risk of SpA
Normal referral process to Rheumatologist (Follow current management guidelines)
Low risk of SpA
Referral to Rheumatologist not necessary (Unlikely to have SpA)

Figure 4.6: Prospective Screening Process.

Criteria of ≥ 3 months of severe lower back pain, age of onset < 45 years. Patients determined various levels of risk.

Conclusion

A large proportion of SpA patients exhibit low back pain which is often misdiagnosed as muscular skeletal problems. SpA represents one of the most common inflammatory rheumatic diseases with axial SpA and psoriatic arthritis being the most representative. Currently, from symptom onset to diagnosis is around 10 years with an estimated health care expenditure of 45 million. This disease is a highly treatable subset of low back pain among individuals less than the age of 45 at symptom onset. A genetic-based screening algorithm can facilitate large changes in the screening and evaluation of axial SpA and its related phenotypes.

This genetic-based screening algorithm has a higher sensitivity and similar specificity when compared to the current ASAS diagnostic evaluation for axial SpA. Importantly, this genetic screening algorithm is relatively inexpensive, can potentially facilitates early diagnosis, has a high negative predictive value and could also limit further diagnostic evaluations for a subset of patients.

References

1. van Tubergen A. The changing clinical picture and epidemiology of spondyloarthritis. Nat Rev Rheumatol. 2015 Feb;11(2):110-8.

2. Australo-Anglo-American Spondyloarthritis Consortium (TASC), Reveille JD, Sims AM, Danoy P, Evans DM, Leo P, et al. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nat Genet. 2010 Feb;42(2):123-7.

3. Smith JA. Update on ankylosing spondylitis: current concepts in pathogenesis. Curr Allergy Asthma Rep. 2015 Jan;15(1):489,014-0489-6.

4. Robinson PC, Brown MA. Genetics of ankylosing spondylitis. Mol Immunol. 2014 Jan;57(1):2-11.

5. Brewerton DA, Hart FD, Nicholls A, Caffrey M, James DC, Sturrock RD. Ankylosing spondylitis and HL-A 27. Lancet. 1973 Apr 28;1(7809):904-7.

6. Caffrey MF, James DC. Human lymphocyte antigen association in ankylosing spondylitis. Nature. 1973 Mar 9;242(5393):121.

7. Schlosstein L, Terasaki PI, Bluestone R, Pearson CM. High association of an HL-A antigen, W27, with ankylosing spondylitis. N Engl J Med. 1973 Apr 5;288(14):704-6.

8. Ranganathan V, Gracey E, Brown MA, Inman RD, Haroon N. Pathogenesis of ankylosing spondylitis - recent advances and future directions. Nat Rev Rheumatol. 2017 Jun;13(6):359-67.

9. International Genetics of Ankylosing Spondylitis Consortium (IGAS), Cortes A, Hadler J, Pointon JP, Robinson PC, Karaderi T, et al. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. Nat Genet. 2013 Jul;45(7):730-8.

10. Chen L, Ridley A, Hammitzsch A, Al-Mossawi MH, Bunting H, Georgiadis D, et al. Silencing or inhibition of endoplasmic reticulum aminopeptidase 1 (ERAP1) suppresses free heavy chain expression and Th17 responses in ankylosing spondylitis. Ann Rheum Dis. 2016 May;75(5):916-23.

11. Garcia-Medel N, Sanz-Bravo A, Van Nguyen D, Galocha B, Gomez-Molina P, Martin-Esteban A, et al. Functional interaction of the ankylosing spondylitis-associated endoplasmic reticulum aminopeptidase 1 polymorphism and HLA-B27 in vivo. Mol Cell Proteomics. 2012 Nov;11(11):1416-29.

12. Haroon N, Tsui FW, Uchanska-Ziegler B, Ziegler A, Inman RD. Endoplasmic reticulum aminopeptidase 1 (ERAP1) exhibits functionally significant interaction with

HLA-B27 and relates to subtype specificity in ankylosing spondylitis. Ann Rheum Dis. 2012 Apr;71(4):589-95.

13. Tran TM, Hong S, Edwan JH, Colbert RA. ERAP1 reduces accumulation of aberrant and disulfide-linked forms of HLA-B27 on the cell surface. Mol Immunol. 2016 Jun;74:10-7.

14. Tsui FW, Haroon N, Reveille JD, Rahman P, Chiu B, Tsui HW, et al. Association of an ERAP1 ERAP2 haplotype with familial ankylosing spondylitis. Ann Rheum Dis. 2010 Apr;69(4):733-6.

15. O'Rielly DD, Uddin M, Codner D, Hayley M, Zhou J, Pena-Castillo L, et al. Private rare deletions in SEC16A and MAMDC4 may represent novel pathogenic variants in familial axial spondyloarthritis. Ann Rheum Dis. 2015 May 8.

16. O'Rielly DD, Rahman P. Advances in the genetics of spondyloarthritis and clinical implications. Curr Rheumatol Rep. 2013 Aug;15(8):347,013-0347-x.

17. Lin Z, Bei JX, Shen M, Li Q, Liao Z, Zhang Y, et al. A genome-wide association study in Han Chinese identifies new susceptibility loci for ankylosing spondylitis. Nat Genet. 2011 Dec 4;44(1):73-7.

18. Lories RJ, de Vlam K. Is psoriatic arthritis a result of abnormalities in acquired or innate immunity? Curr Rheumatol Rep. 2012 Aug;14(4):375-82.

19. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet. 2011 Jul 10;43(8):761-7.

20. Ryan C, Abramson A, Patel M, Menter A. Current investigational drugs in psoriasis. Expert Opin Investig Drugs. 2012 Apr;21(4):473-87.

21. Liang SC, Tan XY, Luxenberg DP, Karim R, Dunussi-Joannopoulos K, Collins M, et al. Interleukin (IL)-22 and IL-17 are coexpressed by Th17 cells and cooperatively enhance expression of antimicrobial peptides. J Exp Med. 2006 Oct 2;203(10):2271-9.

22. Jethwa H, Abraham S. The evidence for microbiome manipulation in inflammatory arthritis. Rheumatology (Oxford). 2016 Oct 27.

23. Martinez-Gonzalez O, Cantero-Hinojosa J, Paule-Sastre P, Gomez-Magan JC, Salvatierra-Rios D. Intestinal permeability in patients with ankylosing spondylitis and their healthy relatives. Br J Rheumatol. 1994 Jul;33(7):644-7.

24. Tito RY, Cypers H, Joossens M, Varkas G, Van Praet L, Glorieus E, et al. Brief Report: Dialister as a Microbial Marker of Disease Activity in Spondyloarthritis. Arthritis Rheumatol. 2017 Jan;69(1):114-21.

25. Taurog JD, Richardson JA, Croft JT, Simmons WA, Zhou M, Fernandez-Sueiro JL, et al. The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats. J Exp Med. 1994 Dec 1;180(6):2359-64.

26. Hammer RE, Maika SD, Richardson JA, Tang JP, Taurog JD. Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human beta 2m: an animal model of HLA-B27-associated human disorders. Cell. 1990 Nov 30;63(5):1099-112.

27. Sherlock JP, Joyce-Shaikh B, Turner SP, Chao CC, Sathe M, Grein J, et al. IL-23 induces spondyloarthropathy by acting on ROR-gammat+ CD3+CD4-CD8- entheseal resident T cells. Nat Med. 2012 Jul 1;18(7):1069-76.

28. Jacques P, Lambrecht S, Verheugen E, Pauwels E, Kollias G, Armaka M, et al. Proof of concept: enthesitis and new bone formation in spondyloarthritis are driven by mechanical strain and stromal cells. Ann Rheum Dis. 2014 Feb;73(2):437-45.

29. Galea GL, Sunters A, Meakin LB, Zaman G, Sugiyama T, Lanyon LE, et al. Sost down-regulation by mechanical strain in human osteoblastic cells involves PGE2 signaling via EP4. FEBS Lett. 2011 Aug 4;585(15):2450-4.

30. Tu X, Rhee Y, Condon KW, Bivi N, Allen MR, Dwyer D, et al. Sost downregulation and local Wnt signaling are required for the osteogenic response to mechanical loading. Bone. 2012 Jan;50(1):209-17.

31. Pedersen OB, Svendsen AJ, Ejstrup L, Skytthe A, Harris JR, Junker P. Ankylosing spondylitis in Danish and Norwegian twins: occurrence and the relative importance of genetic vs. environmental effectors in disease causation. Scand J Rheumatol. 2008 Mar-Apr;37(2):120-6.

32. Brown MA, Kennedy LG, MacGregor AJ, Darke C, Duncan E, Shatford JL, et al. Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment. Arthritis Rheum. 1997 Oct;40(10):1823-8.

33. Brown MA, Kenna T, Wordsworth BP. Genetics of ankylosing spondylitis-insights into pathogenesis. Nat Rev Rheumatol. 2015 Oct 6.

34. Brown MA, Laval SH, Brophy S, Calin A. Recurrence risk modelling of the genetic susceptibility to ankylosing spondylitis. Ann Rheum Dis. 2000 Nov;59(11):883-6.

35. Rudwaleit M, Khan MA, Sieper J. The challenge of diagnosis and classification in early ankylosing spondylitis: do we need new criteria? Arthritis Rheum. 2005 Apr;52(4):1000-8.

36. Khan MA. Polymorphism of HLA-B27: 105 subtypes currently known. Curr Rheumatol Rep. 2013 Oct;15(10):362,013-0362-y.

37. Colbert RA, Tran TM, Layh-Schmitt G. HLA-B27 misfolding and ankylosing spondylitis. Mol Immunol. 2014 Jan;57(1):44-51.

38. Allen RL, O'Callaghan CA, McMichael AJ, Bowness P. Cutting edge: HLA-B27 can form a novel beta 2-microglobulin-free heavy chain homodimer structure. J Immunol. 1999 May 1;162(9):5045-8.

39. Taurog JD. The role of HLA-B27 in spondyloarthritis. J Rheumatol. 2010 Dec;37(12):2606-16.

40. Kollnberger S, Bird L, Sun MY, Retiere C, Braud VM, McMichael A, et al. Cell-surface expression and immune receptor recognition of HLA-B27 homodimers. Arthritis Rheum. 2002 Nov;46(11):2972-82.

41. Bird LA, Peh CA, Kollnberger S, Elliott T, McMichael AJ, Bowness P. Lymphoblastoid cells express HLA-B27 homodimers both intracellularly and at the cell surface following endosomal recycling. Eur J Immunol. 2003 Mar;33(3):748-59.

42. Kollnberger S, Bowness P. The role of B27 heavy chain dimer immune receptor interactions in spondyloarthritis. Adv Exp Med Biol. 2009;649:277-85.

43. Reveille JD. The genetic basis of spondyloarthritis. Ann Rheum Dis. 2011 Mar;70 Suppl 1:i44-50.

44. Cortes A, Pulit SL, Leo PJ, Pointon JJ, Robinson PC, Weisman MH, et al. Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with ERAP1. Nat Commun. 2015 May 21;6:7146.

45. Brown MA, Pile KD, Kennedy LG, Calin A, Darke C, Bell J, et al. HLA class I associations of ankylosing spondylitis in the white population in the United Kingdom. Ann Rheum Dis. 1996 Apr;55(4):268-70.

46. Chandran V, Bull SB, Pellett FJ, Ayearst R, Rahman P, Gladman DD. Human leukocyte antigen alleles and susceptibility to psoriatic arthritis. Hum Immunol. 2013 Oct;74(10):1333-8.

47. Gudjonsson JE, Karason A, Runarsdottir EH, Antonsdottir AA, Hauksson VB, Jonsson HH, et al. Distinct clinical differences between HLA-Cw*0602 positive and

negative psoriasis patients--an analysis of 1019 HLA-C- and HLA-B-typed patients. J Invest Dermatol. 2006 Apr;126(4):740-5.

48. Nair RP, Stuart PE, Nistor I, Hiremagalore R, Chia NV, Jenisch S, et al. Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. Am J Hum Genet. 2006 May;78(5):827-51.

49. Hreggvidsdottir HS, Noordenbos T, Baeten DL. Inflammatory pathways in spondyloarthritis. Mol Immunol. 2014 Jan;57(1):28-37.

50. O'Rielly DD, Uddin M, Rahman P. Ankylosing spondylitis: beyond genome-wide association studies. Curr Opin Rheumatol. 2016 Jul;28(4):337-45.

51. Jung SH, Yim SH, Hu HJ, Lee KH, Lee JH, Sheen DH, et al. Genome-wide copy number variation analysis identifies deletion variants associated with ankylosing spondylitis. Arthritis Rheumatol. 2014 Aug;66(8):2103-12.

52. Uddin M, Maksymowych WP, Inman R, Gladman D, Munn A, Yazdani R, et al. UGT2B17 copy number gain in a large ankylosing spondylitis multiplex family. BMC Genet. 2013 Aug 8;14:67,2156-14-67.

53. Wang L, Yang X, Cai G, Xin L, Xia Q, Zhang X, et al. Association study of copy number variants in FCGR3A and FCGR3B gene with risk of ankylosing spondylitis in a Chinese population. Rheumatol Int. 2016 Mar;36(3):437-42.

54. Duan R, Leo P, Bradbury L, Brown MA, Thomas G. Gene expression profiling reveals a downregulation in immune-associated genes in patients with AS. Ann Rheum Dis. 2010 Sep;69(9):1724-9.

55. Smith JA, Barnes MD, Hong D, DeLay ML, Inman RD, Colbert RA. Gene expression analysis of macrophages derived from ankylosing spondylitis patients reveals interferon-gamma dysregulation. Arthritis Rheum. 2008 Jun;58(6):1640-9.

56. Assassi S, Reveille JD, Arnett FC, Weisman MH, Ward MM, Agarwal SK, et al. Whole-blood gene expression profiling in ankylosing spondylitis shows upregulation of toll-like receptor 4 and 5. J Rheumatol. 2011 Jan;38(1):87-98.

57. Xu L, Sun Q, Jiang S, Li J, He C, Xu W. Changes in gene expression profiles of the hip joint ligament of patients with ankylosing spondylitis revealed by DNA chip. Clin Rheumatol. 2012 Oct;31(10):1479-91.

58. Fang F, Pan J, Xu L, Li G, Wang J. Identification of potential transcriptomic markers in developing ankylosing spondylitis: a meta-analysis of gene expression profiles. Biomed Res Int. 2015;2015:826316.

59. Lee YH, Song GG. Meta-analysis of differentially expressed genes in ankylosing spondylitis. Genet Mol Res. 2015 May 18;14(2):5161-70.

60. Laval SH, Timms A, Edwards S, Bradbury L, Brophy S, Milicic A, et al. Whole-genome screening in ankylosing spondylitis: evidence of non-MHC genetic-susceptibility loci. Am J Hum Genet. 2001 Apr;68(4):918-26.

61. Zhang G, Luo J, Bruckel J, Weisman MA, Schumacher HR, Khan MA, et al. Genetic studies in familial ankylosing spondylitis susceptibility. Arthritis Rheum. 2004 Jul;50(7):2246-54.

62. Chandran V, Rahman P. Update on the genetics of spondyloarthritis--ankylosing spondylitis and psoriatic arthritis. Best Pract Res Clin Rheumatol. 2010 Oct;24(5):579-88.

63. Chang SC, Momburg F, Bhutani N, Goldberg AL. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism. Proc Natl Acad Sci U S A. 2005 Nov 22;102(47):17107-12.

64. York IA, Chang SC, Saric T, Keys JA, Favreau JM, Goldberg AL, et al. The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues. Nat Immunol. 2002 Dec;3(12):1177-84.

65. Birtley JR, Saridakis E, Stratikos E, Mavridis IM. The crystal structure of human endoplasmic reticulum aminopeptidase 2 reveals the atomic basis for distinct roles in antigen processing. Biochemistry. 2012 Jan 10;51(1):286-95.

66. Tsoi LC, Spain SL, Knight J, Ellinghaus E, Stuart PE, Capon F, et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat Genet. 2012 Dec;44(12):1341-8.

67. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 2015 Sep;47(9):979-86.

68. Adamopoulos IE, Tessmer M, Chao CC, Adda S, Gorman D, Petro M, et al. IL-23 is critical for induction of arthritis, osteoclast formation, and maintenance of bone mass. J Immunol. 2011 Jul 15;187(2):951-9.

69. Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2, Strange A, Capon F, Spencer CC, Knight J, Weale ME, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat Genet. 2010 Nov;42(11):985-90.

70. Liu Y, Helms C, Liao W, Zaba LC, Duan S, Gardner J, et al. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. PLoS Genet. 2008 Mar 28;4(3):e1000041.

71. O'Rielly DD, Rahman P. Genetics of psoriatic arthritis. Best Pract Res Clin Rheumatol. 2014 Oct;28(5):673-85.

72. Robinson PC, Claushuis TA, Cortes A, Martin TM, Evans DM, Leo P, et al. Genetic dissection of acute anterior uveitis reveals similarities and differences in associations observed with ankylosing spondylitis. Arthritis Rheumatol. 2015 Jan;67(1):140-51.

73. Lee E, Trepicchio WL, Oestreicher JL, Pittman D, Wang F, Chamian F, et al. Increased expression of interleukin 23 p19 and p40 in lesional skin of patients with psoriasis vulgaris. J Exp Med. 2004 Jan 5;199(1):125-30.

74. Sonder SU, Saret S, Tang W, Sturdevant DE, Porcella SF, Siebenlist U. IL-17-induced NF-kappaB activation via CIKS/Act1: physiologic significance and signaling mechanisms. J Biol Chem. 2011 Apr 15;286(15):12881-90.

75. Huffmeier U, Uebe S, Ekici AB, Bowes J, Giardina E, Korendowych E, et al. Common variants at TRAF3IP2 are associated with susceptibility to psoriatic arthritis and psoriasis. Nat Genet. 2010 Nov;42(11):996-9.

76. TNGAIP3: Genetics Home Reference [Internet]. Bethesda (MD): National Library of Medicine (US); 2017 [updated April 2017; cited April 17th 2017]. Available from: https://ghr.nlm.nih.gov/gene/TNFAIP3#.

77. Vereecke L, Beyaert R, van Loo G. The ubiquitin-editing enzyme A20 (TNFAIP3) is a central regulator of immunopathology. Trends Immunol. 2009 Aug;30(8):383-91.

78. Stuart PE, Nair RP, Tsoi LC, Tejasvi T, Das S, Kang HM, et al. Genome-wide Association Analysis of Psoriatic Arthritis and Cutaneous Psoriasis Reveals Differences in Their Genetic Architecture. Am J Hum Genet. 2015 Dec 3;97(6):816-36.

79. Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D, et al. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. Nat Genet. 2009 Feb;41(2):199-204.

80. Stuart PE, Nair RP, Ellinghaus E, Ding J, Tejasvi T, Gudjonsson JE, et al. Genome-wide association analysis identifies three psoriasis susceptibility loci. Nat Genet. 2010 Nov;42(11):1000-4.

81. LeibundGut-Landmann S, Gross O, Robinson MJ, Osorio F, Slack EC, Tsoni SV, et al. Syk- and CARD9-dependent coupling of innate immunity to the induction of T helper cells that produce interleukin 17. Nat Immunol. 2007 Jun;8(6):630-8.

82. McGovern D, Kugathasan S, Cho JH. Genetics of Inflammatory Bowel Diseases. Gastroenterology. 2015 Aug 7.

83. NOD2: Genetics Home Reference [Internet]. Bethesda (MD): National Library of Medicine (US); 2012 [updated July 2012; cited April 17th 2017]. Available from: https://ghr.nlm.nih.gov/gene/NOD2.

84. Cooney R, Baker J, Brain O, Danis B, Pichulik T, Allan P, et al. NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. Nat Med. 2010 Jan;16(1):90-7.

85. ATG16L1: Genetics Home Reference [Internet]. Bethesda (MD): National Library of Medicine (US); 2017 [updated April 2017; cited April 17th 2017]. Available from: https://ghr.nlm.nih.gov/gene/ATG16L1.

86. Baeten D, Breban M, Lories R, Schett G, Sieper J. Are spondylarthritides related but distinct conditions or a single disease with a heterogeneous phenotype? Arthritis Rheum. 2013 Jan;65(1):12-20.

87. Uddin M, Codner D, Hasan SM, Scherer SW, O'Rielly DD, Rahman P. Integrated genomics identifies convergence of ankylosing spondylitis with global immune mediated disease pathways. Sci Rep. 2015 May 18;5:10314.

88. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet. 2016 Mar 14.

89. Calin A, Porta J, Fries JF, Schurman DJ. Clinical history as a screening test for ankylosing spondylitis. JAMA. 1977 Jun 13;237(24):2613-4.

90. Rudwaleit M, van der Heijde D, Landewe R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis. 2009 Jun;68(6):777-83.

91. Rudwaleit M, Landewe R, van der Heijde D, Listing J, Brandt J, Braun J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. Ann Rheum Dis. 2009 Jun;68(6):770-6.

92. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum. 1984 Apr;27(4):361-8.

93. Sieper J, van der Heijde D. Review: Nonradiographic axial spondyloarthritis: new definition of an old disease? Arthritis Rheum. 2013 Mar;65(3):543-51.

94. Aydin SZ, Maksymowych WP, Bennett AN, McGonagle D, Emery P, Marzo-Ortega H. Validation of the ASAS criteria and definition of a positive MRI of the sacroiliac joint in an inception cohort of axial spondyloarthritis followed up for 8 years. Ann Rheum Dis. 2012 Jan;71(1):56-60.

95. Weber U, Zubler V, Pedersen SJ, Rufibach K, Lambert RG, Chan SM, et al. Development and validation of a magnetic resonance imaging reference criterion for defining a positive sacroiliac joint magnetic resonance imaging finding in spondyloarthritis. Arthritis Care Res (Hoboken). 2013 Jun;65(6):977-85.

96. Spondyloarthritis [Internet].; 2013 []. Available from: www.rheumatology.org/I-Am-A/Patient-Caregiver/Diseases-Conditions/Spondyloarthritis.

97. van den Berg R, Baraliakos X, Braun J, van der Heijde D. First update of the current evidence for the management of ankylosing spondylitis with non-pharmacological treatment and non-biologic drugs: a systematic literature review for the ASAS/EULAR management recommendations in ankylosing spondylitis. Rheumatology (Oxford). 2012 Aug;51(8):1388-96.

98. Yurtkuran M, Ay A, Karakoc Y. Improvement of the clinical outcome in Ankylosing spondylitis by balneotherapy. Joint Bone Spine. 2005 Jul;72(4):303-8.

99. Spondylitis Association of America Medications Page [Internet].; 2017 []. Available from: http://www.spondylitis.org/Medications.

100. McInnes IB, Mease PJ, Kirkham B, Kavanaugh A, Ritchlin CT, Rahman P, et al. Secukinumab, a human anti-interleukin-17A monoclonal antibody, in patients with psoriatic arthritis (FUTURE 2): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet. 2015 Sep 19;386(9999):1137-46.

101. Vander Cruyssen B, Ribbens C, Boonen A, Mielants H, de Vlam K, Lenaerts J, et al. The epidemiology of ankylosing spondylitis and the commencement of anti-TNF therapy in daily rheumatology practice. Ann Rheum Dis. 2007 Aug;66(8):1072-7.

102. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012 Nov 1;491(7422):119-24.

103. O'Rielly DD, Rahman P. Genetics of psoriatic arthritis. Best Pract Res Clin Rheumatol. 2014 Oct;28(5):673-85.

104. Ellinghaus E, Stuart PE, Ellinghaus D, Nair RP, Debrus S, Raelson JV, et al. Genome-wide meta-analysis of psoriatic arthritis identifies susceptibility locus at REL. J Invest Dermatol. 2012 Apr;132(4):1133-40.

105. Xavier RJ, Podolsky DK. Unravelling the pathogenesis of inflammatory bowel disease. Nature. 2007 Jul 26;448(7152):427-34.

106. Uniken Venema WT, Voskuil MD, Dijkstra G, Weersma RK, Festen EA. The genetic background of inflammatory bowel disease: from correlation to causality. J Pathol. 2017 Jan;241(2):146-58.

107. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology. 2012 Jan;142(1):46,54.e42; quiz e30.

108. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet. 2011 Oct 9;43(11):1066-73.

109. Beaudoin M, Goyette P, Boucher G, Lo KS, Rivas MA, Stevens C, et al. Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. PLoS Genet. 2013;9(9):e1003723.

110. Linssen A, Rothova A, Valkenburg HA, Dekker-Saeys AJ, Luyendijk L, Kijlstra A, et al. The lifetime cumulative incidence of acute anterior uveitis in a normal population and its relation to ankylosing spondylitis and histocompatibility antigen HLA-B27. Invest Ophthalmol Vis Sci. 1991 Aug;32(9):2568-78.

111. Wakefield D, Yates W, Amjadi S, McCluskey P. HLA-B27 Anterior Uveitis: Immunology and Immunopathology. Ocul Immunol Inflamm. 2016 Aug;24(4):450-9.

112. Martin TM, Rosenbaum JT. An update on the genetics of HLA B27-associated acute anterior uveitis. Ocul Immunol Inflamm. 2011 Apr;19(2):108-14.

113. Brewerton DA, Caffrey M, Nicholls A, Walters D, James DC. Acute anterior uveitis and HL-A 27. Lancet. 1973 Nov 3;302(7836):994-6.

114. Robinson PC, Leo PJ, Pointon JJ, Harris J, Cremin K, Bradbury LA, et al. The genetic associations of acute anterior uveitis and their overlap with the genetics of ankylosing spondylitis. Genes Immun. 2016 Jan-Feb;17(1):46-51.

115. Hoy D, Brooks P, Blyth F, Buchbinder R. The Epidemiology of low back pain. Best Pract Res Clin Rheumatol. 2010 Dec;24(6):769-81.

116. Steenstra IA, Verbeek JH, Heymans MW, Bongers PM. Prognostic factors for duration of sick leave in patients sick listed with acute low back pain: a systematic review of the literature. Occup Environ Med. 2005 Dec;62(12):851-60.

117. Kent PM, Keating JL. The epidemiology of low back pain in primary care. Chiropr Osteopat. 2005 Jul 26;13:13.

118. Thelin A, Holmberg S, Thelin N. Functioning in neck and low back pain from a 12-year perspective: a prospective population-based study. J Rehabil Med. 2008 Jul;40(7):555-61.

119. Andersson GB. Epidemiological features of chronic low-back pain. Lancet. 1999 Aug 14;354(9178):581-5.

120. GBD 2013 DALYs and HALE Collaborators, Murray CJ, Barber RM, Foreman KJ, Abbasoglu Ozgoren A, Abd-Allah F, et al. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990-2013: quantifying the epidemiological transition. Lancet. 2015 Nov 28;386(10009):2145-91.

121. World Health Organization. Priority Medicines for Europe and the World Update Report, 2013. www.who.int: World Health Organization; 2013.

122. Arthritis in Canada [Internet].; 2013 []. Available from: http://arthritis.ca/getmedia/870886f4-602c-4589-a584-0d582d962706/arthritis-in-Canada-2013.pdf.

123. Cassidy JD, Cote P, Carroll LJ, Kristman V. Incidence and course of low back pain episodes in the general population. Spine (Phila Pa 1976). 2005 Dec 15;30(24):2817-23.

124. Mustard CA, Kalcevich C, Frank JW, Boyle M. Childhood and early adult predictors of risk of incident back pain: Ontario Child Health Study 2001 follow-up. Am J Epidemiol. 2005 Oct 15;162(8):779-86.

125. The Burden of Muscloskeletal Disease in the United States 2014 Report [Internet].; 2014 []. Available from: http://www.boneandjointburden.org/2014-report.

126. Rubin DI. Epidemiology and risk factors for spine pain. Neurol Clin. 2007 May;25(2):353-71.

127. Katz JN. Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. J Bone Joint Surg Am. 2006 Apr;88 Suppl 2:21-4.

128. Feldtkeller E, Khan MA, van der Heijde D, van der Linden S, Braun J. Age at disease onset and diagnosis delay in HLA-B27 negative vs. positive patients with ankylosing spondylitis. Rheumatol Int. 2003 Mar;23(2):61-6.

129. Chen H, Poon A, Yeung C, Helms C, Pons J, Bowcock AM, et al. A genetic risk score combining ten psoriasis risk loci improves disease prediction. PLoS One. 2011 Apr 29;6(4):e19454.

130. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med. 2008 Nov 20;359(21):2208-19.

131. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, et al. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. PLoS Med. 2006 Oct;3(10):e374.

132. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res. 2007 Oct;17(10):1520-8.

133. Paynter NP, Chasman DI, Pare G, Buring JE, Cook NR, Miletich JP, et al. Association between a literature-based genetic risk score and cardiovascular events in women. JAMA. 2010 Feb 17;303(7):631-7.

134. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. Lancet. 2010 Oct 23;376(9750):1393-400.

135. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016 Sep 29;375(13):1216-9.

136. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nat Rev Genet. 2015 Jun;16(6):321-32.

137. van Gaalen FA, Verduijn W, Roelen DL, Bohringer S, Huizinga TW, van der Heijde DM, et al. Epistasis between two HLA antigens defines a subset of individuals at a very high risk for ankylosing spondylitis. Ann Rheum Dis. 2013 Jun;72(6):974-8.

138. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet. 2006 Oct;38(10):1166-72.

139. Robinson PC, Claushuis TA, Cortes A, Martin TM, Evans DM, Leo P, et al. Genetic dissection of acute anterior uveitis reveals similarities and differences in associations observed with ankylosing spondylitis. Arthritis Rheumatol. 2015 Jan;67(1):140-51.

140. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012 Jan;40(Database issue):D930-4.

141. Chow SL, Carter Thorne J, Bell MJ, Ferrari R, Bagheri Z, Boyd T, et al. Choosing wisely: the Canadian Rheumatology Association's list of 5 items physicians and patients should question. J Rheumatol. 2015 Apr;42(4):682-9.

142. Lehr J, Rahman P, O'Rielly DD. High Accuracy and Significant Savings Using Tag-SNP Genotyping to Determine HLA-B*27 Status. J Rheumatol. 2017 Jun;44(6):962-3.

143. Liu Y, Gadepalli K, Norouzi M, Dahl GE, Kohlberger T, Boyko A, et al. Detecting Cancer Metastases on Gigapixel Pathology Images. Cornell University Library. 2017;n/a(Google Publication):1-13.

144. Rudwaleit M, Khan MA, Sieper J. The challenge of diagnosis and classification in early ankylosing spondylitis: do we need new criteria? Arthritis Rheum. 2005 Apr;52(4):1000-8.

145. Thomas GP, Willner D, Robinson PC, Cortes A, Duan R, Rudwaleit M, et al. Genetic diagnostic profiling in axial spondyloarthritis: a real world study. Clin Exp Rheumatol. 2017 Mar-Apr;35(2):229-33.

146. Developing a Risk-scoring Model for Ankylosing Spondylitis Based on a Combination of HLA-B27, Single-nucleotide Polymorphism, and Copy Number Variant Markers. J Rheumatol. 2017 Jan;44(1):132.

147. Kobelt G, Andlin-Sobocki P, Maksymowych WP. Costs and quality of life of patients with ankylosing spondylitis in Canada. J Rheumatol. 2006 Feb;33(2):289-95.

148. Bonter K, Desjardins C, Currier N, Pun J, Ashbury FD. Personalised medicine in Canada: a survey of adoption and practice in oncology, cardiology and family medicine. BMJ Open. 2011 Jul 29;1(1):e000110,2011-000110.

Appendix

Appendix 1.A: Design Summary Report for LBP_1W well.

This is the Design Summary Report generated by ADS. See Methods section 2.3.1c for specific details.

WELL	TERM	SNP_ID	2nd-PCRP	1st-PCRP	AMP_LEN	UP_CONF	MP_CONF	Tm(NN)	PcGC
W1	iPLEX	rs6759298	ACGTTGGATGAGTTGCAGGCTATTGGTGTC	ACGTTGGATGCTTTGTGGTGGTTCTGTAGG	125	97.5	82.8	49.8	52.9
W1	iPLEX	rs10782001	ACGTTGGATGACACAGTTATCTGCTCCCAC	ACGTTGGATGTGTTCCCCTCATAGAGCAAG	128	96.9	82.8	46.5	41.2
W1	iPLEX	rs6871626	ACGTTGGATGGCAGAGAAAGTTACCTGTCC	ACGTTGGATGCATTATGGGCTAAGTGGGTG	94	98.6	82.8	47.5	50
W1	iPLEX	rs1265163	ACGTTGGATGTAACCTGACAGGTGTTCTCG	ACGTTGGATGAGAAACTGGCACATCCAAGG	103	100	82.8	46.9	42.1
W1	iPLEX	rs2032890	ACGTTGGATGTAAAGACCCAGTGGTGGGAG	ACGTTGGATGCATCCTGGCGAAACTCCTTG	120	95.6	82.8	48.9	47.4
W1	iPLEX	rs2910686	ACGTTGGATGAACTTAAATCCCAGCTCACC	ACGTTGGATGACAAGTGACCACAATGTGGC	99	98.6	82.8	50.6	45
W1	iPLEX	rs11209026	ACGTTGGATGGAAATTCTGCAAAAACCTAC	ACGTTGGATGGGGAATGATCGTCTTTGCTG	115	87	82.8	51.3	45
W1	iPLEX	rs33980500	ACGTTGGATGCTGGGATTGGTTTCAGCAAC	ACGTTGGATGTGAACCGAAGCATTCCTGTG	92	99.7	82.8	49.1	40
W1	iPLEX	rs582757	ACGTTGGATGTAGCCTCATGTGGAATAAGC	ACGTTGGATGATAAGGCTACCAAGGCCTAC	119	97.1	82.8	47.8	31.8
W1	iPLEX	rs10781500	ACGTTGGATGTCTCTAACCATATCGGAAGC	ACGTTGGATGATCTGTGGGTTATTTAGCGG	118	95.9	82.8	46.5	31.8
W1	iPLEX	rs3132528	ACGTTGGATGAGCCTTATCTTGACCTGTTC	ACGTTGGATGCCATTTTAAAAACTTGGGCTC	104	88.8	82.8	53.7	41.7
W1	iPLEX	rs6738490	ACGTTGGATGGAGAACTACTGATTTTGCAC	ACGTTGGATGGTAAACCTGACGACTTTCTC	119	91.7	82.8	53.3	37.5

W1	iPLEX	rs116488202	ACGTTGGATGACCAAGCCTCAGACCATGC	ACGTTGGATGCCCGCACCAAATTCAGTACA	112	90.5	82.8	61.8	53.8
W1	iPLEX	rs2066844	ACGTTGGATGAGTGCCAGACATCTGAGAAG	ACGTTGGATGATGGAGTGGAAGTGCTTGCG	111	98.1	82.8	63.1	61.5
W1	iPLEX	rs10456057	ACGTTGGATGGGCACTGCAATATTGAGTTC	ACGTTGGATGTGTTTTCAGAGGTTCTGGAC	103	97.1	82.8	48.8	30.8

Appendix 1.A Continued. Table could not fit on one page this is the table continued.

PWARN	UEP_DIR	UEP_MASS	UEP_SEQ	EXT1_CALL	EXT1_MASS	EXT1_SEQ	EXT2_CALL	EXT2_MASS	EXT2_SEQ
d	F	5090.3	TCTTCCAACACAGTGCC	C	5337.5	TCTTCCAACACAGTGCCC	G	5377.5	TCTTCCAACACAGTGCCG
d	R	5218.4	ATGAAGGCTTGTCAACA	G	5465.6	ATGAAGGCTTGTCAACAC	A	5545.5	ATGAAGGCTTGTCAACAT
d	R	5416.5	CTGTCCTTCATCACTTGG	C	5703.7	CTGTCCTTCATCACTTGGG	A	5743.6	CTGTCCTTCATCACTTGGT
	R	5646.7	TCTCTTTCTGTCCTTTCAC	G	5893.9	TCTCTTTCTGTCCTTTCACC	C	5933.9	TCTCTTTCTGTCCTTTCACG
	F	5836.8	GAGAAACCTGATCCGGTAT	C	6084	GAGAAACCTGATCCGGTATC	A	6108	GAGAAACCTGATCCGGTATA
	F	5980.9	AATCCCAGCTCACCATTTAC	C	6228.1	AATCCCAGCTCACCATTTACC	T	6308	AATCCCAGCTCACCATTTACT
d	R	6030	CTGCAAAAACCTACCCAGTT	G	6277.1	CTGCAAAAACCTACCCAGTTC	A	6357	CTGCAAAAACCTACCCAGTTT
h	F	6169	TGGGTATGGTTCTGATTCAT	C	6416.2	TGGGTATGGTTCTGATTCATC	T	6496.1	TGGGTATGGTTCTGATTCATT
Dh	F	6641.3	CTGCATTTTTATCCTTTTAGCA	C	6888.5	CTGCATTTTTATCCTTTTAGCAC	T	6968.4	CTGCATTTTTATCCTTTTAGCAT
	F	6775.5	GCTAAAAATCGGTAACAGATAT	C	7022.6	GCTAAAAATCGGTAACAGATATC	T	7102.5	GCTAAAAATCGGTAACAGATATT
	F	7231.7	CCTGTTCTATTAAAACCTGCCACA	C	7478.9	CCTGTTCTATTAAAACCTGCCACAC	T	7558.8	CCTGTTCTATTAAAACCTGCCACAT

dH	R	7335.8	ACTGATTTTGCACAATCAGAATGC	T	7607	ACTGATTTTGCACAATCAGAATGCA	C	7623	ACTGATTTTGCACAATCAGAATGCG
d	R	7851.1	TCAGACCATGCCCAGCCTAGCTTACT	T	8122.3	TCAGACCATGCCCAGCCTAGCTTACTA	C	8138.3	TCAGACCATGCCCAGCCTAGCTTACTG
dH	F	7941.2	GCCAGACATCTGAGAAGGCCCTGCTC	C	8188.3	GCCAGACATCTGAGAAGGCCCTGCTCC	T	8268.3	GCCAGACATCTGAGAAGGCCCTGCTCT
D	R	7977.2	CTGCAATATTGAGTTCATATAACAAG	G	8224.4	CTGCAATATTGAGTTCATATAACAAGC	A	8304.3	CTGCAATATTGAGTTCATATAACAAGT

Appendix 2.A: Design Summary Report for HLA well.

This is the Design Summary Report generated by ADS. See Methods section 2.3.1c for specific details.

WELL	TERM	SNP_ID	2nd-PCRP	1st-PCRP	AMP_LEN	UP_CONF	MP_CONF	Tm(NN)	PcGC
W1	iPLEX	rs887466	ACGTTGGATGAATCCTTCCTGACCTAGAGC	ACGTTGGATGTCCGCACCTATCACACCTAC	114	97.8	93	49.1	52.9
W1	iPLEX	rs2853931	ACGTTGGATGGCATAGAATATCATGCTGCAC	ACGTTGGATGACGCTCTTTTCAGGACGATG	86	93.8	93	46.1	42.1
W1	iPLEX	rs6457374	ACGTTGGATGCCTAACAGTATGACACTCG	ACGTTGGATGTTTCAAACCTCCTGCATCTG	102	96.4	93	49.3	45
W1	iPLEX	rs3129944	ACGTTGGATGTGTGCTTATAAGGTACCCAC	ACGTTGGATGCTGTGGAGAACAAGGAAGAG	105	98.5	93	45.5	27.3
W1	iPLEX	rs4349859	ACGTTGGATGAGAGAGCAGTCCTACAAATG	ACGTTGGATGAAGCAGCCTAATCCCCTTAC	133	94.4	93	47.4	30.4

Appendix 2.A Continued. Table could not fit on one page this is the table continued.

PWARN	UEP_DIR	UEP_MASS	UEP_SEQ	EXT1_CALL	EXT1_MASS	EXT1_SEQ	EXT2_CALL	EXT2_MASS	EXT2_SEQ
	R	5090.3	TCTACCCTCTCCGGAAA	G	5337.5	TCTACCCTCTCCGGAAAC	A	5417.4	TCTACCCTCTCCGGAAAT
	R	5869.9	CTGCACATGAAGAAATAGG	T	6141.1	CTGCACATGAAGAAATAGGA	C	6157.1	CTGCACATGAAGAAATAGGG
d	F	6212.1	ACCAGATAGGTTTAGTGGTG	C	6459.2	ACCAGATAGGTTTAGTGGTGC	T	6539.1	ACCAGATAGGTTTAGTGGTGT
	R	6728.4	AGTCAATAGACACTCAATAAAA	G	6975.6	AGTCAATAGACACTCAATAAAAC	C	7015.6	AGTCAATAGACACTCAATAAAAG
d	R	6945.5	TCTTACATGTCTTTGTACTTACT	G	7192.7	TCTTACATGTCTTTGTACTTACTC	A	7272.6	TCTTACATGTCTTTGTACTTACTT