GENOMIC AND FUNCTIONAL ANALYSES OF NEWFOUNDLAND AND LABRADOR FAMILIES WITH INTRACRANIAL ANEURYSMS

by © Julia Pennell

A Thesis submitted

to the School of Graduate Studies in partial fulfilment of the

requirements for the degree of

Master of Science in Medicine (Human Genetics)

Memorial University of Newfoundland

February 2021

St. John's Newfoundland and Labrador

Abstract

The integrity of an arterial wall is impacted by various genetic, environmental, and physiological factors that predispose to intracranial aneurysm (IA) formation in certain individuals. In this study, whole exome sequencing was completed on a large cohort of families from the Newfoundland and Labrador population to identify genetic predispositions contributing to IA development. With the use of Sanger sequencing, *HAL* (c.959C>T; p.T320M) emerged as a favourable candidate variant, and morpholino mediated gene knockdown of *hal* in *Danio rerio* (zebrafish) demonstrated a dose-dependent hemorrhagic phenotype. The re-expression of wildtype *HAL* RNA in the *hal* morpholino background was able to rescue the hemorrhagic phenotype while mutant *HAL* RNA was not. These results suggest that the *HAL* (c.959C>T; p.T320M) variant could be contributing to IA development in Family R1262. Additional analysis of the *HAL* gene could help determine the extent of its involvement and contribute to an improved understanding of IA pathogenesis.

General Summary

Intracranial aneurysms (IA) are frequently not detected as many individuals do not experience symptoms until the aneurysm ruptures. Ruptured aneurysms often have devastating clinical outcomes that may have been preventable if the aneurysm was discovered earlier. Here, we aimed to identify genetic components contributing to IA development in families from Newfoundland and Labrador. By DNA sequencing, we identified five candidate genes that potentially contributed to IA development. In order to help determine if any of these genes cause IA, we used a zebrafish model organism to study these genes as we cannot study them in humans directly. The loss of function of one gene in particular (*HAL)*, caused significant hemorrhages to occur in zebrafish embryos. Our results suggest that the *HAL* gene could be involved in IA development in a family from Newfoundland and Labrador. Additional studies are required to gain a better understanding of its potential contribution to IA development.

Acknowledgements

I would first like to thank The Heart and Stroke Foundation of Canada for funding this research. To my supervisor, Dr. Michael Woods, thank you for your continuous guidance and support; without you, none of this would have been possible. To my committee members, Dr. Ann Dorward and Dr. Curtis French, thank you for your invaluable insights throughout this entire process. I especially want to extend my sincerest gratitude to all members of the French Lab, who guided me through every step of the zebrafish work, I can't thank you enough.

To Amy Powell, whose previous work laid the foundation for this project, I am incredibly grateful. Thank you for helping me with this study design and teaching me everything I know about the lab. To Daniel Evans, I will forever be thankful for the guidance you provided me, but most importantly, for your friendship. To everyone in the Discipline of Genetics, especially Deborah Quinlan, I am extremely appreciative of all of your help along the way. Finally, I want to extend my gratitude to my amazing family. Thank you for always believing in me and reminding me what matters most.

Statement of Co-Authorship

Significant work was already completed prior to the beginning of this project. Participants were enrolled, clinical information was collected, and whole-exome sequencing was completed. Thank you to everyone who contributed to the project along the way, specifically Dr. Falah Maroun, Dr. Bridget Fernandez, Barbara Noble and Carol Negrin. Your contributions made it feasible for me to complete this project.

Amy Powell previously completed a more focused study, concentrating specifically on two families from the patient cohort. DNA samples had already been sent to The Centre for Applied Genomics (TCAG) at The Hospital for Sick Children in Toronto, Ontario, for sequencing. Thus, the exome data required for this project was readily available. Information regarding the whole-exome sequencing platform included in section 2.2 outlines steps that were completed by TCAG. Furthermore, Amy Powell significantly contributed to the study design and development of the variant filtering strategy used in this project. Her knowledge of our patient cohort and familiarization with sequencing analysis was extremely valuable and very much appreciated.

The functional work completed for this project was done in collaboration with Dr. Curtis French at Memorial University in the Faculty of Medicine. The morpholino-mediated gene knockdown and rescue experiments were planned and executed with the help of Dr. French. Furthermore, any required morpholinos that were not previously published were designed by Dr. French via GeneTools software (Gene Tools, LLC).

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAA	Abdominal Aortic Aneurysm
ABCC6	ATP Binding Cassette Subfamily C Member 6
ADAMTS15	A Disintegrin and Metalloproteinase with Thrombospondin Motifs 15
ADGRV1	Adhesion G Protein-Coupled Receptor V1
ADPKD	Autosomal Dominant Polycystic Kidney Disease
ANGPTL6	Angiopoietin Like 6
ARHGEF17	Rho Guanine Nucleotide Exchange Factor 17
bp	Base Pair
CADD	Combined Annotation Dependent Depletion
CCDC80	Coiled-Coil Domain Containing 80
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A
CDKN2B	Cyclin Dependent Kinase Inhibitor 2B
CDKN2B-AS	CDKN2B Antisense RNA 1
CG	Complete Genomics
Chr	Chromosome
COL3A1	Collagen Type III Alpha 1 Chain
COPD	Chronic Obstructive Pulmonary Disease
CRISPR/CAS9	Clustered, Regularly Interspaced, Short Palindromic Repeats
CT	Computed Tomography
CTA	Computed Tomography Angiography
dbSNP	Single Nucleotide Polymorphism Database
DM	Diabetes Mellitus
dNTP	Deoxyribonucleotide Triphosphate
DSA	Digital Subtraction Angiography
DSG1	Desmoglein 1
E-PAP	E.coli Poly(A) Polymerase
ELN	Elastin
ENG	Endoglin
EPKHE	Erythroderma, Congenital with Palmoplantar Keratoderma, Hypotrichosis and Hyper IgE
ExAC	Exome Aggregation Consortium
EXOSAP	Exonuclease Shrimp Alkaline Phosphatase
FARSA	Phenylalanyl-TRNA Synthetase Subunit Alpha
FBN1	Fibrillin 1
FHIT	Fragile Histidine Triad Diadenosine Triphosphatase
FIA	Familial Intracranial Aneurysm
FMD	Fibromuscular Dysplasia
FOXC1	Forkhead Box C1
FOXE3	Forkhead Box E3
g	Gram
GATK	The Genome Analysis Toolkit
GERD	Gastroesophageal Reflux Disease
GWAS	Genome-Wide Association Study
H_2O	Water

HAL	Histidine Ammonia-Lyase
HCl	Hydrogen Chloride
HPF	Hours Post Fertilization
IA	Intracranial Aneurysm
IBD	Identity by Descent
IGSF3	Immunoglobulin Superfamily Member 3
KCl	Potassium Chloride
L	Litre
LOD	Logarithm of The Odds
LOX	Lysyl Oxidase
LOXL2	Lysyl Oxidase Like 2
MA	MutationAssessor
MADD	MAP Kinase Activating Death Domain
MAF	Minor Allele Frequency
MAT2A	Methionine Adenosyltransferase 2A
MEN1	Menin 1
METTL20	Methyltransferase Like 20
mg	Milligram
$MgCl_2$	Magnesium Chloride
ml	Millilitre
mm	Millimetre
mM	Millimolar
MMP	Matrix Metalloproteinase
$MnCl_2$	Manganese Chloride
MO	Morpholino Oligonucleotide
MRA	Magnetic Resonance Angiography
MRI	Magnetic Resonance Imaging
MST1R	Macrophage Stimulating 1 Receptor
MT	MutationTaster
MTAP	Methylthioadenosine Phosphorylase
NF1	Neurofibromin 1
NFCCR	Newfoundland Colorectal Cancer Registry
ng	Nanogram
ng/nl	Nanogram per Nanoliter
ng/ul	Nanogram per Microliter
NGS	Next Generation Sequencing
NHBLI	National Heart Blood and Lung Institute
NL	Newfoundland and Labrador
nmol	Nanomole
nsSNV	Nonsynonymous Single Nucleotide Variant
OD	Optical Density
OMIM	Online Mendelian Inheritance in Man Databases
OR2T35	Olfactory Receptor Family 2 Subfamily T Member 35
ORF	Open Reading Frame
p	P Value
p53	Tumor Protein p53

PCR	Polymerase Chain Reaction
PHACTR1	Phosphatase and Actin Regulator 1
PhyloPMam	PhyloP Score for Placental Mammals
PhyloPVert100	PhyloP Score for 100 Vertebrates
PKD	Polycystic Kidney Disease
PKD1	Polycystin 1, Transient Receptor Potential Channel Interacting
PKD2	Polycystin 2, Transient Receptor Potential Cation Channel
PTPN13	Protein Tyrosine Phosphatase Non-Receptor Type 13
PTU	1-Phenyl 2-Thiourea
RNF213	Ring Finger Protein 213
RO	Reverse Osmosis
rpm	Revolutions per Minute
SAH	Subarachnoid Hemorrhage
SB	Strand Bias
SIFT	Sorting Intolerant from Tolerant
SMAD3	SMAD Family Member 3
SNP	Single Nucleotide Polymorphisms
SOX17	SRY-Box Transcription Factor 17
T1DM	Type 1 Diabetes Mellitus
TAA	Thoracic Aortic Aneurysm
TAAD	Thoracic Aortic Aneurysm and Dissection
TAE	Tris-Acetate-EDTA
TALEN	Transcription Activator–Like Effector Nucleases
TAS1R1	Taste 1 Receptor Member 1
TBC1D7	TBC1 Domain Family Member 7
TCAG	The Center for Applied Genomics
TGFBR1	Transforming Growth Factor-B Receptor 1
TGFBR2	Transforming Growth Factor-B Receptor 2
THSD1	Thrombospondin Type 1 Domain Containing 1
TMEM132B	Transmembrane Protein 132B
TNF-α	Tumor Necrosis Factor Alpha
TRPM1	Transient Receptor Potential Cation Channel Subfamily M Member 1
TTN	Titin
U/ul	Units per Microliter
UIATS	Unruptured IA Treatment Score
ul	Microliter
V	Volt
VSMC	Vascular Smooth Muscle Cells
VQSR	Variant Quality Score Recalibrations
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
x g	Times Gravity
ZNF835	Zinc Finger Protein 835

List of Appendices

1. Introduction

1.1 Intracranial Aneurysms

Intracranial aneurysms (IA) are thinned pouches of the arterial wall frequently found at the bifurcation of arteries in the Proximal Circle of Willis [1]. Unruptured IAs affect 3-5% of the general adult population, however, our knowledge of the exact steps in IA development has yet to be solidified [2,3]. There are various environmental, physiologic and genetic risk factors that predispose to IA in specific individuals, but IAs are approximately twice as likely to develop in females and are more common in patients over 50 years of age [3,4]. A major concern of IAs is that they are frequently asymptomatic and consequently not detected until they rupture [5]. A ruptured IA can lead to bleeding in the subarachnoid space, and these subarachnoid hemorrhages (SAH) are frequently fatal. During an initial SAH, there is a 30% death rate, and those that do survive often have severe neurological deficits as a result [1]. Vasospasm of cerebral vessels and a neuroinflammatory response, with the recruitment of cytokines and immune cells, have been associated with the significant morbidities post SAH [6–8]. Females experience a higher incidence and mortality rate of SAH, but the functional outcomes of those that survive is similar regardless of sex [9,10]. The vast majority of IAs that develop are considered saccular or berry-shaped IAs and thus are the focus of this thesis (Figure 1.1). Fusiform IAs, which involve dilation on all sides of a blood vessel and dissecting IAs, where a longitudinal tear in the vessel wall traps blood between layers also exist, but account for only 10% of the IAs that develop [11].

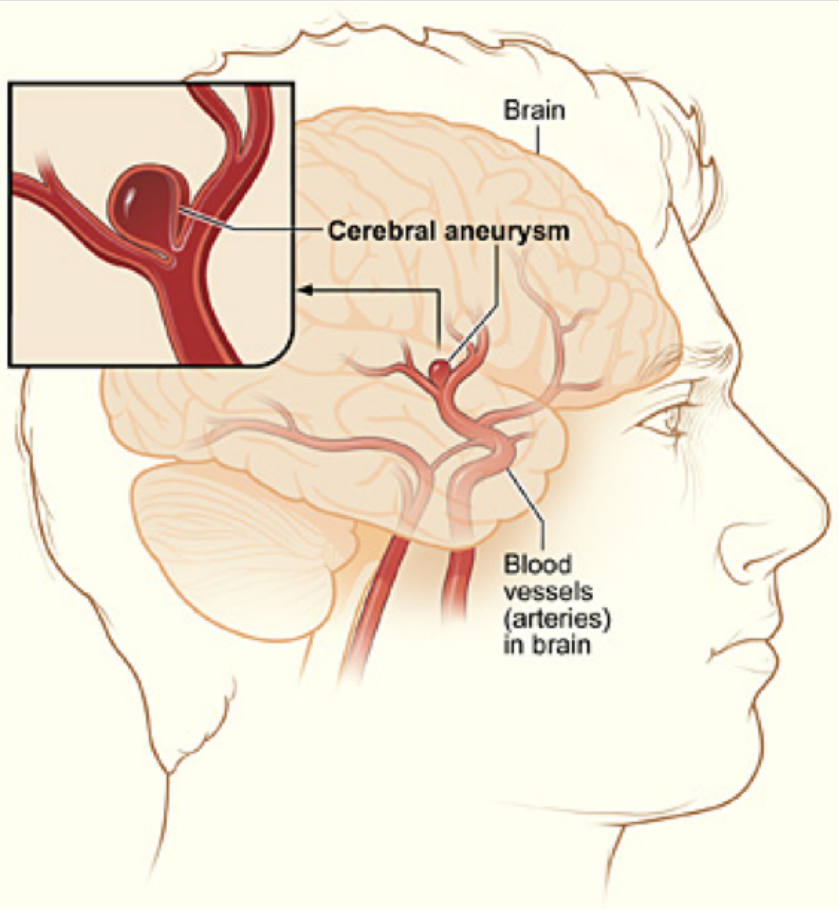


Figure 1.1 Saccular cerebral aneurysm

Adapted from:
https://upload.wikimedia.org/wikipedia/commons/8/80/Cerebral_aneurysm_NIH.jpg
(Public Domain)

1.1.1 Pathophysiology

Despite various efforts, no definite mechanism for IA development has been completely agreed upon [12]. The structure of a typical cerebral artery can be divided into three concentric layers, the tunica intima, tunica media, and tunica externa. Located closest to the vessel lumen, a thin layer of endothelial cells, and a well-developed, thick internal elastic lamina constitutes the tunica intima (Figure 1.2) [13]. An external elastic lamina is characteristic of other arteries throughout the body, but the absence of this layer in cerebral arteries is a distinguishing feature [14]. The middle segment of a cerebral artery is the tunica media and consists mainly of tight bundles of vascular smooth muscle cells (VSMC) [13,15]. Much of the structural integrity of the artery wall originates from this layer. VSMCs, as the primary cellular component, have been shown to play a significant role in aneurysm formation [16]. The tunica externa is the thin, outermost layer of the arterial wall and contains mostly collagen fibers, fibroblasts, and white blood cells [13]. In comparison to the structure of a typical cerebral artery, the vessel wall in IA is markedly different. The tunica media experiences significant atrophy, and thus, the composition of the vessel wall is predominantly externa and intima [17–19]. The internal elastic lamina is also impacted during IA development. This component of the vessel wall may be disrupted or absent completely [19,20].

Identifying arterial bifurcations as a frequent location for IA development prompted researchers to take a closer look at the hemodynamic stress within arterial junctions. A dysfunction in the endothelial cell layer as a result of hemodynamic stress marks the first step in IA formation [21]. The increased amount of wall shear stress at arterial bifurcations is a significant factor in their development [22,23]. Meng *et al.* (2007) demonstrated this by constructing new arterial bifurcations in the carotid vasculature of female dogs and observing the vessel

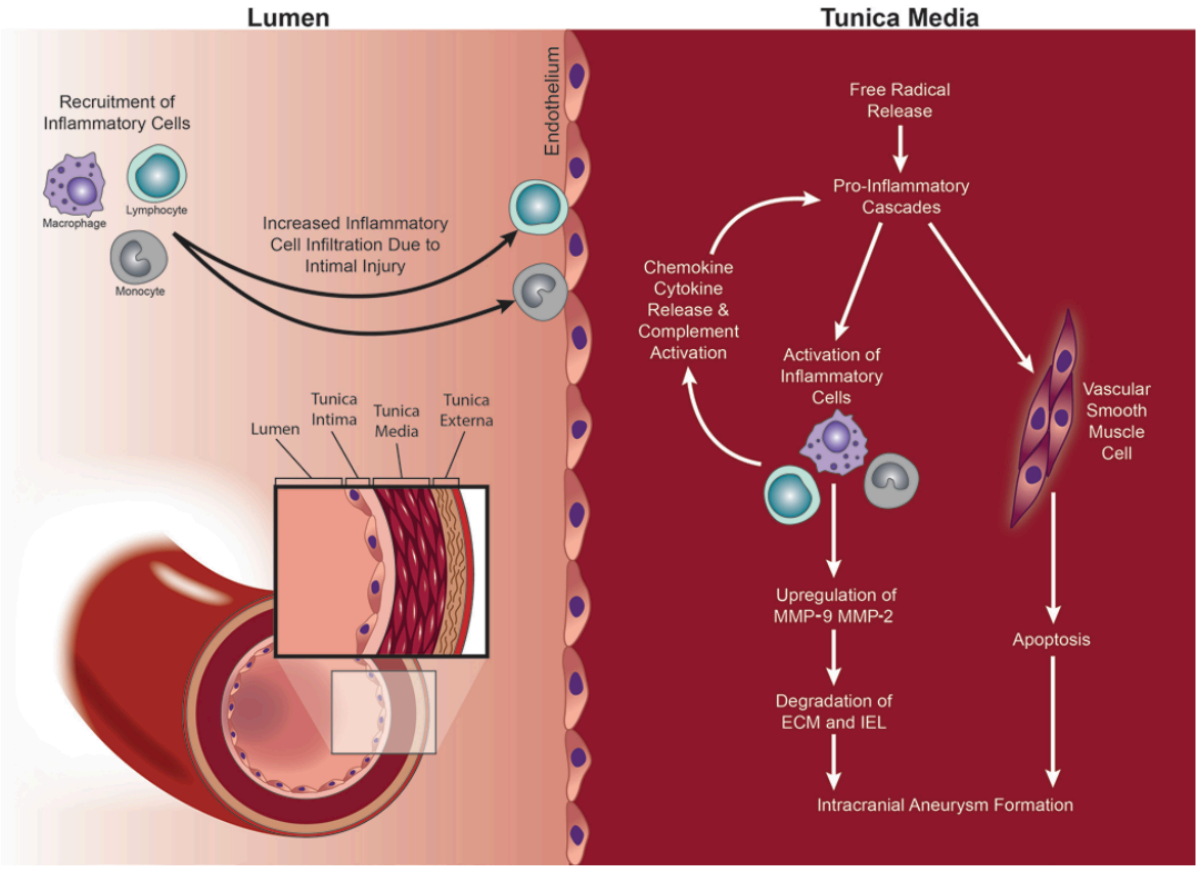


Figure 1.2. Pathogenesis of IA formation in tunica media of cerebral artery

Used with permission from Zhang *et al.* (2019)

remodelling [24]. The exact order of events that occur after endothelial cell dysfunction remains unclear, yet the role of VSMCs is widely accepted. Concentrated in tight bundles within the tunica media, VSMCs are the predominant source of collagen fibers and contribute significantly to the structural integrity of the arterial wall [20,25]. An inflammatory cascade ensues during IA formation causing phenotypic modulation of VSMC [26]. A recent study has demonstrated that phenotypic modulation is triggered by tumor necrosis factor alpha (TNF-α) induced down regulation of contractile genes. The result of this phenotypic modulation is pro-inflammatory VSMCs that move from the tunica media to the inner intima, where they are no longer compacted but proliferate and dissociate [25]. Restructuring of the extracellular matrix ensues as the ability of VSMCs to synthesize collagen is impaired, and matrix metalloproteinase (MMP) production is stimulated [15,20]. The combination of impaired collagen synthesis and increased MMP expression would likely lead to a reduction in the strength of the vessel.

Activation of inflammatory cells such as lymphocytes, monocytes, and macrophages is another widely recognized, critical pathway in IA formation. Studies have demonstrated the infiltration of inflammatory cells, especially macrophages, in the cerebral aneurysm tissue of both rats (drug induced IA formation) and humans [27–29]. Human cerebral tissue was collected during aneurysm repairs and from post-mortem samples. Furthermore, pharmacological depletion of macrophages has shown a reduced incidence of IA in a mouse model (IA formation induced with elastase and angiotensin II), emphasizing the role of macrophages in IA development [30]. In addition to the MMPs synthesized by VSMCs, leukocytes are also significant contributors to MMP production. [20,31]. The up-regulation of MMPs has a considerable effect on both the extracellular matrix and the internal elastic lamina. Many extracellular matrix proteins are known substrates of MMPs. Aoki *et al.* (2007) demonstrated that MMP-9 and MMP-2 play a

role in the degradation of the extracellular matrix components, elastin, and collagen, respectively [28]. The study also demonstrated that the progression of cerebral aneurysm formation could be prevented by selectively inhibiting MMP-9 and MMP-2 in a Sprague-Dawley rat model (drug induced IA formation). *In situ* hybridization has shown MMP-2 expression in blood vessels within the head and trunk of zebrafish [32,33]. Further supporting the role of MMPs in IA development, a recently published genetic association study and meta-analysis found an association between a single nucleotide polymorphism (SNP) thought to influence the MMP-2 gene's transcription levels and IA (Figure 1.2) [34].

Although hemodynamic stress marks the initial step in aneurysm formation, wall shear stress alone is not solely responsible for IA development. Instead, an imbalance between the integrity of the artery wall and the hemodynamic stress within the artery leads to a weakening of the arterial wall [20]. The integrity of the arterial wall is impacted by various genetic, environmental, and physiological factors that predispose to IA formation in certain individuals [35].

1.1.2 Risk Factors

1.1.2.1 Environmental Risk Factors

It has been well established that non-modifiable risk factors for IA development include female sex and an age greater than 50 years old [35,36]. The prevalence of IA is rare in children and is highest in the fifth and sixth decades of life. Numerous modifiable risk factors can lead to *de novo* IA development, increase the risk of growth in a pre-existing IA, or increase the risk of IA rupture. Cigarette smoking increases the risk of developing IA, may hasten the growth of a pre-existing IA and increase the risk of SAH [35]. A case-control study by Feng *et al.* (2018) evaluated 381 cases of IA and found that current smokers had a significantly increased risk of IA rupture

compared to both non-smokers and previous smokers [37]. There is an increased prevalence of IA and an increased risk of SAH in individuals with hypertension and excessive alcohol consumption. These individuals also have an increased risk of aneurysm rupture [36,38,39]. Can *et al.* (2018) found that only current alcohol use is significantly associated with IA rupture and does not persist in individuals with a history of alcohol abuse, thus highlighting the importance of alcohol cessation amongst an IA diagnosis [40]. In addition to smoking, hypertension, and alcohol, the risks associated with oral contraceptive use has also been explored. However, the evidence remains inconclusive [36].

1.1.2.2 Physiologic Risk Factors

Other well-established risk factors contributing to the risk of IA rupture are size and location. Large IAs that are greater than 7 mm in diameter have a significantly higher risk of rupture [41,42]. Various studies have also suggested that IAs located in the anterior communicating artery or posterior communicating artery are at greater risk of rupture [41–43]. Composed of the interconnected anterior cerebral artery, anterior communicating artery, posterior cerebral artery and posterior communicating artery, the anatomical configuration of an individual's circle of Willis varies (Figure 1.3). With 85% of IAs originating in the proximal circle of Willis, 35% of which are located in the anterior communicating artery, understanding the significance of this arterial structure is essential [11]. The possibility an individual's circle of Willis configuration is heritable has also been explored. van Kammen *et al.* (2018) examined the circle of Willis architecture in two unrelated families with a history of intracranial aneurysms. They found that the incomplete posterior communicating artery configuration was more likely to occur within families than between [44]. A recent study by Stojanovic *et al*. (2019) established a significantly

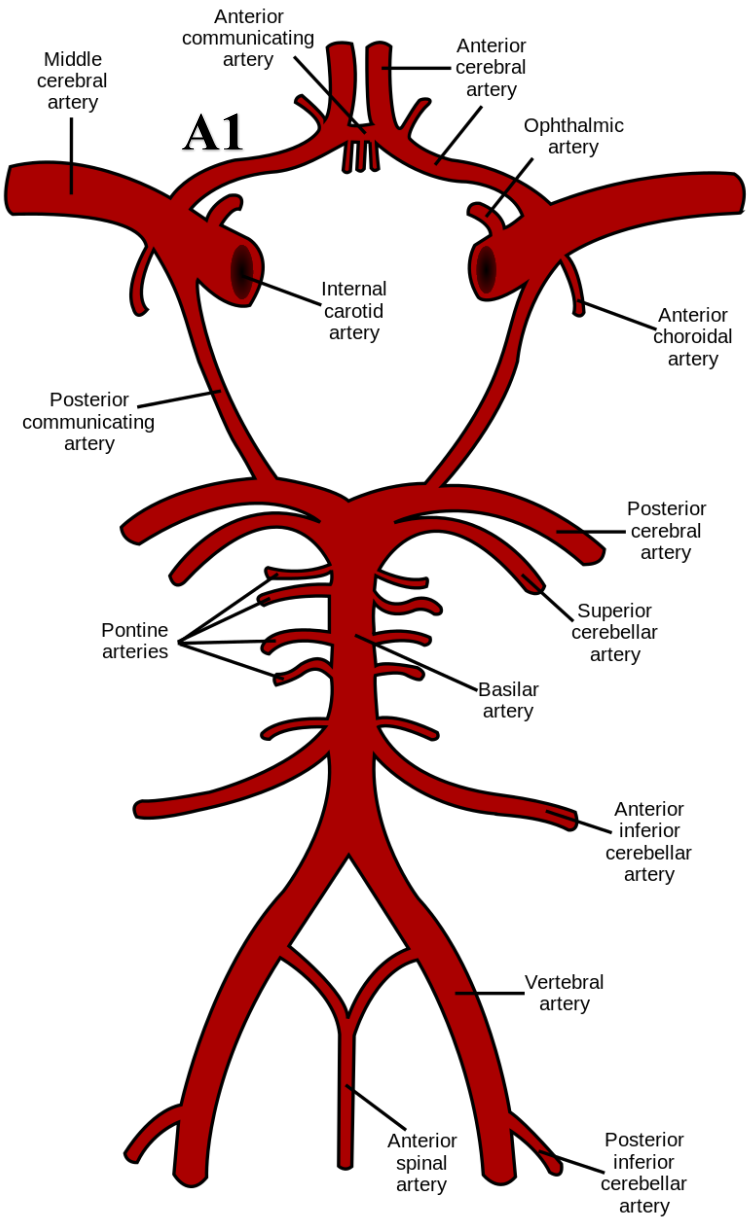


Figure 1.3. Circle of Willis anatomy in cerebral vasculature

The circle of Willis includes the anterior cerebral artery, anterior communicating artery, posterior communicating artery, and posterior cerebral artery

Adapted from: https://commons.wikimedia.org/wiki/File:Circle_of_Willis_en.svg (Public Domain)

increased risk of cerebral aneurysm rupture in patients who have an asymmetrical circle of Willis configuration [45]. The configuration with the strongest link to aneurysm rupture was hypoplasia/aplasia of the A1 segment of the anterior cerebral artery in an asymmetric circle of Willis [45]. As suggested, identifying anatomic characteristics that may predispose to an increased risk of aneurysm rupture could influence the future management plan for patients with unruptured aneurysms [45].

1.1.2.3 Genetic Risk Factors

Considerable efforts have attempted to piece together the genomic contribution to IA development as estimates suggest that up to 25% of IAs have an underlying genetic component [46]. An individual's risk of developing an IA increases by seven times if a first-degree relative has had an IA [46]. Despite significant advancements, there is still a lot to learn about the genomic contribution to IA. There is a strong association between IA and other vascular connective tissue disorders with recognized genetic contributions (Table 1.1) [46]. Anywhere from 4-17% of autosomal dominant polycystic kidney disease (ADPKD) patients will eventually develop IA [46]. Other hereditary Mendelian disorders that are at risk for IA development include Ehlers-Danlos syndrome, Marfan Syndrome, Neurofibromatosis Type 1, and Loeys-Ditz Syndrome [47]. Investigations have found that less than 10% of IA is attributable to these disorders, and thus the focus of this thesis is on non-syndromic IA development [36,41]. The interplay of various physiologic, environmental, and genetic risk factors makes determining the source of non-syndromic IA development challenging and further complicates decisions surrounding screening, diagnosis, and treatment options.

Table 1.1 Prevalence of IA in selected syndromes and their associated genes [a]

Syndrome	Associated Genes	Prevalence of Patients with IA[b]
ADPKD	*PKD1, PKD2*	4 – 17% [46]
Ehlers-Danlos Syndrome	*COL3A1*	12% [48]
Marfan Syndrome	*FBN1*	14% [48]
Neurofibromatosis Type I	*NF1*	9 – 11% [48]
Loeys-Ditz Syndrome	*TGFBR1*, *TGFBR2*, *SMAD3*	11 – 28% [48]
Pseudoxanthoma Elasticum [c]	*ABCC6*	N/A
Hereditary Hemorrhagic Telangiectasia [c]	*ENG*	N/A
Multiple Endocrine Neoplasia Type 1 [c]	*MEN1*	N/A

[a] Adapted from Hitchcock *et al.* (2017) (CC BY 4.0)
[b] Reported prevalence estimates from some studies may include both saccular and fusiform IAs
[c] Individual case reports only

1.1.3 Screening and Diagnosis

As a result of the asymptomatic nature of IA, in combination with the advancements in neuroimaging, the majority of diagnosed IAs are incidental findings [49]. Digital subtraction angiography (DSA) is considered the goal-standard neuroimaging tool for unruptured IA diagnosis. DSA can produce a higher resolution image with excellent specificity and sensitivity. However, the invasive nature, higher cost, associated risks, and radiation exposure makes its use more appropriate when additional structural features are necessary rather than as a primary diagnostic tool [36,49,50]. Therefore, computed tomography angiography (CTA) and magnetic resonance angiography (MRA) continue to be more frequently used methods [36,49]. The sensitivity and specificity of CTA and MRA are also high for the diagnosis of IA, especially for those greater than 3 mm in diameter [50,51]. The less invasive nature and lower costs associated with CTA and MRA also add to their appeal. Guidelines from the American Heart Association suggest that ultimately, both CTA and MRA are appropriate tools for use in unruptured IA detection and follow-up [36].

In families with two or more individuals with IA or SAH, screening of unaffected relatives is recommended [36]. Screening recommendations are particularly important in individuals who have risk factors associated with IA development and a family history of IA. The timeline for beginning screening is ten years before the earliest age of IA diagnosis within a family, according to expert consensus [47]. As with any screening program, it is essential to acknowledge the negative impact an IA diagnosis may have on a patient's quality of life. Aneurysm repair is not always the most appropriate management path for IA. Quite often, conservative management with follow-up is used, and thus patient counseling is of the utmost

importance before screening [52]. Furthermore, family members of individuals with sporadic IA, where there is no suspected familial contribution, are not recommended for screening [36].

The higher risk of IA development in patients with heritable connective tissue disorders makes screening a reasonable consideration [36]. The particularly strong link between IA and ADPKD makes screening with MRA a recommendation in these patients [36,46]. A recent study by Flahault *et al*. (2018) compared the use of targeted screening in which only ADPKD patients with a family history of IA were screened versus systematic screening of all ADPKD patients [53]. They found that systematic screening was, in fact, cost-effective, and a better option considering the high rupture rate of IA in ADPKD patients without a family history [53]. There is a lack of data surrounding the benefits of screening in other IA associated disorders currently, and thus screening recommendations must be evaluated for each case individually [47].

1.1.4 Management of Unruptured IAs

The appropriate management of unruptured IAs is continuously changing. Our understanding of IA pathogenesis continues to evolve, and new technological advances are frequently occurring. Generally, there are three broad categories of treatment options that include conservative management, surgery, or endovascular therapy [50]. The management plan for each unruptured IA is highly individual. Physicians must take into consideration the physiologic characteristics of the aneurysm, such as size, location, and morphology, in addition to patient characteristics such as any environmental and genetic risk factors. Deciding the best management protocol for unruptured IAs can be challenging, partially due to the uncertainties of IA pathogenesis and the numerous risk factors associated with their risk of rupture. The PHASES score was introduced in 2014 to aid in the clinical decision making of unruptured IA [54].

This risk prediction chart uses six predictors of prognosis that include population, hypertension, age, the size of the aneurysm, previous SAH from another aneurysm, and the site of the aneurysm, to predict the five-year aneurysm rupture risk. Each predictor is associated with a specific number of points, and the larger an individual's total score, the higher the risk of aneurysm rupture [54]. In 2015, the unruptured IA treatment score (UIATS) was generated as an additional tool to assist with unruptured IA treatment decisions [55]. Unlike the PHASES score, the UIATS is not a predictive model, and instead, it is a scoring model generated by multidisciplinary specialists in the cerebrovascular field. A physician can use the model to determine if other specialists would support a specific management plan based on various patient and aneurysm characteristics [55]. Most recently, a system was created to predict the three and five–year absolute risks for aneurysm growth [56]. The ELAPSS system uses many of the same predictors as PHASES, except aneurysm shape is included, and hypertension is not considered. Evaluating the risk of growth can help physicians determine the optimal follow-up time for each patient. Although there are limitations associated with these systems and the management of each patient is highly individual, they can be used as supplemental tools when the most appropriate management plan is unclear [56].

A conservative management approach is often used for smaller IAs (less than 7 mm), or where there is a low risk of aneurysm rupture [36,50]. Extensive counseling is required for these patients to emphasize any modifiable factors that will increase the risk of IA rupture. The American Heart Association guidelines suggest that counseling should include information about smoking cessation, moderate alcohol consumption, and hypertension control [36]. Conservative management also requires routine serial imaging with MRA or CTA to track aneurysm growth or morphological changes [36,50].

Surgical occlusion has long been used as a treatment option for unruptured aneurysms, with the first successful clipping of an internal carotid aneurysm taking place in 1937 [57]. Although other surgical techniques such as wrapping and occlusion bypass are occasionally used for complex aneurysms, basic aneurysm clipping is still widely used [36,49]. The procedure generally involves the placement of a metal clip at the neck of the aneurysm during an open craniotomy (Figure 1.4). Complete occlusion of blood flow is the goal of treatment in order to isolate the aneurysm from the parental blood supply [1]. The procedure is quite effective with minimal recurrence risk as total occlusion occurs in the majority of cases [1,58]. Major risk factors associated with surgical management of IA involve the aneurysm size and location [36]. Kotowski *et al*. (2013) found that IAs located in the posterior circulation and those with an increased diameter were associated with unfavorable surgical outcomes such as death and disability [15,59]. Treatment guidelines suggest that long term follow-up should be considered post-surgery, especially cases of incomplete obliteration [36].

Endovascular techniques are another treatment option for unruptured IA. Developed after surgical clipping, endovascular therapy provided a less invasive alternative to surgical techniques with quicker recovery times (Figure 1.5) [49]. Although there are now several different endovascular techniques available such as flow diverters, liquid embolic agents, and more recent medical management strategies, coil embolism remains the most commonly used procedure [1,49]. To isolate an IA from the parental artery, local thrombosis and blocking of blood flow are stimulated by the endovascular insertion of a coil into the aneurysm [1]. To address problems with coils moving out through the neck of larger aneurysms and affecting the parental artery, adjuvant

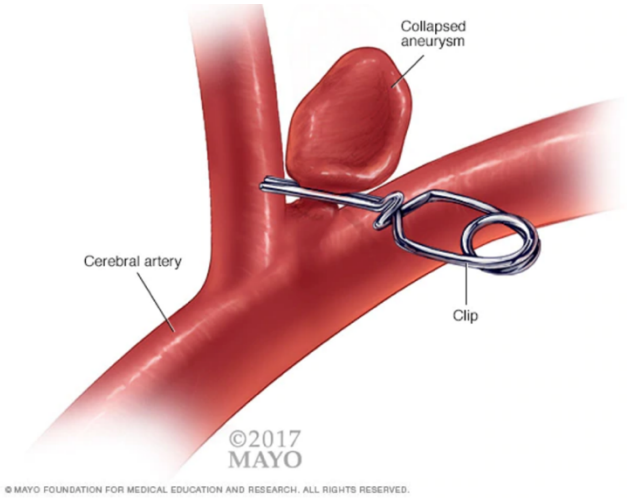


Figure 1.4. Aneurysm clipping

Used with permission of Mayo Foundation for Medical Education and Research, all rights reserved

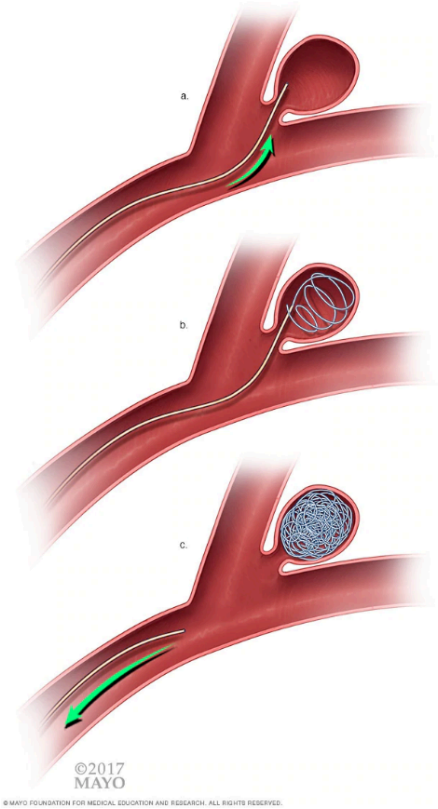


Figure 1.5. Endovascular coiling

Used with permission of Mayo Foundation for Medical Education and Research, all rights reserved

techniques like stent assistance and balloon inflation were generated [50]. Numerous reviews suggest a high quality, large-scale, randomized clinical trial is needed to directly compare surgical clipping and endovascular coiling for IA management [1,15,60]. In comparison to surgical clipping, endovascular techniques have lower associated morbidity and mortality rates [36].

Additional treatment may be necessary for endovascular therapies as there is a greater recurrence risk that should be considered [36]. Guidelines from the American Heart Association suggest that both surgical clipping and endovascular coiling should be performed in higher volume centers as they are associated with better procedural outcomes [36]. Regardless of the treatment selected, all cases of unruptured IAs are unique and must be evaluated on an individual basis. Treatment plans are ultimately the patient's decision. Nonetheless, they must have all the necessary information, including the advantages and disadvantages of each option, to make an informed decision. As more information is uncovered about IA pathogenesis, hopefully, advancements in diagnosis, screening, and management will follow.

1.2 Genetic Component

1.2.1 Family IA Studies

The genetic predisposition contributing to familial IA development has long been recognized. To gain a better understanding of how familial IA differs from sporadic cases, analyzing large family cohorts has been particularly useful. In 2003, a large familial IA study was published by Wills *et al.* (2003) aiming to assess the inheritance patterns of IA, which seemed consistent with Mendelian inheritance [61,62]. The study included 346 Finnish families, each of which had at least two individuals affected with IA, and no other IA associated syndromes were present. Analysis of the 937 IA patients revealed multiple modes of inheritance.

The study found that IA was passed down to both sons and daughters of affected parents, thus indicating an autosomal inheritance pattern. With 57.2% of families demonstrating inheritance patterns consistent with autosomal recessive inheritance, this was the predominant pattern. There was also evidence of autosomal dominant inheritance in 36.4% of families, and 5.5% had autosomal dominant inheritance with incomplete penetrance. There was an additional 0.9% of families that did not align with any known inheritance pattern [61].

The Familial Intracranial Aneurysm (FIA) study was another large, well-known study that utilized large family cohorts to investigate familial IA. The study was a collaborative effort between researchers and clinicians across North America, Australia, and New Zealand with 26 clinical centers, 41 recruitment sites, and over 500 families enrolled [63]. Broderick *et al*. (2009) used 542 families that were recruited as part of the FIA study to investigate the risk of rupture in familial IA. The results indicated that there is a greater risk of rupture in familial IA as compared to sporadic cases of similar size [64]. The study published by Broderick *et al.* (2009) is just one example of countless studies that have since been published using the FIA patient cohort to help understand familial IA.

1.2.2 Linkage Studies

Early investigations evaluating the genetic contribution of familial IA development employed the use of linkage analysis. A useful tool for investigating Mendelian disorders, linkage studies attempt to identify the genomic location of genes that are contributing to disease development [65]. Linkage studies use a family-based approach to evaluate the co-segregation of IA with known genetic markers and have been able to identify several IA susceptibility loci [66]. The FIA study, an extensive collaborative international effort, performed a whole genome SNP

linkage analysis on IA families from across the globe. They first performed a linkage analysis midway through the study on 192 families then completed a subsequent analysis on the total cohort of 333 families once the genotyping was complete [67,68]. The results from both studies demonstrated possible linkage to four chromosomal regions, with linkage to region 4q32 being the most evident, although not statistically significant. The lack of strong linkage detection to one specific chromosomal region in this study is also characteristic of other linkage analyses. A recent review article by Zhou *et al.* (2018) summarized the loci identified via linkage analysis for familial IA (Table 1.2) [66]. The chromosomal regions 1p34.3-36.13, 4q32, 7q11, 19q13 and Xp22 have been validated in more than one study across multiple populations, thus increasing the likelihood of their relevance to familial IA.

In 2001 the first genome-wide linkage analysis of IA was published using a cohort of 84 families of Japanese ethnicity [69]. Results identified three chromosomal regions as IA risk loci, 5q22-31, 7q11, and 14q22. The most substantial evidence of linkage was reported near the marker *D752472* on chromosome (Chr) 7q11. The study reported the closely positioned elastin (*ELN*) as a positional candidate IA gene due to its location and functional relevance as a blood vessel wall component [69]. Three years later, Farnham *et al.* (2004) confirmed 7q11 as a risk locus through linkage analysis in 13 families from the United States [70].

Shortly after the first linkage analysis to identify chromosomal regions related to IA was published, Olson *et al.* (2002) performed a similar study examining 85 Finnish families [71]. Although there was no overlap in the loci identified, Olson *et al.* (2002) found chromosomal regions Xp22 and 19q13 to be suggestive of linkage [71]. The significance of region 19q13 as an IA susceptibly loci was further validated in a follow-up study. In addition to confirmation of previous linkage, the study also defined the region on Chr 19q13 between markers *D19S5454*

Table 1.2. Risk Loci Reported in the Linkage Studies for Familial Intracranial Aneurysms[iv]

Linkage Loci	Populations	Study Cohort: Affected/Family (LOD Score)	References
1p34.3-36.13	North American	10/1 (4.2)	Nahed et al (2005)[72]
	Dutch	7/1 (3.18)	Ruigrok et al (2008)[73]
2p13	Dutch	7/1 (3.55)	Roos et al (2004)[74]
4q32	FIA	482/192 (3.5)	Foroud et al (2008)[67]
		866/333 (2.6)	Foroud et al (2009)[68]
5p15.2-14.3	French-Canadian	12/1 (3.57)	Verlaan et al (2006)[75]
5q22-31	Japanese	104 sib-pairs (2.24)	Onda et al (2001)[69]
7q11	Utah	39/13 (2.34)	Farnham et al (2004)[70]
	Japanese	104 sib–pairs (3.22)	Onda et al (2001)[69]
8p22.2	Korean	13/5 (3.61)	Kim et al (2011)[76]
11q24-25	North American	10/2 (4/3)	Ozturk et al (2006)[77]
12p12.3	FIA	866/333 (3.1)	Foroud et al (2009)[68]
13q14-21	French-Canadian	10/1 (4.56)	Santiago-Sim et al (2009)[78]
14q22	Japanese	104 sib-pairs (2.31)	Onda et al (2001)[69]
14q23-31	North American	10/2 (3.0)	Ozturk et al (2006)[77]
17cen	Japanese	127/29 (3.0)	Yamada et al (2004)[79]
19q13	Finnish	48 sib-pairs (2.6)	Olson et al (2002)[71]
		222 sib-pairs (3.16)	van der Voet et al (2004)[80]

[iv] Used with permission from Zhou *et al.* (2018)

	Japanese	41/9 (4.1)	Mineharu et al (2007)[81]
		127/29 (2.15)	Yamada et al (2004)[79]
Xp22	Dutch	7/1 (4.54)	Ruigrok et al (2008)[73]
	Finnish	48 sib-pairs (2.08)	Olson et al (2002)[71]
	Japanese	127/29 (2.16)	Yamada et al (2004)[79]

and *D19S246* to most likely harbour an IA candidate gene [80]. Further evidence of linkage on Chr 19q13 was identified in a genome-wide non-parametric linkage analysis of 29 Japanese families. The study identified both 19q13 and Xp22 as potentially interesting regions for familial IA [79,82]. A follow-up study used the same cohort of Japanese families but completed a parametric analysis assuming autosomal dominant inheritance of IA. The results confirmed linkage between autosomal dominant IA in Japanese families and chromosome 19q13, suggesting this region may be relevant across populations [81].

Chr region Xp22 is the only familial IA risk loci identified in three different populations [66]. Already acknowledged as an IA susceptibility locus in both Finnish and Japanese populations, Ruigrok *et al.* (2008) replicated these findings when they identified significant linkage to Xp22 in a large Dutch family [73]. They also found suggestive linkage to Chr 1p36, which had already been highlighted as a risk locus in a large family from North America with 11 affected individuals. More specifically, the region on Chr 1 initially identified is a 12.5cM segment from 1p34.3-36.13 [72].

Linkage analysis has previously been successful in identifying disease-causing genes responsible for many Mendelian disorders, including Huntington's disease and cystic fibrosis [65]. Nevertheless, linkage analysis can be problematic as it is often difficult to extrapolate the results to a larger population. Refining a candidate region to a specific gene of interest can also be challenging [83]. While many IA susceptibility loci have been identified via linkage analysis, clarifying which particular variants are relevant to IA development within these chromosomal loci can be difficult.

1.2.3 Genome-wide Association Studies

Seven years after the first linkage study of familial IA, genome-wide association studies (GWAS) gained appeal as an alternative methodology to study IA [84]. Without specifying a candidate gene of interest, GWAS can provide an association between human disease and known SNPs by scanning the genomes of affected individuals and comparing them to unrelated disease–free controls [66,83]. Although linkage analysis has been successful in studying various monogenic disorders, it has been much less beneficial in more common, complex diseases. The emergence of GWAS provided an alternative approach that also addressed some of the limitations surrounding linkage analysis [85].

The initially published GWAS for IA considered patients from Finland, Japan, and the Netherlands. Using these three cohorts and controls, Bilguvar *et al.* (2008) found two new loci on chromosomes 2q and 8q, and the previously highlighted region on chromosome 9q, to be significantly associated with IA [84]. All cases of IA were confirmed via MRA, CTA or DSA. Control cases came from Finland, Japan, and the Netherlands, but only Japanese controls were screened to verify there was no IA. A follow-up GWAS by the same group conducted shortly after increased the number of patients in the Japanese cohort and added two additional European cohorts. This subsequent GWAS identified three new chromosomal regions, including 10q24.32, 13q13.1, and 18q11.2, with strong associations to IA and furthermore confirmed the previously established associations with areas 8q11.23-q12.1 and most significantly 9q21.3 [86]. While many loci have been associated with IA, locus 9q21.3 has been replicated most frequently (Figure 1.6) [47,87]. This locus is located nearby the cyclin dependent kinase inhibitor 2A (*CDKN2A*) gene, which has tumor suppressor functions, and the adjacent cyclin dependent kinase inhibitor 2B (*CDKN2B*) that is involved in cell cycle regulation [47,88]. The *CDKN2A*-*CDKN2B* gene cluster

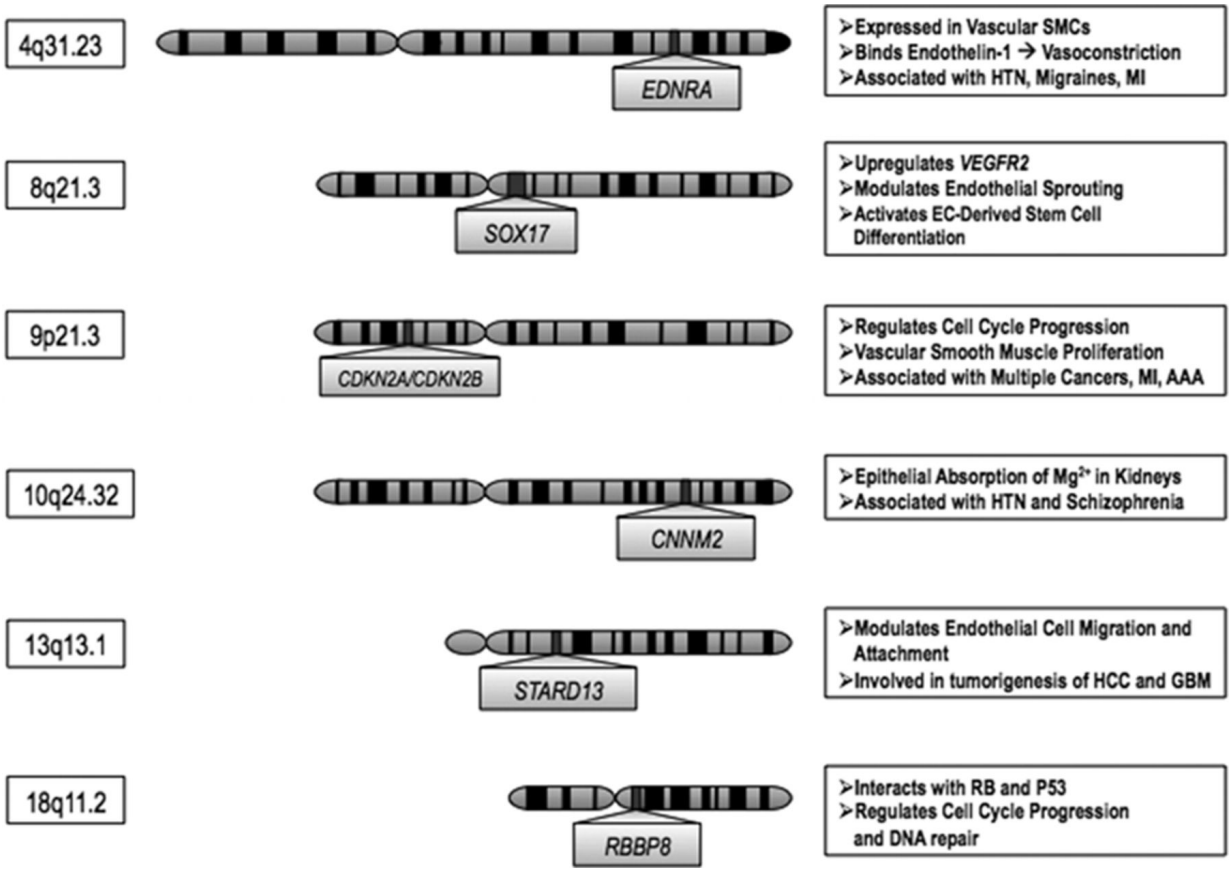


Figure 1.6. Significant IA candidate genes discovered via GWAS

Used with permission from Hussain *et al.* (2013)

also contains methylthioadenosine phosphorylase (*MTAP*) and CDKN2B antisense RNA 1 (*CDKN2B-AS*). *MTAP* is involved in polyamine metabolism and *CDKN2B-AS* is a large antisense non-coding RNA otherwise known as *ANRIL* [89,90]. A GWAS by Foroud *et al*. (2012) found that rather than an association to *CDKN2A* or *CDKN2B* as previously reported, the association between IA and 9q21.3 is more attributable to SNPs in *CDKN2B-AS1* [91]. The structural complexity of this region has hindered any functional validation, and thus, its actual relevance to IA development is still mostly unknown [92]. Furthermore, although the region 9q21.3 has been consistently associated with IA, it has also been classified as a GWAS hot spot. Several other human diseases such as, coronary artery disease, type 2 diabetes, and various carcinomas, have been associated with this locus [89,93].

The locus 8q11.23-q12.1, near the gene SRY-box transcription factor 17 (*SOX17*), was also associated with IA in the initial GWAS, and has subsequently been validated by other independent groups [84,86,91]. The involvement of *SOX17* in embryonic development, and more specifically, angiogenesis has made it a particularly noteworthy candidate in IA development [88]. Studies investigating *SOX17* have demonstrated its selective expression in the arterial endothelial cells of both newborn and adults mice [94,95]. Furthermore, an important role for *SOX17* in endothelial regeneration post vascular injury has been suggested [96]. While initial studies found an association with both the 5' and 3' regions of *SOX17*, subsequent GWAS have demonstrated an association with only one or the other, depending on the population [84]. A more recent GWAS of patients recruited from the FIA study found an association with only the 3' end, whereas previous GWAS of Japanese populations appear to associate with the 5' end [84,86,91]. The recurrence of locus 8q11.23 prompted functional validation of *SOX17* to evaluate its role in IA development further. Lee *et al.* (2015) demonstrated significant IA formation in a *SOX17*

deficient mouse model under hypertensive conditions as compared to controls [95]. IA development occurred spontaneously in 25% of mice with *SOX17* deletions, yet IA occurrence increased to 60% when systemic hypertensive stress was induced via angiotensin II infusions. The combination of GWAS data, endothelial expression, and functional analysis provides evidence to suggest *SOX17* plays a notable role in IA formation [95].

While 8q11 and 9q21.3 are the regions persistently replicated via GWAS, many other noteworthy loci have been significantly associated with IA. A small founder population and genetic drift have resulted in unique genetic variation within the French-Canadian population, including an increased incidence of IA [97]. A recent GWAS using this population identified two new loci with significant IA association, 3p14.2 and 3q13.2, which encompass fragile histidine triad (*FHIT*) and coiled-coil domain containing 80 (*CCDC80*) genes, respectively [97]. The control cohort consisted of primarily unrelated individuals from the French-Canadian population without cerebrovascular diseases, and IA cases were confirmed with MRA or during surgery. Over 40% of IA patients in this population also have hypertension which led Zhou *et al*. (2018) to investigate the relationship between SNPs in these genes and hypertensive IA. The results indicated that *FHIT* is predominantly associated with hypertensive IAs, as multiple SNPs in *FHIT* were found to be significantly associated with this trait. The biological relevance of the *CCDC80* gene makes it particularly interesting due to its involvement in cell adhesion, matrix assembly, and wide expression in VSMCs [88,97].

The co-occurrence of multiple aneurysm types has also prompted investigations into shared genetic contributions. A large meta-analysis published in 2016 analyzed GWAS data from previously published IA, thoracic aortic aneurysm (TAA) and abdominal aortic aneurysm (AAA) cases and controls. The goal was to determine any significantly shared chromosomal loci [98].

While no significant polygenic overlap was found, the majority of analyzed cases were sporadic IA. However, aneurysmal co-occurrence exists predominately in families. The study suggested that a shared genetic contribution may exist in familial cases, and further analysis with larger sample sizes are required [98]. Although the various loci and candidate genes identified by GWAS have laid a solid foundation to help understand the genetic contribution leading to IA development, significant knowledge gaps remain for this complex vascular condition [47,66].

1.2.4 Next Generation Sequencing

Genetic variations that underlie several Mendelian disorders have been identified due to advancements in next generation sequencing (NGS) technologies [99]. With the use of whole exome sequencing (WES), any discrepancies between the protein-coding region of an individual's genome and that of the human reference genome can be identified, and conclusions can be formed about common genetic variants. Although the exome accounts for only 1% of the human genome, variants underlying Mendelian disorders are most often identified in this region [99]. The previous success of using this technique to identify candidate variants in human disease provided confidence that WES could also be an efficient approach to investigate IAs.

The population recruited from the FIA study was utilized in multiple publications employing the use of WES. Seven families were selected from this cohort that had the greatest number of family members affected with IA [62]. The initial study performed WES on both affected and unaffected individuals from the seven families, for a total of 50 exomes. The variants generated via WES were prioritized based on numerous filtering criteria that ultimately highlighted 96 candidate genes [62]. The second publication analyzed sequencing data of 45 individuals from the seven families, 36 affected, and nine unaffected relatives [100]. To organize

the WES data, variants were prioritized that segregated with all affected family members, variants with a minor allele frequency (MAF) <1%, autosomal variants, nonsynonymous SNVs, and variants predicted deleterious by two protein prediction programs. Farlow *et al.* (2015) ultimately published a list of 68 rare variants in 68 genes that segregated with IA in at least one family [100]. The chromosomal location of these 68 candidate genes was also compared with the well-established candidate IA loci previously reported in GWAS. However, there was no overlap identified. RNA expression data was also considered, which involved collecting IA tissue samples from patients undergoing surgical clipping and comparing them to vascular tissue from surgical patients with intractable epilepsy as a control. The only gene differentially expressed in IA tissue as compared to the control was transmembrane protein 132B (*TMEM132B*). The study concluded that further research into this list of candidate genes, specifically through replication with a larger sample size or functional analysis in an appropriate functional model, would be beneficial [100].

A similar study by Yan *et al.* (2015) revealed new candidate genes in IA development [101]. WES was performed on the exomes of 42 individuals from 12 families of Japanese ancestry who had at least three individuals affected with IA. There were 78 candidate variants identified after variant prioritization, which encompassed the inclusion criteria: variants shared by all affected family members, MAF <5%, predicted to affect the protein-coding sequence, and predicted damaging by two protein prediction programs. Of the 78 candidates highlighted, 10 variants from nine genes were selected for further investigation based on biological relevance and increased prevalence in IA patients. Replicate association studies used Sanger sequencing and examined two additional unrelated cohorts, one with familial IA cases and another with sporadic IA cases. The variant p.E133Q in A Disintegrin and Metalloproteinase with

Thrombospondin Type 1 Motif 15 (*ADAMTS15*), was the only variant that significantly aggregated in familial IA cases. Furthermore, the study showed antiangiogenic activity of *ADAMTS15* with overexpression of the p.E133Q variant. The study suggested that while further replication and a functional investigation is required, *ADAMTS15* has emerged as a possible candidate IA gene [101].

The increased prevalence of IA in the French-Canadian population has prompted its use to investigate the genetic contribution to IA pathogenesis [97]. In addition to the susceptibility loci identified by GWAS, the use of WES highlighted mutations in ring finger protein 213 (*RNF213*), as a risk factor for IA in the French-Canadian population [102]. A study by Zhou *et al*. (2016) completed WES on 26 individuals with IA across six French-Canadian families. The variant prioritization strategy aimed to prioritize variants shared in affected individuals across multiple families. *RNF213* emerged as the top candidate after five distinct variants within the gene were identified in four families. Furthermore, it was the only prioritized gene with a mutation burden significantly different between affected individuals and controls. Targeted sequencing in a larger French-Canadian cohort, and association testing by Zhou *et al.* (2018) further supported the possibility that *RNF213* was contributing to IA within this particular population [97,103].

Yang *et al.* (2018) combined the data generated from the three previously described studies by Farlow, Yan, and Zhou in a 2018 paper that identified Rho Guanine Nucleotide Exchange Factor 17 (*ARHGEF17*) as an IA risk gene [100–102,104]. The study began with whole genome sequencing (WGS) of twenty Chinese patients with either familial or sporadic IA [104]. A bioinformatics based variant filtering strategy prioritized 30 candidate variants that were subsequently investigated in cohorts recruited from previously published studies [100–102]. Of the 30 candidate variants, six of them were also identified in the replication cohorts. *ARHGEF17*

emerged as a candidate gene of interest due to its particular high mutation burden in IA cases as compared to controls. Using a zebrafish model, Yang *et al.* (2018) demonstrated significant *ARHGEF17* expression in the head region, and knockdown of *ARHGEF17* induced a significant intracranial hemorrhage phenotype, providing further support of its role in IA development [104].

In contrast to the WES studies previously described that utilized cohorts of multiple IA families, Santiago-Sim *et al.* (2016) focused on one large European-American family [105]. There were twenty-one study participants within the family, three with documented SAH, six with unruptured IA, and 13 unaffected relatives. WES was completed on the exomes of two affected family members, and after the implementation of various filtering criteria, 53 heterozygous variants were eventually identified [105]. Santiago-Sim *et al.* (2009) previously published a genome-wide linkage analysis of the same family, which identified a region on Chr 13q as an IA susceptibility locus [78]. After a comparison of the linkage and WES data, only one of 53 variants fell under the linkage peak that resulted in a truncated protein, the nonsense mutation p.R450X in thrombospondin type-1 domain-containing protein 1 (*THSD1*) [105]. The p.R450X variant in *THSD1* segregated with all affected family members and was absent from all unaffected relatives. Targeted sequencing of *THSD1* in 507 unrelated IA probands revealed eight additional missense mutations. Furthermore, functional analysis of *THSD1* in both zebrafish and mouse models resulted in cerebral bleeding, further substantiating the possibility that *THSD1* is relevant to IA development [105].

An additional WES study by Wu *et al.* (2018) focused on a smaller IA family of Chinese ancestry [106]. There were six study participants within the family, five individuals with IA, and one unaffected relative. After WES of three affected individuals and the one unaffected relative, bioinformatic filters were implemented, and fifteen SNVs along with three indels were

recognized as potentially significant. The five variants deemed most functionally relevant to IA were then Sanger sequenced to visualize segregation within the family. One mutation segregated completely with all affected individuals, p.C133T in lysyl oxidase like 2 (*LOXL2*) [106]. The relevance of *LOX* in IA development had been examined previously, but these investigations produced conflicting results. Sathyan *et al.* (2013) reported no significant association between *LOX* polymorphisms and IA in a south Indian population [107]. Alternatively, associations have been found between *LOXL2* and IA in Japanese and Korean populations [108,109]. Wu *et al.* (2018) proposed that the combination of these results indicates that further functional analysis of *LOXL2* should be considered [106].

A study by Bourcier *et al.* (2018) combined WES and identity by descent (IBD) analysis to generate a list of IA candidate variants from a family recruited as part of the French ICAN project [110,111]. This multicenter nationwide research project recruited both familial and sporadic IA cases from across France, aiming to better understand the pathophysiology of IA [110]. The family used by Bourcier *et al.* (2018) had five individuals affected with IA, four of which DNA was available from [111]. Two affected family members underwent WES and IBD analysis, and ultimately ten common variants were highlighted. Eight of these variants were also found in the two other family members with IA. The only loss of function variant within those prioritized was c.1378A>T in angiopoietin-like 6 (*ANGPTL6*). This gene was previously identified as a pro–angiogenic factor capable of increasing the permeability of the endothelium and stimulating cell migration [112]. Bourcier *et al.* (2018) extended their genetic analysis to five additional IA families. They discovered that rare variants within *ANGPTL6* were present in 92% of individuals affected with IA as compared to 37% of unaffected relatives. The study concluded that the results

generated via WES, IBD, and expression data prompt further functional analysis into *ANGPTL6* [111].

Several studies have recently been able to identify genetic contributions that are potentially responsible for the development of non-syndromic IA with the use of NGS techniques (Table 1.3). Although we have gained valuable insight using these technologies, our knowledge of the genetic predisposition leading to IA development is still lacking. It is becoming increasingly clear that one causative genetic variant is likely not responsible for IA development across all populations. Instead, multiple genetic contributions are more plausible. Future NGS studies will be beneficial in identifying additional pathogenic variants contributing to IA development in other families.

1.3 Zebrafish – *Danio rerio*

1.3.1 Functional Model of Genetic Disease

A common problem that arises is the vast abundance of variants generated from NGS. Even with extensive variant filtering pipelines, it is difficult to classify a genetic variant as causative. When relying on clinical data and analyzing pedigree segregation, there are many underlying factors to consider. External risk factors, incomplete penetrance, and potential phenocopies can all impact how variants are ultimately filtered. With significant advancements in NGS technologies and higher numbers of genetic variants being identified, the use of model organisms has become increasingly important. By using a model organism, the link between genotype and phenotype can be validated, and more accurate functional predictions can be made [113].

In recent years, the zebrafish has emerged as an exceptional model organism for

Table 1.3 Candidate IA genes identified via NGS

Gene	Chromosomal Position	Population	References
TMEM132B	12q24.31	FIA	Farlow *et al.* (2015) [100]
ADAMTS15	11q24.3	Japanese	Yan *et al.* (2015) [101]
RNF213	17q25.3	French-Canadian	Zhou *et al*. (2016)[102]
ARHGEF17	11q13.4	Chinese	Yang *et al.* (2018) [104]
THSD1	13q14.3	European-American	Santiago-Sim *et al.* (2016) [105]
LOXL2	8p21.3	Chinese	Wu *et al.* (2018) [106]
ANGPTL6	19p13.2	French	Bourcier *et al.* (2018) [111]

investigating genetic disease. This tropical, brackish water, vertebrate belongs to the largest family of fish, the Cyprinidae family [114]. Although the benefits of using this model in genetic research have come to light in more recent years, zebrafish have been used as an embryologic model since the 1930s [115,116]. An appealing aspect of a zebrafish model is that they address many of the concerns that arise with traditional animal models. They are relatively inexpensive, less time consuming, and less labor-intensive than a conventional model [117]. Post-embryonic development of zebrafish occurs rapidly, and sexually mature adults are formed within approximately three months [118]. Generating zebrafish embryos is a relatively simple process as laboratory breeding can occur throughout the year and does not follow a marked seasonality. Furthermore, their fertilization is external, and each mating pair can generate up to 300 embryos [117,119]. The small size of these fish minimizes the infrastructure required for storage but also allows larger-scale studies (Figure 1.7) [120]. Completing studies at a larger-scale would typically be characteristic of only invertebrate models as space and support required to develop a large scale experiment in other vertebrates are quite costly and labor-intensive [116]. As a model organism, zebrafish have been able to strike a balance between the advantages and disadvantages of other model organisms. Zebrafish address various practicality issues. They provide the affordability and feasibility associated with an invertebrate model such as a fly, while also providing physiologic attributes characteristic of vertebrate models that are essential to study human disease [116].

Many studies have employed the use of zebrafish to investigate various vascular dysfunctions. The fundamental vasculature of zebrafish is similar to that of humans, and their transparent embryos allow for visualization of this vasculature. To investigate any hemorrhagic phenotype, this provides a non-invasive approach to view bleeding in real-time [116].

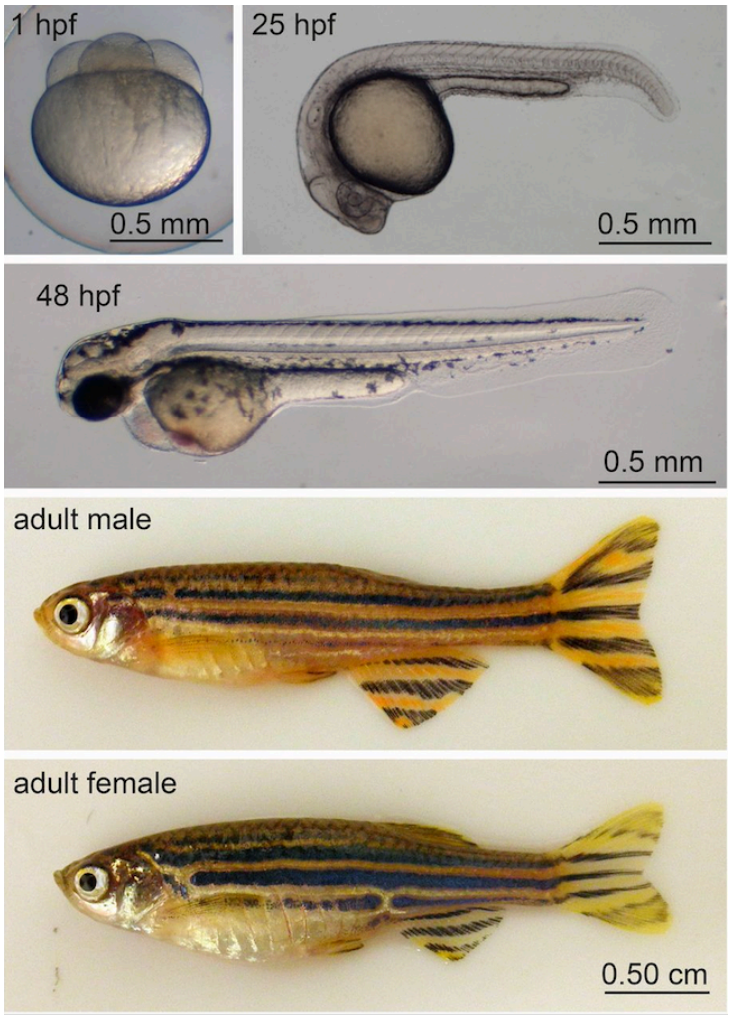


Figure 1.7. Life stages of zebrafish

hpf – hours post fertilization
At 1 hpf the zebrafish embryo is at the four cell stage. At 25 hpf the body axis has formed. At 48 hpf rudimentary organs have formed. The adult male zebrafish often appears more streamlined with a golden color. The adult females often have a protruding stomach with a bluish color.

Used with permission from Holtzman *et al.* (2016)

Particularly relevant to studying IA development, the cerebral vascular framework of zebrafish is conserved among other vertebrates, including humans [121,122]. This anatomical resemblance has allowed various human cerebrovasculature dysfunctions such as cerebral arteriovenous malformations, moyamoya disease, and cerebral amyloid angiopathy to be investigated in a zebrafish model [123]. Similarities also exist between the cardiovascular system of zebrafish and humans. The zebrafish heart consists of two chambers rather than four, but a distinct atrium and ventricle are present [116]. The average heart rate characteristic of zebrafish is much closer to that of humans at 120-180 beats per minute, then the >500 beats per minute typical of mice [117,124,125]. A closed circulatory framework of arteries and veins is another resemblance between zebrafish and humans [116]. During the initial days of development, oxygenation of zebrafish embryos is not dependent on their circulation. Instead, passive diffusion provides an adequate amount of oxygenation for growth [126,127]. This becomes particularly advantageous when studying diseases that impact cardiovascular development during embryogenesis. Maturing independently of the circulation allows zebrafish embryos to survive cardiovascular defects that may be lethal to other traditional models [123].

In addition to the anatomic and physiologic similarities, some homology exists between the genomes of zebrafish and humans. The zebrafish genome is highly conserved with that of humans, with approximately 70% of human genes having an obvious ortholog within the zebrafish genome [128]. Although 47% of these human genes have one distinct ortholog in zebrafish, this is not true for all genes [128]. Specific genes within the zebrafish genome have been duplicated, resulting in some cases of multiple zebrafish genes that are co-orthologous to one single human gene [129]. The duplication of genes within the zebrafish genome is the result of

the teleost specific whole genome duplication that occurred at the origin of the teleost lineage [130–132]. It has been suggested that up to 20% of duplicated genes remain in the zebrafish genome, and thus, the impact of this whole genome duplication is still relevant [132–134]. While the majority of human genes do have a clear zebrafish ortholog, there are some in which an ortholog has yet to be identified, or no such equivalent exists, thus complicating the investigation of such genes in a zebrafish model.

1.3.2 Morpholino Mediated Gene Knockdown

As the success of zebrafish as functional models of genetic disease has grown, the genetic techniques applicable to this organism have also advanced. From the first zebrafish mutant generated in 1983, approaches have ranged from forward genetic screening to directed genetic engineering using methods such as Clustered, Regularly Interspaced, Short Palindromic Repeats (CRISPR/CAS9) mutagenesis and everything in between (Figure 1.8) [135]. The use of morpholino mediated gene knockdown in zebrafish began when Nasevicius *et al.* (2000) showed that by using antisense, morpholino-mediated oligonucleotides, they could both specifically and effectively inhibit translation in zebrafish [136]. Initially designed in 1989, the creation of morpholino oligonucleotides (MO) was a cost-effective alternative to the expensive nucleic acid analogs previously available [137,138]. Composed of an uncharged phosphorodiamidate backbone, MO ring, and twenty-five bases, these synthetic, antisense RNA analogs work in vivo to effectively bind complementary nucleic acids and hinder gene expression [129,138,139]. Splice-blocking MOs inhibit proper splicing of the primary transcript by binding near the splice site while translation-blocking MOs prevent protein synthesis by binding near the start site [129]. The genetic technique employed by MOs is considered gene knockdown, rather than gene knockout.

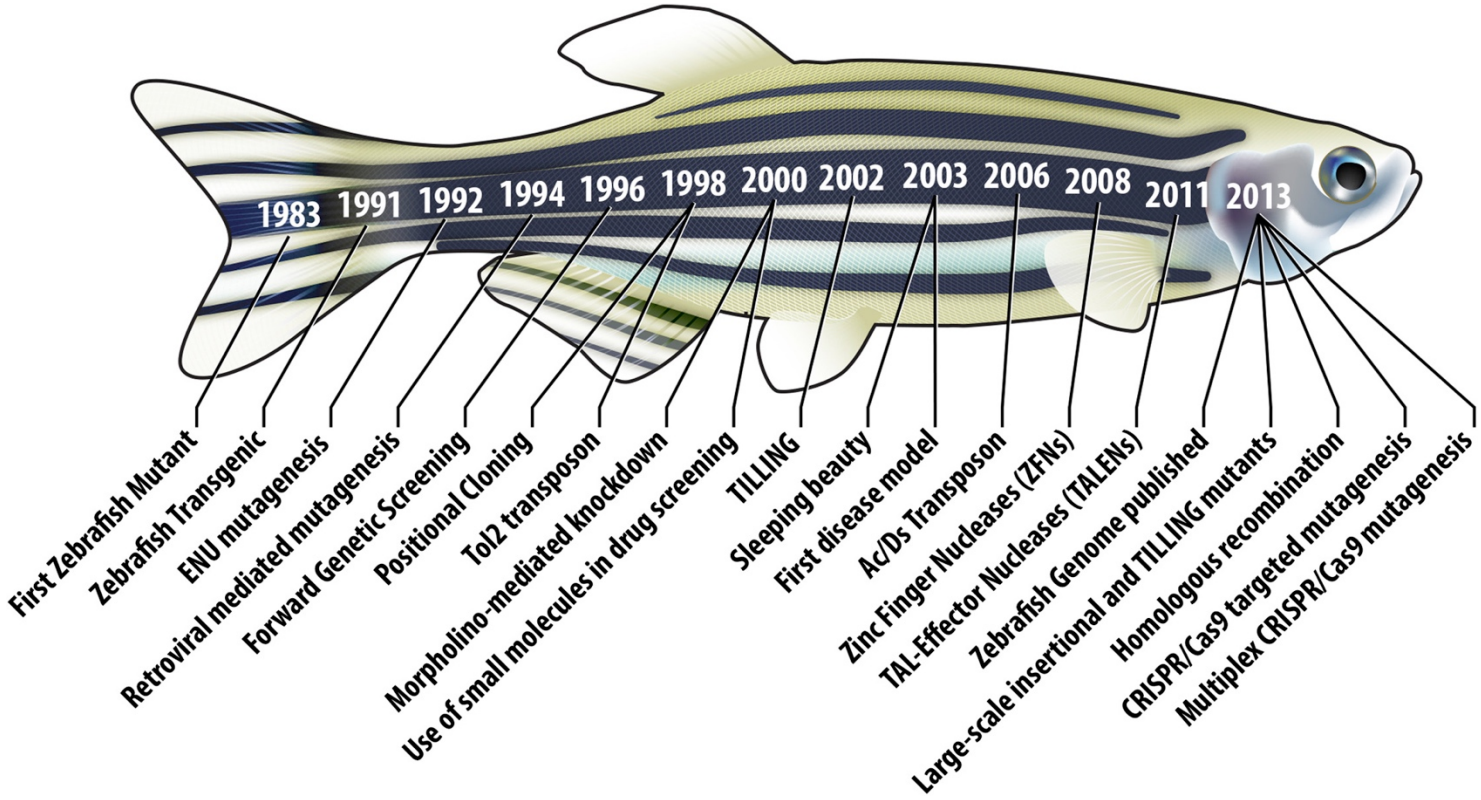


Figure 1.8. Timeline of significant advancement in zebrafish research

Used with permission from Varshney *et al.* (2015)

Knockdown is advantageous as gene expression levels can be reduced but not completely silenced. The phenotype induced by MO gene knockdown is referred to as a morphant phenotype rather than a mutant, which would be the result of CRISPR/Cas9 mutagenesis.

As the use of MO knockdown in zebrafish grew and additional genetic approaches were established, concerns arose regarding the use of MOs as the sole genetic technique for investigating genotype to phenotype correlations. A controversial article published by Kok *et al.* (2015) raised questions regarding the reliability of phenotypes observed in morphant zebrafish after they found many discrepancies between morphant and mutant phenotypes [140]. Although other studies have expressed similar concerns, the value of MOs in genetics research should not be understated [140–142]. Rather than dismiss the use of MOs entirely, many researchers argue that adequately controlled MO experiments are the best solution [143,144]. From a practical perspective, the use of a MO-mediated gene knockdown in a zebrafish model is quite advantageous. MOs are easy to use experimentally. Their uncharged phosphorodiamidate backbone makes them unappealing to enzymes and thus resistant to degradation. These oligonucleotides are usually micro-injected into the yolk of zebrafish embryos anywhere from the 1 to 8 cell stage (Figure 1.9) [136,145]. The whole genome duplication that occurred within the zebrafish genome can make the functional investigation of duplicated genes more challenging. However, MOs provide a solution to knockdown multiple gene products simultaneously [129]. MOs produce genetic knockdown rather than genetic knockout. The effectiveness of a MO can be adjusted based on the concentration that is used. In circumstances where the expression of a gene is necessary to zebrafish survival, the concentration can be titrated to a dose that corresponds with survival [129].

One specific problem identified early on with MOs use in zebrafish was 15-20% of the MOs demonstrated unanticipated off-target effects [146,147]. Although the exact mechanism is

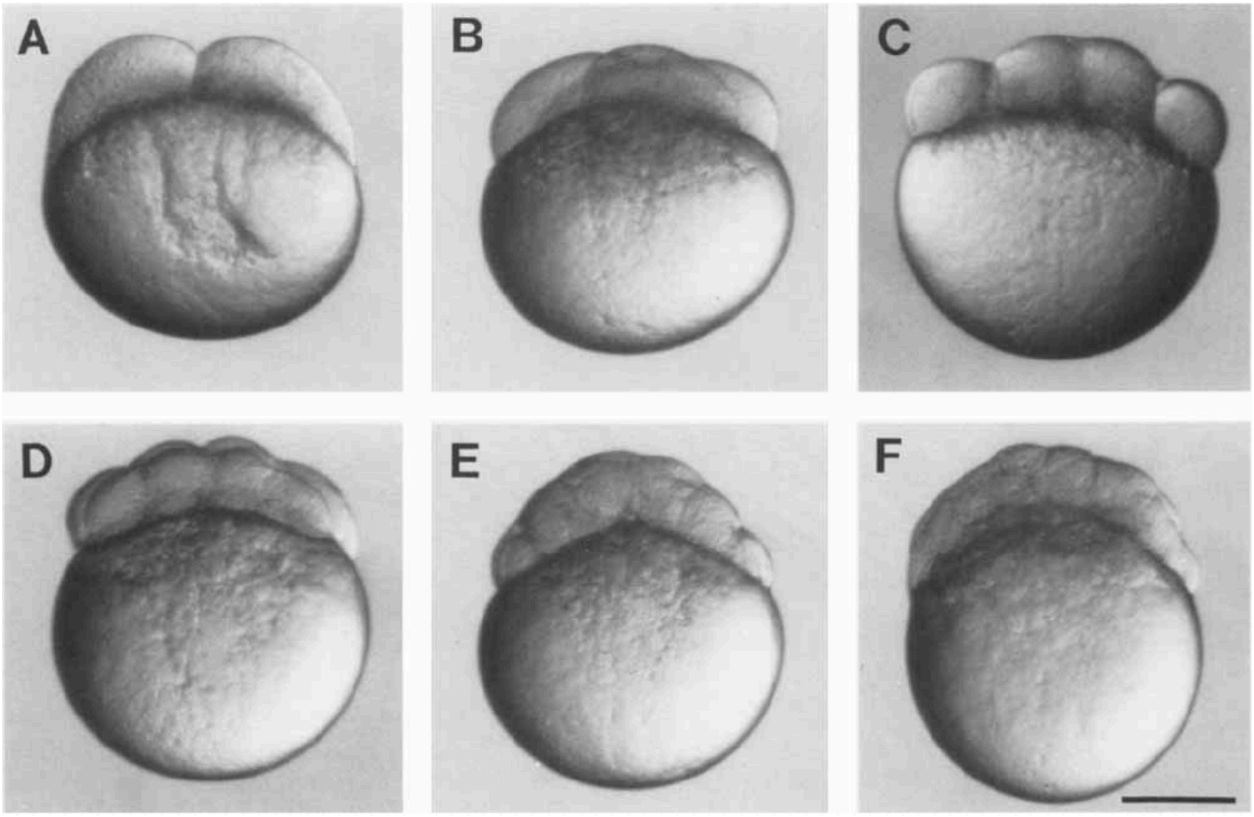


Figure 1.9 Embryos during the cleavage period

(**A**) 2-cell stage (0.75 h). (**B**) 4-cell stage (1 h). (**C**) 8-cell stage (1.25 h). (**D**) 16-cell stage (1.5 h). (**E**) 32-cell stage (1.75 h). (**F**) 64-cell stage (2h). Scale bar: 250 μm

Used with permission from Kimmel *et al.* (1995)

unknown, the majority of unanticipated effects are found to be mediated through tumor protein *p53* (*p53*) induced apoptosis. To combat the up-regulation of *p53* by MOs, the knockdown of *p53* is recommended. Co-injection with a *p53* MO has been found to alleviate the off-target effects and thus can be used as a toxicity management tool [144,147]. Guidelines also suggest that MO phenotypes should be validated by comparing them to a mutant phenotype whenever possible. Without comparison to a mutant, MO experiments should take other appropriate validation measures. This may include using a second MO that differs in sequence, consistently injecting a standard control MO, and incorporating RNA rescue experiments [142]. As is the case with the majority of experimental procedures, there are advantages and disadvantages associated with the use of MOs that should be considered when constructing a study design.

1.3.3 Evaluating Vasculature Dysfunction with Gene Knockdown

Various studies have used MO-mediated gene knockdown methods to evaluate the role of candidate genes on zebrafish vasculature. Not all studies have assessed arterial aneurysm development exclusively but provide confidence in the use of this model to evaluate vascular dysfunction. Investigations into the role of the Forkhead Box C1 (*FOXC1*) gene in cerebral small vessel disease, a condition with an increased risk of stroke, employed the use of MO-mediated gene knockdown in zebrafish [148,149]. French *et al.* (2014) demonstrated that patients with *FOXC1* mutations had increased rates of cerebral hemorrhages, and *FOXC1* zebrafish morphants produced a hemorrhagic phenotype [148]. A paper published by Kiando *et al.* (2016) used zebrafish to investigate fibromuscular dysplasia (FMD) [150]. This vascular disease leads to problems in both renal and cerebrovascular arteries that can also include aneurysm formation. The MO knockdown of Phosphatase and Actin Regulator 1 (*PHACTR1*), a gene previously associated

with FMD, demonstrated subtle vascular impairment in a zebrafish model. In combination with a GWAS, this functional analysis provided evidence for Kiando *et al.* (2016) to report *PHACTR1* as the first identified susceptibility locus of FMD [150].

The strong association between ADPKD and IA has been recognized since 4-17% of patients with ADPKD eventually develop IA[46,151]. The contribution of the *Survivn* gene to polycystic kidney disease (PKD) and IA development was investigated by Aboualaiwi *et al.* (2014). Part of their functional analysis involved the MO-mediated gene knockdown of *Survivn* in a zebrafish model in addition to the PKD mice used for functional analysis. The study concluded that abnormal expression of *Survivn* may be contributing to both the cystic kidney formation and the vascular aneurysm formation [151].

1.3.4 Modeling Intracranial Aneurysms In Vivo

Although various studies have used MO-mediated gene knockdown in zebrafish, few studies have assessed arterial aneurysm development exclusively. The studies published by Santiago-Sim *et al.* (2016) and Yang *et al.* (2018) both employed the use of WES before using zebrafish as a functional model for IA [104,105]. Yang *et al.* (2018) used a cohort of 20 Chinese individuals with IA and 86 previously unsolved cases of familial IA of Japanese, European, American, and French-Canadian ancestry [104]. Their sequencing data suggested *ARHGEF17* as an interesting candidate gene in IA development, and thus, they wanted to investigate further. Functional analysis using MO knockdown in a zebrafish model revealed cerebral hemorrhage. This evidence allowed them to suggest *ARHGEF17* mutations are a risk factor for IA development [104].

Santiago-Sim *et al.* (2016) published a similar study investigating the contribution of *THSD1* mutations to IA development [105]. WES was initially performed on a large family with a strong predisposition to IA. A rare missense mutation in *THSD1* was ultimately identified as a variant of interest, and functional studies were performed in both zebrafish and mouse models. The loss of *THSD1* expression in the zebrafish with MO-mediated gene knockdown revealed a significant cerebrovascular hemorrhage phenotype that was not evident in embryos injected with a negative control MO. Similar findings in *THSD1* knockdown mice provided further evidence to suggest *THSD1* mutations may be contributing to IA development [105]. Another *THSD1* knockdown in zebrafish was also completed by Haasdijk *et al.* (2016) using an alternative MO. They too demonstrated a significant hemorrhagic phenotype [152]. One potential drawback of the models expressed by Santiago-Sim *et al.* (2016) is the lack of aneurysm development observed in zebrafish and mouse knockdowns. Instead, a cerebral hemorrhage phenotype is observed [105] . They suggested that as is the case with humans, small IAs often go unnoticed. However, it is also possible that zebrafish do not develop aneurysms at all and weakening of the vasculature instead leads to a hemorrhagic phenotype [105]. Therefore, instead of looking for cerebral aneurysm formation, a cerebral hemorrhage might be the corresponding phenotype that is observable.

Other studies have used MO-mediated gene knockdown to evaluate the development of thoracic aortic aneurysms and dissections (TAAD), rather than IA. Kuang *et al.* (2016) performed WES on individuals with TAAD from a large family of European descent [153]. Variant filtering identified Forkhead Box E3 (*FOXE3*) as a potential candidate gene in TAAD development, and thus, functional analysis was performed using MO-mediated gene knockdown in embryonic zebrafish. They determined that the aortic arch development was disrupted in embryos with *FOXE3* knockdown. Furthermore, the study also used mouse embryos to evaluate

TAAD development and concluded *FOXE3* mutations are likely contributing [153]. Preceding the publication by Kuang *et al.* (2016), Guo *et al.* (2015) published a similar paper assessing the role of Methionine Adenosyltransferase 2A (*MAT2A*) mutations on TAAD development [154]. They also used WES and functional studies, which demonstrated that *MAT2A*-deficient zebrafish experienced significant cerebrovasculature dysfunction. Thus suggesting *MAT2A* mutations may predispose to TAAD development [154]. The co-morbidity that exists between IA and other arterial aneurysms, and the similarity of their vascular pathology, provides confidence that studies of TAAD are also employing zebrafish models. The detailed methodology outlined does differ within each study, and therefore it is important to note that significant variations are contributing to their respective results. However, the current literature helps speak to the appropriateness of zebrafish as a functional model.

1.4 Newfoundland and Labrador Population

The unique genetic architecture of the Newfoundland and Labrador (NL) population has made it an appealing cohort in genetic research for many years. This 405,212 km^2 island is the most easterly province in Canada and is home to approximately 525,000 people [155]. Immigration to NL from Europe, which was already inhabited by Indigenous People, predominately took place between the eighteenth and nineteenth centuries [156]. The majority of individuals originated from southwestern England and southeastern Ireland, and were drawn to fishing opportunities off the NL coast [156]. Isolated communities began to develop along the shoreline of the province, many of which still exist today. The remote settlement of many outport communities, some of which are only accessible by boat, has caused persistent geographic isolation [157]. The small founder population and minimal immigration, in combination with the geographic isolation, has

resulted in a population with an increased inbreeding coefficient and reduced genetic heterozygosity [157–159].

The NL population has an exceptionally high incidence of various monogenic and complex diseases [157]. Only surpassed by Finland, the NL population has the second-highest incidence of type 1 diabetes mellitus (T1DM) on the planet and the prevalence of colorectal cancer in the NL population is the highest in Canada [160]. In addition to T1DM and colorectal cancer, genetic studies of Bardet-Biedl syndrome, hemophilia A, multiple endocrine neoplasia type 1, psoriatic arthritis, and cardiomyopathies have all been facilitated in the province [161–165]. Although these conditions contribute significantly to the burden of disease, they have also highlighted this population as an appropriate and successful cohort to use for various genetic disease investigations. Within the NL population, numerous families have a strong genetic predisposition to IA and thus this cohort became the focus of this thesis [157,166].

1.5 Hypotheses and Objectives

I hypothesize that highly penetrant pathogenic variants predispose to IA in families from NL. Furthermore, the genetic knockdown of candidate gene(s) in zebrafish will recapitulate a hemorrhagic phenotype.

The goal of this project is to identify any pathogenic variants potentially relevant to the development of IA in families from NL. The first objective is to prioritize the genetic variants generated from the exomes of study participants affected with IA. Secondly, I will use Sanger sequencing to validate these genetic loci and assess the familial segregation. The final objective is to functionally analyze candidate variants that are identified as potentially pathogenic using MO-induced gene knockdown in a zebrafish model.

1.6 Relevance

Bleeding within the subarachnoid space results in devastating clinical outcomes with substantial morbidity and mortality. In 30% of SAHs, the outcomes are fatal, and an additional 30% require nursing care as they are unable to live independently [167]. As the most common non-traumatic cause of SAH is ruptured IA, improving the management of IA is of utmost importance. Identifying genetic variants responsible for familial IA development could lead to new diagnostic techniques and thus reduce the amount of imaging required for screening. A better understanding of IA pathogenesis could also assist in detecting a potential novel therapeutic target. These could eventually lead to advancements in current treatment options and the development of prophylactic strategies.

2. Materials & Methods

2.1 Study Participants

Before my project, significant work was already completed, focusing specifically on two study families, R1352 and R1256 [168]. These two families provided an ideal cohort for initial investigations. There was a particularly strong incidence of IA, and many of the unaffected family members were consenting study participants. With a more focused study already completed, and such an extensive cohort of families available, I expanded my project to focus on 13 of our multiplex families from across NL. This included 46 affected individuals and 308 of their unaffected relatives (Table 2.2). The average age of initial diagnosis for these affected individuals was 50.5 years of age, and 80% of them were females.

The recruitment of study participants affected with IA from across NL occurred before the beginning of this project and has been described previously [168]. Recruitment of individuals began as a collaborative effort between neurosurgeon Dr. Falah Maroun and medical geneticist Dr. Bridget Fernandez (Figure 2.1). Potential participants were referred to the study by neurologists and neurosurgeons across Eastern Health. Consenting patients were considered "affected" if an IA was identified via computed tomography (CT) scan or magnetic resonance imaging (MRI). Subsequently, a three-generation pedigree was collected from all affected participants, and any consenting family members were also recruited to the study.

Further classification of participants was then necessary to identify whether an individual's IA was familial, sporadic, or equivocal in origin. Familial IA was defined as an affected individual with one or more first or second-degree relatives with a known IA. Conversely, sporadic IA was defined as an individual with no known family history of IA. Any undetermined case, in which an IA is not definitively familial or sporadic, was considered

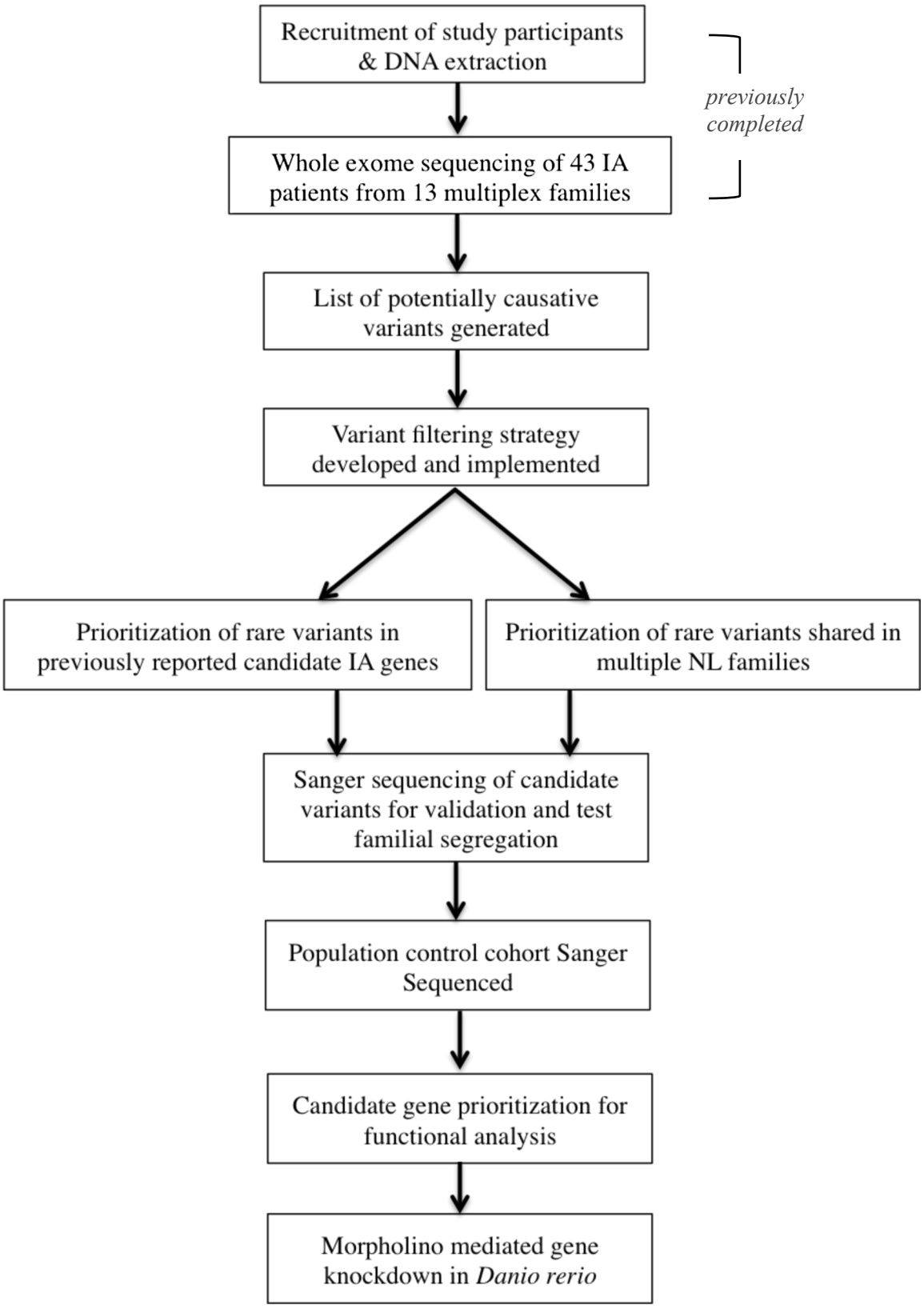


Figure 2.1. Study design

equivocal IA. To date, a total of 53 families have been recruited to the study, encompassing 92 affected individuals and 396 unaffected relatives. An additional 33 sporadic and five equivocal IA cases were also participants in the study.

Clinical data from all affected individuals was collected, including: their date of birth, age of initial diagnosis, number of IAs, size and location of IA, whether a SAH occurred and if surgical intervention was required. Details of the surgical technique performed (clipping or coiling), as well as any surgical complications, were also recorded. Finally, a detailed medical history was collected from all affected study participants, specifically inquiring about risk factors known to lead to IA development such as hypertension, smoking, and DM (Table 2.1).

2.2 Whole Exome Sequencing

The first iteration of WES began in 2015 when 23 exomes, representing six families, were sequenced. In 2016, 19 more exomes, representing seven additional families were also sent for sequencing (Table 2.1). The exomes were sequenced at The Centre for Applied Genomics (TCAG) at The Hospital for Sick Children in Toronto, Ontario. An Illumina HiSeq2500 platform was used for sequencing. All DNA samples were submitted to TCAG in 70 ul aliquots at 100 ng/ul with an optical density ($OD_{260/280}$) ratio between 1.8 and 2.0 to meet sample purity standards. With a previous study already completed, the exome data required for this project was available. The following information regarding the WES platform outlines steps completed by TCAG and has been adapted from the material provided by TCAG with additional information collected from their website (2018) [169].

2.2.1 Library Preparation & Cluster Generation

Samples were prepared by first undergoing ultrasonic vibration that produces random fragmentation of the DNA. Adapters were ligated to both the 5' and 3' ends of the DNA fragments, which can then be polymerase chain reaction (PCR) amplified, followed by purification with gel electrophoresis [170]. Oligos complementary to the ligated adapters are bound to a flowcell where they can capture the DNA fragments. Through bridge amplification, distinct clonal clusters are then generated from each fragment [170].

2.2.2 Sequencing

The HiSeq2500 uses Sequencing by Synthesis technology that utilizes reversible terminator chemistry for base-by-base sequencing [171]. During each sequencing cycle, four reversible terminators are present (A, C, T, G), all of which are fluorescently labeled. After each cycle, the deoxyribonucleotide triphosphate (dNTP) that was incorporated can be identified by the emission wavelength. This technology allows for a natural selection of the base incorporated into the DNA template strand, which is understood to reduce error rates and incorporation bias [172].

2.2.3 Sequence Alignment & Quality Control

Once the base calls have been identified, the sequencing reads were aligned to the genome reference consortium human build 37 (hg19) reference genome using the Burrows-Wheeler Alignment Tool version 0.5.9 [173]. This tool can align short sequence reads against a large reference sequence while adapting to any disruptions and discrepancies between the two sequences [173].

Table 2.1. Clinical Data for all affected study participants[v]

ID	Family	Sex M/F	Age at Diagnosis	Number of IAs	Surgery	SAH Y/N	Other Known Health Issues
Z991*	R1262	F	46	1	No surgery	N	N/A
Z1037*	R1262	F	56	1	Clipping	N	N/A
Z1133*	R1381	F	56	2	Clipping	Y	N/A
Z1137*	R1381	M	52	1	No surgery	Y	Hypertension, Depression, Anxiety
Z1369*	R1381	F	64	1	No surgery	N	N/A
Z994*	R1265	F	50	1	Coiling	Y	N/A
Z995*	R1265	M	63	1	No surgery	N	N/A
Z1148*	R1385	F	47	2	Clipping	Y	Hypertension, Cognitive Impairment
Z1150*	R1385	F	31	1	No surgery	N	N/A
Z1007*	R1276	M	67	1	Clipping	N	N/A
Z1015*	R1276	F	59	1	Clipping	N	N/A
Z1017*	R1276	M	72	1	Clipping	Y	AAA, Hypertension
Z1008*	R1277	M	45	1	No surgery	N	N/A
Z1304*	R1277	M	26	1	No surgery	N	N/A
Z1305*	R1277	M	54	1	No surgery	N	Hypertension, Reflux
Z1353*	R1277	F	51	2	Coiling	Y	Hypertension, Smoker

[v] An asterisk following any Z identification number indicates a participant in which whole exome sequencing has been completed

Z1354*	R1277	M	50	1	No surgery	N	N/A
Z1355*	R1277	F	54	2	Coiling	Y	Hypertension
Z984	R1888	F	68	2	No surgery	N	Hypertension, Breast Cancer
Z1043*	R1888	F	53	1	Clipping	N	Ex-smoker
Z1102*	R1888	F	45	1	No surgery	N	Smoker, Hypercholesterolemia, Crohn's Disease
Z1427*	R1888	M	69	1	No surgery	N	Asthma
Z1054*	R1357	F	46	3	Clipping	N	Hypertension, Smoker, Reflux
Z1069*	R1357	F	34	2	Clipping	Y	Smoker, Hypothyroidism
Z1297	R1357	F	37	1	No surgery	N	Smoker, Hyperlipidemia
Z1365*	R1357	F	57	1	No surgery	N	Hypertension, Pulmonary Fibrosis
Z1366*	R1357	F	45	1	Clipping	N	N/A
Z1159	R1400	F	20	1	No surgery	N	Migraines
Z1360*	R1400	F	36	1	No surgery	N	GERD
Z1364*	R1400	F	47	2	Clipping	Y	Cognitive Impairment
Z1140*	R1383	F	51	1	No surgery	N	Hypertension, DM, Smoker, Depression, Papillary Carcinoma, Hyperlipidemia
Z1317*	R1383	F	31	1	No surgery	N	Migraines
Z1395*	R1893	F	57	3	Coiling	Y	Smoker, GERD, COPD
Z1582*	R2409	F	67	2	No surgery	N	Hypertension, Ex-Smoker, Cardiac Disease, Breast Cancer

Z1584*	R2409	F	67	2	No surgery	N	Hypertension, Smoker, Laryngeal Cancer
Z929*	R1256	F	62	1	No surgery	N	Hypertension, Coronary Artery Bypass Surgery
Z1013	R1256	F	29	2	Clipping	Y	N/A
Z1390	R1256	F	54	1	No surgery	N	Hypertension, Smoker, Heart murmur
Z1405	R1256	F	51	4	Coiling	N	Ex-smoker, Hypercholesterolemia, Asthma, GERD, Fibromyalgia
Z1406	R1256	F	47	1	Clipping	Y	Hypertension, Hyperlipidemia
Z1441	R1256	M	48	1	No surgery	N	Hypertension, Smoker
Z1448	R1256	M	59	1	No surgery	N	Hypertension, Ex-smoker, Hypercholesterolemia
Z1459	R1256	F	47	2	Coiling; Clipping	N	Hypertension, Ex-smoker, D.M, Obesity, Asthma
Z1471*	R1256	F	50	3	No surgery	N	Smoker
Z1501*	R1256	F	61	1	No surgery	N	Hypertension, Ex-Smoker, Hypercholesterolemia, DM, Osteoarthritis, Burkitt's Lymphoma, Diverticulosis
Z1522	R1256	F	44	1	Coiling	N	Smoker, Hypothyroidism, Migraines, Reflux

Table 2.2. Classification of study participants within each familial cohort

Family	Number of Affected	DNA Available	Unaffected Study Participants
R1256	12	11	51
R1262	4	2	0
R1265	2	2	6
R1276	3	3	55
R1277	6	6	7
R1357	5	5	41
R1381	3	3	8
R1383	3	2	14
R1385	3	2	1
R1400	5	3	62
R1888	6	4	23
R1893	3	1	14
R2409	2	2	25
TOTAL	**53**	**46**	**308**

Even with the advancements in WES, there are possible sources of error associated with using this technology. During the library preparation step, ideally only one PCR copy of each original DNA fragment would hybridize to the flowcell. However, this is not always the case, and occasionally, PCR duplicates arise. Any erroneous mutations introduced within these duplicates will then occur more often than appropriate, causing amplification induced errors [174]. Therefore, duplicate reads were removed via Picard version 1.79 [175]. Additionally, the presence of indels within an exome often result in intervals of misalignment and thus inaccurate variant calls. The Genome Analysis Toolkit (GATK) version 1.1-28, indel realigner tool was used to identify these suspicious intervals and realign the reads [176,177]. Systemic technical errors that result from the Illumina instrument itself were also addressed by detecting technical errors using GATK and adjusting the base quality scores accordingly [178].

2.2.4 Variant Calling

After the alignment of sequenced reads to the reference genome, and the quality control steps were performed, TCAG also provided data analysis. Variants were identified by comparing the sequenced exomes to the reference genome; this included both SNVs and indels. Variant calling was implemented via GATK, which utilizes the Bayesian genotype likelihood model to call variants on a per locus basis [178].

Initial variant filters were applied by bioinformatics at TCAG to ensure variant calls were actual genetic variations and not sequencing artifacts. All identified SNVs were subject to variant quality score recalibrations (VQSR) and indels overlapping repeat masked regions were discarded [176]. Variant calls were also hard-filtered based on Quality Depth score, Rank Sum Test, and Strand Bias (SB). The higher the read depth, the more confidence there is in a particular base

call. However, each read contributes to the quality score. Thus regions of the exome that have more in-depth coverage can have artificially inflated quality scores [176]. To account for this, the variant confidence by depth was normalized to less than 10. A filter of less than -20.0 on the Rank Sum test was also applied. Sequences tend to make the most errors at the ends of reads, and therefore seeing an allele only at the end is indicative of a false call. The Rank Sum test evaluates how well the reference allele and the alternative alleles are supported [177]. Finally, sequencing occasionally favors one DNA strand over the other, which can result in an inappropriate amount of evidence in support of one allele [179]. Therefore, a filter of greater than 0.01 SB was also applied.

2.2.5 Annotation

Prior to beginning variant filtering, variant annotation pipelines helped characterize the large amount of genetic information. ANNOVAR was the program used by TCAG to annotate the genetic variants called by GATK and predict their functionalities. There are three significant pipelines used by ANNOVAR. The gene-based annotation identifies any variant affiliation with known genes and their potential functional impact [180]. This provided details such as the location of the variant, the amino acid change, and the functional role of the variant (nonsynonymous, frameshift, etc.) [180]. The region-based annotation identifies variants in specific genomic regions, for example recognizing if a variant resides in a conserved region of the genome. Furthermore, the filter-based annotation provides details on the variant via information retrieved from several external databases [180]. The details provided from the filter-based annotation include whether the variant was previously reported in the Single Nucleotide Polymorphism database (dbSNP), ClinVar, or Online Mendelian Inheritance in Man (OMIM) databases. Allele frequencies were

also provided from the 1000 Genome Project, National Heart Blood and Lung Institute (NHBLI), Exome Aggregation Consortium (ExAC) and Complete Genomics (CG). In addition, mutation prediction algorithm scores such as sorting intolerant from tolerant (SIFT), Polyphen and Combined Annotation Dependent Depletion (CADD) were included to provide predictions on the pathogenicity the amino acid change will have on the protein.

TCAG provided the sequencing data for each of the exomes in annotated variant tables, grouped by family. Each row corresponded to an identified variant within the family and each column corresponded to a specific annotation field. The files were easily transferred to Microsoft Excel, where each table could then be used for variant filtering.

2.3 Variant Filtering

A variant filtering strategy was developed and implemented that could assist in prioritizing the large number of variants generated from WES while excluding those unlikely to be pathogenic (Figure 2.2). Amy Powell, who completed the previous IA project in the Woods lab focusing on two specific families, significantly contributed to the study design and development of the variant filtering strategy used in this project. The approach used considered our current knowledge of IA, previous work in the Woods Lab, filtering strategies of previous WES studies, and suggestions provided by TCAG for variant prioritization.

2.3.1 Preliminary Variant Filtering

The initial filtering steps were suggested by TCAG and were implemented to provide quality control, eliminating sequencing artifacts, and instead identifying only true genetic variants. Any variant with a read depth less than 10 from all exomes within a single family was

excluded. Furthermore, the GATK primary filter that is based on the VQSR provides variants with a "PASS" value if they are more likely to be true genetic variants. Thus, only variants with a PASS value were included for further filtering.

Although the role of synonymous variants in heritable disease is now better understood, synonymous variants were excluded at this point to implement a more stringent filtering strategy, which focused on potentially more damaging variants [181,182]. Recent studies have also demonstrated the importance of noncoding regions of the genome in disease development [183]. However, excluding noncoding variants was another filtering step implemented to reduce the large number of genetic variants prioritized. Finally, any variants with a minor allele frequency greater than 1% in ExAC, NHLBI, 1000 Genomes and CG Datasets were also excluded as IA is a relatively rare disorder.

The late onset of IA, combined with the generally asymptomatic development, makes it difficult to definitively classify an individual as truly unaffected. Small IAs can sometimes resolve or rupture unnoticed, and the brain takes care of the hemorrhagic blood products via normal blood degradation pathways [184,185]. Therefore, even a seemingly negative CT or MRI can inaccurately classify an individual as unaffected. These challenges have made it acceptable to include only affected individuals in WES when studying IA. Thus, variants not shared by all affected individuals were excluded. In two particular families, R1381 and R1277, there are family members who have AAAs but no IAs (Z1371 and Z1304). While this did not classify them as affected participants within the study, there has been sufficient research investigating the link between AAA and IA [186–188]. For this reason, these two families were filtered in two different ways, initially considering Z1371 and Z1304 as affected individuals and then excluding them as affected individuals.

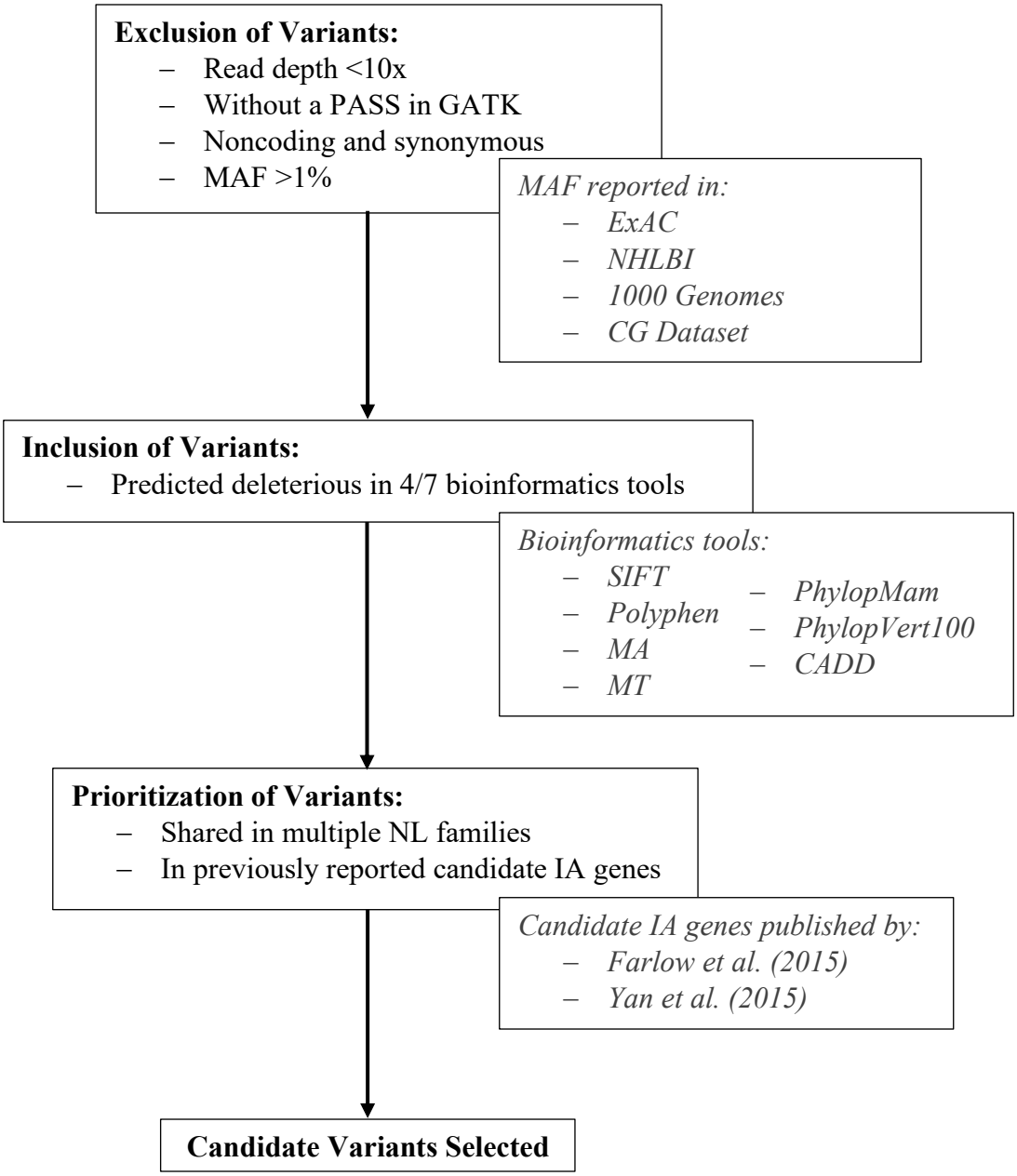


Figure 2.2. Variant filtering strategy

The genome analysis toolkit (GATK) – software version 1.1-28, Minor allele frequency (MAF), Exome Aggregation Consortium (ExAC), National Heart Blood and Lung Institute (NHLBI), Sorting intolerant from tolerant (SIFT), MutationAssessor (MA), MutationTaster (MT), PhyloP placental mammals (PhyloPMam), PhyloP 100 Vertebrates (PhyloPVert100), Combined Annotation Dependent Depletion (CADD)

Further filtering of variants was still required to help prioritize variants for investigation. Different types of genetic variations have variable impacts on a gene. For example, frameshift and stopgain mutations usually cause major disruptions in a gene, whereas nonsynonymous SNVs are thought to have a more moderate impact. For this study, only nonsynonymous SNVs that were predicted deleterious by 4/7 bioinformatic tools were considered further (Table 2.3). The bioinformatics tools used included: MutationAssessor (MA), MutationTaster (MT), SIFT, Polyphen2, PhyloP, and CADD. MA and MT use functional prediction algorithms to provide predictions of the pathogenicity an amino acid change will have on a protein [189–191]. Polyphen2 and SIFT both use scales from 0.1-1.0, with opposite implications. A low Polyphen2 score indicates an amino acid substitution that is likely tolerated, whereas a low SIFT score indicates a deleterious amino acid substitution. PhyloP scores for placental mammals and 100 vertebrates were also considered which measure the evolutionary conservation of an amino acid change [192]. Finally, CADD scores, a tool that integrates various annotations into a single ensemble score, were also considered [193].

Table 2.3. Bioinformatics predictive criteria

Predictive Criteria[vi]	Score
SIFT	<= 0.05
Polyphen2	>= 0.90
MA	>= 1.90
MT	>= 0.5
PhylopMam	>= 2.30
PhylopVert100	>= 4.0
CADD	>= 15

[vi] All nonsynonymous single nucleotide variants (nsSNV) that did not meet at least 4/7 predictive criteria were excluded from further consideration

2.4 Variant Prioritization

2.4.1 Published IA Candidate Genes

Once the variant filtering strategy had been designed and implemented, further prioritization of the remaining variants was required. All variant prioritization strategies described were weighted equally. I searched my list for variants within published IA candidate genes. As mentioned in the introduction, previous studies have also used WES to evaluate the genetic contribution that is leading to IA formation. Farlow *et al.* (2015) performed WES of 45 individuals, representing seven multiplex families that were recruited as part of the FIA study [100]. With the use of this cohort, Farlow *et al.* (2015) published 68 candidate variants in 68 genes that segregated with IA in at least one family. Another study by Yan *et al.* (2015) also performed WES of 42 individuals, representing 12 families of Japanese ancestry [101]. They ultimately published 78 candidate IA variants. I compared both of these IA candidate gene lists with my remaining variants and any variants found in genes that were highlighted in both lists were prioritized for further consideration.

2.4.2 Variants Shared in Multiple NL Families

The reduced genetic heterogeneity of the NL population made it plausible to assume the same genetic variant, or different variants within the same gene, could be contributing to IA development in more than one of our study families. Therefore, any variant that was present in all affected family members, in two or more families was prioritized for further analysis. In addition, different variants in the same gene, present in multiple families, were also prioritized for further consideration.

2.5 Validation and Testing Familial Segregation

To validate the prioritized variants and eliminate any false positives variants introduced by WES, Sanger sequencing was completed on all available affected individuals from each family (Table 2.1). Any variants that did not segregate with all affected individuals within a family were not prioritized. Once the variants were validated, any available unaffected individuals from each respective family were also Sanger sequenced. Locating a variant in a seemingly unaffected relative did not exclude the variant from further investigations as the late onset of IA, and incomplete penetrance, makes it difficult to determine if an individual is an unaffected carrier, or not. Therefore, variants were analyzed individually and needed to be absent from a large percentage of unaffected individuals. The sequencing of both affected and unaffected individuals would provide a more thorough representation of how each variant segregated within a family.

2.5.1 Population Controls

After segregation analysis, 100 NL population controls were Sanger sequenced for any variants that emerged as top candidates to proceed with for functional analysis. By sequencing population controls, the prevalence of these variants in the NL population could be examined, as the reported MAF may not be exemplified in a relatively homogenous population such as this. The controls were available from the Newfoundland Colorectal Cancer Registry Project (NFCCR) and consisted of participants that had no known personal or family history of cancer. These participants were never examined to determine if they had IA or AAA.

2.6 Sanger Sequencing

The Sanger sequencing protocols described here are standard methods used in the Woods lab and were used for all sequencing reactions.

2.6.1 DNA Extraction from Whole Blood

As WES data was already available at the beginning of this project, most DNA extractions were previously performed. However, not all study participants were analyzed in previous projects, and the poor quality of some previously analyzed samples required more DNA extractions to be completed. The extraction of DNA from whole blood was completed in a Laminar Flow Biosafety Containment Cabinet, following the protocol from the Wizard® Genomic DNA Purification Kit (Promega). The first step in DNA extraction requires the lysis of whole white blood cells and their nuclei to access the DNA. To begin, 30 ml of cell lysis solution was added to a 50 ml sterile centrifuge tube. The tube of whole blood was gently rocked until mixed thoroughly and then 10 ml was transferred to the 50 ml sterile centrifuge tube containing the cell lysis solution. The tube was then inverted six times to mix. This mixture was then incubated at room temperature for 10 minutes, inverting six times halfway through incubation to lyse the red blood cells.

After the incubation, the mixture was centrifuged at 2000 x g for 10 minutes in the Sorvall™ Legend™ T+ Centrifuge (Thermo Fisher Scientific) and the supernatant was discarded without disturbing the pellet at the bottom of the tube. The tube was then vortexed to re-suspend the white blood cells and 10 ml of nuclei lysis solution was added prior to the tube being vortexed again. In order to remove any cellular proteins and isolate the DNA, 3.3 ml of protein precipitation solution was added. The solution was mixed using a motorized pipette and the tube

was then centrifuged for 10 minutes at 2000 x g.

The final step in DNA extraction aimed to concentrate the DNA and remove any remaining precipitation solution. The supernatant was transferred to a fresh 50 ml sterile centrifuge tube containing 10 ml of isopropanol at room temperature. The solution was gently mixed by inversion until the white, thread-like strands of DNA formed a visible mass and was then centrifuged at 2000 x g for two minutes. The supernatant was decanted and 10 ml of 70% ethanol was added to the DNA. The tube was gently inverted several times to wash the DNA pellet and the sides of the centrifuge tube. The tube was again centrifuged at 2000 x g for two minutes. The remaining ethanol was carefully aspirated, and the pellet was left to air-dry overnight. Finally, the DNA was dissolved in 400 ul of rehydration buffer and left to incubate at room temperature overnight. The DNA sample could then be mixed, and the concentration could be quantified.

2.6.2 Polymerase Chain Reaction

The PCR completed in the first step of Sanger sequencing allows amplification of the DNA sequence of interest, generating millions of copies. Initially, all primer sets were optimized using a standard temperature gradient to identify their ideal annealing temperature. Specific primer sequences and optimal annealing temperatures can be found in Appendix A. The Invitrogen™ Platinum™ *Taq* DNA Polymerase Kit was used for all PCR (Thermo Fisher Scientific).

All DNA samples were first diluted to standard 100 ng/ul aliquots. A PCR master mix was made containing 10x PCR buffer (200 mM Tris-HCl pH 8.4, 500 mM KCl) (1.5 ul), 100 ng/ul dNTPs (0.375 ul), forward primer at 10 mM (0.5 ul), reverse primer at 10 mM (0.5 ul), 50

mM $MgCl_2$ (0.75 ul), *Taq* DNA polymerase at 5 U/ul (0.15 ul) and 10.225 ul of distilled H_2O. The master mix was made in a 1.5 ml microcentrifuge tube and then 14 ul of mix was individually transferred to each well of a 96 well PCR reaction plate where 1 ul of genomic DNA at 100 ng/ul was added as well. A negative control sample was included in each PCR; containing the PCR master mix without any genomic DNA to ensure there was no contamination. The PCR plate was capped, vortexed and centrifuged for 1 minute at 300 rpms before being placed into the Verti 96-well Thermal Cycler (Applied Biosystems). Specific thermal cycler protocols for each respective primer set can be found in Appendix B.

Agarose gel electrophoresis was used to view the PCR products, ensuring each sample was amplified and the primers were specific to the DNA sequence of interest. The agarose gel was prepared in an Erlenmeyer flask with 50 ml of 1x Tris-Acetate-EDTA (TAE) buffer and 1 g of UltraPure agarose (Thermo Fisher Scientific). This mixture was microwaved for 75 seconds before 3.75 ul of SYBR safe DNA gel stain (Thermo Fisher Scientific) was added. The mixture was then poured into a cassette with a 16-well, 1.5 mm comb and was left to allow the gel to set. After 20 minutes the comb was removed, and each well was loaded with 3.5 ul loading dye and 3 ul PCR product. The first well was reserved for 3.5 ul of loading dye and 1 ul of 100 bp DNA ladder. The gel was then loaded into the electrophoresis chamber and ran at 120 V for 25 minutes. The gel was then viewed via an AlphaImager®EP (SelectScience) under trans ultraviolet light to confirm DNA amplification.

2.6.3 EXOSAP

After ensuring the DNA fragments were amplified, an exonuclease shrimp alkaline phosphatase (exosap) step was performed prior to beginning the cycle sequencing reactions. This

step acts to clean up the PCR products as exonuclease removes any excess primers, while the shrimp alkaline phosphatase removes any unincorporated dNTPs. An exosap master mix was made with 20 U/ul exonuclease (0.25 uL) (Thermo Fisher Scientific), shrimp alkaline phosphatase (0.5 ul) (Thermo Fisher Scientific) and 4.25 ul of distilled H_2O. The mixture was made in a 1.5 ul microcentrifuge tube, and 5 ul was individually transferred to each well of a 96-well PCR reaction plate along with 5 ul of PCR product. Once each well contained 5 ul exosap mix with 5ul of PCR product, the plate was capped, vortexed and centrifuged at 300 rpm for 1 minute. The plate was then placed into the Verti 96-well Thermal Cycler (Applied Biosystems). Specific thermal cycler protocols can be found in Appendix B.

2.6.4 Cycle Sequencing Reactions

After exosap cleanup of the PCR products, cycle sequencing reactions were setup on ice using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific). An individual master mix was made for both the forward and reverse primers in each set. The master mix contained BigDye® Terminator v3.1 Ready Reaction Mix (0.5 ul), 5X Sequencing Buffer (2.0 ul), respective primer at 10 mM (0.67 ul) and 14.83 ul of distilled H_2O. Each well of a 96-well PCR reaction plate was then individually loaded with 18 ul of master mix and 4 ul of exosap product. The plate was then capped, vortexed and centrifuged at 300 rpm for 1 minute before being placed on the Verti 96-well Thermal Cycler (Applied Biosystems). Specific thermal cycler protocols can be found in Appendix B.

2.6.5 Ethanol Precipitation

Prior to loading the cycle sequencing reactions onto the sequencer, a series of ethanol precipitation steps were performed to precipitate the DNA. Initially, 95% ethanol (65 ul) was

added to each well. Subsequently, 125 mM EDTA (5 ul) was added to remove any unincorporated dye terminators that could obscure the sequencing data and help stabilize the extension products throughout the precipitation. The plate was then capped, vortexed and centrifuged at 300 rpm for 1 minute. In order to allow time for precipitation, the plate was then left undisturbed in a drawer overnight, or in the fridge for 30 minutes to 1 hour.

The samples were then removed from the drawer and centrifuged for 30 minutes at 3000 x g. The caps were removed, and the plate was blotted onto paper towel to remove the ethanol without disturbing the DNA pellet. To ensure the removal of any remaining ethanol, the plate was inverted on paper towel and centrifuged at 300 rpm for 1 minute. Thereafter, 70% ethanol (150 ul) was added to each well and the plate was capped, vortexed and centrifuged for 15 minutes at 3000 x g. The caps were again removed, and the plate was blotted onto paper towel to remove the ethanol without disturbing the DNA pellet. The plate was inverted on paper towel and centrifuged at 300 rpm for 1 minute to remove any excess ethanol. To ensure that all of the ethanol had now been removed, the plate was left to dry in a drawer for 30 minutes with the caps removed.

Once it was ensured the plate had completely dried, the samples were re-suspended with the addition of 10 ul of highly deionized formamide to each well. The plate was capped, vortexed and centrifuged at 300 rpm for 1 minute before being placed on the Verti 96-well Thermal Cycler for denaturing (Applied Biosystems). Specific thermal cycler protocols can be found in Appendix B. Once the samples are denatured, the caps are removed, and a 96-well silicon septa matt was placed on top of the plate.

The cycle sequencing reactions were then ready for sequencing and the plate was placed onto the ABI3130 Genetic Analyzer (Applied Biosystems). Once the sequencing was completed,

the data collected was removed from the sequencer and visualized via the sequencing analysis software Sequencher 5.0 (Gene Codes Corporation). This analysis software shows individual bases of the DNA segment and provides the corresponding sequence chromatograms. A reference sequence was then aligned to the DNA sequence of interest and any inconsistencies could be identified.

2.7 Functional Analysis in a *Danio rerio* Model

2.7.1 Candidate Gene Prioritization

In order to determine which variants should be prioritized for further functional analysis in a zebrafish model, a segregation analysis was performed for each candidate variant. To be prioritized, variants needed to segregate with all affected individuals within a family and they also needed to be absent from a high percentage of unaffected individuals within the family.

Up until this point, an unbiased filtering approach was used that did not employ any assumptions about biological processes or pathways. Following segregation analysis, the functionality of each candidate gene was investigated further. The web-based program, Varelect, was used to identify phenotype relationships between candidate genes and IA formation [194]. The program uses information collected from GeneCards and its affiliated databases to draw connections between the genes and phenotype keywords entered by the user. The query terms used to search for IA associations were “intracranial aneurysm”, “aneurysm”, “intracranial”, “subarachnoid hemorrhage”, “blood vessel” and “extracellular matrix”. Varelect then rates the genes as directly or indirectly related to the phenotype keywords and provides information on the relationships drawn [194].

To further aid in candidate gene prioritization, a review of the functional work already completed in model organisms was performed. By using both Mouse Genome Informatics and the Zebrafish Information Network candidate genes could be searched, and any evidence obtained from model organisms could be reviewed [195,196]. Searching a particular gene in either database provides curated reference lists for any research related to these genes in the respective model organism. While the evidence obtained from these model organisms was not used as a definitive filter of any genes, the information was used to review what functional work, if any, had already been completed on a particular gene.

2.7.2 Morpholino Mediated Gene Knockdown in Zebrafish

The functional work completed for this project was done in collaboration with Dr. Curtis French at Memorial University in the Faculty of Medicine. The MO-mediated gene knockdown protocols described here are standard protocols used in the French Lab and were used for all MO injections throughout the project. Prior to beginning injections, 300 nmol MOs were diluted to standard 3 ng/nl aliquots and heated for 10 minutes in a 65°C water bath. Both splice-blocking and translation-blocking MOs were used, and specific oligonucleotide sequences can be found in Table 2.4. There was no MO previously published for *METTL20* and thus a MO was designed via GeneTools software (Gene Tools, LLC).

Initially, each MO was injected at two different dosages into live zebrafish embryos at the one cell-stage. The initial injection was approximately a 3 ng dose, followed by a higher dose at approximately 9 ng. An equivalent dose of standard control MO was also included at each round of injections to ensure any observed phenotype was not the result of off target effects. The standard control MO was purchased from GeneTools (Gene Tools, LLC) and should have very

Table 2.4. Morpholino oligonucleotide details

Gene (Human)	Gene (Zebrafish)	Type of MO	MO Sequence	Published
HAL	*hal*	splice blocking	GCCATTATAAAGGATACAGATGCGA	Han *et al.* (2016)[197]
METTL20	*mettl20*	translation blocking	ATCGATAATTTTGTCCTGTAACCGG	N/A
ADGRV1	*adgrv1*	splice blocking	AAGTTGATCCTGTTACCTCAGATAG	Ebermann *et al.* (2010)[198]
MST1R	*mst1ra (ron1)*	translation blocking	ATGAGTGATGCTATAATAACCTGCA	Huitema *et al.* (2012) [199]
	mst1rb (ron2)	translation blocking	CTAAATGAGTGGCCCAATGGACCAT	Huitema *et al.* (2012) [199]
Standard Control	–	–	CCTCTTACCTCAGTTACAATTTATA	N/A

little biologic activity with no specific target in the zebrafish genome. Injections were completed under a Nikon SMZ18 stereomicroscope, immediately following embryo collection. Following injections, embryos were separated based on dosage, and stored in petri dishes with 1X embryo media, incubated at 28.5°C. Embryo media was made in a 1L bottle containing 50 ml of 20X embryo media, 946 ml reverse osmosis (RO) H_2O, and 4 ml of 500X sodium bicarbonate. The embryo media was made fresh each week and stored at room temperature. The recipe for the 20X embryo media can be found in Appendix C.

At 24 hours post fertilization, embryos were screened using a stereomicroscope to remove any deceased or unfertilized embryos. The embryo media in each petri dish was also removed and replaced with 1-phenyl 2-thiourea (PTU) media mix. PTU mix was prepared in a 1 L glass bottle containing 500 ml of embryo media along with 20 mg of PTU. The mixture was placed on a stir plate and left to mix overnight. PTU media prevents embryo pigmentation, thus the embryos continue to develop transparently, and a hemorrhage phenotype can be readily identified [200]. At 30 hours post fertilization, embryos were removed from their chorions and subsequently at 48 hours post fertilization, embryos were imaged. Embryos were again imaged at 72 hours post fertilization to determine if any additional hemorrhages were observable.

The MOs that produced a significant dose dependent hemorrhagic phenotype were then co-injected with *p53* at a higher dose of MO. The addition of *p53* allows an increase in the dosage of a MO without causing unwanted off target effects [147,201]. An equivalent dose of *p53* MO was also included at each round of injections to ensure any observed phenotype was the result of the MO of interest, and not due to *p53*.

2.7.3 Rescue Experiment

To validate the specificity of the hemorrhagic phenotype characteristic of *hal* morphants, a rescue experiment was performed. Prior to beginning the rescue experiment, mutant and wildtype human *HAL* RNA was synthesized from open reading frame (ORF) expression clones generated by GeneCopoeia (GeneCopoeia, Inc). The ORF cell free expression clones contain a wheat germ cell free vector backbone and T7 promotor (Figure 2.3).

2.7.3.1 Transforming Chemically Competent E.coli Cells

LB agar plates were made by adding 10 g LB agar powder to 250 ml of RO H_2O in a 500 ml flask. The mixture was autoclaved and then spun on a stir plate until cool enough to handle (approximately 20 minutes), being careful not to allow solidification. Ampicillin antibiotic was added, the solution was transferred to petri dishes using a sterile technique, and plates were left at room temperature overnight.

E.coli cells were thawed on ice, and 33 ul of cells were transferred into a 1.7 ml centrifuge tube for each transformation. Three solutions were made, one for each of wildtype DNA, mutant DNA, and a control. Without mixing, 1 ul of DNA (H_20 in control) was added directly to the cells and left on ice for 20-30 minutes. The solution was then heat shocked for 30 seconds at 42°C, then placed on ice for another two minutes. Following the ice, 200 ul of SOC media was added to each solution, and tubes were shaken at 37°C for 1 hour. The tubes were removed from the each solution, and tubes were shaken at 37°C for 1 hour. The tubes were removed from the shaker and spun for three minutes at 3,000 rpm, allowing a cell pellet to form. After removing 175 ul of media, the cells were resuspended by gently pipetting the remaining 50 ul of media. Cells were then spread onto a plate containing carbenicillin, a synthetic form of

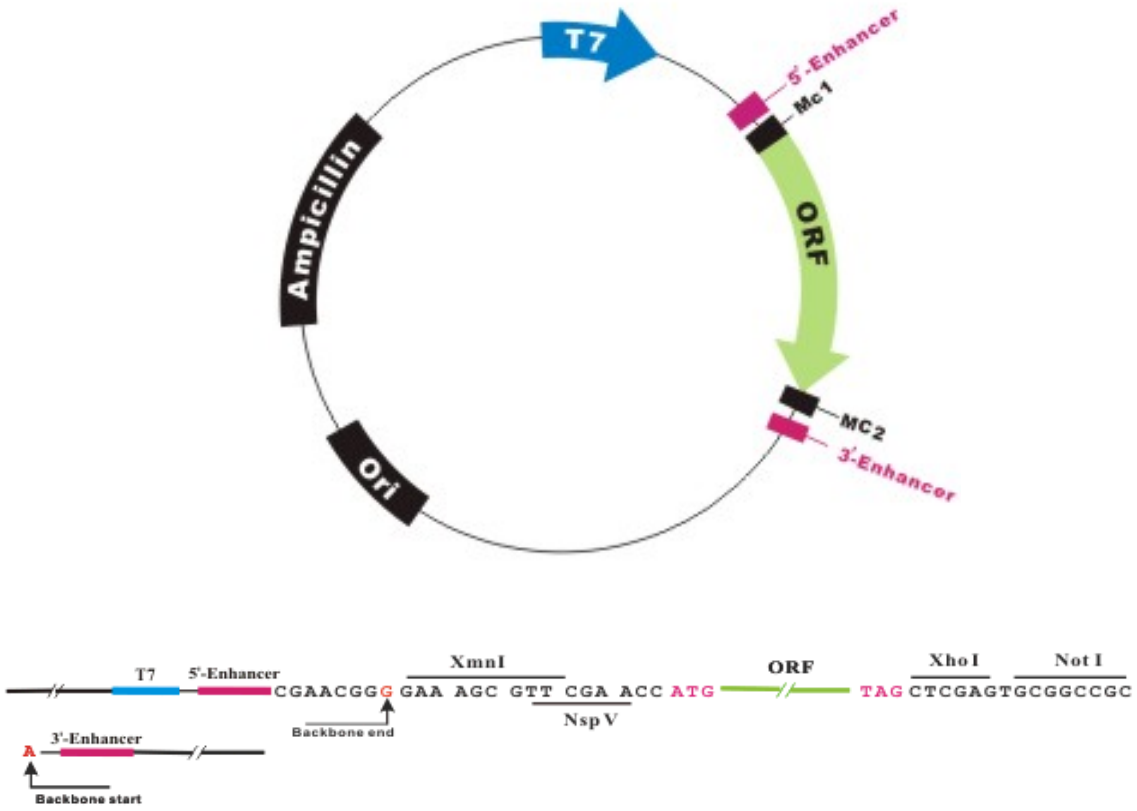


Figure 2.3. Vector information for *HAL* expression clone

The vector information for the wildtype *HAL* transcript and the mutant *HAL* clone is the same. The clone contains a wheat germ cell free vector, the ORF length is 1350 bp, and the antibiotic is ampicillin. The whole plasmid size differs slightly between clones, the wildtype *HAL* plasmid size is 4769 bp, and the mutant *HAL* clone is 4714 bp.

ampicillin that allows for selection of clones that contain the expression plasmid. Plates were covered with parafilm, and stored upside down in incubator at 37°C for 18 hours.

2.7.3.2 Culturing Bacterial Cells

To culture bacterial cells, 3 ml of LB media and 3ul of carbenicillin (a synthetic form of ampicillin) were added to 50 ml centrifuge tubes, one for each of wildtype DNA and mutant DNA. A single colony from the corresponding LB plate was transferred to each tube by adding the entire pipette tip. The tubes were shaken at 37°C for five hours. After shaking, the contents of the tube was added to a 250 ml flask already containing 100 ml of LB media and 100 ul of carbenicillin. The flasks were again shaken at 37°C overnight.

2.7.3.3 Plasmid DNA Purification

DNA purification was completed using the Presto™ Midi Plasmid Kit (Geneaid Biotech Ltd). The first step in plasmid DNA purification was harvesting the cells. The previously cultured bacterial cells were transferred to 50 ml centrifuge tubes and centrifuged at 3,000 x g at room temperature for 15 minutes. During centrifugation, a plasmid midi column was placed in a new 50 ml centrifuge tube. To equilibrate the tube, 5 ml of PEQ buffer was added to the column and allowed to empty by gravity. The flow through was discarded, and the plasmid midi column was set aside.

Once centrifugation was completed, a mixture of 4 ml of resuspension buffer and 40 ul of TrueBlue lysis was transferred to the tube with the cell pellet. The cell was resuspended by vortexing until all traces of the pellet had been dissolved. 4 ml of lysis buffer was added to the resuspended sample then mixed gently by inverting the tube ten times. To neutralize the solution,

4 ml of PM3 buffer was added, and the solution was again mixed by inverting the tube ten times. The entire solution was then transferred to a collection tube with a filter column. The solution was centrifuged at 3,000 x g at room temperature for 2 minutes. The flow through was then transferred to the previously equilibrated plasmid midi column and allowed to empty by gravity flow, discarding the flow through. The plasmid midi column was then washed with 12 ml of PW buffer, and again the column was emptied via gravity flow and the flow through was discarded. The plasmid midi column was then placed in a clean 50 ml centrifuge tube, and 8 ml of PEL buffer was added to elute the DNA.

For precipitation, 6 ml of isopropanol was added to the eluted DNA. The solution was mixed by inverting and then centrifuged at 3,000 x g at 10°C for 60 minutes. The supernatant was then removed, and the DNA pellet was washed with 5 ml of 75% ethanol. The solution was centrifuged one final time at 3,000 x g for 30 minutes at 10°C before allowing the DNA pellet to dry for 10 minutes. Once the pellet was completely dry, the DNA was resuspended in 200 ul RNase free water.

2.7.3.4 Linearization of Plasmid DNA

An individual master mix was made for both the wildtype and mutant DNA in each step. To linearize the purified DNA, 25 ul of DNA, 4ul Reaction Buffer, 2.5 ul NOT1 enzyme, and 8.5 ul RNase free H_2O was added to a 150 ul centrifuge tube, and the solution was incubated at 37°C overnight. Following incubation, a phenol/chloroform extract was performed by adding 160 ul of RNase free H_2O and 200 ul of phenol/chloroform to the incubated solution. The mixture was vortexed for 20 seconds and then centrifuged for five minutes. The upper layer was transferred to a new centrifuge tube. Next, a chloroform extract was performed by adding 190 ul of chloroform

to the upper layer. The mixture was again vortexed for 20 seconds, centrifuged for five minutes and then the top layer was transferred to a new tube. The final step in linearization was ethanol precipitation. 1:10 the volume of sodium acetate and 3x the volume of 100% ethanol was added. This mixture was placed on ice for ten minutes and then centrifuged at 4°C for 20 minutes to precipitate the DNA. The DNA pellet was then washed with 100 ul of 70% ethanol and centrifuged at 4°C for five minutes. The pellet was left to air-dry before resuspension in 10 ul of RNase free H_2O. Agarose gel electrophoresis was used to view the DNA products and ensure the DNA was linearized.

2.7.3.5 RNA Synthesis

RNA was synthesized from linearized DNA using the mMESSAGE mMACHINE™ T7 Transcription Kit (Thermo Fisher Scientific). The first step was assembling the capped transcription reaction at room temperature. A master mix was made for both the wildtype and mutant DNA. The master mix contained 10 ul of 2x NTP/CAP solution, 2 ul of 10x reaction buffer, 2 ul of T7 enzyme mix, 1 ug of linearized DNA, and RNase free H_2O to make 20 ul. The mixture was briefly micro-centrifuged and then incubated at 37°C for two hours. After incubation, 1ul of TURBO DNase was added, the solution was mixed and left to incubate at 37°C for 15 minutes.

A Poly(A) tailing kit was used to polyadenylate the RNA transcript and ensure stability (Thermo Fisher Scientific). The kit adds a >150 base poly(A) tail with the use of *E.coli* Poly(A) Polymerase (E-PAP) and ATP. After incubation, 20 ul of 5xE-PAP buffer, 10 ul of 25 mM $MnCl_2$, 10 ul of 10 mM ATP and 36 ul of RNase free H_2O was added to the transcription reaction. At this point, 1 ul was removed to run on a gel at the end of the experiment to ensure

the polyadenylation reaction was successful. After the addition of 4 ul of E-PAP enzyme, the reaction was left to incubate at 37°C for one hour.

Lithium chloride precipitation was used to recover the RNA and remove unincorporated nucleotides and proteins. 30 ul of nuclease-free H_2O and 30 ul of lithium chloride precipitation solution was added to each reaction. Tubes were mixed thoroughly and chilled overnight at -20°C. The mixtures were removed from the freezer and centrifuged at 4°C for 15 minutes to pellet the RNA. The supernatant was carefully removed, and the RNA pellet was washed with 1 ml of 70% ethanol. Tubes were recentrifuged for another 15 minutes at 4°C to maximize the removal of unincorporated nucleotides. The remaining ethanol was removed, and the RNA was resuspended in 20 ul of nuclease-free H_2O.

Agarose gel electrophoresis with bleach was used to view the RNA products. A bleach gel provides an alternative method of analyzing RNA while preserving its integrity [202]. The agarose gel was prepared as described previously with the addition of 500 ul of bleach for every 50 ul TAE buffer. The final RNA product, and the minus-enzyme control that was reserved before adding the E-PAP enzyme, were run next to each other to ensure polyadenylation had occurred. The gel was loaded into the electrophoresis chamber and ran at 120 V for 25 minutes. The gel was then viewed via an AlphaImager®EP under trans ultraviolet light to assess the results of the RNA synthesis. The RNA concentrations were then quantified and stored at -20°C.

2.7.3.6 MO and RNA Injections

To determine if the previously observed morphant phenotype could be rescued, after the administration of *hal* MO, zebrafish embryos were also injected with the synthesized *hal* wildtype RNA. As a control measure, a separate cohort of embryos were injected with an

equivalent dose of mutant RNA rather than the wildtype, to ensure only the latter demonstrated rescue of the hemorrhagic phenotype and to demonstrate that the mutation causes loss of function.

Prior to beginning injections, the *hal* MO was prepared as previously described. The 300 nmol *hal* MO was combined with *p53* MO and diluted into 4.25 ng/ul aliquots. The aliquots were heated for 10 minutes in a 65°C water bath before injections. Approximately 12 ng of *hal* + *p53* MO was injected into live zebrafish embryos at the one cell-stage. After embryos were injected with *hal* + *p53* MO, RNA injections began. Initially, the RNA was diluted to 200 ng/ul, and injections were aimed at 400 ng/embryo. Injections were performed carefully to inject RNA directly into the cell of each embryo. In total there were four cohorts of embryos, *hal* + *p53* MO, wildtype *HAL* RNA and *hal* + *p53* MO, mutant *HAL* RNA and *hal* + *p53* MO, and uninjected embryos (Figure 2.4).

Once injections were completed, embryos were cared for in the same way as all other MO-mediated gene knockdowns throughout the project. Embryos were stored in petri dishes with 1X embryo media and incubated at 37°C. At 24 hours post fertilization, embryos were screened to remove any deceased or unfertilized embryos. The embryo media in each petri dish was also removed, replaced with PTU media mix, and at 30 hours post fertilization embryos were removed from their chorions. Finally, embryos were imaged at 48 and 72 hours post fertilization to determine if any hemorrhages were observable. The rescue experiment was repeated the same way with a lower concentration of RNA. All subsequent repetitions of the experiment used wildtype and mutant RNA diluted to 100 ng/ul, and injections were aimed at 200 ng/embryo.

2.7.4 Statistics

A Chi-square test with Yates correction was used for statistical analysis of morpholino injections throughout the project. The Chi-square test was calculated using a chi-square calculator for a 2x2 contingency table from the web-based program Social Science Statistics [203]. Associations between cohorts were reported as a chi-square statistic with Yates correction and a corresponding p-value.

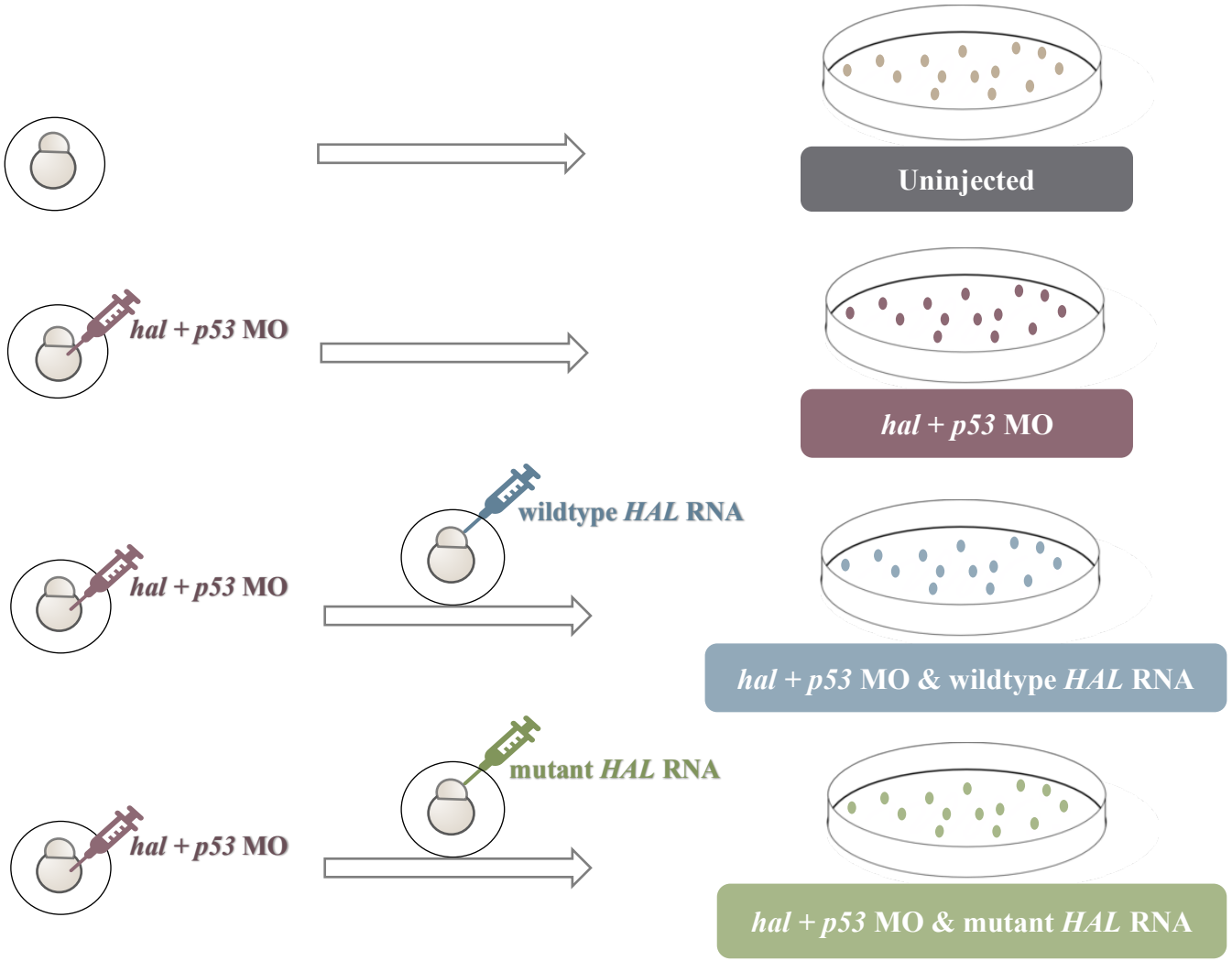


Figure 2.4. MO and RNA injection cohorts

3. Results

3.1 Family Origins

This project focused on a cohort of 13 families that originated from NL. All families had a strong genetic predisposition to IA, with at least two affected individuals. While the number of affected study participants within each family varied, the majority had three or more relatives with IA. This created a cohort of 46 affected individuals within the project; the average age of initial diagnosis is 50.5 years, and the majority of affected individuals are females (80%). The literature suggests that the mode of inheritance of IA is likely variable [61]. However, the majority of families within this cohort demonstrate inheritance patterns consistent with autosomal dominant inheritance. Because the inheritance pattern was more challenging to define in the remaining minority of families, all modes of inheritance were considered a possibility. The following subsections illustrate the pedigrees of all 13 study families, including a general indication of their familial origin and summary of all affected individuals.

3.1.1 Family R1256

The Family R1256 originates from the Happy Valley-Goose Bay area of Labrador (Figure 3.1). With 12 affected individuals, 11 of whom are study participants, R1256 is by far the largest family within the project (Figure 3.2). Only 2/11 affected individuals within R1256 are male, and the average age of initial diagnosis is 50.2 years. The proband, Z1013 has two IAs, one of which was discovered at 29 years of age due to the SAH that ensued, while the second IA was found at age 61. All other affected individuals from R1256 are within the same generation as Z1013 except for Z1459, who is from a subsequent generation. A possible mode of inheritance in

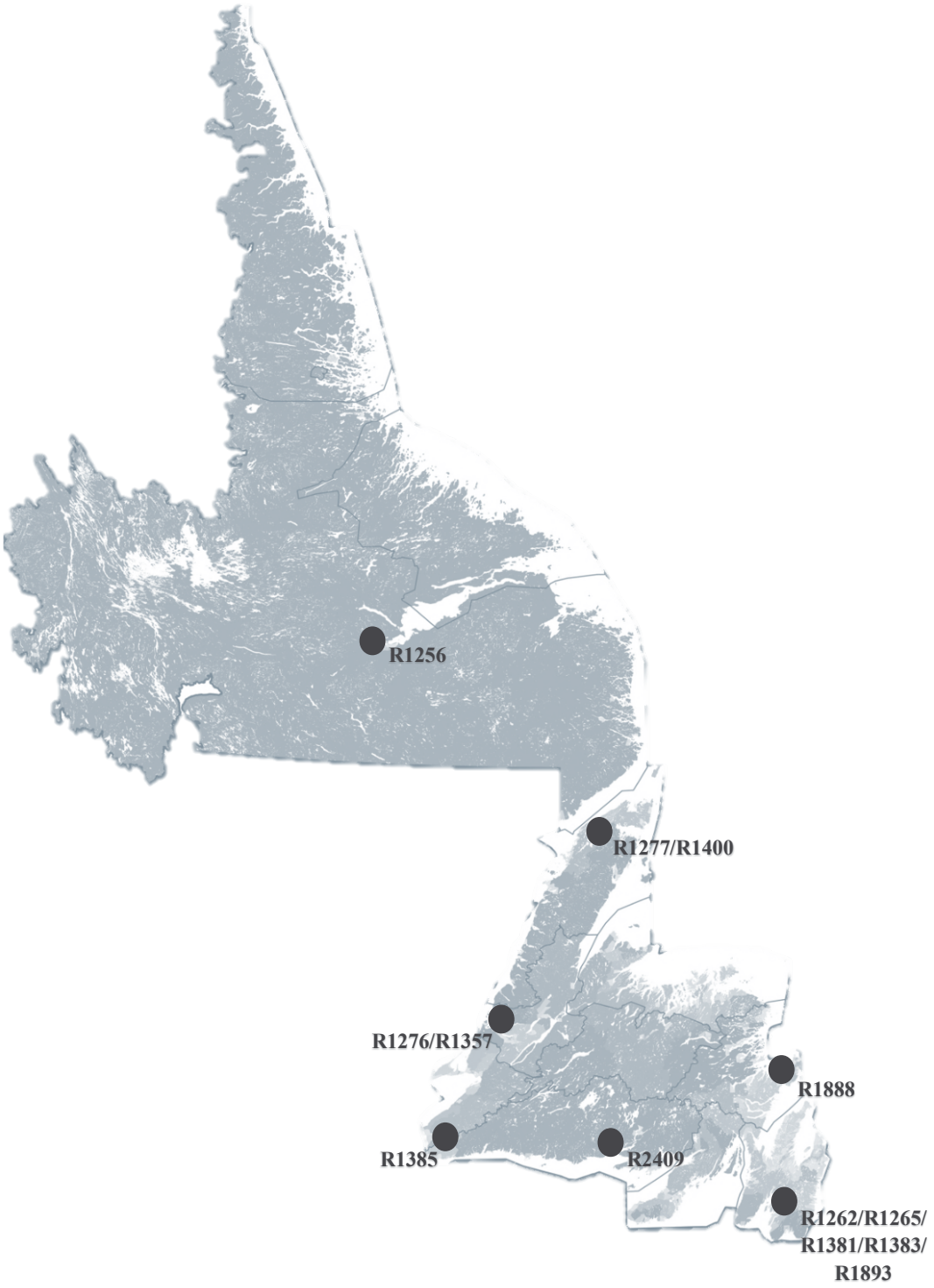


Figure 3.1. Relative geographic family locations

Adapted from: https://commons.wikimedia.org/wiki/File:Canada_Newfoundland_and_Labrador_Density_2016.png
(Public Domain)

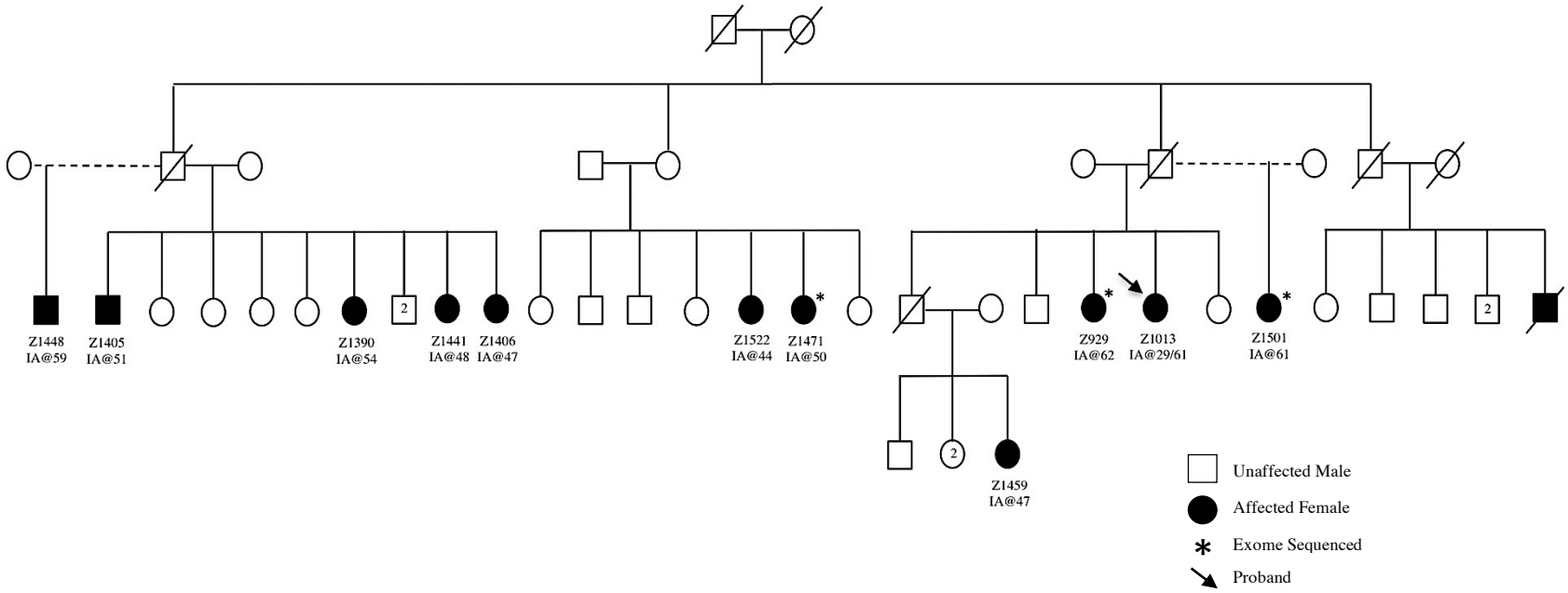


Figure 3.2. Condensed pedigree of Family R1256

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

R1256 is autosomal recessive inheritance. However, it could also be autosomal dominant inheritance with incomplete penetrance. Various other known environmental factors predisposing to aneurysms are prevalent within Family R1256, specifically a history of smoking and hypertension (Table 2.1).

3.1.2 Family R1262

There is a strong history of IA and AAA in Family R1262. Originating from the Avalon Peninsula, the family has four individuals with IA, and two individuals with AAA (Figure 3.3). Only 2/4 individuals with IA are study participants, both of whom are female, the proband Z1037, and her sister Z991. Their age at initial diagnosis was 56 and 46 years, respectively, and neither individual has a known history of any IA associated risk factors. Although not a study participant, the paternal grandmother of Z1037 and Z991 also had IA. Furthermore, while neither of their parents had IA, they both had AAA. The occurrence of AAA within Family R1262 complicates the interpretation of inheritance. If the same genetic variant responsible for the development of IA is also responsible for the paternal AAA, the mode of inheritance is likely autosomal dominant through paternal alleles. However, given the maternal AAA, the possibility of maternal inheritance cannot be ignored.

3.1.3 Family R1265

Family R1265 represents a smaller kindred within the study, which originates from the Avalon Peninsula. DNA was available from two siblings, Z995 and Z994, both who have IA (Figure 3.4). The proband, Z995, was diagnosed at 63 years of age and has no other known IA risk factors. Z994 was diagnosed at 50 years of age when her large basilar tip aneurysm ruptured

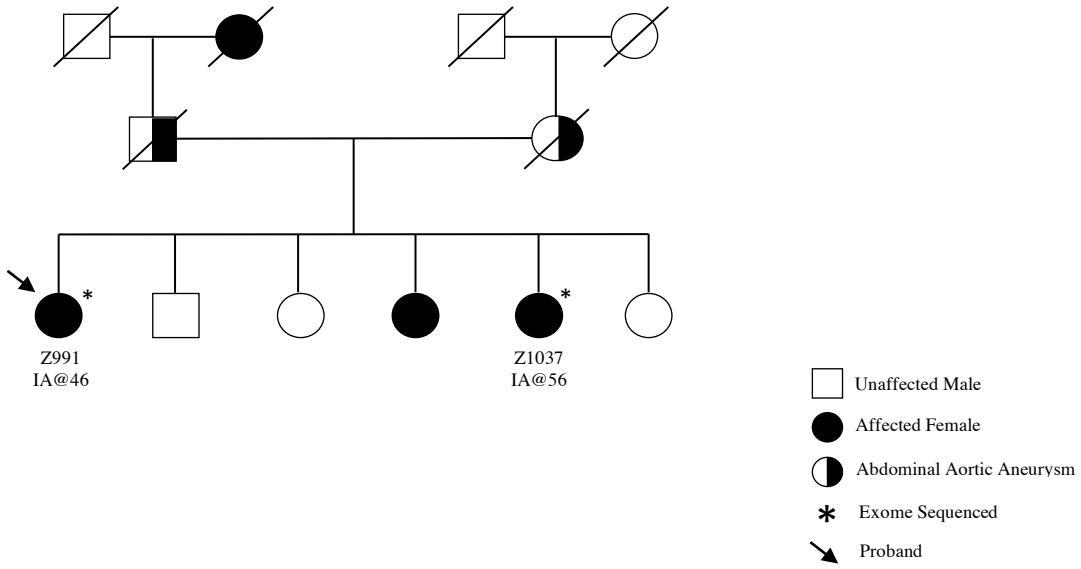


Figure 3.3. Condensed pedigree of Family R1262

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

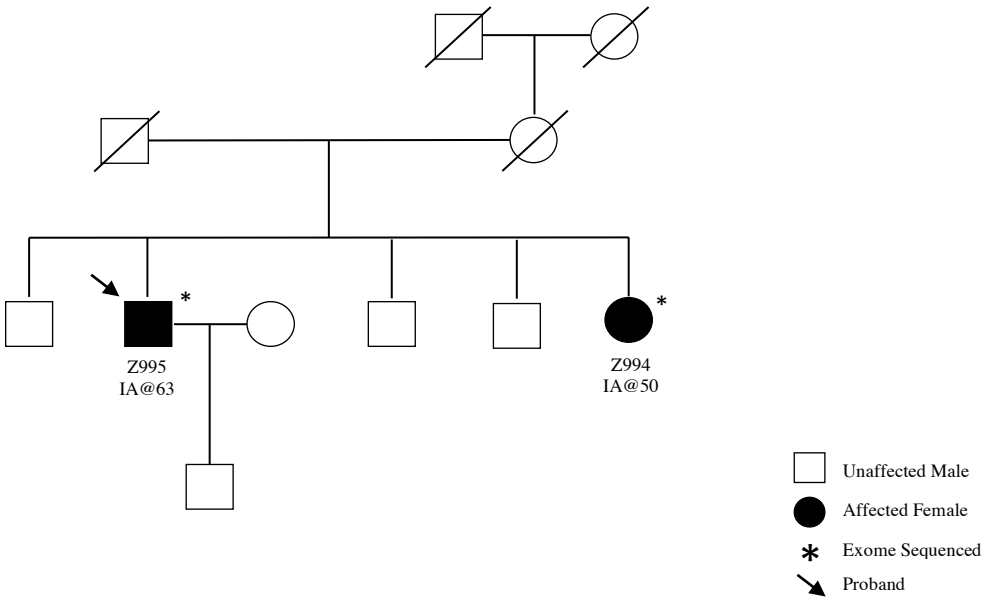


Figure 3.4. Condensed pedigree of Family R1265

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

and she had no other known IA risk factors.

3.1.4 Family R1276

The Family R1276 has origins from the Bay of Islands region, more specifically the north shore. R1276 is one of the larger, well-characterized families within the study. There are three affected siblings Z1007, Z1015, Z1017, and a vast number of unaffected familial study participants as well (Figure 3.5). The average age of initial diagnosis within Family R1276 is 66 years of age, slightly older than the typical age of onset. Individual Z1017 also has AAA in addition to IA, and all three affected siblings underwent aneurysmal clipping.

It was initially understood that there were four affected individuals within Family R1276. Individual Z1415, a first-degree, maternal cousin of the other affected individuals (Z1007, Z1015, and Z1017), was initially classified as an affected family member. Initial investigations identified what appeared to be a tiny middle cerebral artery aneurysm. Subsequent follow-up investigations ruled it was not an IA, and Z1415 was reclassified as unaffected.

3.1.5 Family R1277

Another very large, well-characterized kindred with a strong predisposition to IA is Family R1277. The family has origins in the St. Barbe Straits of the Northern Peninsula. There are eight affected individuals within the family, six of whom are study participants, spanning two generations (Figure 3.6). The inheritance pattern within R1277 may be autosomal dominant inheritance; affected parents have affected offspring without skipping generations. However, the inheritance pattern may also be autosomal recessive. Perhaps the mother and father of Z1304 are related, which is not unheard of in these very small communities. The affected individuals

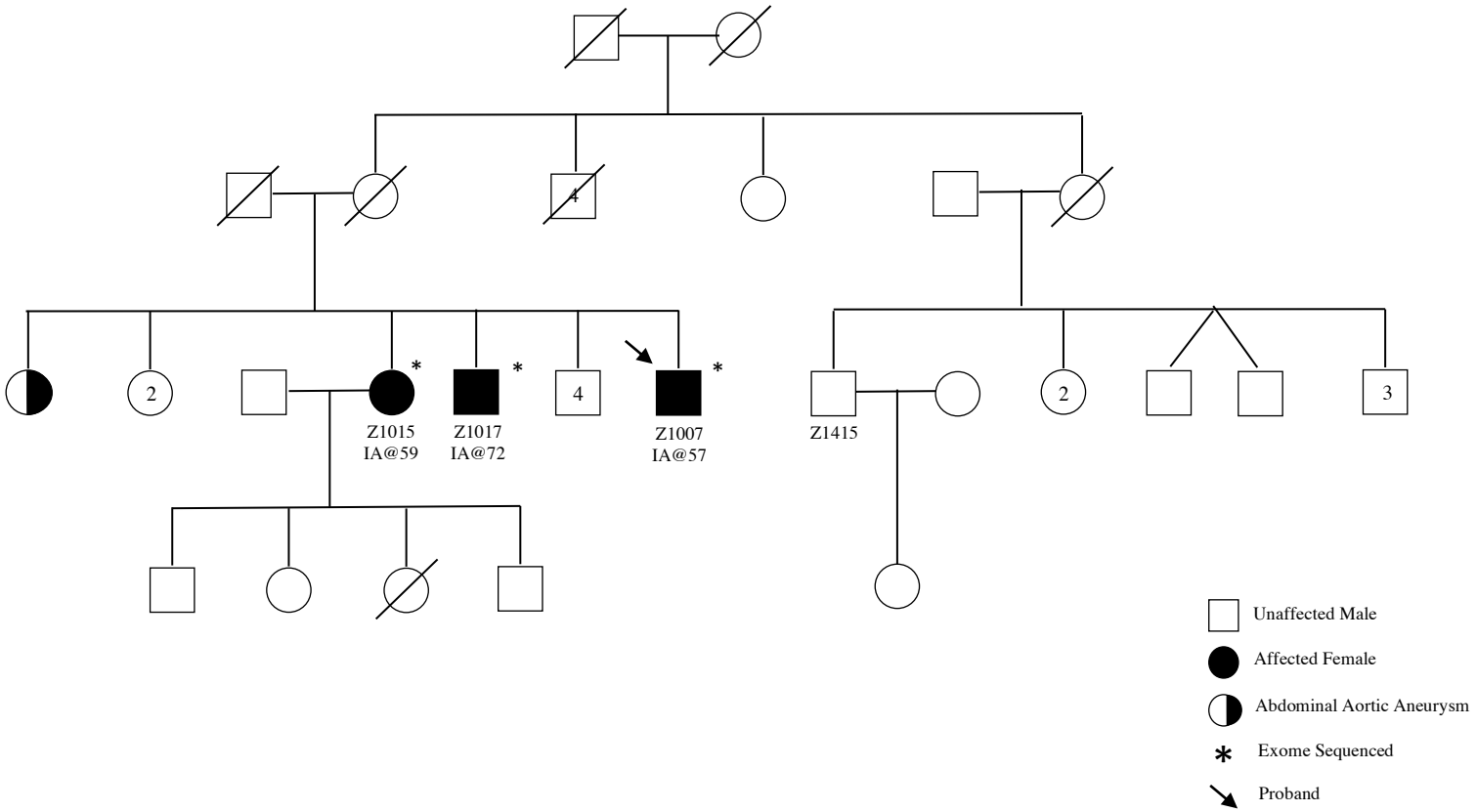


Figure 3.5. Condensed pedigree of Family R1276

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

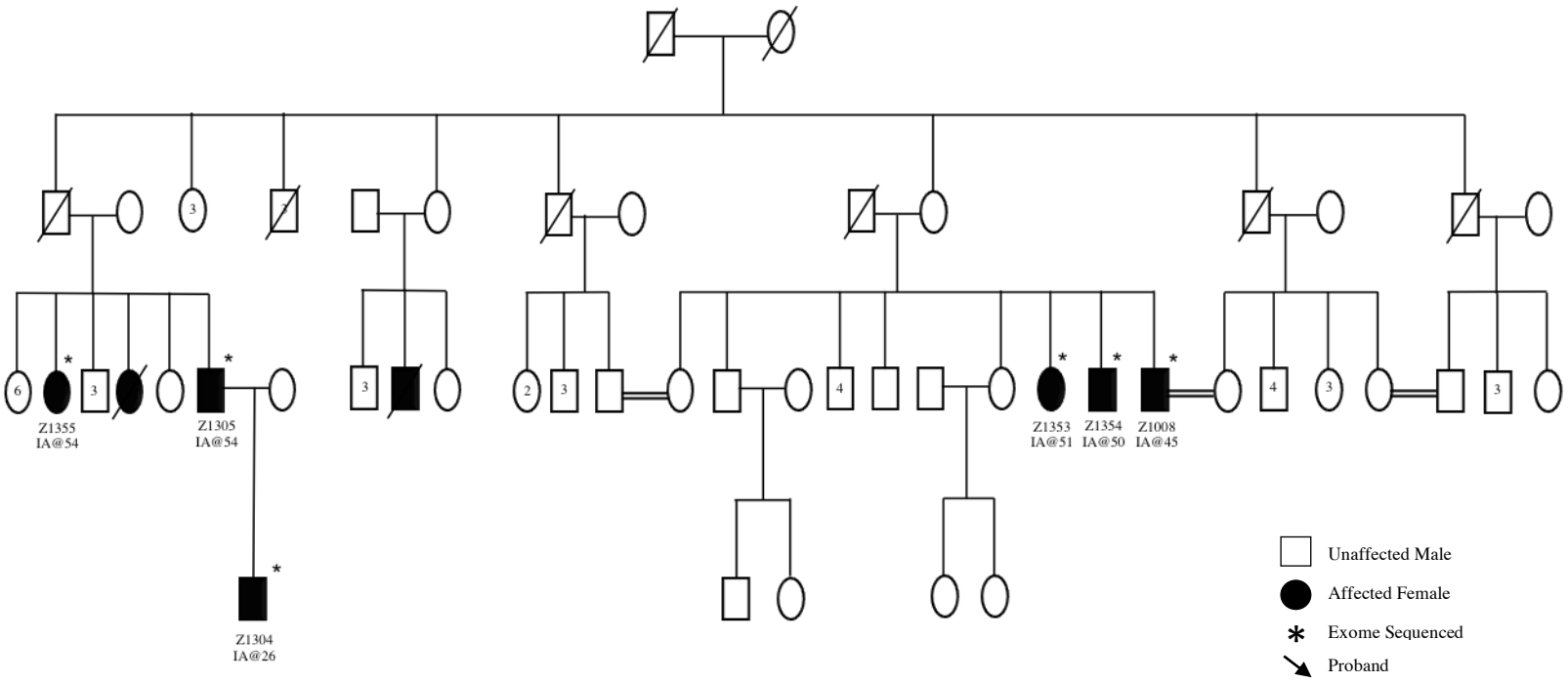


Figure 3.6. Condensed pedigree of Family R1277

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

include four males and two females, with an average age of initial diagnosis of 46.6 years of age. Individual Z1304, son of Z1305 (also affected), was only 26 when his IA was discovered. The majority of affected family members within R1277 did not receive surgical intervention, with the exception of Z1355 and Z1353, both of whom underwent surgical clipping. Both Z1355 and Z1353 had SAH as a result of their aneurysm rupture.

3.1.6 Family R1357

With five family members who have IA, Family R1357 represents another large kindred within the study that originates from the western shores of NL. Many affected individuals within R1357 also have a history of smoking and hypertension (Table 2.1). All of the affected individuals are female, and the average age of initial diagnosis was 51 years of age (Figure 3.7). Two affected individuals were relatively young at initial diagnosis; Z1069 and Z1297, who were 34 and 37, respectively. The mother of Z1297, Z1366, also has an IA and was diagnosed at 45 years of age. The familial proband Z1054 was initially diagnosed at 46 years old and has three IAs. Her maternal cousin, Z1069, also has multiple IAs. One of these IAs ruptured, and a SAH resulted. Affected individuals Z1054, Z1069, and Z1366 had their aneurysms clipped, while Z1297 and Z1365 had no surgical intervention.

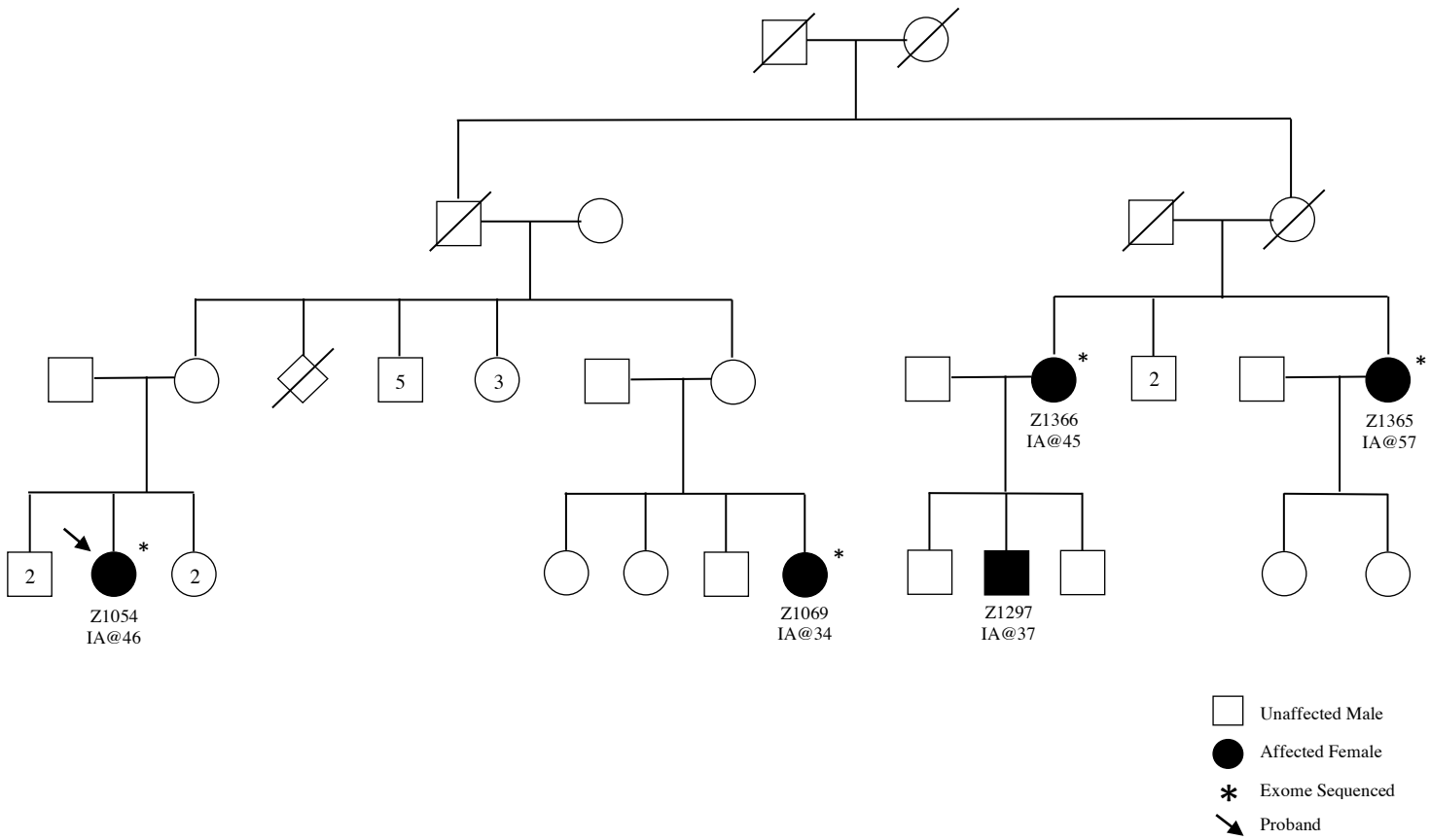


Figure 3.7. Condensed pedigree of Family R1357

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

3.1.7 Family R1381

Originating from the Avalon Peninsula, the inheritance of both IAs and AAAs makes Family R1381 a particularly interesting kindred to study (Figure 3.8). While individuals with AAAs are not classified as affected study participants, the association between these two types of aneurysms is well recognized. The average age of initial diagnosis between the three affected siblings Z1133, Z1137, and Z1369 (two females, one male) was 57.3 years of age. Individual Z1371, an additional sibling, was diagnosed with AAA at 65 years of age. The inheritance of IA within R1381 seems to be coming from the paternal side, as the paternal grandmother of the affected individuals had IA. The father of the affected individuals also had an AAA. The proband Z1133 has two IAs, one of which ruptured, and a SAH followed. The other IA was clipped to prevent further complications. Individual Z1137 also had a SAH due to an IA, before the aid of any surgical intervention.

3.1.8 Family R1383

An additional family with origins from NL's eastern Avalon Peninsula, R1383, has three affected individuals, two of whom are study participants (Figure 3.9). Both Z1140 and Z1317 are female and were diagnosed with IA at 51 and 31, respectively. No surgical intervention was required for either individual. However, Z1140 does have a history of hypertension, DM, and smoking.

3.1.9 Family R1385

Family R1385 originates from the eastern part of the province, in the Bonavista Bay area. The family has three individuals affected with IA, all of whom are female. Of these three

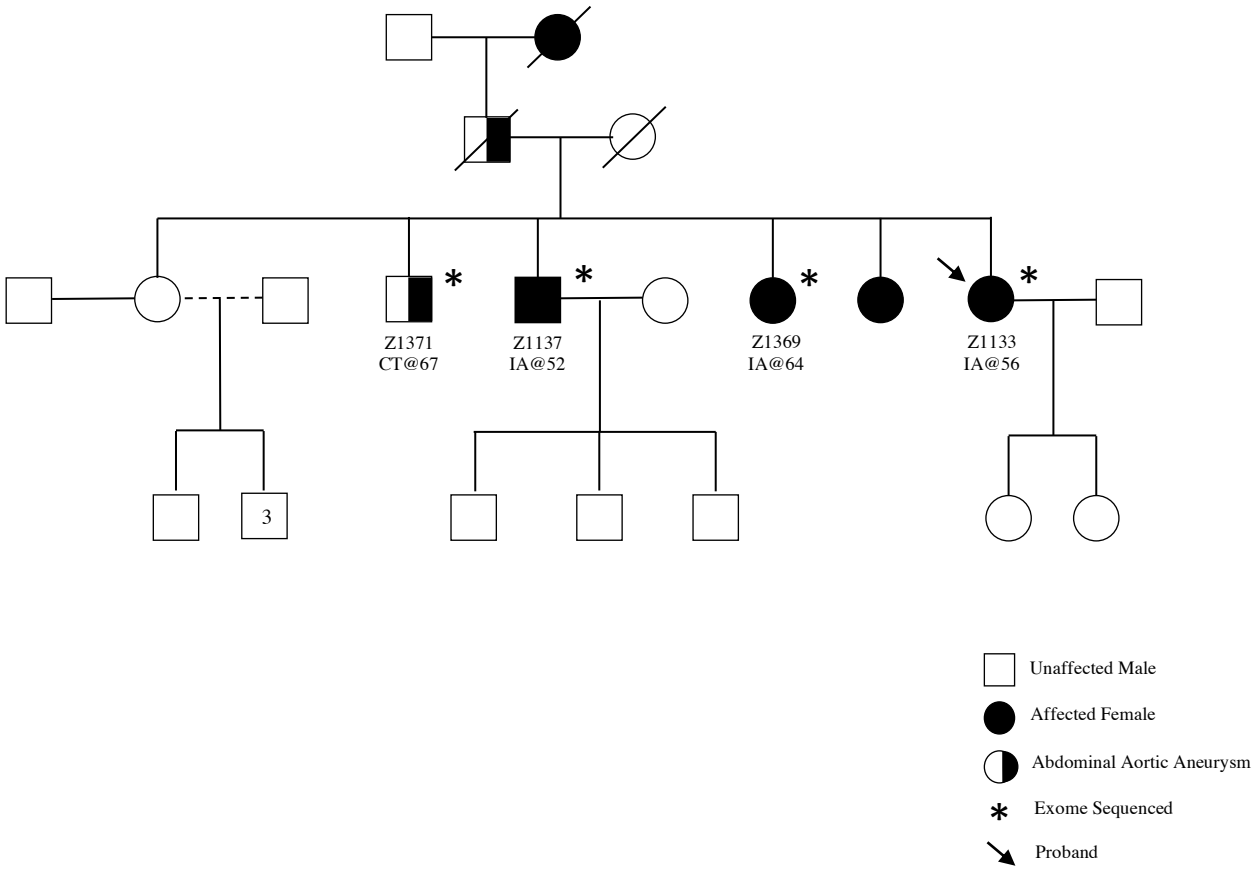


Figure 3.8. Condensed pedigree of Family R1381

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

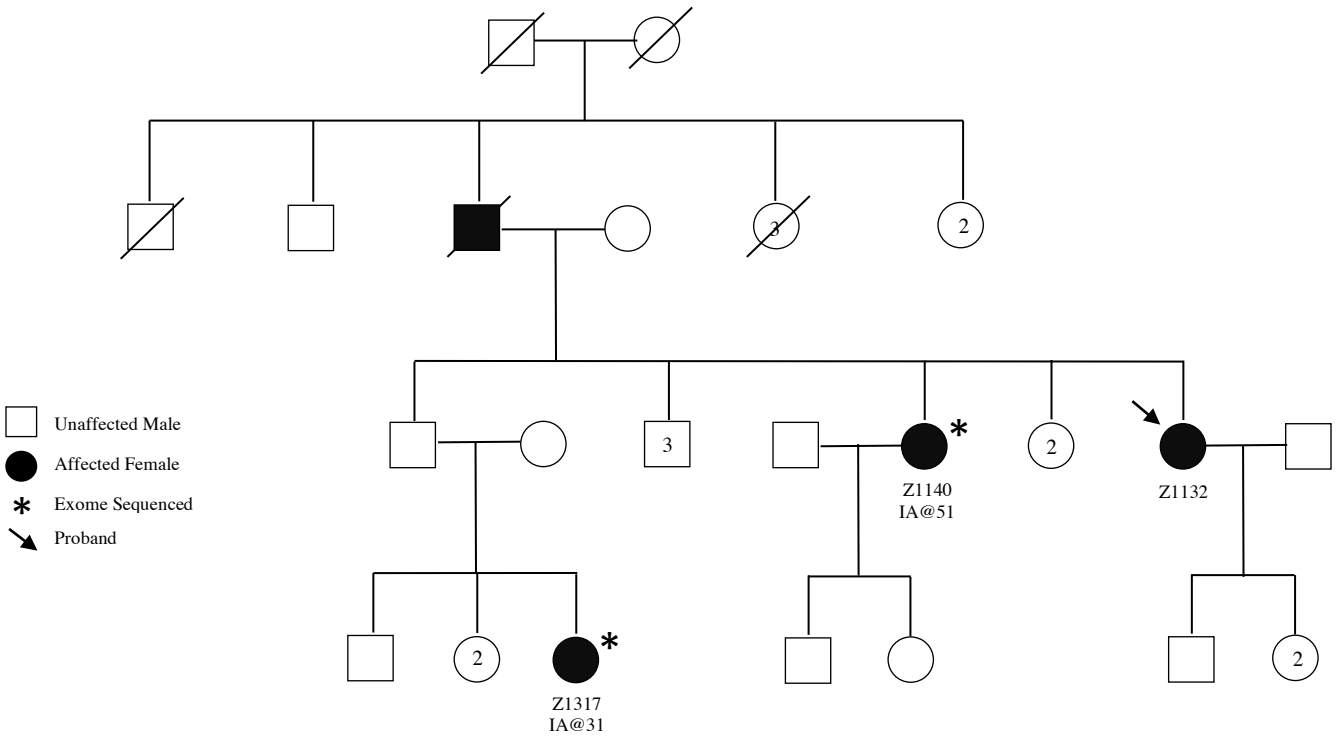


Figure 3.9. Condensed pedigree of Family R1383

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

affected individuals, Z1148 and Z1150 are study participants who were diagnosed with IA at 47 and 31, respectively (Figure 3.10). The proband, Z1148, has two large IA's and experienced a SAH that was repaired by surgical clipping. Individual Z1150 also has a history of hypertension. A small IA was also discovered in individual Z1150, the daughter of Z1148. The third affected individual within Family R1385 is a first cousin once removed of Z1150.

3.1.10 Family R1400

Another large, well-characterized family, R1400, although unrelated to Family R1277, also has origins in the St. Barbe Straits of the Northern Peninsula. Three of the affected study participants, Z1159, Z1360, and Z1364, are female, and their average age of initial diagnosis is relatively young, at 34.3 years of age (Figure 3.11). Individual Z1364 has two IAs, one of which ruptured, causing a SAH. Individuals Z1159 and Z1360 did not require surgical intervention, but Z1364 had her aneurysm clipped. There was no known history of any IA associated risk factors in Family R1400.

3.1.11 Family R1888

Family R1888 originated from the southwestern tip of the island and has a strong family history of IA. Four of six affected family members are study participants, three females and one male (Figure 3.12). The average age of initial diagnosis in Family R1888 is 58.8 years of age. There is a history of some modifiable IA risk factors within the family, such as smoking and hypertension. The proband Z1043 was 53 at initial diagnosis and had her IA surgically clipped.

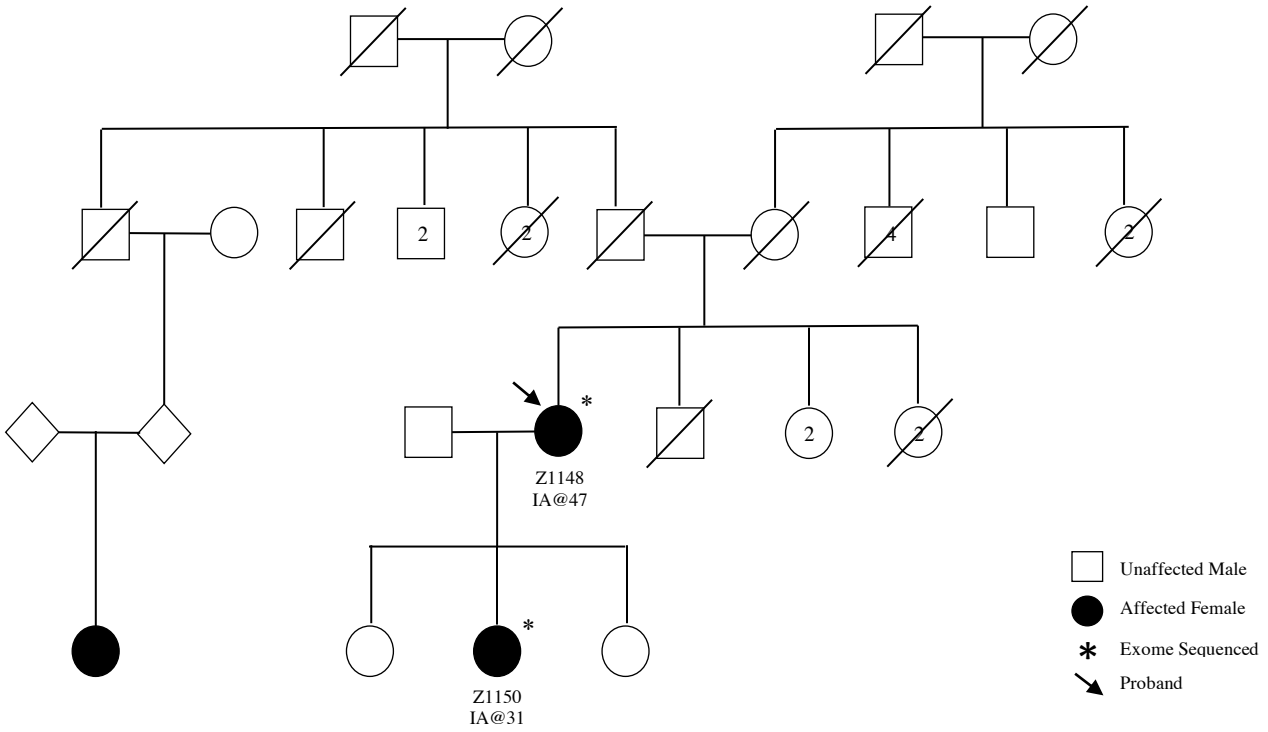


Figure 3.10. Condensed pedigree of Family R1385

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

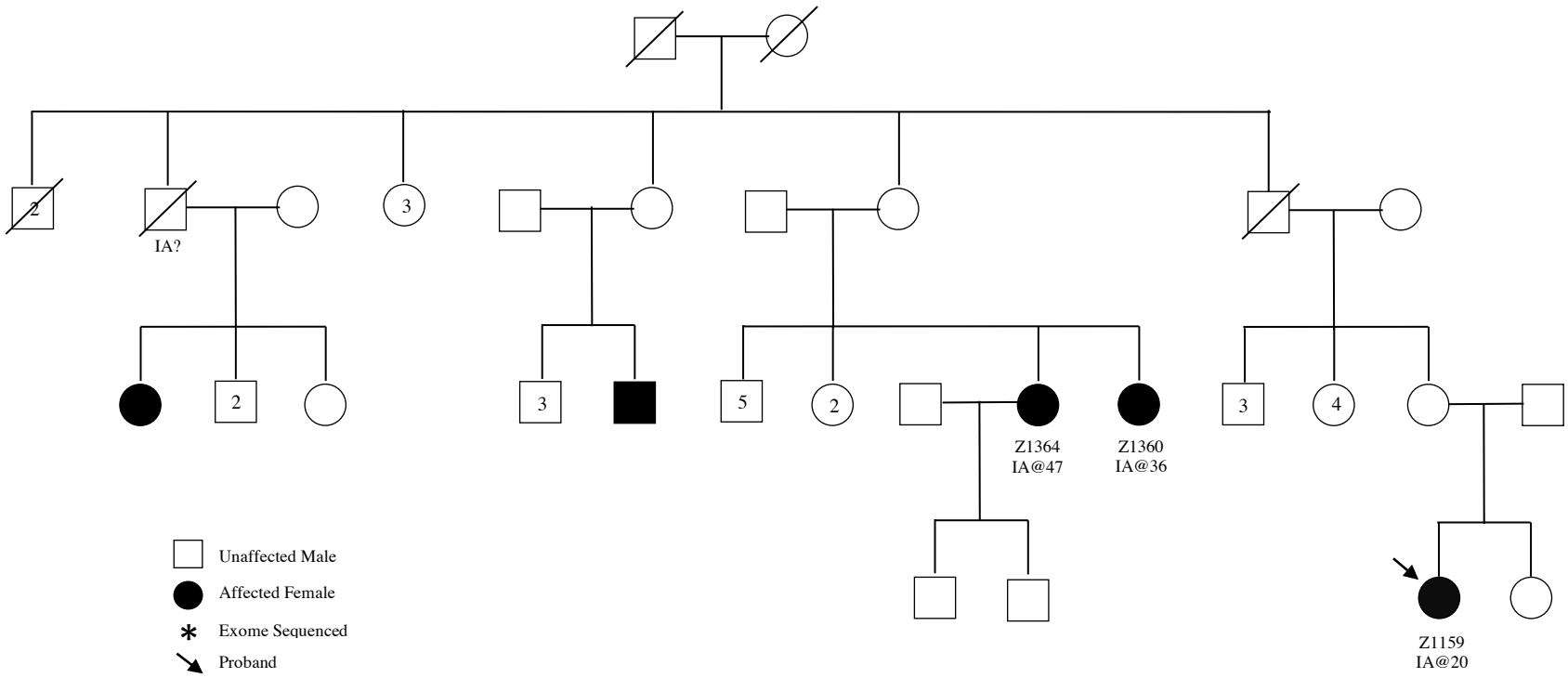


Figure 3.11. Condensed pedigree of Family R1400

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

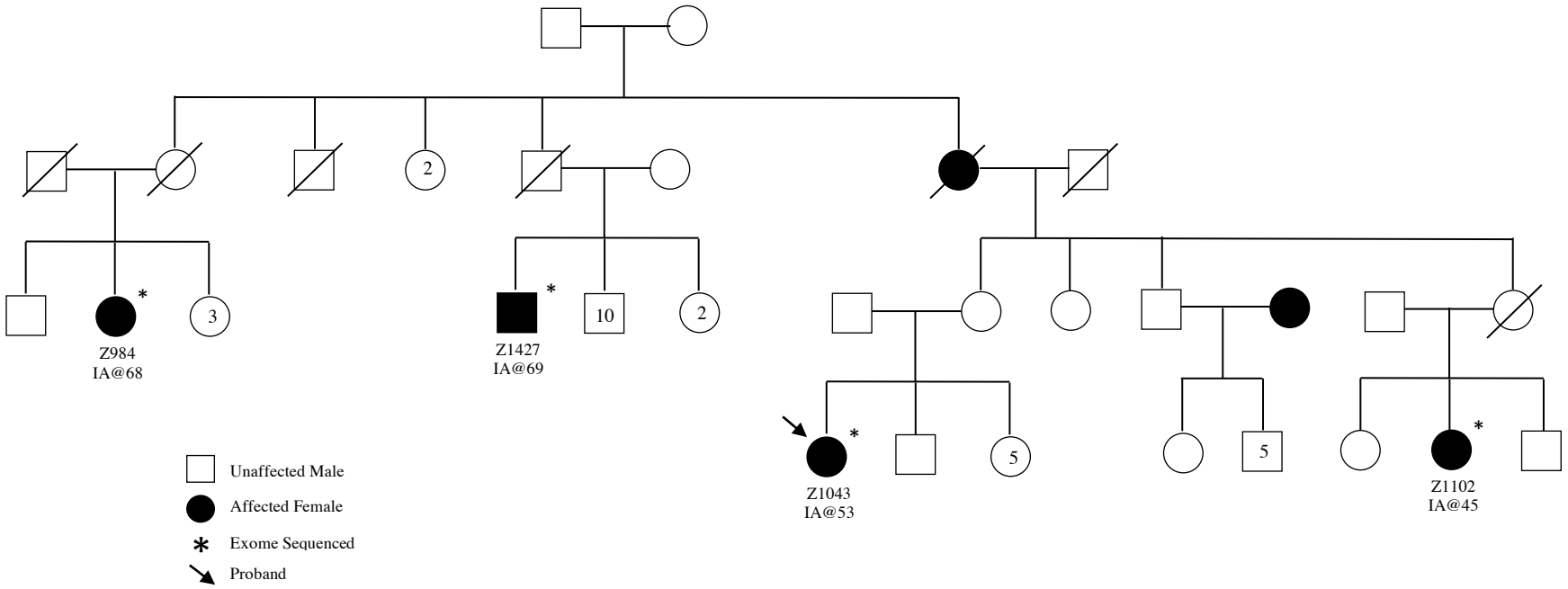


Figure 3.12. Condensed pedigree of Family R1888

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

No other affected individuals within R1888 required surgical intervention for IA. The inheritance pattern within R1888 is difficult to interpret.

3.1.12 Family R1893

Family R1893 is one of the smaller kindreds within the study, who also originates from the Avalon Peninsula. Of the three affected family members within R1893, Z1395 is the only study participant (Figure 3.13). There are also multiple unaffected study participants within R1893. Three IAs were identified in individual Z1395; the first one was recognized when she was 57 years of age. One of her large aneurysms within the anterior communicating artery ruptured, causing a SAH. Individual Z1395 had a history of smoking and underwent surgical coiling to repair her IA. Another affected individual within R1893 is a first-degree relative once removed from Z1395. This individual died of a ruptured IA at only 12 years of age.

3.1.13 Family R2409

Family R2409 has origins from the south coast of NL, in an area known as the Coast of Bays Region. The two affected study participants Z1582 and Z1584 are sisters, both with two IAs that were initially diagnosed at 67 years of age (Figure 3.14). No surgical intervention was required for either affected individual, but they both have a history of smoking and hypertension. There are many unaffected first-degree relatives of Z1584 and Z1582 who are also study participants that have been screened to determine an unaffected phenotype.

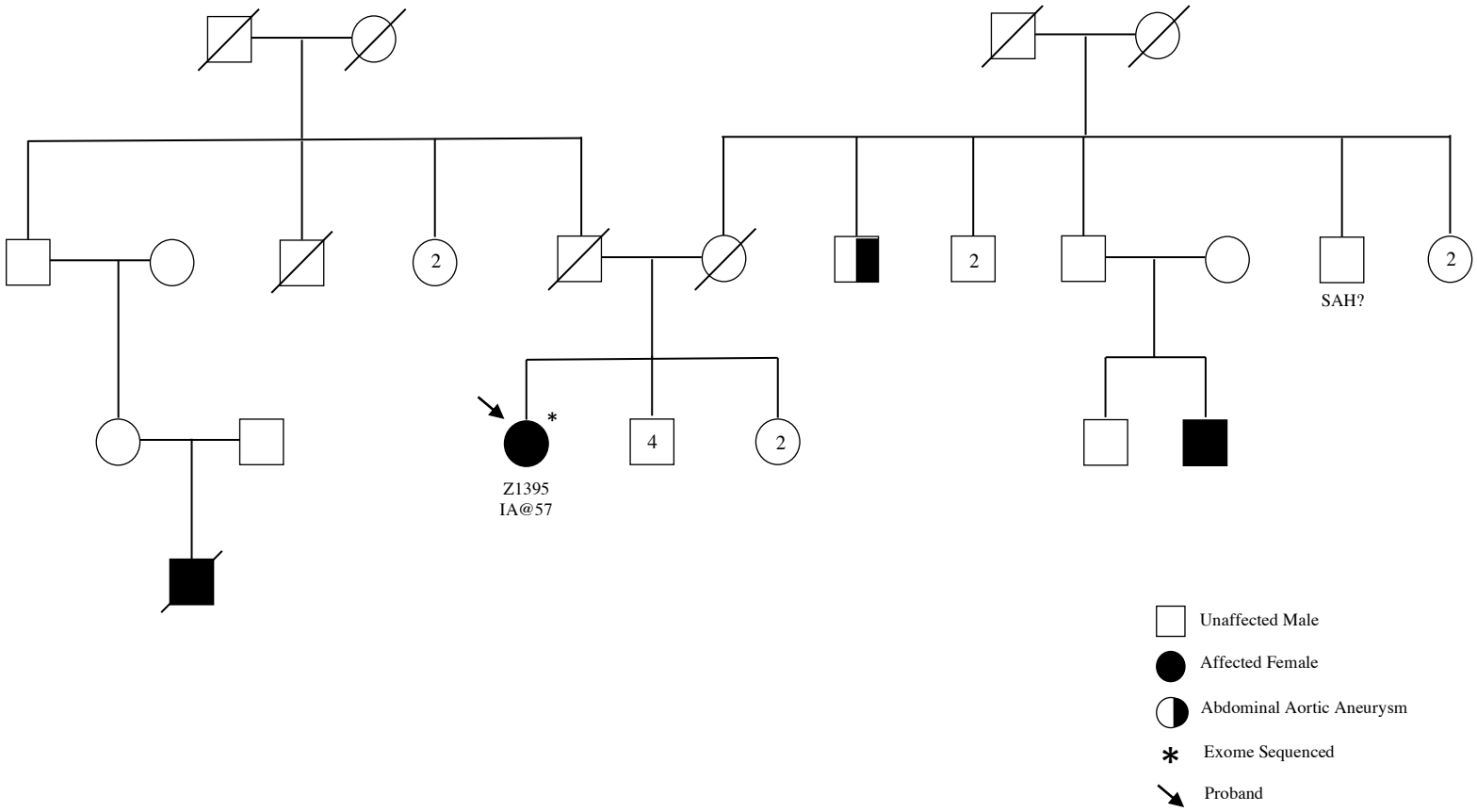


Figure 3.13. Condensed pedigree of Family R1893

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

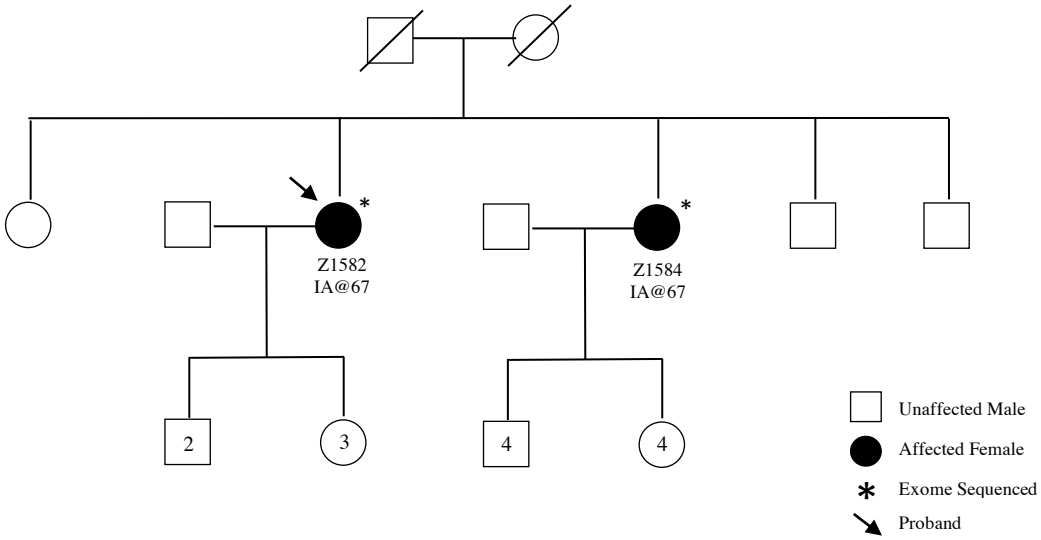


Figure 3.14. Condensed pedigree of Family R2409

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

3.2 Variant Filtering

3.2.1 WES Data

TCAG provided the WES data in annotated variant tables, grouped by family. The first filtering step, implemented to reduce the vast number of genetic variants, was to only include variants with a read depth greater than 10 from all exomes within a single family. While this initial step significantly reduced the number of variants, an average of 42,277 variants per family remained.

3.2.2 Preliminary Variant Filtering

The exact number of variants that passed filtering criteria per family following each of the filtering steps is outlined in Table 3.1. The filtering step most effective in reducing the quantity of genetic variants was only including variants with a MAF less than 5%, which were also shared by all affected individuals within a family. Later it was decided a more stringent 1% allele frequency cut off should be used. Once the filtering criteria was implemented, the genetic variants remaining in each family varied anywhere from 2-72 genetic variants. The greater the number of affected individuals who were whole exome sequenced within a family, the fewer genetic variants that remained. All remaining variants were categorized into high impact or moderate impact based upon the effect of the variant on the coding sequence. The high impact variant group consisted of frame-shift deletions/insertions, stop-gained/lost, and splice-acceptor/donors, whereas the moderate impact group consisted of in-frame insertions/deletions, missense, and protein-altering mutations. The moderate impact group had the majority of variants remaining following initial filtering, ranging anywhere from 2-63 per family. The number of high impact variants was slightly lower, ranging anywhere from 0-11 per family. As

Family R2409 had only two affected individuals whole exome sequenced, it understandably had the highest number of variants remaining post-filtering. There were 72 variants remaining, nine of these were high impact, and 63 were moderate impact variants. As expected with such stringent filtering criteria, not all families had such a large quantity of variants remaining. Family R1357 had the lowest number of variants with only two moderate impacts remaining and no high impact variants.

Table 3.1. Variant filtering results

Study Family	R1256	R1262	R1265	R1276	R1277	R1357	R1381	R1383	R1385	R1400	R1888	R1893	R2409
Read depth > 10x	38436	51912	58309	41663	24943	30125	37960	48252	55167	38064	27735	40386	56649
Shared variants with MAF <5%[vii]	566	976	1094	651	223	242	411	649	1032	401	243	618	1104
Exclude noncoding and synonymous variants	319	574	608	352	122	143	224	367	558	237	137	355	628
"PASS" in GATK	98	308	348	152	17	8	65	122	300	50	9	164	296
MAF <1%	51	138	146	66	11	5	29	54	163	27	7	65	168
Predicted deleterious in 4/7 bioinformatics tools[viii]	**27**	**51**	**54**	**21**	**5**	**2**	**15**	**27**	**58**	**15**	**5**	**29**	**72**
High Impact	2	10	3	3	1	0	4	6	11	1	1	4	9
Moderate Impact	25	41	51	18	4	2	11	21	47	14	4	25	63

[vii] The minor allele frequencies reported in ExAC, NHLBI, 1000 Genomes and CG Datasets were used
[viii] The bioinformatics predictive criterion is listed in Table 2.3

3.2.3 Published IA Candidate Genes

The first step in prioritizing candidate variants was to query the remaining variants from each family for any of the 68 candidate IA genes published by Farlow *et al.* (2015) [100]. Any variants in genes that were identified in both lists were prioritized for further consideration. Four genes were detected in both lists, and the specific variants prioritized are summarized in Table 3.2. The table also includes the corresponding variant highlighted in the Farlow *et al.* (2015) paper for each candidate gene.

The novel, frame-shift mutation *ZNF835* (c.378delC; p.I126fs), and the missense mutation *TAS1R1* (c.1807C>T p.R603C) were detected in all affected family members in Family R1381. The *TAS1R1* (c.1807C>T p.R603C) variant was previously reported in dbSNP as *rs41278022* and has a MAF of 0.6307%. The missense variant *HAL* (c.959C>T; p.T320M), which has a MAF of 0.02472% and *TBC1D7* (c.866C>T; p.P289L), that has a MAF of 0.00244%, were also prioritized as they were detected in all affected family members in R1262 and R1256 respectively. The *HAL* (c.959C>T; p.T320M) variant was reported as *rs200270904* in dbSNP and the *TBC1D7* (c.866C>T; p.P289L) variant was unreported.

In addition to the paper published by Farlow *et al.* (2015), the remaining list of variants was also cross referenced with the list of candidate IA genes published by Yan *et al.* (2015) (Table 3.2) [101]. There were two missense variants identified, one in *DSG1* and the other in *TTN.* The *DSG1* (c.2215A>G; p.I739V) variant is reported in dbSNP as *rs78995670,* has a MAF of 0.4061%, and was detected in all affected family members in Family R1256. Alternatively, the *TTN* (c.22420G>A; p.A7474T) variant is unreported in dbSNP, has a MAF of 0.01256%, and was detected in all affected family members in Family R1265. Table 3.2 also includes the corresponding variant highlighted in the Yan *et al.* (2015) paper for both candidate genes.

Table 3.2. Details of variants in published IA candidate genes

Gene	Family	Chromosome Position	Variant Details (Transcript ID)	rsID	Effect	MAF [ix] (Allele Count)
HAL	R1262	12q23.1	c.959C>T p.T320M (NM_00125833)	*rs200270904*	nsSNV	0.0002472 (30/121346)
	Farlow *et al*. (2015)	12q23.1	c.G1472T p.G491V (NM_002108)	*rs141635447*	nsSNV	0.0008821 (107/121300)
TAS1R1	R1381	1p36.31	c.1807C>T p.R603C (NM_138697)	*rs41278022*	nsSNV	0.006307 (748/118590)
	Farlow *et al*. (2015)	1p36.31	c.C344T p.T115M (NM_177540)	*rs150473918*	nsSNV	0.00009068 (11/121310)
TBC1D7	R1256	6p24.1	c.866C>T p.P289L (NM_001143965)	N/A	nsSNV	0.0000244 (3/119770)
	Farlow *et al*. (2015)	6p24.1	c.C413A p.A138D (NM_001143965)	*rs80189640*	nsSNV	0.002537 (308/121406)
ZNF835	R1381	19q13.43	c.378delC p. I126fs (NM_001005850)	N/A	frameshift deletion	novel
	Farlow *et al*. (2015)	19q13.43	c.G753C p.E251D (NM_001005850)	*rs201104681*	nsSNV	0.0008207 (97/118188)
DSG1	R1256	18q12.1	c.2215A>G p.I739V (NM_001942)	*rs78995670*	nsSNV	0.004061 (493/121408)
	Yan *et al*. (2015)	18q12.1	c.2134C>T p.R712C (NM_001942)	*rs149191001*	nsSNV	0.001697 (206/121398)
TTN	R1265	2q31.2	c.22420G>A p.A7474T (NM_001267550)	N/A	nsSNV	0.0001256 (15/119380)
	Yan *et al*. (2015)	2q31.2	c.28730G>A p.P9577L (NM_133378)	*rs146400809*	nsSNV	0.0009602 (20/20828)

[ix] MAF as reported in ExAC

3.2.4 Variants Shared in Multiple NL Families

Further variant prioritization included any variants detected in all affected family members, in two or more study families. Of the five variants prioritized (summarized in Table 3.3), two of them were nonsynonymous single nucleotide variants, *FARSA* (c.340C>T; p.R114W), and *METTL20* (c.583C>G; p.L195V). The *FARSA* (c.340C>T; p.R114W) variant was previously reported in dbSNP as *rs117345957* and has a MAF of 0.2098%. The *METTL20* (c.583C>G; p.L195V) variant is reported as *rs11051523* and has a MAF of 0.2728%.

Furthermore, three stop-gain mutations were also prioritized. *MST1R* (c.4129C>T; p.Q1377X) is a novel variant, present in two unrelated study families and was previously reported in dbSNP as *rs368400291*. The variant *OR2T35* (c.957_958insTG; p.I320_R321delinsX) was unreported in dbSNP but was identified in three unrelated study families. The MAF of the *OR2T35* (c.957_958insTG; p.I320_R321delinsX) variant, as reported in the exome data, is 0.0298%. Finally, the novel variant *IGSF3* (c.1784G>A; p.W595X) was reported in dbSNP as *rs61730489* but was identified in six study families. Detecting this variant in such a large number of study families prompted the manual search of each variant in dbSNP. This step would avoid prioritizing false-positive variants. The manual search of *IGSF3* (c.1784G>A; p.W595X) detected that the variant was flagged as "suspect" in dbSNP. Variants are flagged as a suspect in dbSNP if located in paralogous regions that could result in mapping artifacts or if there are sequencing errors or computation artifacts. Thus *IGSF3* (c.1784G>A; p.W595X) was not considered for further investigation.

The final variant prioritization step included identifying any different variants, in the same gene, present in multiple families (Table 3.3). Two different variants were prioritized in each of *ADGRV1, MADD*, *PTPN13*, and *TRPM1*. The prioritized variants in *ADGRV1, MADD,*

Table 3.3. Details of variants shared in multiple NL families

Gene	Family	Chromosome Position	Variant Details (Transcript ID)	rsID	Effect	MAF [x] (Allele Count)
FARSA	R1256 R2409	19p13.2	c.340C>T p.R114W (NM_004461)	*rs117345957*	nsSNV	0.002098 (254/121068)
IGSF3	R1256 R1262 R1383 R1400 R1888 R1893	1p13.1	c.1784G>A p.W595X (NM_001542)	*rs61730489*	stopgain	novel
METTL20	R1277 R1385	12p11.21	c.583C>G p.L195V (NM_001135863)	*rs11051523*	nsSNV	0.002728 (331/121356)
MST1R	R1262 R1385	3p21.31	c.4129C>T p.Q1377X (NM_002447)	*rs368400291*	stopgain	novel
OR2T35	R1262 R1277 R2409	1q44	c.957_958insTG p.I320_R321delinsX (NM_001001827)	N/A	stopgain	0.000298
ADGRV1	R1262	5q14.3	c. 1522A>C p. I508L (NM_032119)	*rs61744480*	nsSNV	0.00261 (252/96550)
	R1276	5q14.3	c.6317C>T p.A2106V (NM_032119)	*rs186999408*	nsSNV	0.001622 (195/120242)
MADD	R1262	11p11.2	c.3869G>A p.R1290H (NM_003682)	N/A	nsSNV	novel
	R1265	11p11.2	c.3418G>A p.A1161T (NM_003682)	*rs61755074*	nsSNV	0.001319 (160/121332)
PTPN13	R1381	4q21.3	c.3563C>T p.T1188I (NM_080685)	*rs186651045*	nsSNV	0.001485 (174/117162)
	R1385	4q21.3	c.7336C>T p.R2446C (NM_080685)	*rs367673560*	nsSNV	0.00005014 (6/119672)
TRPM1	R1256	15q13.3	c.413T>C p.L138P (NM_001252020)	*rs191205969*	nsSNV	0.00009802 (12/120766)
	R1381	15q13.3	c.4237G>T p.E1413X (NM_001252020)	N/A	stopgain	0.00000817 (1/120754)

[x] MAF as reported in ExAC

and *PTPN13* are missense mutations. The *ADGRV1* (c.1522A>C; p.I508L) variant was detected in Family R1262 has a MAF of 0.261% and was reported in dbSNP as *rs6174480*. The alternative variant within the *ADGRV1* gene, *ADGRV1* (c.6317C>T; p.A2106V), was detected in Family R1276, has a MAF of 0.1622% and was reported in dbSNP as *rs186999408*. One of the variants prioritized in the *MADD* gene, *MADD* (c.3869G>A; p.R1290H), is a novel variant that was not previously reported in dbSNP but was detected in Family R1262. Alternatively, the variant *MADD* (c.3418G>A; p.A1161T) was detected in Family R1265, has a MAF of 0.1319%, and was reported in dbSNP as *rs61755074*. Both variants in the *PTPN13* gene – *PTPN13* (c.3563C>T; p.T1188I) that has a MAF of 0.1484% and *PTPN13* (c.7336C>T; p.R2446C) that has a MAF of 0.005014% – were previously reported in dbSNP as *rs186651045* and *rs367673560* respectively. Finally, the rare missense mutation *TRPM1* (c.413T>C; p.L138P) was detected in Family R1256, has a MAF of 0.009802%, and was previously reported dbSNP as *rs191205969*. The additional *TRMP1* variant prioritized, *TRPM1* (c.4237G>T; p.E1413X) is a stop-gain mutation that was not previously reported in dbSNP. This variant was detected in Family R1381 and had the lowest MAF of all prioritized variants at 0.000817%.

The stringent steps of the variant filtering strategy combined with the specific variant prioritization approach reduced the thousands of genetic variants to a list of 19 candidate variants. With the exception of *IGSF3* (c.1784G>A; p.W595X), that was excluded from further investigation, all 18 variants listed in Tables 3.2 and 3.3 were Sanger sequenced to test for validation.

3.3 Validation of Candidate Variants

The use of Sanger sequencing in conjunction with WES allows validation that candidate variants are real genetic variants and not false-positive calls that succeeded in passing filtering criteria. In addition to validation purposes, Sanger sequencing can help determine if a candidate variant segregates with all the affected individuals in a family. The following subsections provide the Sanger sequencing results for all 18-candidate variants, in 14 genes, which are summarized in Table 3.4. The initial round of Sanger sequencing was exclusively used to validate each variant, and thus only affected individuals within each family were sequenced. Of the 18 candidate variants, six variants, in six genes (*TBC1D7*, *DSG1*, *FARSA*, *OR2T35*, *MADD*, *TRPM1*) were not validated upon Sanger sequencing and thus were not prioritized for further investigation. A summary of the familial segregation for the additional 12 candidate variants that were validated is described in section 3.4.

3.3.1 DSG1 (c.2215A>G; p.I739V)

The filtering approach applied to the WES data did not employ any assumptions about biological processes or pathways. However, the known functionality of Desmoglein 1 (*DSG1*), in addition to Yan *et al.* (2015) identifying it as a candidate IA gene, made it an exceptionally noteworthy candidate to investigate [101]. As part of the desmoglein family, these cadherin-like adhesion molecules are an integral part of the desmosome [204]. The ability of these intracellular junctions to resist mechanical stress provides strength to various tissues [205]. Mutations in *DSG1* have been associated with the syndrome erythroderma, congenital with palmoplantar keratoderma, hypotrichosis and hyper IgE (EPKHE) which is characterized by extensive dermatitis, metabolic issues and various allergies [206].

The *DSG1* (c.2215A>G; p.I739V) variant was found to segregate with 4/11 affected individuals in Family R1256 via Sanger sequencing (Figure 3.15). Two siblings Z929 and Z1013 are carriers of the variant, as is their half-sister Z1501 (same father). Z1471 is a paternal cousin of Z929, Z1013, and Z1501 and is also a carrier of *DSG1* (c.2215A>G; p.I739V). A niece of Z929 and Z1013 has a daughter, Z1459, who also has an IA, but she does not have the *DSG1* (c.2215A>G; p.I739V) variant. The six additional affected family members, Z1522, Z1406, Z1441, Z1390, Z1405, and Z1448 are not carriers of the variant, and thus *DSG1* (c.2215A>G; p.I739V) was not prioritized for further investigation.

Table 3.4. Summary of segregation for all 18 candidate variants

Gene	Variant Details (Transcript ID)	MAF (Allele Count)		Family	Segregation		Affected Individuals with Variant	Unaffected Individuals with Variant
		ExAC	GnomAD [xi]		Affected	Unaffected		
DSG1	c.2215A>G; p.Ile739Val (NM_001942)	0.004061 (493/121408)	0.003897 (980/251464)	R1256	4/11	N/A	Z929, Z1013, Z1471, Z1501	N/A
HAL	c.959C>T; p.Thr320Met (NM_00125833)	0.0002472 (30/121346)	0.0002148 (54/251426)	R1262	2/2	N/A	Z991, Z1037	N/A
TAS1R1	c.1807C>T; p.Arg603Cys (NM_138697)	0.006307 (748/118590)	0.006115 (1517/248086)	R1381	3/3	4/8	Z1133, Z1137, Z1369	Z1135, Z1145, Z1368, Z1370
TBC1D7	c.866C>T; p.Pro289Leu (NM_001143965)	0.0000244 (3/119770)	0.00002004 (5/249518)	R1256	4/11	N/A	Z929, Z1471, Z1501, Z1522	N/A
TTN	c.22420G>A; p.Ala7474Thr (NM_001267550)	0.0001256 (15/119380)	0.0001212 (30/247460)	R1265	2/2	2/6	Z994, Z995	Z1041, Z1045
ZNF835	c.378delC; p. Ile126fs (NM_001005850)	novel		R1381	3/3 (+AAA)	1/8	Z1133, Z1137, Z1369, Z1371[xii]	Z1370
FARSA	c.340C>T; p.Arg114Trp (NM_004461)	0.002098 (254/121068)	0.002221 (558/251276)	R2409	2/2	N/A	Z1582, Z1584	N/A
				R1256	7/11	N/A	Z1405, Z1406, Z1441, Z1448, Z1471, Z1501, Z1522	N/A
METTL20	c.583C>G; p.Leu195Val (NM_001135863)	0.002728 (331/121356)	0.002947 (741/251462)	R1385	2/2	1/1	Z1148, Z1150	Z1149
				R1277	6/6	1/9	Z1008, Z1304, Z1305, Z1353, Z1354, Z1355	Z1359

[xi] MAF as reported in GnomAD were also included for reference
[xii] Individual Z1371 has an AAA not an IA

MST1R	c.3982C>T; p.Gln1377Ter (NM_002447)	novel	0.00001988 (5/251462)	R1262	2/2	N/A	Z991, Z1037	N/A
				R1385	2/2	1/1	Z1148, Z1150	Z1149
OR2T35	c.957_958insTG p.Ile320_Arg321 delinsX (NM_001001827)	0.000298	0.00003268 (1/30596)	R1262 R1277 R2409	0/2 0/6 0/2	N/A	N/A	N/A
ADGRV1	c. 1522A>C; p. Ile508Leu (NM_032119)	0.00261 (252/96550)	0.002138 (484/226402)	R1262	2/2	N/A	Z991, Z1037	N/A
	c. 6317C>T; p. Ala2106Val (NM_032119)	0.001622 (195/120242)	0.001446 (358/247656)	R1276	3/3	2/52	Z1007, Z1015, Z1017	Z1027, Z1035
MADD	c.3869G>A; p. Arg1290His (NM_003682)	novel		R1262	1/2	N/A	Z1037	N/A
	c. 3418G>A; p. Ala161Thr (NM_003682)	0.001319 (160/121332)	0.001269 (319/251444)	R1265	2/2	N/A	Z994, Z995	N/A
PTPN13	c. 7336C>T; p.Arg2446Cys (NM_080685)	0.00005014 (6/119672)	0.0001300 (15/249026)	R1385	2/2	0/1	Z1148, Z1150	N/A
	c. 3563C>T; p. Thr1188Ile (NM_080685)	0.001485 (174/117162)	0.001333 (330/247592)	R1381	3/3 (+AAA)	4/8	Z1133, Z1137, Z1369, Z1371	Z1134, Z1143, Z1145, Z1368
TRPM1	c.4237G>T; p.Glu1413Ter (NM_001252020)	0.00000817 (1/120754)	0.0001904 (30/249556)	R1381	3/3 (+AAA)	N/A	Z1133, Z1137, Z1369, Z1371	N/A
	c.413T>C; p.Leu138Pro (NM_001252020)	0.00009802 (12/120766)	0.0001202 (30/249556)	R1256	3/11	N/A	Z929, Z1471, Z1501	N/A

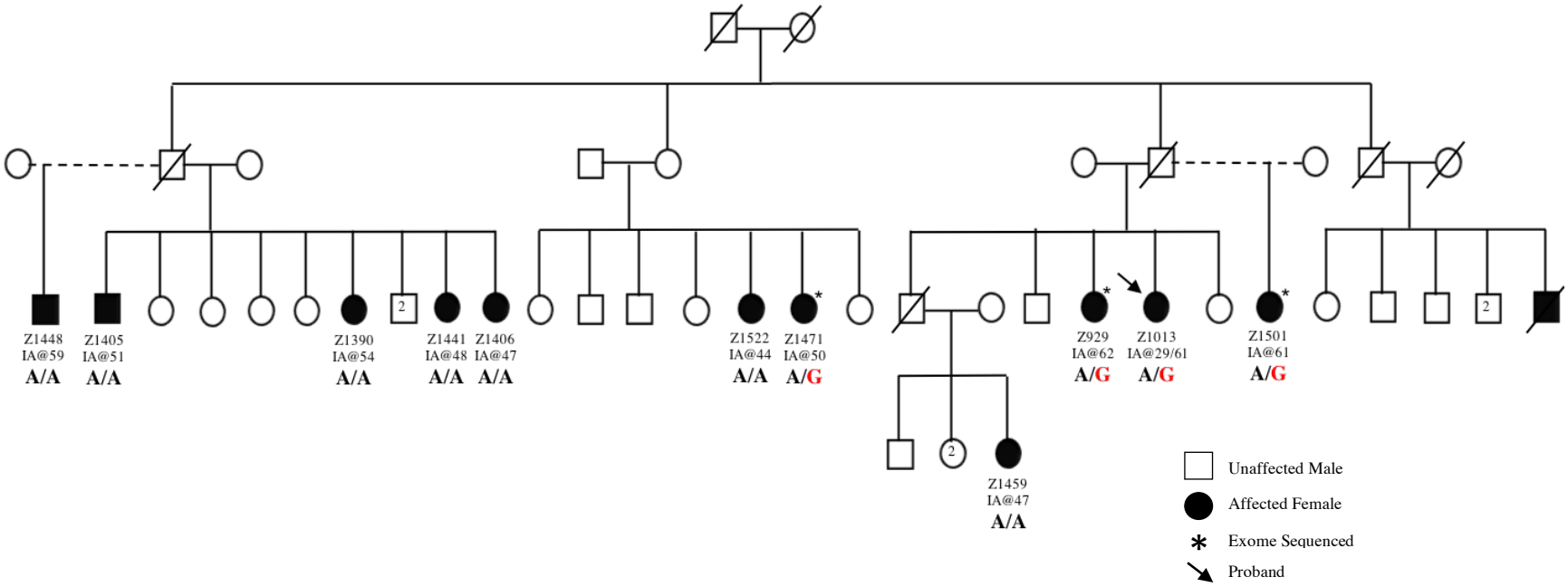


Figure 3.15. *DSG1* (c.2215A>G; p.I739V) segregation in affected family members of R1256

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

3.3.2 TBC1D7 (c.866C>T; p.P289L)

TBC1 Domain Family Member 7 (*TBC1D7*) encodes a subunit of the tuberous sclerosis TSC1-TSC2 complex that is involved in cell growth and differentiation [207]. Previously described mutations in the *TBC1D7* gene are associated with the autosomal recessive growth disorder megalencephaly [208]. Farlow *et al.* (2015) highlighted a mutation in the *TBC1D7* gene as a candidate IA variant that is different from the variant identified in Family R1256 [100]. DNA sequence chromatograms generated from Sanger sequencing revealed *TBC1D7* (c.866C>T; p.P289L) only segregates with 4/11 affected individuals in Family R1256. These four individuals include two half-sisters, Z929 and Z1501 (same father), and two of their paternal cousins, Z1522, and Z1471. The remaining affected individuals, Z1013, Z1459, Z1406, Z1441, Z1390, Z1405, and Z1448, are not carriers of the variant (Figure 3.16). This segregation pattern did not provide confidence in the contribution of *TBC1D7* (c.866C>T; p.P289L) to IA development, and thus it was not considered for further investigation in this population.

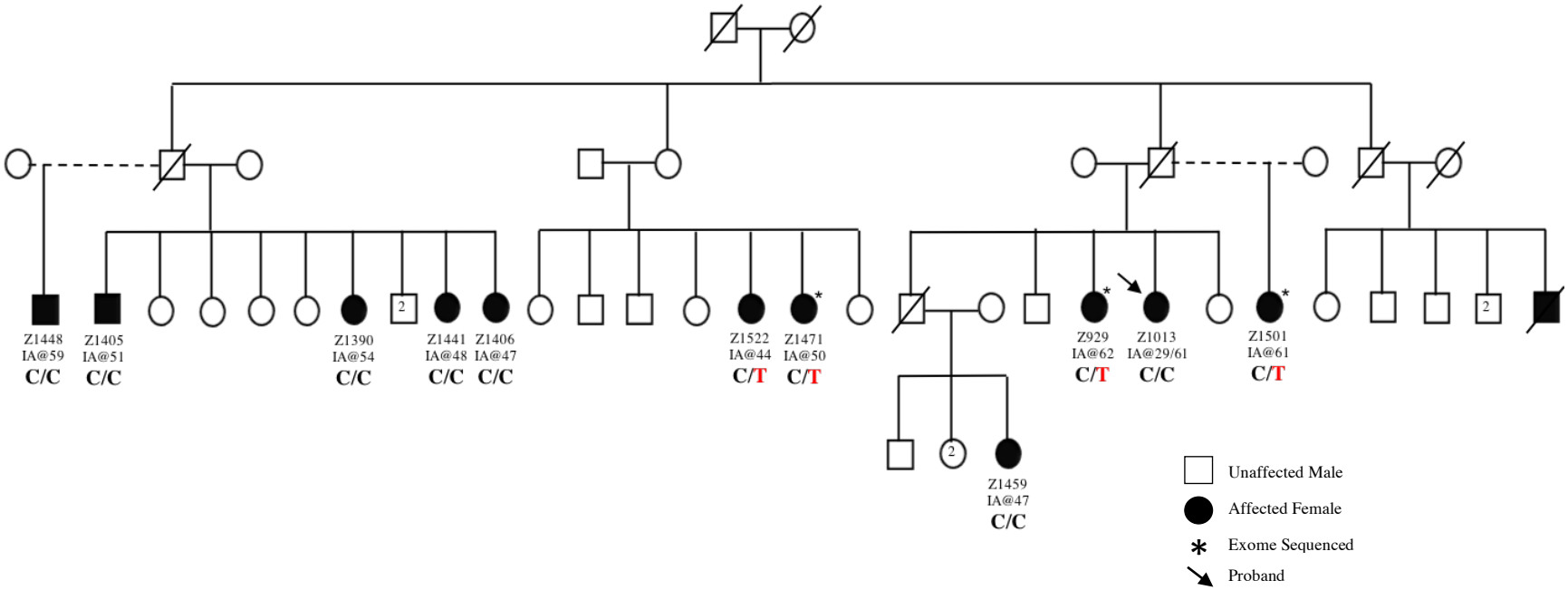


Figure 3.16. *TBC1D7* (c.866C>T; p.P289L) segregation in affected family members in R1256

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

3.3.3 FARSA (c.340C>T; p.R114W)

The variant identified in Phenylalanyl-TRNA Synthetase Alpha Subunit (*FARSA*) (c.340C>T; p.R114W), was found in two unrelated NL families, R2409 and R1256. *FARSA* belongs to a family of enzymes that use amino acids to charge a cognate tRNA [209]. Mutations in *FARSA* are associated with sarcocystosis and myiasis, two related parasitic infections [210,211]. Of the 11 affected individuals Sanger sequenced from Family R1256, the *FARSA* (c.340C>T; p.R114W) variant was found in seven individuals (Figure 3.17). The sisters Z1522 and Z1471 are both carriers of the variant, as well as their maternal cousins, Z1405, Z1406, and Z1441. These three affected siblings have an additional affected sister Z1390 who is not a carrier of the *FARSA* (c.340C>T; p.R114W) variant. Another affected family member, Z1448, is their half–brother (same father), and he too carries the variant. The *FARSA* variant was absent from two other affected siblings, Z929 and Z1013. Conversely, their half-sister Z1501, who is also affected, is a carrier of *FARSA* (c.340C>T; p.R114W). The DNA from Z1459 was also Sanger sequenced, and the *FARSA* (c.340C>T; p.R114W) variant was absent.

In addition to Family R1256, *FARSA* (c.340C>T; p.R114W) was also identified in Family R2409, and Sanger sequencing validated the variant segregated with both affected sisters Z1582 and Z1584 (Figure 3.18). Nonetheless, the segregation of *FARSA* (c.340C>T; p.R114W) within Family R1256 did not provide confidence in prioritizing the variant as a top candidate in IA development.

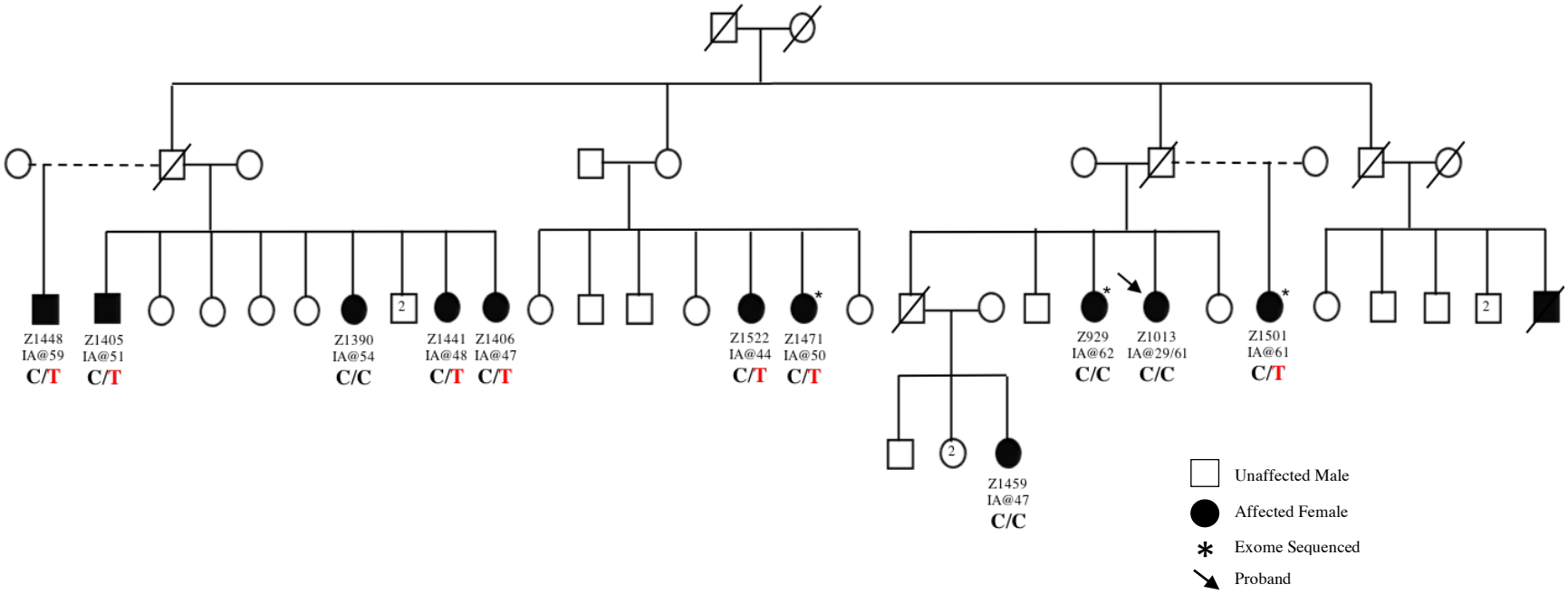


Figure 3.17. *FARSA* (c.340C>T; p.R114W) segregation in affected family members of R1256

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

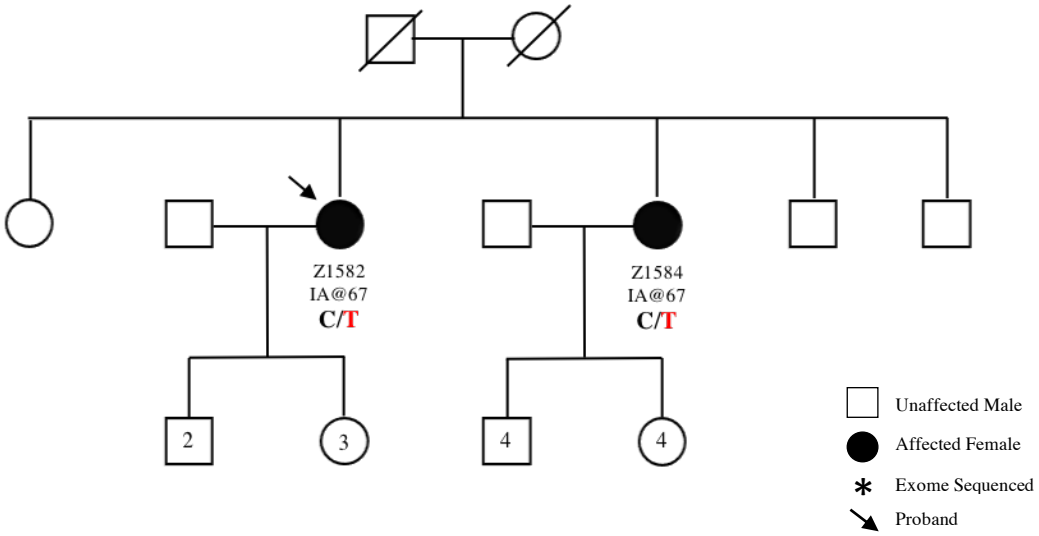


Figure 3.18. *FARSA* (c.340C>T; p.R114W) segregation in affected family members of R2409

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

3.3.4 OR2T35 (c.957_958insTG; p.I320_R321delinsX)

Olfactory Receptor Family 2 Subfamily T Member 35 (*OR2T35*) encodes a family of G-protein coupled receptors, olfactory receptor proteins. There are various families of G-protein coupled receptors, and this particular gene family is one of the largest in the human genome [212,213]. Based on the current literature, no disorders have been associated with the *ORT2T35* gene thus far. WES data identified the *OR2T35* (c.957_958insTG; p.I320_R321delinsX) variant in all affected individuals in three different study families, R1262, R1277, and R2409.

Sanger sequencing of affected individuals in all three families revealed the *OR2T35* (c.957_958insTG; p.I320_R321delinsX) variant was not in any affected individuals. As with any sequencing platform, there are errors associated with TCAG variant calls, and false-positive calls do occur. The variant filtering criteria would ideally eliminate all false-positive calls. However, that is not always a realistic expectation and exemplifies why validating variants with Sanger sequencing is a necessary step in WES projects. The *OR2T35* (c.957_958insTG; p.I320_R321delinsX) variant was not validated via Sanger sequencing and was deemed a false-positive variant call.

3.3.5 MADD (c.3869G>A; p.R1290H) & MADD (c.3418G>A; p.A1161T)

MAP Kinase Activating Death Domain (*MADD*) encodes a death domain adaptor protein that is involved in the activation of the apoptotic signal [214]. Distinctive isoforms of *MADD* play different roles in the regulation of cell death, survival, and proliferation [215]. However, there are no disorders currently associated with *MADD* [216]. Two different mutations in *MADD* passed the variant filtering criteria and segregated with IA in two unrelated study families according to

WES data. The *MADD* (c.3869G>A; p.R1290H) variant was identified in Family R1262 and *MADD* (c.3418G>A; p.A1161T) was identified in Family R1265.

Sanger sequencing confirmed *MADD* (c.3418G>A; p.A1161T) in Family R1265. Both affected individuals Z994 and Z995 are carriers of the variant (Figure 3.19). Despite the validation of *MADD* (c.3418G>A; p.A1161T) in Family R1265, the additional *MADD* (c.3869G>A; p.R1290H) variant did not segregate with all affected individuals in Family R1262. Affected individual Z1037 is a carrier of *MADD* (c.3869G>A; p.R1290H), but the variant was absent from Z991 (Figure 3.20). Because variant prioritization criteria required all affected individuals to be carriers of the variant of interest, *MADD* variants were not prioritized for further investigation.

3.3.6 TRPM1 (c.413T>C; p.L138P) & TRPM1 (c.4237G>T; p.E1413X)

Two different variants in the gene Transient Receptor Potential Cation Channel Subfamily M Member 1 (*TRPM1*) passed all variant filtering criteria and were found in two unrelated NL families. WES data identified the variant *TRPM1* (c.413T>C; p.L138P) in Family R1256, and the variant *TRPM1* (c.4237G>T; p.E1413X) in Family R1381. *TRPM1* encodes a cation channel protein that is thought to play a role in melanin synthesis [217,218]. The gene has various identified biological functions. Decreased expression of the *TRPM1* protein is used as a marker of melanoma metastasis and furthermore, autosomal recessive congenital stationary night blindness is caused by specific *TRPM1* mutations [219–221].

The DNA sequence chromatograms generated from Sanger sequencing the DNA of affected individuals in Family R1381 validated the *TRPM1* (c.4237G>T; p.E1413X) variant. This variant segregated with all three affected siblings Z1133, Z1137, and Z1369. An additional

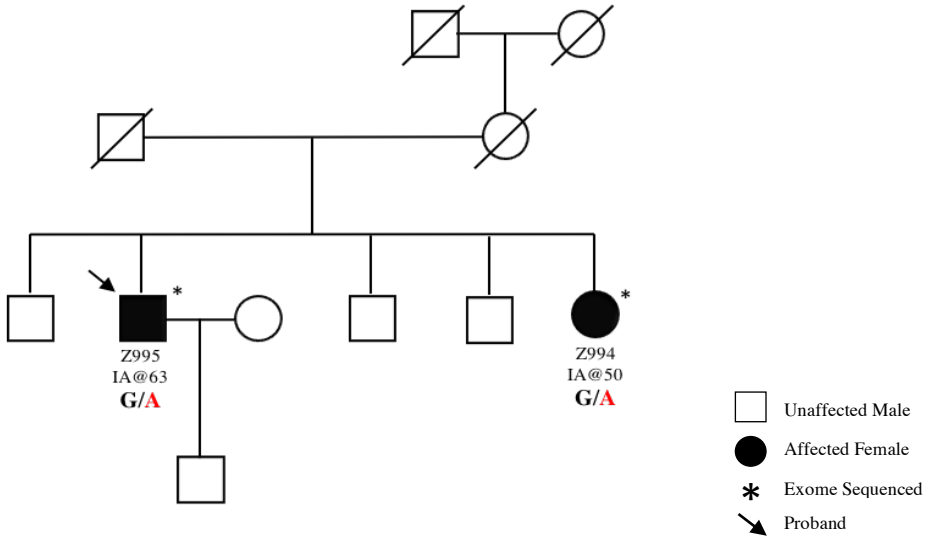


Figure 3.19. *MADD* (c.3418G>A; p.A1161T) segregation in affected family members of R1265

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

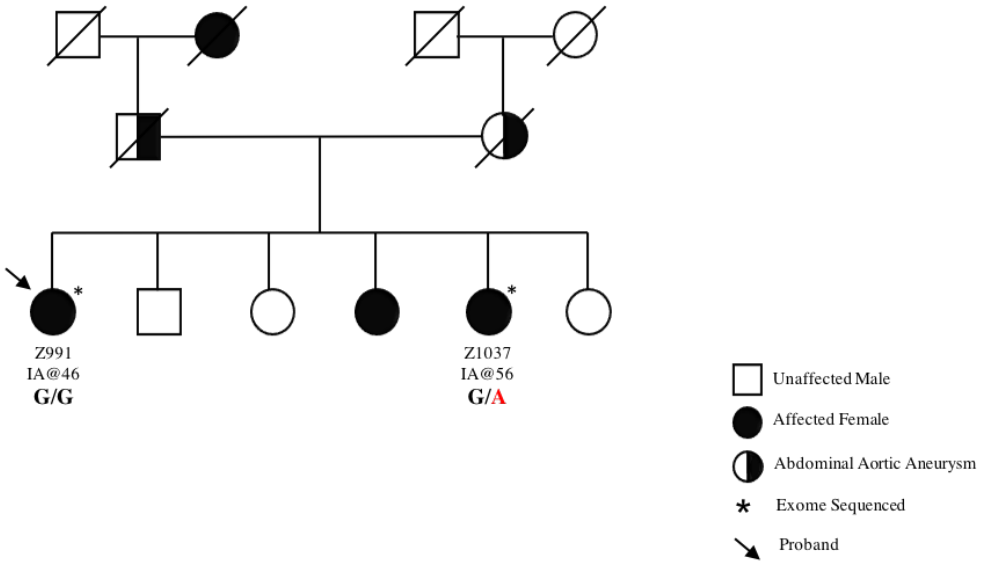


Figure 3.20. *MADD* (c.3869G>A; p.R1290H) segregation in affected family members of R1262

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

sibling, Z1371, who has AAA, is also a carrier of the variant (Figure 3.21).

Sanger sequencing was then completed in Family R1256 to validate the additional variant *TRPM1* (c.413T>C; p.L138P) (Figure 3.22). The DNA was available from 11 affected individuals in Family R1256. Sanger sequencing identified that only 3/11 affected members were carriers of the *TRPM1* (c.413T>C; p.L138P) variant. This included half-sisters Z929 and Z1501, as well as their paternal cousin Z1471. Despite the fact that *TRPM1* (c.4237G>T; p.E1413X) segregated with all affected individuals in Family R1381, the segregation within Family R1256 did not provide supporting evidence of *TRPM1s* role in IA development.

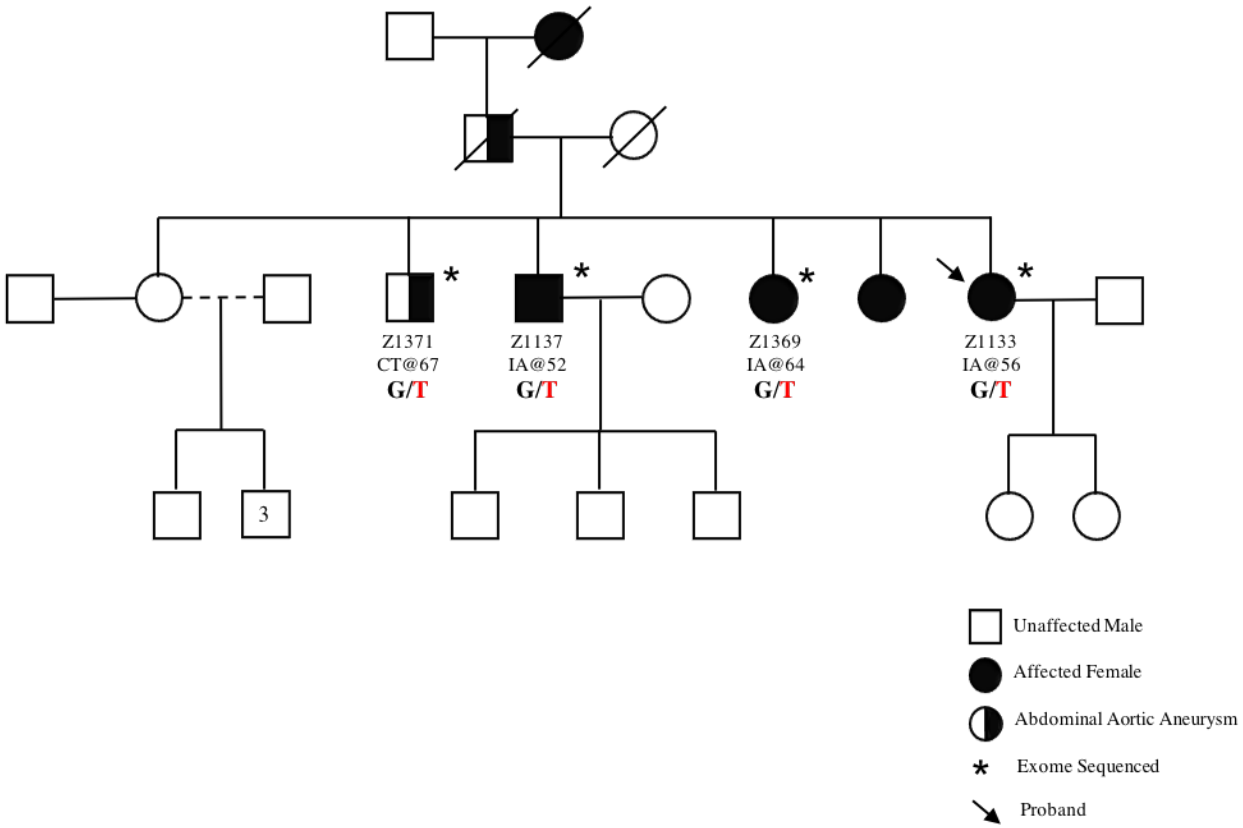


Figure 3.21. *TRPM1* (c.4237G>T; p.E1413X) segregation in affected family members of R1381

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

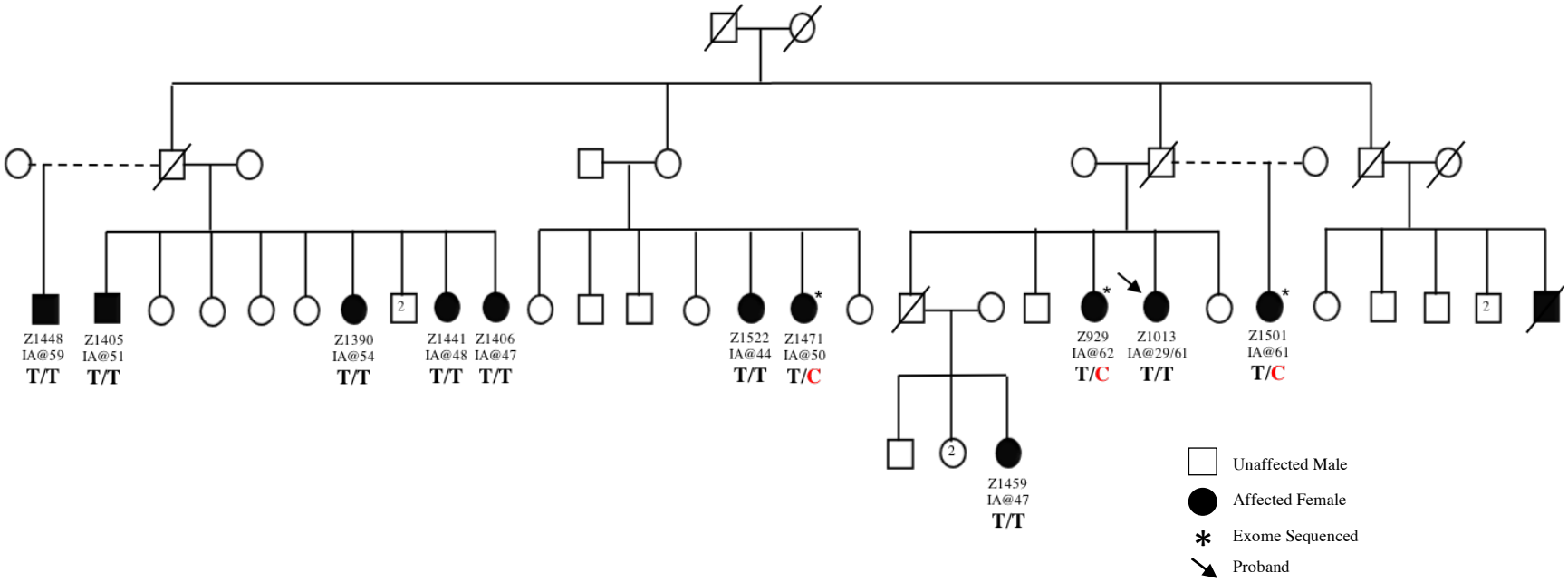


Figure 3.22. *TRPM1* (c.413T>C; p.L138P) segregation in affected family members of R1256

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis.

3.4 Familial Segregation

The remaining 12 candidate variants from eight genes (*TAS1R1*, *HAL*, *ZNF835*, *TTN*, *MST1R*, *METTL20*, *PTPN13*, *ADGRV1*) segregated with all affected individuals within a family, and thus subsequent sequencing of unaffected family members was completed. The following sections summarize the familial segregation for all validated variants.

3.4.1 TAS1R1 (c.1807C>T p.R603C)

Taste 1 Receptor Member 1 (*TAS1R1*) encodes a G-protein coupled receptor, a component of the amino acid taste receptor T1R1+3 that is activated by tastes perceived as sweet [222]. There are currently no disorders associated with this gene, but Farlow *et al.* (2015) highlighted *TAS1R1* as a candidate IA gene. They identified a mutation in *TAS1R1* that segregated with all affected individuals in an IA study family [100]. WES identified a different mutation, *TAS1R1* (c.1807C>T p.R603C), in the study Family R1381.

Sanger sequencing validated the *TAS1R1* (c.1807C>T p.R603C) variant as it was present in 3/3 affected individuals from Family R1381. All three affected siblings Z1133, Z1137, and Z1369 are carriers of the variant. An additional sibling, Z1371, who has an AAA, is not a carrier of the variant. The DNA from eight additional unaffected relatives was also available, and Sanger sequencing results showed that 4/8 of these unaffected individuals are carriers of *TAS1R1* (c.1807C>T p.R603C) (Figure 3.23). Individual Z1135 is another sibling of Z1133, Z1137, and Z1369 (all affected), and is a carrier of *TAS1R1* (c.1807C>T p.R603C) yet showed no IA phenotype as of her last CT scan (68 years old). Affected individual Z1137 has three unaffected sons who also have the *TAS1R1* (c.1807C>T p.R603C) variant, Z1145, Z1368, and Z1370. All three sons were characterized as unaffected. However, they were 25, 28, and 25 at the time of

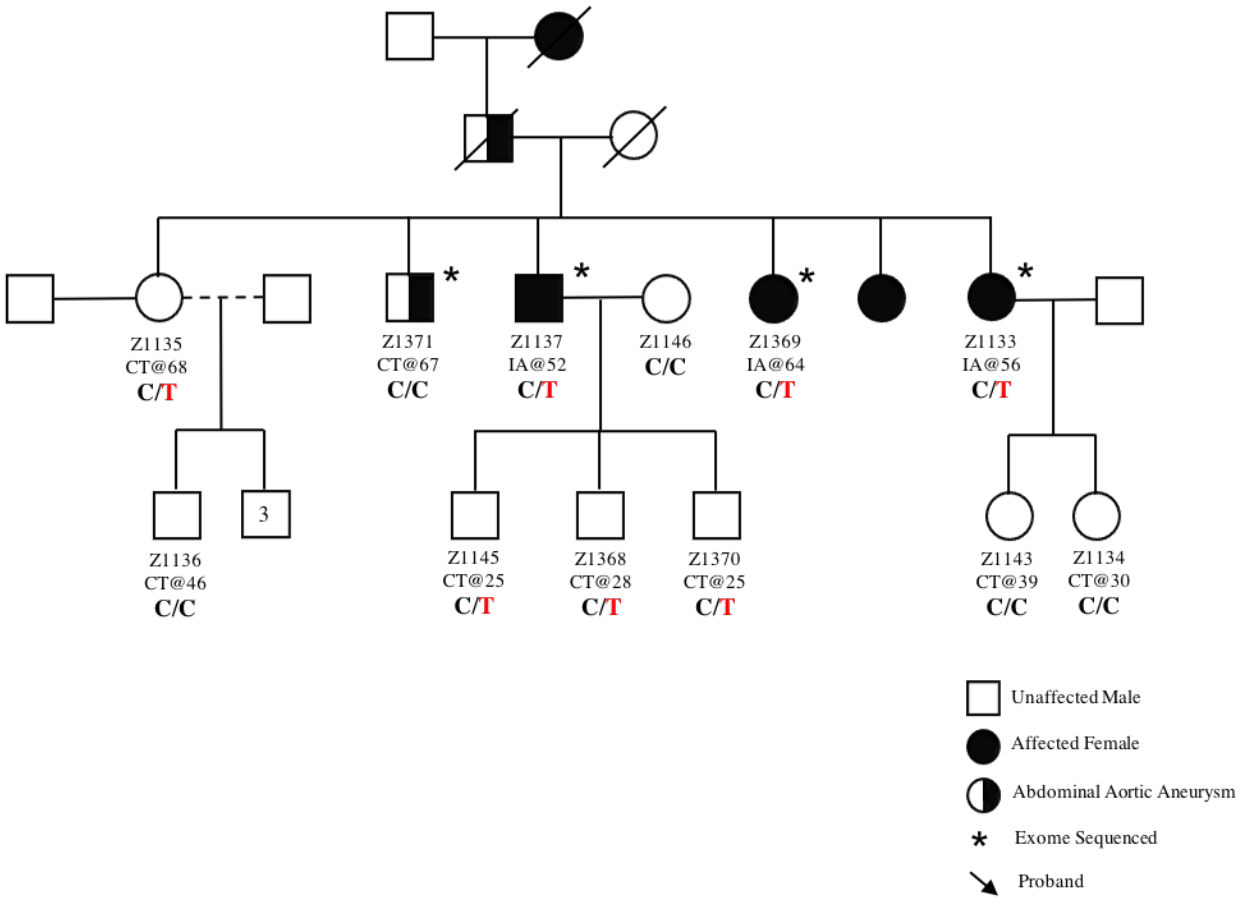


Figure 3.23. *TAS1R1* (c.1807C>T p.R603C) segregation in Family R1381

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

their last CT scans.

3.4.2 TTN (c.22420G>A; p.A7474T)

As it encodes the largest protein in the human body, the Titin (*TTN*) gene is more susceptible to genetic variants by chance, simply due to its size alone [223]. However, this cannot discount the contribution of *TTN* mutations to human disease. The titin protein is an abundant protein of striated muscle, and thus mutations in *TTN* cause various familial cardiomyopathies [224]. Titin plays a significant role in the flexibility and structure of muscle, most notably cardiac and skeletal muscle, depending on the isoform of titin that is present. The discovery of a new titin isoform, the protein sm-titin in smooth muscle demonstrates the potential relevance of titin to the vascular smooth muscle of human blood vessels [225,226]. Farlow and colleagues identified *TTN* as a candidate IA gene as a variant within *TTN* segregated with IA in one of their study families [100]. A different variant than that identified in the Farlow study, *TTN* (c.22420G>A; p.A7474T), was found in Family R1265 after WES.

Sanger sequencing of Family R1265 validated that the *TTN* (c.22420G>A; p.A7474T) variant was present in 2/2 affected siblings, Z994, and Z995. The DNA from four additional unaffected relatives was also available. Of these four, Z1041 and Z1046, who are siblings of Z994 and Z995, are also carriers of the *TTN* (c.22420G>A; p.A7474T) variant (Figure 3.24). The DNA from two additional relatives was also available, Z1050 and Z996, but neither was screened to determine a phenotype. Individual Z1050 is the mother of Z994, Z995, Z1041, and Z1046, and is also a carrier of *TTN* (c.22420G>A; p.A7474T). Individual Z996, who is married to Z995, was the other unscreened participant, and she is not a carrier of the variant.

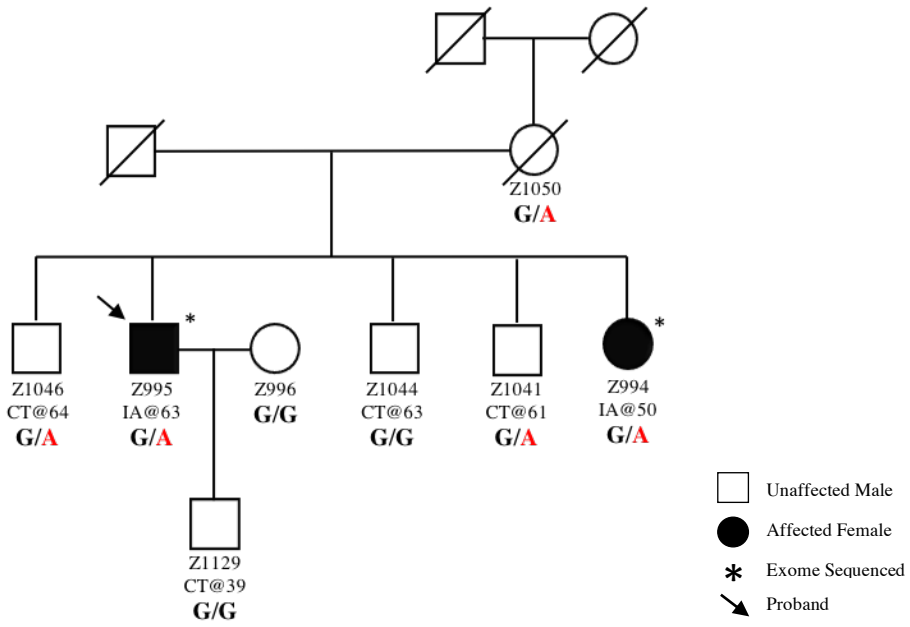


Figure 3.24. *TTN* (c.22420G>A; p.A7474T) segregation in Family R1265

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

3.4.3 HAL (c.959C>T; p.T320M)

Histidine Ammonia-Lyase (*HAL*) encodes the enzyme that catalyzes the first reaction in histidine catabolism. Previously identified mutations in *HAL* cause histidinemia, a disorder characterized by increased levels of histidine that is inherited in an autosomal recessive pattern [227]. The majority of individuals with histidinemia are unaware of their condition, as there is often no observable phenotype exhibited [228]. In rare circumstances, when histidinemia is combined with a lack of oxygen at birth, individuals may be at greater risk of developing intellectual disabilities [211].

Sanger sequencing validated that the *HAL* (c.959C>T; p.T320M) variant segregated with both affected individuals in Family R1262, one of the smaller study families in this study. The affected sisters Z991 and Z1037 are the only study participants (Figure 3.25). The family history of these participants revealed an additional sister with IA, as well as their paternal grandmother. While neither parent of the siblings had a known IA, they both had AAA. The validation of *HAL,* in combination with the identification of *HAL* as a candidate IA gene by Farlow *et al.* (2015), provided confidence in prioritizing *HAL* for further investigation, despite the limited segregation data [100].

3.4.4 ZNF835 (c.378delC; p.I126fs)

Resembling many zinc-finger proteins, Zine Finger Protein 835 (*ZNF835*) is thought to be involved in transcriptional regulation [229]. The data available regarding *ZNF835* is limited, and there are no disorders currently associated with this gene. However, Farlow *et al*. (2015) did identify *ZNF835* as a candidate IA gene [100]. The specific variant identified within Family R1381 is a novel, frame-shift deletion.

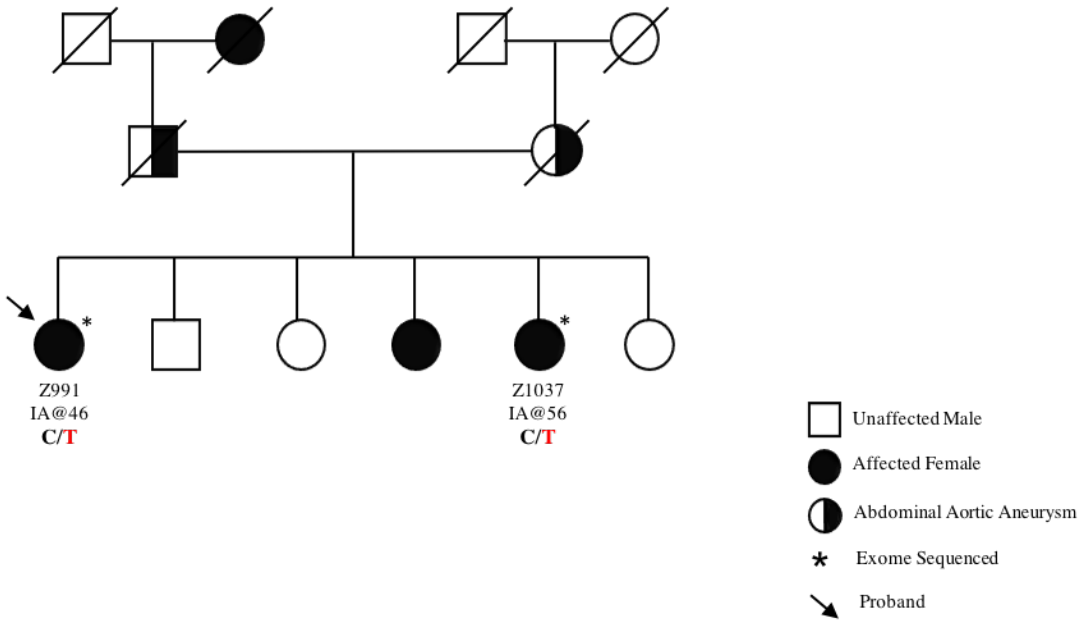


Figure 3.25. *HAL* (c.959C>T; p.T320M) segregation in Family R1262

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

Sanger sequencing results identified that all three affected siblings within Family R1381 are carriers of *ZNF835* (c.378delC; p.I126fs). The affected siblings Z1133, Z1137, and Z1369 also have a brother with an AAA (Z1371), and he too is a carrier of this *ZNF835* variant.

Sanger sequencing of the eight unaffected relatives in Family R1381 was completed following the validation of *ZNF835* (c.378delC; p.I126fs) and revealed the absence of the variant in 7/8 unaffected relatives (Figure 3.26). Individual Z1370, the son of Z1137 (affected), was found to have the *ZNF835* (c.378delC; p.I126fs) variant but was only 25 at the time of his last CT scan.

3.4.5 METTL20 (c.583C>G; p.L195V)

A protein lysine methyltransferase is encoded by *METTL20,* previously known as Electron Transfer Flavoprotein Beta Subunit Lysine Methyltransferase (*ETFBKMT*) [230,231]. Cellular metabolism may be a functional role of the *METTL20* gene, and there are no disorders currently associated with a loss of enzyme activity. The variant within *METTL20* identified by WES, *METTL20* (c.583C>G; p.L195V), was identified in two unrelated NL families, R1385 and R1277.

METTL20 (c.583C>G; p.L195V) segregated with all affected individuals from Family R1277. Three affected siblings, Z1353, Z1354, and Z1008, are all carriers of the variant, as well as their maternal cousins Z1305 and Z1355. Individual Z1305 has an affected son, Z1304, who also has the *METTL20* (c.583C>G; p.L195V) variant. The DNA was also available from eight unaffected relatives in R1277, and Sanger sequencing discovered *METTL20* (c.583C>G; p.L195V) was absent from 7/8 (Figure 3.27). The one unaffected relative with the *METTL20* variant is Z1359. Individual Z1358 is the mother of Z1353, Z1354, and Z1008 and was also a

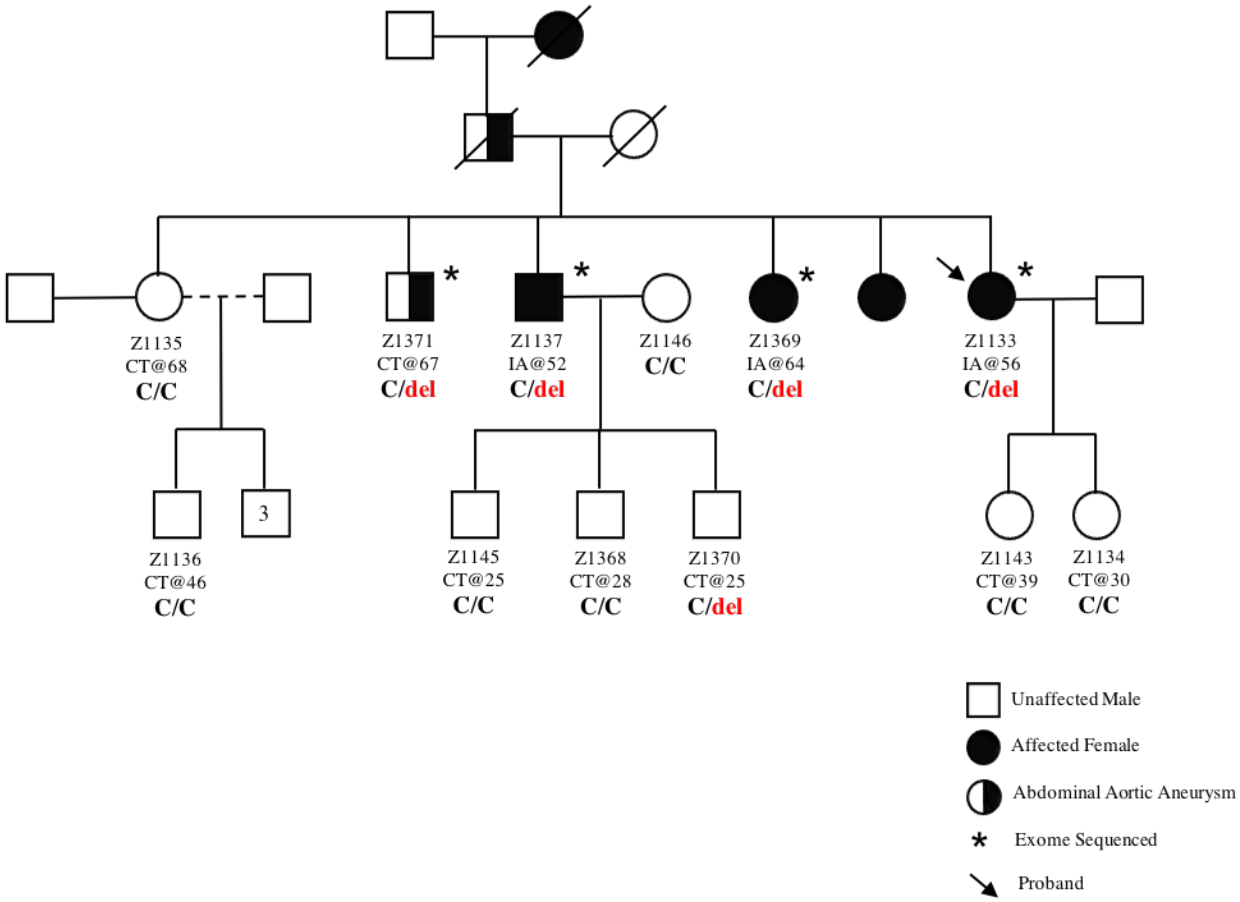


Figure 3.26. *ZNF835* (c.378delC; p.I126fs) segregation in Family R1381

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

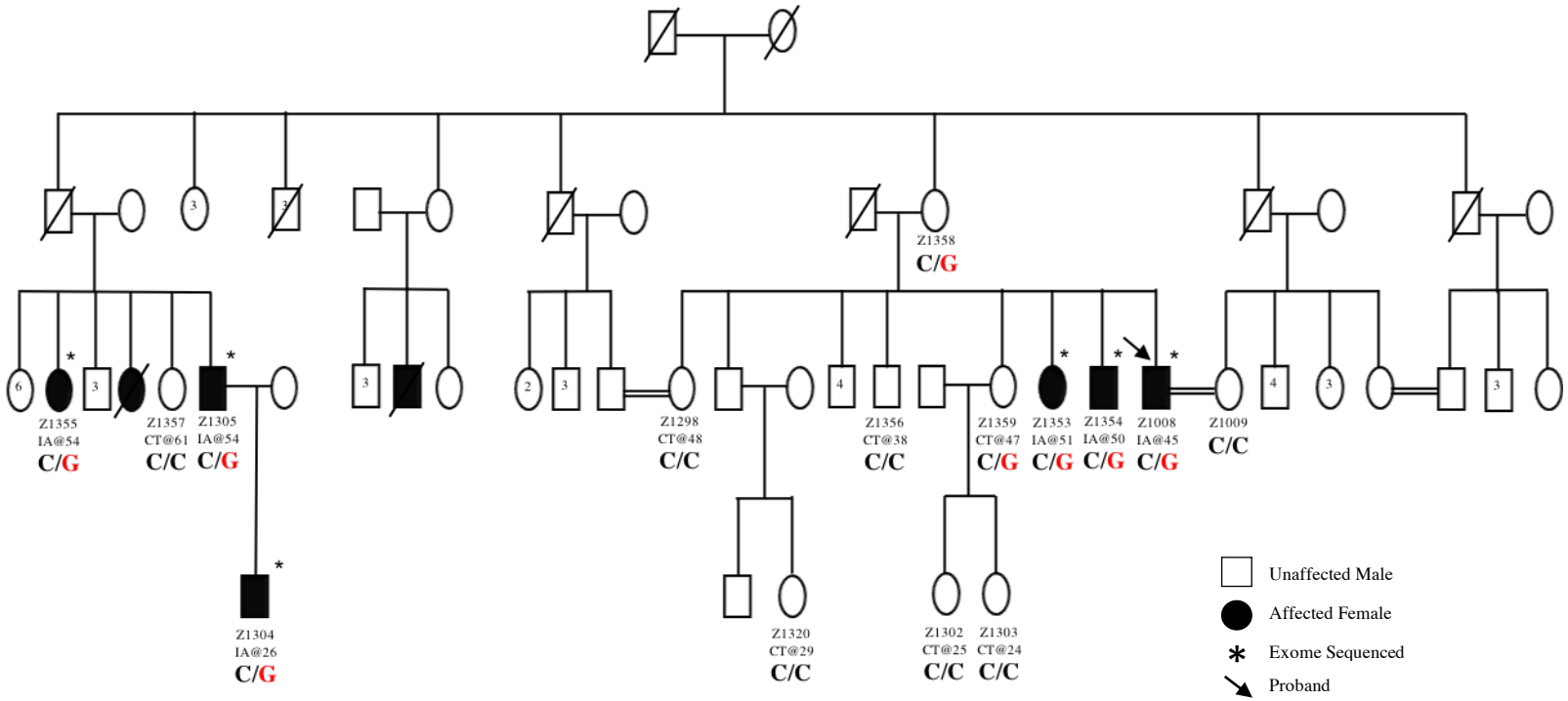


Figure 3.27. *METTL20* (c.583C>G; p.L195V) segregation in Family R1277

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

study participant. She also was found to be a carrier of *METTL20* (c.583C>G; p.L195V) but was never screened to determine a phenotype.

The *METTL20* (c.583C>G; p.L195V) variant was also identified in Family R1385. Furthermore, Sanger sequencing validated *METTL20* (c.583C>G; p.L195V) segregated with both affected individuals, Z1148, and Z1150 (Figure 3.28). Individual Z1149 was a sister of Z1150 and was the only unaffected study participant in R1385. Sanger sequencing the DNA of Z1149 revealed she too is a carrier of *METTL20* (c.583C>G; p.L195V). However, she was only 41 at the time of her last CT scan. The late onset of IA makes it challenging to assume a seemingly unaffected individual would not eventually develop an IA.

3.4.6 MST1R (c.4129C>T; p.Q1377X)

Macrophage Stimulating 1 Receptor (*MST1R*) encodes a cell surface receptor for macrophage stimulating protein [232]. This cell surface receptor is highly expressed in the ciliary epithelia of the lung, where it is thought to be involved in host defense [233]. The novel, stopgain mutation *MST1R* (c.4129C>T; p.Q1377X) was identified in Family R1262, as well as Family R1385, via WES. Sanger sequencing validated the *MST1R* (c.4129C>T; p.Q1377X) variant in both affected individuals Z991 and Z1037 from Family R1262. While R1262 does have a strong family history of both IA and AAA, Z991 and Z1037 are the only familial study participants (Figure 3.29).

Sanger sequencing revealed the variant segregated with both affected individuals Z1148 and Z1150 (Figure 3.30). Individual Z1149 is the only unaffected study participant within Family R1385, and she too is a carrier of the *MST1R* (c.4129C>T; p.Q1377X) variant. Notably, she was 41 at the time of her last CT scan.

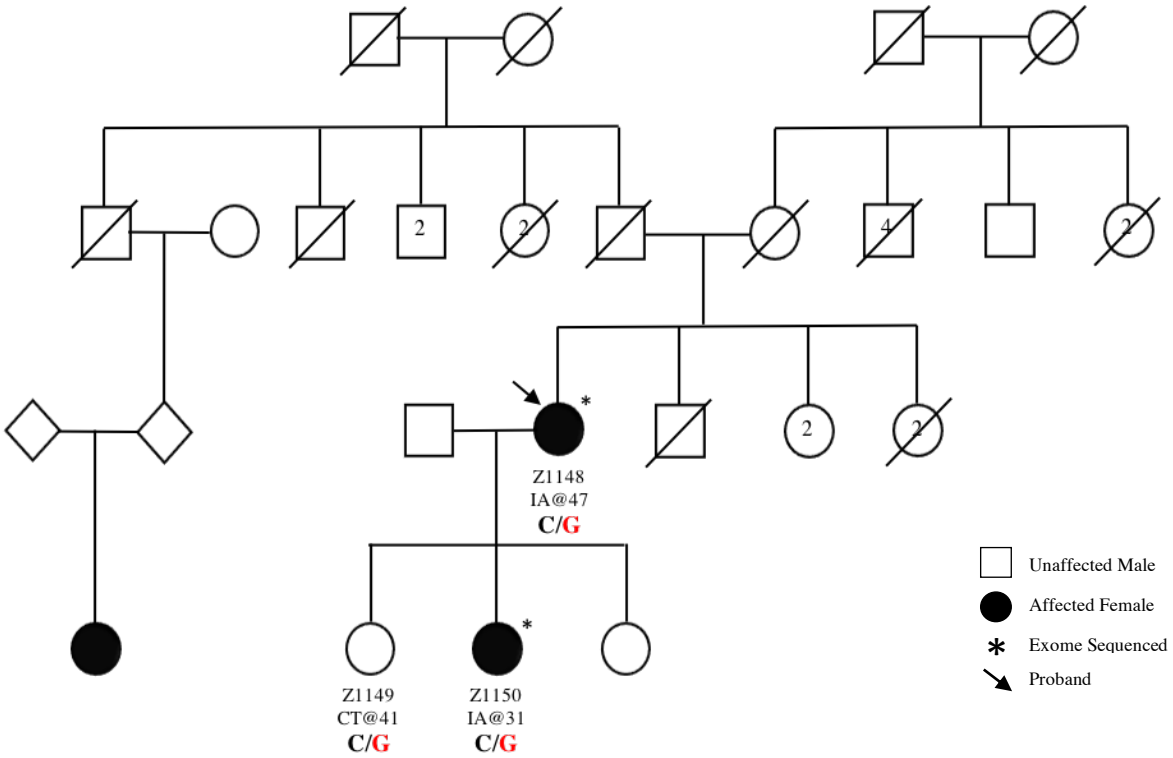


Figure 3.28. *METTL20* (c.583C>G; p.L195V) segregation in Family R1385

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

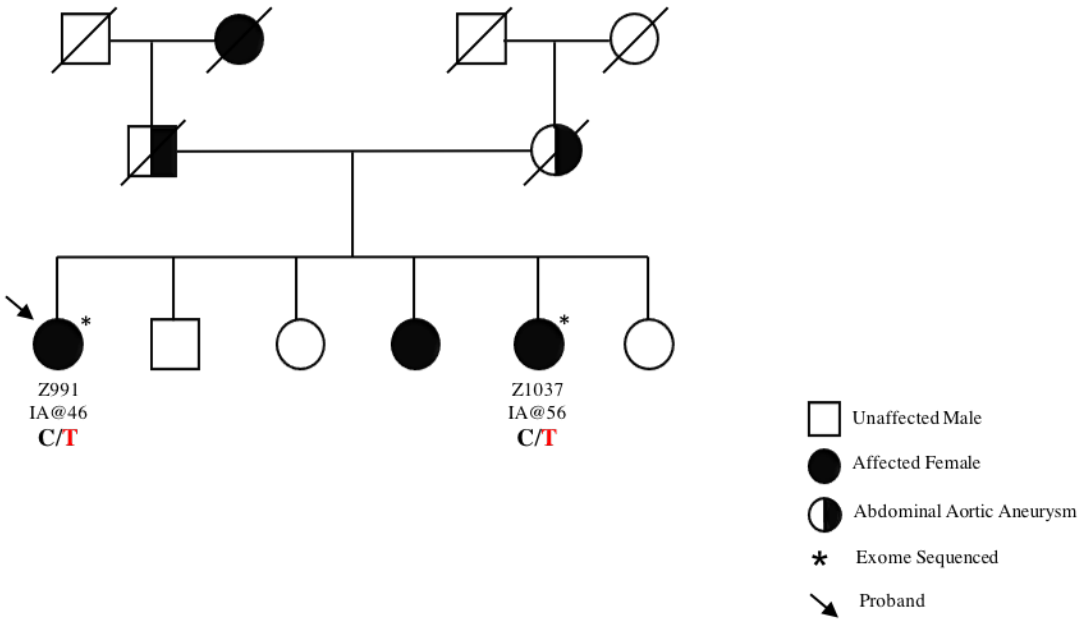


Figure 3.29. *MST1R* (c.4129C>T; p.Q1377X) segregation in Family R1262

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

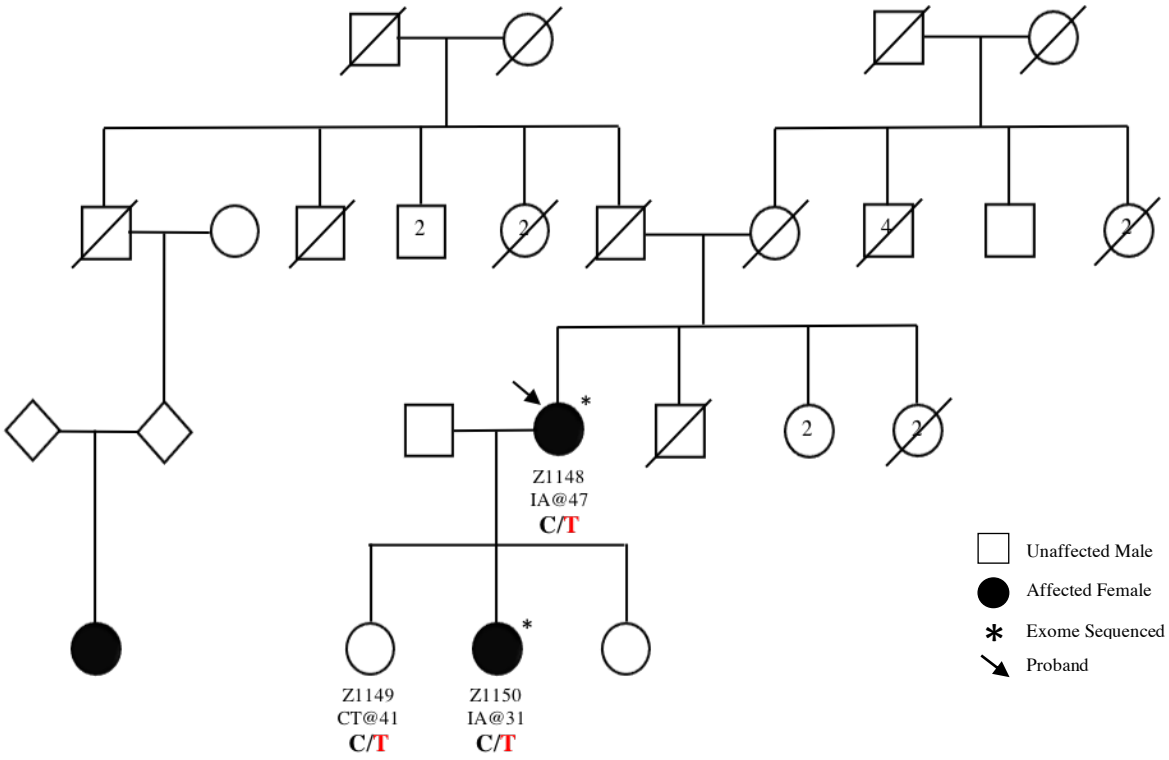


Figure 3.30. *MST1R* (c.4129C>T; p.Q1377X) segregation in Family R1385

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

3.4.7 ADGRV1 (c.1522A>C; p.I508L) & *ADGRV1* (c.6317C>T; p.A2106V)

The ADGRV1 protein is a member of the G-protein coupled receptor superfamily. Mutations in Adhesion G Protein Coupled Receptor V1 (*ADGRV1*) are associated with an autosomal dominant form of familial febrile seizures as well as Usher syndrome, which demonstrates autosomal recessive inheritance. [198,234,235]. Two different variants in the *ADGRV1* gene were found in two unrelated NL families. Sanger sequencing confirmed *ADGRV1* (c.1522A>C; p.I508L) was present in 2/2 affected individuals in Family R1262. The family has three affected sisters, yet DNA was only available from two, Z991 and Z1037, both of whom are carriers of the variant (Figure 3.31).

The alternative *ADGRV1* (c.6317C>T; p.A2106V) variant was identified in Family R1276. Sanger sequencing confirmed the variant segregated in 3/3 affected siblings, Z1015, Z1017, and Z1007 (Figure 3.32). The validation of *ADGRV1* (c.6317C>T; p.A2106V) in all affected individuals from Family R1276, prompted the sequencing of DNA from unaffected relatives as well. This would provide a better representation of the variant segregation within the family (Figure 3.33). Family R1276 exemplifies one of the large, well-characterized families we have available within the study. The DNA was available from 52 unaffected relatives, all of whom were screened to determine a phenotype. Individual Z1415 was initially classified as an affected family member. However, subsequent follow-up investigations ruled there was no IA, and Z1415 was reclassified as unaffected. His original classification as an affected study participant contributed to the widespread number of unaffected study participants, as relatives had already been recruited from that side of the family. Sanger sequencing revealed 2/52 unaffected individuals, Z1035 and Z1027, are carriers of *ADGRV1* (c.6317C>T; p.A2106V). Z1035 is the son of affected participant Z1015. The other unaffected carrier was Z1027, a

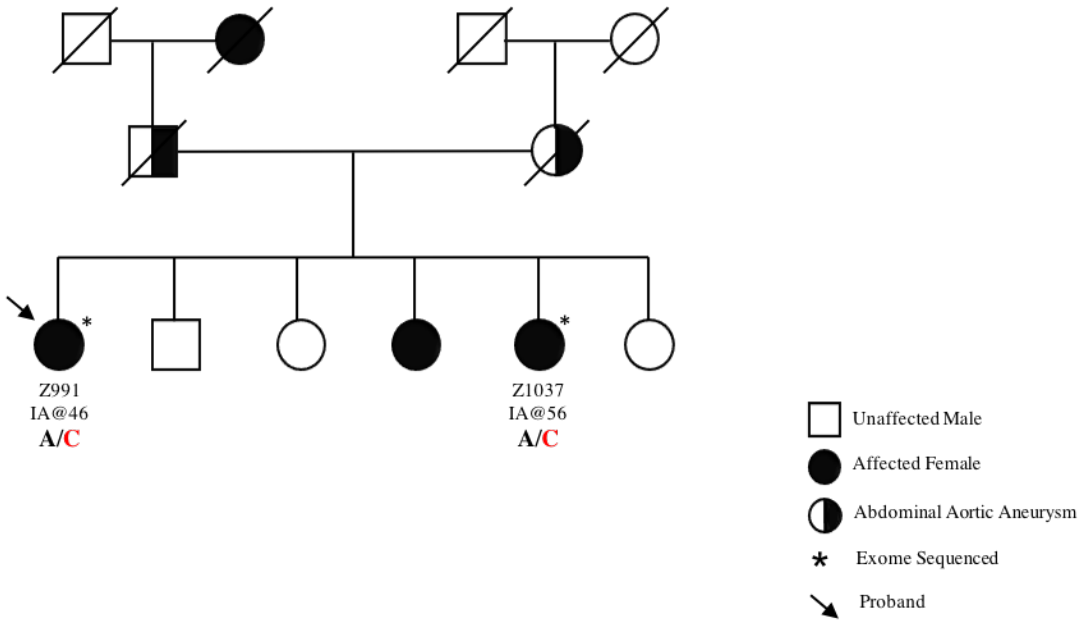


Figure 3.31. *ADGRV1* (c.1522A>C; p.I508L) segregation in Family R1262

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

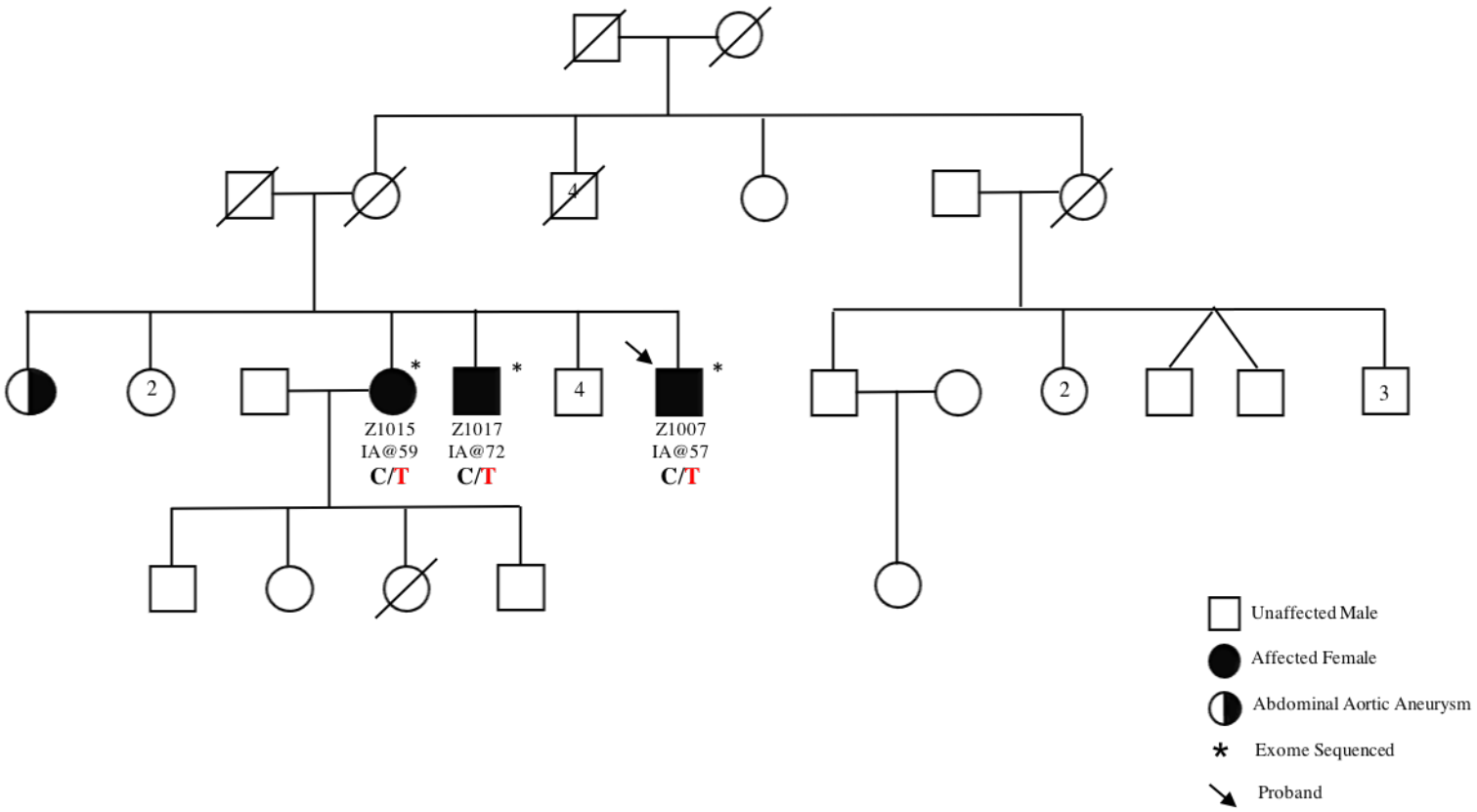


Figure 3.32. ***ADGRV1*** **(c.6317C>T; p.A2106V) segregation in affected family members of Family R1276**

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

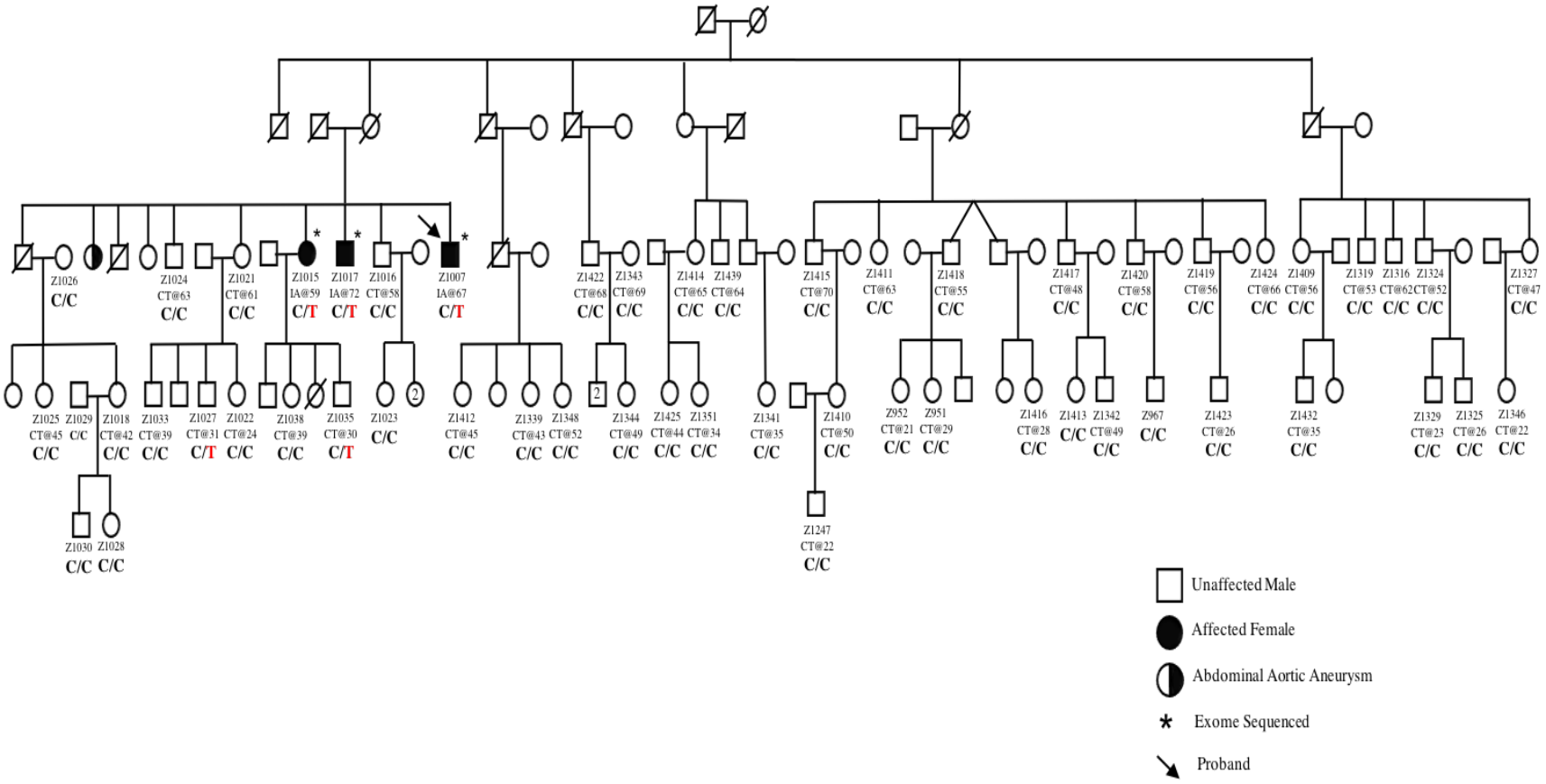


Figure 3.33. *ADGRV1* (c.6317C>T; p.A2106V) segregation in Family R1276

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

nephew of Z1015. These two seemingly unaffected study participants were 31 and 30, respectively, at the time of their latest CT scans. However, the late onset of IA makes it difficult to assume these are truly unaffected individuals who would not eventually develop an IA. The segregation of both *ADGRV1* (c.1522A>C; p.I508L) and *ADGRV1* (c.6317C>T; p.A2106V) provided assurance that further investigation into their possible involvement in IA development was worth pursuing.

3.4.8 PTPN13 (c.3563C>T; p.T1188I) & PTPN13 (c.7336C>T; p.R2446C)

The Protein Tyrosine Phosphatase Nonreceptor Type 13 (*PTPN13*) gene, encodes a member of the protein tyrosine phosphatase family of signaling molecules [236]. While various functions of *PTPN13* have been evaluated, such as its role in apoptosis, tumor suppression, the mitotic cycle, cell migration, and differentiation, gaps still remain in our knowledge of *PTPN13* [236–239]. WES exposed two different mutations in the *PTPN13* gene. The variant *PTPN13* (c.3563C>T; p.T1188I) was identified in Family R1381, and the variant c.7336C>T (p.R2446C) was identified in Family R1385.

The *PTPN13* (c.3563C>T; p.T1188I) variant was validated in 3/3 affected individuals from Family R1381 via Sanger sequencing. Three affected siblings, Z1133, Z1137, and Z1369, are all carriers of *PTPN13* (c.3563C>T; p.T1188I). Their additional sibling, Z1371, who has an AAA, is also a carrier of the *PTPN13* (c.3563C>T; p.T1188I) variant. DNA was also available from eight additional unaffected relatives, and 4/8 of these individuals are carriers of the *PTPN13* (c.3563C>T; p.T1188I) variant (Figure 3.34). Two of the four individuals, Z1145 and Z1368, are sons of Z1137 (affected) who were respectively 25 and 28 years old at the time of their last CT scans. The other two unaffected carriers of *PTPN13* (c.3563C>T; p.T1188I), Z1143

and Z1134, are the daughters of Z1133 (affected) and were 39 and 30 respectively at the time of their last CT scans.

Family R1385 has two affected study participants, Z1148 and Z1150. Sanger sequencing validated the *PTPN13* (c.7336C>T; p.R2446C) variant in both of these affected individuals. Sanger sequencing was then completed on the one unaffected family member, Z1149, and she is not a carrier of the *PTPN13* (c.7336C>T; p.R2446C) variant (Figure 3.35).

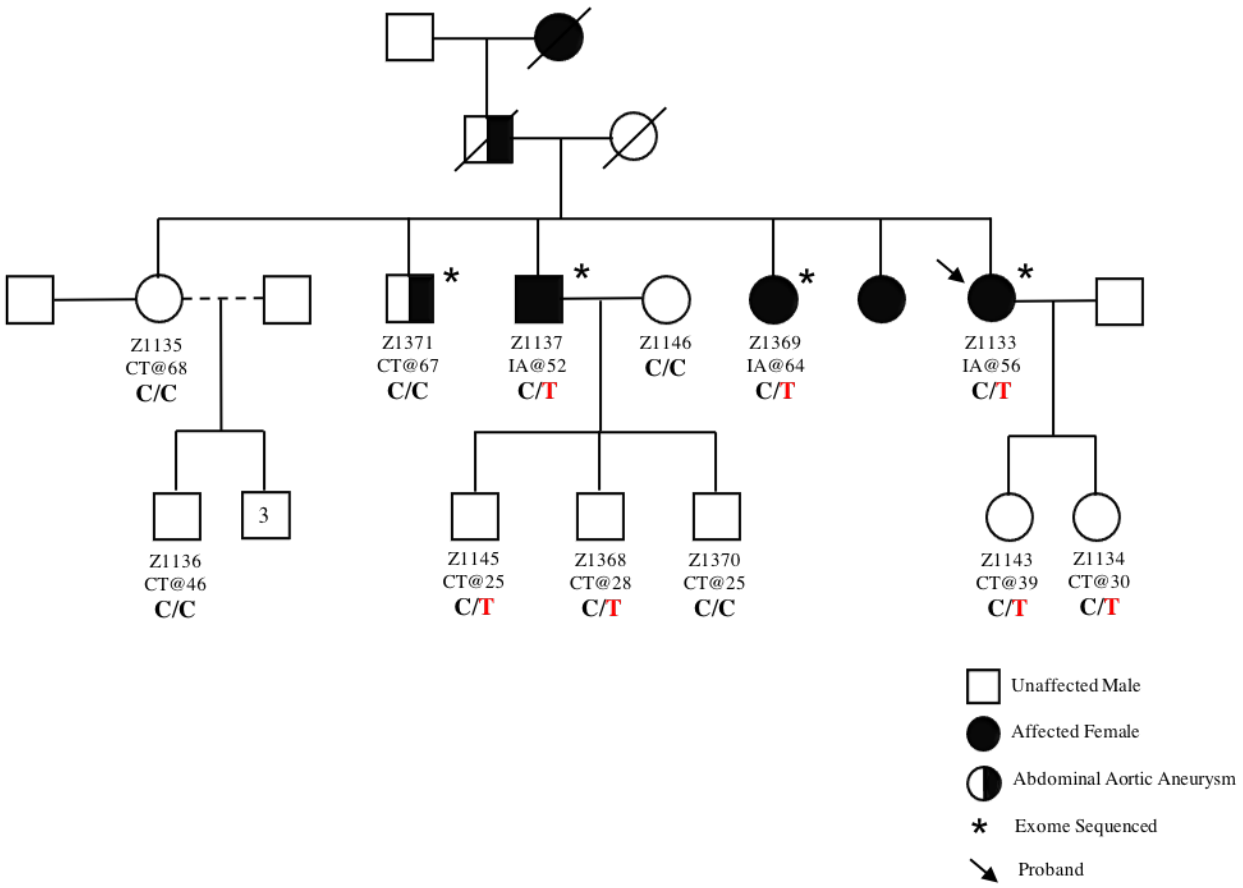


Figure 3.34. *PTPN13* (c.3563C>T; p.T1188I) segregation in Family R1381

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

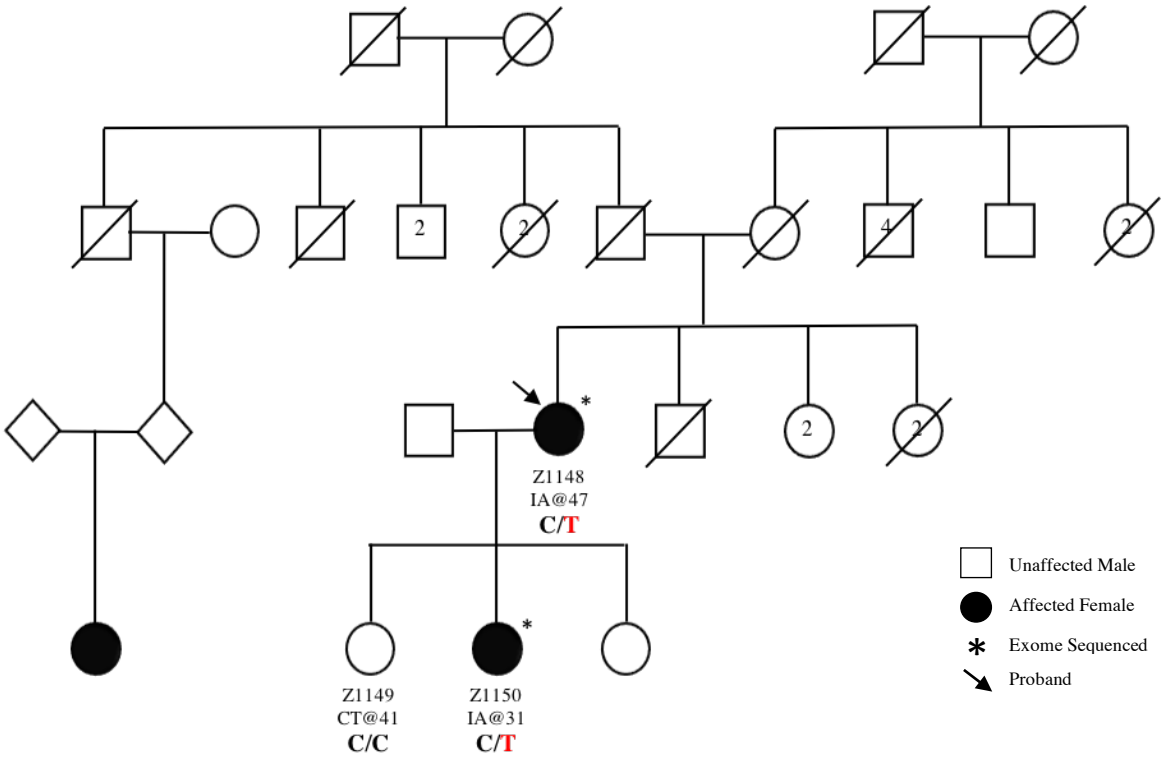


Figure 3.35. *PTPN13* (c.7336C>T; p.R2446C) segregation in Family R1385

Any individual with a Z identification number indicates a study participant in which DNA was available. Individuals with IA and AAA are indicated, as well as the patient age at diagnosis or age at last CT scan.

3.5 Population Controls

To determine the prevalence of the top candidate variants in the NL population, 100 population controls from the NFCCR were Sanger sequenced for each of the six prioritized variants (Table 3.5). Because the *TTN* gene was still being considered for functional analysis, population controls were also sequenced for the variant *TTN* (c.22420G>A; p.A7474T). For the gene *ADGRV1,* analysis of two different mutations was required, and thus 100 population controls were sequenced for each variant. The candidate variants *HAL* (c.959C>T; p.T320M), *METTL20* (c.583C>G; p.L195V), *TTN* (c.22420G>A; p.A7474T) and *ADGRV1* (c.1522A>C; p.I508L) were absent from 100 NL population controls. The other *ADGRV1* variant, *ADGRV1* (c.6317C>T; p.A2106V) was identified in 1/100 NL population controls, as well as the variants in *ZNF835* (c.378delC; p.I126fs) and *MST1R* (c.4129C>T; p.Q1377X).

As the project progressed, it became apparent that *HAL* (c.959C>T; p.T320M) was emerging as the primary candidate variant. Due to the lack of unaffected study participants within Family R1262, 200 additional population controls were Sanger sequenced, which identified *HAL* (c.959C>T; p.T320M) in 1/300 population controls (Table 3.5).

Table 3.5. NL control sequencing data of functional candidate variants

Gene	Variant Details (Transcript ID)	Family	MAF (Allele Count)	NL Controls
TTN	c.22420G>A; p.Ala7474Thr (NM_001267550)	R1265	0.0001256 (15/119380)	0/100
ZNF835	c.378delC; p. Ile126fs (NM_001005850)	R1381	novel	1/100
ADGRV1	c. 1522A>C; p. Ile508Leu (NM_032119)	R1262	0.00261 (252/96550)	0/100
	c. 6317C>T; p. Ala2106Val (NM_032119)	R1276	0.001622 (195/120242)	1/100
METTL20	c.583C>G; p.Leu195Val (NM_001135863)	R1385 R1277	0.002728 (331/121356)	0/100
MST1R	c.4129C>T; p.Gln1377Ter (NM_002447)	R1262 R1385	novel	1/100
HAL	c.959C>T; p.Thr320Met (NM_001258333)	R1262	0.0002472 (30/121346)	1/300[xiii]

[xiii] Functional analysis revealed *HAL* (c.959C>T; p.T320M) was emerging as the primary candidate variant. Due to the lack of unaffected study participants within Family R1262, 200 additional population controls were Sanger sequenced

3.6 Functional Candidates

Once Sanger sequencing was completed for all 18 candidate variants, the familial segregation of each variant was analyzed. In addition to segregating with all affected individuals within a family, variants also needed to be absent from a large percentage of unaffected individuals within the family to be considered for further functional analysis. There are some limitations to such stringent segregation criteria. However, time and financial constraints required that potential candidate variants be prioritized, as not every variant could be functionally analyzed. Of the 12 candidate variants from eight genes (*TAS1R1*, *HAL*, *ZNF835*, *TTN*, *MST1R*, *METTL20*, *PTPN13*, *ADGRV1*) that segregated with all affected individuals within a family, six variants from five genes (*HAL, ZNF835, MST1R, METTL20* and *ADGRV1*) met the segregation criteria, and thus were prioritized for further investigation. Although the *TTN* variant (c.22420G>A; p.A7474T) was found in 4/8 unaffected relatives in Family R1381, the biological function of this gene caused it to remain a noteworthy candidate, and it was initially considered for further functional analysis.

3.7 Morpholino Mediated Genetic Knockdowns

To evaluate the role of each prioritized candidate variant, antisense MOs were used to knockdown the homologous gene in a zebrafish model. Initially, all six candidate genes, including *TTN,* planned to be evaluated using this model. Zebrafish have emerged as an exceptional model organism for studying genetic disease. The cerebral vasculature is similar to that of humans, their transparent embryos allow for visualization of this vasculature, and their genome is highly conserved with that of humans [117]. However, the additional round of whole genome duplication the zebrafish has undergone resulted in several gene duplications. The *TTN* gene has two orthologs within the zebrafish genome, *ttna,* and *ttnb* [240]. Complicating matters further, the large size and alternative splice sites of *TTN* requires the use of four antisense MOs for gene knockdown [241]. Previous work with *TTN* in zebrafish produced morphant embryos with severe cardiac problems, some of which included embryos without a functioning heart [242]. The cardiac issues that arise in these zebrafish morphants make it hard to differentiate if hemorrhaging is the result of the genetic knockdown or secondary to the cardiac dysfunction [240]. The foreseeable complexity of the *TTN* knockdown requires more exploration that could be pursued in the future, but was not done so in this thesis.

3.7.1 ZNF835

Although approximately 70% of all human genes have a clear orthologue within the zebrafish genome, this is not the case for all human genes [128]. There is no clear zebrafish orthologue for *ZNF835,* and thus MO-mediated gene knockdown could not be completed. As the *ZNF835* (c.378delC; p.I126fs) variant appears to be an exceptional candidate variant genetically,

finding an alternative route to functionally evaluate its role in IA development would be worth future pursuit.

3.7.2 ADGRV1 & METTL20

The initial MO-mediated gene knockdowns were completed by injecting zebrafish embryos at the one–cell stage, with two different doses of *adgrv1* or control MO. Embryos were then imaged to see if a dose-dependent hemorrhagic phenotype was observable (Figure 3.36). Knockdown of the *ADGRV1* ortholog (*adgrv1)* did not produce a significant hemorrhagic phenotype (p=0.451), and thus no further MO injections were completed (Table 3.6). The chi-square test was used to determine the difference in proportion of hemorrhaging and non-hemorrhaging embryos. The chi-square statistic with Yates correction was 0.568 with a p-value of 0.451. Similarly, the MO-mediated gene knockdown of the *METTL20* zebrafish ortholog (*mettl20)* also did not produce a significant hemorrhagic phenotype (p=0.807) compared to that of the control morphants (Table 3.6). The chi-square statistic with Yates correction was 0.0595 with a p-value of 0.807. Thus, no additional MO injections were completed.

Table 3.6. Hemorrhage prevalence induced by varying doses of *adgrv1* or *mettl20* MO

Treatment	Dose	Number of Embryos Injected	Number of Embryos with hemorrhages	Hemorrhage Prevalence
***adgrv1* MO**	3 ng	27	2	7.41%
***adgrv1* MO**	9 ng	24	0	0.00%
***mettl20* MO**	3 ng	47	2	4.26%
***mettl20* MO**	9 ng	21	0	0.00%
Control MO	9 ng	62	1	1.61%
Uninjected	–	5	0	0.00%

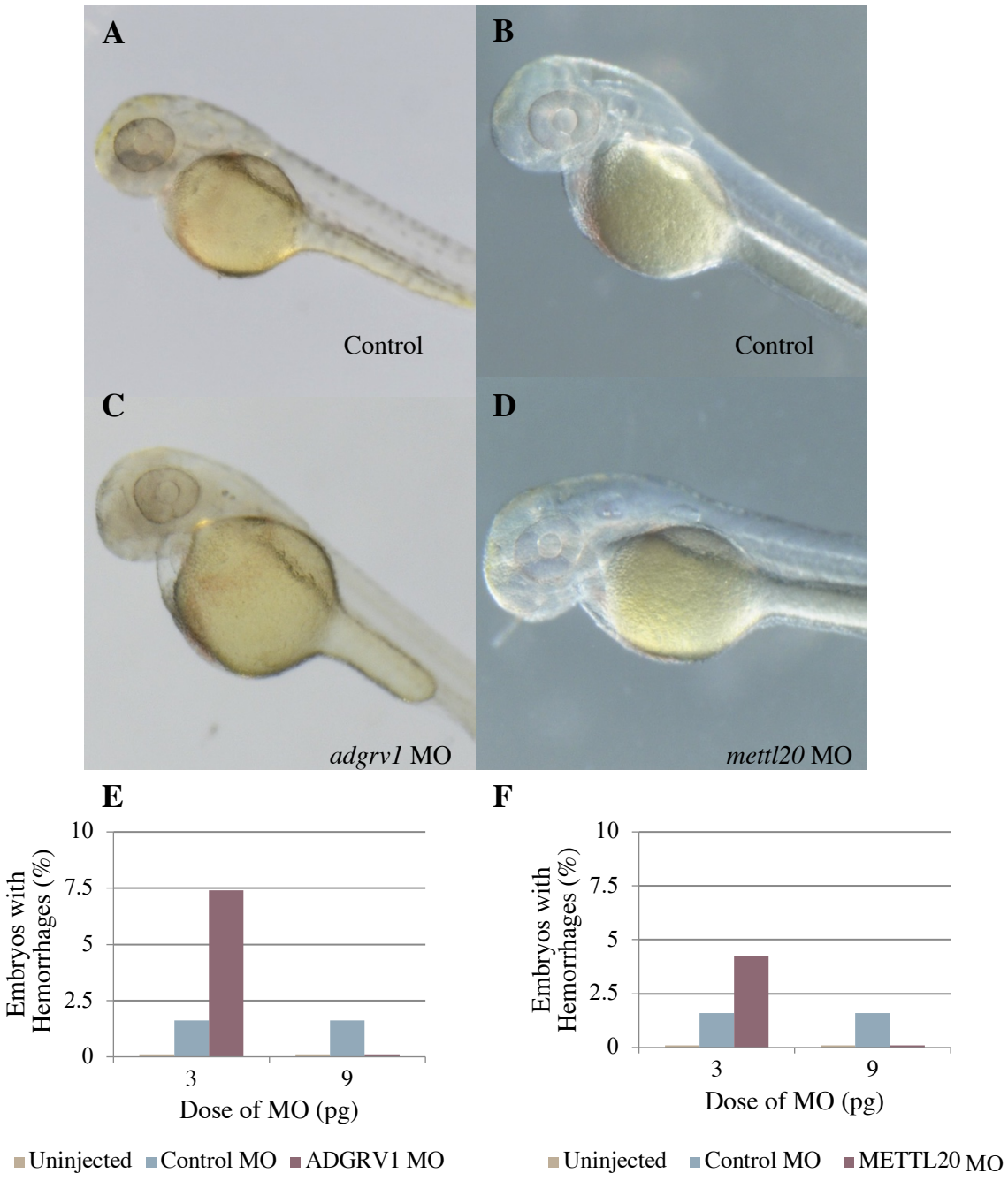


Figure 3.36. Knockdown of *adgrv1* & *mettl20* on zebrafish vasculature

The top images (**A-D**) show zebrafish embryos 48 hours post fertilization that were injected with various doses of control MO (**A** and **B**), *adgrv1* MO (**C**) or *mettl20* MO (**D**). The difference in hemorrhage prevalence (**E**) between control and *adgrv1* morphants was not significant (P=0.451). The difference in hemorrhage prevalence (**F**) between control and *mettl20* morphants was also not significant (P=0.807)

3.7.3 MST1R

The whole genome duplication of the zebrafish genome resulted in two *MST1R* orthologs, *mst1ra* and *mst1rb*. Zebrafish embryos were examined for a hemorrhagic phenotype after the simultaneous knockdown of *mst1ra* and *mst1rb,* as well as after the knockdown of only one paralog (*mst1ra* or *mst1rb*). No hemorrhagic phenotype occurred following individual *mst1ra* or *mstr1rb* knockdowns. However, the co-injection of *mst1ra+mstr1rb* did produce a hemorrhagic phenotype. The dose-dependent response that occurred with the *mst1ra*+b MO as compared to the control MO is evident in Figure 3.37. At a 3 ng dose, 6.67% of *mst1ra+b* MO injected embryos displayed a hemorrhagic phenotype as compared to the 1.61% of control morphants with hemorrhages. The prevalence of hemorrhages increased to 13.46% in *mst1ra+b* morphants when the MO dose was increased to 9 ng (Table 3.7).

Due to the dose-dependent response that occurred in *mst1ra+b* morphants, the dose of MO was increased to 12 ng, and embryos were co-injected with *p53*. When co-injected with *p53*, *mst1ra+b* morphants did not produce a significant hemorrhage phenotype as compared to controls (p=0.578). The chi-square statistic with Yates correction was 0.309 with a p-value of 0.578. Instead, co-injection with *p53* resulted in a substantial decline of hemorrhages in *mst1ra+b* morphants (3.22%) (Figure 3.37).

Table 3.7. Hemorrhage prevalence induced by varying doses of *mst1ra+b* or *hal* MO

Treatment	Dose	Number of Embryos Injected	Number of Embryos with hemorrhages	Hemorrhage Prevalence
***mst1rab* MO**	3 ng	15	1	6.67%
***mst1rab* MO**	9 ng	52	7	13.46%
***mst1rab + p53* MO**	12 ng	62	2	3.22%
***hal* MO**	3 ng	39	3	7.69%
***hal* MO**	9 ng	12	1	8.33%
***hal + p53* MO**	12 ng	43	6	13.95%
Control MO	9 ng	62	1	1.61%
***p53* MO**	12 ng	116	1	0.86%
Uninjected	–	27	0	0.00%

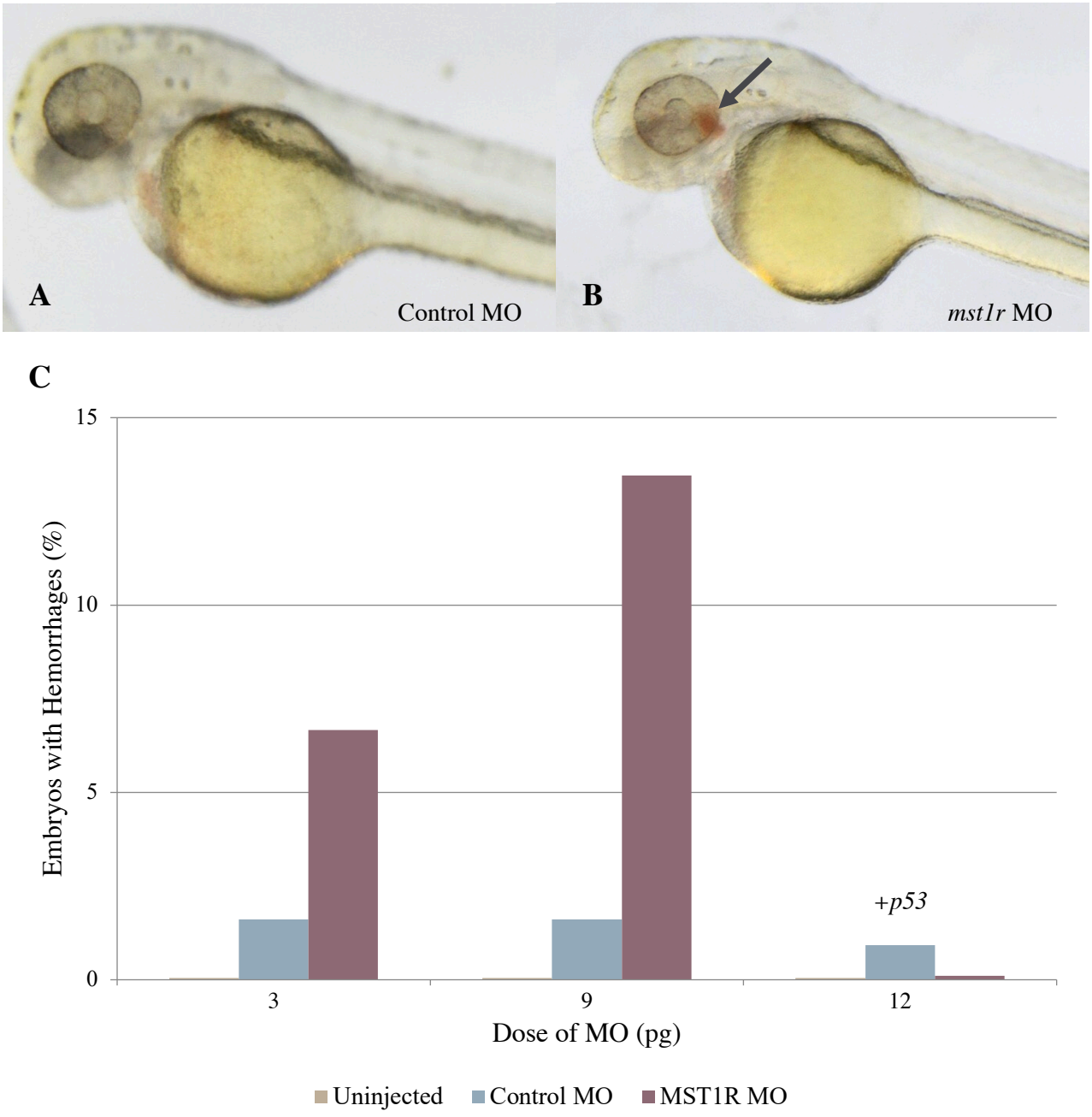


Figure 3.37. Impact of *mst1r* knockdown on zebrafish vasculature

The top images (**A** and **B**) show zebrafish embryos 48 hours post fertilization that were injected with various doses of control MO (**A**) or *mst1r* MO (**B**). The difference in hemorrhage prevalence (**C**) between control (*p53*) morphants and *mst1r*+*p53* morphants was not significant (P=0.578).

3.7.4 HAL

MO-mediated *HAL* knockdown was carried out the same way as the previous MO knockdowns. Embryos were first injected with either *hal* MO or control MO to see if a hemorrhagic phenotype was observed. At a 3 ng dose of *hal* MO, 7.69% of embryos had hemorrhages, whereas only 1.61% of embryos injected with control MO displayed a hemorrhagic phenotype. Increasing the dose of *hal* MO to 9 ng, increased the percentage of embryos with hemorrhages to 8.33% (Table 3.7).

Once an observable hemorrhagic phenotype was observed in *hal* morphants, the MO dose was increased to 12 ng, and embryos were co-injected with *p53*. At a 12 ng dose, the difference in hemorrhage prevalence between *hal+p53* MO injected zebrafish (13.95%), and *p53* MO injected control zebrafish (0.86%) was significant (p=0.0017). The chi-square statistic with Yates correction was 9.85 with a p-value of 0.0017. Figure 3.38 provides visualization of the dose-dependent response that occurred, as the concentration of *hal* MO was increased, the percentage of zebrafish embryos with hemorrhages also increased.

The initial rescue experiment was performed with human RNA concentrations at 200 ng/ul. At this concentration, both the wildtype and mutant *HAL* RNA rescued the hemorrhagic phenotype. At a 12 ng dose of *hal + p53* MO, the prevalence of hemorrhages was 10.26%. Embryos that received a subsequent injection of 400 ng wildtype *HAL* RNA showed a 3.26% hemorrhage prevalence, and embryos that received a subsequent injection of 400 ng mutant *HAL* RNA showed a 2.13% hemorrhage prevalence (Table 3.8). The difference in hemorrhage prevalence between *hal + p53* morphants and *hal + p53* morphants with wildtype *HAL* RNA was not significant. The chi-square statistic with Yates correction was 0.668 with a p-value of 0.414. Similarly, the difference in hemorrhage prevalence between *hal + p53* morphants and *hal + p53*

morphants with mutant *HAL* RNA was not significant. The chi-square statistic with Yates correction was 1.811 with a p-value of 0.178. Furthermore, the difference in hemorrhage prevalence between *hal + p53* morphants with wildtype *HAL* RNA and *hal + p53* morphants with mutant *HAL* RNA was not significant. The chi-square with Yates correction was 0.186 with a p-value of 0.666.

The results of the rescue experiment with a diluted RNA concentration (100 ng/ul) confirm the specificity of the *hal* MO induced phenotype and demonstrate the reduced functionality of *HAL* mRNA with the identified variant (Table 3.8). Injection of embryos with wildtype *HAL* RNA, after the injection of *hal + p53* MO, rescued the hemorrhagic phenotype. At a 12 ng dose of *hal + p53* MO, the prevalence of hemorrhages was 12.90%, whereas embryos that received a subsequent injection of 200 ng of wildtype *HAL* RNA had a hemorrhage prevalence of 1.61% (Figure 3.39). There was a significant difference in hemorrhage prevalence between *hal + p53* morphants and *hal + p53* morphants with wildtype *HAL* RNA. The chi-square statistic with Yates correction was 10.124 with a p-value of 0.00146. Embryos that instead received a subsequent injection of 200 ng of mutant *HAL* RNA demonstrated a hemorrhage prevalence of 12.77%. This was not statistically different than the hemorrhage prevalence of *hal + p53* morphants. The chi-square statistic with Yates correction was 0.0227 with a p-value of 0.880. There was also a significant difference in hemorrhage prevalence between *hal + p53* morphants with wildtype *HAL* RNA and *hal + p53* morphants with mutant *HAL* RNA. The chi-square with Yates correction was 10.218 with a p-value of 0.00139. The control cohort that did not receive any MO or RNA injections had a hemorrhage prevalence of 0.94% (Table 3.8). Figure 3.40 provides a visualization of the morphant embryos from the rescue experiment.

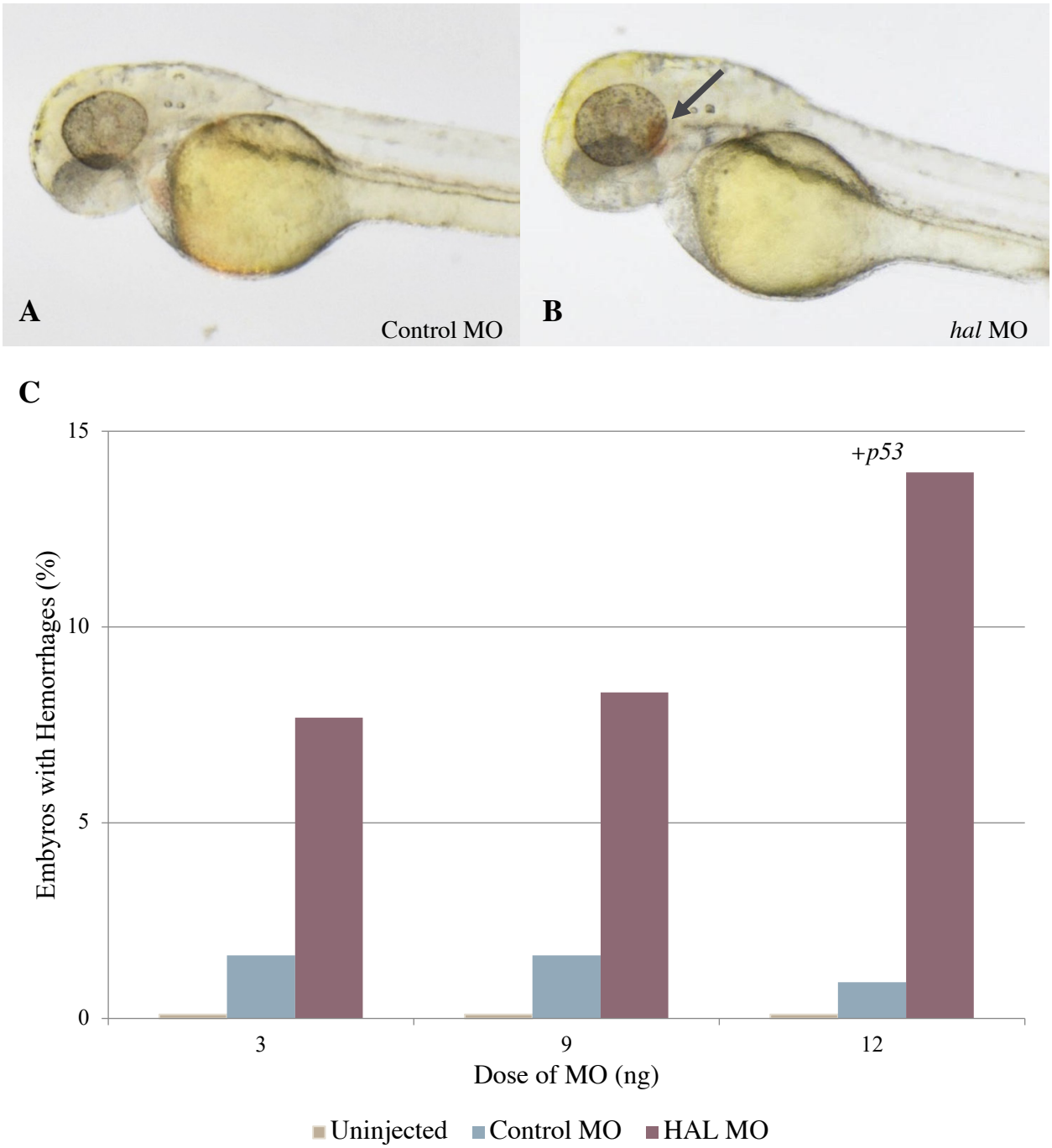


Figure 3.38. Impact of *hal* knockdown on zebrafish vasculature

The top images (**A** and **B**) show zebrafish embryos 48 hours post fertilization that were injected with various doses of control MO (**A**) or *hal* MO (**B**). The difference in hemorrhage prevalence (**C**) between control (*p53*) morphants and *HAL+p53* morphants was significant (P=0.0017).

Table 3.8. Hemorrhage prevalence in *hal+p53* morphants with RNA rescue

Treatment	Dose	Number of Embryos Injected	Number of Embryos with hemorrhages	Hemorrhage Prevalence
***hal + p53* MO**	12 ng	78	8	10.26%
Wildtype *HAL* RNA ***hal + p53* MO**	400 ng 12 ng	31	1	3.26%
Mutant *HAL* RNA ***hal + p53* MO**	400 ng 12 ng	47	1	2.13%
Uninjected	–	81	1	1.23%
***hal + p53* MO**	12 ng	124	16	12.90%
Wildtype *HAL* RNA ***hal + p53* MO**	200 ng 12 ng	124	2	1.61%
Mutant *HAL* RNA ***hal + p53* MO**	200 ng 12 ng	141	18	12.77%
Uninjected	–	106	1	0.94%

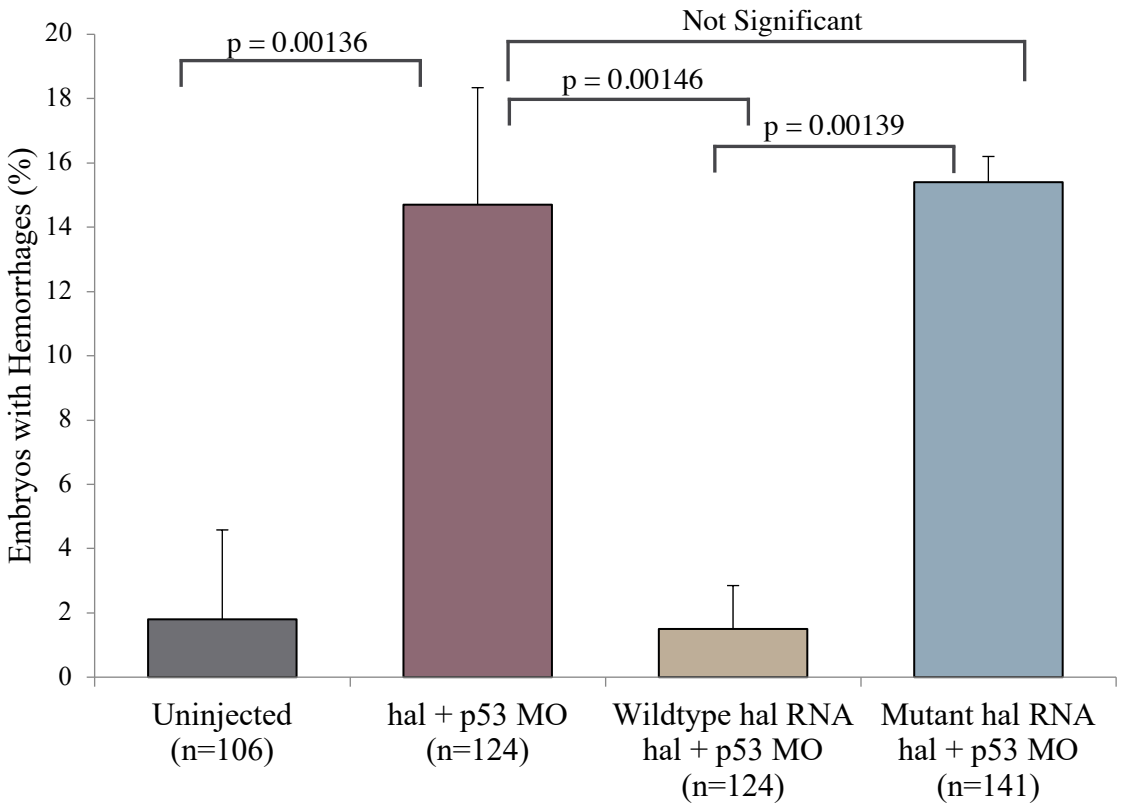


Figure 3.39. Hemorrhage prevalence in rescue experiment with diluted RNA concentration

The hemorrhage prevalence in uninjected embryos was 0.94 %. At a 12ng dose of *hal* + *p53* MO, the prevalence of hemorrhages was 12.90%. Embryos that received a subsequent injection of 200 ng of wildtype *hal* RNA had a hemorrhage prevalence of 1.61%. Embryos that instead received a subsequent injection of 200 ng of mutant *hal* RNA demonstrated a hemorrhage prevalence of 12.77%. Error bars indicate standard deviation.

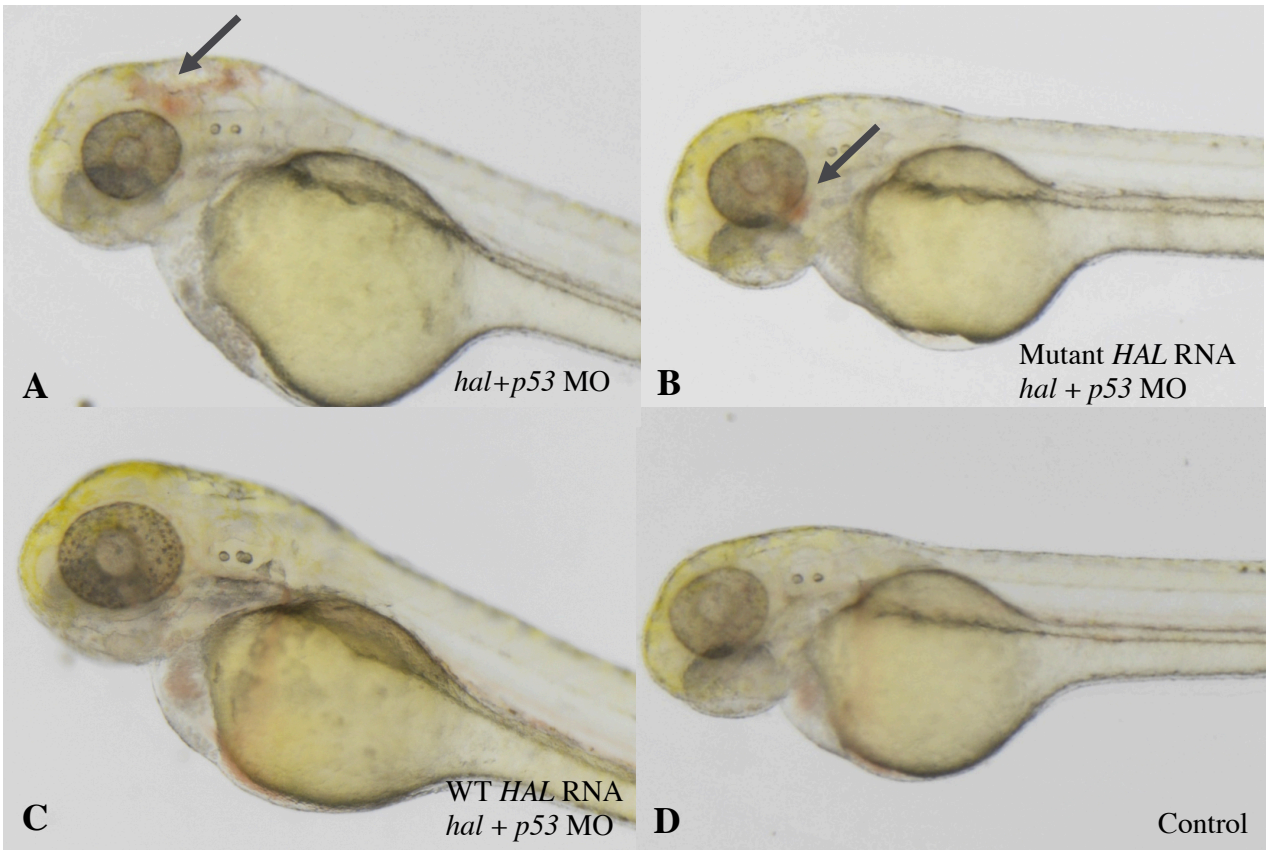


Figure 3.40. Morphant embryos in rescue experiment

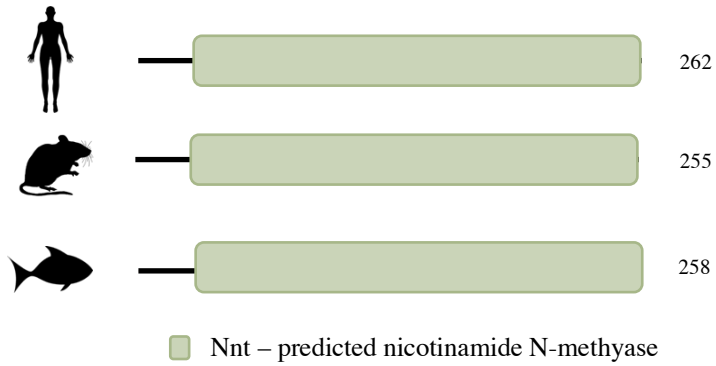
All images show zebrafish embryos at 48 hours post fertilization. The top left image (**A**) shows an embryo injected with *hal+p53* MO. The top right image (**B**) shows an embryo injected with mutant *HAL* RNA and *hal+p53* MO. The bottom left image (**C**) shows an embryo with wildtype *HAL* RNA and *hal+p53* MO. The bottom right image (**D**) shows an uninjected control embryo.

3.7.5 Conservation Between Orthologous Genes

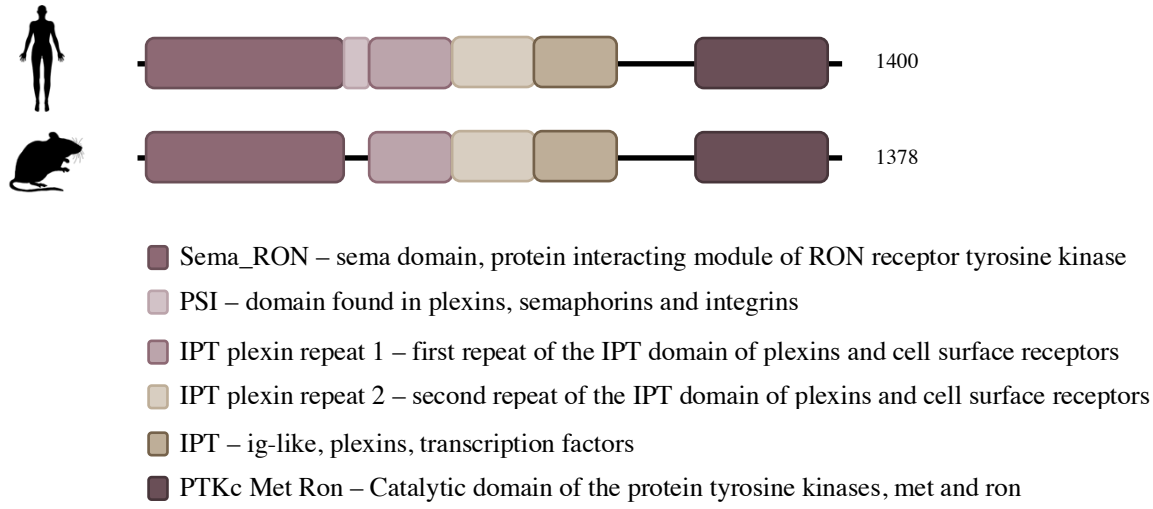
To better understand why *HAL* was the only candidate gene that produced a significant hemorrhagic phenotype, the level of conservation between orthologous genes at the protein domain level was considered (Figure 3.41). Information provided by the National Center for Biotechnology Information (NCBI) was used to compare the protein domains of each candidate gene across human, mouse and zebrafish species. The name and function of each domain, according to the NCBI, is included in Figure 3.41. All protein domains were conserved across species for both *HAL* and *METTL20.* The majority of protein domains for *MST1R* were conserved between human and mouse species, with the exception of the IPT plexin repeat 1 domain. However, there were no protein domains depicted for *MST1R* in zebrafish. Furthermore, there were also no protein domains described for *ADGRV1* in zebrafish or mouse species. The *HAL* and *METTL20* protein domain conservation across species could indicate the importance of these genes from an evolutionary standpoint.

HAL

METTL20



MST1R



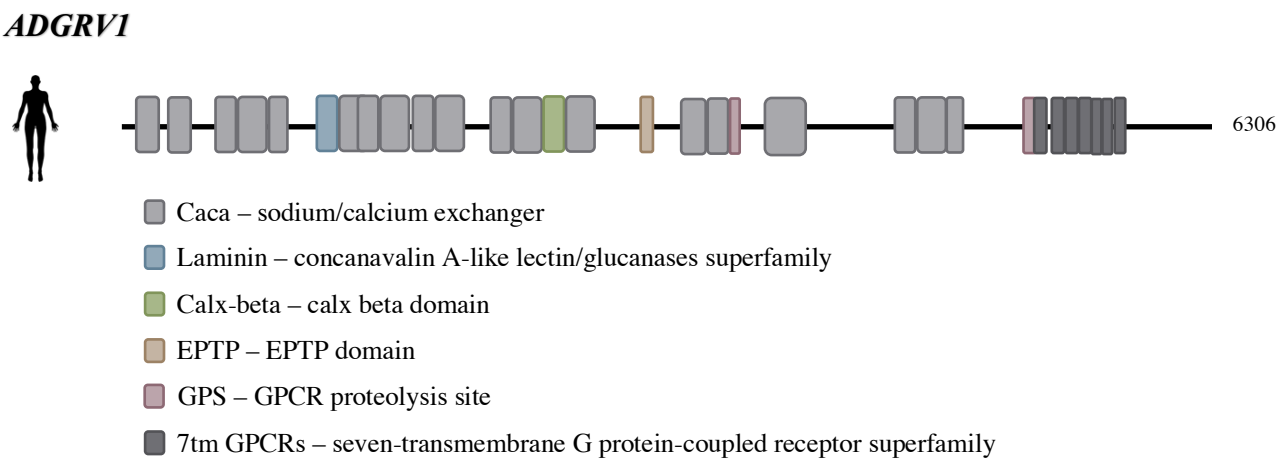


Figure 3.41. Level of conservation between orthologous genes

Analysis of conserved protein domains across species was performed using data from NCBI. The human gene for each candidate was compared to all copies of the mouse and zebrafish genes. There were no protein domains depicted for *MST1R* in zebrafish. There were no protein domains described for *ADGRV1* in zebrafish or mouse species.

4. Discussion

4.1 Reflections on Current Literature

The emergence of NGS has shifted the focus of many studies investigating the genetic component of IA from the initial linkage analyses and GWAS to the newer WES. Although these new techniques have been able to identify several different IA candidate genes, progressing this list of potential candidates into a definitive contributor to IA development has remained a challenge [92,243]. The difficulties associated with classifying genetic variants as causative is multifactorial; however, there is also a distinct lack in functional analyses accompanying WES studies of IA [243]. The impact a candidate variant may have on IA development cannot be determined without first investigating its functional significance. As NGS technologies continue to advance and WGS becomes more accessible, genetic research into IA development will likely follow. Expansion to WGS will broaden the search for genomic variations, but similar challenges in narrowing candidate variants may persist. Gene-environment interactions are another confounding variable that likely contributes to the lack of causative genes identified in IA development. Although risk factors such as smoking and hypertension are well-documented contributors, considering gene-environment interactions has not been the focus of many IA studies [243].

4.2 Overview

Validation of 18 candidate variants was completed via Sanger sequencing, and segregation analysis identified six variants from five genes (*ZNF835, METTL20, ADGRV1, MST1R,* and *HAL*) that met segregation criteria. *ZNF835* has no clear orthologue in the zebrafish genome, and therefore functional analysis was completed on the four other candidate genes. MO-mediated

gene knockdown revealed the loss of *METTL20* or *ADGRV1* orthologs in a zebrafish model does not produce a significant hemorrhagic phenotype. Thus suggesting they are not likely involved in zebrafish vascular stability. Functional analysis of *MST1R* initially revealed a hemorrhagic phenotype, but this phenotype was not apparent with the co-injection of *p53*. Alternatively, MO knockdown of *hal* demonstrated a dose-dependent hemorrhagic phenotype that could be rescued by wildtype *HAL* RNA. The *HAL* gene was previously recognized as a candidate IA gene, and the variant *HAL* (c.959C>T; p.T320M) segregated with all affected family members in Family R1262. Furthermore, *HAL* (c.959C>T; p.T320M) was predicted deleterious by 6/7 bioinformatic tools. The results of Sanger sequencing and functional analysis together provide confidence that the *HAL* (c.959C>T; p.T320M) variant is relevant to the development of IA in Family R1262. However, more evidence is needed to prove that *HAL* is pathogenically relevant in IA.

4.3 Selection of Variant Filtering Criteria

To prioritize the large number of variants identified via WES, the filtering strategy implemented had very stringent exclusion criteria. Determining the best approach to sort through WES data can be challenging with such a large number of genetic variants generated. The strategy that was ultimately developed aimed to use well-known filtering criteria that did not make any assumptions based on a gene's biological relevance. The filtering criteria implemented excluded any non-coding and synonymous variants, to prioritize variants assumed to be more likely pathogenic. Although an effective filtering strategy, it is possible that synonymous variants may have a role in IA development. The impact of synonymous variants on protein has become increasingly apparent, and their role in other human diseases such as melanoma and cystic fibrosis has been established [182,244,245]. Furthermore, the importance of noncoding regions of the

genome in disease development has become more evident in recent years [183,246]. Noncoding regions are understood to house key elements of gene regulation, such as promoters and enhancers, that undoubtedly play a role in human disease. Many Mendelian disorders are still unexplained despite the use of WES, and thus recent studies have redirected their focus to consider the contribution of noncoding regions of the genome [183,246]. The exclusion of variants with a MAF greater than 1% is another characteristic of many WES studies. Rare genetic variants are presumed more likely to be involved in human disease than those that are more common [247]. Alternatively, some WES studies of IA have utilized a less stringent 5% allele frequency cut-off. While a MAF of less than 5% was initially applied to the variants generated from WES within this study, a more stringent filtering criterion was required. Thus a MAF of less than 1% was ultimately used. The use of seven different variant pathogenicity prediction tools in the variant filtering strategy was implemented based on recommendations from TCAG. Although these bioinformatic pipelines are widely used in the filtering of WES data, some studies suggest multiple tools are required to offset the low specificity of certain tools [248]. The filtering criteria used in this project required variants to be predicted pathogenic by 4/7 of bioinformatic tools. Future studies could adjust this ratio to construct a more or less stringent strategy. To evaluate potentially pathogenic variants in families where no variants of interest have been identified, future studies may restructure the current filtering strategy. Relaxation of some of the previously used filtering criteria could broaden the group of candidate variants. Such stringent filtering criteria may have eliminated pathogenic variants involved in IA development.

Even after the implementation of the stringent filtering strategy, further prioritization of the remaining variant list was still required. When potential prioritization strategies were being explored, the WES studies by Farlow *et al*. (2015) and Yan *et al*. (2015) were fairly recent IA

studies that published lists of candidate IA genes. Prioritizing variants in previously reported candidate IA genes was an efficient way to decide which variants should be validated with Sanger sequencing [15]. As the definitive pathophysiology behind IA development has yet to be solidified, the objective was to avoid making premature assumptions based on only gene function. Although there were six variants prioritized from this phase of the filtering strategy, the resources were available to investigate additional candidates. The reduced genetic heterozygosity of the NL population reinforces the possibility that highly penetrant pathogenic variants in a single gene may predispose to IA in more than one family. Therefore, variants shared in multiple families were also prioritized. Ultimately, 18 variants were selected for validation with Sanger sequencing. Deciding on the most appropriate variant filtering strategy can be extremely challenging, primarily due to the copious amount of data generated from WES. Thus, a more stringent filtering approach was decided to be the most appropriate place to start. Future studies could relax some filtering criteria, disregard the significance of genes reported in previous IA studies, or place more importance on the functional relevance of genes. Gene expression data is another criterion that could be utilized in future endeavours. While this project did not filter variants based on a gene's biological relevance, many databases exist that provide lists of genes relevant to brain physiology. Databases such as BrainEXP or The Human Protein Atlas could be used in future studies to include gene expression data when implementing a variant filtering strategy [249–251].

4.4 Variant Segregation Analysis

When analyzing variant segregation after Sanger sequencing, only variants found within the exomes of all affected individuals within a family were considered for further functional

analysis. Such stringent segregation criteria assumes that all IAs within a family are the result of the same pathogenic variant. This strategy does not acknowledge any possible phenocopies or environmental contributions to IA development. Future studies may consider including variants shared by all but one affected individual within a family, but for this project, more stringent criteria were used.

If variants were found in all affected individuals within a family, unaffected relatives were also sequenced to evaluate variant segregation further. The use of unaffected relatives in variant prioritization in studies of IA can be challenging. The late-onset nature of IA makes it difficult to classify an individual as truly unaffected as they may eventually develop an IA. To be considered unaffected within the study, individuals had to be screened to confirm no IA was present. However, some individuals were well below the average age of IA onset. Although not used as a definitive filter, considering the age of seemingly unaffected individuals when analyzing variant segregation did introduce some subjectivity. Furthermore, it's unlikely that all individuals carrying a causative genetic variant will display the corresponding phenotype, and thus incomplete penetrance should be taken into account. For these reasons, the presence of a candidate variant in a seemingly unaffected individual was not grounds for exclusion from further functional analysis. Instead, the familial segregation for each variant was analyzed independently with the requirement that prioritized variants were absent from a large percentage of unaffected relatives.

Family R1262 was of particular interest within this study due to the strong history of both IA and AAA. Variant identification within unaffected relatives could not be used as an exclusion criterion within this family as there were only affected study participants. With one less filtering criterion applied to potential candidate variants within this family, a higher number of candidate

variants may have resulted as compared to other families. However, to compensate for the lack of unaffected study participants within Family R1262, once *HAL* (c.959C>T; p.T320M) emerged as the primary variant of interest, 200 additional population controls were Sanger sequenced.

4.5 Implications of Functional Analysis Results

Although Sanger sequencing results ultimately identified six variants from five genes (*HAL, ZNF835, MST1R, METTL20,* and *ADGRV1*) that met segregation criteria, *ZNF835* has no clear orthologue in the zebrafish genome. Therefore functional analysis was completed on the four other candidate genes.

The nonsynonymous SNV *METTL20* (c.583C>G; p.L195V) was initially deemed a candidate variant of interest. Segregation analysis revealed it segregated with all affected individuals in two unrelated NL families, R1385 and R1277. Sequencing of unaffected relatives showed the variant was found in 1/8 unaffected relatives in Family R1277 (Z1359) but was also present in the one unaffected study participant in Family R1385 (Z1149). The CT scans of both Z1359 and Z1149 were completed when these participants were slightly younger than the average age of IA onset, and thus functional analysis was still completed. The MO-mediated knockdown of the *METTL20* zebrafish ortholog did not produce a significant hemorrhagic phenotype at two different MO doses compared to that of controls. These results do not support the involvement of the *METTL20* gene in zebrafish vascular stability, and alternative genetic contributions should be considered for these families.

Functional analysis into *ADGRV1* produced similar results to those of *METTL20*. The *ADGRV1* gene has been previously associated with familial febrile seizures and Usher syndrome, a disorder characterized by vision and hearing loss [234,235]. There were two different variants

within this gene that were found in two unrelated NL families. The variant *ADGRV1* (c.1522A>C; p.I508L) segregated with both affected individuals in Family R1262 and the variant *ADGRV1* (c.6317C>T; p.A2106V) segregated with all affected individuals in Family R1276 and was found in only 2/52 unaffected study participants. The extensive number of unaffected study participants recruited within Family R1276 is due to the initial classification of Z1415 as an affected family member. Subsequent investigations revealed that Z1415 is an unaffected individual who does not carry the *ADGRV1* (c.6317C>T; p.A2106V) variant. Taking a closer look at the unaffected carriers of *ADGRV1* (c.6317C>T; p.A2106V), specifically Z1027, produced some interesting results. The strong family history of IA is from the maternal side of the family (relative to Z1027). The mother of Z1027, Z1021, is also an unaffected study participant, but she does not carry the *ADGRV1* (c.6317C>T; p.A2106V) variant. Sanger sequencing was completed multiple times on DNA from Z1027 to confirm the presence of *ADGRV1* (c.6317C>T; p.A2106V). The father of Z1027 is not a study participant. While the possibility remains that *ADGRV1* (c.6317C>T; p.A2106V) was transmitted paternally to Z1027, a sample mix-up during DNA extraction is another plausible explanation and can occur at a rate >1%. The lack of a significant hemorrhagic phenotype with the genetic knockdown of the *ADGRV1* zebrafish ortholog does not provide confidence that this gene is contributing to vascular stability in zebrafish.

The predicted impact of *MST1R* (c.4129C>T; p.Q1377X) as a stopgain mutation made it an enticing candidate from the beginning. Furthermore, this was a novel variant not previously described and was found to segregate with all affected individuals in two unrelated NL families, R1262 and R1385. The only unaffected study participant within Family R1381, Z1149, was also a carrier of *MST1R* (c.4129C>T; p.Q1377X) but again was only 41 years old at the time of her

last CT scan. Non-IA phenotypes are also associated with the *MST1R* gene. Increased expression of *MST1R* has been noted in many epithelial neoplasms, and WES has identified it as a susceptibility gene in nasopharyngeal carcinoma [252–254]. During the initial stages of functional analysis, *MST1R* still appeared to be a favorable candidate gene in IA development. The co-injection of *mst1ra+b* produced a dose-dependent hemorrhagic phenotype at the two lowest doses of MO (3 ng and 9 ng). To avoid the off-target effects observed at higher doses of MO, embryos were co-injected with *p53* when the dose of MO was increased. The possible contribution of *MST1R* mutations to IA development became less plausible as *mst1ra+b* morphants did not produce a significant hemorrhage phenotype when co-injected with *p53*. These results suggested that the initially observed phenotype was not specific to the *mst1ra+b* knockdown and instead was due to off-target effects of *p53*. These results did not produce confidence moving forward that *MST1R* (c.4129C>T; p.Q1377X) is a causative variant in IA development. Consequently, *HAL* became the only remaining candidate gene of interest. While the *HAL* gene became the focus of further investigations, it is important to recognize that the absence of a visible hemorrhagic phenotype in zebrafish morphants cannot definitively rule out the contribution of a gene to IA development. The results from this project do not support a role for *METTL20, ADGRV1,* or *MST1R* in the vascular stability of zebrafish, but further investigations are required to examine their connection to human IA.

The *HAL* gene was initially highlighted as a candidate IA gene by Farlow *et al*. (2015) who performed WES on 45 individuals affected with IA that were recruited as part of the FIA study. The variant *HAL* (c.G1472T; p.G491V) was prioritized after meeting all variant filtering criteria and segregated with IA in one of the seven study families. An alternative variant was identified in Family R1262, *HAL* (c.959C>T; p.T320M), that segregated with all affected family

members. The *HAL* (c.959C>T; p.T320M) variant was identified in 1/300 NL population controls, a slightly higher prevalence than the 0.024% MAF reported by ExAC. While it is possible *HAL* (c.959C>T; p.T320M) is more prevalent in the NL population, it is also a possibility this individual has an IA. The population controls were available from the NFCCR and were never screened for IAs.

Once functional analysis began, knockdown of the *HAL* zebrafish ortholog at lower MO doses produced an observable hemorrhagic phenotype. The dose of *hal* MO was subsequently increased, embryos were co-injected with *p53* MO, and a significant dose-dependent hemorrhagic phenotype was still observed. Although these results indicate that the hemorrhagic phenotype was not due to off-target effects of *p53*, previous research suggests that co-injection of *p53* alone is not enough to ensure a phenotype is gene specific, and additional control measures should be taken [140,142]. The results of the RNA rescue experiment confirm the specificity of the *hal* MO. Injection of wildtype human *HAL* RNA in the *hal* MO background was able to rescue the hemorrhagic phenotype. This demonstrates that there must be some functional overlap between human and zebrafish *HAL* as human RNA can rescue the zebrafish phenotype. The mutant *HAL* RNA was not able to rescue the morpholino phenotype, suggesting that the variant causes a change in protein function. During the initial rescue experiment, RNA concentrations of 200 ng/ul resulted in both the wildtype *HAL* RNA and mutant *HAL* RNA to rescue the hemorrhagic phenotype. Such a high concentration of RNA is much greater than the physiologic levels typical of a zebrafish embryo. Overloading the embryo with excess RNA could result in both wildtype and mutant RNA rescuing the phenotype, assuming that the variant does not generate a null allele. Once the experiment was repeated with a diluted RNA

concentration (100 ng/ul) at 200 ng/embryo, it became apparent that only the wildtype *HAL* RNA was able to rescue the hemorrhagic phenotype.

The phenotype observed in all MO-mediated gene knockdown experiments completed throughout this project was hemorrhaging within the zebrafish and not the development of IAs specifically. This lack of aneurysm development was a drawback discussed by Santiago-Sim *et al*. (2016) and is not specific to this project [105]. There was no aneurysm development observed in their mouse or zebrafish knockdowns, and instead, a cerebral hemorrhage phenotype was observed [105]. They suggested that as is the case with humans, small IAs often go unnoticed. The size of an IA may render it indistinguishable, but it is also possible that zebrafish do not develop aneurysms at all, and weakening of the vasculature instead leads to a hemorrhagic phenotype [105]. A lack of visible aneurysm in this model system makes it challenging to determine the clinical significance of the variants that have been identified. The vascular dysfunction that was demonstrated in *HAL* morphants could potentially represent the vascular dysfunction that is involved in human IA development, but further functional work is needed to confirm this. Furthermore, the hemorrhages produced by *hal* knockdown were not always localized to the head of the zebrafish. Hemorrhages identified within the body of the zebrafish were also considered a hemorrhagic phenotype. In humans, the co-occurrence between AAA and IA has been widely recognized but not well understood. With such a strong family history of IA and AAA in Family R1262, the same genetic variant underlying IA development could also be contributing to AAA development. The results of the functional analysis, as well as the rescue experiment, suggest that *HAL* (c.959C>T; p.T320M) could be contributing to IA development in Family R1262. Future research is needed to investigate the mechanism underlying *HAL* function and assess the broader contribution of *HAL* mutations to aneurysm formation.

4.6 Limitations

The many appealing aspects of WES that have been very beneficial in gene discovery also underlie many of the challenges associated with this technology. By using of WES, all genetic variants within an individual's exome can be identified. However, even after initial filtering steps thousands of genetic variants remained. The challenge becomes determining an appropriate filtering approach that is stringent enough to narrow the search while still broad enough to not prematurely exclude possible candidates. There are also limitations associated with exome sequencing coverage itself. Although the majority of Mendelian diseases have been identified in the protein-coding region of the genome, this still only makes up a small proportion of the human genome. Data generated from WES will not be able to identify genomic variations outside this region that may be contributing to IA development. The current study design was based on the assumption IA follows a monogenic model of inheritance. But the possibility remains that a much more complex inheritance pattern involving the interaction of multiple genes could be contributing to IA development in some families. Various environmental risk factors such as smoking, and hypertension have also been documented in many study participants that could be contributing to the development of IA. These risk factors were not considered when analyzing segregation data, yet they may have had a significant impact on IA formation in some patients.

While this project had numerous large, well-documented families available, there is a lack of recent clinical follow-up data available. The majority of CT scans performed on these participants were completed almost ten years ago. Many seemingly unaffected individuals at the time of screening may have since developed IA, and thus follow-up data would have been beneficial for segregation analysis. Clinical data regarding known IA risk factors such as

smoking and hypertension were recorded during participant recruitment. However, this data was not significantly considered throughout the study, as only minimal details were available. More detailed, and up to date clinical data could have helped distinguish possible phenocopies during segregation analysis.

The controls used in this project were from the NFCCR. The only information available about these individuals is that participants had no known personal or family history of cancer. Ideally, population controls would have been recruited exclusively for this project. However, financial constraints made NFCCR controls the next best option. Future studies could look at sequencing an additional number of controls to increase accuracy or recruit controls specific to the project.

In addition to the advantages of using zebrafish as a functional model of IA development, there are also limitations specific to a morpholino-based technique. Significant dosage based, off-target effects have been consistently seen at high doses of MO [201]. Although specific precautions were followed to ensure accurate functional analysis, such as co-injection with *p53*, possible off-target effects should always be considered when using MOs. Furthermore, it cannot be disregarded that a limited number of studies have used MO-mediated gene knockdown to investigate an IA phenotype specifically. Therefore, it is important to be aware of the possible challenges in relying on a zebrafish as the sole functional model, and additional functional assays should be explored.

4.7 Future Directions

Due to the inability to complete MO-mediated gene knockdown of *ZNF835* (c.378delC; p.I126fs) within this project, an alternative method of functional analysis would be worth

pursuing in future investigations. There is limited information available about the functional relevance of the *ZNF835* gene. However, it was identified as a candidate IA gene by Farlow *et al*. (2015) and a novel frame-shift deletion that segregated with all affected individuals was identified in Family R1381. The only seemingly unaffected relative identified to have the *ZNF835* (c.378delC; p.I126fs) variant was only 25 at the time of his last CT scan. Without any more recent clinical data, it is difficult to know that he is truly unaffected. The identification of a novel variant within a previously identified candidate gene, combined with the variant segregation within Family R1381, resulted in *ZNF835* (c.378delC; p.I126fs) as the initial favorable candidate. Future investigations should explore alternative functional investigations for *ZNF835* as it is becoming increasingly clear that multiple variants, in different genes, are likely contributing to IA within different families.

To examine the *HAL* (c.959C>T; p.T320M) variant frequency within the general NL population, Sanger sequencing was completed on 300 population controls. However, the larger contribution of *HAL* to IA formation could also be investigated by sequencing unrelated IA probands for the *HAL* (c.959C>T; p.T320M) variant or any other potentially pathogenic variants within this gene. Future studies could also use statistical analysis to provide valuable insight into the likelihood *HAL* (c.959C>T; p.T320M) is relevant to IA development.

Additionally, the involvement of *HAL* in histidine catabolism could be the focus of future investigations. Histidine ammonia-lyase catalyzes the conversion of l-histidine to urocanate, and thus, a loss of function *HAL* variant can result in excess levels of histidine (histidinemia). Most patients with histidinemia are thought to be asymptomatic. However, as an essential amino acid that crosses the blood-brain barrier, histidine may have many additional functions not yet established [227,228]. Histidinemia can be identified by measuring the histidine concentration in a

blood or urine sample. Consequently, future projects could measure histidine levels to determine if the *HAL* (c.959C>T; p.T320M) variant causes an increased concentration of histidine in affected individuals. Furthermore, other intermediates in histidine catabolism could also be explored in future projects to help clarify histidine's role in IA development.

Guidelines on the use of MOs in zebrafish provide many suggestions that could be the focus of future directions to verify the contribution of *HAL* mutations in IA development. Ideally, morphant phenotypes would be compared to mutant phenotypes generated via genome–editing tools. Such tools include CRISPR/CAS9 or Transcription Activator–Like Effector Nucleases (TALEN) [142]. Future studies could aim to generate a mutant line using a genome-editing tool. Mutant phenotypes could then be assessed to ensure the MO induced morphant phenotype is similar and not more severe. Another validation suggestion that was recently described is the use of multiple MOs [142]. Future studies could use a *HAL* MO targeted to an alternative oligonucleotide sequence or use a translation blocking MO in addition to the splice blocking type used in this project. Functional analysis of *HAL* in a different animal model is another future goal of the project. This could provide additional evidence that *HAL* knockdown results in a hemorrhagic phenotype.

Although the results of this study provide evidence that *HAL* (c.959C>T; p.T320M) could be contributing to IA development in Family R1262, there are twelve other families within this cohort, without *HAL* variants. Likely multiple genes are contributing to IA development, and future studies could use a similar project framework with a modified variant filtering strategy. Another possibility includes looking for a genetic contribution outside the exome by employing the use of whole genome sequencing analysis. Furthermore, the impact of environmental factors on IA development should not be overlooked. Underlying factors, in addition to those previously

identified, such as smoking, hypertension, and alcohol consumption, may have a much more significant impact on IA development than what is currently presumed.

4.8 Conclusions

Focusing on a large cohort of multiple NL families with a strong genetic predisposition to IA, this project was able to identify a potentially pathogenic variant relevant to the development of IA. The variant filtering strategy that was implemented in 43 individuals affected with IA across 13 NL families identified 18 candidate variants. With the use of Sanger sequencing, segregation analysis of all 18 variants identified five candidate genes worth pursuing further functional analysis. Without a clear zebrafish ortholog, *ZNF835* was not able to undergo functional analysis; however, *ZNF835* (c.378delC; p.I126fs) remains an exceptional candidate variant in IA development worth pursuing in future studies. The MO-mediated gene knockdown of *ADGRV1, METTL20, MST1R,* and *HAL* established that *hal* morphants demonstrate a dose-dependent hemorrhagic phenotype that can be rescued with wildtype *HAL* RNA. These results suggest that the *HAL* (c.959C>T; p.T320M) variant could be contributing to IA development in Family R1262. Future analysis of the *HAL* gene could be beneficial in determining the extent of its involvement in IA development and contribute to an improved understanding of IA pathogenesis.

References

1. Williams LN, Brown RD. Management of unruptured intracranial aneurysms. *Neurol Clin Pract*. 2013;3(2):99-108. doi:10.1212/CPJ.0b013e31828d9f6b
2. Vernooij MW, Ikram MA, Tanghe HL, et al. Incidental Findings on Brain MRI in the General Population. *N Engl J Med*. 2007. doi:10.1056/nejmoa070972
3. Vlak MHM, Algra A, Brandenburg R, Rinkel GJE. Prevalence of unruptured intracranial aneurysms, with emphasis on sex, age, comorbidity, country, and time period: A systematic review and meta-analysis. *Lancet Neurol*. 2011. doi:10.1016/S1474-4422(11)70109-0
4. Imaizumi Y, Mizutani T, Shimizu K, Sato Y, Taguchi J. Detection rates and sites of unruptured intracranial aneurysms according to sex and age: An analysis of MR angiography–based brain examinations of 4070 healthy Japanese adults. *J Neurosurg*. 2019. doi:10.3171/2017.9.JNS171191
5. Leonardo A, Manoel O, Turkel-parrella D, Duggal A, Murphy A, Marotta TR. Managing aneurysmal subarachnoid hemorrhage : It takes a team. *Cleve Clin J Med*. 2015;82(3):177-192. doi:10.3949/ccjm.82a.14021
6. Dorsch NWC, King MT. A review of cerebral vasospasm in aneurysmal subarachnoid haemorrhage Part I: Incidence and effects. *J Clin Neurosci*. 1994. doi:10.1016/0967-5868(94)90005-1
7. Lucke-Wold BP, Logsdon AF, Manoranjan B, et al. Aneurysmal subarachnoid hemorrhage and neuroinflammation: A comprehensive review. *Int J Mol Sci*. 2016. doi:10.3390/ijms17040497
8. Lawton MT, Vates GE. Subarachnoid hemorrhage. *N Engl J Med*. 2017. doi:10.1056/NEJMcp1605827
9. Frösen J, Tulamo R, Paetau A, et al. Saccular intracranial aneurysm: Pathology and mechanisms. *Acta Neuropathol*. 2012. doi:10.1007/s00401-011-0939-3
10. Turan N, Heider RAJ, Zaharieva D, Ahmad FU, Barrow DL, Pradilla G. Sex Differences in the Formation of Intracranial Aneurysms and Incidence and Outcome of Subarachnoid Hemorrhage: Review of Experimental and Human Studies. *Transl Stroke Res*. 2016. doi:10.1007/s12975-015-0434-6
11. Keedy A. An overview of intracranial aneurysms. *McGill J Med*. 2006.
12. Signorelli F, Sela S, Gesualdo L, et al. Hemodynamic Stress, Inflammation, and Intracranial Aneurysm Development and Rupture: A Systematic Review. *World Neurosurg*. 2018. doi:10.1016/j.wneu.2018.04.143
13. Chipolla. Chapter 2 Anatomy and ultrastructure. *Boek*. 2009.
14. Lee RMKW. Morphology of cerebral arteries. *Pharmacol Ther*. 1995. doi:10.1016/0163-7258(94)00071-A
15. Etminan N, Rinkel GJ. Unruptured intracranial aneurysms: Development, rupture and preventive management. *Nat Rev Neurol*. 2016. doi:10.1038/nrneurol.2016.150
16. Chalouhi N, Hoh BL, Hasan D. Review of cerebral aneurysm formation, growth, and rupture. *Stroke*. 2013;44(12):3613-3622. doi:10.1161/STROKEAHA.113.002390
17. Fox JL, Stehbens WE. The Pathology of Intracranial Arterial Aneurysms and Their Complications. In: *Intracranial Aneurysms*. ; 1983. doi:10.1007/978-1-4612-5437-9_11
18. Austin GM, Fisher S, Dickson D, Anderson D, Richardson S. The significance of the extracellular matrix in intracranial aneurysms. *Ann Clin Lab Sci*. 1993.

19. Schievink W. Intracranial Aneurysms. *N Engl J Med*. 1997.
20. Chalouhi N, Ali MS, Jabbour PM, et al. Biology of intracranial aneurysms: role of inflammation. *J Cereb Blood Flow Metab*. 2012;32(9):1659-1676. doi:10.1038/jcbfm.2012.84
21. Stehbens WE. Etiology of intracranial berry aneurysms. *J Neurosurg*. 2009. doi:10.3171/jns.1989.70.6.0823
22. Kondo S, Hashimoto N, Kikuchi H, Hazama F, Nagata I, Kataoka H. Cerebral aneurysms arising at nonbranching sites: An experimental study. *Stroke*. 1997. doi:10.1161/01.STR.28.2.398
23. Shojima M, Oshima M, Takagi K, et al. Magnitude and role of wall shear stress on cerebral aneurysm: Computational fluid dynamic study of 20 middle cerebral artery aneurysms. In: *Stroke*. ; 2004. doi:10.1161/01.STR.0000144648.89172.0f
24. Meng H, Wang Z, Hoi Y, et al. Complex hemodynamics at the apex of an arterial bifurcation induces vascular remodeling resembling cerebral aneurysm initiation. *Stroke*. 2007. doi:10.1161/STROKEAHA.106.481234
25. Starke RM, Chalouhi N, Ding D, et al. Vascular Smooth Muscle Cells in Cerebral Aneurysm Pathogenesis. *Transl Stroke Res*. 2014. doi:10.1007/s12975-013-0290-1
26. Nakajima N, Nagahiro S, Sano T, Satomi J, Satoh K. Phenotypic modulation of smooth muscle cells in human cerebral aneurysmal walls. *Acta Neuropathol*. 2000. doi:10.1007/s004010000220
27. Bruno G, Todor R, Lewis I, Chyatte D. Vascular extracellular matrix remodeling in cerebral aneurysms. *J Neurosurg*. 2009. doi:10.3171/jns.1998.89.3.0431
28. Aoki T, Kataoka H, Morimoto M, Nozaki K, Hashimoto N. Macrophage-derived matrix metalloproteinase-2 and -9 promote the progression of cerebral aneurysms in rats. *Stroke*. 2007;38(1):162-169. doi:10.1161/01.STR.0000252129.18605.c8
29. Hashimoto N, Handa H, Hazama F. Experimentally induced cerebral aneurysms in rats. *Surg Neurol*. 1978.
30. Kanematsu Y, Kanematsu M, Kurihara C, et al. Critical roles of macrophages in the formation of intracranial aneurysm. *Stroke*. 2011. doi:10.1161/STROKEAHA.110.590976
31. Zhang X, Ares W, Taussky P, Ducruet A, Grandhi R. Role of matrix metalloproteinases in the pathogenesis of intracranial aneurysms. *Neurosurg Focus*. 2019.
32. Keow JY, Pond ED, Cisar JS, Cravatt BF, Crawford BD. Activity-Based Labeling of Matrix Metalloproteinases in Living Vertebrate Embryos. *PLoS One*. 2012. doi:10.1371/journal.pone.0043434
33. Janssens E, Gaublomme D, de Groef L, et al. Matrix Metalloproteinase 14 in the Zebrafish: An Eye on Retinal and Retinotectal Development. *PLoS One*. 2013. doi:10.1371/journal.pone.0052915
34. Alg V, Ke X, Grieve J, et al. Association of functional MMP-2 gene variant with intracranial aneurysms: case-control genetic association study and meta-analysis. *Br J Neurosurg*. 2018.
35. Juvela S, Poussa K, Porras M. Factors affecting formation and growth of intracranial aneurysms: A long-term follow-up study. *Stroke*. 2001. doi:10.1161/01.STR.32.2.485
36. Thompson BG, Brown RD, Amin-Hanjani S, et al. Guidelines for the Management of Patients With Unruptured Intracranial Aneurysms. *Stroke*. 2015. doi:10.1161/str.0000000000000070
37. Feng X, Qian Z, Zhang B, et al. Number of cigarettes smoked per day, smoking index, and

intracranial aneurysm rupture: A case-control study. *Front Neurol*. 2018. doi:10.3389/fneur.2018.00380
38. Lindgren AE, Kurki MI, Riihinen A, et al. Hypertension predisposes to the formation of saccular intracranial aneurysms in 467 unruptured and 1053 ruptured patients in Eastern Finland. *Ann Med*. 2014. doi:10.3109/07853890.2014.883168
39. Klatsky AL, Armstrong MA, Friedman GD. Alcohol use and subsequent cerebrovascular disease hospitalizations. *Stroke*. 1989. doi:10.1161/01.STR.20.6.741
40. Can A, Castro VM, Ozdemir YH, et al. Alcohol Consumption and Aneurysmal Subarachnoid Hemorrhage. *Transl Stroke Res*. 2018. doi:10.1007/s12975-017-0557-z
41. Wiebers DO. Unruptured intracranial aneurysms: Natural history, clinical outcome, and risks of surgical and endovascular treatment. *Lancet*. 2003. doi:10.1016/S0140-6736(03)13860-3
42. Juvela S, Poussa K, Lehto H, Porras M. Natural history of unruptured intracranial aneurysms: A long-term follow-up study. *Stroke*. 2013. doi:10.1161/STROKEAHA.113.001838
43. Adams S. The Natural Course of Unruptured Cerebral Aneurysms in a Japanese Cohort. *J Emerg Med*. 2012. doi:10.1016/j.jemermed.2012.10.009
44. Van Kammen MS, Moomaw CJ, Van Der Schaaf IC, et al. Heritability of circle of Willis variations in families with intracranial aneurysms. *PLoS One*. 2018. doi:10.1371/journal.pone.0191974
45. Stojanovic N, Kostic A, Mitic R, Berilazic L, Radisavljevic M. Association between Circle of Willis Configuration and Rupture of Cerebral Aneurysms. *Medicina (B Aires)*. 2019.
46. Zhang J, Claterbuck RE. Molecular genetics of human intracranial aneurysms. *Int J Stroke*. 2008;3(4):272-287. doi:10.1111/j.1747-4949.2008.00224.x
47. Hitchcock E, Gibson WT. A Review of the Genetics of Intracranial Berry Aneurysms and Implications for Genetic Counseling. *J Genet Couns*. 2017. doi:10.1007/s10897-016-0029-8
48. Kim ST, Brinjikji W, Kallmes DF. Prevalence of intracranial aneurysms in patients with connective tissue diseases: A retrospective study. *Am J Neuroradiol*. 2016. doi:10.3174/ajnr.A4718
49. Ellis JA, Nossek E, Kronenburg A, Langer DJ, Ortiz RA. Intracranial Aneurysm: Diagnostic Monitoring, Current Interventional Practices, and Advances. *Curr Treat Options Cardiovasc Med*. 2018. doi:10.1007/s11936-018-0695-y
50. Ajiboye N, Chalouhi N, Starke RM, Zanaty M, Bell R. Unruptured Cerebral Aneurysms: Evaluation and Management. *Sci World J*. 2015. doi:10.1155/2015/954954
51. White PM, Teasdale EM, Wardlaw JM, Easton V. Intracranial Aneurysms: CT Angiography and MR Angiography for Detection—Prospective Blinded Comparison in a Large Patient Cohort. *Radiology*. 2013. doi:10.1148/radiology.219.3.r01ma16739
52. Williams L, Brown R. Management of unruptured intracranial aneurysms. *Neurol Clin Pract*. 2013.
53. Flahault A, Trystram D, Nataf F, et al. Screening for intracranial aneurysms in autosomal dominant polycystic kidney disease is cost-effective. *Kidney Int*. 2018. doi:10.1016/j.kint.2017.08.016
54. Greving JP, Wermer MJH, Brown RD, et al. Development of the PHASES score for prediction of risk of rupture of intracranial aneurysms: A pooled analysis of six

prospective cohort studies. *Lancet Neurol*. 2014. doi:10.1016/S1474-4422(13)70263-1
55. Etminan N, Brown RD, Beseoglu K, et al. The unruptured intracranial aneurysm treatment score: A multidis ciplinary consensus. *Neurology*. 2015. doi:10.1212/WNL.0000000000001891
56. Backes D, Rinkel GJE, Greving JP, et al. ELAPSS score for prediction of risk of growth of unruptured intracranial aneurysms. *Neurology*. 2017. doi:10.1212/WNL.0000000000003865
57. Louw DF, Asfora WT, Sutherland GR. A brief history of aneurysm clips. *Neurosurg Focus*. 2008. doi:10.3171/foc.2001.11.2.5
58. David CA, Vishteh AG, Spetzler RF, Lemole M, Lawton MT, Partovi S. Late angiographic follow-up review of surgically treated aneurysms. *J Neurosurg*. 2009. doi:10.3171/jns.1999.91.3.0396
59. Kotowski M, Naggara O, Darsaut TE, et al. Safety and occlusion rates of surgical treatment of unruptured intracranial aneurysms: A systematic review and meta-analysis of the literature from 1990 to 2011. *J Neurol Neurosurg Psychiatry*. 2013. doi:10.1136/jnnp-2011-302068
60. Brown RD, Broderick JP. Unruptured intracranial aneurysms: Epidemiology, natural history, management options, and familial screening. *Lancet Neurol*. 2014;13(4):393-404. doi:10.1016/S1474-4422(14)70015-8
61. Wills S, Ronkainen A, Van der Voet M, et al. Familial intracranial aneurysms: An analysis of 346 multiplex Finnish families. *Stroke*. 2003;34(6):1370-1374. doi:10.1161/01.STR.0000072822.35605.8B
62. Foroud T. Whole exome sequencing of intracranial aneurysm. In: *Stroke*. ; 2013. doi:10.1161/STROKEAHA.113.001174
63. Broderick JP, Sauerbeck LR, Foroud T, et al. The Familial Intracranial Aneurysm (FIA) study protocol. *BMC Med Genet*. 2005. doi:10.1186/1471-2350-6-17
64. Broderick JP, Brown RD, Sauerbeck L, et al. Greater rupture risk for familial as compared to sporadic unruptured intracranial aneurysms. *Stroke*. 2009;40(6):1952-1957. doi:10.1161/STROKEAHA.108.542571
65. Bailey-Wilson JE, Wilson AF. Linkage analysis in the next-generation sequencing era. *Hum Hered*. 2011. doi:10.1159/000334381
66. Zhou S, Dion P, Rouleau G. Genetics of Intracranial Aneurysms. *Stroke*. 2018;49(3):780-787.
67. Foroud T, Sauerbeck L, Brown R, et al. Genome screen to detect linkage to intracranial aneurysm susceptibility genes: The familial intracranial aneurysm (FIA) study. *Stroke*. 2008;39(5):1434-1440. doi:10.1161/STROKEAHA.107.502930
68. Foroud T, Sauerbeck L, Brown R, et al. Genome screen in familial intracranial aneurysm. *BMC Med Genet*. 2009;10:3. doi:10.1186/1471-2350-10-3
69. Onda H, Kasuya H, Yoneyama T, et al. Genomewide-linkage and haplotype-association studies map intracranial aneurysm to chromosome 7q11. *Am J Hum Genet*. 2001;69(4):804-819. doi:10.1086/323614
70. Farnham JM, Camp NJ, Neuhausen SL, et al. Confirmation of chromosome 7q11 locus for predisposition to intracranial aneurysm. *Hum Genet*. 2004;114(3):250-255. doi:10.1007/s00439-003-1044-z
71. Olson J, Vongpunsawad S, Kuivaniemi H, et al. Search for intracranial aneurysm susceptibility gene(s) using Finnish families. *BMC Med Genet*. 2002;3(1):7.

doi:10.1186/1471-2350-3-7
72. Nahed B V, Seker A, Guclu B, et al. Mapping a Mendelian form of intracranial aneurysm to 1p34.3-p36.13. *Am J Hum Genet*. 2005;76(1):172-179. doi:10.1086/426953
73. Ruigrok YM, Wijmenga C, Rinkel GJE, et al. Genomewide Linkage in a Large Dutch Family With Intracranial Aneurysms: Replication of 2 Loci for Intracranial Aneurysms to Chromosome 1p36.11-p36.13 and Xp22.2-p22.32. *Stroke*. 2008;39(4):1096-1102. doi:10.1161/STROKEAHA.107.495168
74. Roos YBWEM, Pals G, Struycken PM, et al. Genome-wide linkage in a large Dutch consanguineous family maps a locus for intracranial aneurysms to chromosome 2p13. *Stroke*. 2004. doi:10.1161/01.STR.0000141415.28155.46
75. Verlaan DJ, Dubé MP, St-Onge J, et al. A new locus for autosomal dominant intracranial aneurysm, ANIB4, maps to chromosome 5p15.2-14.3. *J Med Genet*. 2006. doi:10.1136/jmg.2005.033209
76. Kim CJ, Park SS, Lee HS, et al. Identification of an autosomal dominant locus for intracranial aneurysm through a model-based family collection in a geographically limited area. *J Hum Genet*. 2011. doi:10.1038/jhg.2011.27
77. Ozturk AK, Nahed B V., Bydon M, et al. Molecular genetic analysis of two large kindreds with intracranial aneurysms demonstrates linkage to 11q24-25 and 14q23-31. *Stroke*. 2006;37(4):1021-1027. doi:10.1161/01.STR.0000206153.92675.b9
78. Santiago-Sim T, Depalma SR, Ju KL, et al. Genomewide linkage in a large caucasian family maps a new locus for intracranial aneurysms to chromosome 13q. In: *Stroke*. Vol 40. ; 2009. doi:10.1161/STROKEAHA.108.534396
79. Yamada S, Utsunomiya M, Inoue K, et al. Genome-wide scan for Japanese familial intracranial aneurysms: Linkage to several chromosomal regions. *Circulation*. 2004;110(24):3727-3733. doi:10.1161/01.CIR.0000143077.23367.18
80. van der Voet M, Olson JM, Kuivaniemi H, et al. Intracranial Aneurysms in Finnish Families: Confirmation of Linkage and Refinement of the Interval to Chromosome 19q13.3. *Am J Hum Genet*. 2004. doi:10.1086/382285
81. Mineharu Y, Inoue K, Inoue S, et al. Model-based linkage analyses confirm chromosome 19q13.3 as a susceptibility locus for intracranial aneurysm. *Stroke*. 2007. doi:10.1161/01.STR.0000259657.73682.03
82. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet*. 1996.
83. Tromp G, Weinsheimer S, Ronkainen A, Kuivaniemi H. Molecular basis and genetic predisposition to intracranial aneurysm. *Ann Med*. 2014;46(8):597-606. doi:10.3109/07853890.2014.949299
84. Bilguvar K, Yasuno K, Niemelä M, et al. Susceptibility loci for intracranial aneurysm in European and Japanese populations. *Nat Genet*. 2008. doi:10.1038/ng.240
85. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 2005. doi:10.1038/nrg1521
86. Yasuno K, Bilguvar K, Bijlenga P, et al. Genome-wide association study of intracranial aneurysm identifies three new risk loci. *Nat Genet*. 2010. doi:10.1038/ng.563
87. Hussain I, Duffis EJ, Gandhi CD, Prestigiacomo CJ. Genome-wide association studies of intracranial aneurysms: An update. *Stroke*. 2013. doi:10.1161/STROKEAHA.113.001753
88. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards suite: From gene data mining to disease genome sequence analyses. *Curr Protoc Bioinforma*. 2016. doi:10.1002/cpbi.5

89. Pasmant E, Sabbagh A, Vidaud M, Bièche I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J*. 2011. doi:10.1096/fj.10-172452
90. Holdt LM, Teupser D. Recent studies of the human chromosome 9p21 locus, which is associated with atherosclerosis in human populations. *Arterioscler Thromb Vasc Biol*. 2012. doi:10.1161/ATVBAHA.111.232678
91. Foroud T, Koller DL, Lai D, et al. Genome-wide association study of intracranial aneurysms confirms role of anril and SOX17 in disease risk. *Stroke*. 2012. doi:10.1161/STROKEAHA.112.656397
92. Xu Z, Rui YN, Hagan JP, Kim DH. Intracranial Aneurysms: Pathology, Genetics, and Molecular Mechanisms. *NeuroMolecular Med*. 2019. doi:10.1007/s12017-019-08537-7
93. Helgadottir A, Thorleifsson G, Magnusson KP, et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet*. 2008. doi:10.1038/ng.72
94. Corada M, Orsenigo F, Morini MF, et al. Sox17 is indispensable for acquisition and maintenance of arterial identity. *Nat Commun*. 2013. doi:10.1038/ncomms3609
95. Lee S, Kim IK, Ahn JS, et al. Deficiency of endothelium-specific transcription factor Sox17 induces intracranial aneurysm. *Circulation*. 2015. doi:10.1161/CIRCULATIONAHA.114.012568
96. Liu M, Zhang L, Marsboom G, et al. Sox17 is required for endothelial regeneration following inflammation-induced vascular injury. *Nat Commun*. 2019. doi:10.1038/s41467-019-10134-y
97. Zhou S, Gan-Or Z, Ambalavanan A, et al. Genome-wide association analysis identifies new candidate risk loci for familial intracranial aneurysm in the French-Canadian population. *Sci Rep*. 2018. doi:10.1038/s41598-018-21603-7
98. van't Hof FNG, Ruigrok YM, Lee CH, et al. Shared genetic risk factors of intracranial, abdominal, and thoracic aneurysms. *J Am Heart Assoc*. 2016. doi:10.1161/JAHA.115.002603
99. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*. 2010. doi:10.1038/ng.499
100. Farlow JL, Lin H, Sauerbeck L, et al. Lessons learned from whole exome sequencing in multiplex families affected by a complex genetic disorder, intracranial aneurysm. *PLoS One*. 2015;10(3). doi:10.1371/journal.pone.0121104
101. Yan J, Hitomi T, Takenaka K, et al. Genetic study of intracranial aneurysms. *Stroke*. 2015. doi:10.1161/STROKEAHA.114.007286
102. Zhou S, Ambalavanan A, Rochefort D, et al. RNF213 Is Associated with Intracranial Aneurysms in the French-Canadian Population. *Am J Hum Genet*. 2016. doi:10.1016/j.ajhg.2016.09.001
103. Liu W, Morito D, Takashima S, et al. Identification of RNF213 as a susceptibility gene for moyamoya disease and its possible role in vascular development. *PLoS One*. 2011. doi:10.1371/journal.pone.0022542
104. Yang X, Li J, Fang Y, et al. Rho Guanine Nucleotide Exchange Factor ARHGEF17 Is a Risk Gene for Intracranial Aneurysms. *Circ Precis Med*. 2018;11(7):e002099. doi:10.1161/CIRCGEN.117.002099 [doi]
105. Santiago-Sim T, Fang X, Hennessy ML, et al. THSD1 (Thrombospondin Type 1 Domain Containing Protein 1) Mutation in the Pathogenesis of Intracranial Aneurysm and Subarachnoid Hemorrhage. *Stroke*. 2016;47(12):3005-3013.

doi:STROKEAHA.116.014161 [pii]
106. Wu YQ, Li Z, Shi Y, et al. Exome Sequencing Identifies LOXL2 Mutation as a Cause of Familial Intracranial Aneurysm. *World Neurosurg*. 2018. doi:10.1016/j.wneu.2017.10.094
107. Sathyan S, Koshy L, Lekshmi KRS, et al. Lack of association of Lysyl oxidase (LOX) gene polymorphisms with intracranial aneurysm in a south Indian population. *Mol Biol Rep*. 2013. doi:10.1007/s11033-013-2693-1
108. Akagawa H, Narita A, Yamada H, et al. Systematic screening of lysyl oxidase-like (LOXL) family genes demonstrates that LOXL2 is a susceptibility gene to intracranial aneurysms. *Hum Genet*. 2007. doi:10.1007/s00439-007-0333-3
109. Hong EP, Jeon JP, Kim SE, et al. A novel association between lysyl oxidase gene polymorphism and intracranial aneurysm in koreans. *Yonsei Med J*. 2017. doi:10.3349/ymj.2017.58.5.1006
110. Bourcier R, Chatel S, Bourcereau E, et al. Understanding the pathophysiology of intracranial aneurysm: The ICAN project. *Clin Neurosurg*. 2017. doi:10.1093/neuros/nyw135
111. Bourcier R, Le Scouarnec S, Bonnaud S, et al. Rare Coding Variants in ANGPTL6 Are Associated with Familial Forms of Intracranial Aneurysm. *Am J Hum Genet*. 2018. doi:10.1016/j.ajhg.2017.12.006
112. Oike Y, Ito Y, Maekawa H, et al. Angiopoietin-related growth factor (AGF) promotes angiogenesis. *Blood*. 2004. doi:10.1182/blood-2003-04-1272
113. Lehner B. Genotype to phenotype: Lessons from model organisms for human genetics. *Nat Rev Genet*. 2013. doi:10.1038/nrg3404
114. Meyers JR. Zebrafish: Development of a Vertebrate Model Organism. *Curr Protoc Essent Lab Tech*. 2018. doi:10.1002/cpet.19
115. Laale HW. The biology and use of zebrafish, Brachydanio rerio in fisheries research. A literature review. *J Fish Biol*. 1977. doi:10.1111/j.1095-8649.1977.tb04049.x
116. Lieschke GJ, Currie PD. Animal models of human disease: Zebrafish swim into view. *Nat Rev Genet*. 2007. doi:10.1038/nrg2091
117. Walcott BP, Peterson RT. Zebrafish models of cerebrovascular disease. *J Cereb Blood Flow Metab*. 2014;34(4):571-577. doi:10.1038/jcbfm.2014.27 [doi]
118. Parichy DM, Elizondo MR, Mills MG, Gordon TN, Engeszer RE. Normal table of postembryonic zebrafish development: Staging by externally visible anatomy of the living fish. *Dev Dyn*. 2009. doi:10.1002/dvdy.22113
119. Spence R, Gerlach G, Lawrence C, Smith C. The behaviour and ecology of the zebrafish, Danio rerio. *Biol Rev*. 2008. doi:10.1111/j.1469-185X.2007.00030.x
120. Holtzman NG, Kathryn Iovine M, Liang JO, Morris J. Learning to fish with genetics: A primer on the vertebrate model Danio rerio. *Genetics*. 2016. doi:10.1534/genetics.116.190843
121. Ungos J, Weinstein BM. Vascular Development in the Zebrafish. *Adv Dev Biol*. 2007. doi:10.1016/S1574-3349(07)18012-1
122. Isogai S, Horiguchi M, Weinstein BM. The vascular anatomy of the developing zebrafish: An atlas of embryonic and early larval development. *Dev Biol*. 2001. doi:10.1006/dbio.2000.9995
123. Bournele D, Beis D. Zebrafish models of cardiovascular disease. *Heart Fail Rev*. 2016. doi:10.1007/s10741-016-9579-y
124. K. B, K.S. W, G. Y. Defective “pacemaker” current (I(h)) in a zebrafish mutant with a

slow heart rate. *Proc Natl Acad Sci U S A*. 1997. doi:http://dx.doi.org/10.1073/pnas.94.9.4554
125. Ho D, Zhao X, Gao S, Hong C, Vatner DE, Vatner SF. Heart Rate and Electrocardiography Monitoring in Mice. In: *Current Protocols in Mouse Biology*. ; 2011. doi:10.1002/9780470942390.mo100159
126. Pelster B, Burggren WW. Disruption of hemoglobin oxygen transport does not impact oxygen- dependent physiological processes in developing embryos of zebra fish (Danio rerio). *Circ Res*. 1996. doi:10.1161/01.RES.79.2.358
127. Nguyen CT, Lu Q, Wang Y, Chen JN. Zebrafish as a model for cardiovascular development and disease. *Drug Discov Today Dis Model*. 2008. doi:10.1016/j.ddmod.2009.02.003
128. Howe K, Clark MD, Torroja CF, et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*. 2013;496(7446):498-503. doi:10.1038/nature12111 [doi]
129. Timme-Laragy AR, Karchner SI, Hahn ME. Gene knockdown by morpholino-modified oligonucleotides in the zebrafish (danio rerio) model: Applications for developmental toxicology. *Methods Mol Biol*. 2012. doi:10.1007/978-1-61779-867-2_5
130. Taylor JS, Van de Peer Y, Braasch I, Meyer A. Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc B Biol Sci*. 2001. doi:10.1098/rstb.2001.0975
131. Zhang X, Zhang Y, Zheng X, et al. A Consensus Linkage Map Provides Insights on Genome Character and Evolution in Common Carp (Cyprinus carpio L.). *Mar Biotechnol*. 2013. doi:10.1007/s10126-012-9485-9
132. Glasauer SMK, Neuhauss SCF. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics*. 2014. doi:10.1007/s00438-014-0889-2
133. Postlethwait J, Amores A, Cresko W, Singer A, Yan YL. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet*. 2004. doi:10.1016/j.tig.2004.08.001
134. Woods IG, Wilson C, Friedlander B, et al. The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res*. 2005. doi:10.1101/gr.4134305
135. Varshney GK, Sood R, Burgess SM. Understanding and Editing the Zebrafish Genome. *Adv Genet*. 2015. doi:10.1016/bs.adgen.2015.09.002
136. Nasevicius A, Ekker SC. Effective targeted gene "knockdown" in zebrafish. *Nat Genet*. 2000. doi:10.1038/79951
137. Summerton J. Uncharged nucleic acid analogs for therapeutic and diagnostic applications: oligomers assembled from ribose-derived subunits. In: Brackel C, ed. *Discoveries in Antisense Nucleic Acids*. The Portfolio Publishing Co; 1989:71-80.
138. Summerton J, Weller D. Morpholino Antisense Oligomers: Design, Preparation, and Properties. *Antisense Nucleic Acid Drug Dev*. 2011. doi:10.1089/oli.1.1997.7.187
139. Bill BR, Petzold AM, Clark KJ, Schimmenti LA, Ekker SC. A Primer for Morpholino Use in Zebrafish. *Zebrafish*. 2009. doi:10.1089/zeb.2008.0555
140. Kok FO, Shin M, Ni CW, et al. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev Cell*. 2015. doi:10.1016/j.devcel.2014.11.018
141. Rossi A, Kontarakis Z, Gerri C, et al. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature*. 2015. doi:10.1038/nature14580

142. Stainier DYR, Raz E, Lawson ND, et al. Guidelines for morpholino use in zebrafish. *PLoS Genet*. 2017. doi:10.1371/journal.pgen.1007000
143. Blum M, De Robertis EM, Wallingford JB, Niehrs C. Morpholinos: Antisense and Sensibility. *Dev Cell*. 2015. doi:10.1016/j.devcel.2015.09.017
144. Eisen JS, Smith JC. Controlling morpholino experiments: don't stop making antisense. *Development*. 2008. doi:10.1242/dev.001115
145. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. Stages of embryonic development of the zebrafish. *Dev Dyn*. 1995. doi:10.1002/aja.1002030302
146. Ekker SC, Larson JD. Morphant technology in model developmental systems. *Genesis*. 2001. doi:10.1002/gene.1038
147. Robu ME, Larson JD, Nasevicius A, et al. p53 activation by knockdown technologies. *PLoS Genet*. 2007. doi:10.1371/journal.pgen.0030078
148. French C, Seshadri S, Destefano A, et al. Mutation of FOXC1 and PITX2 induces cerebral small-vessel disease. *J Clin Invest*. 2014. doi:10.1172/JCI75109DS1
149. Wardlaw JM, Smith C, Dichgans M. Mechanisms of sporadic cerebral small vessel disease: Insights from neuroimaging. *Lancet Neurol*. 2013. doi:10.1016/S1474-4422(13)70060-7
150. Kiando SR, Tucker NR, Castro-Vega LJ, et al. PHACTR1 Is a Genetic Susceptibility Locus for Fibromuscular Dysplasia Supporting Its Complex Genetic Pattern of Inheritance. *PLoS Genet*. 2016;12(10):e1006367. doi:10.1371/journal.pgen.1006367 [doi]
151. Aboualaiwi WA, Muntean BS, Ratnam S, et al. Survivin-induced abnormal ploidy contributes to cystic kidney and aneurysm formation. *Circulation*. 2014;129(6):660-672. doi:10.1161/CIRCULATIONAHA.113.005746 [doi]
152. Haasdijk RA, Den Dekker WK, Cheng C, et al. THSD1 preserves vascular integrity and protects against intraplaque haemorrhaging in ApoE -/- mice. *Cardiovasc Res*. 2016. doi:10.1093/cvr/cvw015
153. Kuang SQ, Medina-Martinez O, Guo DC, et al. FOXE3 mutations predispose to thoracic aortic aneurysms and dissections. *J Clin Invest*. 2016;126(3):948-961. doi:10.1172/JCI83778 [doi]
154. Guo DC, Gong L, Regalado ES, et al. MAT2A mutations predispose individuals to thoracic aortic aneurysms. *Am J Hum Genet*. 2015;96(1):170-177. doi:10.1016/j.ajhg.2014.11.015 [doi]
155. Statistics Canada DD. *Annual Estimates of Population for Canada, Provinces and Territories, from July 1, 1971 to July 1, 2018*.; 2018.
156. Neary P. The peopling of Newfoundland. Essays in historical geography. *J Hist Geogr*. 2004. doi:10.1016/0305-7488(80)90066-3
157. Rahman P, Jones A, Curtis J, et al. The Newfoundland population: a unique resource for genetic investigation of complex diseases. *Hum Mol Genet*. 2003;(suppl 2):R167-72. doi:10.1093/hmg/ddg257
158. Bear JC, Nemec TF, Kennedy JC, et al. Persistent genetic isolation in outport Newfoundland. *Am J Med Genet*. 1987. doi:10.1002/ajmg.1320270410
159. Zhai G, Zhou J, Woods MO, et al. Genetic structure of the Newfoundland and Labrador population: Founder effects modulate variability. *Eur J Hum Genet*. 2016. doi:10.1038/ejhg.2015.256
160. Kosseim P, Pullman D, Perrot-Daley A, Hodgkinson K, Street C, Rahman P. Privacy protection and public goods: Building a genetic database for health research in

Newfoundland and Labrador. *J Am Med Informatics Assoc*. 2013. doi:10.1136/amiajnl-2012-001009
161. Moore SJ, Green JS, Fan Y, et al. Clinical and genetic epidemiology of Bardet-Biedl syndrome in Newfoundland: A 22-year prospective, population-based, cohort study. *Am J Med Genet*. 2005. doi:10.1002/ajmg.a.30406
162. Scully MF, Stoffman J, Boyd S. The unusual pattern of hereditary bleeding disorders in the province of Newfoundland and Labrador—Canada's most Eastern Province. *Transfus Apher Sci*. 2018. doi:10.1016/j.transci.2018.10.010
163. Olufemi SE, Green JS, Manickam P, et al. Common ancestral mutation in the MEN1 gene is likely responsible for the prolactinoma variant of MEN1 (MEN1(Burin)) in four kindreds from Newfoundland. *Hum Mutat*. 1998. doi:10.1002/(SICI)1098-1004(1998)11:4<264::AID-HUMU2>3.0.CO;2-V
164. O'Rielly DD, Rahman P. Genetic, Epigenetic and Pharmacogenetic Aspects of Psoriasis and Psoriatic Arthritis. *Rheum Dis Clin North Am*. 2015. doi:10.1016/j.rdc.2015.07.002
165. Merner ND, Hodgkinson KA, Haywood AFM, et al. Arrhythmogenic Right Ventricular Cardiomyopathy Type 5 Is a Fully Penetrant, Lethal Arrhythmic Disorder Caused by a Missense Mutation in the TMEM43 Gene. *Am J Hum Genet*. 2008. doi:10.1016/j.ajhg.2008.01.010
166. Woods MO, Younghusband HB, Parfrey PS, et al. The genetic basis of colorectal cancer in a population-based incident cohort with a high rate of familial disease. *Gut*. 2010. doi:10.1136/gut.2010.208462
167. Kundra S, Mahendru V, Gupta V, Choudhary AK. Principles of neuroanesthesia in aneurysmal subarachnoid hemorrhage. *J Anaesthesiol Clin Pharmacol*. 2014. doi:10.4103/0970-9185.137261
168. Powell A. Whole Exome Sequencing Analysis of Intracranial Aneurysm In Multiplex Families From Newfoundland and Labrador. 2016.
169. TCAG DNA Sequencing/Synthesis. The Centre for Applied Genomics.
170. An Introduction to Next-Generation Sequencing Technology. illumina.
171. System Specification Sheet: Sequencing. HiSeq® 2500 Sequencing System. illumina.
172. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008. doi:10.1038/nature07517
173. H L, R D. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*. 2009.
174. Ebbert MTW, Wadsworth ME, Staley LA, et al. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC Bioinformatics*. 2016. doi:10.1186/s12859-016-1097-3
175. Broadinstitute. Picard.
176. Depristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011. doi:10.1038/ng.806
177. Van der Auwera GA, Carneiro MO, Hartl C, et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr Protoc Bioinforma*. 2013. doi:10.1002/0471250953.bi1110s43
178. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010. doi:10.1101/gr.107524.110

179. Guo Y, Li J, Li CI, Long J, Samuels DC, Shyr Y. The effect of strand bias in Illumina short-read sequencing data. *BMC Genomics*. 2012. doi:10.1186/1471-2164-13-666
180. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010. doi:10.1093/nar/gkq603
181. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat Rev Genet*. 2002. doi:10.1038/nrg775
182. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet*. 2011. doi:10.1038/nrg3051
183. Perenthaler E, Yousefi S, Niggl E, Barakat TS. Beyond the Exome: The Non-coding Genome and Enhancers in Neurodevelopmental Disorders and Malformations of Cortical Development. *Front Cell Neurosci*. 2019. doi:10.3389/fncel.2019.00352
184. Yoon NK, McNally S, Taussky P, et al. Imaging of cerebral aneurysms: a clinical perspective. *Neurovascular Imaging*. 2016;2(1):6. doi:10.1186/s40809-016-0016-3
185. Perry JJ, Stiell IG, Sivilotti MLA, et al. Sensitivity of computed tomography performed within six hours of onset of headache for diagnosis of subarachnoid haemorrhage: Prospective cohort study. *BMJ*. 2011. doi:10.1136/bmj.d4277
186. Kim DH, Van Ginhoven G, Milewicz DM. Familial aggregation of both aortic and cerebral aneurysms: Evidence for a common genetic basis in a subset of families. *Neurosurgery*. 2005. doi:10.1227/01.NEU.0000156787.55281.53
187. Ruigrok YM, Elias R, Wijmenga C, Rinkel GJE. A comparison of genetic chromosomal loci for intracranial, thoracic aortic, and abdominal aortic aneurysms in search of common genetic risk factors. *Cardiovasc Pathol*. 2008. doi:10.1016/j.carpath.2007.06.001
188. Worrall BB, Foroud T, Brown RD, et al. Genome screen to detect linkage to common susceptibility genes for intracranial and aortic aneurysms. *Stroke*. 2009. doi:10.1161/STROKEAHA.108.522631
189. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010. doi:10.1038/nmeth0810-575
190. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010. doi:10.1038/nmeth0410-248
191. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812-3814. doi:10.1093/nar/gkg509
192. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. 2010. doi:10.1101/gr.097857.109
193. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014. doi:10.1038/ng.2892
194. Stelzer G, Plaschkes I, Oz-Levi D, et al. VarElect: The phenotype-based variation prioritizer of the GeneCards Suite. *BMC Genomics*. 2016. doi:10.1186/s12864-016-2722-2
195. Bult CJ, Blake JA, Smith CL, et al. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*. 2019. doi:10.1093/nar/gky1056
196. Ruzicka L, Howe DG, Ramachandran S, et al. The Zebrafish Information Network: New support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources. *Nucleic Acids Res*. 2019. doi:10.1093/nar/gky1090
197. Han T, Yang C, Chang K, Zhang D, Imam FB, Rana TM. Identification of novel genes and networks governing hematopoietic stem cell development. *EMBO Rep*. 2016.

doi:10.15252/embr.201642395
198. Ebermann I, Phillips JB, Liebau MC, et al. PDZD7 is a modifier of retinal disease and a contributor to digenic Usher syndrome. *J Clin Invest*. 2010. doi:10.1172/JCI39715
199. Huitema LFA, Renn J, Logister I, et al. Macrophage-stimulating protein and calcium homeostasis in zebrafish. *FASEB J*. 2012. doi:10.1096/fj.11-202663
200. Karlsson J, Von Hofsten J, Olsson PE. Generating transparent zebrafish: A refined method to improve detection of gene expression during embryonic development. *Mar Biotechnol*. 2001. doi:10.1007/s1012601-0053-4
201. Bedell VM, Westcot SE, Ekker SC. Lessons from morpholino-based screening in zebrafish. *Brief Funct Genomics*. 2011. doi:10.1093/bfgp/elr021
202. Aranda PS, Lajoie DM, Jorcyk CL. Bleach gel: A simple agarose gel for analyzing RNA quality. *Electrophoresis*. 2012. doi:10.1002/elps.201100335
203. Social Science Statistics. https://www.socscistatistics.com/.
204. Wheeler GN, Parker AE, Thomas CL, et al. Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules. *Proc Natl Acad Sci U S A*. 1991. doi:10.1073/pnas.88.11.4796
205. Amagai M, Stanley JR. Desmoglein as a target in skin disease and beyond. *J Invest Dermatol*. 2012. doi:10.1038/jid.2011.390
206. Schlipf NA, Vahlquist A, Teigen N, et al. Whole-exome sequencing identifies novel autosomal recessive DSG1 mutations associated with mild SAM syndrome. *Br J Dermatol*. 2016. doi:10.1111/bjd.14079
207. Dibble CC, Elis W, Menon S, et al. TBC1D7 Is a Third Subunit of the TSC1-TSC2 Complex Upstream of mTORC1. *Mol Cell*. 2012. doi:10.1016/j.molcel.2012.06.009
208. Capo-Chichi JM, Tcherkezian J, Hamdan FF, et al. Disruption of TBC1D7, a subunit of the TSC1-TSC2 protein complex, in intellectual disability and megalencephaly. *J Med Genet*. 2013. doi:10.1136/jmedgenet-2013-101680
209. Rodova M, Ankilova V, Safro MG. Human phenylalanyl-tRNA synthetase: Cloning, characterization of the deduced amino acid sequences in terms of the structural domains and coordinately regulated expression of the α and β subunits in chronic myeloid leukemia cells. *Biochem Biophys Res Commun*. 1999. doi:10.1006/bbrc.1999.0141
210. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. In: *Current Protocols in Bioinformatics*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2016:1.30.1-1.30.33. doi:10.1002/cpbi.5
211. Rappaport N, Twik M, Plaschkes I, et al. MalaCards: An amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res*. 2017. doi:10.1093/nar/gkw1012
212. Gilad Y, Bustamante CD, Lancet D, Pääbo S. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am J Hum Genet*. 2003. doi:10.1086/378132
213. Verbeurgt C, Wilkin F, Tarabichi M, Gregoire F, Dumont JE, Chatelain P. Profiling of olfactory receptor gene expression in whole human olfactory mucosa. *PLoS One*. 2014. doi:10.1371/journal.pone.0096333
214. Schievella AR, Chen JH, Graham JR, Lin LL. MADD, a novel death domain protein that interacts with the type 1 tumor necrosis factor receptor and activates mitogen-activated protein kinase. *J Biol Chem*. 1997. doi:10.1074/jbc.272.18.12069
215. Mulherkar N, Ramaswamy M, Mordi DC, Prabhakar BS. MADD/DENN splice variant of the IG20 gene is necessary and sufficient for cancer cell survival. *Oncogene*. 2006.

doi:10.1038/sj.onc.1209650
216. Al-Zoubi AM, Efimova E V., Kaithamana S, et al. Contrasting Effects of IG20 and Its Splice Isoforms, MADD and DENN-SV, on Tumor Necrosis Factor α-induced Apoptosis and Activation of Caspase-8 and -3. *J Biol Chem*. 2001. doi:10.1074/jbc.M104835200
217. Farooqi AA, Javeed MK, Javed Z, et al. TRPM channels: Same ballpark, different players, and different rules in immunogenetics. *Immunogenetics*. 2011. doi:10.1007/s00251-011-0570-4
218. Oancea E, Vriens J, Brauchi S, Jun J, Splawski I, Clapham DE. TRPM1 forms ion channels associated with melanin content in melanocytes. *Sci Signal*. 2009. doi:10.1126/scisignal.2000146
219. Guo H, Carlson JA, Slominski A. Role of TRPM in melanocytes and melanoma. *Exp Dermatol*. 2012. doi:10.1111/j.1600-0625.2012.01565.x
220. van Genderen MM, Bijveld MMC, Claassen YB, et al. Mutations in TRPM1 Are a Common Cause of Complete Congenital Stationary Night Blindness. *Am J Hum Genet*. 2009. doi:10.1016/j.ajhg.2009.10.012
221. Nakamura M, Sanuki R, Yasuma TR, et al. TRPM1 mutations are associated with the complete form of congenital stationary night blindness. *Mol Vis*. 2010.
222. Nelson G, Chandrashekar J, Hoon MA, et al. An amino-acid taste receptor. *Nature*. 2002. doi:10.1038/nature726
223. Eilbeck K, Quinlan A, Yandell M. Settling the score: Variant prioritization and Mendelian disease. *Nat Rev Genet*. 2017. doi:10.1038/nrg.2017.52
224. Lewinter MM, Granzier HL. Titin is a major human disease gene. *Circulation*. 2013. doi:10.1161/CIRCULATIONAHA.112.139717
225. Chi RJ, Simon AR, Bienkiewicz EA, Felix A, Keller TCS. Smooth muscle titin Zq domain interaction with the smooth muscle α-actinin central rod. *J Biol Chem*. 2008. doi:10.1074/jbc.M709621200
226. Chi RJ, Olenych SG, Kim K, Keller TCS. Smooth muscle α-actinin interaction with smitin. *Int J Biochem Cell Biol*. 2005. doi:10.1016/j.biocel.2005.02.014
227. Kawai Y, Moriyama A, Asai K, et al. Molecular characterization of histidinemia: Identification of four missense mutations in the histidase gene. *Hum Genet*. 2005. doi:10.1007/s00439-004-1232-5
228. Lam WK, Cleary MA, Wraith JE, Walter JH. Histidinaemia: A benign metabolic disorder. *Arch Dis Child*. 1996. doi:10.1136/adc.74.4.343
229. Laity JH, Lee BM, Wright PE. Zinc finger proteins: New insights into structural and functional diversity. *Curr Opin Struct Biol*. 2001. doi:10.1016/S0959-440X(00)00167-6
230. Małecki J, Ho AYY, Moen A, Dahl HA, Falnes P. Human METTL20 is a mitochondrial lysine methyltransferase that targets the β subunit of electron transfer flavoprotein (ETFβ) and modulates its activity. *J Biol Chem*. 2015. doi:10.1074/jbc.M114.614115
231. Rhein VF, Carroll J, He J, Ding S, Fearnley IM, Walker JE. Human METTL20 methylates lysine residues adjacent to the recognition loop of the electron transfer flavoprotein in mitochondria. *J Biol Chem*. 2014. doi:10.1074/jbc.M114.580464
232. Gaudino G, Follenzi A, Naldini L, et al. RON is a heterodimeric tyrosine kinase receptor activated by the HGF homologue MSP. *EMBO J*. 1994. doi:10.1002/j.1460-2075.1994.tb06659.x
233. Sakamoto O, Iwama A, Amitani R, et al. Role of macrophage-stimulating protein and its receptor, RON tyrosine kinase, in ciliary motility. *J Clin Invest*. 1997.

doi:10.1172/JCI119214
234. Nakayama J, Fu YH, Clark AM, et al. A nonsense mutation of the MASS1 gene in a family with febrile and afebrile seizures. *Ann Neurol*. 2002. doi:10.1002/ana.10347
235. Besnard T, Vaché C, Baux D, et al. Non-USH2A mutations in USH2 patients. *Hum Mutat*. 2012. doi:10.1002/humu.22004
236. Sardina JL, López-Ruano G, Prieto-Bermejo R, et al. PTPN13 regulates cellular signalling and β-catenin function during megakaryocytic differentiation. *Biochim Biophys Acta - Mol Cell Res*. 2014. doi:10.1016/j.bbamcr.2014.08.014
237. Freiss G, Chalbos D. PTPN13/PTPL1: An Important Regulator of Tumor Aggressiveness. *Anticancer Agents Med Chem*. 2012. doi:10.2174/187152011794941262
238. Vermeer PD, Bell M, Lee K, et al. ErbB2, EphrinB1, Src kinase and PTPN13 signaling complex regulates MAP kinase signaling in human cancers. *PLoS One*. 2012. doi:10.1371/journal.pone.0030447
239. van Eekelen M, Runtuwene V, Masselink W, den Hertog J. Pair-wise regulation of convergence and extension cell movements by four phosphatases via RhoA. *PLoS One*. 2012. doi:10.1371/journal.pone.0035913
240. Seeley M, Huang W, Chen Z, Wolff WO, Lin X, Xu X. Depletion of zebrafish titin reduces cardiac contractility by disrupting the assembly of Z-discs and A-bands. *Circ Res*. 2007. doi:10.1161/01.RES.0000255758.69821.b5
241. Shih YH, Dvornikov A V., Zhu P, et al. Exon-and contraction-dependent functions of titin in sarcomere assembly. *Dev*. 2016. doi:10.1242/dev.139246
242. Huttner IG, Wang LW, Santiago CF, et al. A-Band Titin Truncation in Zebrafish Causes Dilated Cardiomyopathy and Hemodynamic Stress Intolerance. *Circ Genomic Precis Med*. 2018. doi:10.1161/CIRCGEN.118.002135
243. Samuel N, Radovanovic I. Genetic basis of intracranial aneurysm formation and rupture: Clinical implications in the postgenomic era. *Neurosurg Focus*. 2019. doi:10.3171/2019.4.FOCUS19204
244. Bartoszewski RA, Jablonsky M, Bartoszewska S, et al. A synonymous single nucleotide polymorphism in ΔF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J Biol Chem*. 2010. doi:10.1074/jbc.M110.154575
245. Gartner JJ, Parker SCJ, Prickett TD, et al. Whole-genome sequencing identifies a recurrent functional synonymous mutation in melanoma. *Proc Natl Acad Sci U S A*. 2013. doi:10.1073/pnas.1304227110
246. Smith M, Flodman PL. Expanded insights into mechanisms of gene expression and disease related disruptions. *Front Mol Biosci*. 2018. doi:10.3389/fmolb.2018.00101
247. Bao R, Huang L, Andrade J, et al. Review of current methods, Applications, And data management for the bioinformatics analysis of whole exome sequencing. *Cancer Inform*. 2014. doi:10.4137/CIN.S13779
248. Dashti MJS, Gamieldien J. A practical guide to filtering and prioritizing genetic variants. *Biotechniques*. 2017. doi:10.2144/000114492
249. The Human Protein Atlas. https://www.proteinatlas.org/. Accessed October 31, 2020.
250. Thul PJ, Akesson L, Wiking M, et al. A subcellular map of the human proteome. *Science (80-)*. 2017. doi:10.1126/science.aal3321
251. Jiao C, Yan P, Xia C, et al. BrainEXP: A database featuring with spatiotemporal expression variations and co-expression organizations in human brains. *Bioinformatics*. 2019. doi:10.1093/bioinformatics/bty576

252. Pires-Luís AS, Vieira-Coimbra M, João Ferreira M, et al. Prognostic significance of MST1R dysregulation in renal cell tumors. *Am J Cancer Res*. 2016.
253. Babicky ML, Harper MM, Chakedis J, et al. MST1R kinase accelerates pancreatic cancer progression via effects on both epithelial cells and macrophages. *Oncogene*. 2019. doi:10.1038/s41388-019-0811-9
254. Dai W, Zheng H, Cheung AKL, et al. Whole-exome sequencing identifies MST1R as a genetic susceptibility gene in nasopharyngeal carcinoma. *Proc Natl Acad Sci U S A*. 2016. doi:10.1073/pnas.1523436113

Appendices

Appendix A – Primer Sequences

Primer	Sequence	Product Size	PCR Protocol (Optimal Annealing Temp)
ADGRV1_1262 Fwd **ADGRV1_1262 Rev**	CTGTTTTTGGTTCATCTTTTGTCC TTTAGTCCCTCCTAGTCTTG	299	Touchdown A
ADGRV1_1276 Fwd **ADGRV1_1276 Rev**	CTGTGTGGATCTTCTGTCTTTCA TGTCCCTCTGCTTACATCC	297	Touchdown A
DSG1_Fwd **DSG1_Rev**	GGTCCTCCTGAAACCAAT TGTCCACTAACAACGGGCTC	598	Touchdown A
FARSA_Fwd **FARSA_Rev**	TTTTAAGTTGCAGGAATGGCTA GAAAACCCCACACTGTGCTT	343	Touchdown A
HAL_Fwd **HAL_Rev**	CCATCCCTCGTCTGTTGACT CCATCCCTCGTCTGTTGACT	298	Touchdown A
MADD_1262 Fwd **MADD_1262 Rev**	TGAAGAGCCTTGGGTGTTTT CCAGACGTGCTTACCAGGTT	449	Standard (58°)
MADD_1265 Fwd **MADD_1265 Rev**	CCTTTGTGCACGGAGTAACA CCGGCTGTCTGACTGTTCTA	299	Standard (60°)
METTL20_Fwd **METTL20_Rev**	CCAAGTCTCGTTTCCCTGT CCCAGGGTCACCAATCAGTAC	527	Touchdown A
MST1R_Fwd **MST1R_Rev**	CCCTGGAGGAGAGTCCATAC CTCAAGGCAGCTAAGCAGGT	338	Standard (58°)
OR2T35_Fwd **OR2T35_Rev**	CCGTGCTGAAGTTGTCTTGC TGATCAGTCACCACCCCTGA	511	Touchdown A
PTPN13_1381 Fwd **PTPN13_1381 Rev**	GCAATTTGGAGTTGCCTTCCA CCATGTCTGAATCTCCACGGT	832	Touchdown A
PTPN13_1385 Fwd **PTPN13_1385 Rev**	GTCCATTCGCCAAGTCATTT CAGCCATGACTCACCTCTGT	240	Standard (58°)

TAS1R1_Fwd **TAS1R1_Rwd**	CAGTATCTTGCAGGCCCCTA TTGGAATGAGCGAACTGTCA	298	Touchdown A
TBC1D7_Fwd **TBC1D7_Rev**	GAGCCCTTTCATGTCCTCAG TTCTAACAACCCTTTTGAAATGC	499	Standard (58°)
TRPM1_1256 Fwd **TRPM1_1256 Rev**	CCATGGTACCAGGCATCTTT GGGGACCACAGTGAGTTCTG	497	Touchdown A
TRPM1_1381 Fwd **TRPM1_1381 Rev**	TGGCAGTCACCTTGCAGTAG TCGTTTCATCGGGGAAATAG	275	Standard (62.5°)
TTN_Fwd **TTN_Rev**	GGTGAACTTCTCATTTCATCTCC GATTTCTTGGGTTCTGGAGGA	480	Touchdown A
ZNF835_Fwd **ZNF835_Rev**	TCCCCGACGACAGCAGTT CACTCGTGGCAGGCGTAG	293	Touchdown A

Appendix B – Thermal Cycler Protocols

Step	Temperature (°C)	Duration
Standard PCR		
1	95	5 minutes
2a	95	30 seconds
2b	Annealing Temp	30 seconds
2c	72	1 minutes
3	-	Return to step 2 (30 cycles)
4	72	7 minutes
5	4	Hold
Touchdown A		
1	95	2 minutes
2a	95	20 seconds
2b	67*	45 seconds
2c	72	45 seconds
3	-	Return to step 2 (14 cycles)
4a	95	20 seconds
4b	60	45 seconds
4c	72	45 seconds
5	-	Return to step 4 (19 cycles)
6	72	5 minutes
7	4	Hold

ExoSAP		
1	37	30 minutes
2	80	15 minutes
3	4	Hold
ABI Sequencing		
1	96	1 minute
2a	96	10 seconds
2b	50	5 seconds
2c	60	4 minutes
3	-	Return to step 2 (24 cycles)
4	4	Hold
Denature		
1	95	2 minutes
2	4	Hold

* After the initial cycle the annealing temperature is decreased by 0.5°C increments for subsequent cycles

Appendix C – 20X Embryo Media Recipe

- 17.5g NaCl
- 0.75g KCl
- 2.9g $CaCl\text{-}2H_2O$
- Add MiliQ water to 800ml
- 0.41g KH_2PO_4
- 0.142g Na_2HPO_4 anhydrous
- 4.9g $MgSO_4\text{-}7H_2O$
- Add MiliQ water to 1 liter
- Vacuum filter
- Store at 4°C

Appendix D – Ethics Approval Documentation

Monday, July 6, 2020 at 7:41:20 PM Newfoundland Daylight Time

Subject: HREB - Approval of Ethics Renewal 500066
Date: Monday, June 29, 2020 at 7:23:03 PM Newfoundland Daylight Time
From: administrator@hrea.ca
To: Fernandez, Bridget
CC: Negrijn, Carol, Maroun Falah B.(Co-Principal Investigator), Hreaadministrator

Researcher Portal File #: 20162579

Dear Dr. Bridget Fernandez:

This e-mail serves as notification that your ethics renewal for study HREB # 2004.089 – The Clinical and Genetic Epidemiology of Intracranial Aneurysms in Newfoundland – has been **approved**. Please log in to the Researcher Portal to view the approved event.

Ethics approval for this project has been granted for a period of twelve months effective from June 18, 2020 to June 18, 2021.

Please note, it is the responsibility of the Principal Investigator (PI) to ensure that the Ethics Renewal form is submitted prior to the renewal date each year. Though the Research Ethics Office makes every effort to remind the PI of this responsibility, the PI may not receive a reminder. The Ethics Renewal form can be found on the Researcher Portal as an "Event".

The ethics renewal [will be reported] to the Health Research Ethics Board at their meeting dated June 30, 2020.

Thank you,

Research Ethics Office

(e) info@hrea.ca
(t) 709-777-6974
(f) 709-777-8776
(w) www.hrea.ca
Office Hours: 8:30 a.m. – 4:30 p.m. (NL TIME) Monday-Friday