ORIGINAL ARTICLE

Copyright © 2019 The Canadian Journal of Neurological Sciences Inc.

Familial Intracranial Aneurysm in Newfoundland: Clinical and Genetic Analysis

Amy E. Powell, Bridget A. Fernandez, Falah Maroun, Barbara Noble, Michael O. Woods

ABSTRACT: ***Objective:*** Intracranial aneurysm (IA) is an expansion of the weakened arterial wall that is often asymptomatic until rupture, resulting in subarachnoid hemorrhage. Here we describe the high prevalence of familial IA in a cohort of Newfoundland ancestry. We began to investigate the genetic etiology of IA in affected family members, as the inheritance of this disease is poorly understood. ***Methods:*** Whole exome sequencing was completed for a cohort of 12 affected individuals from two multiplex families with a strong family history of IA. A filtering strategy was implemented to identify rare, shared variants. Filtered variants were prioritized based on validation by Sanger sequencing and segregation within the families. ***Results:*** In family R1352, six variants passed filtering; while in family R1256, 68 variants remained, so further filtering was pursued. Following validation by Sanger sequencing, top candidates were investigated in a set of population controls, namely, *C4orf6* c.A1G (p.M1V) and *SPDYE4* c.C103T (p.P35S). Neither was detected in 100 Newfoundland control samples. ***Conclusion:*** Rare and potentially deleterious variants were identified in both families, though incomplete segregation was identified for all filtered variants. Alternate methods of variant prioritization and broader considerations regarding the interplay of genetic and environmental factors are necessary in future studies of this disease.

RÉSUMÉ: Prévalence d’anévrismes intracrâniens au sein de familles terre-neuviennes : une analyse clinique et génétique. ***Objectif*** **:** Un anévrisme intracrânien (AI) consiste en une expansion, souvent asymptomatique, d’une paroi artérielle affaiblie. La rupture qui peut s’ensuivre résultera en une hémorragie sous-arachnoïdienne. Nous voulons décrire ici la forte prévalence d’AI au sein de familles terre-neuviennes ayant des ancêtres communs. Nous avons débuté cette étude en étudiant l’étiologie génétique de l’AI chez les membres de ces familles affectés par cette déformation car l’hérédité des AI demeure encore mal comprise. ***Méthodes*** **:** Nous avons tout d’abord procédé au séquençage entier de l’exome d’un groupe de 12 sujets appartenant à deux familles dites « multiplexes » présentant des antécédents notables d’AI. À cet égard, une stratégie de filtrage a été mise de l’avant afin d’identifier des variantes génétiques à la fois peu fréquentes et partagées. Nous avons ensuite privilégié et validé ces variantes filtrées en nous fondant sur la méthode de séquençage et de ségrégation de Sanger. ***Résultats*** **:** Dans la famille R1352, 6 variantes ont été sélectionnées par filtrage alors que 68 variantes l’ont été dans le cas de la famille R1256, ce qui fait que des tâches additionnelles de filtrage ont été menées. Une fois complétée notre validation par la méthode de Sanger, les meilleurs sujets ont fait l’objet d’un travail d’analyse au sein d’un groupe de témoins de la population, à savoir *C4orf6* c.A1G (p. M1V) et *SPDYE4* c.C103T (p. P35S). À cet égard, aucune variante génétique n’a été détectée parmi 100 échantillons de témoins de Terre-Neuve. ***Conclusion*** **:** Bien qu’une ségrégation incomplète ait été effectuée pour toutes les variantes filtrées, des variantes génétiques peu fréquentes et potentiellement délétères ont été détectées au sein de ces deux familles. D’autres méthodes de priorisation des variantes génétiques, de même que des considérations d’ordre plus général en ce qui a trait à l’interaction entre les facteurs génétiques et les facteurs environnementaux, sont nécessaires si l’on veut étudier les AI dans le futur.

Keywords: Genetics – human, Cerebral aneurysms, Genetics – molecular biology

doi:10.1017/cjn.2019.230 Can J Neurol Sci. 2019; 46: 518–526

From the Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John’s, Newfoundland and Labrador, Canada (AEP, BAF, BN, MOW); Discipline of Surgery, Faculty of Medicine, Memorial University of Newfoundland, St. John’s, Newfoundland and Labrador, Canada (FM)

Received November 27, 2018. Final Revisions Submitted May 3, 2019. Date of Acceptance June 5, 2019.

Correspondence to: Michael O. Woods, Craig L. Dobbin Genetics Research Centre, Memorial University of Newfoundland, 300 Prince Philip Drive, St. John’s, NL, A1B 3V6, Canada. Email: mwoods@mun.ca

https://doi.org/10.1017/cjn.2019.230 Published online by Cambridge University Press

Introduction

Though intracranial aneurysms (IAs) can occur sporadically, they are more common in individuals that have a family history of the disease, and on average, familial IAs develop at an earlier age.[1,2] To date, several common variants have been statistically associated with IA through genome-wide association, candidate gene, and linkage studies.[3] As the next generation sequencing technology has become more accessible, research groups have been able to search for rare variants in exomes and explore the possible modes of inheritance for this disease in families with non-syndromic IA.[4,5] However, no causative variants have been identified, and our knowledge of the genetic etiology of familial IA is still limited.

There appears to be a high prevalence of familial IA in the province of Newfoundland and Labrador (NL), Canada. Over a 5-year recruitment period, we ascertained 53 IA families containing 99 affected individuals. The unique population of NL has provided geneticists with a wealth of opportunity to study familial disease. The current population has grown from a group of initial immigrants of mainly Irish and English descent who settled in isolated communities around the island in the mid-1700 s.[6] This settlement has led to the relatively high prevalence of several monogenic disorders, including Bardet–Biedl syndrome, Lynch syndrome, and arrhythmogenic right ventricular cardiomyopathy.[7–9] The presence of familial IA in the population was first referenced in a case study of familial central nervous system disorders, which described three families with seven individuals that had been diagnosed with subarachnoid hemorrhage (SAH).[10] Previous studies have also reported an increased prevalence of familial IA in similar genetic isolates, such as the French-Canadian population of Québec and the Finnish population.[11–13]

In order to elucidate the genetic risk factors involved with familial IA development in this province, we investigated the presence of shared, rare variants in affected family members from two large kindreds, through the use of whole exome sequencing (WES). It was predicted that one or more strongly penetrant variants would be present in each family. Pathogenic variants identified in this cohort could lead to the identification of causative genes and affected pathways in other populations.

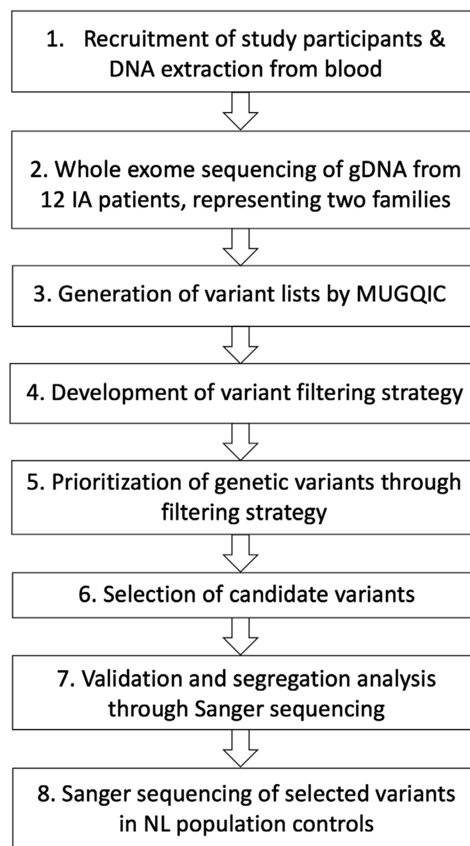


Figure 1: *Study design for genetic investigations. Abbreviations: genomic DNA (gDNA), Newfoundland and Labrador (NL), McGill University and Genome Quebec Innovation Centre (MUGQIC).*

Methods

Participant Recruitment and Phenotyping

The Health Research Ethics Board (Reference Number 04.89) for Memorial University of Newfoundland approved this study, and all participants provided informed consent. The overall study design for this pilot study is shown in Figure 1. Our cohort was assembled through collaboration between Memorial University's Discipline of Genetics and the Department of Surgery (Division of Neurosurgery) at Eastern Health in St. John's, NL, Canada, which is the only neurosurgical unit in the province. Patients were referred to our study if they had presented with either ruptured or unruptured IA. Diagnosis of IA was confirmed through computed tomography (CT) imaging of the Circle of Willis or by magnetic resonance (MR) angiography. Patients were asked whether they had a family history of IA, and additional family members were then contacted for participation. At the date of publication, 154 affected individuals and 415 unaffected family members had been consented to this study. Our database also includes 105 individuals with unknown affection status. Relatives of patients that provided consent were offered CT angiography, underwent a comprehensive medical and family history assessment, and a blood sample was drawn if possible. Genomic DNA was extracted from whole blood using either a simple salting out method or the Wizard® Genomic DNA Purification kit (Promega, Madison, WI, USA). For some participants, sufficient clinical data were obtained (age of diagnosis, rupture status, etc.), but molecular investigations could not be pursued. These circumstances included individuals who were deceased prior to participant recruitment, who were consented by proxy by a relative, so that we could use their clinical data as part of our larger study. As well, participants living outside the province that did not arrange to have blood provided were not included in the genetic portion of our study.

For all affected participants, histories encompassed as much detailed information as possible such as age at diagnosis, number of IAs, dimensions and location of any IAs, rupture incidence as well as any courses of treatment provided. Information regarding two known risk factors, cigarette smoking and hypertension, was also recorded. Affected participants were ultimately placed into one of the three categories: familial (≥2 first-degree relatives with IA), sporadic (no known family history), or equivocal (not enough data to categorize). Our cohort includes over 50 families with possible shared genetic risk factors. Eight families contained four or more affected individuals.

https://doi.org/10.1017/cjn.2019.230 Published online by Cambridge University Press

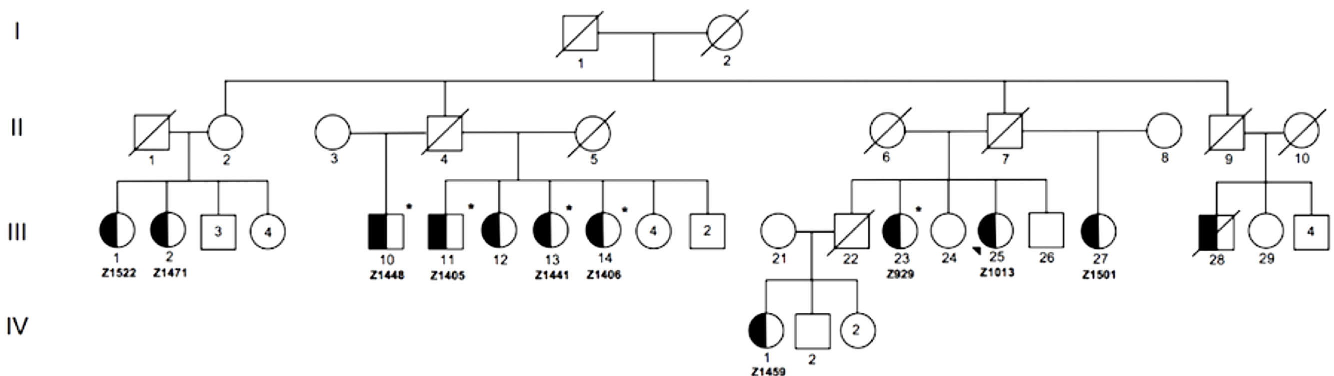
Table 1: Phenotype characteristics for affected study participants from families R1256 and R1352

Family	Patient ID	Pedigree ID	Sex	Age at Dx	# of IAs	Size of IA(s)	Site of IA(s)	Treatment given	IA rupture	Known risk factors
R1256	Z929	III-23	F	62	1	4.5 mm	Right MCA	Observation only	N	Hypertension
	Z1013	III-25	F	29	2	Unknown; 4.5 × 5.5 mm	Right MCA × 2	Clipping	Y	Unknown
	Z1390	III-12	F	54	1	1.5 × 2.5 mm	Left A1–A2 junction	Observation only	N	Hypertension, smoker
	Z1405	III-11	F	51	4	13–14 mm; 4–5 mm; 2 mm; 2 mm	Paraophthalmic; left paraclinoid; right ICA × 2	Coiling	N	Ex-smoker
	Z1406	III-13	F	47	1	5 × 10 mm	Left ACOMM	Clipping	Y	Hypertension
	Z1441	III-14	M	48	1	3.6 × 3.2 mm	Left ICA	Observation only	N	Hypertension, smoker
	Z1448	III-10	M	59	1	1.5 mm	Right MCA	Observation only	N	Hypertension, ex-smoker
	Z1459	IV-1	F	47	2	7 × 8 mm; 9 mm	Left ACOMM; right PICA	Coiling; clipping	N	Hypertension, ex-smoker
	Z1471	III-1	F	50	3	1.9 × 1.3 mm; 2.4 × 1.7 mm; 1.3 × 0.8 mm	PCOMM; bilateral distal ICA; unknown	Observation only	N	Smoker
	Z1501	III-27	F	61	1	3 × 4 mm	ACOMM	Observation only	N	Hypertension, ex-smoker
	Z1522	III-2	F	44	1	10 mm	Right ICA	Coiling	N	Smoker
R1352	Z1039	IV-11	F	43	2	10 × 7 mm; 3 mm	Left ICA; right distal ICA	Clipping	N	Hypertension
	Z1040	IV-15	F	50	2	Unknown	Left ICA x 2	Observation only	N	Unknown
	Z1495	III-11	F	60	1	Moderate	R MCA	Clipping	Y	Smoker
	Z1496	III-10	M	79	2	Small; small	L MCA; R MCA	Diagnosed in autopsy (no treatment given)	N	Smoker, heavy alcohol intake
	Z1497	IV-16	F	50	1	Small	Paraclinoid	Observation only	N	Hypertension
	Z1507	IV-19	M	44	1	2.8 × 4.8 mm	ACOMM	Observation only	N	Smoker
	Z1508	IV-17	F	47	2	Small; 4 mm	Left MCA; right ICA	Coiling	N	Smoker
	Z1533	III-19	F	65	1	9.4 × 5.6 mm	Right MCA	Clipping	N	Smoker
	Z1651	IV-1	F	56	1	1.5-2 mm	Left paraophthalmic	Observation only	N	Unknown

F = female; M = male; ICA = internal carotid artery; ACA = anterior carotid artery; MCA = middle cerebral artery; PCOMM = posterior communicating artery; PCA = posterior cerebral artery; ACOMM = anterior communicating artery; Dx, diagnosis; Y = yes; N = no; mm = millimeters.

https://doi.org/10.1017/cjn.2019.230 Published online by Cambridge University Press

Family 1256



Family 1352

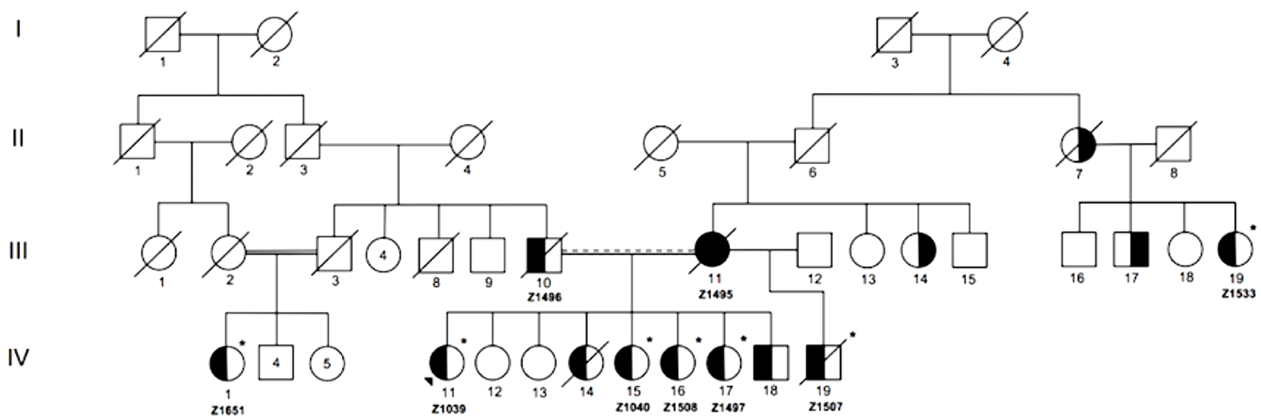


Figure 2: *Condensed pedigrees for R1256 and R1352. Study numbers starting with "Z" are located under each affected individual who consented to participate. III-1 and II-5 of R1352 represent the same individual. Dotted marriage line indicates possible consanguinity. Symbols: squares (males), circles (females), left side shaded (IA diagnosis), right side shaded (AAA diagnosis), unshaded (unaffected or phenotype unknown), asterisk (exome sequenced), diagonal line (known to be deceased), shaded triangle (proband).*

Selection of Study Families

For this pilot study, we chose to focus on two of the families that have a particularly strong family history of IA, specifically eight or more first-degree relatives with IA. Each family has more than 11 affected family members in total, either living or deceased. Eligible affected participants for our pilot study must have had (1) a diagnosis of unruptured IA through CT or MR imaging, or a history of ruptured IA or SAH and (2) at least one affected relative (first or second degree) with IA. Exclusion criteria included (1) individuals that did not provide informed consent, (2) individuals that did not provide a blood sample for genetic analysis, and (3) individuals under the age of 18. Phenotype characteristics for the 20 affected individuals that met these criteria are provided in Table 1.

Figure 2 illustrates condensed pedigrees for these two multiplex families. A younger generation beyond the four shown has been documented but was eliminated from our molecular research, as children were not recruited into the study. For the purpose of this study, we acted under the assumption that familial IA reflected an unknown monogenic inheritance pattern, with the knowledge that additional risk factors play a role in disease development and rupture. Likewise, due to IA's variable expressivity and penetrance, unaffected family members were also not included in initial molecular analyses. They were however used to test the segregation of candidate variants.

WES Analysis

From the two families, a total of 12 affected individuals were selected for WES, 5 from R1256 and 7 from R1352 (Figure 2). As shown in Table 1, the additional eight affected individuals were excluded from WES as in some cases it was determined through quality control that there was no sufficient high-quality DNA needed for robust sequencing. As well, we had limitations with regard to available personnel and resources, which limited our study size. WES was completed at McGill University and Genome Quebec Innovation Center (MUGQIC, Montreal, QC, Canada). The Agilent SureSelect 50 Mb All Exon kit was used for exome capture and paired-end sequencing was completed on the Illumina® HiSeq 2000 platform. Following sequencing, initial bioinformatics analyses were completed at MUGQIC using an

https://doi.org/10.1017/cjn.2019.230 Published online by Cambridge University Press

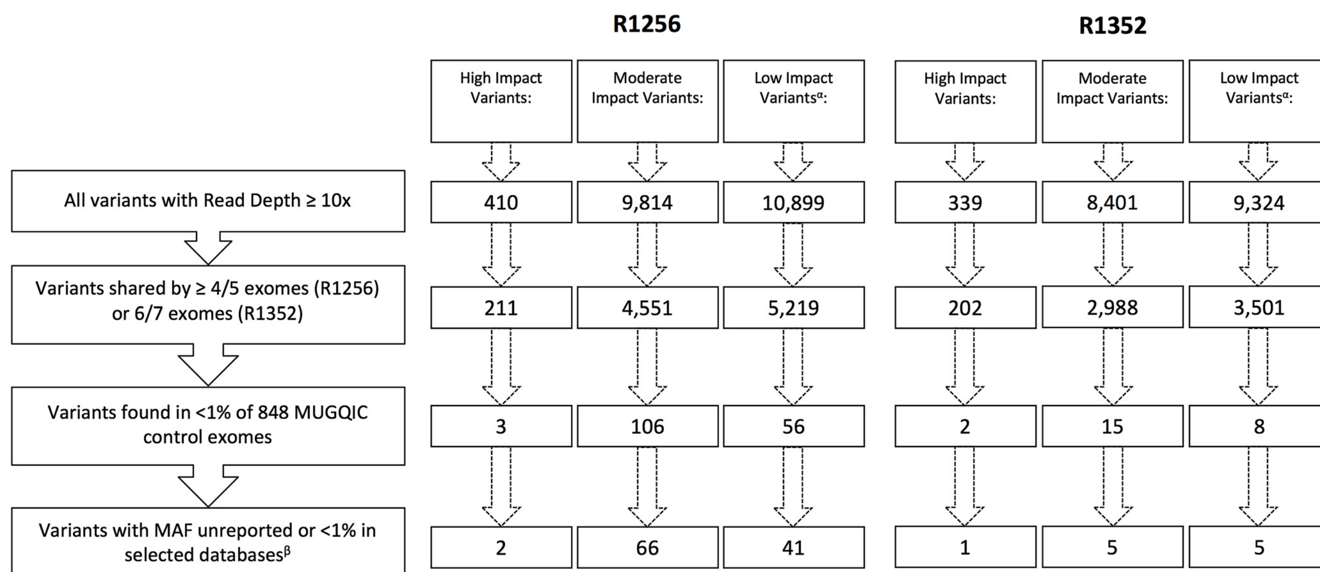


Figure 3: *Remaining variants in each impact category following application of filtering strategy. [α]Low-impact variants were excluded from further analyses beyond filtering. [β]Selected databases include dbSNP, ExAC Browser, Ensembl Genome Browser, and NHLBI Exome Variant Server.*

in-house pipeline for alignment, variant calling, and annotation. Further details are provided in the online supplementary methods.

Variant Filtering Strategy

A filtering strategy was developed to eliminate any variants that were unlikely to play a role in IA pathogenesis. First, annotated variant lists for each family were grouped by predicted impact. For the purpose of our analyses, we filtered each type but focused on variants of high and moderate effect, which excluded synonymous variation from consideration (low impact). High-impact variants included splice-site acceptor, splice-site donor, start-lost, frameshift, stop-gain, stop-lost, and rare amino acid substitution types, while moderate-impact variants were classified as missense variants, according to the provided categorization from MUGQIC. Next, variants with a read depth less than 10× were eliminated. Variants were then prioritized if they were shared by the majority of affected individuals in at least one family. Variants were filtered out if they were present in less than 4/5 exomes in R1256 or less than 6/7 exomes in R1352. Common variants were eliminated next, through the use of population frequency data from the MUGQIC in-house control exome set, and the following data sets: dbSNP, ExAC Browser, NHLBI Exome Variant Server, and Ensembl Genome Browser.[14–17] Variants with a minor allele frequency greater than or equal to 1% in any of these databases were eliminated. Following this step, there were still a high number of moderate-effect variants remaining for family R1256. To suit the parameters and constraints of this pilot project, an additional step was added to focus solely on variants found in 5/5 exomes in this family that were also unreported in variant databases. We modeled the disease inheritance as both recessive and dominant when filtering the variants in both families.

Validation and Segregation in Families

Sanger sequencing was employed to validate filtered variants in affected individuals and eliminate any false positive findings. In this step, an additional five affected family members from R1256 were included, for whom sufficient DNA was available (Z1522, Z1471, Z1459, Z1013, and Z1501 in Figure 2). Primer sequences and specific protocols for polymerase chain reaction are available upon request.

After a list of top candidates was created, additional measures of prioritization were applied. Several statistical algorithms for predicting variant pathogenicity were employed, specifically PolyPhen2, sorting intolerant from tolerant (SIFT), and genomic evolutionary rate profiling, which gives a measure of evolutionary conservation of an amino acid.[18–20] All top candidates were then reviewed manually to ensure that filtering steps were applied properly and that no obviously benign or common variants were remaining. Top candidates were sequenced in a selection of 100 population control samples from NL. These control DNA samples were from NL residents who were recruited through random-digit dialing as part of the Newfoundland Colorectal Cancer Registry project.[21]

Results

Summary of Results

After consolidating our clinical data for the 20 participants in this study, we were able to generate some basic statistics for our cohort. This led into the successful completion of WES, which provided us with 12 exomes for analysis. Through our variant filtering strategy, 6 genetic variants of interest were prioritized in family R1352, and 17 were prioritized in family R1256. It was determined that there were no shared variants between the two families. Segregation analysis with Sanger sequencing showed that the missense variant *SPDYE4* c.C103T (p.P35S) was present in 10 of the 11 affected individuals from family R1256 and was absent in 100 control samples from the NL population. This led us to categorize it as a top candidate in this family. The high-impact variant *C4orf6* c.A1G (p.M1V) was present in six of seven affected individuals and was absent in 100 control samples as well in family R1352. The fact that this variant was classified as high impact led us to categorize it as a top candidate in the family.

https://doi.org/10.1017/cjn.2019.230 Published online by Cambridge University Press

Table 2: Variants remaining following the application of filtering strategy and segregation testing by Sanger sequencing in family R1256 exomes

Chr position	Gene	Nucleotide change	Amino acid change	Transcript	rs ID	Minor allele frequency	Polyphen2*	SIFT*	GERP*	Number of affected with variant	Affected family members with variant	MUGQIC impact category
4:48833549	*OCIAD1*	c.G-6+1A	N/A	NM_001168254.1	*rs144048911*	0.02851%	N/A	N/A	N/A	6/8	Z929, Z1405, Z1406, Z1448, Z1459, Z1501	High
10:12940429	*CCDC3*	c.425delA	p.Y267TfsTer21	NM_001282658.1	*Unreported*	0.2677%	N/A	N/A	N/A	5/10	Z1390, Z1405, Z1406, Z1441, Z1448	High
6:83869646	*DOPEY1*	c.C6902T	p.A2301V	NM_001199942.1	*Unreported*	N/A	0.004	1	5.77	7/10	Z929, Z1390,Z1405, Z1406, Z1441, Z1448, Z1459	Moderate
10:13043354	*CCDC3*	c.C217G	p.L73V	NM_031455.3	Unreported	N/A	0.998	0.06	4.33	7/10	Z929, Z1390, Z1405, Z1406, Z1441, Z1448, Z1471	Moderate
10:135012698	*KNDC1*	c.G2686T	p.A896S	NM_152643.6	Unreported	N/A	0.689	0.01	3.83	7/10	Z929, Z1013, Z1405, Z1406, Z1441, Z1448, Z1471	Moderate
10:135215690	*MTG1*	c.C611T	p.P204L	NM_138384.2	Unreported	N/A	1	0	5.59	3/10	Z929, Z1013, Z1448	Moderate
17:8661598	*SPDYE4*	c.C103T	p.P35S	NM_001128076.1	Unreported	N/A	0.106	0.07	2.65	10/11	Z929, Z1013, Z1390, Z1405, Z1406, Z1441, Z1448, Z1459, Z1471, Z1501	Moderate
17:38029359	*ZPBP2*	c.A622T	p.T208S	NM_198844.2	Unreported	N/A	0.015	0.35	5.67	8/11	Z929, Z1013, Z1390, Z1405, Z1406, Z1441, Z1448, Z1501	Moderate

N/A = not applicable; Het = heterozygous; Hom = homozygous; MUGQIC = McGill University and Genome Quebec Innovation Centre; chr = chromosome; SIFT = sorting intolerant from tolerant; GERP = genomic evolutionary rate profiling.

*PolyPhen2 scores range from 0 (benign) to 1 (damaging), SIFT scores range from 0 (damaging) to 1 (benign), and GERP scores range from −12.3 to 6.17, where 6.17 is the most conserved (Adzhubei et al.[18]; Kumar et al.[19]; Cooper et al.[20]; Exome Variant Server[16]).

https://doi.org/10.1017/cjn.2019.230 Published online by Cambridge University Press

Table 3: Variants remaining following the application of filtering strategy and segregation testing by Sanger sequencing in family R1352 exomes

Chr position	Gene	Nucleotide change	Amino acid change	Transcript	rs ID	Minor allele frequency	Polyphen2*	SIFT*	GERP*	Number of affected with variant	Affected family members with variant	MUGQIC impact category
4:5527058	*C4orf6*	c.A1G	p.M1V	NM_005750.2	*rs144117694*	0%	0	0	0.225	6/7	Z1039, Z1040; Z1497, Z1507; Z1508, Z1651	High
1:160141491	*ATP1A4*	c.C1798T	p.P600S	NM_144699.3	*rs142338502*	0.2002%	1	0	4.19	6/7	Z1039, Z1040,Z1497, Z1507(homozygous), Z1508, Z1533	Moderate
2:233712109	*GIGYF2*	c.A3494G	p.H1165R	NM_001103148.1	*rs72554081*	0.1586%	0.999	0.19	5.43	6/7	Z1039, Z1040; Z1497, Z1507;Z1508, Z1533	Moderate
8:10480510	*RP1L1*	c.C202T	p.L68F	NM_178857.5	Unreported	N/A	1	0	4.78	6/7	Z1039, Z1040; Z1497, Z1507; Z1508, Z1651	Moderate

N/A = not applicable; Het = heterozygous; Hom = homozygous; MUGQIC = McGill University and Genome Quebec Innovation Centre; chr = chromosome; SIFT = sorting intolerant from tolerant; GERP = genomic evolutionary rate profiling.

*PolyPhen2 scores range from 0 (benign) to 1 (damaging), SIFT scores range from 0 (damaging) to 1 (benign), and GERP scores range from −12.3 to 6.17, where 6.17 is the most conserved (Adzhubei et al.[18]; Kumar et al.[19]; Cooper et al.[20]; Exome Variant Server[16]).

Phenotyping of Study Cohort

Based on the data collected in Table 1, we were able to provide a detailed phenotype for all 20 affected participants in our study. In the process, several observations were made, and characteristics of the cohort were defined. For the 20 affected study participants across both families, the average age of diagnosis was 52.3, while the age range of initial diagnosis was 29–79. The 79-year-old male (Z1495) from family R1352 was an outlier in our set, who was found to have two small IAs during autopsy. Our cohort consists of 16 females (80%) and 4 males (20%). Nine individuals from our cohort had been previously diagnosed with hypertension, and 13 had cigarette smoking (previously or currently) as a risk factor. Five of these individuals had both hypertension and a smoking-related history. These two risk factors were not disclosed for three individuals (Z1013, Z1040, and Z1651). As well, eight individuals were reported as having more than one IA, all of whom were females. There was no clear pattern across each family in terms of aneurysm size and/or location, as diagnoses were quite variable.

WES and Variant Filtering

After WES, a range of 40–63 million reads were generated for each library in our set of 12 exomes. The average coverage for the whole genome was 2.8% (total number of aligned reads/size of the genome) and the ratio of surviving reads following trimming to the number of raw reads was 91.85% across all 12 exomes.

Following the application of our stringent filtering strategy, the number of surviving variants was greatly reduced in each family. Figure 3 outlines the number of variants remaining after each filtering step. After step four, only 6 variants in family R1352 (1 high impact and 5 moderate impact), and 68 variants in family R1256 (2 high impact and 66 moderate impact) were prioritized for further investigation (see Supplementary Tables 1 and 2). The 66 moderate-impact variants were filtered further and reduced to a more manageable list of 15 that were present in 5 of 5 exomes and were apparently novel according to our variant annotation, leaving us with 17 variants to investigate in this family (see Supplementary Table 2).

Validation and Segregation

Following Sanger sequencing validation and elimination of false positives, it was acknowledged that there were no shared variants between the two families. As well, there were no variants in the same gene across both families.

In family R1256, two high-impact variants were confirmed: a frameshift variant in *CCDC3* and a splice-site variant in *OCIAD1* (Table 2). For this family, we were fortunate to have a high recruitment rate for affected individuals and their DNA. Therefore, we were able to sequence these variants in additional affected relatives. *CCDC3* c.425delA (p. Tyr267ThrfsTer21) was only validated in 5 of 10 sequenced individuals, though it had been detected in six exomes. This could have been an error in variant calling and was flagged for further review. *OCIAD1* c.G-6+1A was present in six of eight individuals that were successfully sequenced for this difficult region of the gene.

Of the 15 moderate impact variants prioritized, validation of only six was successful (Table 2), as the remainder were located in a highly polymorphic region of the exome. Missense variants in *DOPEY1*, *CCDC3*, and *KNDC1* were confirmed in 7 of 10 affected individuals. *MTG1* c.C611T (p.P204L) was found in

https://doi.org/10.1017/cjn.2019.230 Published online by Cambridge University Press

only 3 of 10, possibly due to sequencing or read depth quality. We were able to successfully obtain sequence from family member Z1522 to test the two remaining variants. Therefore, we were able to validate *ZPBP2* c.A622T (p.T208S) in 8 of 11 affected individuals, and *SPDYE4* c.C103T (p.P35S) in 10 of 11. This variant had the highest degree of segregation with IA incidence, compared to the other candidates. Following sequencing in unaffected relatives, 8 of 14 individuals had the *SPDYE4* variant. Again, this step allowed us to develop a better picture of the variant segregation but was not used as a filtering step to exclude variants. Sanger sequencing of 100 population controls also revealed an absence of the *SPDYE4* variant, and it was identified as the top candidate for IA susceptibility in this data set.

In family R1352, a single high-impact variant, *C4orf6* c.A1G (p.M1V), passed the filtering criteria and was confirmed to be a true variant following Sanger sequencing; it was present in six of seven exomes (Table 3). Additional sequencing determined that 5 out of 14 selected unaffected relatives were also carriers of this variant. This variant had conflicting results with regard to *in silico* predictive tools, as it was reported to be benign by PolyPhen2 and deleterious by SIFT. As this was the only high-impact candidate remaining in the family, it was chosen to be Sanger sequenced in the NL population control data set. *C4orf6* c.A1G (p.M1V) was not found in any of the 100 control samples and was identified as a top candidate in this family.

Additionally, validation proved that three missense variants in *ATP1A4*, *GIGYF2*, and *RP1L1* were true positives and were present in six out of seven exomes (Table 3). Genetic function was explored for all three variants, using databases such as UniProt and GeneCards.[22,23] All three variants were present in 4–5 unaffected relatives; therefore, the results of sequencing these individuals did not add any value to variant prioritization.

Discussion

It is apparent that IA is a genetic disorder in the featured multiplex families, due to the number of closely related affected individuals and the presence of IA in multiple generations. Other analyses using this abundance of phenotype data could be beneficial and could be used to rank phenotype severity. For example, family members with larger aneurysms or rupture history could have additional environmental or genetic risk factors compared to individuals with smaller IA dimensions. The risk factors of hypertension and smoking could also be included in the filtering process, and further investigation into these factors in the unaffected relatives could provide insight into the complexity of this disease.

The use of a geographically isolated population provides several advantages to our study. It is widely known that the NL population is distributed throughout a collection of outports and that families and neighbors are proficient at maintaining information about their genealogy. Isolated populations also have a more consistent genetic landscape due to low rates of migration and a reduced number of disease-causing alleles due to the presence of a genetic bottleneck. They also have a high tendency to participate in research, which is shown through the strong recruitment to our study. Clinical data and active participation from families prove invaluable when completing molecular investigations in a complex disease background. In light of the fact that the genetic landscape of NL is isolated and unique, the variants detected in this study may not be detected at the same allele frequencies in other populations.[24]

During the course of this pilot study, manuscripts from Farlow et al. and Yan et al. were released, which detailed the use of WES in families affected by IA. Each provided a list of prioritized variants for further investigation. The data from these studies was compared to our variant lists, and there was no overlap across these three studies.[4,5]

With the aid of next generation sequencing technology, we were able to identify and prioritize rare variants in our own set of 12 exomes. Following validation, it was confirmed that neither variant from our filtered list was detected in all affected participants from a single family. However, the design of our filtering strategy provided some room for incomplete segregation. As it is known that IA can occur sporadically, and is heavily influenced by lifestyle factors, we chose to examine variants that were detected in all exomes but also variants that were found in four of five exomes from R1256 and six of seven exomes from R1352. The presence of phenocopies could account for the lack of a shared variant in affected individuals. Unruptured IAs have a prevalence of approximately 3.2% and thus occur relatively commonly in the general population.[1] This prevalence could be elevated in our isolated and homogenous population, and it is possible that some family members with strong environmental risk factors could have a separate cause entirely.[24]

Though incomplete segregation was present, further functional investigations could be pursued to determine the biological implications of the *C4orf6* and *SPDYE4* variants. Currently, there is limited information about the function of these genes. At this point in our project, we have concluded that it is unlikely that a single rare variant is responsible for predisposition to IA in our cohort from NL. Through our preliminary genetic investigations in this cohort, several new questions have been stimulated with regard to the genetic etiology of familial IA. Through the interpretation of results and structure of the pedigree in family R1352, we considered the possibility of digenic inheritance in this disease. For example, in this family, four siblings (Z1039, Z1040, Z1497, and Z1508) and one half-sibling (Z1507) shared both *C4orf6* c.A1G (p.M1V) and *GIGYF2* c.A3494G (p.H1165R). Individual Z1533 had the *GIGYF2* variant, while Z1651 had the *C4orf6* variant. There was no clear connection between IA severity (size and rupture) and the combination of these two variants. However, it is an area that could be pursued further as digenic or polygenic inheritance has been largely unexplored in familial IA studies. As well, it is likely that there is genetic heterogeneity in this disease that will result in various causative IA genes throughout our group of families.

There were several limitations to our study, which can be improved upon in future analyses of our unique cohort. In any exome sequencing project, there are technical limitations to consider. Coverage of the exome can be impacted by DNA quality and stringent quality control surrounding read depth. It is also possible that genetic factors influencing IA development are found outside of the exome entirely. As our knowledge of the non-coding region of the genome increases, we will be able to confidently pursue whole genome-wide methods of analysis in the future. The filtering strategy that we designed for this study

https://doi.org/10.1017/cjn.2019.230 Published online by Cambridge University Press

could also be a limitation. It is possible that our methods were too stringent for this cohort and the characteristics of the disease. It is possible that variants involved in IA predisposition are not as universally rare as we predicted and may have a global minor allele frequency higher than 1% and are thus less penetrant in general. As well, there are gaps in the available data that hindered our ability to interpret and prioritize variants. It is possible that some of the genes that we prioritized could have IA-related functionality that is yet unknown.

Overall, this study provided a starting point for further study of familial IA in a unique genetic isolate. In conjunction with researchers that are pursuing WES analysis in other populations, we hope to add to the growing body of knowledge surrounding this disease. The identification of any common biological pathways or gene categories prioritized through different WES studies could be a significant step forward.

Acknowledgments

We would like to thank all of the study participants and their families, without whom this research would not be possible. We would like to acknowledge Krista Mahoney for her assistance with preliminary data collection as well as Carol Negrijn and Marian Crowley for their assistance in acquiring clinical information.

Funding

This project was primarily funded by the Heart and Stroke Foundation of Canada (Grant Number: G-15-0008943), and a grant from the Medical Research Fund at Memorial University's Faculty of Medicine.

Disclosures

None.

Statement of Authorship

AEP assembled the manuscript, created and applied the variant filtering strategy to exome data, and completed Sanger sequencing. BAF and FM collaborated in the creation of this study and are responsible for patient recruitment and phenotyping, along with the assistance of BN. FM also coordinated and performed screening and diagnosis of individuals. MOW directed the design of the molecular genetics study, supervised data collection and analysis, and assisted in primary manuscript revisions. All authors approved the final manuscript.

Supplementary Material

To view supplementary material for this article, please visit https://doi.org/10.1017/cjn.2019.230.

References

1. Vlak MH, Algra A, Brandenburg R, et al. Prevalence of unruptured intracranial aneurysms, with emphasis on sex, age, comorbidity, country, and time period: a systematic review and meta-analysis. Lancet Neurol. 2011;10(7):626–36.
2. Broderick JP, Brown RD Jr, Sauerbeck L, et al. Greater rupture risk for familial as compared to sporadic unruptured intracranial aneurysms. Stroke. 2009;40(6):1952–7.
3. Tromp G, Weinsheimer S, Ronkainen A, et al. Molecular basis and genetic predisposition to intracranial aneurysm. Ann Med. 2014; 46(8):597–606.
4. 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): e0121104.
5. Yan J, Hitomi T, Takenaka K, et al. Genetic study of intracranial aneurysms. Stroke. 2015;46(3):620–6.
6. Rahman P, Jones A, Curtis J, et al. The Newfoundland population: a unique resource for genetic investigation of complex diseases. Hum Mol Genet. 2003;12:R167–72.
7. Young TL, Penney L, Woods MO, et al. A fifth locus for Bardet–Biedl syndrome maps to chromosome 2q31. Am J Hum Genet. 1999;64(3):900–4.
8. Woods MO, Hyde AJ, Curtis FK, et al. High frequency of hereditary colorectal cancer in Newfoundland likely involves novel susceptibility genes. Clin Cancer Res. 2005;11(19):6853–61.
9. Merner ND, Hodgkinson KA, Haywood AF, 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;82(4):809–21.
10. Maroun FB, Murray GP, Jacob JC, et al. Familial intracranial aneurysms: report of three families. Surg Neurol. 1986;25(1): 85–8.
11. Zhou S, Ambalavanan A, Rochefort D, et al. RNF213 is associated with intracranial aneurysms in the French-Canadian population. Am J Hum Genet. 2016;99(5):1072–85.
12. Mathieu J, PØrusse L, Allard P, et al. Epidemiological study of ruptured intracranial aneurysms in the Saguenay-lac-Saint-Jean region (Quebec, Canada). Can J Neurol Sci. 1996;23(3):184–8.
13. Wills S, Ronkainen A, van der Voet M, et al. Familial intracranial aneurysms: an analysis of 346 multiplex Finnish families. Stroke. 2003;34(6):1370–4.
14. Smigielski EM, Sirotkin K, Ward M, et al. dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res. 2000; 28(1):352–5.
15. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60, 706 humans. Nature. 2016;536:285–91.
16. Exome Variant Server, *NHLBI GO Exome Sequencing Project (ESP)*, Seattle, WA. Available at: http://evs.gs.washington.edu/EVS/.
17. Cunningham F, Amode MR, Barrell D, et al. Ensembl 2015. Nucleic Acids Res. 2015;43:D662–9.
18. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7(4):248–9.
19. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073–81.
20. Cooper GM, Stone EA, Asimenos G, et al. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005;15(7):901–13.
21. Wang P, Dicks E, Gong X, et al. Validity of random-digit-dialing in recruiting controls in a case–control study. Am J Health Behav. 2009;33(5):513–20.
22. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics. 2016;54:1.30.1–33.
23. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45(D1):D158–69.
24. Kristiansson K, Naukkarinen J, and Peltonen L. Isolated populations and complex disease gene identification. Genome Biol. 2008; 9:109.1–9.

https://doi.org/10.1017/cjn.2019.230 Published online by Cambridge University Press