Identifying the Genetic Etiology of Spontaneous Coronary Artery Dissection Using Exome and Whole Genome Sequencing Data

Hannah Burr

Division of Experimental Medicine

McGill University

Montreal, Quebec, Canada

August 2020

A thesis submitted to McGill University

in partial fulfillment of the requirements of the degree of

Master of Science in Experimental Medicine.

© Hannah Burr, 2020.

Table of Contents

Abstract

Background: Spontaneous coronary artery dissection (SCAD) is a rare presentation of Acute Coronary Syndrome (ACS), which primarily affects individuals without cardiovascular risk factors. Previous studies indicate that connective tissue disorders (CTDs), systemic inflammatory disorders, and milder vasculopathies are prevalent among SCAD patients, however, less than 1% of SCAD patients are diagnosed with these predisposing conditions. Previous genetic studies of SCAD that sequenced exclusively CTD genes yielded a low rate of genetic diagnosis (5-8%). In addition, due to the low penetrance of SCAD with only 1.2% of cases showing familial inheritance, the use of family data may not be practical. New approaches are needed to facilitate improved diagnosis, treatment, and genetic counselling for SCAD patients.

Methods: I investigated the rare genetic causes of angiographically diagnosed SCAD in 45 patients from the multi-center GENESIS-PRAXY study of early-onset ACS, who provided a detailed medical history, laboratory measurements, and one-year follow up outcomes. I first applied Whole Exome Sequencing (WES) to five patients to determine if this method combined with the Exomiser software could capture causative variants in patients without traditional risk factors. I next applied Whole Genome Sequencing (WGS) to all 45 SCAD patients in GENESIS-PRAXY. I designed a filtering approach based on allele frequency, variants in public databases, predictions of functional effect for each of the variants, and similarity between SCAD and phenotypes caused by each gene in the genome. Our approach examined coding and non-coding variation, did not require familial data, and was not restricted to genes that had previously been associated with SCAD.

Results: In the WES study, I identified three Variants of Uncertain Significance (VUS) in *FBN1, LEMD3,* and *ADAR*. *LEMD3* and *ADAR* had not previously been associated with SCAD, but

have strong phenotypic links to SCAD and are in related pathways. In the WGS study, I identified pathogenic and likely pathogenic variants in three patients (6.7%) and VUS in 28 patients (62.2%). Six patients (13.3%) had VUS in regulatory regions. This is the first report that rare regulatory variants may contribute to SCAD. Of the coding VUS, 12 patients (26.6%) had variants in genes previously associated with SCAD while 11 patients (24.4%) had variants in other genes. These genes include *COL7A1, COL4A5, COL12A1, COL27A1, LTBP4, LEMD3, GDF5, ITGB8, ITGB4, ITGB5, CECR1, BLK, ADAR,* and *KDM6A.* These genes are associated with CTDs or inflammatory conditions, and rank in the top 6% of genes in the genome for phenotypic relevance to SCAD. Four patients (8.9%) had variants in more than one gene, the first report that SCAD may be oligogenic in some patients. 14 patients (31.1%) had no identified rare variant. These patients were significantly more likely to have multivessel dissection, suggesting that multivessel SCAD may have a different etiology.

Conclusions: This study identified rare variants in collagen, TGF-β signaling pathway, and inflammation genes, which are well-known to be associated with SCAD, and identified new genes within these same groups. The high rate of VUS findings in CTD genes needs to be investigated in larger studies to determine if a higher proportion of SCAD patients may have variation in CTD genes. A thorough investigation of CTD-related phenotypes in these SCAD patients could also determine if some patients with VUS findings have mild forms of a CTD that had gone undiagnosed. In the GENESIS-PRAXY cohort, all three individuals with a pathogenic or likely pathogenic variant probably had a recurrent SCAD event, a trend that supports a similar observation in the related vasculopathy thoracic aortic aneurysm. Most of my results are VUS and thus not considered clinically actionable under ACMG guidelines, which could be improved to include regulatory variation and rely less heavily on familial segregation data.

Résumé

Contexte: La dissection spontanée de l'artère coronaire (SCAD) est une présentation rare du syndrome coronarien aigu (SCA) qui affecte majoritairement des individus sans des facteurs de risque cardiovasculaire. Des études effectués précédemment indiquent que des troubles du tissus conjonctif (TTC), des troubles inflammatoires systémiques, ainsi que des autres vasculopathies sont courants chez les patients SCAD, cependant, peu de patients SCAD ont un diagnostic de ces conditions. Des études antérieures de SCAD n'ont mené qu'à un faible taux de diagnostic génétique (5-8%). Puisque la pénétrance familiale de SCAD est limitée à 1.2% des cas, l'utilisation des séquences familiales peut s'avérer inefficace. Le développement de nouvelles approches est nécessaire pour faciliter le diagnostic, le traitement et les conseils génétiques chez les patients SCAD.

Méthode: J'ai étudié les causes génétiques rares de SCAD diagnostiqué par angiographie chez 45 patients de l'étude multicentrique GENESYS-PRAXY portant sur le SCA d'apparition précoce. Les patients ont fourni des antécédents médicaux détaillés, des mesures de laboratoire et des résultats de suivi d'un an. J'ai d'abord soumis cinq patients à un séquençage de l'exome (WES) pour déterminer si cette méthode, combinée au logiciel Exomiser peut capturer la variante causale chez des patients sans des facteurs de risque cardiovasculaires. J'ai ensuite soumis les 45 patients SCAD dans GENESIS-PRAXY à un séquençage complet de leur génome (WGS). J'ai conçu une approche de filtrage basée sur la fréquence allélique, les variantes dans les bases de données publiques, les prédictions de l'effet de toutes les variantes, et la similarité entre SCAD et les phénotypes causés par chacun des gènes du génome. Notre approche comprenait les variantes codantes et non-codantes, ne nécessitait pas de données familiale et n'était pas restreinte aux gènes précédemment associés à SCAD.

Résultats: Chez les cinq patients WES, j'ai identifié trois variantes génétiques de signification inconnue (VUS) dans les gènes *FBN1, LEMD3,* and *ADAR. LEMD3* et *ADAR* n'ont jamais été associés à SCAD, mais ces gènes ont de forts liens phénotypiques avec SCAD. Dans l'étude du WGS, j'ai identifié des variantes pathogènes et probablement pathogènes chez trois patients (6.7%) et des VUS chez 28 patients (62.2%). Six patients (13.3%) avaient des VUS dans des régions régulatrices. Parmi les VUS codantes, 12 patients (26.6%) avaient des variantes de gènes déjà associées avec SCAD, tandis que 11 patients (24.4%) avaient des variantes dans d'autres gènes. Ces gènes incluent *COL7A1, COL4A5, COL12A1, COL27A1, LTBP4, LEMD3, GDF5, ITGB8, ITGB4, ITGB5, CECR1, BLK, ADAR* et *KDM6A*. Ces gènes se classent dans le 6% supérieur parmi tous gènes du génome en termes de pertinence phénotypique pour SCAD. Quatre patients (8.9%) avaient des variantes dans plus d'un gène. Je n'ai trouvé aucune variante rare chez 14 patients (31.1%). Ceux-ci étaient plus susceptibles d'avoir une dissection de plusieurs vaisseaux.

Conclusions: Cette étude a identifié des variantes génétiques rares dans les collagènes, la signalisation TGF-β et l'inflammation qui sont déjà connues pour être associées au SCAD et a identifié des nouveaux gènes d'intérêt dans ces mêmes voies. Le taux élevé de VUS trouvées dans les gènes des TTC doit être étudié plus à l'avenir dans des études plus grandes afin de déterminer si une proportion plus élevée de patients SCAD peuvent avoir des variations dans les gènes liés aux TTC. Une étude approfondie des phénotypes liés aux TTC chez ces patients SCAD pourrait également déterminer si certains patients avec des VUS ont des formes bénignes d'une CTD qui n'avaient pas été diagnostiquées. La majorité de mes résultats sont des VUS et ne sont donc pas cliniquement exploitables selon les lignes directrices de l'ACMG, qui pourraient

être améliorées par l’ajout de variantes régulatrices et s'appuient moins fortement sur les données de ségrégation familiale.

Acknowledgements

I would like to personally thank my co-supervisors, Dr. Engert and Dr. Thanassoulis, for their commitment to their students and their efforts and guidance in supervising this work. I especially appreciate the freedom and encouragement I was given to follow my own instincts about the direction this project should take, and your extreme dedication to helping me achieve my future career goals. It has been an absolute pleasure to work with you during the four years that I have been in the lab, both in my undergraduate and now Master's degree.

I would also like to personally thank Dr. Pilote for establishing the cohort used in this study and her valuable suggestions based on clinical insights. I would also like to thank the additional members of my thesis supervisory committee, Dr. Simone Chevalier and Dr. Jean-Baptiste Rivière for their thoughtful questions and constructive criticisms that helped establish our analysis plan. I would like to thank all other members of the Engert-Thanassoulis lab (Katia Desbiens, Line Dufresne, Maedeh Mohammadnetaaj, Hao Yu Chen, Mary Hoekstra, Thomas Marsh, and Yoshihiko (Tiger) Nagai) for their support, insights, and cherished comradery. The McGill University and Génome Québec Innovation Centre was also an integral part of the success of these studies, and I would like to specifically thank Dr. Mark Lathrop, Dr. Hans Markus Münter, and Dr. Pierre Lepage for the contribution of their vast expertise in genome sequencing technologies.

I would finally like to thank my funding sources, including recruitment and travel awards from the McGill University Division of Experimental Medicine and an Internal Award from the McGill University Faculty of Medicine.

Contribution of Authors

The General Introduction (Chapter 1), Transition (Chapter 3), and General Discussion (Chapter 5) were written by **Hannah Burr** and revised by **Dr. Jamie Engert** and **Dr. George Thanassoulis**.

Katia Desbiens translated the Abstract to the French Résumé.

Chapter 2: Identification of Rare Genetic Determinants of Spontaneous Coronary Artery Dissection (Manuscript to be submitted)

Burr: Performed all statistical genetic analysis, drafted the manuscript

Pilote: Created the GENESIS-PRAXY cohort, revised manuscript

Cheema: Identification and description of SCAD and coronary artery hyper-tortuosity within GENESIS-PRAXY

Dufresne: Data extraction and cleaning in GENESIS-PRAXY, revised the manuscript

Desbiens: DNA extraction and preparation of samples for sequencing, revised the manuscript

Lepage: Sanger sequencing for confirmations

Engert, Thanassoulis: Supervisory support, procured funding, revised the manuscript

Chapter 4: Identifying Genetic Determinants of Spontaneous Coronary Artery Dissection with Genome Sequencing (Manuscript in preparation to submit to Genetics in Medicine)

Burr: Designed the analysis plan and performed all genetic analysis, drafted the manuscript

Pilote: Created the GENESIS-PRAXY cohort

Cheema: Identification and description of SCAD and coronary artery hyper-tortuosity within GENESIS-PRAXY

Dufresne: Data extraction and cleaning in GENESIS-PRAXY, revised the manuscript

Desbiens: DNA extraction and preparation of samples for sequencing, revised the manuscript

Münter: Supervised genome sequencing, quality control, and variant calling

Lathrop: Study design, procured of funding

Engert, Thanassoulis: Supervisory support, procured funding, revised the manuscript

Chapter 1: General Introduction

Overview of the Guidelines Used to Identify the Cause of Rare Genetic Diseases and Their Limitations

Since the completion of the Human Genome Project, researchers have been able to pinpoint specific genes and variants responsible for a wide variety of diseases. In the context of rare Mendelian diseases, a single nucleotide variant unique to a particular family can often be identified as the cause of the disease. Scientific literature has filled databases such as the Online Mendelian Inheritance of Man (OMIM, https://www.omim.org/), ClinVar,[1] and the Human Gene Mutation Database[2] (HGMD) with disease-gene and disease-variant relationships, allowing clinicians and genetic counsellors to order sequencing of a targeted panel of genes consistent with an individuals suspected phenotype to inform treatment options and identify a specific genetic test for immediate family members. Once enough data amassed, researchers defined a general set of rules describing well-validated, disease-causing variation and compiled them into the 2015 American College of Medical Genetics (ACMG) and Association for Molecular Pathology (AMP) guidelines for variant classification to standardize the protocol used to define a definitive genetic diagnosis among different labs.[3] These guidelines set forth 28 criteria (16 in support of a pathogenic rating and 12 in support of a benign rating) of varying levels of strength (very strong, strong, moderate, and supporting) that are used in different combinations to determine if a variant is considered clinically actionable and very likely to cause disease (pathogenic or likely pathogenic), uncertain (variant of uncertain significance (VUS)), or unlikely to cause disease (likely benign or benign).

Even though the ACMG/AMP guidelines increased the agreement between labs on the genetic diagnosis,[4] many patients were still left without a genetic diagnosis, evidenced by the

over 4,500 applications to the Undiagnosed Diseases Network since its advent in 2008 (https://undiagnosed.hms.harvard.edu/). Some diseases have a degree of genetic heterogeneity, meaning that the same phenotype can be a result of a variant in one of many different genes, usually in the same pathway.[5] Therefore, in some undiagnosed cases the patient may have a causal genetic variant in a gene that has not yet been studied in relation to the disease and thus was not included in the targeted sequencing panel. A more thorough sequencing method such as whole exome sequencing is required to achieve a genetic diagnosis for such a patient.[6] Exome sequencing is also commonly applied when the patient has a novel phenotype, or an unusual clinical presentation of a known condition leading to diagnostic uncertainty.[6] In these cases, there is no prior knowledge of which gene could be causal so all genes must be considered.

Whole exome sequencing (WES) is a method to capture only exonic regions of the genome, allowing researchers studying disease to efficiently sequence a much larger number of clinically relevant genomic regions compared to Sanger sequencing, but for a fraction of the cost and the time required for whole genome sequencing (WGS). The most common way to perform WES is to fragment the sample DNA and hybridize it to a commercially available microarray containing a library of cDNA probes specific for all exons in the genome, amplify the hybridized DNA by PCR, and then sequence these fragments with massively parallel short read sequencing.[7] While this technology has revolutionized genetic diagnosis, in recent years some important limitations have been identified. First, the library capture method of WES results in extremely variable read depth across the genome. The variable depth leads to difficulty calling variants in lower depth regions, and thus WGS detects 3% more coding variants. Therefore, WES has a slightly higher false negative rate than WGS for coding variants.[8] In addition, different genes are affected by poor sequencing depth in different capture kits, therefore it is

important to use a capture kit that will perform well for the genes of interest.[9] Finally, although UTRs are exonic and thus should be included by WES, a recent study indicates that only 20% of UTRs are included in WES results generated by popular capture kits.[10]

If a suspicious variant is identified in a new gene or in a novel domain of a known gene through WES and bioinformatic methods, the ACMG/AMP guidelines ensure that the result is well-validated and has sufficient evidence of causality before it is reported to the patient, used to prescribe treatments, or accepted by the scientific community.[3] Causal confirmation can be achieved in a few ways. One possibility is to perform functional assays to confirm that the variant perturbs the protein function (ACMG criteria PS3).[3] Functional validation can take the form of a measurement from a patient derived sample, observation of phenotypes in a model organism with the homolog of the variant under study introduced, or assessing the molecular consequence of the variant in an *in vitro* cell line.[11] Thus, achieving the PS3 criteria faces the limitations that the molecular mechanism of the disease must be known in order to select an appropriate functional measurement, and in the case of an animal model, assumes that the model organism has the same sensitivity to this particular perturbation.[11] Recently, the ClinGen Variant Curation Expert Panel[12], which was created to improve the consistency of variant interpretations currently available in ClinVar,[1] released more specific guidelines regarding the acceptability of functional assays to achieve criteria PS3 dictating that 10-11 variant controls including a mix of known pathogenic and known benign variants should be included in each assay to validate its utility in the clinical genetics setting.[11] For diseases that have not been rigorously studied, an appropriate number of control variants may not be available to develop a suitable assay. Thorough functional validation in this way remains prohibitively expensive and time-consuming for many diseases,[11] especially if multiple candidate variants are identified in a single patient.

Another validation option is to sequence both affected and unaffected family members to confirm that the variant segregates according to the appropriate inheritance pattern (criteria PS2, PM3, PM6, PP1). Segregation studies can easily be performed with only a trio in the context of classical Mendelian diseases, however, they become more difficult when the disease has incomplete penetrance or variable expressivity. In the case of incomplete penetrance, only a certain percentage of the individuals harboring the variant will express the disease phenotype.[13] Incomplete penetrance has been reported for a number of diseases.[13] The rate of penetrance can depend on the variant type (for example, reduced penetrance of loss-of-function variants compared to missense variants in Ehlers-Danlos syndrome), genetic background (for example, patients with variants in both *BRCA1* and one of *PALB2, MLH1,* or *BRCA2* have increased rates of cancer compared to patients with *BRCA1* variants alone), or environmental factors (for example, the patient's age).[13] In the case of variable expressivity, the disease has a range of phenotypes and severity, and thus not all individuals harboring the same variant will have identical symptoms.[13] In order to perform a segregation study in either of these cases, many more family members need to be sequenced and thoroughly evaluated for mild manifestations of the disease. In the context of personalized medicine, this approach is unrealistic because it relies on the proband having a large family willing to consent to the study. The ACMG/AMP criteria rely heavily upon this validation option.[3]

Another possibility, when both targeted sequencing and whole exome sequencing do not yield a diagnosis for a genetic disease, is that the variant is located in a non-coding regulatory region of the gene. In this case, it will only be discovered with whole genome sequencing (WGS). The current ACMG/AMP guidelines emphasize coding variants, with five out of 16 criteria supporting pathogenicity specific to coding variation.[3] A robust schema describing features that

could lead to prioritization of pathogenic non-coding mutations has not yet been developed due to a number of challenges, including insufficient annotation tools and lack of robust functional studies in non-coding domains.[14]

Of the 14 total combinations of criteria that can lead to a pathogenic or likely pathogenic rating, four of them require familial segregation data.[3] Of the 10 remaining combinations, five require the mutation to be a loss-of-function (truncation, frameshift, or splicing change), four require functional validation or that the specific variant has previously been identified as pathogenic, and the final one requires that the gene has been well studied and has a low rate of benign missense variation.[3] For these reasons, many patients are left with a VUS finding, especially if the variant is in a non-coding regulatory region. In fact, of the 1,527,368 variants contained in the most recent version of ClinVar (July 28, 2020), 687,055 (45%) are VUS. The ACMG/AMP guidelines state that VUS should not influence clinical decision-making.[3] Furthermore, VUS findings can increase a patient's distrust of physicians and counselors, cause additional confusion, and/or heighten anxiety about their condition and genetic risk to their family members.[15]

Spontaneous Coronary Artery Dissection as a Rare Genetic Cause of Heart Attack

Complex diseases are influenced by multiple genetic and environmental factors.[16] One example is acute coronary syndrome (ACS), a condition that includes myocardial infarction (MI), often referred to as a "heart attack". The most common cause of ACS is coronary artery disease (CAD), which causes a build-up of atherosclerotic plaque in artery walls. These plaques can rupture, leading to *in situ* thrombosis that blocks the coronary arteries, limits blood flow, and reduces oxygen perfusion to the heart. In many patients, CAD is attributed to environmental

factors such as advanced age, male sex, increased BMI, smoking, hypertension, and increased lipids. There is also a genetic component to CAD, as evidenced by the rare condition Familial Hypercholesterolemia (FH).[17] FH patients have markedly genetically elevated LDL-cholesterol levels, which puts them at risk of experiencing more severe, early-onset CAD.

In addition to the contribution of rare genetic variants, numerous Genome Wide Association Studies (GWAS) have identified loci common in the general population that slightly increase an individual's risk of CAD. Because these risk loci are present in significant proportions of the population, certain individuals may, by chance, have a higher than usual accumulation of these loci.[18] This risk accumulation can be quantified with a polygenic risk scores (PRS), which is often calculated as the sum of an individual's risk alleles weighted by the effect size calculated in the GWAS.[18] A PRS is especially useful for oligogenic diseases, or disease where multiple genes simultaneously contribute to the development of a phenotype. Recent work indicates that individuals in the highest categories of polygenic risk for five common diseases have a risk profile similar to that of individuals with a rare Mendelian cause of disease.[19]

Despite the fact that CAD is well studied, it remains difficult to pinpoint whether the cause of disease for an individual patient is environmental factors, a rare genetic variation, a high burden of common genetic variation or a combination of these etiological factors.[20] Thorough preventative assessments of a patient's CAD risk in the context of personalized medicine will be challenging even when taking all of these contributors into account.

A less common type of MI is Myocardial Infarction with Nonobstructive Coronary Arteries (MINOCA), which is thought to cause approximately 5% of acute MI events.[21] The diagnosis of MINOCA requires an invasive imaging procedure, coronary artery angiography,

where the coronary arteries are visualized by X-ray after injecting dye through a catheter. In the case of an MI with CAD, the physician will frequently identify a blockage or significant stenosis (≥70% narrowing of an artery due to atherosclerotic plaque build-up) and will likely perform an angioplasty or stenting procedure treatment. In the case of MINOCA, however, evidence of significant atherosclerotic plaque is not found.[21] Often the cause of MINOCA can be observed on the angiogram and confirmed with another test or imaging modality. Causes of MINOCA include microvascular dysfunction, spontaneous thrombus formation (with subsequent dissolution), coronary artery spasm, and spontaneous coronary artery dissection.[21] Some of these causes have known genetic origins, such as platelet disorders underlying spontaneous thrombosis, and others have known environmental origins, such as coronary spasm and cocaine use. Each of these underlying etiologies has distinct treatment options, thus it is important for clinicians to undertake these additional investigations when possible.[21] Patients presenting with MINOCA tend to be younger at age of presentation, have a lower prevalence of established risk factors, and are twice as likely to be female compared to MI patients with CAD.[21] MINOCA patients are also less likely to present with traditional chest pain and dyspnea often associated with ACS.[22]

Spontaneous coronary artery dissection (SCAD) is a cause of MINOCA that similarly affects younger females without traditional cardiovascular risk factors. SCAD occurs when any of the three layers of the coronary artery, the intima, media, or adventitia, separate from each other, resulting in the formation of a false lumen that fills with blood forming an intramural hematoma (**Figure 1**).[23] The false lumen compresses the true lumen, which causes an obstruction of the affected coronary artery and presentation of ACS. SCAD can be effectively diagnosed on an angiogram and has three distinct appearances (**Figure 2**), which do not appear to correlate with any clinical features or outcomes.[23] 29.1% of patients have the most obvious angiographic

presentation of SCAD (Type 1), in which the double lumen can be clearly seen.[24] 67% of patients have the Type 2 SCAD presentation, which manifests as a sudden long section of stenosis.[24] This presentation can be more difficult to identify for a clinician not extensively trained in angiography and may require a repeat angiography or another imaging modality that can provide a detailed view of the vessel wall, such as intravascular ultrasound (IVUS) or optical coherence tomography (OCT) to confirm. Finally, 3.9% of patients have the Type 3 SCAD presentation, which is indistinguishable from CAD and requires IVUS or OCT to confirm.[24] The relative difficulty of identifying all different types of SCAD, the increased expense and training required to perform follow up IVUS or OCT, and the fact that patients without a complete artery blockage at low risk of requiring an immediate revascularization procedure often do not receive an angiogram, all lead to the underdiagnosis of SCAD among ACS patients.

Early and accurate diagnosis of SCAD at the time of presentation is important because the treatment differs from that of ACS caused by CAD. In one cohort of 162 patients, 95% exhibited spontaneous healing of the dissection after 30 days or more.[25] SCAD patients undergoing the common revascularization procedure, Percutaneous Coronary Intervention (PCI, also commonly known as stent placement) have a 69.7% rate of complications from the procedure including extension of the dissection requiring another PCI procedure, stent thrombosis (excessive blood clotting at the site of stent placement often leading to another ACS), or stent restenosis.[24] Therefore, conservative treatment for SCAD is recommended. A surgical intervention that removes the intramural hematoma (with or without subsequent stent placement), has recently been described; however, larger studies must be performed to determine if this procedure reduces the risk of future complications.[26]

To date there have been no randomized control trials regarding effective management of SCAD post-discharge. A significant number of SCAD patients receive medical therapies similar to those prescribed for patients with ACS due to CAD, including aspirin, beta-blockers, and antiplatelet drugs, as well as ACE inhibitors and statins where indicated. SCAD has a high rate of recurrence, with rates up to 30% reported in a 10-year follow-up period.[27] The only clinical parameter currently identified that predisposes a patient to recurrent SCAD is hypertension, which increases recurrence risk by two-fold, whereas the prescription of a blood pressure lowering beta-blocker reduces risk of recurrence by 60% in a retrospective cohort.[27]

SCAD is hypothesized to be genetic in origin because it is rare and mostly affects younger individuals without traditional cardiovascular risk factors.[28] Indeed, SCAD has been associated with a number of rare genetic conditions, including connective tissue disorders such as Marfan syndrome (MFS), Ehlers-Danlos syndrome (EDS), and Loeys-Dietz syndrome (LDS), vasculopathies such as arterial tortuosity syndrome, fibromuscular dysplasia (FMD), polycystic kidney disease (PKD), and systemic inflammatory conditions known to affect arteries such as systemic lupus erythematosus (SLE), Kawasaki disease, and polyarteritis nodosa.[29] These conditions have also been reported in conjunction with similar conditions affecting vascular integrity, including cervical artery dissection (CeAD)[30] and thoracic aortic aneurysm and dissection (TAAD).[31] Recently, SCAD has also been associated with common variants in GWAS.[32,33] In addition, SCAD has low penetrance, with only an estimated 1.2% of cases have familial inheritance.[34] Over 50% of SCAD patients report a precipitating event, including pregnancy, childbirth, extreme emotion, or extreme physical exertion.[24] Therefore, it is believed that SCAD and related vasculopathies are manifestations of genetically weakened vascular

integrity, where environmental factors serve as a trigger. The vessel location and environmental triggers are thus probabilistic elements, which contribute to reduced penetrance.

Genetic studies of SCAD[35] and TAAD[36] evaluating the same panel of 23 CTD related genes to diagnose patients according to ACMG/AMP pathogenic guidelines had low diagnostic yields (14% and 18% respectively) and also reported a high percentage of VUS (27% and 10% respectively). Both of these studies and others[37] identified patients with pathogenic variants in a CTD gene who were not previously diagnosed with a syndrome, and furthermore exhibited only mild features of these diseases, such as translucent skin and easy bruising, which could be easily missed in a typical primary care setting.

Objectives

The phenotypic variability within CTDs leads us to believe that variants in CTD genes could be a more common cause of SCAD than previously believed, as individuals may not possess any characteristics that could lead to an early diagnosis. I sought to use methods that would accurately identify genetic variants in genes that had not been previously associated with SCAD but had strong evidence of phenotypic relevance to gain novel insight into the etiology of SCAD and elucidate new disease mechanisms and contributing pathways. We first performed Whole Exome Sequencing (WES) on a subset of the cohort to determine the sensitivity of these methods. We then performed Whole Genome Sequencing (WGS) to also include non-coding, regulatory variants in our analysis to perform the most thorough possible assessment of genetic variation that could lead to SCAD and assess whether the added utility warrants the extra expense of this method. Finally, I sought to determine if there were differences in clinical

features between patients with an identified rare genetic variant and those without and identified variant.

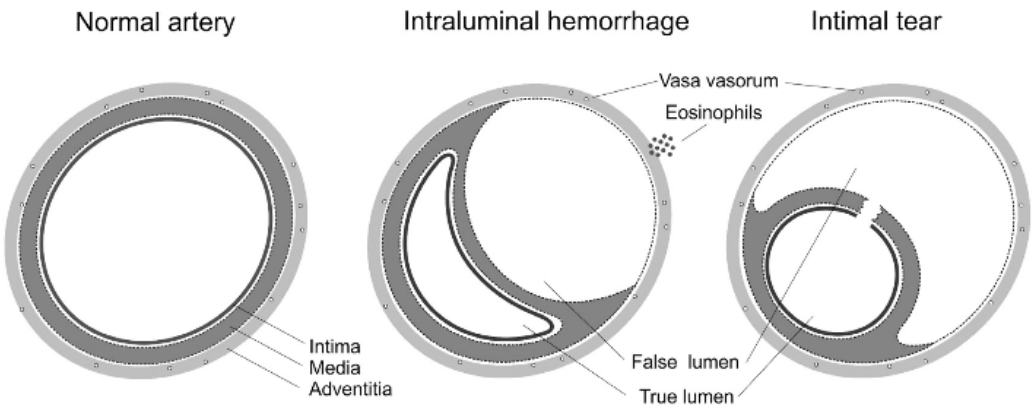
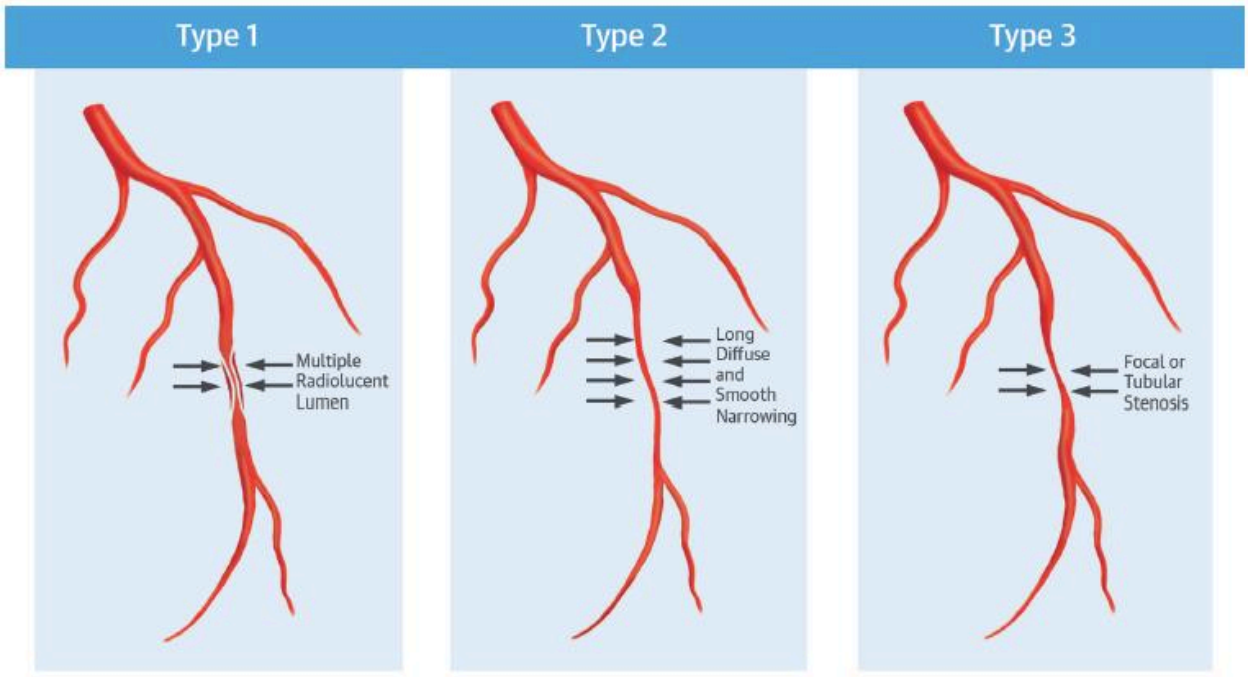


Figure 1. Schematic diagram of the mechanisms of spontaneous coronary artery dissection.

Figure 1

Diagram of the mechanism of SCAD adapted from Saw, J *Can J Cardiol* 2013[38]



Saw, J. et al. J Am Coll Cardiol. 2017;70(9):1148-58.

Figure 2

Drawing depicting the three presentations of SCAD on an angiogram adapted from Saw, J *et al. J Am Coll Cardiol* 2017[27]

Chapter 2

Identification of Rare Genetic Determinants of Spontaneous Coronary Artery Dissection

Hannah A Burr, BSc[1,2]; Louise Pilote, MD, MPH, PhD[3]; Asim N Cheema, MD, PhD, MSc[4]; Line Dufresne, MSc[2]; Katia Desbiens, MSc[2]; Pierre Lepage, PhD[6]; George Thanassoulis*, MD[1,2,5]; James C Engert, PhD* [1,2,5,7]

*co-senior authors

[1]Division of Experimental Medicine, McGill University, Montreal, Canada

[2]Preventative Genomic Cardiology, McGill University Health Centre and Research Institute, Montreal, Canada

[3]Division of General Internal Medicine, McGill University, Montreal, Canada

[4]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada

[5]Division of Cardiology, Department of Medicine, McGill University, Montreal, Canada

[6]Genome Quebec Innovation Centre, McGill University, Montreal, Canada

[7]Department of Human Genetics, McGill University, Montreal, Canada

Corresponding Authors:

James C. Engert, PhD

McGill University Health Centre

1001 Decarie Blvd., Room EM1.2218

Montreal, Quebec H4A 3J1

Canada

Phone: 1-514-934-1934 ext. 35325

George Thanassoulis, MD, MSc

McGill University Health Centre

1001 Decarie Blvd., Room D05.5120

Montreal, Quebec H4A 3J1

Canada

Phone: 1-514-934-1934 ext. 35465

Abstract

Background: It is difficult to offer SCAD patients a genetic confirmation based on American College of Medical Genetics criteria due to low penetrance, genetic heterogeneity, and missing pedigree information. Whether whole exome sequencing (WES) can identify pathogenic variants in genes not previously associated with SCAD is unknown.

Methods: WES was performed on five early-onset SCAD patients with few cardiovascular risk factors from the GENESIS-PRAXY cohort recruited between 2009-2013 who underwent coronary angiography.

Results: We identified potentially causal variants using an approach, ranking variants according to how deleterious they are predicted to be, and their likelihood to cause SCAD. This approach uses population allele frequencies, computational predictions of variant intolerance, protein interactome data, and similarity between diseases and animal phenotypes to determine rankings.

We identified variants in three of the five selected SCAD patients including one case with a variant in *FBN1*, which encodes fibrillin-1. Variants in *FBN1* cause Marfan syndrome. In another patient, we observed a novel variant in *LEMD3*, which codes for Man1, a TGF-β pathway repressor. We also identified a variant in *ADAR*, in which mutations can cause Aicardi-Goutières Syndrome, previously associated with intracerebral large artery disease.

Conclusions: In three patients, we identified mutations in genes that are consistent with vasculopathy and potential novel monogenic etiologies for SCAD.

Introduction

Spontaneous Coronary Artery Dissection (SCAD) is a rare type of Acute Coronary Syndrome (ACS). SCAD causes 1.7-4% of all ACS cases and is particularly prevalent in women under 50, where it accounts for 20-40% of ACS cases.[1] Previous genetic and epidemiologic studies indicate that SCAD has multiple distinct etiologies.[2,3] Connective tissue disorders (CTDs) such as Marfan syndrome; arteriopathies such as fibromuscular dysplasia (FMD) and polycystic kidney disease; and inflammatory diseases such as Kawasaki disease and systemic lupus erythematosus are more prevalent among SCAD patients.[3] These conditions combined, however, explain less than 1% of SCAD cases.[3]

Approximately 1.2% of SCAD cases exhibit familial inheritance[1], however, this figure is likely an underestimate due to phenotypic heterogeneity in vascular pathology. In addition, using familial data to rule out variants present in unaffected family members can lead to false negatives due to reduced penetrance. Currently, a genetic diagnosis for SCAD is only given if a variant pathogenic for a CTD by ACMG guidelines is identified and therefore, this approach results in a low diagnostic yield. To improve the identification of genetic causes of SCAD, we tested an approach that broadens the possible candidate genes to include genes that have not previously been linked to SCAD or a CTD.

Methods

GENESIS-PRAXY study

The GENESIS-PRAXY study (Gender and Sex Determinants of Cardiovascular Disease: From Bench to Beyond Premature Acute Coronary Syndrome) is a multi-center prospective study of patients age 18-55 admitted to hospital with ACS that has been previously described in

detail.[4] Briefly, patients were recruited in 24 participating hospitals across Canada, the United States, and Switzerland. All participating sites received ethics approval from their respective ethics review boards. Patients were eligible for inclusion if they were aged 18 to 55 years, admitted with ACS to the coronary care unit of participating hospitals, and able to provide informed consent.

After patients consented, they completed a self-administered questionnaire assessing sociodemographic data and medical history, including previous acute MI, stroke, coronary artery bypass grafting, and percutaneous coronary intervention, smoking, diabetes, and hypertension, as well as physical activity and health-related behaviours such as alcohol and drug consumption. These data were supported by laboratory tests, a baseline medical chart review conducted by the research nurse, and outcome data at one year.

Weight, height, blood pressure, and heart rate were measured by a research nurse within 24 hours of admission. Blood samples collected within the first 24 hours of hospital admission were immediately centrifuged and stored at -80°C for DNA extraction and biochemical tests.

Angiograms were performed according to standard practices and reviewed independently by two cardiologists to assess for the presence of thrombus formation, SCAD, FMD, myocardial bridging, coronary arterial spasm, or Takotsubo cardiomyopathy.[5]

WES was performed on five SCAD patients with an age of onset <51 and no obesity, dyslipidemia, diabetes, current smoking, or hypertension. A summary of the sample selection workflow is found in Supplementary Figure 1. Clinical characteristics of the selected patients are provided in Table 1.

Sequencing, Alignment, and Variant Calling

DNA was extracted from buffy coats using the FlexiGene DNA Kit (Qiagen, Venlo, Netherlands) and quantification was performed with PicoGreen™ (ThermoFisher, Waltham, Ma, U.S.) according to the manufacturers' instructions. Exomes were captured using the Agilent SureSelect Low Throughput library preparation and Exome 50mb capture. The Illumina TruSeq DNA v1 kit was used to multiplex the DNA. The DNA was sequenced at 50X using the Illumina HiSeq paired ends 100bp system with three samples in each sequencing lane. Reads were aligned to GRCh37 using bwa, quality controlled using GATK, and called using samtools mpileup with standard parameters resulting in a total of 349,363 variants. The average read depth was 46, with 58.5% of variants having a read depth >9. Analysis was restricted to variants with >9 reads. The resulting VCF files were analyzed by Exomiser.

Exomiser

To rank the variants according to functional consequence and phenotypic relevance, we used the latest publicly available Exomiser[6] release (version 12.1.0 Oct 1, 2019) and phenotype database version 1909 (Oct 30, 2019) from the repository at https://data.monarchinitiative.org/exomiser/latest/. As described by Smedley *et al*[6], Exomiser generates a variant-based score using population allele frequencies and computational predictions of pathogenicity and a gene-based score using a variety of phenotypic data reported in the Online Mendelian Inheritance of Man (OMIM), Human Phenotype Ontology (HPO)[7], Mouse Genome Database (MGD), and Zebrafish Model Organism (ZFIN) databases as well as protein-protein interaction data from STRING[8]. The variant-based and gene-based scores are combined by logistic regression to generate an overall score which is used to rank each variant

based on both the predicted consequence of the variant and how closely related the gene is to the list of phenotypes.

While Exomiser is designed to operate using only specific phenotypes of the patients, the sparsity of information about SCAD in the HPO and OMIM databases limits Exomiser's ability to determine the closest related phenotypes. However, Exomiser ranks the causal variant as the #1 variant in >75% of imprecise phenotyping scenarios.[9] To extend the possible overlapping phenotypes, we designed our list to utilize information from other known CTDs and genetic dissection conditions. We selected the hiPHIVE algorithm, which has the best performance in cases of imprecise phenotyping compared to similar software.[9] We used the HPO terms "Coronary artery dissection" (HP:0006702), "Arterial dissection" (HP:0005294), "Arterial tortuosity" (HP:0005116), "Aortic dissection" (HP:0002647), "Thoracic aortic aneurysm" (HP:0012727), and "Joint hypermobility" (HP:0001382). The heterozygous and homozygous allele frequency thresholds were set to $1x10^{-4}$ (0.01%) and .01 (1%) respectively.

We evaluated the five top ranked variants to identify likely candidate variants, which were verified with Sanger sequencing.

Results

The Exomiser score incorporates functional effects, protein domains, biochemical pathways, and associated conditions in the HPO database. Based on the strength of Exomiser's closest matching phenotype terms, we identified three top ranked variants in *FBN1*, *LEMD3*, and *ADAR* that were related to vasculopathies and thus highly plausible for SCAD etiology (Table 2). The top five variants for each patient are reported in Table S1 (Appendix 1).

Discussion

Our study identified three highly deleterious candidate mutations in genes that are all related to CTDs or vasculopathies.

FBN1

A heterozygous variant in the *FBN1* gene (p.N156S), which encodes the extracellular matrix protein fibrillin-1, was identified in a female with no cardiovascular risk factors who had an ACS at age 32. Premature stop codons and cysteine substitutions in *FBN1* are frequently the cause of Marfan syndrome, which can lead to arterial or aortic dissections through increased TGF-β signaling and matrix deposition.[10] It is possible that the N156S variant could cause a less severe, undiagnosed form of Marfan where SCAD would be the first clinical presentation. This has been previously reported by von Hundelshausen *et al*, who identified an *FBN1* missense variant in a SCAD patient without any physical characteristics of Marfan and normal TGF-β levels.[10] An alternative disease mechanism was identified that implicates elevated smooth muscle progenitor cell migration.

The N156 residue is within a highly conserved domain that is required for inter-domain interactions between adjacently located EGF-like domains in other regions of the protein.[11] NMR studies also indicate that N156 likely interacts with an aromatic molecule, further supporting a functional role.[12] Variants of N156 are extremely rare, present in only 9/125,669 individuals in the gnomAD database (which contains SCAD cases). N156S and N156H have been reported as variant of uncertain significance (VUS) in ClinVar for cardiovascular phenotypes.

LEMD3

A novel heterozygous variant (p.S728P) in the C-terminal domain of the nuclear membrane protein Man-1 (*LEMD3* gene), was identified in a male without cardiovascular risk factors who had an ACS at age 44. Man-1 is a negative regulator of the TGF-β pathway[13] and homozygous knock outs of *Lemd3* in mice leads to elevated TGF-β signaling, and disorganized vascular development including malformed aorta in developing embryos.[14] Mutations in TGF-β repressors lead to CTD Loeys-Dietz syndrome, with some patients having SCAD or arterial dissections and aneurysms. Heterozygous truncating mutations in the C-terminal domain of Man-1 are classified as pathogenic in ClinVar for the CTD Buschke-Ollendorff syndrome (BOS), and result in collagen fiber nevi in the skin and bone abnormalities.[15] In ClinVar, missense variants are classified as VUS for BOS, and one missense variant in the C-terminal domain (T879S) is classified pathogenic for cerebral arteriovenous malformation, suggesting a role in vasculopathy. The S728P variant occurs in a linker region in a DNA-binding domain, which is highly sensitive to structural changes.[16] We hypothesize that S728P could cause a phenotype similar to mild BOS in this patient, with the first presentation being SCAD.

ADAR

A heterozygous variant (p.K974E) in the RNA-editase domain of RNA adenosine deaminase 1 (*ADAR*), was identified in a male without risk factors who had an ACS (due to SCAD) at age 45. A comparison of the editase domains of ADAR2 and ADAR1 indicates that K974 is likely in a 17-residue RNA binding loop that interacts with the RNA phosphodiester backbone.[17] Mutations in *ADAR* can cause Aicardi-Goutières syndrome (AGS), a recessive neurodegenerative disease with pediatric onset and Dyschromatosis Symmetrica Hereditaria (DSH), a dominant, benign

skin pigmentation abnormality. K974E has been reported in ClinVar as a VUS for DSH. Importantly, some cases of AGS present with cerebral arteriopathy.[18]

In the two other patients Exomiser produced lower phenotype similarity scores, and we did not identify strongly deleterious mutations in genes linked to any CTDs (Table S1).

We have provided evidence that variants in CTD genes may contribute to SCAD. This is consistent with the work of Kaadan *et al.*[2] and von Hundelshausen *et al.*[10] demonstrating that SCAD may represent the first obvious clinical presentation of a mild form of a CTD. As such, VUS in CTD genes should be given extra scrutiny in SCAD patients.

Our work has several limitations. First, our bioinformatic analysis does not confirm causality. We did not have access to genetic data from family members, which is often required for an ACMG "pathogenic" rating for novel variants. However, cases of SCAD seen in the clinic may also not have family members available to aid in a genetic diagnosis. Second, while WES is cost effective, exon capture technology has regions of poor coverage and thus a higher rate of false negatives in certain genes and does not allow for analysis outside of exons compared to Whole Genome Sequencing. In addition, large-scale deletions and duplications (Copy Number Variations and Structural Variations) cannot be detected. Third, the GENESIS-PRAXY study does not include information on previous diagnoses of FMD and CTDs, or family history specifically for SCAD and related conditions.

Conclusions

We have demonstrated that WES in conjunction with Exomiser can discover potential new genetic etiologies for SCAD. Our results demonstrate that possible monogenic etiology should be considered for SCAD patients, especially for early-onset cases without risk factors. Additional functional work is necessary to further validate variants and their specific disease mechanisms.

Tables

Table 1 – Study Subject Characteristics

	101-0069	401-0050	503-0003	505-0033	701-0002
Sex	F	M	F	F	M
Ethnicity	European	European	European	European	European
Age	32	44	42	50	45
Coronary hypertortuosity	Yes	Yes	Yes	Yes	Yes
MI type	NSTEMI	STEMI	STEMI	NSTEMI	NSTEMI
N CAD risk factors	0/5	0/5	0/5	1/5 (former smoker)	0/5
BMI	28.08	23.84	23.77	24.56	29.01
Systolic BP (at admission)	128	108	105	170	143
LDL-C (mmol/L)	3.78	3.20	2.80	3.02	2.77
MI/stroke family history	Unknown	Father stroke (age<54)	No	No	No

CAD risk factors, determined by questionnaire, chart review, or medication, are defined as obesity (BMI>30 at admission), dyslipidemia, diabetes, smoking, and hypertension. All clinical measurements were taken post-MI. Family history of MI or stroke were determined based on patient questionnaires.

Table 2 – Top Ranked Exomiser Variant for Three Patients

Patient	Variant (g.)	Variant (p.)	Gene	gnomAD popmax	Pheno Score	ClinVar	OMIM	ACMG	Rank
101-0069	15:48888551 T>C	N156S	*FBN1*	6.15x10-4	0.794	VUS	Marfan	VUS	1
701-0002	1:154560700 T>C	K974E	*ADAR*	NP	0.615	VUS	Aicardi-Goutières	VUS	1
401-0050	12:65634744 T>C	S728P	*LEMD3*	NP	0.571	NP	Buschke-Ollendorf	VUS	1

Variant (g.) – coordinates of the variant in GRCh37, *Variant (p.)* – amino acid consequence of the variant according to the canonical transcript, *Rank* – rank variant was assigned in Exomiser software, *gnomAD popmax* – the maximum population frequency of the variant in gnomAD, *Pheno Score* – the phenotype similarity score assigned by Exomiser (maximum 1 indicates perfect match of input phenotypes to known disease gene), *ClinVar* – previously reported consequence of the variant in the ClinVar database, *OMIM* – disease associated with the gene in the OMIM database, *ACMG* – assessment of variant pathogenicity according to ACMG guidelines. NP = Not present.

References

1. Saw J, Mancini GBJ, Humphries, KH. Contemporary Review on Spontaneous Coronary Artery Dissection. *JACC*. 2016;68(3):297-312.
2. Kaadan MI, MacDonald C, Ponzini F, et al. Prospective Cardiovascular Genetics Evaluation in Spontaneous Coronary Artery Dissection. *Circ Genom Precis Med*. 2018;11(4):e001933.
3. Krittanawong C, Kumar A, Johnson KW, et al. Conditions and Factors Associated with Spontaneous Coronary Artery Dissection (From a National Population-Based Cohort Study). *Am J Cardiol*. 2018;123(2):249–253.
4. Pilote L, Karp I. GENESIS-PRAXY (GENdEr and Sex determInantS of cardiovascular disease: From bench to beyond-Premature Acute Coronary SYndrome). *Am Heart J*. 2012;163(5):741-746.
5. Raparelli, V, Elharram, M, Shimony, A et al. Myocardial Infarction With No Obstructive Coronary Artery Disease: Angiographic and Clinical Insights in Patients With Premature Presentation. *Can J Cardiol.* 2018;34(4):468-476.
6. Smedley D, Jacobsen JOB, Jäger, M et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc*. 2015;10(12), 2004–2015.
7. Köhler S, Carmody L, Vasilevsky N, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res*. 2018;47(D1):D1018-27.
8. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-D13.
9. Smedley D, & Robinson PN. Phenotype-driven strategies for exome prioritization of human Mendelian disease genes. *Genome Med*. 2015;7(81), 1–11.
10. von Hundelshausen P, Oexle K, Bidzhekov K, et al. Recurrent spontaneous coronary dissections in a patient with a de novo fibrillin-1 mutation without Marfan syndrome. *Thromb Haemost*. 2014;113(03):668–670.
11. Yadin DA, Robertson IB, McNaught-Davis J, et al. Structure of the Fibrillin-1 N-Terminal Domains Suggests that Heparan Sulfate Regulates the Early Stages of Microfibril Assembly. *Structure*. 2013;21(10):1743–1756.
12. Robertson, IB, Osuch, I, Yadin, DA, et al. 1 H, 13 C and 15 N resonance assignments for the fibrillin-1 EGF2-EGF3-hybrid1-cbEGF1 four-domain fragment. *Biomol NMR Assign*. 2014;8(1):189–194.
13. Lin F, Morrison JM, Wu W, Worman HJ. MAN1, an integral protein of the inner nuclear membrane, binds Smad2 and Smad3 and antagonizes transforming growth factor-β signaling. *Hum Mol Genet*. 2004;14(3):437-45.
14. Cohen TV, Kosti O, Stewart CL. The nuclear envelope protein MAN1 regulates TGFbeta signaling and vasculogenesis in the embryonic yolk sac. Development. 2007;134(7):1385-95.
15. Hellemans J, Preobrazhenska O, Willaert A, et al. Loss-of-function mutations in LEMD3 result in osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. *Nat Genet*. 2004;36(11):1213.
16. Caputo S, Duband-goulet I, Konde E, et al. The Carboxyl-terminal Nucleoplasmic Region of MAN1 Exhibits a DNA Binding Winged Helix Domain. *J Bio Chem*. 2006;281(26):18208–18215.

17. Fisher AJ, Beal PA. Effects of Aicardi-Goutieres syndrome mutations predicted from ADAR-RNA structures. *RNA Biol*. 2017;14(2):164-70.
18. Ramesh V, Bernardi B, Stafa A, Garone C, et al. Intracerebral large artery disease in Aicardi-Goutières syndrome implicates SAMHD1 in vascular homeostasis. *Dev Med Child Neurol*. 2010;52(8):725-32.

Supplementary Material

Figure 1

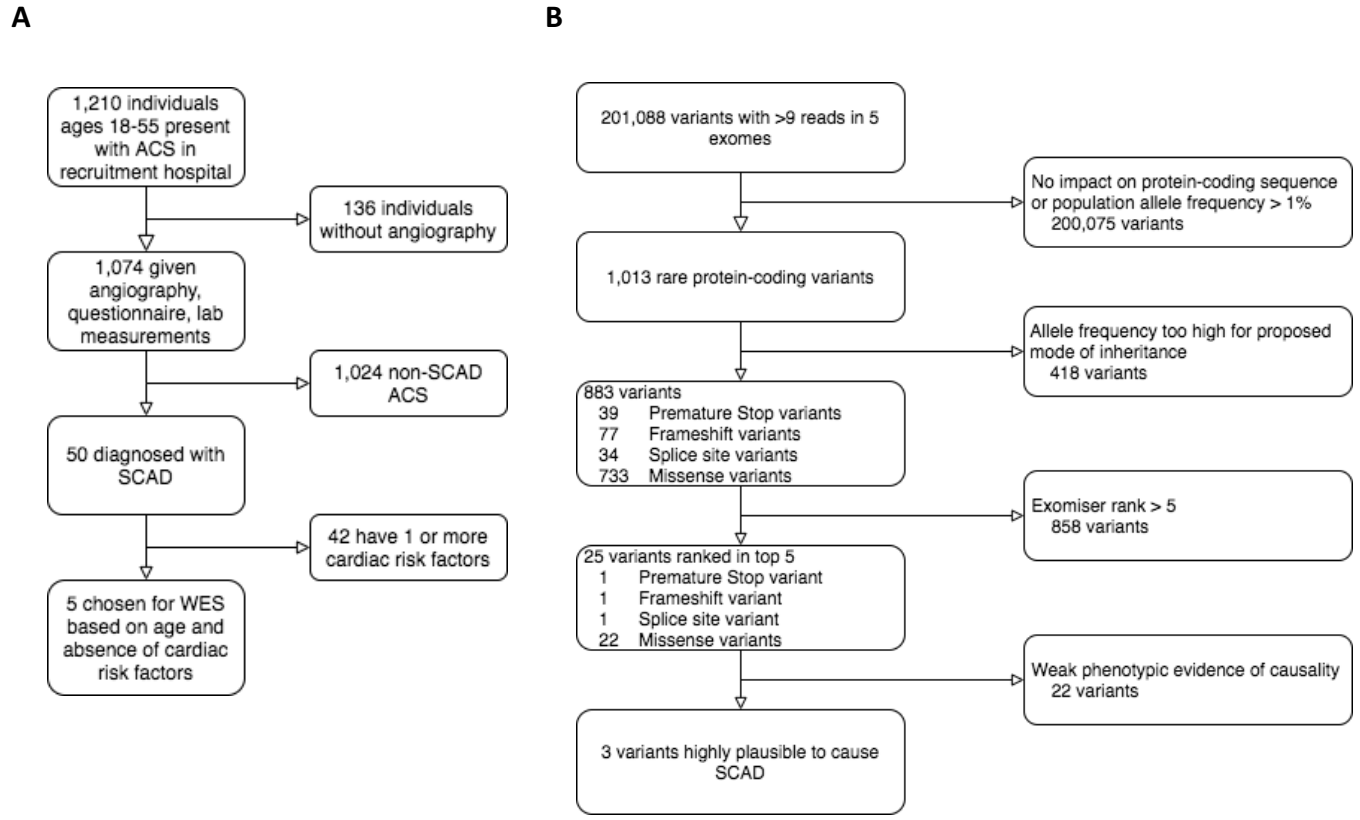


Figure Legend

A. Selection of samples for Whole Exome Sequencing.

B. Flowchart describing how Exomiser filtered variants obtained from Whole Exome Sequencing.

Table S1 – See Appendix 1

Chapter 3: Transition

Whole Exome Sequencing (WES) was able to identify a genetic cause in three out of five patients included in the pilot analysis. WES is extremely cost effective; however, it suffers from important limitations. First, the exon capture step of WES creates non-uniform coverage of the exome. In particular the first exon of genes tends to have lower coverage, making it more likely that a relevant variant in this region will be missed. For example, many proteins have signal peptides required for proper localization that are coded in exon one, or cleavage sites that must be recognized to convert the protein to an active form.

Another limitation of WES is that it cannot be used to call large-scale duplications and deletions, also known as Copy Number Variants (CNV). CNV calling methods often rely on analyzing read depth to determine how many copies of the gene exist. WES read depth is often highly variable depending on the genomic region, and the short length of exons does not provide enough context to reliably separate a region of increased or reduced depth from background.

Finally, WES does not capture non-coding and regulatory variation. Non-coding variants can change the affinity of transcription factor binding, thus altering transcription activation or repression.[39] 5' UTRs may contain a number of elements controlling translation initiation, including secondary structures, upstream open reading frames (uORFs), and internal ribosome entry sites (IRES).[39] 3' UTRs also have an important role in determining mRNA localization, mRNA stability through control of poly-adenylation signals, miRNA binding, and AU-rich elements, and translational repression through miRNA binding.[39] While 5' and 3' UTRs are considered exonic, and thus included in some WES capture platforms, recent analysis suggests that popular capture platforms only capture the UTR for approximately 20% of genes.[10] In the context of SCAD, the predisposing conditions of Loeys-Dietz Syndrome (LDS) and Marfan

Syndrome (MFS) cause elevated TGF-β signaling, which results in excess collagen deposition around the arteries leading to increased arterial tortuosity and fragility that can result in SCAD or another arterial dissection.[40] Reduced expression levels of collagen can also affect vascular architecture and integrity, as seen in Ehlers-Danlos Syndrome (EDS) patients with truncation mutations leading to nonsense mediated decay.[41] Therefore, regulatory mutations in UTRs or promoters that increase or reduce expression of TGF-β pathway or extracellular matrix genes are plausible mechanisms of SCAD that warrant further investigation. Outside of variants that are predicted to affect splicing, rare regulatory variants have not, to my knowledge, been examined in the context of a vasculopathy, and it is unknown how often such variants may contribute to the etiology of SCAD.

Chapter 4

Identifying Genetic Determinants of Spontaneous Coronary Artery Dissection with Genome Sequencing

Hannah A Burr, BSc[1,2]; Louise Pilote, MD, MPH, PhD[3]; Asim N Cheema, MD, PhD, MSc[4]; Line Dufresne, MSc[2]; Katia Desbiens, MSc[2]; Hans Markus Münter, PhD[6]; Mark Lathrop, PhD[6,7]; George Thanassoulis*, MD[1,2,5]; James C Engert, PhD* [1,2,5,7]

*co-senior authors

[1]Division of Experimental Medicine, McGill University, Montreal, Canada

[2]Preventative Genomic Cardiology, McGill University Health Centre and Research Institute, Montreal, Canada

[3]Division of General Internal Medicine, McGill University, Montreal, Canada

[4]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada

[5]Division of Cardiology, Department of Medicine, McGill University, Montreal, Canada

[6]Genome Quebec Innovation Centre, McGill University, Montreal, Canada

[7]Department of Human Genetics, McGill University, Montreal, Canada

Corresponding Authors:

James C. Engert, PhD
McGill University Health Centre
1001 Decarie Blvd., Room EM1.2218
Montreal, Quebec H4A 3J1
Canada
Phone: 1-514-934-1934 ext. 35325
Email: jamie.engert@mcgill.ca

George Thanassoulis, MD, MSc
McGill University Health Centre
1001 Decarie Blvd., Room D05.5120
Montreal, Quebec H4A 3J1
Canada
Phone: 1-514-934-1934 ext. 35465
Email: george.thanassoulis@mcgill.ca

Abstract

Purpose: Spontaneous coronary artery dissection (SCAD) is a significant cause of early-onset acute coronary syndrome (ACS) in young, often female patients with few traditional cardiovascular risk factors. SCAD is genetically heterogeneous and has low familial penetrance, making it difficult to identify clinically actionable variants in patients.

Methods: We applied genome sequencing to 45 angiographically confirmed SCAD patients in the GENESIS-PRAXY cohort. We used an analysis approach that considered candidate genes as well as and genes not previously associated with SCAD.

Results: Clinically actionable pathogenic and likely pathogenic variants were identified in 3 patients (6.7%) while Variants of Uncertain Significance were identified in an additional 28 (62.2%) patients. 14 patients (31%) did not have a rare genetic finding. Our results suggest that patients with an identified rare genetic variant are at increased risk of SCAD recurrence and that SCAD may have an oligogenic etiology in some patients.

Conclusion: This study replicated the finding that variants in connective tissue genes are associated with SCAD but also implicates nine genes not previously associated with SCAD. We suggest modifications to ACMG guidelines to improve the detection of clinically actionable variants in SCAD that may be applicable to other diseases with overlapping genetic etiologies.

Keywords

Spontaneous coronary artery dissection; connective tissue disorders

Introduction

Spontaneous Coronary Artery Dissection (SCAD) is a rare presentation of Acute Coronary Syndrome (ACS) that is increasingly recognized as a significant cause of ACS in younger patients without traditional cardiovascular risk factors. Recent epidemiologic studies indicate that SCAD has multiple etiologies, including connective tissue disorders (CTDs), polycystic kidney disease, fibromuscular dysplasia (FMD), vasculopathies, as well as inflammatory diseases such as systemic lupus erythematosus (SLE).[1,2] Many of these syndromes have known genetic determinants, however, previous genetic studies of SCAD, focused on identifying variants classified as pathogenic by the ACMG/AMP guidelines, [3] frequently failed to yield a genetic diagnosis.[4,5]

One explanation for the low yield is that common genetic variants also contribute to SCAD. Indeed, recent genome wide association (GWAS) and gene burden studies have identified a number of common variants and novel genes, including the *PHACTR1/EDN1*,[6] *TSR1*,[7] *TLN1*,[8] *PTGIR*,[9] *ADAMTSL4, LRP1,* and *LINC00310*[10] loci. Another explanation could be that some SCAD cases are caused by rare regulatory variants, which to our knowledge no studies have analyzed to date. Our previous work (see **Chapter 2**) identified pathogenic variants in the exomes of early-onset SCAD patients most likely to have a genetic cause due to the absence of traditional cardiovascular risk factors. A recent study of the related vasculopathy familial thoracic aortic aneurysm (TAAD) indicates that non-syndromic TAAD has a correlated genetic and phenotypic spectrum[11] in which ACMG/AMP pathogenic and likely pathogenic variants are associated with a more severe phenotype (increased familial penetrance, recurrent dissections, and interventional surgery). We hypothesize that SCAD will follow a similar pattern of individuals with causative variants having a more severe phenotype.

To thoroughly examine the genetic etiology of SCAD, we applied genome sequencing to angiographically confirmed SCAD patients. Using whole genome sequencing data, we can analyze both coding and non-coding variants for evidence of pathogenicity in genes that would be part of a traditional gene panel, as well as identify novel genes. To efficiently parse the thousands of rare variants captured by genome sequencing, including non-coding variants, which are not routinely included, and improve specificity over existing methods, we applied an approach that considered both the consequence of each variant and the phenotypic evidence for each gene.

Methods

GENESIS-PRAXY Cohort

The GENESIS-PRAXY study (Gender and Sex Determinants of Cardiovascular Disease: From Bench to Beyond Premature Acute Coronary Syndrome) has been previously described in detail.[12,13] Briefly, patients aged 18 to 55 years old admitted with an ACS to 24 participating hospitals were enrolled. If clinically indicated, patients underwent coronary angiography at the time of presentation. Angiograms were reviewed independently by two cardiologists to assess for SCAD and were later assessed for coronary artery hyper-tortuosity consistent with fibromuscular dysplasia (FMD). Patients completed a detailed questionnaire regarding personal and family medical history, medications, and health habits. These data were augmented by a chart review conducted by a research nurse who also obtained additional data such as height, weight, and blood pressure within 24 hours of admission. A blood sample for biochemical tests and DNA extraction was collected within 24 hours of admission, centrifuged, and stored at -80°C.

Cardiac risk factors were defined as obesity (BMI $\geqslant$ 30), dyslipidemia (based on a self-report using a standardized questionnaire, chart review, or use of relevant medication), diabetes (based on questionnaire, chart review, or use of relevant medication), smoking (based on questionnaire), and hypertension (based on questionnaire, chart review, or use of relevant medication). Family risk of early MI was defined as a male parent or sibling with an MI before age 45 or a female parent or sibling with an MI before age 55 (both based on questionnaire). Recurrence was defined as a previous MI (based on questionnaire or chart review) or re-hospitalization for a cardiac event within the 12-month follow-up period.

Whole Genome Sequencing and Annotation

Full sequencing, SNV calling, quality control (QC), and annotation methods can be found in the Supplemental Methods. Briefly, we performed paired-end sequencing at 30x coverage and QC and joint calling according to GATK best practices.[14] Variants were annotated using WGSA[15] and coding, UTR, and promoter variants were retained. The Exomiser hi-PHIVE algorithm[16] was used to generate gene-phenotype scores, which are a calculation of the similarity of phenotypes associated with each gene in the Human Phenotype Ontology (HPO)[17] and SCAD or similar vasculopathies. Further QC measures included read depth filters and allele frequency thresholds based on the prevalence of SCAD. After QC and allele frequency pruning, 25,081 unique variants were included in the analysis.

Predicting Likelihood of Pathogenicity

Dozens of prediction scores have been trained to distinguish benign and pathogenic variation on the basis of evolutionary conservation, protein structure, functional effect, or

combinations of these data.[18] Unfortunately, these scores tend to have a high false positive rate and no uniform benchmarking methodology or dataset exists to compare the predictive values of these scores.[19] Pathogenic predictions from multiple algorithms are used as "supporting" strength criteria in the ACMG guidelines.

Recent work by Ghosh *et al.* demonstrated that the choice of prediction algorithms affects the rates of concordance, true positives, and false positives detected in ClinVar.[20] They observed that the popular combination of SIFT,[21] PolyPhen,[22] and CADD[23] had a high rate of false concordance, with all three algorithms assigning 22.5% of benign reports in ClinVar a pathogenic rating. They further demonstrated that VEST3,[24] REVEL,[25] and MetaSVM[19] had the optimal performance out of all possible combinations of three algorithms. This combination had only a 2.8% false positive rate, retained 70% of true pathogenic variants, and assigned only 5.4% of pathogenic variants a benign rating (false negative). The algorithms were not concordant for the remaining 21.8% of the dataset. The study also revealed that the software M-CAP[26] had superior sensitivity for pathogenic variants, however, it had a higher false positive rate than the other three algorithms. A recent publication by Lassmann *et al.* indicates that the performance of scoring algorithms varies by phenotype.[27] M-CAP and MetaSVM demonstrate a strong performance for cardiovascular phenotypes and VEST3 shows a moderate performance. Based on the need to balance sensitivity and specificity, we selected VEST4, REVEL, MetaSVM, and M-CAP prediction scores to filter missense variants.

Variants were excluded if more than two scores were missing. To standardize the variant scores, we used the WGSA rank score, which is the percentile of a variant in the distribution of all scores assigned to all possible missense variants in the genome. We used the median of all available rank scores (median rank score) to assign one score per variant. See Supplemental

Results and Figure S2 for a comparison of performance between our method and the popular method (SIFT/PolyPhen/CADD) in our dataset. Additionally, information from MutPred2[28] was used to develop hypotheses about functional effects.

To identify high-consequence promoter, UTR, and "disruptive" (frameshift, splicing, and in-frame insertion/deletion) coding variation, we used CADD because it can score novel insertion/deletion variants and thus had the lowest missingness. We calculated a percentile of the CADD score to use as a threshold for each class of variants (Table S1). For splicing, UTR, and promoter variants we applied the median LINSIGHT score reported by Huang *et al.* as a threshold [29] in addition to the CADD threshold. Variants with experimental evidence of miRNA binding from the TargetScan database [30] and binding evidence from RegulomeDB (score 1a-f, 2a-c, 3a-b) [31] were also included, regardless of CADD and LINSIGHT score.

Variant Prioritization

All prioritization was performed using custom scripts in R. **Figure 1** describes the workflow. In general, we varied the percentile thresholds of *in silico* prediction scores based on the likelihood of the gene to cause SCAD. Variants that had been previously reported as pathogenic or likely pathogenic in ClinVar were subject to the most liberal set of thresholds to capture conditions that may pre-dispose an individual to SCAD. Variants in "candidate genes" that had previously been associated with SCAD (defined as 125 genes associated with "dissection", "aneurysm" and "arterial tortuosity" in the HPO database) were included if they were above the 50th percentile of most damaging predicted variants for a given class of variation (Table S1).

If the gene had not been previously associated with SCAD or a related condition ("non-candidate genes"), we required that the gene have a strong phenotypic relationship to SCAD (defined as an Exomiser gene-phenotype score $\geqslant$ 0.5017, which retained approximately 10% of genes in the genome (2,229 genes)). In addition, coding variants were included if they were above the 80th percentile of most damaging predicted variants for a given class of variation (Table S1). Non-coding variants in non-candidate genes had the strictest set of thresholds, requiring an Exomiser gene-phenotype score in the 95th percentile (score $\geqslant$ 0.504) and a CADD score $\geqslant$ 20.

This approach resulted in a list of 231 variants (1 to 11 variants per individual) before adding potential compound heterozygous variants. The full list of variants identified with this approach is found in **Appendix 2**.

In a final step, we manually reviewed each of the 231 identified variants to determine if it was a likely cause of SCAD. General exclusion criteria used in this step can be found in the Supplemental Methods. After manual curation, 41 high-confidence variants in 31 patients remained. The InterVar software[32] was used to estimate the ACMG status of coding variants and the status was confirmed manually.

Results

Clinical Features of SCAD Patients in the GENESIS-PRAXY Cohort

The prevalence of SCAD in GENESIS-PRAXY is 4%, consistent with estimates from other early-onset ACS cohorts.[33,2] **Table 1** provides a summary of relevant clinical characteristics for non-SCAD GENESIS-PRAXY ACS patients in comparison to the 45 sequenced individuals with SCAD. SCAD patients were more likely to be female ($p<0.001$),

younger at the time of ACS (mean age 46, p=0.008), and have significantly lower rates of smoking (p=0.021). Other cardiovascular risk factors (hypertension, diabetes, dyslipidemia, and BMI) were lower in SCAD cases compared to non-SCAD ACS cases, however, these differences were not statistically significant (**Table 1**).

SCAD patients in GENESIS-PRAXY had a high rate of coronary artery hyper-tortuosity (95%), consistent with other SCAD studies documenting rates of FMD up to 86%.[2] In addition, 24.4% of SCAD patients had a potential recurrence. While the date of previous MI and thus mean follow up time is unknown, this result is consistent with other studies of SCAD, which estimate a recurrence rate of 10.4% within 3 years, and up to 30% within 10 years.[34] Therefore, we believe these events are highly likely to represent recurrent SCAD, especially given the low cardiovascular risk profile of these patients.

Coding Variants in Candidate Genes

Our methodology identified 12 coding variants (1 pathogenic, 1 likely pathogenic, 10 VUS) in 12 individuals in 11 candidate genes **(Table 2)**.

The identified pathogenic variant is a novel heterozygous frameshift variant in the *SLC2A10* gene leading to premature truncation 83 amino acids later. Homozygous truncations of the *SLC2A10* gene have been reported as pathogenic for arterial tortuosity syndrome in ClinVar, which often has a childhood onset and causes a phenotype similar to Loeys-Dietz syndrome including aortic aneurysm.[35] Interestingly, vascular abnormalities in large vessels have not been previously reported in heterozygous carriers,[35] however, the individual harboring this variant displayed coronary artery hyper-tortuosity consistent with the phenotype. The individual also had

a variant in the *ITGB4* gene (see Table S3), and thus the SCAD and observed arterial tortuosity may be the result of digenic inheritance.

The likely pathogenic variant that we identified is in the *MYH11* gene, which encodes a myosin heavy chain protein predominantly expressed in smooth muscle. Missense variants in this gene cause familial TAAD in an autosomal dominant fashion.[36] VUS were found in *NOTCH1, MYLK, FBN1, FBN2, TGFBR1, SMAD3, COL3A1, COL5A2,* and *COL1A1*. Two of these variants have been previously reported in ClinVar as VUS for TAAD; rs863223613 in *FBN2* (SCV000250309.11) and rs770798158 in *SMAD3* (SCV001355004.1).

Non-Coding Variants in Candidate Genes

A total of eight heterozygous non-coding variants in six individuals were identified in the candidate genes *MYLK, LMX1B, LOX, LRP1, COL5A1, TGFBR2,* and *TGFB2* (**Table 3**). These variants represent a potential genetic cause for 13.3% of our cohort and would have been missed using exome sequencing. The ACMG/AMP have not issued formal recommendations regarding the pathogenicity of non-coding variants; however, we can extend the concepts of allele frequency and computational predictions of deleteriousness to assign a VUS classification. Detailed proposed molecular mechanisms and a discussion of the plausibility of regulatory variation in these genes causing a phenotype can be found in the Supplemental Results. Potential compound heterozygous results for *MLYK* and *TGFBR2* are summarized in Table S2.

Variants in Non-Candidate Genes

In addition to the 20 coding and non-coding variants identified in candidate genes, 21 variants of interest were discovered in non-candidate genes. 1 variant is pathogenic and 20 are

VUS. Most of these genes are collagens (5 genes), TGF-β pathway proteins (3 genes), integrins (3 genes), or genes related to inflammation (3 genes). These variants are summarized in **Table 4**. No non-coding variants in non-candidate genes remained after manual curation. For possible biochemical mechanisms of these variants as well as additional clinical manifestations of the associated conditions, see the Supplemental Results.

Collagens Five patients have VUS heterozygous missense variants in four collagen genes that have not yet been associated with SCAD, but that have been associated with other CTDs: *COL7A1, COL4A5, COL12A1,* and *COL27A1*. These collagens are all expressed in coronary artery tissue according to GTEx (https://gtexportal.org/home/). Recent work indicates that SCAD patients have an increased burden of disruptive variants across all collagen genes compared to controls. [37] The *COL7A1* variant has been reported as pathogenic in ClinVar for recessive dystrophic epidermolysis bullosa (SCV000490486.1), in which approximately 20% of patients show aortic dilation. [38] Variants in *COL4A5* cause Alport syndrome, which case reports have previously linked to SCAD. [39,40] *COL12A1* causes muscular Ehlers-Danlos syndrome and *COL27A1* causes Steel syndrome, neither of which have been characterized for vasculopathies or cardiovascular risk.

TGF-β Pathway The TGF-β pathway is one of the most well studied pathways in arterial dissection disorders. Variants leading to Marfan syndrome and Loeys-Dietz syndrome elevate TGF-β activation and signaling, causing an excess of ECM deposition in blood vessels.[41] Our work has identified three patients with VUS heterozygous missense variants in TGF-β pathway genes that have not been associated with SCAD: *LTBP4, GDF5,* and *LEMD3*. Variants in these

genes have been associated with cutis laxa, brachydactyly, and Buschke-Ollendorf syndrome, respectively.

Integrins Three patients have VUS heterozygous integrin gene variants that have not previously been associated with SCAD. These are in the genes *ITGB4, ITGB5,* and *ITGB8*. Integrins are cell-adhesion molecules that connect the ECM to the cytoskeleton and regulate TGF-β signaling. Recent work by Turley *et al* [8] revealed rare missense variants in the *TLN1* gene of SCAD patients, and a subsequent GWAS found signals in three other proteins in close network proximity [10]. The protein encoded by *TLN1* binds to certain integrin β subunits and increases affinity for ECM ligands [42] but also connects integrins to the actin cytoskeleton via an interaction with vinculin. Heterozygous missense variants in *ITGB4* have previously been linked to epidermolysis bullosa, [43] and the specific identified variant has been reported as a VUS in ClinVar for an unspecified phenotype (SCV000853811.1). *ITGB5* and *ITGB8* also play a role in TGF-β signaling by controlling the availability of the TGF-β1 ligand.[44] *ITGB5* was recently implicated as a risk locus for myocardial infarction in a GWAS study,[45] and *ITGB8* has been linked to a vasculopathy, brain arteriovenous malformation.[46]

Inflammatory Pathways Systemic inflammation has been identified as a contributor to SCAD.[1] We observed a heterozygous VUS in the *CECR1* gene, which has been reported as pathogenic in ClinVar for polyarteritis nodosa (SCV000994598.1) in a biallelic state, as well as VUS in the *BLK* gene associated the systemic lupus erythematosus and in the *ADAR* gene associated with Aicardi-Goutières Syndrome.

Other Pathways A heterozygous ACMG/AMP pathogenic premature stop variant was identified in the *KDM6A* gene, which encodes a lysine specific histone demethylase. This type of variation causes Kabuki syndrome, which has a wide range of phenotypes and is inherited in an X-linked fashion.[47] Females often exhibit less severe phenotypes [48] and thus may go undetected.[49] Additionally, there are case reports of cervical artery dissection in a patient with mild Kabuki syndrome [50] and childhood-onset aortic dilation with disrupted elastic fibers in other patients. [51] We believe this loss of function variant in a SCAD patient is consistent with these other Kabuki syndrome phenotypes.

Evidence of Oligogenicity

Four of the 45 individuals sequenced (8.9%) displayed evidence of oligogenicity, or variants in multiple genes that could contribute to SCAD etiology. The results are summarized in **Table S3**. In two of the individuals, the two affected genes are in different pathways. For two other patients the affected genes are in the same pathway; *TGFBR1* and *LEMD3* are both regulators of the TGF-β pathway, as are *SMAD3* and *LTBP4*. All four patients could be considered a more severe SCAD phenotype as they are under the mean age of 46 and three of the four had a previous ACS. This finding is consistent with the idea that SCAD exists on a phenotypic spectrum, with a higher genetic burden leading to a more severe phenotype.

Multivessel SCAD is Significantly Associated with No Rare Genetic Finding

Of the 14 patients in the cohort who did not have an identified pathogenic, likely pathogenic, or VUS variant, seven had a multivessel dissection accounting for 70% of multivessel dissection in our cohort. This enrichment is statistically significant (p=0.0053,

Fisher's exact test), suggesting that multivessel SCAD could have a different genetic etiology. Multivessel SCAD is also enriched in pregnancy associated SCAD,[52] which could represent another mechanism for these patients.

Discussion

In summary, we have described a method leveraging modern bioinformatic tools (VEST4, REVEL, M-CAP, MetaSVM, CADD, and LINSIGHT) to identify DNA variants with a high likelihood of altering protein function or expression. Our method relies on the software Exomiser to identify variants in genes not previously associated with SCAD but that cause related phenotypes or are in related pathways.

We identified 41 variants with a high likelihood of causality in 31 SCAD patients. The genes harboring these variants were part of previously known pathways such as collagens, TGF-β signaling, and inflammation, and replicated recent results implicating integrin pathways. Furthermore, our results are supported by previous case reports of SCAD and other vasculopathies associated with Alport and Kabuki syndromes. Our results also suggest that SCAD could be part of a clinical spectrum for epidermolysis bullosa, Buschke-Ollendorf syndrome, muscular Ehlers-Danlos syndrome, and Steel syndrome. We report that genetic variants in regulatory regions of *MYLK, LMX1B, TGFBR2, LOX, TGFB2, LRP1,* and *COL5A1* could lead to SCAD in some patients. We also report that SCAD could have an oligogenic etiology in some patients.

An important limitation of our study is the lack of deep phenotyping of the SCAD patients. The GENESIS-PRAXY protocol did not include an evaluation of symptoms consistent with a CTD, now recommended by the AHA.[53] In addition, biomarker assays (e.g. circulating

TGF-β levels) may have provided functional evidence to corroborate some VUS findings. Another limitation is the incomplete knowledge of genetic dissection and aneurysm conditions available in public databases such as ClinVar and OMIM, which could lead to a lower Exomiser score. However, the inclusion of many genes not previously associated with SCAD likely reduced our false negative rate compared to other genetic studies of SCAD.[4,5] In addition, the small size of our cohort limited our ability to validate common variants identified by GWAS.

Only three of the variants identified in our study are pathogenic or likely pathogenic according to the ACMG/AMP guidelines, a yield of 6.7% consistent with other genetic studies of SCAD.[4,5] Arnaud *et al.* demonstrated that monogenic pathogenic and likely pathogenic variations in TAAD correlate with increased penetrance, risk of recurrence, and surgical intervention.[11] Our dataset is too small to statistically replicate this result, however, we observe a similar trend. All three patients with pathogenic and likely pathogenic variants in had potential recurrence, and two out of three had a self-reported history of an immediately family member with an early-onset ACS. In addition, 28.6% of patients with a VUS had a possible recurrent event and all rare genetic findings (pathogenic, likely pathogenic, and VUS findings together explain 91% of recurrence in our cohort. This trend should be further investigated in larger studies to characterize if patients with rare genetic findings including pathogenic, likely pathogenic, and VUS could have an increased rate of recurrence, which would impact clinical management.

The ACMG/AMP pathogenicity guidelines rely heavily on familial segregation. In the context of a low penetrance and highly heterogeneous disease such as SCAD, determining segregation will require sequencing and thoroughly evaluating a large number of family members, which may not be available for every patient. Pathogenic or likely pathogenic

classifications are also not possible for non-coding variants unless regulatory variation is a well-established mechanism of disease in the gene or experimental validation of the variant is performed. Thus, many patients with SCAD will have a non-actionable VUS finding.

Our study identified 24 coding VUS across candidate and non-candidate genes. 14 of these 24 variants have a combination of ACMG/AMP evidence giving them an 81% probability of pathogenicity according to a Bayesian model of the guidelines, very close to the 90% probability required for a likely pathogenic rating.[54] Future studies of the functional implications of the VUS identified in this study, and a large genetic studies of SCAD patients using a gene burden methodology could further our understanding of the contribution of VUS in connective tissue genes to. In addition, more detailed phenotyping in a larger genetic study of patients could determine whether patients with a pathogenic, likely pathogenic, or VUS are more likely to have phenotypes consistent with a mild CTD that would likely be missed in a primary care setting. ACMG/AMP guidelines could ultimately be improved to better reflect the reality that some of the identified VUS may actually be causal and thus clinically actionable.

Criteria PP2, which dictates a "missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease", could be modified to apply to nine VUS (four in candidate genes, five in non-candidate genes). Most genes that were previously known or newly identified in this study are associated with CTDs. The classical forms of the disease are usually only caused by a loss-of-function variant or very specific set of missense mutations (e.g. cysteine in *FBN1*, glycine in collagens). Missense variants outside of these pre-defined domains are usually not considered pathogenic for a CTD without additional data, however, they could still be pathogenic for SCAD, a phenotypic variant of the full syndromic disease.

Criteria PP5, which dictates a "reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation" could also be modified to better suit these results. Due to the observation that VUS in sporadic, non-syndromic dissection conditions are common, we suggest that this criterion could reasonably be applied if the variant has been reported as a VUS for a dissection condition or CTD. In this case, an additional three VUS could be re-classified as likely pathogenic.

With these modifications, aggregate reporting of VUS findings in CTD genes could improve the identification of clinically actionable variants in the absence of familial information, an important goal of personalized medicine in the context of other low penetrance, highly heterogeneous diseases as well. Including the VUS reclassified through these modifications would increase the diagnostic yield of our study from 6.7% to 33.3%. Further study will be required to determine if different follow-up treatments may be more effective for individuals with variants in different pathways.

Figure Legends

Figure 1 – A flowchart outlining the approach and thresholds used for variants previously reported as pathogenic in ClinVar, candidate genes that had been previously linked to SCAD or a related vasculopathy, and non-candidate genes that had not been previously linked to SCAD.

Tables

Table 1 – Population Characteristics

Trait	**Non-SCAD**	**SCAD**	**p**
N (%)	1,074 (96)	45 (4)	-
Female (%)	341 (32)	38 (84)	<0.001
Age at admission [sd]	48.3 [5.8]	45.9 [5.9]	0.008
European (%)	825 (77)	34 (76)	0.86
Ever Smoked (%)	754 (70)	25 (56)	0.021
Dyslipidemia (%)	598 (56)	19 (42)	0.092
Hypertension (%)	521 (49)	18 (40)	0.29
Diabetes (%)	184 (17)	6 (13)	0.68
Obesity (%)	445 (41)	11 (24)	0.028
Family History (%)	231 (22)	6 (13)	0.81
Recurrence (%)	221 (21)	11 (24)	0.58
FMD (%)	n/a	43 (95)	n/a

Variable definitions are described in the methods section. Discrete variables are presented as N (%) and continuous variables are presented as mean [sd]. P values were obtained for discrete variables using a Fisher's exact test and for continuous variables using an independent t-test.

Table 2 – Coding Variants in Candidate Genes

Patient	Age	Sex	CVD Risk	Recurrent	MV	rsID	Gene	HGVSp	ACMG Criteria	CADD	Median Rank	gnomAD popmax
401.0082	45	F	5/6	Y	N	chr20:45354153 AC>A	*SLC2A10*	W162G fs*83	PVS1; PM2; PP3	16.6	n/a	0
101.0132	39	F	1/6	Y	N	rs769614526	*MYH11*	R1511Q	PM1; PM2; PP2; PP3	27.7	0.94	2.63e-5
101.0058	55	F	1/6	N	N	rs777423973	*NOTCH1*	R1946H	PM1; PM2; PP3	33	0.73	8.70e-5
101.0130	45	F	1/6	N	N	rs1052930526	*MYLK*	R179Q	PM1; PM2; PP3	26.8	0.54	2.66e-5
101.0133	41	F	3/6	N	N	rs752117227	*FBN1*	T968S	PM1; PM2; PP3	14.83	0.57	2.64e-5
201.0106	52	F	2/6	Y	N	chr15:48725182 A>G	*FBN1*	I2207T	PM1; PM2; PP3	24.8	0.82	0
401.0050	44	M	0/6	N	N	chr9:101867573 C>T	*TGFBR1*	P29L	PM2; PP3	18.88	0.54	0
401.0149	51	F	4/6	N	N	rs863223613	*FBN2*	G2317S	PM1; PM2; PP3	32	0.95	0
401.0158	49	F	2/6	N	Y	rs1190691997	*COL3A1*	N1230K	PM2; PP3	19.4	0.58	8.80e-6
407.0059	43	F	1/5	Y	N	rs770798158	*SMAD3*	A112T	PM1; PM2; PP3	27.1	0.92	3.27e-5
407.0146	38	F	3/6	Y	N	chr2:189927755 G>A	*COL5A2*	A635V	PM2; PP3	24.7	0.80	0
507.0027	41	F	3/6	Y	N	rs769106952	*COL1A1*	A327S	PM2; PP3	22.3	0.67	6.17e-5

Detailed description of coding variants identified in a candidate gene and the clinical characteristics of individuals harboring each variant.

Patient – patient ID#, *Age* – age at admission to hospital with ACS, *Sex* – biological sex, *CVD Risk* – number of cardiovascular risk factors (defined as obesity, dyslipidemia, smoking, hypertension, diabetes, and where available, family history of early MI), *Recurrent* – patient had a previous MI or a cardiac rehospitalization within a one year follow up period, *MV (multivessel)* – angiogram revealed multiple vessels were dissected, *rsID* – rs number of the variant or GRCh37 coordinates and allele if variant is novel, *Gene* - affected gene, *HGVSp* – amino acid change according to the VEP canonical transcript, *ACMG Criteria* – summary of lines of evidence used to determine ACMG status, *CADD* – the CADD phred score (see table S1 for 50th percentile thresholds), *Median Rank* – the median rank score of four algorithms selected to determine pathogenicity of missense variants (minimum 0.5), *gnomAD popmax* – the maximum allele frequency of all populations in the gnomAD database

Table 3 – Non-coding variants in candidate genes

Patient	**Age**	**Sex**	**CVD Risk**	**Recurrent**	**MV**	**rsID**	**Gene**	**Variant Type**	**ACMG Criteria**	**CADD**	**LINSIGHT**	**gnomAD popmax**
101.0130	45	F	1/6	N	N	rs1040685992	*MYLK*	5' UTR	PM2; PP3	15.13	0.167	2.11e-4
401.0021	46	M	5/6	Y	N	rs571597357	*TGFB2*	3' UTR	PM2; PP3	14.8	0.262	0
501.0019	49	F	0/6	N	Y	rs779024919	*LMX1B*	3' UTR	PM2; PP3	9.063	0.883	6.49e-5
501.0038	50	F	2/6	Y	Y	chr3:30733401 TTA>TTATA; rs199931498	*TGFBR2*	3' UTR	PM2; PP3	n/a; 14.03	0.134; 0.953	0; 0.017
601.0030	26	F	3/5	Stroke	N	rs577885381	*LOX*	3' UTR	PM2; PP3	6.334	0.0990	0
202.0004	36	F	0/6	Y	N	rs35282763	*LRP1*	Promoter	PM2; PP3	20.6	0.965	6.68e-5
						chr9:137533804 G>A	*COL5A1*	5' UTR	PM2; PP3	16.37	0.426	0

Detailed description of non-coding variants identified in a candidate gene and the clinical characteristics of individuals harboring each variant.

Patient – patient ID#, *Age* – age at admission to hospital with ACS, *Sex* – biological sex, *CVD Risk* – number of cardiovascular risk factors (defined as obesity, dyslipidemia, smoking, hypertension, diabetes, and where available, family history of early MI), *Recurrent* – patient had a previous MI or a cardiac rehospitalization within a one year follow up period, *MV (multivessel)* – angiogram revealed multiple vessels were dissected, *rsID* – rs number of the variant or GRCh37 coordinates and allele if variant is novel, *Gene* - affected gene, *Variant Type* – regulatory region where variant is located according to the VEP canonical transcript, *ACMG Criteria* – summary of lines of evidence used to determine ACMG status, *CADD* – the CADD phred score (see table S1 for 50th percentile thresholds), *LINSIGHT* – the LINSIGHT score used as another line of evidence for pathogenicity of non-coding variants (see table S1 for thresholds), *gnomAD popmax* – the maximum allele frequency of all populations in the gnomAD database

Table 4 – Coding variants in non-candidate genes

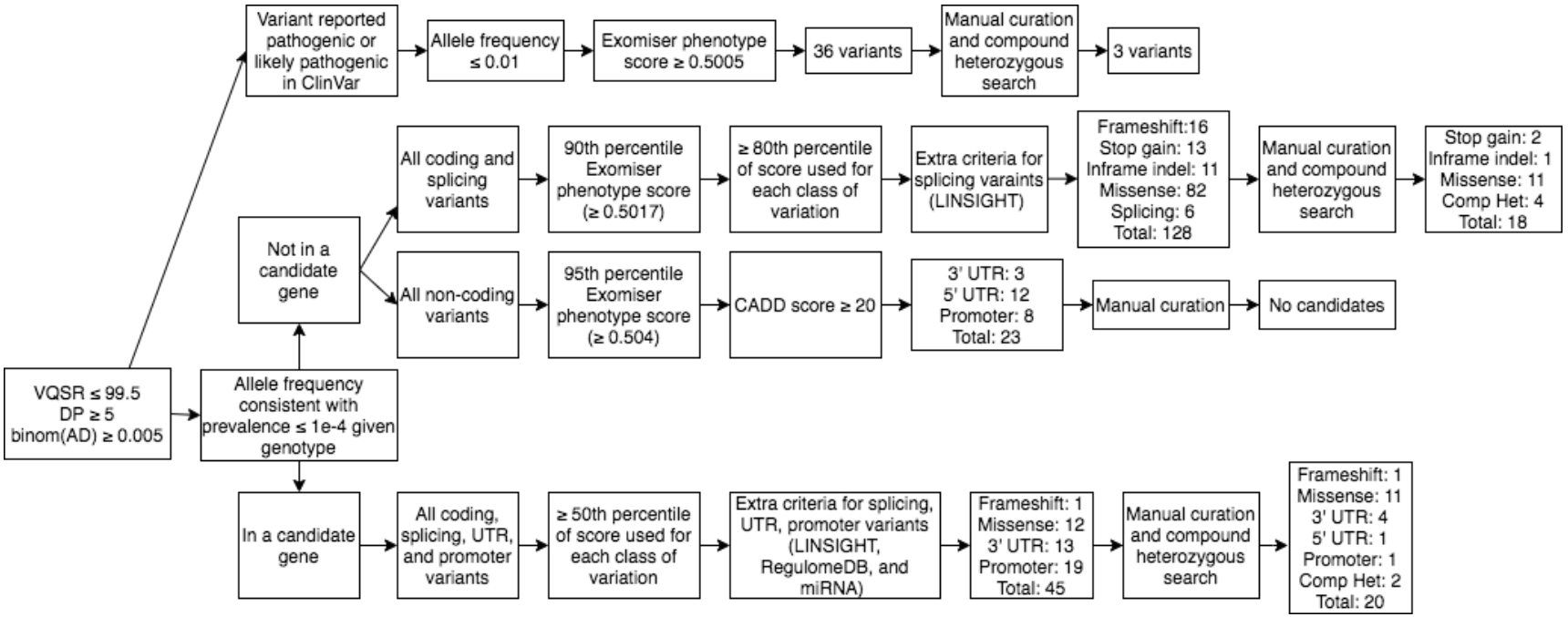
Patient	Age	Sex	CVD Risk	Recurrent	MV	rsID	Gene	HGVSp	Exomiser Score	ACMG Criteria	CADD	Median Rank	gnomAD popmax
102.0015	51	M	4/6	N	N	rs79378857	*COL7A1*	Pro1458Leu	0.5038	PM2; PP3; PP5	26.8	0.91	3.6e-3
301.0031	49	F	1/6	N	N	rs1263254683	*COL4A5*	Lys288Asn	0.504	PM2; PP3	23.5	0.89	3.65e-5
407.0040	45	M	3/6	N	N	rs377150636	*COL27A1*	Gly1412Val	0.504	PM1; PM2; PP3	24	0.98	6.20e-5
407.0059	43	F	1/5	Y	N	rs1440709540	*COL27A1*	Gly1415Ser	0.504	PM1; PM2; PP3	25	0.97	0
						rs573310430; rs200667255	*LTBP4*	Arg954Cys; Pro1037Arg	0.5778	PM2; PP3	29.7; 25.7	0.88; 0.81	6.46e-5; 1.74e-3
509.0023	48	F	2/5	N	N	rs370544100	*COL12A1*	His595Asn	0.5039	PM1; PM2; PP3	28.2	0.84	2.65e-5
401.0050	44	M	0/6	N	N	chr12:65634744 T>C	*LEMD3*	Ser728Pro	0.5707	PM1; PM2; PP3	28.4	0.88	0
407.0106	53	F	5/6	N	N	rs748141103	*GDF5*	Thr469Arg	0.5053	PM2; PP3	27.1	0.92	8.68e-5
101.0053	46	F	2/6	Y	N	rs147908281	*ITGB8*	Arg182Cys	0.5031	PM1; PM2; PP3	19.99	0.80	8.82e-5
401.0082	45	F	5/6	Y	N	rs1179032756	*ITGB4*	Cys598Arg	0.5033	PM2; PP3	28.3	0.97	0
505.0033	50	F	1/6	N	N	rs755511649; rs149090091	*ITGB5*	Glu711_Pro712 delinsAla; Glu711*	0.5032	PM2; PM4; PP3	22.6; 46	n/a	1.77e-5
407.0140	52	F	3/5	N	N	rs77563738	*CECR1*	Arg169Gln	#N/A	PM1; PM2; PP3	21.8	0.94	6.07e-4
601.0057	53	F	4/6	Y	N	chrX:44938589 C>G	*KDM6A*	Ser1046*	0.5815	PVS1; PM2; PP3	44	0.92	0
408.0034	40	F	2/5	N	N	rs146505280	*BLK*	Arg359Cys	0.5007	PM1; PM2; PP3; PP5	29.4	0.96	1.99e-3
701.0002	45	M	0/6	N	N	rs886045339	*ADAR*	Lys974Glu	0.6146	PM1; PM2; PP3	31	0.91	0

Detailed description of coding variants identified in a non-candidate gene and the clinical characteristics of individuals harboring each variant.

Patient – patient ID#, *Age* – age at admission to hospital with ACS, *Sex* – biological sex Male or Female, *CVD Risk* – number of cardiovascular risk factors (defined as obesity, dyslipidemia, smoking, hypertension, diabetes, and where available, family history of early MI), *Recurrent* – Yes or No previous MI or a cardiac rehospitalization within a one year follow up period, *MV (multivessel)* – Yes or No angiogram revealed multiple vessels were dissected, *rsID* – rs number of the variant or GRCh37 coordinates and allele if variant is novel, *Gene* - affected gene, *HGVSp* – amino acid change according to the VEP canonical transcript, *Exomiser Score* – gene-phenotype score assigned by Exomiser (minimum 0.5017), *ACMG Criteria* – summary of lines of evidence used to determine ACMG status, *CADD* – the CADD phred score (see table S1 for 80th percentile thresholds), *Median Rank* – the median rank score of four algorithms selected to determine pathogenicity of missense variants (minimum 0.8), *gnomAD popmax* – the maximum allele frequency of all populations in the gnomAD database

Figures

Figure 1 – Variant Prioritization Flowchart



References

1. Krittanawong C, Kumar A, Johnson KW, et al. Conditions and Factors Associated With Spontaneous Coronary Artery Dissection (from a National Population-Based Cohort Study). Am J Cardiol. 2019;123(2):249-253.
2. Saw J, Mancini GBJ, Humphries KH. Contemporary Review on Spontaneous Coronary Artery Dissection. J Am Coll Cardiol. 2016;68(3):297-312.
3. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-424.
4. Kaadan MI, MacDonald C, Ponzini F, et al. Prospective Cardiovascular Genetics Evaluation in Spontaneous Coronary Artery Dissection. Circ Genom Precis Med. 2018;11(4):e001933.
5. Henkin S, Negrotto SM, Tweet MS, et al. Spontaneous coronary artery dissection and its association with heritable connective tissue disorders. Heart. 2016;102(11):876-881.
6. Adlam D, Olson TM, Combaret N, et al. Association of the PHACTR1/EDN1 Genetic Locus With Spontaneous Coronary Artery Dissection. J Am Coll Cardiol. 2019;73(1):58-66.
7. Sun Y, Chen Y, Li Y, et al. Association of TSR1 Variants and Spontaneous Coronary Artery Dissection. J Am Coll Cardiol. 2019;74(2):167-176.
8. Turley TN, Theis JL, Sundsbak RS, et al. Rare Missense Variants in TLN1 Are Associated With Familial and Sporadic Spontaneous Coronary Artery Dissection. Circ Genom Precis Med. 2019;12(4):e002437.
9. Georges A, Albuisson J, Berrandou T, et al. Rare Loss-of-function Mutations of PTGIR are enriched in Fibromuscular Dysplasia. Cardiovasc Res. 2020.
10. Turley TN, O'Byrne MM, Kosel ML, et al. Identification of Susceptibility Loci for Spontaneous Coronary Artery Dissection. JAMA Cardiol. 2020.
11. Arnaud P, Hanna N, Benarroch L, et al. Genetic diversity and pathogenic variants as possible predictors of severity in a French sample of nonsyndromic heritable thoracic aortic aneurysms and dissections (nshTAAD). Genet Med. 2019;21(9):2015-2024.
12. Pilote L, Karp I. GENESIS-PRAXY (GENdEr and Sex determInantS of cardiovascular disease: From bench to beyond-Premature Acute Coronary SYndrome). Am Heart J. 2012;163(5):741-746 e742.
13. Raparelli V, Elharram M, Shimony A, Eisenberg MJ, Cheema AN, Pilote L. Myocardial Infarction With No Obstructive Coronary Artery Disease: Angiographic and Clinical Insights in Patients With Premature Presentation. Can J Cardiol. 2018;34(4):468-476.
14. Poplin R, Ruano-Rubio V, DePristo MA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. 2018:201178.
15. Liu X, White S, Peng B, et al. WGSA: an annotation pipeline for human genome sequencing studies. J Med Genet. 2016;53(2):111-112.
16. Smedley D, Jacobsen JO, Jager M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc. 2015;10(12):2004-2015.
17. Kohler S, Carmody L, Vasilevsky N, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res. 2019;47(D1):D1018-D1027.
18. Eilbeck K, Quinlan A, Yandell M. Settling the score: variant prioritization and Mendelian disease. Nat Rev Genet. 2017;18(10):599-612.

19. Dong C, Wei P, Jian X, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet. 2015;24(8):2125-2137.
20. Ghosh R, Oak N, Plon SE. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. Genome Biol. 2017;18(1):225.
21. 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):1073-1081.
22. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-249.
23. 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;46(3):310-315.
24. Carter H, Douville C, Stenson PD, Cooper CN, Karchin R. Identifying Mendelian disease genes with the Variant Effect Scoring Tool. BMC Genomics. 2013;14:S3.
25. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet. 2016;99(4):877-885.
26. Jagadeesh KA, Wenger AM, Berger MJ, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet. 2016;48(12):1581-1586.
27. Anderson D, Lassmann T. A phenotype centric benchmark of variant prioritisation tools. NPJ Genom Med. 2018;3:5.
28. Pejaver V, Urresti J, Lugo-Martinez J, et al. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. bioRxiv. 2017:134981.
29. Huang YF, Gulko B, Siepel A. Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. Nat Genet. 2017;49(4):618-624.
30. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015;4.
31. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22(9):1790-1797.
32. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am J Hum Genet. 2017;100(2):267-280.
33. Nishiguchi T, Tanaka A, Ozaki Y, et al. Prevalence of spontaneous coronary artery dissection in patients with acute coronary syndrome. Eur Heart J Acute Cardiovasc Care. 2016;5(3):263-270.
34. Saw J, Humphries K, Aymong E, et al. Spontaneous Coronary Artery Dissection: Clinical Outcomes and Risk of Recurrence. J Am Coll Cardiol. 2017;70(9):1148-1158.
35. Callewaert BL, Willaert A, Kerstjens-Frederikse WS, et al. Arterial tortuosity syndrome: clinical and molecular findings in 12 newly identified families. Hum Mutat. 2008;29(1):150-158.
36. Wang L, Guo DC, Cao J, et al. Mutations in myosin light chain kinase cause familial aortic dissections. Am J Hum Genet. 2010;87(5):701-707.
37. Zekavat SM, Kii E, Chaffin M, et al. Whole Exome Sequencing of Spontaneous Coronary Artery Dissection (SCAD) Cases Discovers Enrichment of Rare Disruptive Variants in Fibrillar Collagen Genes. Circulation. 2018;138.

38. Jefferies JL, Towbin J, Ryan T, Lucky A. Cardiovascular Findings in Recessive Dystrophic Epidermolysis Bullosa. Journal of the American College of Cardiology. 2013;61(10).
39. Diez-del Hoyo F, Sanz-Ruiz R, Diez-Villanueva P, et al. A novel cardiovascular presentation of Alport Syndrome: spontaneous coronary artery dissection. Int J Cardiol. 2014;177(3):e133-134.
40. Anuwatworn A, Sethi P, Steffen K, Jonsson O, Petrasko M. Spontaneous Coronary Artery Dissection: A Rare Manifestation of Alport Syndrome. Case Rep Cardiol. 2017;2017:1705927.
41. Loeys BL, Chen J, Neptune ER, et al. A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. Nat Genet. 2005;37(3):275-281.
42. Das M, Ithychanda S, Qin J, Plow EF. Mechanisms of talin-dependent integrin signaling and crosstalk. Biochim Biophys Acta. 2014;1838(2):579-588.
43. Turcan I, Pasmooij AM, van den Akker PC, et al. Heterozygosity for a Novel Missense Mutation in the ITGB4 Gene Associated With Autosomal Dominant Epidermolysis Bullosa. JAMA Dermatol. 2016;152(5):558-562.
44. Munger JS, Sheppard D. Cross talk among TGF-beta signaling pathways, integrins, and the extracellular matrix. Cold Spring Harb Perspect Biol. 2011;3(11):a005017.
45. Verweij N, Eppinga RN, Hagemeijer Y, van der Harst P. Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure. Sci Rep. 2017;7(1):2761.
46. Su H, Kim H, Pawlikowska L, et al. Reduced expression of integrin alphavbeta8 is associated with brain arteriovenous malformation pathogenesis. Am J Pathol. 2010;176(2):1018-1027.
47. Miyake N, Mizuno S, Okamoto N, et al. KDM6A point mutations cause Kabuki syndrome. Hum Mutat. 2013;34(1):108-110.
48. Lederer D, Shears D, Benoit V, Verellen-Dumoulin C, Maystadt I. A three generation X-linked family with Kabuki syndrome phenotype and a frameshift mutation in KDM6A. Am J Med Genet A. 2014;164A(5):1289-1292.
49. Bogershausen N, Gatinois V, Riehmer V, et al. Mutation Update for Kabuki Syndrome Genes KMT2D and KDM6A and Further Delineation of X-Linked Kabuki Syndrome Subtype 2. Hum Mutat. 2016;37(9):847-864.
50. Gatto LAM, Sousa LHA, Koppe GL, Demartini ZJ. Carotid artery occlusion in Kabuki syndrome: Case report and literature review. Surg Neurol Int. 2017;8:88.
51. Dyamenahalli U, Abraham B, Fontenot E, Prasad V, Imamura M. Pathologic Aneurysmal Dilation of the Ascending Aorta and Dilation of the Main Pulmonary Artery in Patients with Kabuki Syndrome: Valve-sparing Aortic Root Replacement. Congenit Heart Dis. 2007;2:424–428.
52. Tweet MS, Hayes SN, Codsi E, Gulati R, Rose CH, Best PJM. Spontaneous Coronary Artery Dissection Associated With Pregnancy. J Am Coll Cardiol. 2017;70(4):426-435.
53. Hayes SN, Kim ESH, Saw J, et al. Spontaneous Coronary Artery Dissection: Current State of the Science: A Scientific Statement From the American Heart Association. Circulation. 2018;137(19):e523-e557.

54. Tavtigian SV, Greenblatt MS, Harrison SM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genet Med. 2018;20(9):1054-106.

Supplemental Methods

Whole Genome Sequencing and Annotation

DNA extraction from blood samples was performed with the Qiagen Flexigene kit. We performed paired-end sequencing on an Illumina machine at 30x coverage. Alignment was performed using BWA with default parameters. Duplicates were marked, joint calling was performed with HaplotypeCaller, and insertion/deletion variants were realigned according to the GATK best practices.[1] VQSR was used to perform quality control according to GATK best practices using the 99.5% truth tranche as a threshold. Variants were annotated using WGSA version 0.8[2] and analysis was performed according to VEP worst consequence annotations on Ensembl transcripts (release 94).[3] The Eukaryotic Promoter Database (EPD)[4] was used to define promoter variants, as these are not a standard part of Ensembl annotations.

The Exomiser (version 12.0) hi-PHIVE algorithm[5] was used to generate gene-phenotype association scores for each gene in the genome with the Human Phenotype Ontology (HPO) terms[6] "Coronary artery dissection" (HP:0006702), "Arterial dissection" (HP:0005294), "Arterial tortuosity" (HP:0005116), "Aortic dissection" (HP:0002647), "Thoracic aortic aneurysm" (HP:0012727), and "Joint hypermobility" (HP:0001382). This method generates higher gene-phenotype scores for known genes associated with SCAD and CTDs (median score: 0.7138) and all genes associated with the terms "dissection", "aneurysm", and "arterial tortuosity" in the HPO database (median score: 0.6244) compared to all other genes (median score: 0.5) (**Fig S1**). Other gene level annotations such as diseases and inheritance patterns reported in the Online Mendelian Inheritance in Man (OMIM) database (https://www.omim.org/), GTEx median

expression level in coronary artery tissue (https://gtexportal.org/home/), and gnomAD pLI scores[7] were added using custom R scripts.

In addition to VQSR, we added strict read depth filters to ensure that only variants called with very high confidence were included in our analysis. Variants with a read depth ⩽ 5 were excluded from analysis. In addition, Tian *et al.* presented evidence that HaplotypeCaller has a lower sensitivity below 40x coverage.[8] Therefore, variants with a read depth ⩽ 40 were excluded if the binomial likelihood of the observed read distribution was ⩽ 0.5%. Variants were also subject to allele frequency thresholds. Variants were excluded if they had an allele count > 5 (MAF>0.055) within the sequenced cohort. Heterozygous variants were included in the analysis if they had an allele frequency ⩽ 0.0001 in ExAC[9] and gnomAD (population maximum of exomes and genomes).[7] Homozygous variants were included if they had an allele frequency ⩽ 0.01 in the same databases. Exceptions were variants that had been previously reported as pathogenic or likely pathogenic in ClinVar,[10] which only required an allele frequency threshold of ⩽ 0.01.

Potential compound heterozygous variants were also included if they had an allele frequency ⩽ 0.01 in the same databases (the two variants had a multiplied allele frequency of ⩽ 0.0001). If potentially compound heterozygous variants were close enough to be observed on the same read (approximately 100 base-pairs), we inspected the BAM file in Integrative Genomics Viewer to confirm compound heterozygous status. Results of this analysis are found in **Table S2**.

After QC and allele frequency thresholding, 25,081 unique variants were included in the analysis.

Manual Curation

Manual exclusion criteria include heterozygosity for a variant that is pathogenic in ClinVar only in biallelic state, Exomiser identified closest matching HPO term is based on a deletion syndrome, failure of variant alignment or calling (for example, multiple variants identified in the repetitive region of a gene), instances where 1 of the 4 missense scores was skewing the median rank score, a ClinVar report of benign or likely benign, variants that failed gnomAD filters, loss-of-function variants in genes determined to be tolerant to loss of function by the gnomAD pLI score, no expression in artery tissues according to GTEx, and variants where the closest HPO term match provided by Exomiser was a mechanism distinct from SCAD and other vasculopathies (for example, familial hypercholesterolemia, arterial calcification, congenital heart defects). After manual curation we found that an Exomiser threshold of 0.503, including approximately 6% of all genes, might be appropriate for future SCAD studies using this tool.

Supplemental Results

Pathogenicity Prediction Algorithms

To validate our choice of VEST4, M-CAP, MetaSVM, and REVEL to score missense variants, we compared the distributions of the median rank scores obtained from these algorithms against the median rank scores of the combination of SIFT, PolyPhen, CADD, and MutationTaster (**Fig S2**). The two methods are correlated, with a Spearman coefficient of 0.53. On average, the popular method produced a median rank score 9.6% higher than that of our method. Our method scores only 4% of missense variants above the 80th percentile, compared to 14.9% using the popular method. Because our dataset contains only 45 individuals, it is unlikely

that we would observe this proportion of all possible variation, indicating a high false positive rate in the popular method. Thus, our method likely has increased specificity, possibly at the expense of sensitivity.

Non-Coding Variants in Candidate Genes

Many non-coding elements are not included in popular annotation softwares and have not been characterized at the single base-pair resolution required for functional predictions. Currently, WGSA includes information on transcription factor binding and miRNA, which was included in our analysis.

One non-coding variant is in the 5' UTR of the *MYLK* gene, which controls smooth muscle cell contractility. This variant is potentially compound heterozygous with a VUS missense variant in *MYLK* (**Table S2**), although the variants were too far apart to confirm phase. Heterozygous missense variants in this gene can be pathogenic for TAAD and ENCODE ChIP-seq data shows a strong signal of RNA Polymerase 2 binding at this site, possibly affecting expression.

A variant in the 3' UTR of the *LMX1B* gene, which causes the autosomal dominant CTD Nail-patella syndrome associated with SCAD,[11] was observed in one patient. LMX1B is a transcription factor that promotes expression of basement membrane proteins *COL4A3* and *COL4A4* in mice.[12] While the exact mechanism of action is unknown, LMX1B deficiency could result in reduced levels of arterial cartilage, weakening the vessel wall structure. Similarly, an overabundance of LMX1B and arterial collagen could also weaken vascular integrity similar to the phenotype observed in Loeys-Dietz syndrome, where artery walls are too thick to maintain integrity.

A 3' UTR variant in the *LOX* gene (lysyl oxidase) was observed in a very young (26 yrs) patient. In addition, this individual reported a previous stroke. The *LOX* gene controls cross-linking between collagens and the ECM and coding variants are associated with familial TAAD under an autosomal dominant model.[13] Certain polymorphisms are enriched in ischemic stroke patients,[14] suggesting that this variant could be the source of both events. While the exact mechanism of action is unknown, both deficiency and overabundance of the *LOX* protein product could lead to a disorganized extracellular matrix.

One patient had a variant in the promoter region of the *LRP1* gene, which encodes LDL Receptor Related Protein 1. A variant within *LRP1* was recently implicated in a GWAS of SCAD[15] with a hypothesized mechanism of focal adhesion dysregulation. The same patient also has a variant in the 5' UTR of *COL5A1*. ENCODE ChIP-seq data shows a strong signal of RNA Polymerase 2 binding for both of these sites, suggesting the variants could result in reduced expression.

One patient has phase-verified compound heterozygous variants in the 3' UTR of the *TGFBR2* gene (**Table S2**); two variants that occur 3 base pairs apart. TargetScan predicts one of these variants is within a miR-410-3p binding site. Studies in U251 and A172 human glioma cell lines confirm that miR-410-3p binding represses *TGFBR2* expression,[16] suggesting that these variants could increase *TGFBR2* expression consistent with the mechanism of Loeys-Dietz syndrome. This individual had a previous ACS, a multivessel SCAD event at age 50, and coronary artery hyper-tortuosity that further supports the functional impact of these variants.

One patient has a 3' UTR variant in *TGFB2* that is predicted to affect the binding site of miR-323-3p. Thus, this variant has the potential to cause SCAD through a mechanism similar to

the one outlined above. The individual has a similarly severe phenotype, exhibiting coronary hyper-tortuosity, SCAD at age 46, and a previous ACS.

Variants in Non-Candidate Genes

Collagens

Most pathogenic collagen variants are loss-of-function or amino acid substitutions affecting glycine residues of the triple helix Gly-X-Y repeat structure. While these variants are most often identified because the causal mechanism is clear, certain configurations of the Xaa and Yaa are more stable, prone to post-translational modification, and even dictate the exact angle of the helix.[17] Most of the identified variants in our cohort are in the triple helix, however, we hypothesize that they may also affect other aspects of the molecule such as the rate of assembly, post-translational modifications, or interactions with other ECM molecules.

COL7A1

COL7A1 forms anchoring fibrils that connect the basal lamina to the reticular lamina in the basement membrane.[18] We identified a heterozygous variant in one patient substituting the Xaa proline to a leucine (p.P1458L) in the "interrupted" collagenous domain, meaning there are imperfections in the rigid Gly-X-Y structure. Because the glycine pattern is slightly disrupted in this domain and proline in the Xaa position enhances stability of the triple helix,[17] we believe this proline substitution could result in destabilization. In addition, slower self-assembly of a collagen molecule is associated with an increased number of post-translational modifications on the final product.[17] Thus, a destabilizing variant may also cause over-modification. In addition, this variant has previously been reported as pathogenic in ClinVar (SCV000490486.1) for

compound heterozygous recessive dystrophic epidermolysis bullosa (RDEB), which causes severe blistering of the skin.[19] Approximately 20% of RDEB patients show mild aortic dilation suggesting vascular involvement.[20] Phenotyping heterozygous RDEB variant carriers could confirm this result.

COL4A5

COL4A5 is an integral part of the basement membrane, in which premature truncation and mostly glycine missense variants cause X-linked Alport syndrome.[21] The variant identified in this female patient is a Yaa lysine to asparagine substitution (p.K288N) close to the N-terminus of the collagen triple helical domain. MutPred predicts an "actionable hypothesis" that this substitution could alter post-translational modifications at this site by loss of glycosylation, methylation and/or ubiquitination. Indeed, lysine residues at the N-terminal and C-terminal ends are often glycosylated before they are cross-linked to other helices by lysyl oxidase.[17] Thus, the loss of post-translational modification at this site could destabilize the interactions of *COL4A5* with itself and other collagens. Heterozygous females usually present with a milder phenotype, making it possible this female patient was an undiagnosed carrier.[22] Case reports have previously documented SCAD as the first presentation of undiagnosed Alport syndrome. [23,24]

COL12A1

One patient had a heterozygous missense variant in *COL12A1* that substituted a histidine for an asparagine (p.H595N) in the second von Willebrand factor type A (VWA) domain of the protein. *COL12A1* variants can cause myopathic Ehlers-Danlos syndrome (mEDS), which manifests as joint hypermobility and hypotonia in the most severe cases,[25] but can also have a

subtle phenotype.[26] Heterozygous missense variants can be pathogenic,[25] however, a potential compound heterozygous variant was identified in the 5' UTR as well (**Table S2**). Although glycine residues are most often altered to cause Ehlers-Danlos like phenotypes, pathogenic non-glycine variants are relatively common in ClinVar for mEDS. The main function of *COL12A1* is to anchor *COL1A1* fibrils to other ECM proteins and cells via fibronectin and VWA domains.[27] *COL12A1* has two documented isoforms; the "long isoform" includes all domains and the "short isoform" excludes this VWA domain. Both isoforms are expressed in cultured endothelial cells[28] and the long isoform appears more prevalent. Furthermore, *COL12A1* binds to and influences the expression of tenascin x, the protein product of *TNXB*.[26] Variants in *TNXB* can cause Ehlers-Danlos syndrome[29] concurrent with arterial dissections[30], offering another link to vasculopathy phenotypes. This is the first report of a cardiovascular complication in a patient with a genetic variant consistent with mEDS.[31]

COL27A1

Another collagen gene implicated in our results is *COL27A1*; one patient had a glycine to serine substitution (p.G1415S) and another had a glycine to valine substitution (p.G1412V). Both variants occur in the triple helix domain, separated by only three amino acids. In addition, MutPred predicts a "very confident" gain of phosphorylation for the glycine to serine variant. Biallelic variants in *COL27A1* cause Steel syndrome, which primarily affects bones and cartilage,[32] however, *COL27A1* is prevalent in the coronary arteries of developing mice[33] and was identified as one of a class of genes enriched in disruptive variants in SCAD patients.[34] Both patients had potential compound heterozygous variants in the 3' or 5' UTR (**Table S2**), however, both patients are average in height, ruling out a classical manifestation of biallelic Steel

syndrome. A mild carrier phenotype has been documented.[35] Variants in this region of the triple helix have not been reported in ClinVar. Further work is needed on cardiovascular events in Steel Syndrome patients and variant carriers to determine if SCAD is part of the manifestation of Steel Syndrome, and if this region has a cardiovascular function.

TGF-β Pathway

Two rare missense variants in *LTBP4* (p.R954C, p.P1037R) were identified in a compound heterozygous state (**Table S2**) in one patient who also had an identified VUS missense variant in *SMAD3*. The *LTBP4* gene product binds to Latency-associated protein and sequesters inactive TGF-β ligand in the ECM. One of the variants introduces a cysteine into a calcium binding EGF domain with multiple sulfide bonds, and thus may be a gain of function variant affecting disulfide bonds. Biallelic variants in *LTBP4* cause cutis laxa (loose skin) with severe pulmonary, gastrointestinal, and urinary abnormalities due to defects in the formation of elastic fibers.[36] The severe form, which often leads to premature death from respiratory failure, is caused by the presence of at least one loss-of-function variant.[36] The PRAXY patient did not have a loss-of-function variant, and thus may have a less severe phenotype with minimal pulmonary involvement allowing them to survive to age 43. However, the patient did show coronary artery hyper-tortuosity, and may be the first report of arterial tortuosity and vasculopathy in a patient with a biallelic *LTBP4* variant. While some forms of cutis laxa are associated with aortic aneurysm and arterial tortuosity,[37] most *LTBP4* cutis laxa patients do not exhibit this even though patients have elevated TGF-β levels.[36,37]

One patient had a missense variant in the *GDF5* gene, which codes for a TGF-β ligand. According to ClinVar and OMIM, heterozygous variants in the gene can cause a range of

skeletal and articular phenotypes, including brachydactyly or shortening of the fingers, DuPan syndrome, which manifests as brachydactyly and absence of the fibula, and multiple synostoses syndrome, which causes progressive joint fusion through aberrant chondrogenesis. For the identified threonine to arginine variant (p.T469R), MutPred predicts a "very confident" loss of glycosylation. Nearby variants at amino acid position 441 and 475 have been experimentally validated to reduce SMAD 1/5/8 activity, indicating that this protein region regulates TGF-β signaling.[38] GDF5 also interacts with FBN1 at its N-terminal domain, and thus a gain or loss of affinity of GDF5 could lead to increased sequestering of other TGF-β ligands.[39] While a variant in *GDF5* should lead to decreased TGF-β signaling, feedback mechanisms could actually increase signaling as observed with variants in other TGF-β ligands causing LDS.[40]

LEMD3 encodes a nuclear membrane protein that binds to DNA and SMAD proteins and is as a repressor of the TGF-β pathway.[41] The specific variant observed in this SCAD patient (p.S728P), which has been previously described in detail (see Chapter 2), is likely to disrupt the DNA binding domain. MutPred predicts an "actionable hypothesis" of a gain of helix, gain of sheet, and/or gain of loop for this variant. Truncation variants in this gene cause the CTD Buschke-Ollendorf syndrome,[42] which primarily affects bones and can cause collagen deposits in the skin. Homozygous knockout mice have elevated TGF-β signaling consistent with an LDS-like phenotype, as well as a disorganized developing vasculature, fewer smooth muscle cells in blood vessels, and a malformed aorta.[43]

Integrins

One patient had a heterozygous amino acid deletion immediately followed by a premature stop codon in *ITGB5* (p.E711_P712delinsA; p.E711*); part of the same rare haplotype.

The stop codon removes the transmembrane and cytoplasmic domains, which would interrupt cellular signaling in response to fibronectin and binding to TLN1, and thus its connection to the actin cytoskeleton. ITGB5 also interacts with latency-associated protein (LAP) to sequester TGF-β1.[44] Therefore, defective ITGB5 could result in increased TGF-β signaling consistent with Marfan- or Loeys-Dietz-like phenotypes.

Another variant of interest was identified in the integrin gene *ITGB4*, which encodes a protein that binds to laminins. Laminin gene variants cause periventricular nodular heterotopia (PVNH), a seizure disorder often accompanied by artery dissections.[45] Other interruptions of this pathway could have an overlapping phenotypic spectrum. *ITGB4* has been linked to a mild form of epidermolysis bullosa,[46] which could be related to SCAD (see *COL7A1* result). The amino acid substitution, a cysteine to arginine (p.C598R), is at a disulfide bond site according to Uniprot (https://www.uniprot.org/), consistent with the MutPred prediction of a "gain of disorder" for this protein. In addition, this variant has been reported in ClinVar as a VUS for an unspecified phenotype (SCV000853811.1).

Another patient had a missense variant in *ITGB8*, which is also a fibronectin receptor. The variant introduces a cysteine into the extracellular domain in a region with multiple disulfide bonds (p.R182C). ITGB8 sequesters TGF-β1,[44] and it is possible that a misfolded ITGB8 will alter levels of TGF-β signaling. Indeed, common variants in *ITGB8* are associated with a vasculopathy, brain arteriovenous malformations, and affected tissues display decreased expression of *ITGB8*.[47] Homozygous *ITGB8* knock-out mice have reduced signaling and enlarged, tortuous brain capillaries,[47] consistent with our observation of coronary artery hyper-tortuosity in this patient.

Inflammatory Pathways

We observed a heterozygous variant in *CECR1* (p.R169Q) that has been reported in ClinVar as pathogenic for polyarteritis nodosa (SCV000994598.1) in a biallelic state. Polyarteritis nodosa is a systemic inflammatory condition that attacks blood vessels in affected children,[48] leading to lifelong cardiovascular complications including SCAD.[49] There is some evidence that heterozygous carriers have increased rates of recurrent cardiovascular events.[50] This SCAD patient had no evidence of coronary hyper-tortuosity, consistent with the proposed mechanism.

One SCAD patient had a heterozygous missense variant in the *BLK* gene, which has been reported to cause Systemic Lupus Erythematosus (SLE) in ClinVar (SCV000886651.1), a condition enriched in SCAD patients.[49] Jiang *et al* report that missense variants in *BLK* cause approximately 10% of SLE.[51] Furthermore, they demonstrate that this specific amino acid change (p.R359C) interferes with repression of INFβ in human cells.[51] Despite the known association of SLE and SCAD, this gene was not included in the candidate gene list derived from other studies[11] and the HPO database. There are case reports documenting SCAD as the first presentation of SLE,[52] however, we do not have information indicating if this patient had other symptoms consistent with SLE. Therefore, we consider this variant a VUS for SCAD.

The heterozygous *ADAR* variant (p.K974E) found in one patient has been described in detail (see Chapter 2). Briefly, the variant occurs in the RNA editase domain and is predicted to directly interact with the RNA.[53] Biallelic variants in this protein region cause Aicardi-Goutières Syndrome, which can cause cerebral arteriopathy and SLE in some patients.[54] Carriers have not been thoroughly evaluated for cardiovascular risk or presence of SLE. Therefore, we suggest that

this *ADAR* variant could lead to SCAD through vasculopathy associated with being a carrier of an AGS causing variant, or through SLE.

Supplemental Tables

Table S1 – Score Thresholds

Variation Class	**50% CADD**	**80% CADD**	**LINSIGHT**
Disruptive	15.68	27.3	N/A
Splicing	15.68	27.3	0.9
Inframe Indel	13.46	17.80	N/A
5' UTR	8.134	12.66	0.128
3' UTR	3.19	8.46	0.076
*Promoter**	8.198	14.28	0.106

Thresholds used to filter each class of variant. The 50th percentile was used for genes previously associated with SCAD or related phenotypes, whereas the 80th percentile was used for non-candidate genes.

* Only non-coding promoter variants (intronic, upstream, downstream, and intergenic) were used to calculate the score threshold

Table S2 – Summary of Potential Compound Heterozygous Variants

Patient	Age	Sex	CVD risk	Recurrent	**M.V.**	**rsID**	**Gene**	**Variant Type**	**Phased**	**CADD**	**LINSIGHT**	**gnomAD popmax**	**miRNA**
407.0040	45	M	3/6	N	N	rs377150636	COL27A1	Missense	N	24	n/a	6.20e-5	
						rs368463689	COL27A1	5' UTR	N	17.01	0.88	9.36e-3	
407.0059	43	F	1/5	Y	N	rs1440709540	COL27A1	Missense	N	25	n/a	0	
						rs779351956	COL27A1	3' UTR	N	16.31	0.95	1.20e-3	let-7-5p/98-5p
407.0059	43	F	1/5	Y	N	rs573310430	LTBP4	Missense	Y	29.7	n/a	6.46e-5	
						rs200667255	LTBP4	Missense	Y	25.7	n/a	1.74e-3	
509.0023	48	F	2/5	N	N	rs370544100	COL12A1	Missense	N	28.2	n/a	2.65e-5	
						rs554201004	COL12A1	3' UTR	N	15.95	0.96	1.18e-3	
101.0130	45	F	1/6	N	N	rs1052930526	MYLK	Missense	N	26.8	n/a	2.66e-5	
						rs1040685992	MYLK	5' UTR	N	15.13	0.17	2.11e-4	
501.0038	50	F	2/5	Y	Y	chr3:30733401 TTA>TTATA	TGFBR2	3' UTR	Y	n/a	0.13	0	miR-410-3p
						rs199931498	TGFBR2	3' UTR	Y	14.03	n/a	0.017	

Detailed description of variants identified in compound heterozygous analysis and the clinical characteristics of individuals harboring each variant.

Patient – patient ID#, *Age* – age at admission to hospital with ACS, *Sex* – biological sex Male or Female, *CVD Risk* – number of cardiovascular risk factors (defined as obesity, dyslipidemia, smoking, hypertension, diabetes, and where available, family history of early MI), *Recurrent* – Yes or No previous MI or a cardiac rehospitalization within a one year follow up period, *MV (multivessel)* – Yes or No angiogram revealed multiple vessels were dissected, *rsID* – rs number of the variant or GRCh37 coordinates and allele if variant is novel, *Gene* - affected gene, *Variant Type* – variant consequence according to the VEP canonical transcript, *Phased* – Yes or No the phase of the two variants was confirmed by visual inspection in IGV, *CADD* – the CADD phred score (see table S1 for 80$_{th}$ percentile thresholds), *LINSIGHT* – the LINSIGHT score used as another line of evidence for pathogenicity of non-coding variants (see table S1 for thresholds), *gnomAD popmax* – the maximum allele frequency of all populations in the gnomAD database, *miRNA* – miRNAs predicted to bind at this site from the TargetScan database

Table S3 – Evidence of Oligogenicity

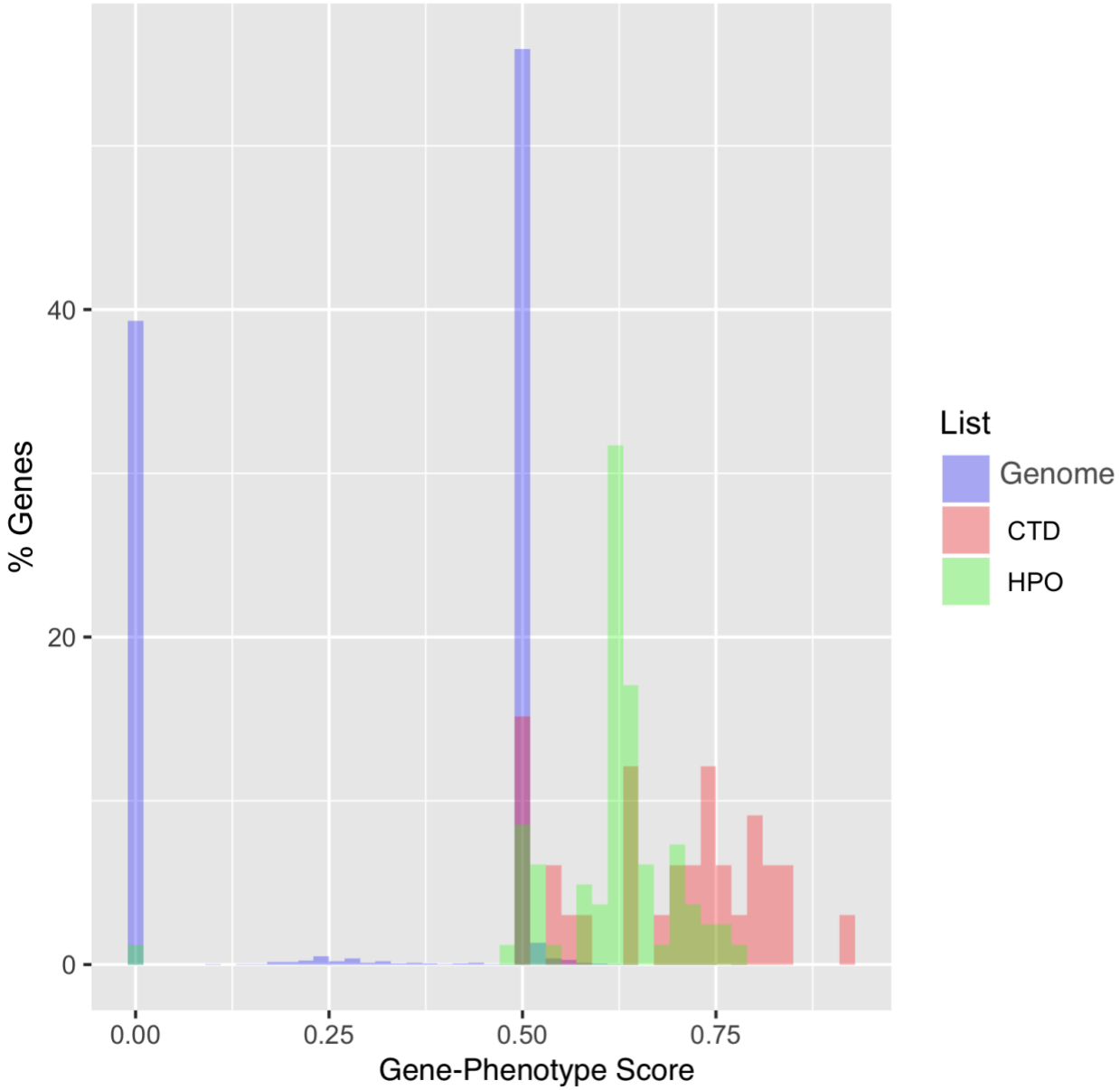
Patient	Age	Sex	CVD risk	Recurrent	**M.V.**	**Variant Type**	**Gene**	**rsID**	**HGVSp**	**CADD**	**Median Rank**	**gnomAD popmax**
401.0082	45	F	5/6	Y	N	Frameshift	SLC2A10	chr20:45354153 AC>A	Trp162GlyfsTer83	16.6	n/a	0
						Missense	ITGB4	rs1179032756	Cys598Arg	28.3	0.97049	0
401.0050	44	M	0/6	N	N	Missense	TGFBR1	chr9:101867573 C>T	Pro29Leu	18.88	0.535625	0
						Missense	LEMD3	chr12:65634744 T>C	Ser728Pro	28.4	0.88414	0
407.0059	43	F	1/5	Y	N	Missense	SMAD3	rs770798158	Ala112Thr	27.1	0.91843	3.27e-5
						Missense	COL27A1	rs1440709540	Gly1415Ser	25	0.968545	0
						3' UTR	COL27A1	rs779351956	n/a	16.31	n/a	1.20e-3
						Missense	LTBP4	rs573310430	Arg954Cys	29.7	0.881545	6.46e-5
						Missense	LTBP4	rs200667255	Pro1037Arg	25.7	0.806315	1.74e-3
202.0004	36	F	0/6	Y	N	Promoter	LRP1	rs35282763	n/a	20.6	n/a	6.68e-5
						5' UTR	COL5A1	chr9:137533804 G>A	n/a	16.37	n/a	0

Detailed description of variants that could lead to SCAD in an oligogenic manner and the clinical characteristics of individuals harboring each variant.

Patient – patient ID#, *Age* – age at admission to hospital with ACS, *Sex* – biological sex Male or Female, *CVD Risk* – number of cardiovascular risk factors (defined as obesity, dyslipidemia, smoking, hypertension, diabetes, and where available, family history of early MI), *Recurrent* – Yes or No previous MI or a cardiac rehospitalization within a one year follow up period, *MV (multivessel)* – Yes or No angiogram revealed multiple vessels were dissected, *Variant Type* – variant consequence according to the VEP canonical transcript, *Gene* - affected gene, *rsID* – rs number of the variant or GRCh37 coordinates and allele if variant is novel, *HGVSp* – amino acid change according to the VEP canonical transcript, *CADD* – the CADD phred score (see table S1 for 80$_{th}$ percentile thresholds), *Median Rank* – the median rank score of four algorithms selected to determine pathogenicity of missense variants, *gnomAD popmax* – the maximum allele frequency of all populations in the gnomAD database

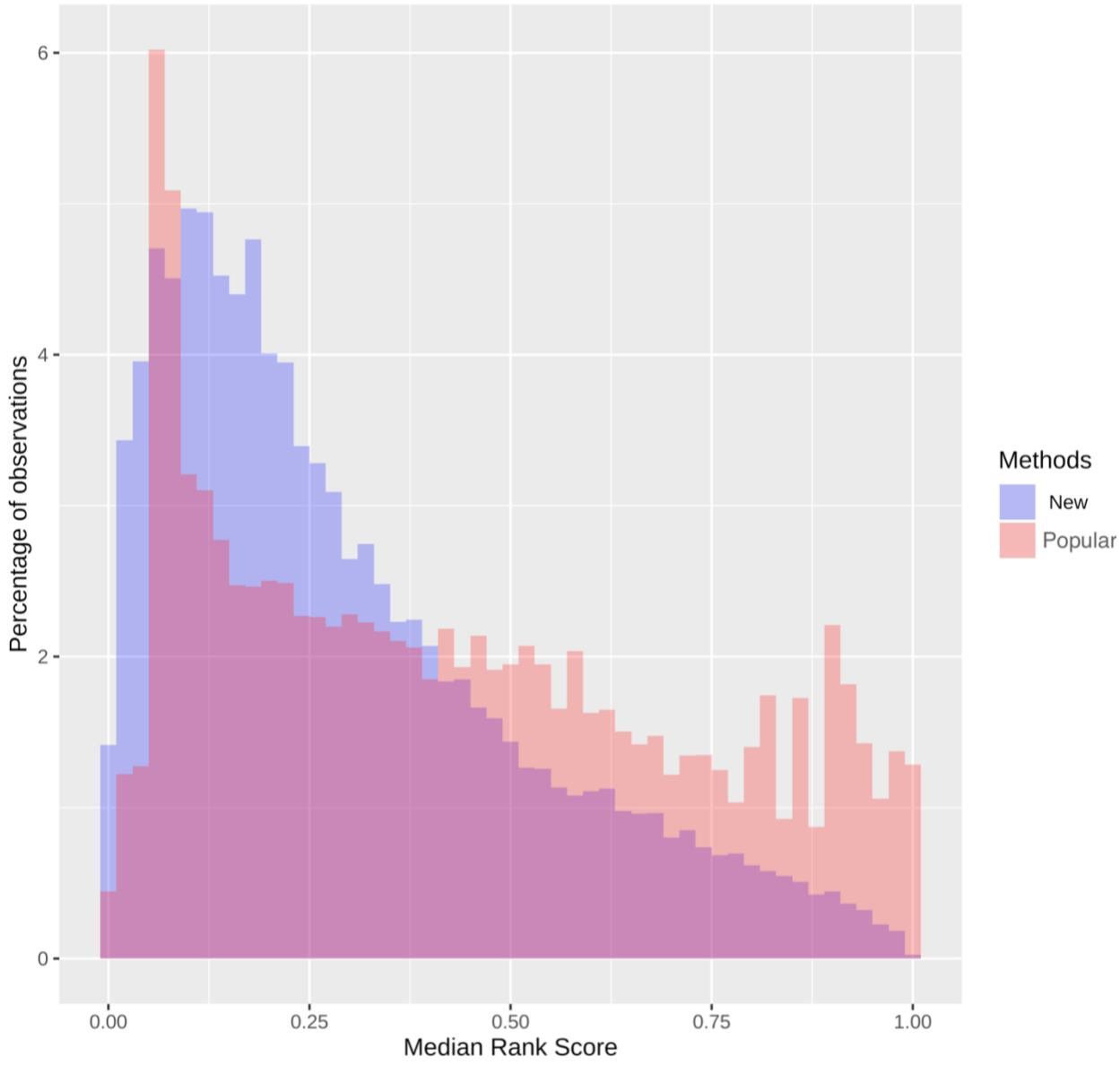
Supplemental Figures

Figure S1 - Distribution of Exomiser Gene-Phenotype Scores



Distribution of Exomiser gene-phenotype scores for SCAD according to prior evidence of a causal relationship between the gene and SCAD. The CTD list was taken from the panel used by Kaadan *et al*, which is a list of 29 genes with the highest evidence of causality. As expected, this list had the highest median score (0.7138). The HPO list was generated by seraching for all genes associated with the terms "dissection", "aneurysm", and "arterial tortuosity" in the HPO database. The 29 genes in the CTD list were excluded from this list for display purposes. The median score of genes on this list was 0.6244. Finally, the Genome list is all genes in the genome not represented in the HPO or CTD list. As expected, this list had the lowest median score (0.5). These distributions demonstrate that the Exomiser gene-phenotype scores are well-calibrated, and validates the method used to generate them.

Figure S2 – Median Rank Score Distribution for Missense Variants Using Popular Method and New Method



Distributions of the median rank score assigned to each missense variant by the New method (VEST4, MetaSVM, REVEL, M-CAP) and by the popular method (SIFT, PolyPhen, CADD, and MutationTaster2).

Supplemental References

1. Poplin R, Ruano-Rubio V, DePristo MA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. 2018:201178.
2. Liu X, White S, Peng B, et al. WGSA: an annotation pipeline for human genome sequencing studies. J Med Genet. 2016;53(2):111-112.
3. Hunt SE, McLaren W, Gil L, et al. Ensembl variation resources. Database (Oxford). 2018;2018.
4. Dreos R, Ambrosini G, Groux R, Cavin Perier R, Bucher P. The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. Nucleic Acids Res. 2017;45(D1):D51-D55.
5. Smedley D, Jacobsen JO, Jager M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc. 2015;10(12):2004-2015.
6. Kohler S, Carmody L, Vasilevsky N, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res. 2019;47(D1):D1018-D1027.
7. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-443.
8. Tian S, Yan H, Neuhauser C, Slager SL. An analytical workflow for accurate variant discovery in highly divergent regions. BMC Genomics. 2016;17:703.
9. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-291.
10. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-D1067.
11. Kaadan MI, MacDonald C, Ponzini F, et al. Prospective Cardiovascular Genetics Evaluation in Spontaneous Coronary Artery Dissection. Circ Genom Precis Med. 2018;11(4):e001933.
12. Hamano Y, Grunkemeyer JA, Sudhakar A, et al. Determinants of vascular permeability in the kidney glomerulus. J Biol Chem. 2002;277(34):31154-31162.
13. Guo DC, Regalado ES, Gong L, et al. LOX Mutations Predispose to Thoracic Aortic Aneurysms and Dissections. Circ Res. 2016;118(6):928-934.
14. Zhang HF, Zhao KJ, Xu Y, et al. Lysyl oxidase polymorphisms and ischemic stroke--a case control study. Mol Biol Rep. 2012;39(10):9391-9397.
15. Turley TN, O'Byrne MM, Kosel ML, et al. Identification of Susceptibility Loci for Spontaneous Coronary Artery Dissection. JAMA Cardiol. 2020.
16. Li F, Li F, Chen W. Propofol Inhibits Cell Proliferation, Migration, and Invasion via mir-410-3p/Transforming Growth Factor-beta Receptor Type 2 (TGFBR2) Axis in Glioma. Med Sci Monit. 2020;26:e919523.
17. Shoulders MD, Raines RT. Collagen structure and stability. Annu Rev Biochem. 2009;78:929-958.
18. Ryyanen J, Sollbert S, Parente MG, Chung LC, Christiano AM, Uitto J. Type VII collagen gene expression by cultured human cells and in fetal skin. Abundant mRNA and protein levels in epidermal keratinocytes. Journal of Clinical Investigation. 1992;89(1):163-168.
19. Varki R, Sadowski S, Uitto J, Pfendner E. Epidermolysis bullosa. II. Type VII collagen mutations and phenotype-genotype correlations in the dystrophic subtypes. J Med Genet. 2007;44(3):181-192.

20. Jefferies JL, Towbin J, Ryan T, Lucky A. Cardiovascular Findings in Recessive Dystrophic Epidermolysis Bullosa. Journal of the American College of Cardiology. 2013;61(10).
21. Barker DF, Hostikka SL, Zhou J, et al. Identification of mutations in the COL4A5 collagen gene in Alport syndrome. Science. 1990;248(4960):1224-1227.
22. Savige J, Colville D, Rheault M, et al. Alport Syndrome in Women and Girls. Clin J Am Soc Nephrol. 2016;11(9):1713-1720.
23. Diez-del Hoyo F, Sanz-Ruiz R, Diez-Villanueva P, et al. A novel cardiovascular presentation of Alport Syndrome: spontaneous coronary artery dissection. Int J Cardiol. 2014;177(3):e133-134.
24. Anuwatworn A, Sethi P, Steffen K, Jonsson O, Petrasko M. Spontaneous Coronary Artery Dissection: A Rare Manifestation of Alport Syndrome. Case Rep Cardiol. 2017;2017:1705927.
25. Zou Y, Zwolanek D, Izu Y, et al. Recessive and dominant mutations in COL12A1 cause a novel EDS/myopathy overlap syndrome in humans and mice. Hum Mol Genet. 2014;23(9):2339-2352.
26. Delbaere S, Dhooge T, Syx D, et al. Novel defects in collagen XII and VI expand the mixed myopathy/Ehlers-Danlos syndrome spectrum and lead to variant-specific alterations in the extracellular matrix. Genet Med. 2020;22(1):112-123.
27. Whittaker CA, Hynes RO. Distribution and Evolution of von Willebrand/Integrin A Domains: Widely Dispersed Domains with Roles in Cell Adhesion and Elsewhere. Molecular Biology of the Cell. 2002;13:3369–3387.
28. Gerecke DR, Olson PF, Koch M, et al. Complete Primary Structure of Two Splice Variants of Collagen XII, and Assignment of a1(XII) Collagen (COL12A1), a1(IX) Collagen (COL9A1), and a1(XIX) Collagen (COL19A1) to Human Chromosome 6q12–q13. Genomics. 1997;41:236–242.
29. Kaufman CS, Butler MG. Mutation in TNXB gene causes moderate to severe Ehlers-Danlos syndrome. World J Med Genet. 2016;6(2):17-21.
30. Wang K, Zhao S, Zhang Q, et al. Whole-exome sequencing reveals known and novel variants in a cohort of intracranial vertebral-basilar artery dissection (IVAD). J Hum Genet. 2018;63(11):1119-1128.
31. D'Hondt S, Van Damme T, Malfait F. Vascular phenotypes in nonvascular subtypes of the Ehlers-Danlos syndrome: a systematic review. Genet Med. 2018;20(6):562-573.
32. Gonzaga-Jauregui C, Gamble CN, Yuan B, et al. Mutations in COL27A1 cause Steel syndrome and suggest a founder mutation effect in the Puerto Rican population. Eur J Hum Genet. 2015;23(3):342-346.
33. Gonzaga-Jauregui C, Yesil G, Nistala H, et al. Functional biology of the Steel syndrome founder allele and evidence for clan genomics derivation of COL27A1 pathogenic alleles worldwide. Eur J Hum Genet. 2020.
34. Zekavat SM, Kii E, Chaffin M, et al. Whole Exome Sequencing of Spontaneous Coronary Artery Dissection (SCAD) Cases Discovers Enrichment of Rare Disruptive Variants in Fibrillar Collagen Genes. Circulation. 2018;138.
35. Belbin GM, Odgis J, Sorokin EP, et al. Genetic identification of a common collagen disease in puerto ricans via identity-by-descent mapping in a health system. Elife. 2017;6.

36. Urban Z, Hucthagowder V, Schurmann N, et al. Mutations in LTBP4 cause a syndrome of impaired pulmonary, gastrointestinal, genitourinary, musculoskeletal, and dermal development. Am J Hum Genet. 2009;85(5):593-605.
37. Callewaert B, Su CT, Van Damme T, et al. Comprehensive clinical and molecular analysis of 12 families with type 1 recessive cutis laxa. Hum Mutat. 2013;34(1):111-121.
38. Schwaerzer GK, Hiepen C, Schrewe H, et al. New insights into the molecular mechanism of multiple synostoses syndrome (SYNS): mutation within the GDF5 knuckle epitope causes noggin-resistance. J Bone Miner Res. 2012;27(2):429-442.
39. Sengle G, Charbonneau NL, Ono RN, et al. Targeting of bone morphogenetic protein growth factor complexes to fibrillin. J Biol Chem. 2008;283(20):13874-13888.
40. Loeys BL, Chen J, Neptune ER, et al. A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. Nat Genet. 2005;37(3):275-281.
41. Lin F, Morrison JM, Wu W, Worman HJ. MAN1, an integral protein of the inner nuclear membrane, binds Smad2 and Smad3 and antagonizes transforming growth factor-beta signaling. Hum Mol Genet. 2005;14(3):437-445.
42. Hellemans J, Preobrazhenska O, Willaert A, et al. Loss-of-function mutations in LEMD3 result in osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. Nat Genet. 2004;36(11):1213-1218.
43. Cohen TV, Kosti O, Stewart CL. The nuclear envelope protein MAN1 regulates TGFbeta signaling and vasculogenesis in the embryonic yolk sac. Development. 2007;134(7):1385-1395.
44. Munger JS, Sheppard D. Cross talk among TGF-beta signaling pathways, integrins, and the extracellular matrix. Cold Spring Harb Perspect Biol. 2011;3(11):a005017.
45. Chen MH, Choudhury S, Hirata M, Khalsa S, Chang B, Walsh CA. Thoracic aortic aneurysm in patients with loss of function Filamin A mutations: Clinical characterization, genetics, and recommendations. Am J Med Genet A. 2018;176(2):337-350.
46. Turcan I, Pasmooij AM, van den Akker PC, et al. Heterozygosity for a Novel Missense Mutation in the ITGB4 Gene Associated With Autosomal Dominant Epidermolysis Bullosa. JAMA Dermatol. 2016;152(5):558-562.
47. Su H, Kim H, Pawlikowska L, et al. Reduced expression of integrin alphavbeta8 is associated with brain arteriovenous malformation pathogenesis. Am J Pathol. 2010;176(2):1018-1027.
48. Zhou Q, Yang D, Ombrello AK, et al. Early-onset stroke and vasculopathy associated with mutations in ADA2. N Engl J Med. 2014;370(10):911-920.
49. Krittanawong C, Kumar A, Johnson KW, et al. Conditions and Factors Associated With Spontaneous Coronary Artery Dissection (from a National Population-Based Cohort Study). Am J Cardiol. 2019;123(2):249-253.
50. Caorsi R, Penco F, Grossi A, et al. ADA2 deficiency (DADA2) as an unrecognised cause of early onset polyarteritis nodosa and stroke: a multicentre national study. Ann Rheum Dis. 2017;76(10):1648-1656.
51. Jiang SH, Athanasopoulos V, Ellyard JI, et al. Functional rare and low frequency variants in BLK and BANK1 contribute to human lupus. Nat Commun. 2019;10(1):2201.
52. Reddy S, Vaid T, Ganiga Sanjeeva NC, Shetty RK. Spontaneous coronary artery dissection as the first presentation of systemic lupus erythematosus. BMJ Case Rep. 2016;2016.

53. Fisher AJ, Beal PA. Effects of Aicardi-Goutieres syndrome mutations predicted from ADAR-RNA structures. RNA Biol. 2017;14(2):164-170.
54. Crow Y, Chase D, Schmidt J, et al. Characterization of Human Disease Phenotypes Associated with Mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. American Journal of Medical Genetics Part A. 2015;162:296-312.

Chapter 5: General Discussion

Our study of the genetic etiology of SCAD has identified ACMG/AMP pathogenic or likely pathogenic variant in three patients in the GENESIS-PRAXY cohort (6.7%). In addition, high-confidence VUS with an 81% probability of pathogenicity according to a Bayesian model of the ACMG criteria were identified in another 14 patients in the cohort (31%). Nine lower-confidence VUS (50% probability of pathogenicity under Bayesian ACMG) were also identified. We are the first to report potential oligogenic causes of SCAD in four patients (8.9%). 14 patients did not have any identified rare variants (31%).

The identified variants are primarily in collagen and TGF-β signaling genes, which are known to be associated with SCAD, and introduce new proteins of interest within these biochemical pathways. Additionally, our results in integrin genes support the hypothesis of integrin and focal adhesion pathways as a mechanism of SCAD recently hypothesized by Turley *et al*.[33]

We have also established that a number of conditions including epidermolysis bullosa, muscular Ehlers-Danlos syndrome, Steel syndrome, cutis laxa, brachydactyly, Buschke-Ollendorff syndrome, and brain arteriovenous malformation may be related to SCAD and associated vasculopathies. To further support these findings, patients with these conditions (as well as heterozygous carriers in some cases) could be evaluated for vascular phenotypes. In addition, future SCAD cohorts should undertake a thorough examination for clinical characteristics consistent with mild forms of these diseases to corroborate potential genetic findings. In addition to detailed phenotyping, biomarker assays could assist in validating clinical findings of a VUS. For example, the finding of abnormal collagen organization in conjunction with a VUS in a collagen gene would provide functional evidence for the variant to increase the

likelihood that a finding is definitive in the absence of familial segregation data. Another simple, cost-effective validation assay would be to measure circulating TGF-β levels in patients. Recent work in the field of aortic dissection has identified a number of other biomarker candidates that would be interesting to investigate for their potential utility in SCAD.[42]

Our work demonstrates the possible utility of WGS in a clinical setting. Six patients (13.3%) in our cohort had a non-coding variant of interest, which would have been missed using exome sequencing. We are the first to report that rare non-coding variation could contribute to SCAD in a Mendelian fashion. Many identified non-coding variants were in the regulatory regions of genes in the TGF-β pathway, suggesting future work could focus on the functional validation of these variants. Importantly, WGS identified a missense variant in exon one of *TGFBR1* that was not identified when WES was applied to the same patient.

My work has implications for the ACMG/AMP pathogenicity criteria across all diseases, not just SCAD. The guidelines, which were developed in 2015, could be updated to reflect certain advances in bioinformatic methods. The guidelines currently do not issue guidance on which combinations of *in silico* prediction scores and thresholds maximize the sensitivity of pathogenic variant detection while ensuring a sufficiently low false positive rate that researchers can be more confident in the results from these algorithms. In addition, research is required to understand the contribution of non-coding and regulatory variation to pathogenesis. Better annotation tools for non-coding variants including information regarding functional domains in UTRs such as upstream open reading frames (uORF) and internal ribosome entry sites would allow researchers to more easily propose a potential disease mechanism that could be confirmed biochemically. A thorough set of ACMG guidelines describing steps to confirm pathogenic regulatory variation in genes where a disease-gene relationship has already been established for

coding variants will aid future researchers in identifying such variants for many diseases. In addition, updating the language in certain criteria to reflect incomplete penetrance and variable expressivity could increase the yield of clinically actionable results of this study from 7% to 33%.

Although the small size of the cohort does not allow for a statistical comparison, I found that all three patients with a pathogenic or likely pathogenic genetic finding had a potential SCAD recurrence. This trend is in line with the findings of Arnaud *et al*.[36] for TAAD. Future large-scale studies of SCAD patients with longitudinal data regarding SCAD recurrence could help validate our observation that patients with a pathogenic or likely pathogenic variant are more likely to have a recurrent event, and determine if patients with a VUS may also have an elevated risk of recurrence. In addition, I discovered that patients without an identified rare variant were significantly more likely to have a multivessel event than patients with a rare genetic variant, suggesting a possibly distinct etiology of multivessel SCAD.

We did not identify any variants in the *TSR1* or *PTGIR* genes, which were recently reported to have rare variants in SCAD patients.[43,44] This could be due to the small sample size, or perhaps varying ethnicities among the three cohorts. Our sample size of 45 SCAD patients was not appropriately powered to replicate any findings of common variants[32,33] or leverage certain analytical approaches such as gene burden testing, which determines if cases have a statistical enrichment of variants with certain characteristics within certain genes compared to controls. Gene burden testing could be used to further test the hypothesis that SCAD patients are enriched specifically for VUS in CTD genes compared to controls or determine if regulatory variation in certain pathways is more common among cases than controls, validating it as a mechanism of SCAD. An increased sample size would also allow the creation and assessment of a polygenic risk score (PRS) of common variants contributing to SCAD, which could be an

explanation for SCAD in some of the patients without an identified rare variant. Furthermore, it may be possible that a high genetic burden of common variation could increase penetrance in the presence of a rare variant. Future studies could investigate this potential interaction.

Our work does not address the dramatic enrichment of SCAD cases in females. Exogenous hormone therapy and women who have had more than four children are both relatively common among female SCAD cases (10% each) and could both present interesting avenues of investigation.[24] Turley *et al.*[33] suggest a potential mechanism for estrogen in the regulation of MMP9 and LRP1, which influence matrix remodeling. Therefore, variants in these genes may have a larger effect in women. My study did not identify any rare variants in these genes; however, this interaction could be mediated by common variants that my study did not investigate. In addition, pregnancy could serve as an underappreciated trigger of SCAD. Because some cases of SCAD happen during pregnancy, labor, or shortly peripartum,[28] it is possible that pregnancy accelerates the weakening of vessels in women with an underlying vasculopathy, causing these variants to be more penetrant in females than in males.

14 patients (31%) of our cohort did not have an identified rare genetic variant. One potential explanation is that some of these patients had a Copy Number Variation, which we were not able to detect in our study. We attempted CNV calling using the GATK gCNV software[45] on our SCAD cohort with seven additional samples from the Genome in a Bottle project,[46] which have had CNV calling performed on the basis of long read sequencing data from PacBio,[47] as a validation set. Unfortunately, the gCNV calling method captured only 50% of the deletion events detected using PacBio and had an extremely high false discovery rate (95%). By constraining the data to only calls with quality metrics within a region enriched for validated calls, we were able to achieve only 7.2% sensitivity with a 27.4% false discovery rate. I

hypothesize that the poor performance could be due to the fact that my analysis included roughly half the 100 individuals suggested by the gCNV documentation, or that multiple ethnicities within our cohort significantly decreased the model's performance. While it is unknown how prevalent CNV variations will be for the etiology SCAD, an improvement of CNV calling methods to accommodate small cohorts will benefit not only SCAD research, but all rare disease research.

Another reason that patients may not have had an identified variant is that SCAD and related vasculopathies are not well represented in the ClinVar, OMIM, and HPO databases that Exomiser relies upon to generate gene-phenotype scores. Without sufficient information, novel genes related to SCAD may have gone undetected using our methodology. Similar to other diseases, additional studies of rare variation will contribute the necessary data to give Exomiser and similar tools better power.

Finally, future studies of SCAD must account for the effects of the COVID-19 pandemic. COVID-19 is known to cause a range of cardiac manifestations. Recently, case reports of SCAD with concurrent COVID-19 infection have been reported.[48,49,50] It is unclear whether this co-occurrence is coincidental given the widespread nature of COVID-19, however, there is etiological evidence supporting a link through systemic inflammation and enhanced thrombosis in COVID-19. It will be important to further study patients presenting with SCAD and COVID-19 to understand if the infection served as a trigger in a patient with an existing genetic pre-disposition, or if the infection itself is the sole cause of SCAD. In any event, future genetic studies should include the COVID-19 history of the patients to aid in the discovery of genetic causes. Furthermore, there are recent reports of a Kawasaki-like disease response in some

children infected with COVID-19.[51] These children could now have an increased life-long risk of SCAD, which make it more prevalent in the future.

In summary, my study is a thorough evaluation of the contribution of rare genetic variation to SCAD. I developed an approach to investigate both coding and non-coding variation in genes likely to be relevant to SCAD, whether they had been previously reported in the literature or not, utilizing modern bioinformatic tools. My approach did not rely on data from family members, as the incomplete penetrance and variable expressivity of SCAD makes the collection of this data unrealistic in a routine healthcare context. I have provided avenues for future research, including thorough characterization of the rates of VUS in CTD genes among SCAD patients, the rates of recurrence in SCAD patients with a monogenic finding versus those without, and confirmation that regulatory variation, especially in the TGF-β pathway contributes to SCAD pathogenesis. Ultimately, my study supports the idea that SCAD and other vasculopathies are part of the spectrum of phenotypes related to connective tissue disorders with variable expressivity.

Chapter 6: Additional References

1. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-D1067.
2. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet. 2014;133(1):1-9.
3. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-424.
4. Amendola LM, Jarvik GP, Leo MC, et al. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. Am J Hum Genet. 2016;98(6):1067-1076.
5. McClellan J, King MC. Genetic heterogeneity in human disease. Cell. 2010;141(2):210-217.
6. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A. 2009;106(45):19096-19101.
7. Albert TJ, Molla MN, Muzny DM, et al. Direct selection of human genomic loci by microarray hybridization. Nat Methods. 2007;4(11):903-905.
8. Belkadi A, Bolze A, Itan Y, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci U S A. 2015;112(17):5473-5478.
9. Wang VG, Kim H, Chuang JH. Whole-exome sequencing capture kit biases yield false negative mutation calls in TCGA cohorts. PLoS One. 2018;13(10):e0204912.
10. Barbitoff YA, Polev DE, Glotov AS, et al. Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage. Sci Rep. 2020;10(1):2057.
11. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med. 2019;12(1):3.
12. Rivera-Munoz EA, Milko LV, Harrison SM, et al. ClinGen Variant Curation Expert Panel experiences and standardized processes for disease and gene-level specification of the ACMG/AMP guidelines for sequence variant interpretation. Hum Mutat. 2018;39(11):1614-1622.
13. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum Genet. 2013;132(10):1077-1130.
14. Gloss BS, Dinger ME. Realizing the significance of noncoding functionality in clinical genomics. Exp Mol Med. 2018;50(8):97.

15. Makhnoon S, Garrett LT, Burke W, Bowen DJ, Shirts BH. Experiences of patients seeking to participate in variant of uncertain significance reclassification research. J Community Genet. 2019;10(2):189-196.
16. Mitchell KJ. What is complex about complex disorders? Genome Biol. 2012;13.
17. Youngblom E, Pariani M, Knowles J. Familial Hypercholesterolemia. GeneReviews® https://wwwncbinlmnihgov/books/NBK174884/. 2014.
18. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018;19(9):581-590.
19. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219-1224.
20. Katsanis N. The continuum of causality in human genetic disorders. Genome Biol. 2016;17(1):233.
21. Tamis-Holland JE, Jneid H, Reynolds HR, et al. Contemporary Diagnosis and Management of Patients With Myocardial Infarction in the Absence of Obstructive Coronary Artery Disease: A Scientific Statement From the American Heart Association. Circulation. 2019;139(18):e891-e908.
22. Raparelli V, Elharram M, Shimony A, Eisenberg MJ, Cheema AN, Pilote L. Myocardial Infarction With No Obstructive Coronary Artery Disease: Angiographic and Clinical Insights in Patients With Premature Presentation. Can J Cardiol. 2018;34(4):468-476.
23. Yip A, Saw J. Spontaneous coronary artery dissection-A review. Cardiovasc Diagn Ther. 2015;5(1):37-48.
24. Saw J, Aymong E, Sedlak T, et al. Spontaneous coronary artery dissection: association with predisposing arteriopathies and precipitating stressors and cardiovascular outcomes. Circ Cardiovasc Interv. 2014;7(5):645-655.
25. Hassan S, Prakash R, Starovoytov A, Saw J. Natural History of Spontaneous Coronary Artery Dissection With Spontaneous Angiographic Healing. JACC Cardiovasc Interv. 2019;12(6):518-527.
26. Bonnet M, Toleva O, Ybarra LF, Rinfret S. Successful Treatment of Spontaneous Coronary Artery Dissection With Subintimal Tracking and Re-Entry Technique. JACC: Case Reports. 2019;1(4):553-559.
27. Saw J, Humphries K, Aymong E, et al. Spontaneous Coronary Artery Dissection: Clinical Outcomes and Risk of Recurrence. J Am Coll Cardiol. 2017;70(9):1148-1158.
28. Hayes SN, Kim ESH, Saw J, et al. Spontaneous Coronary Artery Dissection: Current State of the Science: A Scientific Statement From the American Heart Association. Circulation. 2018;137(19):e523-e557.
29. Krittanawong C, Kumar A, Johnson KW, et al. Conditions and Factors Associated With Spontaneous Coronary Artery Dissection (from a National Population-Based Cohort Study). Am J Cardiol. 2019;123(2):249-253.
30. Debette S, Markus HS. The genetics of cervical artery dissection: a systematic review. Stroke. 2009;40(6):e459-466.
31. Brownstein AJ, Ziganshin BA, Kuivaniemi H, Body SC, Bale AE, Elefteriades JA. Genes Associated with Thoracic Aortic Aneurysm and Dissection: An Update and Clinical Implications. Aorta (Stamford). 2017;5(1):11-20.

32. Adlam D, Olson TM, Combaret N, et al. Association of the PHACTR1/EDN1 Genetic Locus With Spontaneous Coronary Artery Dissection. J Am Coll Cardiol. 2019;73(1):58-66.
33. Turley TN, O'Byrne MM, Kosel ML, et al. Identification of Susceptibility Loci for Spontaneous Coronary Artery Dissection. JAMA Cardiol. 2020.
34. Saw J, Mancini GBJ, Humphries KH. Contemporary Review on Spontaneous Coronary Artery Dissection. J Am Coll Cardiol. 2016;68(3):297-312.
35. Kaadan MI, MacDonald C, Ponzini F, et al. Prospective Cardiovascular Genetics Evaluation in Spontaneous Coronary Artery Dissection. Circ Genom Precis Med. 2018;11(4):e001933.
36. Arnaud P, Hanna N, Benarroch L, et al. Genetic diversity and pathogenic variants as possible predictors of severity in a French sample of nonsyndromic heritable thoracic aortic aneurysms and dissections (nshTAAD). Genet Med. 2019;21(9):2015-2024.
37. von Hundelshausen P, Oexle K, Bidzhekov K, et al. Recurrent spontaneous coronary dissections in a patient with a de novo fibrillin-1 mutation without Marfan syndrome. Thromb Haemost. 2015;113(3):668-670.
38. Saw J. Spontaneous coronary artery dissection. Can J Cardiol. 2013;29(9):1027-1033.
39. Barrett LW, Fletcher S, Wilton SD. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cell Mol Life Sci. 2012;69(21):3613-3634.
40. Loeys BL, Chen J, Neptune ER, et al. A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. Nat Genet. 2005;37(3):275-281.
41. Schwarze U, Schievink WI, Petty E, et al. Haploinsufficiency for One COL3A1 Allele of Type III Procollagen Results in a Phenotype Similar to the Vascular Form of Ehlers-Danlos Syndrome, Ehlers-Danlos Syndrome Type IV. Am J Hum Genet. 2001;69:989-1001.
42. Bossone E, Czerny M, Lerakis S, et al. Imaging and Biomarkers in Acute Aortic Syndromes: Diagnostic and Prognostic Implications. Current Problems in Cardiology. 2020.
43. Sun Y, Chen Y, Li Y, et al. Association of TSR1 Variants and Spontaneous Coronary Artery Dissection. J Am Coll Cardiol. 2019;74(2):167-176.
44. Georges A, Albuisson J, Berrandou T, et al. Rare Loss-of-function Mutations of PTGIR are enriched in Fibromuscular Dysplasia. Cardiovasc Res. 2020.
45. Babadi M, Lee SK, Smirnov A, et al. Precise common and rare germline CNV calling with GATK. Cancer Res. 2018;78:Abstract nr 2287.
46. Zook JM, Catoe D, McDaniel J, et al. Extensive sequencing of seven human genomes to characterize benchmark reference materials. Sci Data. 2016;3:160025.
47. Zook JM, Hansen NF, Olson ND, et al. A robust benchmark for detection of germline large deletions and insertions. Nat Biotechnol. 2020.
48. Kumar K, Vogt JC, Divanji PH, Cigarroa JE. Spontaneous coronary artery dissection of the left anterior descending artery in a patient with COVID-19 infection. Catheter Cardiovasc Interv. 2020.
49. Courand PY, Harbaoui B, Bonnet M, Lantelme P. Spontaneous Coronary Artery Dissection in a Patient With COVID-19. JACC Cardiovasc Interv. 2020;13(12):e107-e108.

50. Fernandez Gasso L, Maneiro Melon NM, Sarnago Cebada F, Solis J, Garcia Tejada J. Multivessel spontaneous coronary artery dissection presenting in a patient with severe acute SARS-CoV-2 respiratory infection. Eur Heart J. 2020.
51. Viner RM, Whittaker E. Kawasaki-like disease: emerging complication during the COVID-19 pandemic. The Lancet. 2020;395(10239):1741-1743.

Appendix 1

Table Caption

Patient – patient ID

Exomiser Rank – Ranking that Exomiser automatically assigned the variant

Variant (g.) – Coordinates of the variant in GRCh37

Variant (p.) – Amino acid consequence of the variant in the Ensembl canonical transcript

Gene – Gene name

Genotype – Whether variant is found in heterozygous or homozygous state in the patient

Variant Score – Exomiser assigned score for the relative consequence of the variant (0 is least damaging, 1 is most damaging)

Phenotype Score – Exomiser assigned score for the similarity of the input Human Phenotype Ontology (HPO)[1] list to the HPO list of the proposed condition (0 is least similar, 0.5 is similarity through protein interaction, 1 is perfect match of all HPO terms)

Exomiser Strongest Phenotypic Evidence – Exomiser's proposal for relationship between variant and disease

Closest HPO match – HPO term that Exomiser identified as most similar to SCAD and other vasculopathies

gnomAD popmax – Maximum of all population allele frequencies in gnomAD (v2.1.1)

gnomAD controls popmax – Maximum of all population allele frequencies within gnomAD (v2.1.1) calculated from the controls only

ClinVar – Previous reports of disease causality for this amino acid position within the ClinVar database (P = pathogenic, LP = likely pathogenic, VUS = variant of unknown significance, LB = likely benign, B = benign)

ACMG – American College of Medical Genetics rating[4] automatically assigned by the software InterVar and verified by one of us (HB). P = pathogenic, LP = likely pathogenic, VUS = variant of unknown significance, LB = likely benign, B = benign

Patient	Rank	Variant (g.)	Variant (p.)	Gene	Genotype	Variant Score	Phenotype Score
101-0069							
	1	15:48888551 T>C	Asn156Ser	FBN1	Het	0.999	0.794
	2	22:19969502 G>A	Ser108Phe	ARVCF	Het	0.999	0.539
	3	5:82837788 G>T	Gly2989Val	VCAN	Het	1	0.506
	4	16:84256595 A>G	Phe263Ser	KCNG4	Het	1	0.504
	5	3:33725854 G>C	Ala214Gly	CLASP2	Het	1	0.504
401-0050							
	1	12:65634744 T>C	Ser728Pro	LEMD3	Het	1	0.571
	2	2:47641558 GTAAAAA>G	Splice donor	MSH2	Het	1	0.508
	3	3:25666190 C>A	Ala772Ser	TOP2B	Het	1	0.507
	4	5:83657823 G>A	Arg18Lys	RAMAC	Het	1	0.503
	5	15:44062746 G>A	Ala458Thr	PDIA3	Het	1	0.502
503-0003							
	1	3:12660070 G>A	Pro51Ser	RAF1	Het	1	0.541
	2	9:78790187 G>C	Trp681Ser	PCSK5	Het*	1	0.507
	2	9:78973431 G>A	Glu1726Lys	PCSK5	Het*	1	0.507
	3	10:24873805 C>CT	Gly1805Argfs*19	ARHGAP21	Het	1	0.501
	3	14:96707428 G>T	Glu255*	BDKRB2	Het	1	0.501
505-0033							
	1	3:12447537 C>T	Ala259Val	PPARG	Het	1	0.528
	2	3:126724000 C>T	Thr604Met	PLXNA1	Het	0.999	0.516
	3	16:2138583 C>T	Ser1799Leu	TSC2	Het	0.999	0.508
	4	12:49959389 T>C	Thr84Ala	MCRS1	Het	0.999	0.507
	5	3:124485075 GGCT>G	Glu711_Pro712delinsAla	ITGB5	Het	1	0.503
701-0002							
	1	1:154560700 T>C	Lys974Glu	ADAR1	Het	1	0.615
	2	1:118166560 A>G	Tyr357Cys	TENT5C	Het	1	0.525
	3	9:134357163 G>T	Ala1651Ser	PRRC2B	Het	1	0.524
	4	1:47882541 C>T	Pro185Leu	FOXE3	Het	0.775	0.72
	5	9:78790187 G>C	Trp681Ser	PCSK5	Het	1	0.507
* Potential compound heterozygous							

Patient	Rank	Exomiser Strongest Phenotypic Evidence	Closest HPO match
101-0069			
	1	HPO similarity to FBN1 mouse mutant and Marfan syndrome	Aortic dissection
	2	HPO similarity to 22q11.2 deletion syndrome	Abnormal aortic arch morphology
	3	Interaction with FBN1 (Marfan)	Aortic dissection
	4	Interaction with KCNH1 (Zimmerman-Laband syndrome)	Aortic arch aneurysm
	5	Interaction with ARFGEF2 (Periventricular nodular heterotopia)	Aortic aneurysm
401-0050			
	1	HPO similarity to Buschke-Ollendorff syndrome	Abnormal aortic morphology
	2	Interaction with RNASEH2A (Aicardi-Goutières syndrome)	Aortic aneurysm
	3	Interaction with NSMCE2 (Seckel syndrome)	Abdominal aortic aneurysm
	4	Interaction with FMR1 (Fragile X syndome)	Ascending tubular aorta aneurysm
	5	Interaction with HLA-B (Takayasu arteritis)	Ascending tubular aorta aneurysm
503-0003			
	1	HPO similarity to Noonan Syndrome	Pulmonary artery stenosis
	2	Interaction with PCSK9 (Familial hypercholesterolemia)	Coronary artery aneurysm
	2	Interaction with PCSK9 (Familial hypercholesterolemia)	Coronary artery aneurysm
	3	Interaction with ARF1 (Periventricular nodular heterotopia)	Aortic aneurysm
	3	Interaction with S1PR1 (mouse mutant phenotype)	Abnormal aorta morphology (mouse)
505-0033			
	1	Interaction with PRDM16 (1p36 deletion syndrome)	Aortic arch aneurysm
	2	Interaction with SEMA3E (CHARGE syndrome)	Aortic arch aneurysm
	3	Interaction with PKD1 (Polycystic Kidney Disease)	Cerebral berry aneurysm
	4	Interaction with KANSL1 (Koolen-De Vries syndrome)	Aortic aneurysm
	5	Interaction with COL5A1 (Ehlers-Danlos syndrome)	Arterial dissection
701-0002			
	1	HPO similarity to Aicardi-Goutières syndrome	Aortic aneurysm
	2	HPO similarity to TENT5C mouse mutant	Additional anastomosis between intracranial vertebral arteries
	3	Interaction with ATP7A (Occiptal Horn Syndrome)	Carotid artery tortuosity
	4	HPO similarity to Familial thoracic aortic aneurysm 11	Aortic dissection
	5	Interaction with PCSK9 (Familial hypercholesterolemia)	Coronary artery aneurysm

Patient	Rank	gnomAD popmax	gnomAD controls popmax	ClinVar	ACMG
101-0069					
	1	6.2E-04	1.4E-04	VUS cardiovascular (449605, 519790, 200203)	VUS
	2	6.2E-05	2.3E-05	-	VUS
	3	-	-	-	VUS
	4	-	-	-	VUS
	5	4.9E-05	7.0E-05	-	VUS
401-0050					
	1	-	-	-	VUS
	2	-	-	-	VUS
	3	2.6E-05	-	-	VUS
	4	-	-	-	VUS
	5	-	-	-	VUS
503-0003					
	1	-	-	-	VUS
	2	-	-	-	VUS
	2	1.2E-02	1.4E-02	-	VUS
	3	-	-	-	VUS
	3	-	-	-	VUS
505-0033					
	1	-	-	-	VUS
	2	8.9E-06	-	-	VUS
	3	5.7E-05	4.7E-05	LB (450528)	VUS
	4	5.4E-05	1.1E-04	-	VUS
	5	1.8E-05	2.3E-05	-	VUS
701-0002					
	1	-	-	VUS Dyschromatosis Symmetrica Hereditaria (292763)	VUS
	2	-	-	-	VUS
	3	-	-	-	VUS
	4	-	-	-	VUS
	5	-	-	-	VUS

Appendix 2

Table Caption

Patient – Patient ID

Chr – Chromosome

Position – Genomic coordinates in GRCh37

Reference Allele – Reference allele in GRCh37

Alternate Allele – Alternate allele in GRCh37

Gene – Gene in which the variant occurs, according to VEP Ensembl worst consequence transcript

VEP Worst Consequence – Variant function on the VEP Ensembl worst consequence transcript

EPD Promoter – Gene the variant regulates according to the EPD database

Exomiser Gene-Phenotype Score – Exomiser assigned score for the similarity of the input Human Phenotype Ontology (HPO) list to the HPO list of the proposed matching condition (0 is least similar, 0.5 is similarity through protein interaction, 1 is perfect match of all HPO terms)

GTEx Coronary Artery (rpkm) – Median expression level of the gene in coronary artery tissue reported in GTEx in units of reads per kilobase of transcript, per million (rpkm)

Patient	Chr	Position	Reference Allele	Alternate Allele	Gene	VEP Worst Consequence	EPD Promoter	Exomiser Gene-Phenotype Score	GTEx Coronary Artery (rpkm)
101.0053	2	9630311	ACT	A	ADAM17	frameshift_variant	#N/A	0.5045	13.97
101.0053	7	20418829	C	T	ITGB8	missense_variant	#N/A	0.5031	5.128
101.0058	6	51935807	T	C	PKHD1	missense_variant	#N/A	0.5079	0.005456
101.0058	13	52518281	G	T	ATP7B	missense_variant	#N/A	0.5028	3.434
101.0058	9	139395101	C	T	NOTCH1	missense_variant	#N/A	0.6415	22.41
101.0058	15	58830599	C	A	LIPC	missense_variant	#N/A	0.514	0.2772
101.0058	4	15709246	G	A	BST1	stop_gained	#N/A	0.5033	9.351
101.0058	17	7189147	C	T	SLC2A4	missense_variant	#N/A	0.5175	7.039
101.0058	1	120056975	G	A	HSD3B1	missense_variant	#N/A	0.5086	0
101.0058	4	183675802	C	T	TENM3	missense_variant	#N/A	0.5036	1.395
101.006	2	178969169	GTC	G	PDE11A	frameshift_variant	#N/A	0.5028	0.2422
101.006	17	79767715	G	A	GCGR	missense_variant	#N/A	0.502	0.2379
101.006	16	81298282	C	T	BCMO1	missense_variant	#N/A	#N/A	0.01195
101.006	3	12475396	G	A	PPARG	splice_acceptor_variant	#N/A	0.528	11.85
101.006	19	40321181	C	T	DYRK1B	missense_variant	#N/A	0.5002	13.69
101.006	3	51743300	C	T	GRM2	stop_gained	#N/A	0.5026	0.4911
101.013	20	5294957	TG	T	PROKR2	frameshift_variant	#N/A	0.5056	0
101.013	3	123457796	C	T	MYLK	missense_variant	#N/A	0.7482	138.8
101.013	1	218519398	GCA	GCACA	RP11-224O19.2	promoter/upstream_gene_variant	TGFB2	0.8167	9.631
101.013	1	36766517	ATCC	A	THRAP3	inframe_deletion	#N/A	0.5017	72.9
101.0132	9	136218824	CAG	C	SURF1	frameshift_variant	#N/A	0.5036	56.83
101.0132	16	15815346	C	T	MYH11	missense_variant	#N/A	0.7207	2514
101.0132	15	42115876	C	A	MAPKBP1	stop_gained	#N/A	0.5035	12.74
101.0132	16	3708230	A	T	TRAP1	missense_variant	#N/A	0.5509	33.61
101.0133	X	77302388	A	G	ATP7A	3_prime_UTR_variant	#N/A	0.7558	5.608
101.0133	X	146993567	C	CGCGGCGGCGGCGGCGGCGGCGGCGGAGGCG	FMR1	promoter/5_prime_UTR_variant	FMR1	0.6576	22.46

101.0133	8	72128947	G	A	EYA1	missense_variant	#N/A	0.5739	0.07986
101.0133	14	50628210	G	A	SOS2	missense_variant	#N/A	0.5411	19.58
101.0133	17	1381984	C	A	MYO1C	missense_variant	#N/A	0.5017	150.9
102.0015	3	164764786	A	C	SI	missense_variant	#N/A	0.5236	0
102.0015	3	48621017	G	A	COL7A1	missense_variant	#N/A	0.5038	9.315
102.0015	X	66765463	G	A	AR	missense_variant	#N/A	0.5026	7.796
102.0015	6	15496952	C	T	JARID2	missense_variant	#N/A	0.5023	7.005
201.0088	2	233406191	C	CA	CHRNG	frameshift_variant	#N/A	0.6233	0.01986
201.0088	8	11617152	T	C	GATA4	3_prime_UTR_variant	#N/A	0.5388	41.37
201.0088	16	89598369	G	A	SPG7	missense_variant	#N/A	0.5034	27.87
201.0088	1	218519398	GCA	GCACACACACACA	RP11-224O19.2	promoter/upstream_gene_variant	TGFB2	0.8167	9.631
201.0088	15	90628270	G	T	IDH2	missense_variant	#N/A	0.5025	47.72
201.0106	15	48725182	A	G	FBN1	missense_variant	#N/A	0.7943	63.59
201.0106	16	53636007	CTG	C	RPGRIP1L	stop_gained	#N/A	0.5201	1.568
201.0106	3	148897356	A	G	CP	missense_variant	#N/A	0.5111	22.73
201.0106	11	102667834	G	T	MMP1	missense_variant	#N/A	0.5043	0.1381
201.0106	16	23593613	A	C	NDUFAB1	missense_variant	#N/A	0.5036	62.72
201.0106	2	183070763	C	T	PDE1A	missense_variant	#N/A	0.5027	24.06
202.0004	7	151878445	C	T	KMT2C	missense_variant	#N/A	0.5077	13.22
301.0031	14	81610363	TCA	T	TSHR	frameshift_variant	#N/A	0.502	0.07103
301.0031	X	107823941	G	C	COL4A5	missense_variant	#N/A	0.504	23.58
301.0031	6	43488016	A	G	POLR1C	missense_variant	#N/A	0.5029	18.01
401.0021	20	18505624	C	T	SEC23B	stop_gained	#N/A	0.5035	21.07
401.0021	6	80878706	CAG	C	BCKDHB	frameshift_variant	#N/A	0.5006	15.87
401.0021	1	218615763	T	A	TGFB2	3_prime_UTR_variant	#N/A	0.8167	9.631
401.0021	19	4171788	G	A	CREB3L3	stop_gained	#N/A	0.5022	0.03988
401.0027	16	1636277	C	T	IFT140	splice_acceptor_variant	#N/A	0.5486	10.63
401.0027	7	74072002	AAAAAG	GAAAAG	GTF2I	promoter/non_coding_transcript_exon_variant	GTF2I	0.6115	16.16
401.0027	7	99705216	AGCCGGGGGCCGAGGT	A	TAF6	inframe_deletion	#N/A	0.5066	30.47

401.0027	2	130877828	GTCGTGGTCTCCAGAAGCGCCCACGTTGCTCTTGCCA	G	POTEF	inframe_deletion	#N/A	0.502	0.07754
401.0027	5	176662827	C	T	NSD1	missense_variant	#N/A	0.5082	12.83
401.0027	11	71184686	G	A	NADSYN1	missense_variant	#N/A	0.5034	22.84
401.0027	8	124031444	G	A	DERL1	missense_variant	#N/A	0.5018	44.22
401.005	17	41055964	C	T	G6PC	missense_variant	#N/A	0.5044	0.03242
401.005	9	101867573	C	T	TGFBR1	missense_variant	#N/A	0.8445	34.94
401.005	12	65634744	T	C	LEMD3	missense_variant	#N/A	0.5707	16.19
401.0081	9	94487082	A	G	ROR2	missense_variant	#N/A	0.53	1.758
401.0081	16	30100437	G	A	TBX6	missense_variant	#N/A	0.5024	1.451
401.0082	16	16244583	TC	T	ABCC6	frameshift_variant	#N/A	0.5801	0.7091
401.0082	20	45354153	AC	A	SLC2A10	frameshift_variant	#N/A	0.7333	11.75
401.0082	19	55608200	CGTCTTCGCCCCTTCCCCGCGCAGTTCACCACCGTCTTCGCCCCTTCCCCGCAGTTCACCACCGTCTTCGCCCCTTCCCCGCAGTTCACCACCGTCTTCGCCCCTTCCCCTGCAGTTCACCACT	C	PPP1R12C	splice_donor_variant	#N/A	0.5023	195.1
401.0082	20	18505108	A	G	SEC23B	missense_variant	#N/A	0.5035	21.07
401.0082	17	73732399	T	C	ITGB4	missense_variant	#N/A	0.5033	25.29
401.0127	1	21546501	G	A	ECE1	missense_variant	#N/A	0.5442	123.2
401.0127	17	12608444	G	T	MYOCD	splice_acceptor_variant	#N/A	0.5028	35.57

401.0127	9	126133173	G	C	CRB2	missense_variant	#N/A	0.5109	0.05506
401.0144	13	52548577	T	TG	ATP7B	frameshift_variant	#N/A	0.5028	3.434
401.0144	13	52548830	CTT	C	ATP7B	frameshift_variant	#N/A	0.5028	3.434
401.0144	6	41875003	AC	A	MED20	frameshift_variant	#N/A	0.5022	9.052
401.0144	1	92944283	C	A	GFI1	missense_variant	#N/A	0.5032	0.7923
401.0149	7	117227860	G	A	CFTR	missense_variant	#N/A	0.5025	0.04356
401.0149	5	127622473	C	T	FBN2	missense_variant	#N/A	0.6323	0.178
401.0149	12	132325368	G	C	MMP17	missense_variant	#N/A	0.5033	3.059
401.0158	10	91005469	G	A	LIPA	stop_gained	#N/A	0.5245	33.51
401.0158	2	189873814	C	G	COL3A1	missense_variant	#N/A	0.7772	681.2
401.0158	14	69369232	C	T	ACTN1	missense_variant	#N/A	0.5027	482.8
401.0158	16	53301342	G	T	CHD9	missense_variant	#N/A	0.5019	13.02
402.0084	1	981398	G	A	AGRN	missense_variant	#N/A	0.5169	26.32
402.0084	12	4796141	G	C	NDUFA9	missense_variant	#N/A	0.5035	24.16
407.004	17	79767715	G	A	GCGR	missense_variant	#N/A	0.502	0.2379
407.004	1	155649800	C	T	YY1AP1	splice_acceptor_variant	#N/A	0.5068	39.03
407.004	1	8412855	T	C	RERE	3_prime_UTR_variant	#N/A	0.6244	69.22
407.004	14	30194748	A	AAG	PRKD1	frameshift_variant	#N/A	0.5166	19.69
407.004	5	124080672	T	TGAGATGAA	ZNF608	frameshift_variant	#N/A	0.5141	6.217
407.004	12	1906588	G	A	CACNA2D4	stop_gained	#N/A	0.503	2.347
407.004	9	117052366	G	T	COL27A1	missense_variant	#N/A	0.504	16.08
407.004	19	55667592	C	T	TNNI3	missense_variant	#N/A	0.5033	1.023
407.0059	15	67457360	G	A	SMAD3	missense_variant	#N/A	0.7483	29.35
407.0059	2	44065779	C	T	ABCG5	missense_variant	#N/A	0.6233	0.06449
407.0059	6	152422926	G	A	ESR1	3_prime_UTR_variant	#N/A	0.5099	2.472
407.0059	1	218519398	GCA	GCACACACACACACA	RP11-224O19.2	promoter/upstream_gene_variant	TGFB2	0.8167	9.631
407.0059	X	146993567	C	CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG	FMR1	promoter/5_prime_UTR_variant	FMR1	0.6576	22.46
407.0059	17	39766445	C	A	KRT16	stop_gained	#N/A	0.504	4.709

407.0059	19	41122842	C	T	LTBP4	missense_variant	#N/A	0.5778	539.8
407.0059	15	41222142	G	A	DLL4	missense_variant	#N/A	0.5376	11.66
407.0059	7	80300345	T	C	CD36	missense_variant	#N/A	0.5042	43.71
407.0059	9	117052374	G	A	COL27A1	missense_variant	#N/A	0.504	16.08
407.0059	4	119652620	G	A	SEC24D	missense_variant	#N/A	0.5023	40.73
407.0084	7	151878445	C	T	KMT2C	missense_variant	#N/A	0.5077	13.22
407.0096	1	218519398	GCA	GCACACACACA	RP11-224O19.2	promoter/upstream_gene_variant	TGFB2	0.8167	9.631
407.0096	6	170159101	G	T	ERMARD	missense_variant	#N/A	0.6744	20.37
407.0096	1	234510132	G	C	COA6	missense_variant	#N/A	0.5033	15.66
407.0096	22	50518424	C	G	MLC1	missense_variant	#N/A	0.5031	1.151
407.0106	X	146993567	C	CGCGGCGGCGGCGGCGGCGGCGGCGGCGG	FMR1	promoter/5_prime_UTR_variant	FMR1	0.6576	22.46
407.0106	7	21901469	A	G	DNAH11	splice_acceptor_variant	#N/A	0.515	0.1181
407.0106	19	11660360	CG	C	CNN1	frameshift_variant	#N/A	0.502	749.4
407.0106	20	34021807	G	C	GDF5	missense_variant	#N/A	0.5053	0.341
407.014	5	56171018	G	A	MAP3K1	missense_variant	#N/A	0.5011	7.1
407.014	22	17687997	C	T	CECR1	missense_variant	#N/A	#N/A	12.45
407.014	8	61655340	T	G	CHD7	missense_variant	#N/A	0.6303	2.457
407.014	22	45898215	TC	T	FBLN1	frameshift_variant	#N/A	0.5149	388.7
407.0146	2	189927755	G	A	COL5A2	missense_variant	#N/A	0.7592	86.17
407.0146	12	1919476	C	T	CACNA2D4	missense_variant	#N/A	0.503	2.347
408.0023	20	48140626	C	T	PTGIS	missense_variant	#N/A	0.6005	227.6
408.0023	14	45658326	C	T	FANCM	stop_gained	#N/A	0.5559	1.898
408.0023	1	218519398	GCA	GCACA	RP11-224O19.2	promoter/upstream_gene_variant	TGFB2	0.8167	9.631
408.0023	10	17201151	AG	A	TRDMT1	frameshift_variant	#N/A	0.5027	2.529
408.0034	6	161127501	A	G	PLG	missense_variant	#N/A	0.5027	0.02842
408.0034	8	11418856	C	T	BLK	missense_variant	#N/A	0.5007	0.1521
408.0034	12	65141614	T	C	GNS	missense_variant	#N/A	0.5044	93.81
408.0057	7	44576017	A	C	NPC1L1	stop_gained	#N/A	0.517	1.437

408.0057	7	44576004	C	CT	NPC1L1	frameshift_variant	#N/A	0.517	1.437
408.0057	17	60060300	GAGGAGTCCGAGGAGTCCTTGGAGT	G	MED13	inframe_deletion	#N/A	0.5053	17.12
408.0057	2	166010984	T	TTCATTTTATATA	SCN3A	inframe_insertion	#N/A	0.5029	2.294
408.0057	7	44576019	A	T	NPC1L1	missense_variant	#N/A	0.517	1.437
408.0057	7	44576006	T	A	NPC1L1	missense_variant	#N/A	0.517	1.437
408.0057	2	220285664	C	A	DES	missense_variant	#N/A	0.503	818.3
408.0057	2	166010987	A	G	SCN3A	missense_variant	#N/A	0.5029	2.294
408.0057	17	10427107	G	T	MYH2	missense_variant	#N/A	0.5025	0.3145
501.0019	9	129461956	C	T	LMX1B	3_prime_UTR_variant	#N/A	0.5353	0.4518
501.0019	10	89487074	T	G	PAPSS2	missense_variant	#N/A	0.5167	35.97
501.0037	17	10559406	C	T	MYH3	splice_donor_variant	#N/A	0.5033	4.001
501.0037	15	45427354	C	CA	DUOX1	frameshift_variant	#N/A	0.5036	2.963
501.0037	8	100874102	C	T	VPS13B	missense_variant	#N/A	0.5366	10.29
501.0037	3	48789710	A	G	PRKAR2A	missense_variant	#N/A	0.5272	25.21
501.0037	17	78082318	T	C	GAA	missense_variant	#N/A	0.5035	53.03
501.0038	16	75664390	AG	A	KARS	frameshift_variant	#N/A	#N/A	90.93
501.0038	19	11218180	C	G	LDLR	missense_variant	#N/A	0.6233	19.02
501.0038	3	30733401	TTA	TTATA	TGFBR2	3_prime_UTR_variant	#N/A	0.8445	252.3
501.0038	1	218519398	GCA	GCACACACACACA	RP11-224O19.2	promoter/upstream_gene_variant	TGFB2	0.8167	9.631
501.0038	17	29182252	C	T	ATAD5	missense_variant	#N/A	0.5037	1.001
501.0038	20	36759606	C	T	TGM2	missense_variant	#N/A	0.502	1712
502.0005	1	40756649	A	G	ZMPSTE24	missense_variant	#N/A	0.6091	43.43
502.0005	3	196433611	C	CA	CEP19	3_prime_UTR_variant	#N/A	#N/A	2.836
502.0005	1	218519398	GCA	GCACA	RP11-224O19.2	promoter/upstream_gene_variant	TGFB2	0.8167	9.631
502.0005	5	172662391	AAAAG	A	NKX2-5	promoter/upstream_gene_variant	NKX2-5	0.6415	0.07835
502.0005	X	111019931	TTTC	T	TRPC5	inframe_deletion	#N/A	0.5027	0

502.0005	16	89818553	C	G	FANCA	missense_variant	#N/A	0.5559	0.7294
502.0005	3	183558362	C	G	PARL	missense_variant	PARL	0.506	63.8
503.0003	1	155630950	AAAGT	A	YY1AP1	3_prime_UTR_variant	#N/A	0.5068	39.03
503.0003	15	60666835	G	GCT	ANXA2	frameshift_variant	#N/A	0.5083	339.7
503.0003	15	60666830	ACC	A	ANXA2	frameshift_variant	#N/A	0.5083	339.7
503.0003	15	60666837	AC	A	ANXA2	frameshift_variant	#N/A	0.5083	339.7
503.0003	15	60666846	C	A	ANXA2	stop_gained	#N/A	0.5083	339.7
503.0003	15	60666843	T	TACAA	ANXA2	frameshift_variant	#N/A	0.5083	339.7
503.0003	15	60666832	C	CGGTGATGA	ANXA2	frameshift_variant	#N/A	0.5083	339.7
503.0003	15	60666825	A	AAGAGCTGAAGAG	ANXA2	inframe_insertion	#N/A	0.5083	339.7
503.0003	15	60666848	AGTG	A	ANXA2	inframe_deletion	#N/A	0.5083	339.7
505.0033	3	12447537	C	T	PPARG	missense_variant	#N/A	0.528	11.85
505.0033	3	196433611	C	CA	CEP19	3_prime_UTR_variant	#N/A	#N/A	2.836
505.0033	3	124485079	C	A	ITGB5	stop_gained	#N/A	0.5032	254
505.0033	3	124485075	GGCT	G	ITGB5	inframe_deletion	#N/A	0.5032	254
505.0033	16	2138583	C	T	TSC2	missense_variant	#N/A	0.5083	44.28
505.0033	13	52548568	A	G	ATP7B	missense_variant	#N/A	0.5028	3.434
505.0033	18	25532172	T	C	CDH2	missense_variant	#N/A	0.5019	40.24
507.0027	5	73981198	AG	A	HEXB	frameshift_variant	#N/A	0.5053	95.98
507.0027	17	48273539	C	A	COL1A1	missense_variant	#N/A	0.7851	497.7
507.0027	1	55512252	G	C	PCSK9	missense_variant	#N/A	0.6233	0.08579
507.0027	X	146993567	C	CGCGGCGGCGGCGGCGGCGGCGGCGGAG	FMR1	promoter/5_prime_UTR_variant	FMR1	0.6576	22.46
507.0027	18	9102961	G	T	NDUFV2	splice_donor_variant	#N/A	0.5037	60.85
507.0027	17	41179212	A	G	RND2	missense_variant	#N/A	0.5186	5.956
507.0027	16	2130319	C	T	TSC2	missense_variant	#N/A	0.5083	44.28
507.0027	9	5066712	G	A	JAK2	missense_variant	#N/A	0.5023	34.09
507.003	X	153579350	C	A	FLNA	missense_variant	#N/A	0.6744	2748
507.003	22	21304236	G	A	CRKL	3_prime_UTR_variant	#N/A	0.6481	36.4
507.003	3	196433611	C	CA	CEP19	3_prime_UTR_variant	#N/A	#N/A	2.836

507.003	7	74072002	AAAAAG	GAAAAG	GTF2I	promoter/non_coding_transcript_exon_variant	GTF2I	0.6115	16.16
507.003	16	84100154	A	G	MBTPS1	missense_variant	#N/A	0.5063	67.03
507.003	7	120387826	C	T	KCND2	missense_variant	#N/A	0.5045	0.2384
508.0013	17	56348226	T	G	MPO	splice_acceptor_variant	#N/A	0.5096	0.3406
508.0013	11	71146886	C	G	DHCR7	splice_acceptor_variant	#N/A	0.5006	11.59
508.0013	19	11546962	G	GCTA	PRKCSH	inframe_insertion	#N/A	0.5196	117.7
508.0013	2	175619037	G	C	CHRNA1	missense_variant	#N/A	0.5133	0.05583
509.0023	4	148876485	CAT	C	ARHGAP10	frameshift_variant	#N/A	0.5017	90.92
509.0023	9	135802688	C	T	TSC1	missense_variant	#N/A	0.5044	18.27
509.0023	6	75892874	G	T	COL12A1	missense_variant	#N/A	0.5039	40.69
509.0023	21	28296787	G	T	ADAMTS5	missense_variant	#N/A	0.5038	4.784
601.003	5	121399702	T	A	LOX	3_prime_UTR_variant	#N/A	0.7452	30.79
601.003	X	146993567	C	CGCGGCGGCGGCGGCGGCGGCGGCG	FMR1	promoter/5_prime_UTR_variant	FMR1	0.6576	22.46
601.003	3	33138333	G	A	GLB1	promoter/intron_variant	GLB1	0.629	26.64
601.003	1	201030442	C	A	CACNA1S	stop_gained	#N/A	0.5024	0.1669
601.0057	2	44102553	G	GT	ABCG8	splice_donor_variant	#N/A	0.6233	0
601.0057	1	218519398	GCA	G	RP11-224O19.2	promoter/upstream_gene_variant	TGFB2	0.8167	9.631
601.0057	X	146993567	C	CGCGGCGGCGGCGGCGGCGGCGGCG	FMR1	promoter/5_prime_UTR_variant	FMR1	0.6576	22.46
601.0057	X	44938589	C	G	KDM6A	stop_gained	#N/A	0.5815	10.64
601.0057	10	79769420	C	T	POLR3A	stop_gained	#N/A	0.5242	7.187
601.0057	18	3131436	T	C	MYOM1	missense_variant	#N/A	0.5024	23.11
701.0002	1	47882541	C	T	FOXE3	missense_variant	#N/A	0.7201	0.06706
701.0002	2	29129444	AGTG	A	WDR43	inframe_deletion	#N/A	0.5154	17.37
701.0002	1	154560700	T	C	ADAR	missense_variant	#N/A	0.6146	86.07
701.0005	10	31810782	A	C	ZEB1	missense_variant	#N/A	0.5007	44.15

701.0005	1	218519398	GCA	GCACA	RP11-224O19.2	promoter/upstream_gene_variant	TGFB2	0.8167	9.631
101.013	6	32936713	G	A	BRD2	5_prime_UTR_variant	#N/A	0.5268	123.2
101.0132	17	45727488	C	T	KPNB1	5_prime_UTR_variant	#N/A	0.5144	56.43
101.0133	11	108093387	G	A	ATM	promoter/5_prime_UTR_variant	NPAT	0.5143	6.034
202.0004	19	50310326	A	T	FUZ	3_prime_UTR_variant	#N/A	0.5168	15.59
202.0004	6	108487305	T	C	NR2E1	promoter/5_prime_UTR_variant	NR2E1	0.517	0
401.005	1	170632772	A	C	PRRX1	5_prime_UTR_variant	#N/A	0.5746	37.52
401.0127	15	96869468	T	C	NR2F2	5_prime_UTR_variant	#N/A	0.517	129.5
401.0127	17	60142653	G	A	MED13	promoter/upstream_gene_variant	MED13	0.5053	17.12
401.0158	15	74218810	G	A	LOXL1	promoter/5_prime_UTR_variant	LOXL1	0.5137	66.71
402.0084	21	34915328	C	A	SON	5_prime_UTR_variant	#N/A	0.5835	73.41
402.0084	5	92918936	T	C	NR2F1-AS1	promoter/intron_variant	NR2F1	0.5173	4.98
407.0059	1	110881385	G	C	RBM15	5_prime_UTR_variant	#N/A	0.5352	3.575
407.0084	6	108487305	T	C	NR2E1	promoter/5_prime_UTR_variant	NR2E1	0.517	0
407.0106	17	43225854	T	C	HEXIM1	5_prime_UTR_variant	#N/A	0.5711	40.03
407.014	10	80828795	G	C	ZMIZ1	5_prime_UTR_variant	#N/A	0.5053	30.64
502.0005	3	183558362	C	G	PARL	promoter/missense_variant	PARL	0.506	63.8
503.0003	7	150945824	A	ACGCCGCCGCCGCCCGCC	SMARCD3	promoter/5_prime_UTR_variant	SMARCD3	0.5053	70.64
507.003	2	163695223	C	G	KCNH7	5_prime_UTR_variant	#N/A	0.5042	0.09984
601.003	16	69600155	C	T	NFAT5	5_prime_UTR_variant	#N/A	0.5122	12.96
601.003	4	99850227	G	A	EIF4E	5_prime_UTR_variant	#N/A	0.5115	5.591
601.003	3	57583177	G	A	ARF4	5_prime_UTR_variant	#N/A	0.5083	195.3
601.003	12	120649495	G	A	PXN	3_prime_UTR_variant	#N/A	0.5081	51.39
701.0002	1	150782554	CACACA	C	ARNT	3_prime_UTR_variant	#N/A	0.5055	45.57