

# Somatic Mutations Drive Vascular Malformation Initiation, Progression, and Predisposition

by

Daniel Aaron Snellings

Department of Molecular Genetics and Microbiology  
Duke University

Date: \_\_\_\_\_

Approved:

---

Douglas Marchuk, Supervisor

---

Beth Sullivan

---

Michael Hauser

---

Timothy Reddy

---

Craig Lowe

Dissertation submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy in the Department of Molecular Genetics and Microbiology  
in the Graduate School of Duke University  
2022

## ABSTRACT

### Somatic Mutations Drive Vascular Malformation Initiation, Progression, and Predisposition

by

Daniel Aaron Snellings

Department of Molecular Genetics and Microbiology  
Duke University

Date: \_\_\_\_\_

Approved:

---

Douglas Marchuk, Supervisor

---

Beth Sullivan

---

Michael Hauser

---

Timothy Reddy

---

Craig Lowe

An abstract of a dissertation submitted in partial fulfillment of the requirements for  
the degree of Doctor of Philosophy in the Department of Molecular Genetics and  
Microbiology  
in the Graduate School of Duke University  
2022

Copyright © 2022 by Daniel Aaron Snellings  
All rights reserved except the rights granted by the  
Creative Commons Attribution-Noncommercial Licence

# Abstract

Vascular malformations are a diverse class of focal lesions that may occur throughout the body and affect different vascular beds. Since the advent of next-generation sequencing it has become clear that virtually all vascular malformations are caused by postzygotic genetic changes occurring in a single cell: somatic mutations.

Somatic mutations are widely accepted to be the initial, catalyzing event for vascular malformation. These mutations are a mix of gain-of-function and loss-of-function and occur in canonical vascular genes, or known oncogenes. From the past two decades of research, we have identified numerous genes which contribute to vascular malformation; yet, despite this genetic diversity, the mutations identified in individual malformations has thus far been—without exception—monogenic.

In this document I present the first type of vascular malformation where digenic somatic mutations are a common and critical component of lesion pathogenesis. Furthermore, I propose that somatic mutations cause not only lesion initiation, but may drive the initiation, progression, and predisposition to vascular malformation.

In support of this hypothesis, I first identify that vascular malformations in hereditary hemorrhagic telangiectasia are initiated by somatic mutations via a two-hit mechanism. Second, I determine that cerebral cavernous malformations harbor up to 3 distinct somatic mutations that synergize to fuel lesion progression. Finally, I show that developmental venous anomalies harbor a somatic mutation which creates a mosaic field of mutant cells that predisposes to cerebral cavernous malformations.

for my friend  
LUCA FRANZINI

# Contents

<b>Abstract</b>	<b>iv</b>
<b>List of Tables</b>	<b>x</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Abbreviations and Symbols</b>	<b>xiii</b>
<b>Acknowledgements</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Overview and Significance . . . . .	2
1.2 Background . . . . .	5
1.2.1 Hereditary Hemorrhagic Telangiectasia . . . . .	7
1.2.2 Cerebral Cavernous Malformation . . . . .	11
<b>2 Two-Hit Mechanism of Hereditary Hemorrhagic Telangiectasia</b>	<b>19</b>
2.1 Premise . . . . .	20
2.2 Results . . . . .	21
2.2.1 Telangiectasia Harms from a Somatic Mutation in <i>ENG</i> or <i>ACVRL1</i> . . . . .	22
2.2.2 Somatic and Germline Mutations are Biallelic . . . . .	26
2.2.3 Mutations are Consistent with Homozygous Loss of Function . . . . .	27
2.2.4 Telangiectasia from the Same Individual Harbors Unique Somatic Mutations . . . . .	32
2.3 Discussion . . . . .	35

2.3.1	Evidence for a Genetic Two-Hit Mechanism . . . . .	35
2.3.2	Necessary, but Not Sufficient . . . . .	38
2.3.3	Sensitivity for Detecting Somatic Mutations . . . . .	38
2.3.4	Mutant Cell Metastasis . . . . .	39
2.3.5	Probability of Multiple Somatic Mutations . . . . .	40
2.3.6	Two-Hit Mechanism for <i>SMAD4</i> & JP-HHT . . . . .	40
2.3.7	Compound Heterozygosity . . . . .	41
2.3.8	Therapeutic Potential . . . . .	41
2.4	Methods . . . . .	42
2.5	Contributions and Acknowledgements . . . . .	48
<b>3</b>	<b><i>PIK3CA</i> Mutations Fuel Cerebral Cavernous Malformations</b>	<b>50</b>
3.1	Premise . . . . .	51
3.2	Results . . . . .	54
3.2.1	<i>PIK3CA</i> Mutations Occur in Familial and Sporadic CCMs . .	54
3.2.2	CCMs Harbor Multiple Somatic Mutations in Different Genes	55
3.2.3	<i>PIK3CA</i> and CCM/ <i>MAP3K3</i> Mutations in the Same Cell . .	57
3.3	Discussion . . . . .	61
3.3.1	Three-Hit Model of CCM Pathogenesis . . . . .	61
3.3.2	Role of Clonal Expansion in Mutagenesis . . . . .	62
3.3.3	Therapeutic Implications . . . . .	64
3.3.4	DVA Predispose to CCM and Other PI3K-Related Diseases . .	64
3.4	Methods . . . . .	65
3.5	Contributions and Acknowledgements . . . . .	72
<b>4</b>	<b>Developmental Venous Anomalies Predispose to Malformation</b>	<b>74</b>
4.1	Premise . . . . .	75

4.2	Results . . . . .	76
4.2.1	Mutations in <i>MAP3K3</i> and CCM Genes are Mutually Exclusive	76
4.2.2	Cellular Phase of Somatic <i>MAP3K3</i> and <i>PIK3CA</i> Mutations .	78
4.2.3	CCMs and DVA Harbor a Shared Mutation in <i>PIK3CA</i> . . . .	80
4.2.4	Plasma miRNAs Reflect PI3K Activation in DVA Patients . .	81
4.2.5	Relative Probability of Multiple Somatic Mutations . . . . .	82
4.3	Discussion . . . . .	85
4.3.1	<i>MAP3K3</i> GOF and CCM LOF Have Similar Molecular Effects	85
4.3.2	DVA Clonal Expansion Promotes Additional Mutations . . . .	85
4.3.3	DVA May Contribute to Additional Phenotypes . . . . .	88
4.4	Methods . . . . .	88
4.5	Contributions and Acknowledgements . . . . .	93
<b>5</b>	<b>Conclusions &amp; Musings</b>	<b>94</b>
5.1	Missing Mutations . . . . .	97
5.1.1	Somatic Hypermethylation . . . . .	98
5.1.2	Somatic Loss of Heterozygosity . . . . .	102
5.2	Connection Between Sporadic Brain AVMs and HHT . . . . .	108
5.3	Syndromic Brain AVMs in the Amish . . . . .	110
5.4	Sturge-Weber Syndrome and Somatic Mutations in <i>GNAQ</i> . . . .	114
5.5	Somatic Mutations in Infantile Hemangioma . . . . .	117
5.5.1	GLUT1 in IH endothelium . . . . .	118
5.5.2	Efforts to find somatic mutations in IH . . . . .	118
5.6	CCM & Meningioma . . . . .	119
5.6.1	Kruppel-like factor 4 ( <i>KLF4</i> ) . . . . .	119

<b>A Probability of Somatic Mutations</b>	<b>123</b>
A.1 Two-Hit Mutations in HHT . . . . .	124
<b>Bibliography</b>	<b>126</b>
<b>Biography</b>	<b>154</b>

# List of Tables

2.1	Mutations Identified in HHT Telangiectasia . . . . .	24
2.2	Phase of Somatic and Germline Mutation Pairs . . . . .	29
2.3	Predicted Consequences of Germline and Somatic Mutations . . . . .	33

# List of Figures

1.1	Vascular Malformations in HHT . . . . .	7
1.2	Cerebral Cavernous Malformations . . . . .	12
1.3	Developmental Venous Anomalies . . . . .	13
1.4	Two-Hit Mechanism of CCM Pathogenesis . . . . .	17
2.1	Expectations of a Two-Hit Model of HHT . . . . .	22
2.2	Low Frequency Somatic Mutations Detected in Telangiectasia . . . . .	25
2.3	Establishing Phase of Germline and Somatic Mutations . . . . .	28
2.4	Mutations in <i>ACVRL1</i> Disrupt Splicing . . . . .	31
2.5	Each Telangiectasia is Seeded by a Unique Somatic Mutation . . . . .	34
2.6	Two-Hit Model of HHT Pathogenesis . . . . .	35
3.1	Significant Variability and Rapid Growth of CCMs . . . . .	53
3.2	Familial and Sporadic CCM Harbor Somatic Mutations in <i>PIK3CA</i> .	56
3.3	Single-Nucleus DNA Sequencing to Resolve Cellular Phase . . . . .	58
3.4	Single-Nucleus Sequencing of CCMs with Three Pathogenic Mutations	59
3.5	Three-Hit Model of CCM Pathogenesis . . . . .	63
3.6	Correspondence of Bulk and Single-Nucleus Sequencing . . . . .	70
4.1	<i>MAP3K3</i> Somatic Mutations and Cellular Phase in CCM . . . . .	77
4.2	CCM and Adjacent DVA Have Identical <i>PIK3CA</i> Mutations . . . . .	79
4.3	Differentially Expressed micro-RNAs in CCM and DVA . . . . .	82
4.4	Genetic Model of CCM Pathogenesis . . . . .	83

4.5	Clonal Expansion Predisposes to Additional Mutations . . . . .	86
5.1	Predicted Somatic Hypermethylation Profile . . . . .	99
5.2	CpG Islands in HHT Gene Promoters . . . . .	100
5.3	Primers Targeting CpG Islands in <i>ENG</i> and <i>ACVRL1</i> . . . . .	101
5.4	Promoter Methylation in <i>ENG</i> and <i>ACVRL1</i> . . . . .	102
5.5	Somatic LOH Detection with snDNA-seq . . . . .	104
5.6	Linked Allelic Dropout . . . . .	105
5.7	Differences Between Linked ADO and LOH . . . . .	106
5.8	Leveraging Known Somatic Mutation for LOH Detection . . . . .	107
5.9	Putative Somatic LOH Event in a CCM . . . . .	108
5.10	Autozygosity in Amish with Syndromic Brain AVMs . . . . .	112
5.11	Structural Evaluation of Activating <i>GNAQ</i> Mutations . . . . .	115
5.12	Transcriptional Effects of Activating <i>GNAQ</i> Mutations . . . . .	116
5.13	<i>KLF4</i> Mutations in COSMIC . . . . .	121

# List of Abbreviations and Symbols

## Abbreviations

VM	Vascular Malformation, a blood vessel that is malformed, though functional.
CCM	Cerebral Cavernous Malformation, a common vascular malformation that occurs in the brain and spinal cord. As is standard in the CCM research community, the abbreviation ‘CCM’ is used to refer to the familial disorder, the vascular malformation itself, and occasionally the ‘CCM Genes’ (referring to <i>KRIT1</i> , <i>CCM2</i> , and <i>PDCD10</i> ) and the ‘CCM Complex’ (the signaling complex formed by the proteins: KRIT1, CCM2, and PDCD10).
HHT	Hereditary Hemorrhagic Telangiectasia, an inherited vascular malformation disorder characterized by telangiectasia and arteriovenous malformations.
AVM	Arteriovenous Malformation, a vascular malformation consisting of a direct shunt between an artery and vein, bypassing the capillary bed.
DVA	Developmental Venous Anomaly, a common, benign vascular malformation that is often present with a cerebral cavernous malformation.
SWS	Sturge-Weber Syndrome, a sporadic vascular malformation disorder characterized by a mosaic port-wine stain on the face and neurological involvement.
PWS	Port-Wine Stain, a benign vascular malformation typically on the face. May form independently, or as part of Sturge-Weber syndrome.
FFPE	Formalin-Fixed Paraffin-Embedded, a common tissue fixation strategy in pathology. The fixation process partially degrades DNA and RNA in the tissue.

COSMIC	Catalog Of Somatic Mutations In Cancer, a web database containing somatic mutations that have been identified in cancers.
LOH	Loss Of Heterozygosity, the loss of one allele at a heterozygous site. LOH can result from various genetic mechanisms (e.g. deletion, recombination, mitotic nondisjunction).

# Acknowledgements

First, I would like to acknowledge everyone that donated tissue and participated in these studies. This is ultimately for them, and without them, would not be possible.

I would like to thank my advisor, Douglas Marchuk, for his mentorship and guidance throughout my research and for his boundless enthusiasm for science. I am also thankful for the support of the current and former members of the Marchuk Lab: Han Kyu Lee, Sarah Wetzel-Strong, Leandro Barbosa Do Prado, Matthew Detter, Francesca Galeffi, Carol Gallione, Christian Benavides, Mary Patrucco, Catherine Neilson, and Erin Griffin. I would also like to thank Craig Lowe for his mentorship and for always welcoming me into his lab. I also thank the members of the Lowe Lab for always being open to talk science and for their friendship: Eric Au, Riley Mangan, Christiana Fauci, Luke Bartelt, Chelsea Shoben, Seth Weaver, Yanting Luo, Kat Bendt, and Juliana Carvalho. I would also like to acknowledge the members of my thesis advisory committee for their advice throughout these studies: Beth Sullivan, Michael Hauser, Timothy Reddy, and Craig Lowe.

I could not have completed any of this work without the support of our many collaborators. In particular I would like to acknowledge the contributions of Aileen Ren, Mark Kahn, Issam Awad, Mark Ginsberg, Paul Oh, Michael Lawton, Marie Faughnan, Romuald Girard, Amy Akers, the Angioma Alliance Foundation, and the Cure HHT Foundation. I would also like to acknowledge funding from the National

Heart Lung and Blood Institute and the American Society for Human Genetics.

I would like to acknowledge multiple core research facilities at Duke University that were critical for the completion of this work: Sequencing and Genomic Technologies Core (Nicolas Devos, Holly Dressman, Laura-Leigh Rowlette, Fangfei Ye), Duke Human Vaccine Institute Flow Cytometry Core (Derek Cain, Evan Trudeau, Steven Slater), and the Duke Center for Genomic and Computational Biology High-throughput Applied Research Data Analysis Cluster (Hilmar Lapp, Gregory Steffen, Dan Somers).

Lastly, I thank my family and friends for their loving support throughout this journey.

# 1

## Introduction

---

Part of this chapter is adapted from a review in *Circulation Research* (Snellings et al., 2021)

## 1.1 Overview and Significance

The growth and remodeling of blood vessels is a tightly regulated process that must be maintained from early embryogenesis until death. When this process goes awry blood vessels can become malformed leading to altered, and sometimes pathological function. These malformed blood vessels are termed ‘vascular malformations’ (VMs). VMs may occur anywhere throughout the body, but virtually always occur as focal lesions—affecting one region of the vasculature, rather than presenting as a systemic vascular defect. Even though VMs are focally restricted, they are a source of significant morbidity and mortality in affected individuals. The clinical phenotypes produced by VMs vary substantially, though they are generally prone to rupture (due to increased flow through a fragile vascular bed) or hemorrhage (due to leaky endothelial junctions). The VMs that I will discuss in the following chapters commonly form in the brain and may cause a range of neurological phenotypes including: stroke, epilepsy, seizures, and neurological deficit.

Understanding molecular pathogenesis of VMs is critical for developing methods of treatment and prevention. Although VMs exhibit massive phenotypic heterogeneity, it is becoming increasingly clear that many—if not all—VMs form as a result of somatic mutations.

Somatic mutations are well established as a critical, initiating factor in VM development; however, there remain many types of VM for which the genetic etiology is poorly understood. My dissertation seeks to understand the contribution of somatic mutations to the genetic etiology of Hereditary Hemorrhagic Telangiectasia. Furthermore, I seek to identify novel mechanisms by which somatic mutations fuel the initiation, progression, and predisposition to Cerebral Cavernous Malformations.

In Chapter 2, I determine that VMs associated with Hereditary Hemorrhagic Telangiectasia (HHT) form via a genetic two-hit mechanism whereby a somatic

mutation in *ENG* or *ACVRL1* results in biallelic loss of function. HHT is a mendelian disorder inherited by an autosomal dominant allele containing a mutation either *ENG*, *ACVRL1*, or *SMAD4*. Previous studies established that these inherited mutations result in loss of function. This finding lead to the predominant theory that the VMs associated with HHT are the result of haploinsufficiency. However, this theory is inconsistent with the presentation of VMs in HHT. The VMs associated with HHT present as focal lesions, rather than manifesting as a systemic vascular defect as might be expected by haploinsufficiency. This disconnect suggests that a local event is necessary for the development of VMs in HHT. I tested whether this local event may be a somatic mutation in the single remaining wild-type allele of the mutated gene. Such a mutation would cause in biallelic loss of function and result in the complete loss of functional gene product in the cells harboring the somatic mutation. I sequenced 19 telangiectasia (a cutaneous VM) from individuals with HHT and found a loss of function somatic mutation in 9 of 19 telangiectasia. Each somatic mutation I identified occurred in the same gene as the respective inherited mutation. Using long-read sequencing I showed that pairs of somatic and germline mutations exist in a *trans* configuration, confirming that these mutations resulted in biallelic loss of function. Furthermore I showed that multiple telangiectasia from the same individual harbored unique somatic mutations, suggesting that each telangiectasia is initiated by a different mutational event rather than a metastasis-like mechanism.

In Chapter 3, I identify a novel somatic mutation in Cerebral Cavernous Malformations (CCMs) that occurs in the *same cell* as additional somatic mutations that synergize to fuel CCM progression. Similar to HHT, CCM is a mendelian disorder inherited by an autosomal dominant allele containing a mutation in either *KRIT1*, *CCM2*, or *PDCD10*. Previous studies established that CCMs are caused by somatic mutations via a genetic two-hit mechanism similar to what I describe in Chapter 2. The two-hit mechanism tidily explains the variable penetrance of

the inherited disorder and explains why sporadic cases have a single CCM whereas inherited cases have many; however, the two-hit mechanism does not account for the extreme variation in the severity of CCMs in the same individual, nor why some CCMs rapidly grow after years of quiescence. I hypothesized that some of the most aggressive CCMs may harbor somatic mutations in other genes that fuel CCM growth. To test this hypothesis, I sequenced 79 sporadic and inherited CCMs and found that 71% of these CCMs all had a gain of function somatic mutation in the same gene: *PIK3CA*. Surprisingly, many of the CCMs had mutations in *PIK3CA* and either *KRIT1*, *CCM2*, or *PDCD10*; as many as three distinct somatic mutations within a single lesion. This finding prompted a new question: are all of the mutations present within the same cell, or are two different populations of cells merging to form a mosaic CCM? To answer this question, I collected nuclei from frozen CCM samples and sequenced the DNA from individual nuclei to determine the cellular phase of the somatic mutations. I sequenced a total of 21,221 nuclei across 2 inherited and 6 sporadic CCMs and all 8 confirmed that the somatic mutations are present within a single clonal population of cells. This study is the first to find that multiple genes contribute to the pathogenesis of a vascular malformation and highlights *PIK3CA* as a major contributor to CCM progression. In addition, the identification of *PIK3CA* mutations suggested rapamycin (an inhibitor of PI3K signaling) as a potential therapeutic for CCM and has since proved highly effective in preventing CCMs in mice.

In Chapter 4, I find that Developmental Venous Anomalies (DVA) are caused by a somatic mutation in *PIK3CA* and predisposes to the formation of CCMs. The finding in Chapter 3 that some CCMs harbor as many as three distinct somatic mutations all within the same cell was somewhat surprising. Multiple mutations are common in cancer, but cancer develops in the context of uncontrolled growth and genomic instability. To understand how multiple mutations were occurring in CCMs

we looked for signs of genomic instability in CCMs via sequencing and stains for DNA damage but found no evidence of elevated mutation rates. Instead, I found an answer in the form of DVA. DVA are the most common vascular malformation, present in up to 16% of the population. DVA are considered to be harmless; however, almost all sporadic CCM form directly adjacent to a DVA. The association of DVA and CCM has been known for decades, but the cause has remained a mystery. I hypothesized that DVA could result from somatic *PIK3CA* mutation during developmental angiogenesis, creating a field of mutant cells that may develop into a CCM upon subsequent mutations. I tested this hypothesis using droplet digital PCR and found that *PIK3CA* mutation was present in both the CCM and the associated DVA, but that other pathogenic somatic mutations were found exclusively in the CCM. These data suggest that DVA are acting as a genetic primer that predisposes to CCM formation.

My results demonstrate that somatic mutations play an integral role in the development of VMs associated with HHT and CCM, and constitutes the first evidence of a VM that develops as a result of digenic somatic mutations. This opens new avenues of research into the mechanisms VM development, and identifies promising therapeutics that are currently being pursued for clinical trials. Before describing my results in detail, the remainder of this chapter provides a brief introduction to vascular malformations and previous research into HHT and CCM.

## 1.2 Background

Vascular malformation is a general term that encompasses a wide breadth of vascular phenotypes ranging from simple birthmarks to disfiguring, and life-threatening lesions. Despite these differences, vascular malformations share several important properties that are useful to examine in the context of their counterpart: vascular tumors. Vascular tumors, like other types of tumors, are characterized by rapid

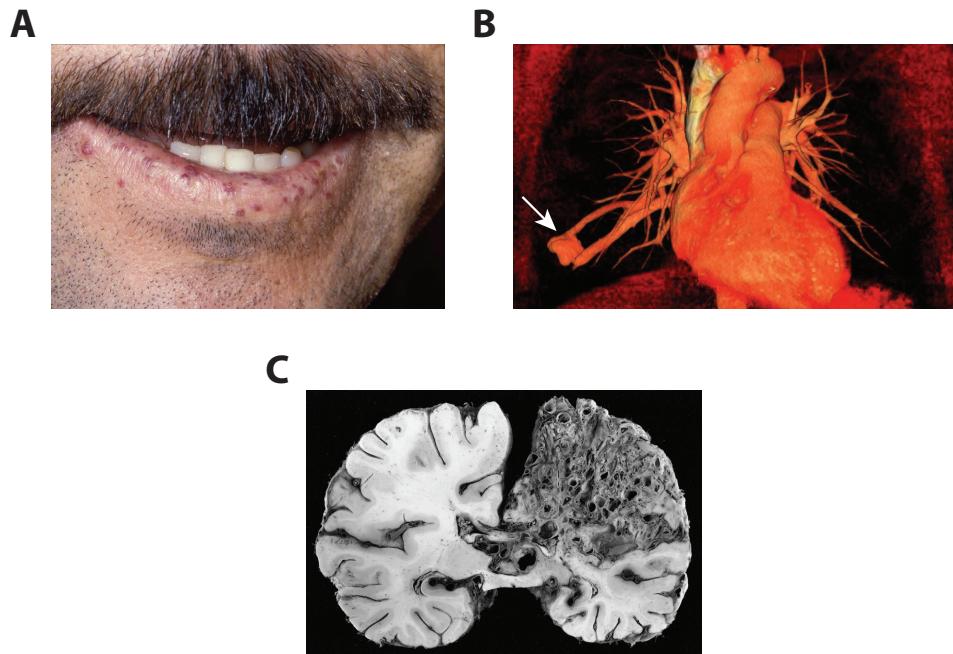
and uncontrolled growth. In contrast, vascular malformations typically do not grow uncontrollably but rather tend to grow proportionally to the individual. There are of course exceptions of vascular malformations that grow rapidly, which will be a focus of Chapter 3. In addition, vascular malformations are generally more stable than vascular tumors. Vascular tumors have a prominent growth phase which is often followed by spontaneous involution resulting in complete regression of the lesion. This is especially true of infantile hemangiomas, the most common type of vascular tumor which I will briefly discuss in Chapter 5. In contrast, the spontaneous regression of vascular malformations is very rare.

Vascular malformations are further divided by the type of vascular bed affected: arteries, veins, capillaries, or lymphatic vessels. While a vascular malformation may be restricted to a single type of vessel, they are often observed as mixed vascular malformations (e.g. arterio-venous malformation) and mixed vascular-lymphatic malformations (e.g. capillary-lymphatic-venous malformation). Vascular malformations are also identified by the rate of flow through the lesion, namely low-flow and high-flow. The rate of flow is generally a function of the type of vascular bed affected; with arterial and mixed-arterial malformations generally being high-flow; and venous, capillary, lymphatic malformations being low-flow. The rate of flow is a useful identifier as it helps inform the type of complications the individual may encounter: high-flow lesions are prone to rupture, whereas low-flow lesions prone to hemorrhage.

While many vascular malformations form sporadically, they may also occur in association with an inherited disorder. The chapters that follow focus on two such disorders: Hereditary Hemorrhagic Telangiectasia, and Cerebral Cavernous Malformations.

### 1.2.1 Hereditary Hemorrhagic Telangiectasia

Hereditary Hemorrhagic Telangiectasia (HHT) is a genetic disease characterized by abnormal direct connections between arteries and veins resulting in high-flow lesions called arteriovenous malformations (AVMs). Individuals with HHT often have small AVMs called telangiectasia on mucosal (GI, lips, tongue, inner eyelids) and dermal (face, fingers, nail beds) surfaces (Figure 1.1A). In addition, some individuals with HHT develop pulmonary (Figure 1.1B), hepatic, and brain AVMs (Figure 1.1C). These visceral AVMs are typically larger and carry considerably more risk compared to telangiectasia.



**FIGURE 1.1: Vascular Malformations in HHT.**

**A**, Telangiectasia on the lip of an individual with HHT. Case courtesy of Herbert L. Fred and Hendrik A. van Dijk (CC-BY 2.0). **B**, 3D render of a CT scan showing a pulmonary AVM creating a shunt between a pulmonary arterial branch and a pulmonary vein tributary (arrow). Case courtesy of Vikas Shah (CC BY-NC-SA 3.0). **C**, The cross section of a large brain AVM in the right hemisphere. Case courtesy of The Armed Forces Institute of Pathology (public domain).

## *Genetics*

HHT is an autosomal dominant disease known to be caused by mutations in one of three causal genes: *ENG* (McAllister et al., 1994), *ACVRL1* (Johnson et al., 1996), and *SMAD4* (Gallione et al., 2004). Mutations in *ENG* and *ACVRL1* account for roughly 85% of HHT cases, though as high as 96% has been reported when the Curaçao criteria are strictly followed (McDonald et al., 2020). Mutations in *SMAD4* cause the combined disorder of juvenile polyposis and HHT (JP-HHT); these mutations account for only 2% of HHT cases. Mutations that cause HHT are known to result in the loss of function of the affected gene. In *ENG* roughly 80% of pathogenic mutations result in premature termination either through frameshift or nonsense mutations. In contrast, 53% of pathogenic mutations in *ACVRL1* are missense changes, the majority of which occur in modest hotspots throughout the intracellular kinase domain (Abdalla et al., 2003). This difference may be attributed to the relative intolerance of missense changes in *ACVRL1* compared to *ENG* as shown by the low number of common missense variants in the gnomad database (*ENG*: expected 389.9, observed 338; *ACVRL1*: expected 311.5, observed 190). Compared to the other genes, *ACVRL1* also has a high number of splice variants, including a 300bp CT-rich region in intron 9 (Woorderchak-Donahue et al., 2018). In *SMAD4*, mutations causing JP-HHT are a mix of missense and nonsense/frameshift; the vast majority of which are found in exons 8–11, overlapping the MH2 domain of *SMAD4* (Gallione et al., 2004).

There remain a minority of individuals who fulfill all of the clinical diagnostic criteria of HHT (the Curaçao criteria) yet sequencing and CNV analysis reveal no pathogenic mutation in *ENG*, *ACVRL1*, or *SMAD4*. In these cases, one possibility is that a pathogenic mutation does exist in one of these genes, however it is labeled as variant of unknown significance, or the mutation occurs in regions not covered

by the sequencing panel (i.e. variants in introns or promoter regions). One other alternative is that mutations in other unidentified genes may cause HHT. In the past there have been some indications that genes besides *ENG*, *ACVRL1*, and *SMAD4* may cause HHT. In 2005 and 2006 two new loci linked to HHT were mapped to chromosome 5 and 7 respectively (Cole et al., 2005; Bayrak-Toydemir et al., 2006). Despite this finding, no additional cases have been identified mapping to these regions, nor have any causal genes been identified. Recent work has identified the gene *GDF2* which encodes the protein BMP9 as a potential causal gene for HHT (Wooderchak-Donahue et al., 2013). There are currently two published reports of *GDF2* variants in individuals with HHT-like phenotypes found by sequencing samples sent for HHT diagnostic testing, but with no identified mutation in the known causal genes (Wooderchak-Donahue et al., 2013; Hernandez et al., 2015). Combining these studies, a total of 5 individuals with HHT-like phenotypes have been reported with five different *GDF2* variants. However, two of the 5 variants (c.997C>T and c.-51C>A) are present in the general population with frequency higher than HHT (HHT:  $1 \times 10^{-4}$ ; variant gnomAD v2 population frequencies:  $4.07 \times 10^{-4}$ , and  $2.48 \times 10^{-4}$  respectively). A third variant, c.950G>A is present in the general population at  $0.44 \times$  the frequency of HHT ( $4.38 \times 10^{-5}$ ). These data suggest that these three variants do not cause HHT. The remaining two variants (c.254C>T, c.203G>T) are extremely rare in the general population and may potentially be pathogenic. Specifically, c.203G>T has been found to segregate in a family with 3 individuals with epistaxis. Despite this evidence, the rarity of these mutations, and the incomplete phenotypic overlap with HHT suggests that if these variants are indeed pathogenic, they cause a rare vascular disease overlapping with HHT but should not be considered an HHT causal gene, though *GDF2* would be an appropriate addition to vascular anomaly diagnostic sequencing panels.

### *Animal Models*

Our understanding of the genetics of HHT has led to the development of mouse models that recapitulate the disease phenotype. The first mice created were made with heterozygous germline loss of either *Eng* or *Acvrl1* (*Eng*<sup>+-</sup> and *Acvrl1*<sup>+-</sup> respectively). These heterozygous mice match the genotype of humans with HHT, however these mice have a very mild phenotype that only presents in a small fraction of mice (Bourdeau et al., 1999; Torsney et al., 2003; Srinivasan et al., 2003). As the heterozygous mice did not produce a robust and reproducible phenotype, it was necessary to create a more aggressive model via homozygous knockout mice: *Eng*<sup>-/-</sup> and *Acvrl1*<sup>-/-</sup>. However, this was challenging as germline homozygous loss of these genes results in embryonic lethality due to vascular defects (Li et al., 1999; Bourdeau et al., 1999; Oh et al., 2000; Arthur et al., 2000). To circumvent the embryonic lethality, Allinson et al. and Park et al. generated mice where gene deletion could be induced by injecting the mice with tamoxifen using the Cre-Lox system (Allinson et al., 2007; Park et al., 2008). These homozygous inducible mice allowed deletion of both alleles of *Eng* or *Acvrl1* postnatally and therefore circumvent embryonic lethality. Surprisingly, inducing gene deletion after developmental angiogenesis had almost no affect on the vasculature. However, several groups found that pairing induced gene deletion while also stimulating vascular growth (e.g. wound healing, VEGF) resulted in robust and reproducible AVM formation (Park et al., 2008; Walker et al., 2011; Choi et al., 2012; Chen et al., 2013).

### *Sporadic Brain AVMs*

While AVMs are common in individuals with HHT, the majority of all AVMs are sporadic, occurring in unrelated individuals with no known genetic predisposition. These sporadic AVMs disproportionately occur in the brain. Molecular studies of sporadic brain AVMs has found that the majority harbor a somatic activating mu-

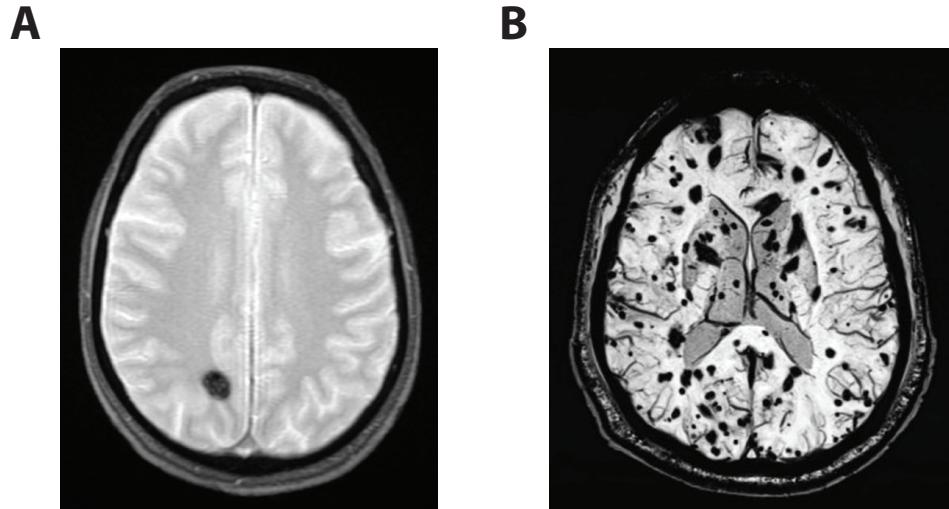
tation in *KRAS*, a commonly mutated oncogene (Nikolaev et al., 2018; Starke et al., 2018; Al-Olabi et al., 2018). Studies comparing the severity of sporadic and HHT-associated AVMs are sparse, however I have discussed this topic with several surgeons who operate on AVMs and the consensus is that sporadic AVMs tend to be larger and more aggressive than HHT-associated AVMs. This observation may suggest that sporadic and HHT-associated AVMs develop via different genetic mechanisms. HHT-associated AVMs are rarely surgically resected (in favor of embolism) therefore it has been challenging to assess the presence of *KRAS* mutations in HHT-associated AVMs; however, given the clear heredity of brain AVMs in HHT, it is unlikely that HHT-associated AVMs develop from a somatic *KRAS* mutation.

### *1.2.2 Cerebral Cavernous Malformation*

Cerebral Cavernous Malformation (CCM) is a genetic disease characterized by the formation of the eponymous vascular malformations—CCMs—in the brain and spinal cord. CCMs form in the capillaries that connect arteries and veins, specifically in postcapillary venules. CCMs consist of multiple ‘caverns’ that fill with blood resulting in a slow-flow vascular malformation similar in appearance to a raspberry. Similar to AVMs, CCMs may form as a result of the familial disease, or sporadically in otherwise healthy individuals. Those with familial CCM typically develop numerous CCMs (Figure 1.2A) whereas sporadic cases of CCM almost always present as a single solitary lesion (Figure 1.2B).

### *Developmental Venous Anomalies*

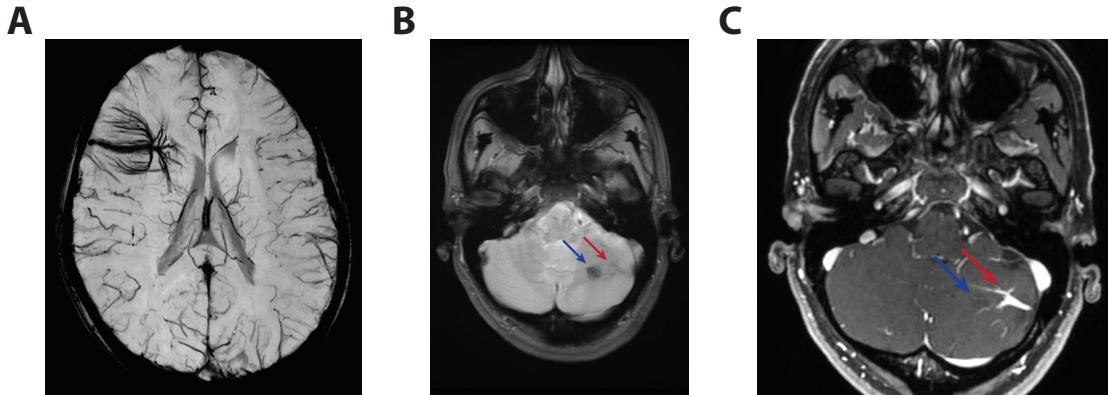
Although sporadic CCMs form in individuals with no known genetic predisposition, sporadic CCMs are known to be closely associated with developmental venous anomalies (DVA). DVA are the single most common vascular malformations affecting up to 16% of people (Brinjikji et al., 2017), and are often found to directly abut a sporadic



**FIGURE 1.2: Cerebral Cavernous Malformations.**

**A**, A sporadic CCM presenting as a solitary lesion in the hindbrain. Case courtesy of Bruno Di Muzio (CC BY-NC-SA 3.0). **B**, The brain of an individual with familial CCM presenting with numerous lesions throughout their brain. Case courtesy of Frank Gaillard (CC BY-NC-SA 3.0).

CCM. Several studies have reported that 24–32% of sporadic CCMs abut a DVA based on MR imaging (Abdulrauf et al., 1999; Wurm et al., 2005; Porter et al., 1999); however, in surgically resected CCM lesions, Porter et al. identified an intimately associated DVA in every of 86 resected lesions (Porter et al., 1999). In addition, several groups have reported that use of more powerful imaging techniques reveal DVA associated with CCM lesions that were otherwise missed with standard MR imaging (Dammann et al., 2017; Kamezawa et al., 2005)—suggesting the frequency of associated DVA and CCM is underestimated. Despite the common association between sporadic CCM and DVA, DVA are rarely ever found in association with familial CCM (Petersen et al., 2010), suggesting that DVA have a yet unknown role in the pathogenesis of specifically sporadic—but not familial—CCM.



**FIGURE 1.3: Developmental Venous Anomalies.**

**A**, DVA with a characteristic caput medusae (palm tree-like appearance) in the left hemisphere. Case courtesy of Neil Lall (CC BY-NC-SA 3.0). **B–C**, MR imaging of one case of an associated DVA and CCM visualized with either an Axial Gradient Echo (B) or an Axial T1 C+ (C). The Axial Gradient Echo shows the CCM best (blue arrow), whereas the Axial T1 C+ shows the DVA best (red arrow). Case courtesy of Frank Gaillard (CC BY-NC-SA 3.0).

#### *Germline Genetics*

Familial CCM follows an autosomal dominant pattern of inheritance (Bicknell et al., 1978; Clark, 1970; Kidd & Cumings, 1947) and is caused by heterozygous loss of function (LOF) mutations in either *KRIT1* (Laberge-le Couteulx et al., 1999a; Sahoo et al., 1999), *CCM2* (Liquori et al., 2003; Denier et al., 2004), or *PDCD10* (Bergametti et al., 2005). The protein products of these genes bind to each other to form the ‘CCM Complex’ which is involved in signal transduction related to vascular development. Mutations in these genes primarily consist of nonsense, frameshift, and canonical splice site mutations. Pathogenic missense mutations have been reported in all three genes; however, they make up the minority of reported mutations. Notably, there are 4 known founder mutations in the CCM genes that account for substantial fraction of familial CCM cases. The most common of these is NM\_194456.1(*KRIT1*):c.1363C>T(p.Gln455Ter) present in most Hispanic American cases of familial CCM (Gunel et al., 1996; Sahoo et al., 1999). Another less

frequent *KRIT1* founder mutation: NM\_194456.1(*KRIT1*):c.987C>A(p.Cys329Ter), has been found in several kindreds with Sardinian lineage (Cau et al., 2009). A deletion spanning 77.6kb of *CCM2* resulting in deletion of exons 2–10 has been found in several Caucasian kindreds, it remains unclear whether all cases with this mutation are related or whether it has occurred independently in other families (Liquori et al., 2007; Gallione et al., 2022). A splice mutation affecting *CCM2*: NM\_031443.3(*CCM2*):c.30+5\_30+6delinsTT, has been found in seemingly unrelated Ashkenazi Jewish probands, possibly due to an historically ancient mutation (Gallione et al., 2011).

#### *Genotype-Phenotype Correlation*

The clinical manifestations of familial CCM are highly heterogeneous. In many genetic diseases, the primary determinant of progression and severity is the causal mutation. However, this does not seem to be the case for CCM. A prime example of this is the common Hispanic founder mutation. Disease progression and severity in individuals with the Hispanic founder mutation is highly variable, despite sharing an identical pathogenic mutation (Gault et al., 2006; Laurans et al., 2003; Denier et al., 2004; Gianfrancesco et al., 2007). The lack of correlation between clinical manifestations and mutation identity is consistent with the finding that the majority of germline mutations found in *KRIT1*, *CCM2*, and *PDCD10*—primarily frameshift, nonsense, and splice site mutations—result in loss of function. This suggests that the identity of mutations within the same gene does not have a significant impact on disease severity.

Though different mutations within a gene do not seem to impact disease severity, the identity of the mutated gene has been associated with several clinical characteristics. Many groups have noted an association between individuals with *KRIT1* mutations and cutaneous vascular lesions (Gianfrancesco et al., 2007; Sirvente et al.,

2009; Musunuru et al., 2003; Grippaudo et al., 2013; Wang et al., 2013; Eerola et al., 2000; Labauge et al., 1999). Individuals with a mutation in *CCM2* are more likely to be asymptomatic and have lower number of lesions compared to individuals with *KRIT1* or *PDCD10* mutations (Denier et al., 2006). Familial CCM is significantly more aggressive in individuals with a mutation in *PDCD10* (Fauth et al., 2015; Shenkar et al., 2015; Riant et al., 2013). The severity of CCM associated with *PDCD10* mutations is attributable to the role of *PDCD10* in the gut epithelium not shared with *KRIT1* or *CCM2* (Tang et al., 2019).

#### *Genetic Modifiers*

To identify genetic variants associated with disease severity, one group performed a large genetic association study of individuals with the *KRIT1* founder mutation that identified several genetic polymorphisms within inflammatory and immune response genes that are associated with total lesion count, number of large lesions, and intracerebral hemorrhage (Choquet et al., 2014). This analysis revealed associations between clinical disease presentation and variants in several genes including: *TGFBR2*, *CD14*, *IL6R*, *MSR1*, *IGH*, and *TLR4*. Some of these genes have been shown to have critical roles in CCM pathogenesis, highlighting the importance of further evaluating the roles of these genes, and demonstrating the power of association studies in a genetically homogenous cohort. Identification of *TLR4* variants associated with disease severity are of particular interest owing to the direct role of *TLR4* in propagating CCM signaling (Tang et al., 2019). These data suggest that polymorphisms in genes other than *KRIT1*, *CCM2*, and *PDCD10* may be important modifiers of CCM pathogenesis.

### *Radiation-Induced CCMs*

While the majority of sporadic CCMs occur in otherwise healthy individuals, numerous reports have shown that ionizing radiation is a potent inducer of CCM formation (Cutsforth-Gregory et al., 2015; Heckl et al., 2002; Jain et al., 2005; Burn et al., 2007; Strenger et al., 2008; Vinchon et al., 2011; Koike et al., 2012; Martinez-Lage et al., 2008; Novelli et al., 1997; Baumgartner et al., 2003). While pathologically similar to non-radiation induced sporadic CCMs, they have several distinct characteristics that hint at the underlying mechanisms driving CCM pathogenesis. One such characteristic is that individuals with radiation-induced CCMs often present with multiple lesions, in stark contrast to non-radiation induced sporadic CCMs which almost always occur as a solitary lesion. Furthermore, it was found that occurrence of multiple radiation-induced CCMs is significantly associated with younger age at time of radiation treatment (Cutsforth-Gregory et al., 2015) and that the presence of multiple radiation induced CCMs may be related to higher doses of ionizing radiation (Novelli et al., 1997). Ionizing radiation has long been recognized as a potent source of DNA damage leading to genomic instability reviewed elsewhere (Little, 1998). The observation that radiation treatment may induce CCM formation—and that the multiplicity of lesions is related to radiation dose and age at radiation—support a key role for somatic mutations in CCM pathogenesis.

### *Somatic Mutation Cause CCM Via a Two-Hit Mechanism*

Although germline LOF mutations in *KRIT1*, *CCM2*, or *PDCD10* cause familial CCM as a clinical entity, they do not explain why CCMs present as focal lesions rather than a systemic vascular defect as might be expected if CCMs were the result of haploinsufficiency. This observation led to the hypothesis that a secondary, local event is necessary to initiate lesion formation; specifically, a somatic mutation in a CCM gene resulting in biallelic LOF. Somatic mutations resulting in biallelic LOF of

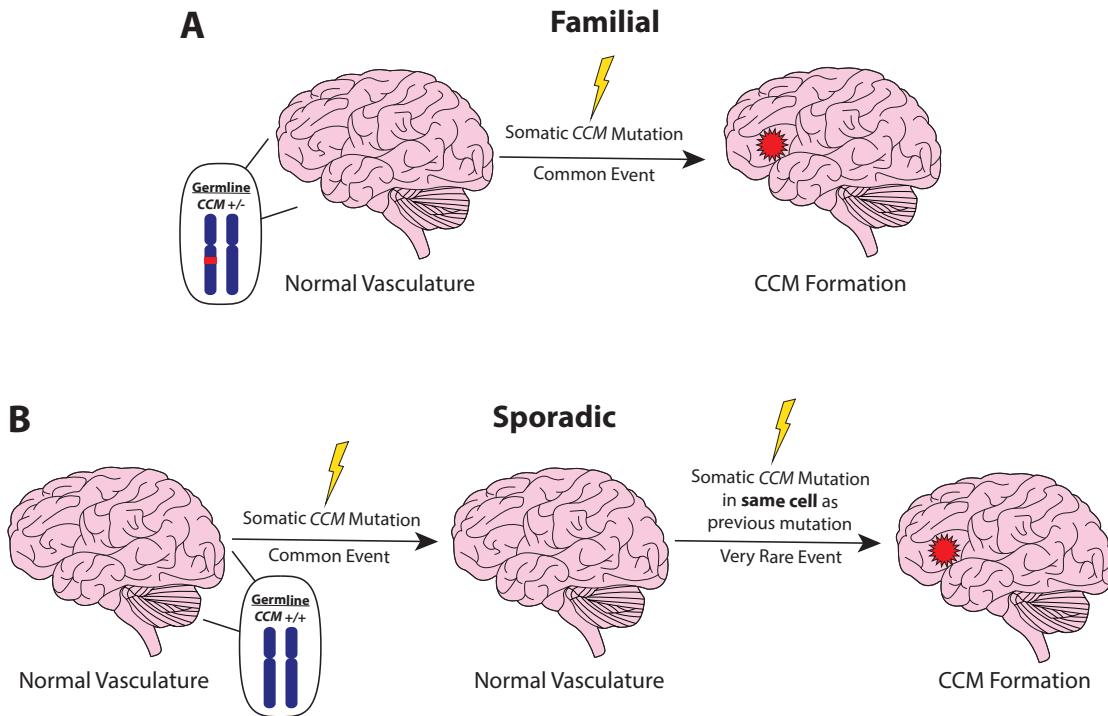


FIGURE 1.4: Two-Hit Mechanism of CCM Pathogenesis.

**A**, Two-Hit model for familial CCM where the starting state is heterozygous and a somatic mutation results in loss of the second allele. **B**, Two-Hit model for sporadic CCM where the starting state is wild type and two somatic mutations *in the same cell* are required to initiate lesion formation.

*KRIT1*, *CCM2*, or *PDCD10* have been reported in both familial (Gault et al., 2005; Akers et al., 2009; Gault et al., 2009) and sporadic (McDonald et al., 2014) CCMs. Furthermore, laser capture microdissection and immunohistochemical staining of the CCM proteins in the lesion endothelium has established that these mutations occur in endothelial cells (Akers et al., 2009; Pagenstecher et al., 2009; Rath et al., 2020). Together these data show that while loss of a CCM gene is dominant at the level of an individual, it is recessive on a cellular level, requiring loss of the normal allele via somatic mutation prior to lesion formation (Figure 1.4).

### *Animal Models*

The two-hit mechanism of CCM pathogenesis has been implemented in mice to generate a phenotype closely resembling the human disease. Inducible homozygous deletion of *Krit1*, *Ccm2*, or *Pcdh10* leads to robust formation of CCMs. However, one limitation of these models is that the capacity to form CCM is highly dependent on the time of gene deletion. If gene deletion occurs on postnatal day 1 (P1), the mice develop numerous large CCMs with strong enrichment in the cerebellum. After P1 the lesion burden steadily decreases and if gene deletion occurs later than P10, the mice do not develop CCMs even when aged (Boulday et al., 2011; Detter et al., 2020). This phenomenon likely reflects a dependence on active angiogenesis for CCM formation in these mice. During development there is a well-established ‘angiogenic window’ where the vasculature in the brain is actively growing (Boulday et al., 2009). Notably, angiogenesis is strongest in the cerebellum in the first few days after birth, which is consistent with the strong enrichment of CCMs in the cerebellum at this timepoint (Boulday et al., 2009). This facet of the CCM mouse models is at odds with the human phenotype. While many CCMs are congenital, there are also many cases of CCM development in adulthood, a time when the vasculature in the human brain is largely quiescent. This suggests that there may be additional events in human CCMs that fuel angiogenesis independent of development, which will be a focus of Chapter 3.

## 2

### Two-Hit Mechanism of Hereditary Hemorrhagic Telangiectasia

---

This chapter is adapted from a study published in *AJHG* (Snellings et al., 2019)

## 2.1 Premise

Hereditary Hemorrhagic Telangiectasia (HHT) is a Mendelian disease characterized by the development of multiple focal vascular malformations consisting of arteriovenous malformations in visceral organs and telangiectasia in mucosal and cutaneous tissue. The genetic etiology of HHT has been established and is caused by mutations in *ENG* (McAllister et al., 1994), *ACVRL1* (Johnson et al., 1996), and rarely *SMAD4* (Gallione et al., 2004); all of which follow an autosomal dominant inheritance pattern. Despite our understanding of the genetics of and downstream pathways involved in HHT, the molecular mechanisms that initiate HHT-related vascular malformation are poorly understood. Early studies of the functional consequences of HHT causal mutations established that these result in the loss of function of the gene product. These findings, in combination with autosomal dominant inheritance, led to the presumption that vascular malformations result from haploinsufficiency of the mutated gene product (Pece et al., 1997; Abdalla et al., 2000; Ola et al., 2018). However, haploinsufficiency does not account for why HHT-related vascular malformations occur as strictly focal lesions, despite the systemic presence of the causal germline mutation. This disconnect between genotype and phenotype led to an alternative long-standing hypothesis that HHT-related vascular malformations result from a Knudsonian two-hit mechanism, where a local somatic mutation in the wild type allele of the affected gene seeds the formation of focal lesions.

The only published study which directly addresses the two-hit hypothesis attempted to determine whether Endoglin was present on the endothelial lining of arteriovenous malformations from an individual with HHT with a causal mutation in the corresponding gene *ENG* (Pece-Barbara et al., 1999). Endoglin immunostaining was visible in the vessel lining, albeit at low levels. The presence of Endoglin in HHT-associated vascular malformations would contradict a two-hit mechanism,

however complete loss of staining might not be predicted to occur, especially with the heterogeneous—and potentially mosaic—tissue of an arteriovenous malformation that may only contain a minority of cells that harbor the somatic mutation. Previous attempts to address this hypothesis at the DNA level have been hampered by the limitations of past sequencing technology. The advent of next-generation sequencing has drastically increased our sensitivity for detecting low-frequency somatic mutations. Somatic mutations have been identified in a diverse array of vascular malformations (Al-Olabi et al., 2018; Soblet et al., 2017; Limaye et al., 2015, 2009; Shirley et al., 2013; Couto et al., 2015; Luks et al., 2015) including recent evidence that sporadic arteriovenous malformations, not associated with HHT, harbor somatic activating mutations in *KRAS* or *MAP2K1* (Nikolaev et al., 2018; Couto et al., 2017). Notably, a genetic two-hit mechanism is known to contribute to Cerebral Cavernous Malformations (CCM) (Akers et al., 2009; McDonald et al., 2014; Gault et al., 2009) and Capillary Malformation-Arteriovenous Malformation Syndrome (CM-AVM) (Macmurdo et al., 2016); like HHT, both diseases are caused by autosomal dominant loss of function mutations. Here we demonstrate that HHT-related telangiectasia contain biallelic mutations in *ENG* or *ACVRL1*, resulting in homozygous loss of function; evidence in support of the long-standing hypothesis that telangiectasia pathogenesis follows a genetic two-hit mechanism.

## 2.2 Results

To determine whether a genetic two-hit mechanism underlies HHT pathogenesis, we tested three underlying expectations of the two-hit mechanism: 1) telangiectasia contain a somatic mutation in the same gene as a germline mutation which causes HHT (Figure 2.1A), 2) the somatic and germline mutations are biallelic (Figure 2.1B), and 3) both mutations result in loss-of-function (Figure 2.1C).

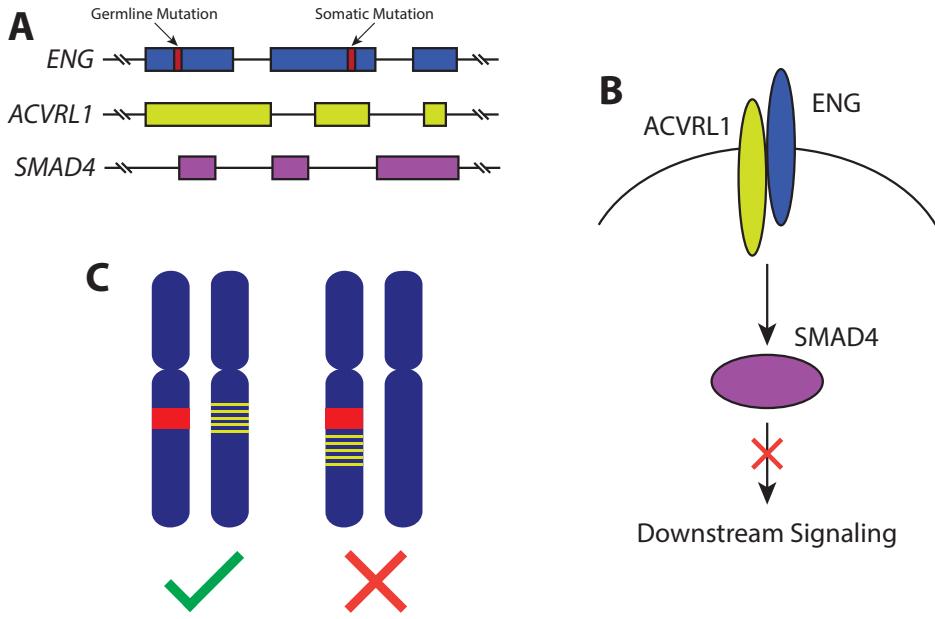


FIGURE 2.1: Expectations of a Two-Hit Model of HHT.

The expectations of a two-hit model are: (A) a germline and somatic mutation in the same gene, (B) both the germline and somatic mutations result in loss of function, and (C) that the germline and somatic mutations exist in a *trans* configuration (i.e. biallelic).

#### 2.2.1 Telangiectasia Harms or Somatically Mutated in *ENG* or *ACVRL1*

We used capture-based library preparations to sequence 19 telangiectasia for the three genes mutated in HHT (*ENG*, *ACVRL1*, and *SMAD4*) and 13 other vascular malformation-related genes (see Methods for identity of genes). The 13 non-HHT vascular malformation genes were chosen with the possibility that they also might harbor somatic mutations, but primarily to serve as control genes, since the two-hit mechanism requires a mutation in the corresponding HHT gene harboring the causal germline mutation. Somatic mutations in these other genes may or may not contribute to HHT pathogenesis, but absence of a somatic mutation in the casual HHT gene would violate the first expectation of the genetic two-hit mechanism.

In each telangiectasia we identified a pathogenic germline mutation in either *ENG*

or *ACVRL1*. Although in most cases the individuals' germline mutation was already known from clinical diagnostic sequencing, we intentionally remained blinded to this information until after our own sequence analysis of the tissue samples. In 6003-1, the individual harbors a silent germline mutation which was found by the clinical lab and noted as a variant of unknown significance. Below we show that this variant is indeed the pathogenic germline variant in this individual.

We used the MuTect2 variant caller to detect variants present in the sequence data. To identify candidate somatic mutations, we removed variants based on several stringent filtering criteria including briefly; intronic or intergenic variants, population frequency >0.01%, <0.1% or <5 total supporting reads, low coverage, strand specificity, and low base quality scores. To validate or refute the authenticity of each candidate somatic mutation we performed an independent round of amplification using primers flanking each putative variant position for each sample, and sequenced to >10000x coverage. In each tissue sample, the identical somatic variant was re-identified. Thus, these variants were bona fide somatic mutations existing in the telangiectatic tissue. In total, we identified somatic variants in 9 of 19 telangiectasia; 5 in *ENG*(NM\_001114753.1 (*ENG\_v001*)) 4 in *ACVRL1*(NM\_000020.2 (*ACVRL1\_v001*)) (Table 2.1) (Figure 2.2) (See Methods). In each case, the somatic mutation was found in the same gene as the pathogenic germline mutation. Somatic mutations were not found in any of the other 15 genes sequenced, not even in one of the other HHT casual genes. Importantly, no telangiectasia harbored more than a single somatic mutation. The lack of mutational noise suggests these mutations are pathobiologically significant. Importantly, all are consistent with strong mutations; five of the variants are small indels resulting in a frameshift, three are in-frame indels, and one is a point mutation 4 bases after an exon-intron boundary that is predicted to impact RNA splicing.

Although these variants fall well below the 50% allele frequency expected for

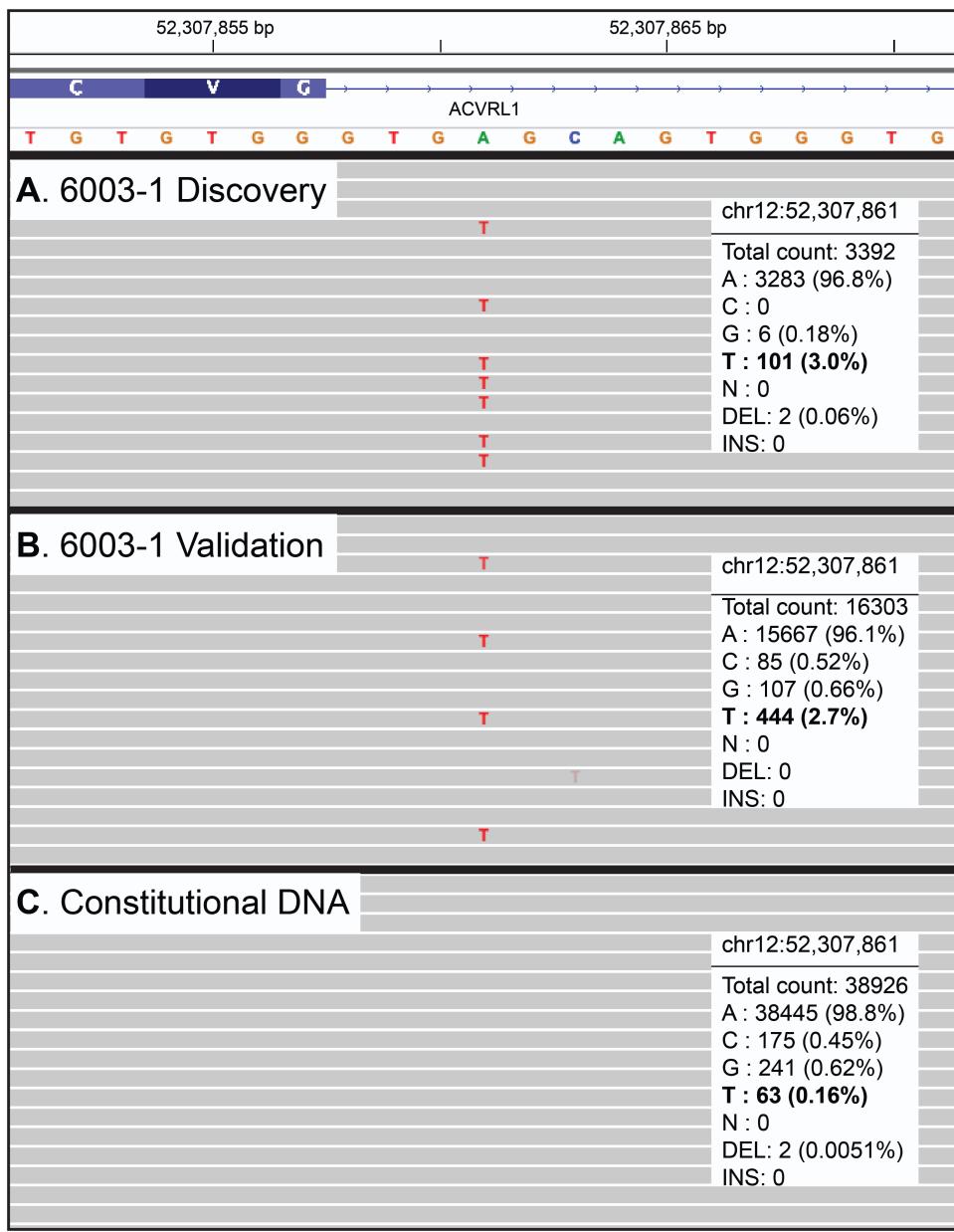
Table 2.1: Mutations Identified in HHT Telangiectasia.

Sample	Germline Mutation	Somatic Mutation	Discovery Reads <sup>a</sup>	Validation Reads <sup>a</sup>	Constitutional Reads <sup>a</sup>
6001-1	<i>ENG</i> c.1080_1083del	<i>ENG</i> c.293_304del	33/1318 (2.5%)	1067/100268 (1.1%)	0/26462 (0%)
6001-3	same as above	<i>ENG</i> c.1195_1196delAGfsX2	5/1080 (0.46%)	723/115963 (0.62%)	0/24357 (0%)
6001-7	same as above	<i>ENG</i> c.1237_1238insCAfsX7	27/5127 (0.53%)	341/115570 (0.30%)	0/23066 (0%)
6001-8	same as above	<i>ENG</i> c.578delCinsTGC p.T193MR	111/4845 (2.3%)	1142/142572 (0.80%)	0/21315 (0%)
6001-10	same as above	<i>ENG</i> c.205delGfsX6	33/3389 (1.0%)	3575/326894 (1.1%)	0/22098 (0%) <sup>b</sup>
6001-*	same as above	NF			
6002-1	<i>ACVRL1</i> c.1451G>A p.R484Q	<i>ACVRL1</i> c.349delGinsTTfsX52	20/2217 (0.90%)	309/24018 (1.3%)	0/65818 (0%)
6002-2	same as above	<i>ACVRL1</i> c.1378_3_1402del19ins9	26/1649 (1.6%)	3189/202550 (1.6%)	6/155855 (0.0038%)
6003-1	<i>ACVRL1</i> c.474A>T p.G158G	<i>ACVRL1</i> c.625+4A>T	101/3392 (3.0%)	372/16303 (2.3%)	2/38924 (0.0051%)
6004-1	<i>ACVRL1</i> c.1232G>A p.R411Q	NF			
6004-2	same as above	NF			
6005-1	<i>ACVRL1</i> c.1232G>A p.R411Q	<i>ACVRL1</i> c.1206delCfsX12	133/1664 (8.0%)	2671/189690 (1.4%)	N/A

\*Eight additional telangiectasia from patient 6001 with no identified somatic mutation

For multiple telangiectasia collected from one individual the sample ID is listed as (Patient#)-(Telangiectasia#), NF = None Found)

<sup>a</sup>Allele frequency in other telangiectasia from 6001



**FIGURE 2.2: Low Frequency Somatic Mutations Detected in Telangiectasia.**  
 Visualization with IGV of next-generation sequencing data for one representative sample with a somatic mutation in *ACVRL1*. The somatic mutation in 6003-1, an A>T point mutation, is present at low frequency in DNA from telangiectatic tissue used for (A) discovery (capture-based sequencing) and (B) validation (amplicon-based sequencing). The mutation is below the level of sequencing noise in (C) constitutional DNA (amplicon-based sequencing) confirming the mutation is somatic.

germline variants, it is formally possible that these variants exist constitutionally as very rare, somatic mosaic variation in the individual. We next investigated whether these somatic mutations were present in constitutional DNA from the individual. A source of constitutional DNA (saliva) was available for three of the nine mutation-positive samples, and in each we find that the somatic mutation was completely absent or present at a level no higher than technical sequencing noise in that sample (see Methods). Saliva was not available for 6001, but we obtained and sequenced DNA from multiple telangiectasia collected from this same individual. This enabled us to determine whether any of the five somatic mutations that we identified in individual samples was present in tissue of near identical pathobiology from the same individual; compared with saliva as a control, this is a more powerful test for somatic mosaicism. We found that the somatic mutations identified in five of the telangiectasia for which we identified a mutation were entirely absent in all other telangiectasia from this same individual. Finally, Sample 6005-1 is a single archived FFPE telangiectasia and no source of constitutional DNA is available. In total, we found that 9 of 19 telangiectasia harbor a somatic mutation specifically in the same gene as a pathogenic germline mutation and that these mutations are not present constitutionally. The presence of somatic mutations in telangiectasia fulfills the 1st expectation of the genetic two-hit mechanism.

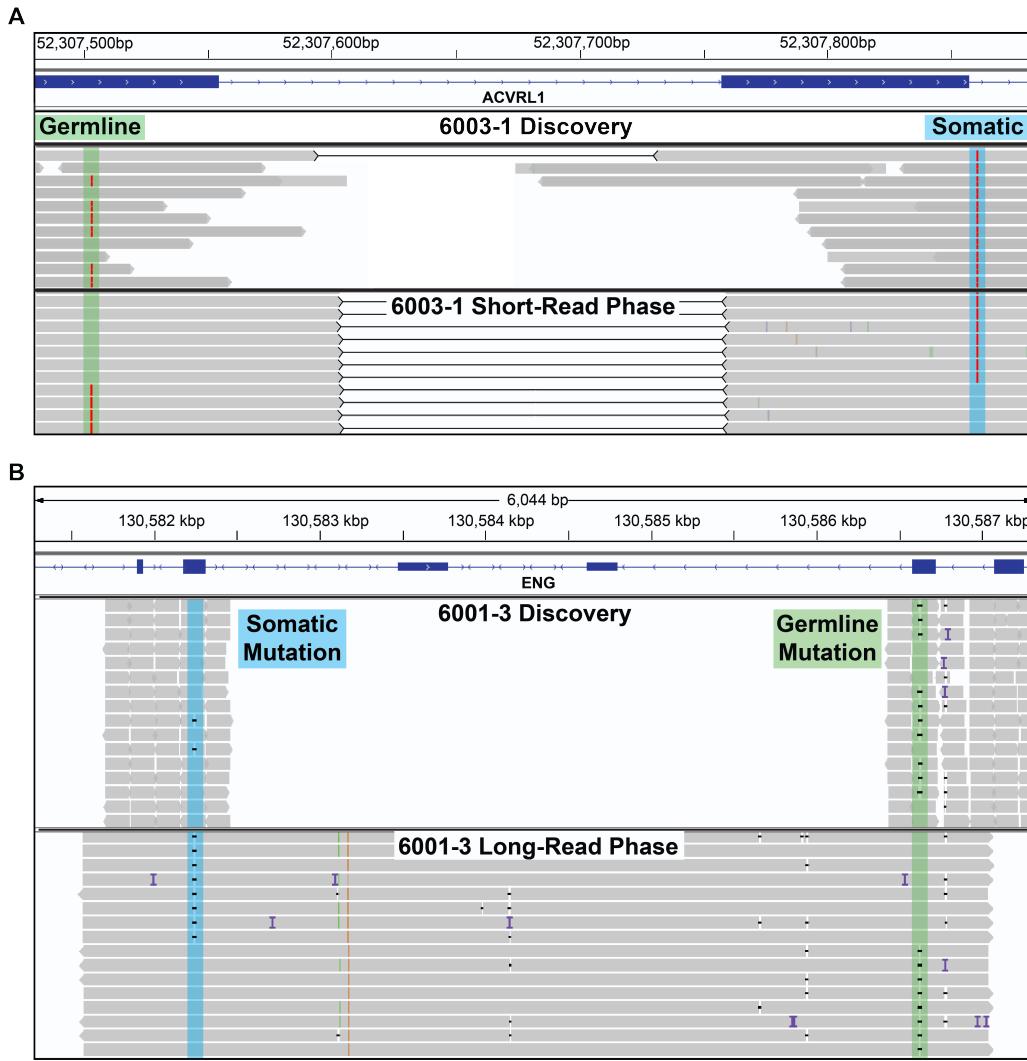
### *2.2.2 Somatic and Germline Mutations are Biallelic*

The 2nd expectation of the genetic two-hit mechanism is that the somatic and germline mutations are biallelic; such that the somatic mutation occurs on the wild-type allele of the affected gene, in trans with the germline mutation. To determine if the mutations are biallelic we examined whether they were arranged in a cis or trans configuration by sequencing amplicons that cover the nucleotide positions of both somatic and germline mutations in a single molecule. The amplicons

were sequenced with either short-read (Illumina) or long-read (PacBio) chemistry, depending on the amplicon size, in order to generate reads that would span the two mutations. In contrast to traditional Sanger sequencing which measures the population average at each position, both Illumina and PacBio chemistries output sequences of single DNA molecules. In total we generated mutation-spanning reads for 7 telangiectasia, each with more than 100 reads that contained the somatic mutation. From these mutation-spanning reads we established that >95% of reads with the somatic mutation possessed the wild type allele at the position of the germline mutation, showing that all 7 mutation pairs are in trans configuration (Table 2.2) (Figure 2.3A-B). Any two variants in a chromosome must be arranged in cis or trans with an equal probability of either arrangement. Considering this, our observation that 7/7 mutation pairs are arranged in trans corresponds to a p-value of 0.008 demonstrating significant bias towards a trans configuration. These data show that the somatic and germline mutations are biallelic, fulfilling the 2nd expectation of the genetic two-hit mechanism.

#### *2.2.3 Mutations are Consistent with Homozygous Loss of Function*

The 3rd expectation of the genetic two-hit mechanism is that the biallelic somatic and germline variants both result in loss of function. Due to the functional studies and extensive allelic series of mutations in each of the HHT genes, HHT is known to be caused by loss of function mutations. The germline mutation in 4 of the 5 individuals in this study has been identified previously in an individual with HHT and are reported in ClinVar (6001:VCV000213214.2, 6002:VCV000212796.2, 6004/6005:VCV000008243.2). There are also several publications supporting the pathogenicity of these mutations (Johnson et al., 1995; Bossler et al., 2006; Gallione et al., 1998; Ricard et al., 2010; Olivieri et al., 2007). These are all therefore bona fide loss of function mutations. The germline mutation in 6003-1, a silent mutation



**FIGURE 2.3: Establishing Phase of Germline and Somatic Mutations.**  
 IGV visualization of two samples showing both methods of establishing phase. Each panel shows reads from the initial discovery sequencing and the reads used to establish phase. **A**, Somatic and germline mutations in 6003-1 are both A>T point mutations and highlighted by the blue and green regions respectively. Since the distance between these mutations is relatively small (357bp) phase was established using Illumina short-reads (also used for validation). Black lines between reads denote read pairs, showing that both reads originate from a single molecule of DNA. Each molecule with the somatic mutation contains the wild-type allele at the germline mutation position proving the mutations are biallelic. **B**, Somatic and germline mutations in 6001-3, both small deletions. The genomic distance between these mutations is 4377bp. In long reads that span the two mutations, each read with the somatic mutation contains the wild-type allele at the position of the germline mutation.

Table 2.2: Phase of Somatic and Germline Mutation Pairs.

Sample	Total Reads	Trans Reads	Cis Reads	P-value
6001-1	N/A	N/A	N/A	N/A
6001-3	112	112 (100%) <sup>a</sup>	0 (0%)	$1.9 \times 10^{-34}$
6001-7	155	153 (98.7%) <sup>a</sup>	2 (1.3%)	$2.6 \times 10^{-43}$
6001-8	593	590 (99.5%) <sup>a</sup>	3 (0.5%)	$1.0 \times 10^{-171}$
6001-10	N/A	N/A	N/A	N/A
6002-1	125	120 (96.0%) <sup>a</sup>	5 (4.0%)	$5.5 \times 10^{-30}$
6002-2	3189	3160 (99.0%) <sup>bc</sup>	29 (1.0%)	$4.2 \times 10^{-890}$
6003-1	372	364 (97.8%) <sup>bc</sup>	6 (1.4%)	$1.4 \times 10^{-99}$
6005-1	2671	2653 (99.3%) <sup>bc</sup>	18 (0.7%)	$6.3 \times 10^{-759}$

<sup>a</sup>Reads generated with PacBio long-read chemistry)

<sup>b</sup>Reads generated with Illumina short-read chemistry

<sup>c</sup>These reads were also used for validation shown in Table 2.1

in *ACVRL1* exon 4, has been identified before in an individual with HHT, however it was classified as a variant of unknown significance (VUS). We used the in silico tool Human Splicing Finder 3.1 (Desmet et al., 2009) to analyze this variant and found that it was predicted to both disrupt an exonic splice enhancer and create an internal splice donor site, potentially activating a cryptic splice site. Based on this prediction, we extracted RNA from peripheral blood leukocytes of 6003 and a control individual and performed RT-PCR to examine the splicing of *ACVRL1* transcripts. RNA from 6003 shows a new splice variant that is not present in control wild-type RNA (Figure 2.4B). As predicted by the Splice Finder program, the aberrant transcript is spliced precisely at the internal splice donor created by the mutation. The resulting transcript is missing the portion of exon 4 downstream of the germline mutation and skips exon 5 resulting in the in-frame deletion of 52 amino acids (Figure 2.4D-E). This deleted region contains several codons with known pathogenic missense mutations, suggesting that the 52 amino acid deletion would likely also result in loss of function.

It is possible that the skipping of exon 5 is due to alternative splicing observed only in peripheral blood leukocytes, rather than a result of the mutation. If exon 5 is retained, the mutation would then generate a protein lacking 17 amino acid residues from exon 4 but then be frameshifted for the remainder of the transcript. With this evidence, all of the identified germline mutations meet the American College of Medical Genetics (ACMG) criteria for pathogenic mutations (Richards et al., 2015), fulfilling the first half of the 3rd expectation of the genetic two-hit mechanism.

In contrast to the germline mutation, many of which have been previously identified, the somatic mutations we identified are all novel. The ACMG guidelines for establishing pathogenicity are not applicable to somatic variants, however several lines of evidence support that all of the somatic mutations result in loss of function. Five of the nine somatic mutations result in a frameshift; all resulting in premature termination codons which would generate transcripts susceptible to nonsense-mediated decay. Frameshift mutations in *ENG* or *ACVRL1* are the most common mechanism for loss of function leading to HHT. Based on this, the 5 somatic frameshift mutations likely result in loss of function. Other than frameshift mutations, the other 4 somatic mutations we identified consisted of 3 in-frame deletions and 1 intronic mutation predicted to impact splicing. These 4 mutations are not present in the genome aggregation database (gnomAD) showing that the population allele frequency of these variants is extremely low or zero. For the somatic in-frame deletion mutation found in 6001-8, there are two reports of different in-frame deletions with overlap at this position which are known to cause HHT, suggesting that the somatic deletion in 6001-8 is likely to result in loss of function. The somatic mutations in 6001-1 and 6002-2 also result in in-frame deletions, which delete 4 and 7 amino acids respectively. Comparing the crystal structures of *ENG* and *ACVRL1* we determined that the somatic mutations in 6001-1 and 6002-2 delete portions of a beta strand and helix respectively, potentially impacting protein folding (Table 2.3). The

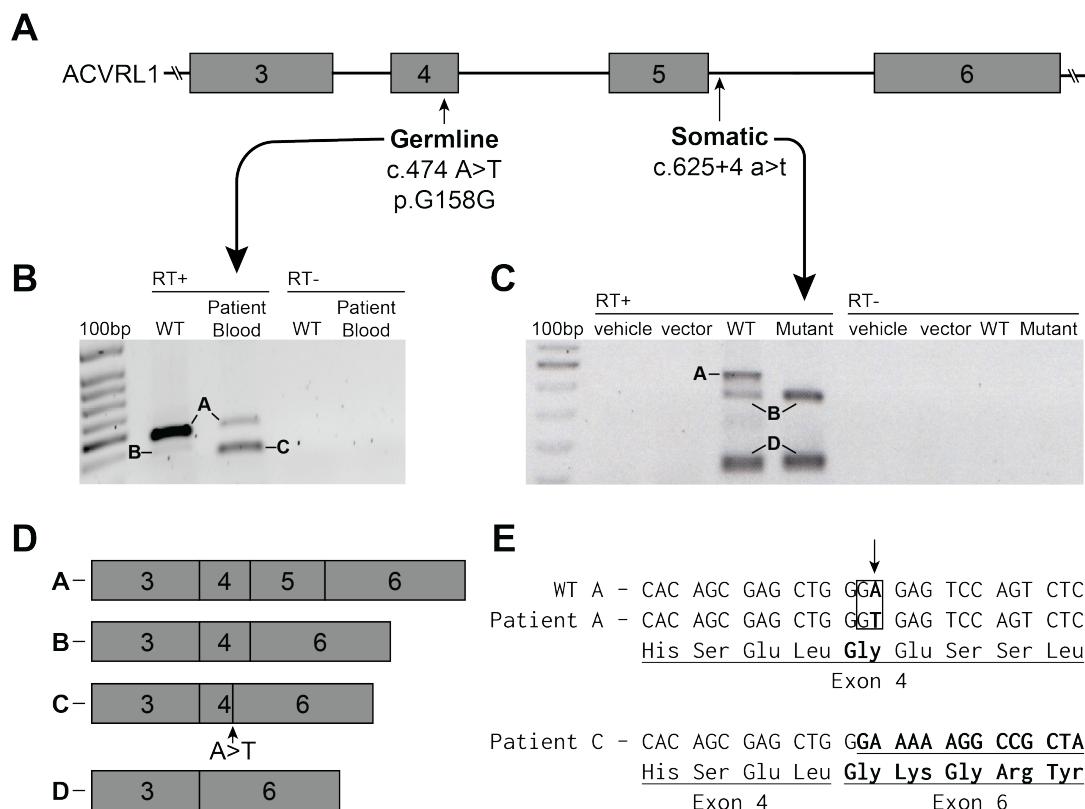


FIGURE 2.4: Mutations in *ACVRL1* Disrupt Splicing.

**A**, The gene structure of *ACVRL1* exons 3–6 marked with the location of the germline and somatic mutations found in 6003-1. **B**, RT-PCR showing *ACVRL1* transcripts from peripheral blood leukocytes taken from wild-type (WT) and peripheral blood leukocytes containing the germline mutation. The labeled bands were excised and sequenced. Full length transcript (Band A) is present in both the control and leukocytes from 6003 although the level in leukocytes from 6003 is greatly reduced. Band B in the control contains complete exons 3, 4 and 6. This splice variant has been seen previously and differs from band C in 6003 which splices from the newly created splice donor site within exon 4 directly to exon 6. **C**, As the somatic mutation is only present in 3.0% of reads it would be challenging to detect misspliced RNA from the biopsied tissue. Therefore, we inserted wild-type and mutant sequence of *ACVRL1* into an in vitro splicing vector, pSPL3-*ACVRL1*, and used RT-PCR to visualize the impact of the mutation on splicing. Only WT band A shows the full-length transcript containing exon 5; there is no corresponding full-length transcript from the plasmid containing the somatic mutation. **D**, Exon structure of *ACVRL1* transcripts determined by sequencing the excised bands. **E**, Sequence of DNA showing the nature of the germline mutation. In *ACVRL1* transcripts containing the germline mutation, exon 4 is shortened due to the activation of a cryptic splice site.

in silico tool PROVEAN was used to predict how the protein would tolerate these deletions. The threshold of -2.5 or lower (more negative) is considered a deleterious change. The scores for 6001-8, 6001-1, and 6002-2 were -6.106, -14.903, and -25.903 respectively, strongly suggesting that all three are deleterious. The remaining somatic mutation is intronic, and occurs 4 nucleotides from the exon-intron boundary. We used Human Splicing Finder 3.1 to predict the effect of this variant on splicing, and found that it is likely to disrupt the donor site. This prediction was confirmed by RT-PCR using an in vitro splicing construct which revealed that the somatic mutation prevents the formation of full-length *ACVRL1* transcripts (Figure 2.4C). In summary we present evidence supporting that the biallelic germline and somatic mutations all likely result in loss of function, fulfilling the 3rd expectation of the genetic two-hit mechanism.

#### 2.2.4 *Telangiectasia from the Same Individual Harbor Unique Somatic Mutations*

We next sought to determine whether mutant cells in different telangiectasia derive from a somatic mutation in a common ancestor cell, or whether the mutant cell population in each telangiectasia derives from an independent somatic mutation event. To test this, we examined the somatic mutations present in multiple telangiectasia from single individuals. In 6001, for which we had obtained 13 different telangiectasia, we identified a somatic mutation in 5 telangiectasia tissue samples. In each case the somatic mutation in each telangiectasia was unique. Likewise, for 6002, for which we had two telangiectasia, we identified a unique somatic mutation in each (Figure 2.5). These results are consistent with independent mutation events rather than the somatic mutation occurring in a progenitor cell or clonality due to a metastasis from a single initial lesion.

Table 2.3: Predicted Consequences of Germline and Somatic Mutations.

Sample	Germline Mutation	Somatic Mutation
6001-1	Frameshift: PVS1 6 supporting publications	In-Frame Deletion (-4 residues) PROVEAN: Deleterious (-14.903) deletes region in $\beta$ -sheet gnomAD AF: 0
6001-3	same as above	Frameshift common ENG LOF mechanism, expect NMD
6001-7	same as above	Frameshift common ENG LOF mechanism, expect NMD
6001-8	same as above	In-Frame Delins (-1 +2 residues) 2 pathogenic in-frame indels overlapping this codon (Shovlin et al., 1997; Argyriou et al., 2006) PROVEAN: Deleterious (-6.106) gnomAD AF: 0
6001-10	same as above	Frameshift common ENG LOF mechanism, expect NMD
6002-2	same as above	In-Frame Delins (-7 +1 residues) PROVEAN: Deleterious (-25.903) deletes region in $\alpha$ -helix gnomAD AF: 0
6003-1	Cryptic Splice Site: PS3 in silico predicted to activate cryptic site in vitro evidence (Figure 2.4)	Splice Site in silico predicted to disrupt donor site gnomAD AF: 0
6005-1	Missense: PS1 16 supporting publications	Frameshift common <i>ACVRL1</i> LOF mechanism, expect NMD

AF = Allele Frequency; NMD = Nonsense Mediated Decay

Germline variant classification according to ACMG guidelines (Richards et al., 2015)

PVS1 = Very strong evidence for pathogenicity; PS1-4 = Strong evidence

PROVEAN scores below -2.5 are predicted deleterious

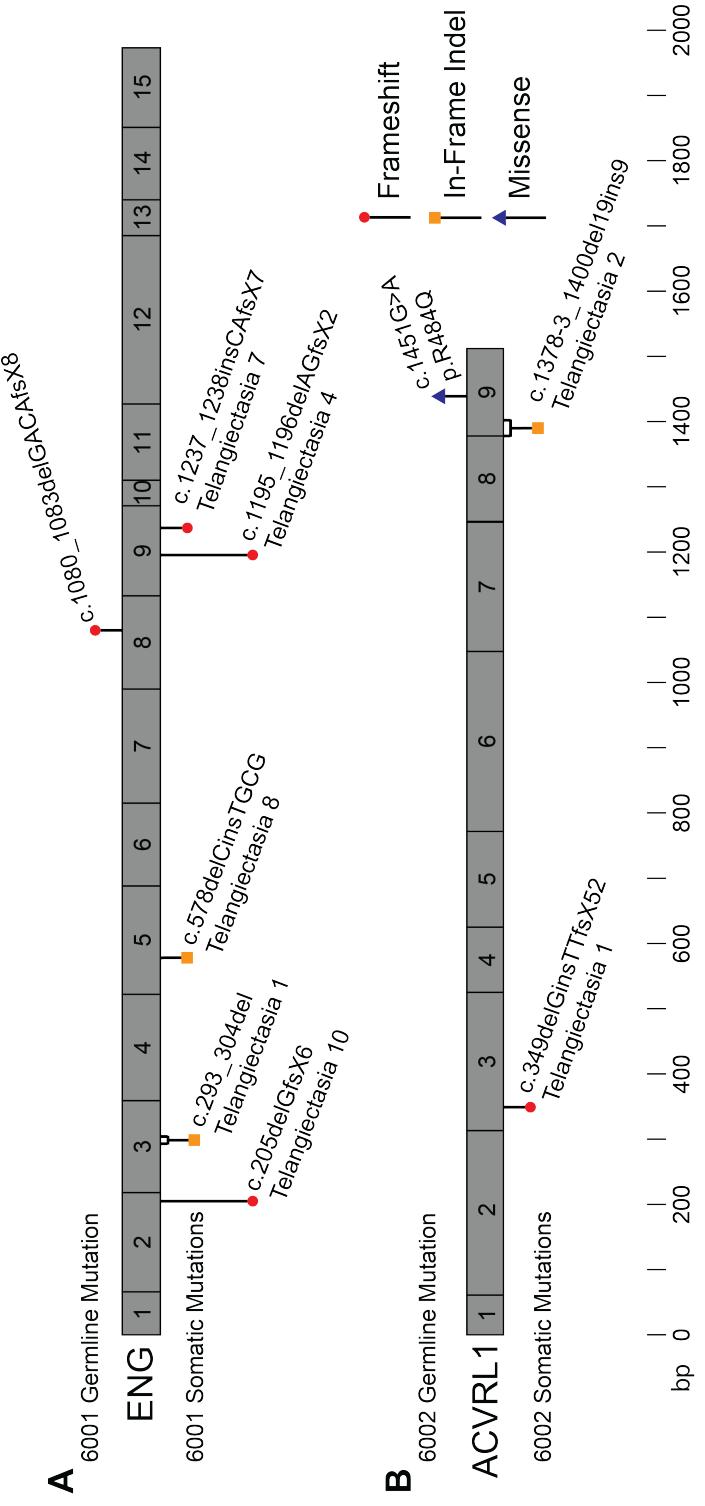
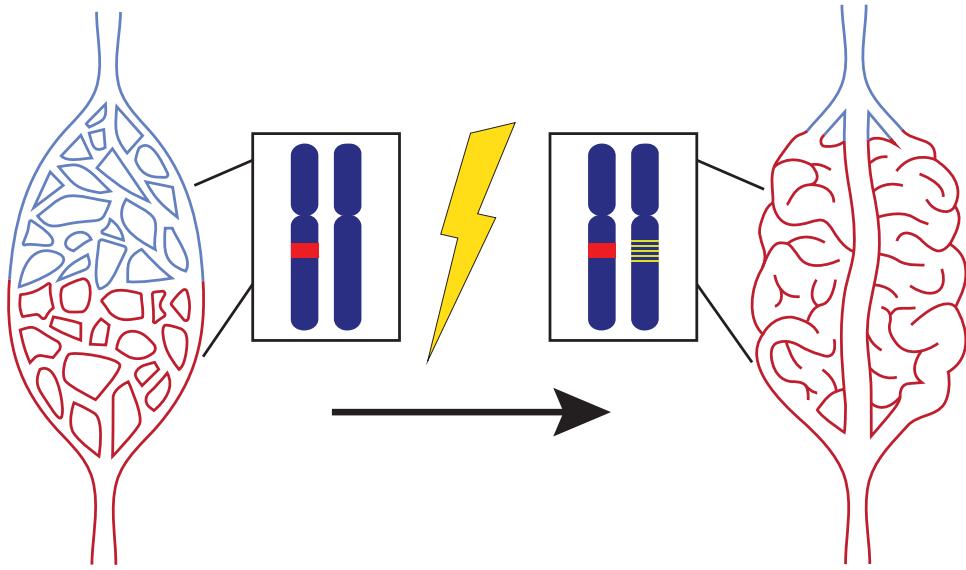


FIGURE 2.5: Each Telangiectasia is Seeded by a Unique Somatic Mutation.

Schematic representation of exons in *ENG* and *ACVR1L* with germline and somatic mutations identified in (A) 5 telangiectasia collected from 6001 and (B) 2 telangiectasia collected from 6002. In each panel the common germline mutation is listed above the gene and somatic mutations in each telangiectasia below the gene. Gene structure and mutation position are drawn to scale.



**FIGURE 2.6: Two-Hit Model of HHT Pathogenesis.**

Each endothelial cell in someone with HHT is heterozygous for either *ENG*, *ACVRL1*, or *SMAD4* (red bar). A somatic mutation (yellow stripes) occurs in the wild type allele of the same gene with the inherited mutation to induce the formation of an AVM.

## 2.3 Discussion

### 2.3.1 Evidence for a Genetic Two-Hit Mechanism

In this study we present strong evidence that vascular malformations associated with HHT, specifically, cutaneous telangiectasia, follow a genetic two-hit mechanism of pathogenesis (Figure 2.6). HHT is also associated with arteriovenous malformations in lung, liver, brain and the gastrointestinal tract, but these tissues are not available for prospective collection. We postulate that the visceral, deeper vascular malformations that occur in HHT also follow this two-hit mechanism.

The two-hit hypothesis for HHT pathogenesis has persisted for decades without evidence, but these low frequency somatic mutations are challenging to identify using traditional sequencing methods. The only published study to address this topic employed immunohistochemical staining in an attempt to identify endothelial cells

lacking staining in the lining of HHT-related arteriovenous malformations (Bourdeau et al., 2000). However, the absence of staining as a proxy for the gain of a mutation could be difficult to discover, especially if only a fraction of a cells would exhibit this lack of signal.

Using next-generation sequencing with unique molecular identifiers we successfully identified somatic mutations in multiple telangiectasia from different individuals with HHT. The somatic mutations we identified were present at frequencies ranging between 0.46% and 8.0% in the tissue, with an average of 2.3%. The low allele frequency is likely a result of two main contributing factors; the presence of normal tissue in the skin biopsy, and somatic mosaicism within the telangiectasia. The telangiectasias in this study were sampled as skin punch biopsies and although some of the surrounding skin tissue was removed before DNA extraction, an undetermined amount of normal tissue invariably remained. However, after removing the surrounding tissue, the enrichment for the vascular component of the tissue was subjectively greater than the often low somatic mutation allele frequency might suggest. We posit that a second explanation for the low mutant allele frequency is that telangiectasia are mosaic for the somatic mutation. This agrees with existing data from mouse models of HHT showing that induced retinal AVMs are mosaic: consisting of both heterozygous and homozygous null cells (Jin et al., 2017). It is also consistent with the heterogeneity seen in Cerebral Cavernous Malformations (Detter et al., 2018; Malinverno et al., 2019) and the low mutant allele frequency in somatic mutations of other vascular malformation disorders (Al-Olabi et al., 2018; Soblet et al., 2017; Limaye et al., 2015, 2009; Shirley et al., 2013; Couto et al., 2015; Luks et al., 2015; Nikolaev et al., 2018; Couto et al., 2017; Akers et al., 2009; McDonald et al., 2014). Vascular malformations, similar to tumors in cancer, appear to be seeded by somatic mutations; however unlike tumors, vascular malformations do not appear to consist of pure populations of clonally expanded mutant cells, but contain a substantial

percentage of unmutated cells. The probability of a single cell acquiring biallelic LOF in an individual with HHT is roughly estimated in Appendix A.

In addition to the data presented here, the two-hit hypothesis of HHT-related vascular malformations is consistent with observations from mouse models of the disease. Whereas constitutional loss of both copies of *Eng* or *Acvrl1* in mice is embryonic lethal, mice heterozygous for constitutional deletion of either gene show extremely mild phenotypes with relatively few, if any, detectable vascular malformations(Bourdeau et al., 1999; Srinivasan et al., 2003). Robust mouse models of HHT that recapitulate vascular malformation phenotypes require the use of Cre-Lox technology to delete both copies of *ENG* in a temporally controlled (postnatal), cell-type specific (endothelial cells) manner.

In these mouse models of HHT it is also required that biallelic KO of the HHT gene occur in endothelial cells. Several groups have experimented with expressing Cre recombinase in different vascular-related cell types including pericytes (*NG2*-Cre), vascular smooth muscle cells (*Myh11*-Cre), and endothelial cells (*Scl*-Cre & *Pdgfb*-Cre); however, mice only develop vascular malformations when Cre is expressed in endothelial cells (Tual-Chalot et al., 2015; Choi et al., 2014; Garrido-Martin et al., 2014; Mahmoud et al., 2010). We have attempted to confirm that the somatic mutations we identified in human lesions occur in the endothelium by using laser capture microdissection, however these efforts have been hampered by the small quantity of tissue in telangiectasia biopsies and the difficulty of isolating a single layer of cells by microdissection. This question may be more easily addressed in larger arteriovenous malformations; however, these samples have been thus far inaccessible due to the rarity of their removal in individuals with HHT.

### 2.3.2 Necessary, but Not Sufficient

Interestingly, in addition to the local requirement for loss of both alleles, vascular malformations in this model only develop after injury such as an ear punch, or by VEGF injection (Choi et al., 2012). This requirement of an angiogenic stimulus is consistent in mouse models for all HHT genotypes: *Eng*, *Acvrl1*, and *Smad4* (Kim et al., 2018). The requirement for knockout of both copies of the gene supports the genetic two-hit mechanism we describe here. In addition, the necessity for an angiogenic stimulus suggests that loss of both copies of the relevant HHT gene is necessary, but not sufficient, for the development of the vascular malformation.

### 2.3.3 Sensitivity for Detecting Somatic Mutations

We might have expected to find somatic mutations in every telangiectasia, however we only found somatic mutations in 9 of the 19 we sequenced. The next-generation sequencing strategy we employ for discovering somatic mutations is extremely sensitive for the detection of point mutations and small indels. However, there are several other types of genetic alterations that would result in biallelic loss of function due to loss of heterozygosity (LOH). LOH is a common occurrence in many tumors in cancer, and is a predominant mechanism of somatic loss/mutation. LOH can occur due to a variety of genetic mechanisms: large deletions, chromosome loss, and mitotic recombination. Given the apparent capacity for even a low fraction of somatically mutant cells to initiate the vascular malformation, it follows that the same would be true for LOH-associated mutational events; therefore, these mutations would appear instead as allelic imbalance rather than outright LOH. But if the level of allelic imbalance is as low as the frequency of somatic mutations we have observed in this study, we might expect linked marker haplotype ratios in the range of 48% to 52% at nearby markers; this slight and even trivial imbalance would be difficult if not impossible to detect and validate. It is also possible that non-genetic mechanisms

such as loss of expression due to epigenetic silencing account for biallelic loss of function. This process, like the LOH associated events, would not be detected by our sequencing strategy; a problem that is only exacerbated by low allele frequency. Thus, it may not be surprising that we identified a somatic mutation in approximately only 9 of the VMs that were sequenced. We postulate it is highly likely that all of the 10 telangiectasia with no identified somatic mutation have biallelic loss by one of these other mechanisms.

#### 2.3.4 Mutant Cell Metastasis

One consequence of the genetic two-hit mechanism that might appear to be improbable is that, if true, a new somatic mutation must occur in every one of the numerous vascular malformations in HHT. For example, some affected individuals have dozens or more visible telangiectasia on the skin and mucocutaneous surfaces alone (Gonzalez et al., 2019; Letteboer et al., 2008; Plauchu et al., 1989). An attractive hypothesis to reconcile this conundrum would be if a somatic mutation first occurs in a circulating progenitor cell which then proliferates and seeds the formation of multiple telangiectasia. There is precedence for this mechanism in another vascular malformation syndrome, Blue Rubber Bleb Nevus syndrome. These individuals display multiple small vascular lesions which harbor an identical somatic double-mutation in the *TEK* gene. These vascular lesions appear to be anatomically-dispersed clones arising from an original, dominant, large lesion (Soblet et al., 2017). In HHT-related vascular malformations, we report evidence that contradicts this hypothesis: different telangiectasia collected from the same individual harbor different, unique somatic mutations. This observation does not exclude the possibility that circulating cells may in some cases spread telangiectasia, however the data thus far suggests that the primary mechanism is independent somatic mutation events.

### 2.3.5 Probability of Multiple Somatic Mutations

The dilemma of the requirement for numerous independent somatic mutation events in a single gene can be resolved with a probabilistic argument. Considering the size of the human genome ( $3.23 \times 10^9$  bp), the probability that a random somatic mutation occurs in the coding sequence of *ENG* (3201 bp) in trans (50% likelihood) with a pathogenic *ENG* germline mutation is  $\sim 0.00005\%$ . Compounding this value with empirical evidence that single cells have anywhere from 100 to 1500 somatic mutations per cell (Milholland et al., 2017; Lodato et al., 2015; Lo Sardo et al., 2017), and an estimate that 5.66% of exonic somatic mutations result in LOF (Milholland et al., 2017); we calculate a conservative estimate that 0.00028% of cell have biallelic LOF *ENG* mutations. An adult human has at least  $6 \times 10^{11}$  endothelial cells (Sender et al., 2016), therefore we estimate that an individual with HHT and a germline mutation in *ENG* has biallelic LOF *ENG* mutations in  $\sim 1.5$  million endothelial cells. It is clear that each cell with *ENG* biallelic LOF does not result in vascular malformation as individuals with HHT have at most hundreds, not millions, of telangiectasia. This disparity is consistent with the idea that telangiectasia only develop under very specific conditions: likely that the somatic mutation must occur in a specific type of vascular bed, in an endothelial cell, and must be followed by local angiogenic stimulus.

### 2.3.6 Two-Hit Mechanism for *SMAD4* & JP-HHT

Our samples consisted of telangiectasia from individuals with HHT from a single HHT Centre of Excellence. This cohort had germline mutations in either *ENG* or *ACVRL1*, and somatic mutations were identified in telangiectasia from both genotypes. Mutations in *SMAD4* cause the combined syndrome HHT and Juvenile Polyposis (JP-HHT), however individuals with *SMAD4* mutations only account for  $\sim 2\%$  of HHT cases (Gallione et al., 2004). Unfortunately, no individuals with JP-

HHT were present in our cohort, but we believe it is likely that JP-HHT-related telangiectasia follow an identical genetic two-hit mechanism, resulting from somatic mutations in *SMAD4*.

### 2.3.7 Compound Heterozygosity

As HHT is caused by germline mutations in any one of 3 genes, an interesting question is whether the compound effect of a LOF mutation in two different HHT genes could drive pathogenesis (e.g. a germline mutation in *ENG* and a somatic mutation in *ACVRL1*). However, thus far all of the somatic mutations we identified in HHT-related telangiectasia occur in the same gene as that harboring the germline mutation. This human genetic data is consistent with the observation that mice with combined deficiency for one allele each of *Acvrl1* and *Eng* are fertile, viable and are not teeming with vascular malformations (Eleftheriou et al., 2016) (Srinivasan and Marchuk, unpublished), as might be expected if trans-heterozygosity of mutations in these two HHT-genes could initiate vascular malformation development.

### 2.3.8 Therapeutic Potential

A genetic two-hit mechanism for HHT pathogenesis has therapeutic implications. Certain efforts to develop therapies for HHT assume a model of haploinsufficiency of the relevant HHT gene. These strategies attempt to increase the amount of the affected gene product by increasing the level of transcript/protein arising from the wild-type allele (Ruiz-Llorente et al., 2017). We show here that some fraction of cells within the malformation do not possess a wild-type allele. Even if expression in surrounding heterozygous cells could be increased, the null cells would remain devoid of protein, suggesting that this avenue of therapy may be ineffective. By contrast, a more effective strategy may be gene replacement, reintroducing a fully wild-type allele into the mutated cells (Seki et al., 2003), as this would simultaneously provide

an extra copy of the gene to both the heterozygous and null cells. This strategy is particularly attractive as it might inhibit new VM formation by adding back a second wild-type copy of the mutated gene.

## 2.4 Methods

### *Sample Collection*

Individuals were enrolled in the study after giving informed consent (approved by either the St. Michael's Hospital IRB committee or the Duke University Health System IRB Committee). Diagnosis of HHT was based on identification of a pathogenic germline mutation, or by exhibiting at least three of the four symptoms as per the Curaçao criteria (Supp Table 1) (Shovlin et al., 2000). Telangiectasia were resected using a 3mm punch biopsy, after local anesthesia (1% xylocaine with epinephrine), with standard aseptic technique. Sample 6005-1 was immediately formalin fixed (10% formalin) and paraffin embedded (FFPE), and then shipped at room temperature. All other samples were immediately frozen at -80 Celsius, and then shipped on dry ice. Saliva samples were obtained using Oragene DNA saliva kits at the time of tissue collection. Blood from individual 6003 was obtained during a subsequent visit, shipped at room temperature, and immediately used for RNA extraction.

### *DNA and RNA Extraction*

DNA from telangiectasia samples was extracted using the DNeasy Blood and Tissue Kit (Qiagen). DNA from FFPE sample 6005-1 was extracted using QIAamp DNA FFPE Tissue Kit (Qiagen). Genomic DNA and RNA were extracted from peripheral blood leukocytes from individual 6003 and from a non-HHT control individual using the Gentra PureGene Blood Kit (Qiagen) and TRIzol Reagent (Invitrogen) extraction protocols, respectively, as per the manufacturers' directions.

### *Targeted Sequencing*

To enable the detection of somatic mutations in telangiectasia we used a next generation sequencing strategy. Somatic mutations involved in the pathogenesis of other vascular malformation diseases such as cerebral cavernous malformations often have a low allele frequency due to somatic mosaicism in the malformation. Somatic mosaicism is also present in retinal AVMs from mouse models of HHT suggesting that low allele frequency may be a confounder when identifying somatic variants in telangiectasia. In addition, the telangiectasia samples collected for this study consist of bulk biopsied tissue which have not been enriched for any particular cell type. Considering these potential sources of normal (non-mutant) cell contamination, we sequenced telangiectasia to >1000x coverage and incorporated a unique molecular identifier to enable the detection of variants as low as 0.1% allele frequency. Eighteen fresh-frozen samples and one FFPE sample were sequenced using a custom Agilent SureSelect panel covering 16 genes implicated in various vascular malformation disorders: *ENG*, *ACVRL1*, *SMAD4*, *BRAF*, *CCM2*, *FLT1*, *FLT4*, *GNAQ*, *KDR*, *KRAS*, *KRIT1*, *MAP2K1*, *NRAS*, *PDCD10*, *PIK3CA*, and *PTEN*. Though *GDF2* variants have been identified in an HHT-like phenotype, these cases are extremely rare. Moreover, mutations in *GDF2* have not been identified in any individual at the Toronto HHT Centre for Excellence therefore *GDF2* was not included in the panel. To ensure the generation of high-quality sequencing libraries, the Agilent NGS FFPE QC Kit was used to determine the extent of DNA degradation in the FFPE sample. Samples with  $\Delta Cq$  values >2 were excluded from the study as per manufacturer recommendation. Sequencing libraries were generated using the Agilent SureSelect XT HS Kit. Samples were then pooled and sequenced on an iSeq 100 (Illumina) with paired-end 150bp reads. Across all samples, target regions were sequenced to a mean depth of 2803x with 78% of the target region at >1000x and 96% of the target

region at >100x.

### *Mutation Detection*

Sequencing data was processed and analyzed using a custom pipeline based on the GATK best practices for somatic short variant discovery. Briefly, after analyzing the raw data with fastQC to ensure high quality data, the adapter sequences were trimmed from reads using bbdduk, reads were aligned to the hg19 human reference genome using bowtie2, duplicates were removed based on UMI sequence using fgbio, variants were called using MuTect2 in tumor-only mode, and variants were annotated using snpSift. The resulting variant call file (VCF) was filtered through several steps to identify somatic mutations. To identify variants that may change the protein sequence or impact splicing, we selected for variants that occur within exons or within 10bp of an exon. From this set, we removed variants that are present in the population at >0.01% frequency by comparing to 3 databases (dbSNP, 1000 Genomes project, and the Exome Aggregation Consortium [ExAC]); as these variants are more common than the frequency of HHT (1–5 in 10000) (Grosse et al., 2014). We removed any variants present in <0.02% of sequence reads, as this is the reported technical limit of detection for the SureSelect XT technology. We also removed variants in regions with <100x coverage, variants with <5 supporting reads, variants that were strand specific, and variants where <50% of alternative bases had a quality score >30 (>Q30). Candidate somatic variants identified in the targeted sequencing data were then validated by sequencing amplicons generated during a second, independent round of PCR amplification. We designed primers for each sample to specifically amplify the position of the somatic mutation and 100–200bp of flanking sequence. When possible, the primers were designed such that they would capture the position of both the germline and somatic mutations within a single amplicon. Each primer was synthesized with the following Illumina flow cell adapter sequences such that the

amplicons could be easily indexed and sequenced:

For: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[primer]-3'

Rev: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[primer]-3'

Amplicons were prepared from telangiectasia DNA and constitutional DNA if available (Supp Table 1) using two rounds of PCR. The first round of PCR was 25 cycles and served to amplify the target region from genomic DNA. Amplicons from the first round were purified using AMPure XP beads (Beckman Coulter) and used for a second round of PCR using 8 cycles to attach a sample index using the Nextera XT Index Kit. Amplicons from the second round were purified, pooled, and sequenced on an iSeq 100 with 150bp paired-end reads to a depth of >10000x. The frequency of the somatic mutation in telangiectasia and constitutional DNA was determined using custom scripts, excluding bases <Q15. As these amplicons were sequenced in the same run, it is possible for a low level of index misassignment causing switching of reads between samples. This could cause some reads from telangiectasia to be assigned as constitutional reads and vice versa. To estimate the rate of index misassignment, we examined pairs of samples that target different genomic locations and quantified the proportion of misassigned reads between these samples. Based on this we estimate that the rate of misassignment is 0.2–0.8%, relative to the sample of origin. For example, if one sample has a somatic mutation with a frequency of 1% and was sequenced to 100000x coverage, then 2–8 reads containing the somatic mutation would be misassigned to each of the other samples in the pool.

#### *Establishing Phase*

To establish the phase of the somatic and germline mutations we used either short-read sequencing with Illumina chemistry, or long-read sequencing with PacBio chem-

istry depending on the distance between the two mutations. For mutations <500bp apart, we generated amplicons during the validation process that cover the positions of both the somatic and germline mutations. These amplicons were sequenced on an iSeq 100 as described above. For mutations >500bp we designed primers such that the resulting amplicon would cover the position of both the germline and somatic mutations for each sample. Each primer was synthesized with the PacBio ‘universal tag’ as follows to enable indexing and sequencing:

For: 5’-/5AmMC6/GCAGTCGAACATGTAGCTGACTCAGGTCAC-[primer]-3’

Rev: 5’-/5AmMC6/TGGATCACTTGTGCAAGCATCACATCGTAG-[primer]-3’

We generated amplicons spanning the somatic and germline mutations using the LongAmp Taq DNA Polymerase Kit as per manufacturer instructions. These amplicons were purified with AMPure XP beads and used for a second round of PCR to attach the sample index. This process used no more than 30 cycles of PCR total. These amplicons were pooled and sequenced across one SMRT cell on a PacBio Sequel System. The sequence reads were aligned to the hg19 human genome using Minimap2 in ava-pb mode. The single molecule resolution of these technologies allowed us to determine how the mutant alleles are arranged; if the mutations are in trans then reads will have either the somatic mutant allele or the germline mutant allele, if the mutations are in cis then reads will have either no mutant alleles or both mutant alleles. In total we generated mutation-spanning reads for 7 telangiectasia, each with >100 reads which contained the somatic mutation. The p-values reported for phase status were calculated using a binomial distribution with the null hypothesis that a random mutation has an equal probability of cis or trans configuration with a nearby variant. The genomic distance between mutations varied greatly with the closest mutations in sample 6005-1 with 26 bases between mutations, and the most

distant in 6001-10 with 18.7 kilobases between mutations. Of the 9 telangiectasia with identified somatic mutations, the distance between mutations in 6002-2, 6003-1, and 6005-1 was small enough to allow for mutation-spanning reads using illumina chemistry (Table. 2). Mutation-spanning reads were generated for 6001-3, 6001-7, 6001-8, and 6002-1 using PacBio chemistry. We were unable to generate amplicons spanning the mutations for 6001-1 and 6001-10. The sequence downstream of the somatic mutation in these telangiectasia contains several repetitive regions which, combined with the genomic distance and limited quantity of input DNA may have contributed to PCR failure. One notable confounder in this analysis is the generation of chimeric reads resulting from template switching during PCR. The generation of chimeric reads is known to interfere with amplicon-based haplotype phasing by switching a variant from one strand to another, potentially generating new haplotypes not present in the original sample (Laver et al., 2016). In practice, chimeric reads randomize the arrangement of the somatic and germline mutations. The frequency of chimeric arrangements is highly dependent on the distance between mutations and the number of PCR cycles used to make the amplicons. To reduce the number of chimeric reads in our libraries we used no more than 30 cycles for amplification. A previous study reports a chimeric arrangement frequency of 6.5% for 29 cycles of amplification for mutations 9kbp apart (Laver et al., 2016). Chimeric reads may account for the very few discordant reads in some of our samples, as shown in Table 2.2. Nonetheless, these were so minor in comparison to the great majority of the reads that phase could be unequivocally determined.

#### *in vitro Splicing*

A 3.8kb fragment of *ACVRL1* genomic DNA spanning from 431 bases upstream of exon 3 to 215 bases downstream from exon 8 was amplified and this entire insert was ligated into the MCS of pSPL3, a splicing vector (Church et al., 1994). Clones

were sequenced to ensure that no PCR-induced errors were present in the exons and adjacent intronic regions of the insert. The specific mutation, c.625+4A>T, was introduced using site directed mutagenesis and again, clones were sequenced to verify that the only sequence difference was at the intended site. Plasmid DNA from empty vector, wild type (control) vector and mutation-containing vector were transfected into HEK293T cells using Lipofectamine 3000 (ThermoFisher Scientific), incubated for 24 hours, and then the RNA was extracted using TRIzol and Direct-zol RNA miniprep kit (Zymo Research).

#### *Reverse-Transcription PCR*

RNA extracted from peripheral blood leukocytes and from transfected cells was used as template for cDNA synthesis using the Maxima H Minus First Strand cDNA kit (ThermoFisher Scientific). An RT primer in *ACVRL1* exon 8 was used in the RNA from blood while a vector-specific RT primer was used for the RNA from transfected cells to ensure that only RNA from the transfected vectors was being used as template for cDNA synthesis. cDNA from peripheral blood leukocytes was PCR amplified using primers in exons 3 and 6 while cDNA from the transfected cells was PCR amplified using primers in exons 3 and 8. PCR reactions were run on 1% agarose gels, the bands excised and Sanger sequenced.

## 2.5 Contributions and Acknowledgements

This chapter is adapted from a study published in AJHG (Snellings et al., 2019) with the following authors: Daniel A. Snellings, Carol J. Gallione, Dewi S. Clark, Nicholas T. Vozoris, Marie E. Faughnan, and Douglas A. Marchuk. DAS performed all sequencing experiments. CJG performed in vitro splicing experiments. DSC, NTV, and MEF provided tissue samples. DAM and DAS designed experiments.

We thank all participants and their families who donated samples for this study.

We thank Dr. Nicolas Devos and the Duke Sequencing and Genomic Technologies Core for assistance designing experiments and sequencing. This study was supported by U.S. Department of Defense grant (W81XWH-17-1-0429) and a Fondation Leducq Transatlantic Network of Excellence Grant in Neurovascular Disease (17 CVD 03). MEF also received financial support from the Nelson Arthur Hyland Foundation.

# 3

## *PIK3CA* Mutations Fuel Cerebral Cavernous Malformations

---

This chapter is adapted from a study published in *Nature* (Ren & Snellings et al., 2021)

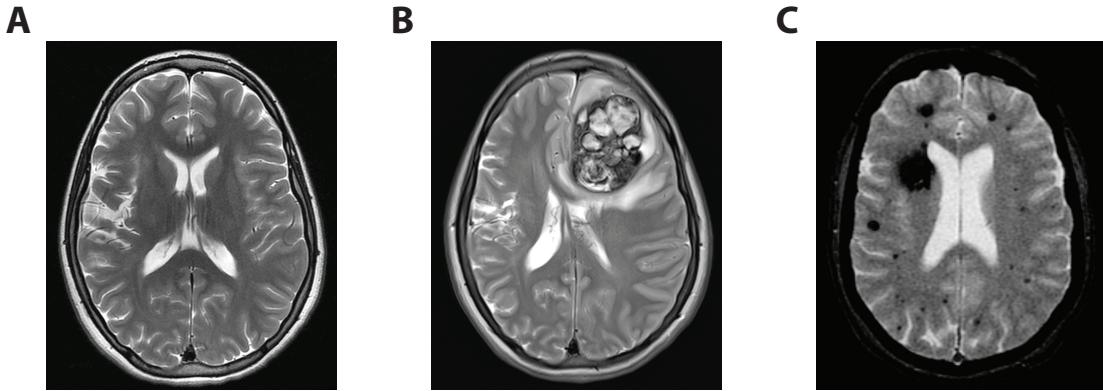
### 3.1 Premise

Vascular malformations such as cerebral cavernous malformations (CCMs) that arise in the central nervous system are an important cause of stroke and disability in younger individuals (Heiskanen, 1993; Fischer et al., 2013). Most CCMs arise sporadically as single lesions, but a minority present as part of a familial, autosomal dominant form of the disease that is associated with multiple lesions (Cavalcanti et al., 2012). Classic genetic studies have associated familial CCM disease with heterozygous germline loss of function mutations in three genes, *KRIT1*, *CCM2*, and *PDCD10*, that encode the components of a heterotrimeric protein complex (the “CCM complex”) (Fisher & Boggon, 2014; Plummer et al., 2005). Subsequent studies have demonstrated that CCM lesions harbor an additional somatic mutation in the same gene as the germline mutation, implicating biallelic loss of function of the affected CCM gene as the cause of the disease (Gault et al., 2005; Akers et al., 2009). Consistent with this monogenic loss of function mechanism, sporadic CCMs harbor biallelic somatic mutations in one of the CCM genes, resulting in homozygous loss of function (McDonald et al., 2014). Mouse models confirm that deletion of any of the CCM genes in the brain endothelial cells of neonatal mice confers CCM lesions, proving a causal role for loss of CCM complex function in this disease. These studies have generated the current genetic model of CCM pathogenesis in which biallelic loss of function mutations in a single CCM gene is sufficient for CCM lesion development.

Serial imaging studies to define the natural history of human CCMs has revealed that most are slow-growing and clinically silent (Akers et al., 2017; Al-Shahi Salman et al., 2012; Horne et al., 2016). In contrast, those that cause stroke and seizure are typically fast-growing (Figure 3.1A,B) and associated with repeated lesional hemorrhage (Awad & Polster, 2019; Porter et al., 1997). Such aggressive, symptomatic lesions are surgically resected if possible to prevent or treat associated neurologic

complications, but surgery is associated with high morbidity and cost and is impractical for patients with multiple lesions or lesions in less reachable locations such as the spinal cord. Furthermore, individuals with familial CCM often have a wide range of lesion size and severity; even lesions that are anatomically close (Figure 3.1C). Why a subset of CCM lesions exhibits rapid growth and intraindividual variability associated with clinical symptoms is unknown. Recent mouse and human studies suggest that a gut microbiome containing more invasive gram negative bacteria or an impairment of the gut barrier that blocks translocation of bacterial products such as lipopolysaccharide may modulate CCM growth through effects on TLR4-MEKK3-KLF2/4 signaling in brain endothelial cells (Tang et al., 2019, 2017; Polster et al., 2020). Plasma biomarkers of angiogenesis have also been correlated with lesional clinical activity (Girard et al., 2018; Lyne et al., 2019). However, individuals with familial CCM disease who harbor numerous silent lesions identified by MRI imaging also often manifest symptomatic hemorrhage and aggressive growth of a single lesion (Polster et al., 2019). Thus the current understanding of the environmental and genetic factors that contribute to CCM growth fails to explain important aspects of the disease natural history, especially the emergence of rapidly growing symptomatic lesions which account for the majority of clinically significant outcomes.

Recent studies of sporadic vascular malformations have identified acquired gain of function mutations in a number of central signaling pathways, including the RAS/MAPK/ERK pathway in congenital hemangiomas and capillary malformations and the PI3K/AKT/mTOR pathway in venous and lymphatic malformations (Ten Broek et al., 2019; Rodriguez-Laguna et al., 2019; Castillo et al., 2019; Wetzel-Strong et al., 2017; Luks et al., 2015; Limaye et al., 2015) (and reviewed in (Queisser et al., 2018)). Many of these gain of function mutations, e.g. those in *PIK3CA*, the catalytic subunit of PI3K, are identical to those that have been identified in cancer cells (Castillo et al., 2016; Castel et al., 2016; Limaye et al., 2015; Koren



**FIGURE 3.1: Significant Variability and Rapid Growth of CCMs.**

**A–B**, Axial T2 MRI of an individual at 18yo (A) and 20yo (B) showing rapid growth of a large CCM over the course of 28 months. Case courtesy of Brian Alan Perry, II (CC BY-NC-SA 3.0). **C**, Axial Gradient Echo MRI of an individual with familial CCM showing lesions of varying sizes. Case courtesy of Frank Gaillard (CC BY-NC-SA 3.0).

et al., 2015; Samuels et al., 2005). However, unlike cancer, in which mutations in multiple driver genes such as tumor suppressor genes and oncogenes combine to promote growth (Bailey et al., 2018; McGranahan et al., 2015), the pathogenesis of vascular malformations has been considered monogenic. The studies described below reveal that symptomatic CCM disease arises through a cancer-like paradigm in which the accumulation of multiple somatic mutations in the same cell results in both the loss of a vascular malformation suppressor gene (i.e. the CCM gene) and the gain of vascular malformation growth gene (i.e. *PIK3CA*). These studies reveal that clinically significant cavernous malformations arise through a compound genetic mechanism like that previously described for cancer, identify PI3K signaling as a major downstream effector pathway in CCM disease, and suggest that symptomatic CCMs may be effectively treated with the approved drug rapamycin (aka Sirolimus).

## 3.2 Results

### 3.2.1 *PIK3CA* Mutations Occur in Familial and Sporadic CCMs

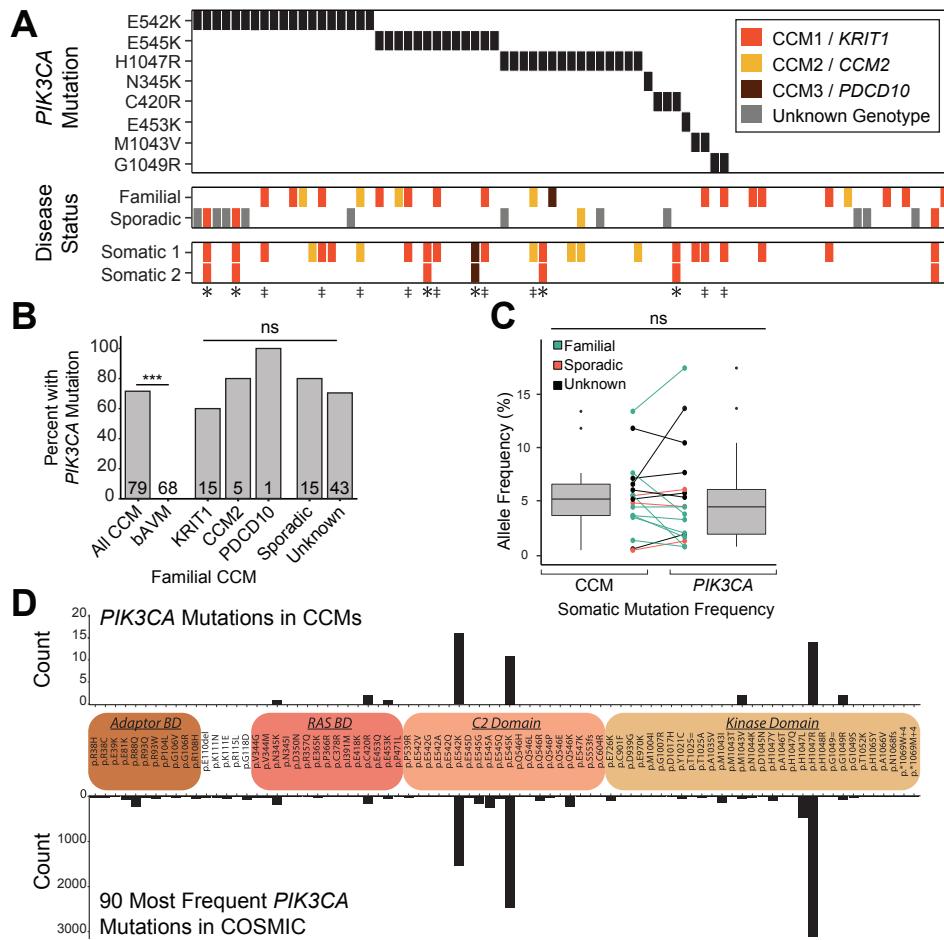
To determine whether human CCM lesions harbor gain of function mutations in *PIK3CA* or other genes that have been associated with increased cell growth and proliferation, 79 surgically resected CCM lesions (a single lesions per individual) were sequenced with a targeted panel of 66 genes, including the three causal CCM genes, genes involved in PI3K signaling and associated pathways, other oncogenic pathway genes, and other genes found to be mutated in vascular malformations (full gene list in methods). The collected CCM lesions were classified as “familial”, “sporadic” or “unknown” based on genetic and clinical evaluations (described in the methods). To ensure that any sequence variants identified in the CCM lesions were specific to CCM disease, 68 distinct surgically resected human brain arteriovenous malformations (bAVMs) were collected and sequenced. Like CCM lesions, bAVMs are neurovascular malformations enriched in vascular endothelial cells; thus the cellular composition of bAVMs is similar to that of CCM lesions. Since bAVMs and CCMs share a similar biological organization but arise due to distinct pathogenic mechanisms bAVMs provide a control with which to identify mutations in CCM lesions that are specific to CCM pathogenesis.

Variants called from the sequencing data were filtered to select for those with at least 5 supporting alternate reads, a variant allele frequency greater than 0.5%, predicted functional consequence, and several other filtering criteria. Remarkably, sequencing revealed that 56/79 (71%,  $P=1.23 \times 10^{-12}$ ) resected human CCM lesions harbor a somatic mutation in *PIK3CA* (Figure 3.2A). By contrast, none of the 68 bAVM samples harbored a somatic mutation in *PIK3CA*. The variant allele frequency of the *PIK3CA* mutations in CCM lesions ranged between 0.7% and 17.5% with a mean of 4.7%, suggesting mosaicism within the CCM lesion. All of the *PIK3CA*

mutations occurred at known hotspots in the catalogue of somatic mutations in cancer (COSMIC and Figure 3.2D). Significantly, analysis of 62 other genes with listings in the COSMIC database failed to identify mutations in any genes other than *PIK3CA* and the CCM genes. No mutations were found in other components of the PI3K pathway, including *PTEN* and *AKT1/2/3*, revealing strong specificity for *PIK3CA* mutations in CCM. The three most common *PIK3CA* mutations identified in CCM lesions (E542K, E545K, and H1047R) were validated with droplet digital PCR and SNaPshot (single nucleotide extension) assays. Mutations in *PIK3CA* were detected in 14/21 known familial lesions (9/15 *KRIT1*, 4/5 *CCM2*, 1/1 *PDCD10*), and 12/15 known sporadic lesions (Figure 3.2B). Each CCM lesion harbored no more than one somatic mutation in *PIK3CA*, and all of the *PIK3CA* mutations identified in CCM lesions have previously been determined to activate PI3K signaling (Dogrukul et al., 2015).

### 3.2.2 CCMs Harbor Multiple Somatic Mutations in Different Genes

Previous studies have demonstrated that human CCM lesions harbor somatic mutations in one of the three causal CCM genes. In this cohort, somatic mutations in CCM genes were identified in 24/79 (30%) of CCM lesions. The relatively low discovery rate of somatic CCM mutations may reflect types of mutations that are not detectable with short-read sequencing, such as large indels or chromosomal rearrangements. Notably, in the CCM lesions in which we positively identified a somatic loss of function CCM mutation, 21/24 (88%) also harbored a somatic gain of function *PIK3CA* mutation. This apparent enrichment in the co-detection of CCM and *PIK3CA* somatic mutations is consistent with poor sample quality that reduced sensitivity and/or low variant allele frequency in many lesions. Thus the true frequency of *PIK3CA* mutations in CCM lesions is likely to be higher than the 71% reported above. The identification of multiple mutations in CCM genes is consistent

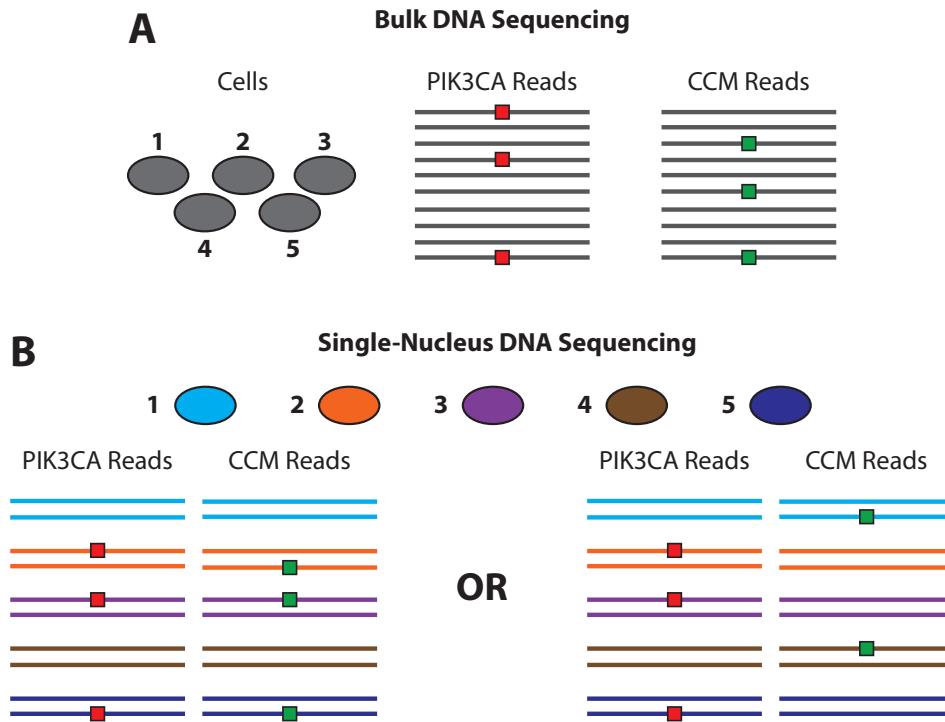


**FIGURE 3.2: Somatic Activating *PIK3CA* Mutations Detected in CCMs.** **A**, A schematic summary of the germline and somatic mutations in *KRIT1*, *CCM2* and *PDCD10*, and the somatic mutations in *PIK3CA*. Color denotes the affected CCM gene. Samples listed as neither familial nor sporadic are deidentified banked CCMs lacking either clinical information or genetic evidence supporting either classification. \* indicates familial CCMs with an activating mutation in *PIK3CA* and both germline and somatic mutations in a CCM gene. ‡ indicates known or presumed sporadic CCMs with an activating mutation in *PIK3CA* and two somatic mutations in a CCM gene. **B**, Distribution of activating mutations in *PIK3CA* present in sporadic CCMs, all three forms of familial CCMs, and control brain AVMs. **C**, The relationship between somatic *PIK3CA* and CCM mutations and *PIK3CA* activating mutations is graphed. Points indicate individual mutations in either a CCM gene or *PIK3CA*. Lines connect the CCM gene and *PIK3CA* gene mutations present in a single sample. **D**, The distributions of somatic *PIK3CA* mutations identified in human CCMs (top) and cancer (bottom) as reported in COSMIC. ns indicates P not significant; P>0.05. \*\*\* indicates P<1<sup>-16</sup>

with the previously described two-hit model of CCM pathogenesis with the addition of a third hit in *PIK3CA*. In 9 samples of familial CCM (‡ on Figure 3.2A), we detected distinct loss of function germline and somatic mutations in the same CCM gene in addition to a gain of function mutation in *PIK3CA* for a total of 3 genetic hits. In 6 samples of presumed sporadic CCM (\* on Figure 3.2A), we detected two distinct loss of function somatic mutations in the same CCM gene in addition to a gain of function mutation in *PIK3CA* for a total of 3 genetic hits. These data indicate that no fewer than three independent somatic mutation events contributed to the pathogenesis of those lesions. A comparison of variant allele frequency between CCM and *PIK3CA* somatic mutations in lesions where both mutations were found revealed no significant correlation ( $P > 0.15$ ) (Figure 3.2C). This balanced mutation frequency is consistent with a pathogenic mechanism in which both CCM and *PIK3CA* somatic mutations arise in a single cell during lesion formation, or one in which CCM mutant cells and *PIK3CA* mutant cells co-exist in similar numbers. Finally, the specific mutant *PIK3CA* alleles identified in resected human CCM lesions closely mirrored those identified in human cancers in the COSMIC database (Figure 3.2D), consistent with a shared molecular and cellular mechanism.

### 3.2.3 *PIK3CA* and CCM/MAP3K3 Mutations in the Same Cell

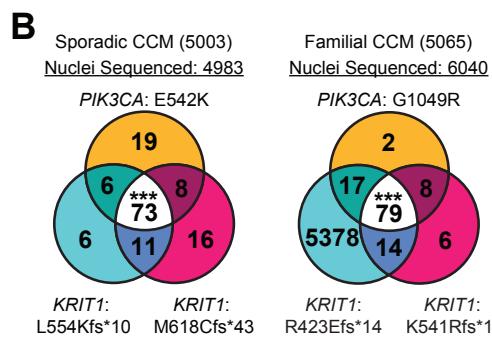
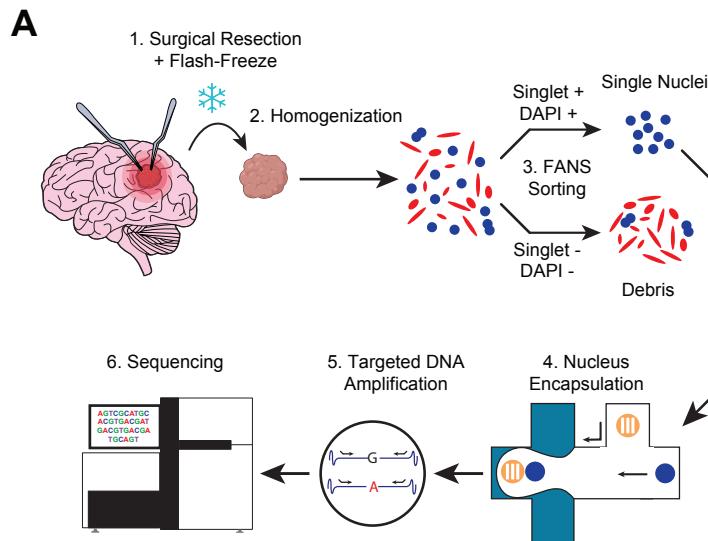
The identification of both *PIK3CA* and CCM gene somatic mutations in CCMs raises a critical question: Do these mutations occur in the same cell, or are these mutations in two distinct clonal populations that intermix to form a CCM. This question cannot be addressed with traditional bulk sequencing strategies, however if the cellular phase of mutations could be determined if we could assign each sequencing read to a cell of origin (Figure 3.3). We were able achieve this with single-nucleus DNA sequencing (snDNA-seq) on 3 sporadic and 2 familial CCMs using the Tapestri platform (Xu et al., 2019). Nuclei isolated from frozen tissue were



**FIGURE 3.3: Single-Nucleus DNA Sequencing to Resolve Cellular Phase.**

**A**, Schematic of reads retrieved from bulk sequencing. It is apparent that there is a somatic mutation in both reads from *PIK3CA* (red square) and reads from a CCM gene (green square), however it cannot be determined if mutant reads came from a the same cell or different cells. **B**, Single-nucleus DNA sequencing barcodes reads so we can determine the cell of origin, and therefore easily determine if the mutations are present in the same cell (left) or in different cells (right).

stained with DAPI and subjected to fluorescence-activated nucleus sorting to isolate single nuclei for input into the Tapestri instrument, where the nuclei were partitioned into droplets and the exons of *KRIT1*, *CCM2*, *PDCD10*, and *PIK3CA* were amplified (Figure 3.4A). Bulk sequencing of sample 5003 identified one somatic mutation in *PIK3CA* and two somatic mutations in *KRIT1*. These same somatic mutations were identified in the snDNA-seq data which show that the majority of somatic mutant nuclei harbored all three mutations. In sporadic CCMs 5038 and 5079 only one somatic CCM gene mutation was called in addition to the *PIK3CA* somatic mutation. In sample 5079 the second somatic CCM gene mutation is clearly present in total



**C**

Sample	Disease Type	Total Nuclei	PIK3CA / CCM Mutation 1 / CCM Mutation 2								p
			+/+/*	+/*/*	*/*/+	*/*/*	+/*/*	*/*/+	+/*/*	+/*/*	
5003	Sporadic	4983	4844	73	6	8	11	19	6	16	2.59E-16
5038	Sporadic	877	834	NA	25	NA	NA	13	5	NA	6.22E-19
5079	Sporadic	2182	1335	NA	391	NA	NA	312	144	NA	7.42E-39
5065	Familial	6040	536	79	17	8	14	2	5378	6	1.90E-18
5073	Familial	2383	165	181	50	20	156	5	1797	9	5.39E-29

**FIGURE 3.4: Single-Nucleus Sequencing of CCMs with Multiple Mutations.**

**A**, Schematic of workflow for processing frozen surgically-resected human CCM lesions for single-nucleus DNA sequencing. **B**, Representative data for sporadic and familial CCMs detailing the number of nuclei with each combination of *PIK3CA* and CCM mutations. \*\*\* indicates  $P < 1^{-16}$ . **C**, A summary of snDNA-seq results for 3 sporadic and 2 familial CCMs analyzed is shown. The number of nuclei with each possible genotype are listed. + indicates a wild-type allele; \* indicates a mutant allele. Note that only 1 somatic CCM mutation was identified in samples 5038 and 5079.  $P$  values were determined by  $\chi^2$ -squared test between the observed and expected triple mutant nuclei (or double mutant for lesions 5038 and 5079) determined by Poisson distribution (see Methods).

reads, however due to poor efficiency of the amplicon there were insufficient reads per nuclei to reliably establish cellular phase with the other two mutations. Data from both 5038 and 5079 show that the majority of mutant cells harbor both the somatic *PIK3CA* and CCM gene mutations. Likewise snDNA-seq data for familial CCMs 5065 and 5073 show the majority of somatic mutant nuclei (excluding nuclei with only the germline CCM gene mutation) harbor the *PIK3CA* mutation and both CCM gene mutations (Figure 3.4B). While the majority of somatic mutant nuclei in all 5 samples harbor all of the identified somatic mutations, notably there is a smaller number of nuclei observed with each possible combination of genotypes. This observation is highly unlikely to reflect genuine biology as the creation of all genotypes would require identical somatic mutations to occur in multiple clonal populations within the CCM. Some of the observed genotype combinations may represent intermediate clonal populations that formed prior to acquiring the full set of mutations, however the majority of these genotype combinations are likely due to allelic dropout (ADO)—a common technical artifact in single-nucleus/cell DNA sequencing data and has been noted by many previous studies (Xu et al., 2019; Szulwach et al., 2015; Satas & Raphael, 2018). To estimate the rate of ADO for each sample we identified heterozygous SNPs called in the snDNA-seq data and evaluated the ratio of heterozygous to homozygous nuclei and determined the rate of ADO to be  $8.4\% \pm 4.1\%$ . We also observed dropout of the constitutional CCM pathogenic allele as evidenced by the presence of WT nuclei in familial CCM samples 5065 and 5073. As a result of ADO, the number of nuclei with all somatic mutations is likely underestimated in each sample. Despite the confounding effects of ADO, all 5 samples clearly indicate that the *PIK3CA* and CCM gene somatic mutations occur in the same cell.

### 3.3 Discussion

These studies demonstrate powerful synergy between mutations that confer loss of CCM function and gain of *PIK3CA* function in both mouse models of CCM disease and a majority of resected human CCMs. These findings provide new insight into the mechanisms underlying the puzzling natural history of this vascular disease, explain why the disease has not been successfully modeled in mature mice with CCM loss of function alone, and reveal a compound genetic disease mechanism for vascular malformation that is highly analogous to that elucidated for human cancer. Translationally, they also provide strong evidence that approved drugs capable of inhibiting the downstream PI3K effector mTOR, and perhaps the PI3K pathway generally, may be used to block the growth and neurologic complications of clinically symptomatic CCM lesions.

#### 3.3.1 Three-Hit Model of CCM Pathogenesis

The most significant conceptual advance of this study is the discovery of a compound genetic mechanism of vascular malformation pathogenesis. To date, vascular malformations have been considered monogenic in origin, due either to bi-allelic loss of function mutations (e.g HHT (Snellings et al., 2019), CM-AVM (Lapinski et al., 2018), previously CCM (Akers et al., 2009; McDonald et al., 2014), or mono-allelic gain of function mutations (e.g. Sturge-Weber syndrome (Shirley et al., 2013), sporadic bAVM (Nikolaev et al., 2018), venous malformations (Limaye et al., 2015), lymphatic malformations (Luks et al., 2015), and blue rubber bleb nevus (Soblet et al., 2017; Tang et al., 2017; Goss et al., 2019; Davis et al., 2020; Francis et al., 2019; Couto et al., 2017, 2015)). Our studies identify a digenic, “triple-hit” mechanism involving the acquisition of as many as three distinct genetic mutations that culminate in loss of CCM gene function and gain of *PIK3CA* function as the

basis for rapidly growing, clinically symptomatic CCMs. Thus, aggressive CCM lesions arise via the acquisition of multiple somatic mutations that synergize in a manner similar to that of established cancer drivers. By analogy to cancer, the CCM genes may be considered vascular “suppressor genes”, required to constrain vessel growth, while *PIK3CA* may be considered a vascular “oncogene”, capable of driving excess vascular growth. As in cancer, the combined loss of a vascular suppressor and gain of a vascular activator is a potent combination that culminates in aggressive, symptomatic disease. Significantly, the present study demonstrates that either CCM loss of function or *PIK3CA* gain of function alone are sufficient to confer a modest vascular phenotype. These less aggressive vascular lesions are analogous to benign tumors that become malignant after mutation of another driver gene (Figure 3.5), a scenario consistent with clinical reports of sudden aggressive growth in small, pre-existing CCM lesions. Whether distinct synergistic driver gene combinations underlie other vascular malformations is an important future question that may be addressed by deeper genomic sequencing of human vascular malformations.

### 3.3.2 Role of Clonal Expansion in Mutagenesis

A key question raised by our studies is how CCM loss of function and *PIK3CA* gain of function interact at the cellular and molecular levels during lesion genesis. Insights gleaned from both genetic analysis of human CCM lesions and mouse genetic models are highly complementary and support an endothelial cell autonomous mechanism in which the two pathways are both directly and indirectly linked. Single cell genomic DNA sequencing of both sporadic and familial human CCM lesions reveals that a majority of mutant cells harbor acquired mutations in both a CCM gene and *PIK3CA*, strong evidence of interaction at the single cell level and that the majority of clonal expansion occurs after acquisition of all three mutations. Aileen Ren showed that sporadic lesions arise in *Slco1c1*(BAC)-CreERT2;*Krit1*<sup>f/f</sup>;R26-LSL-

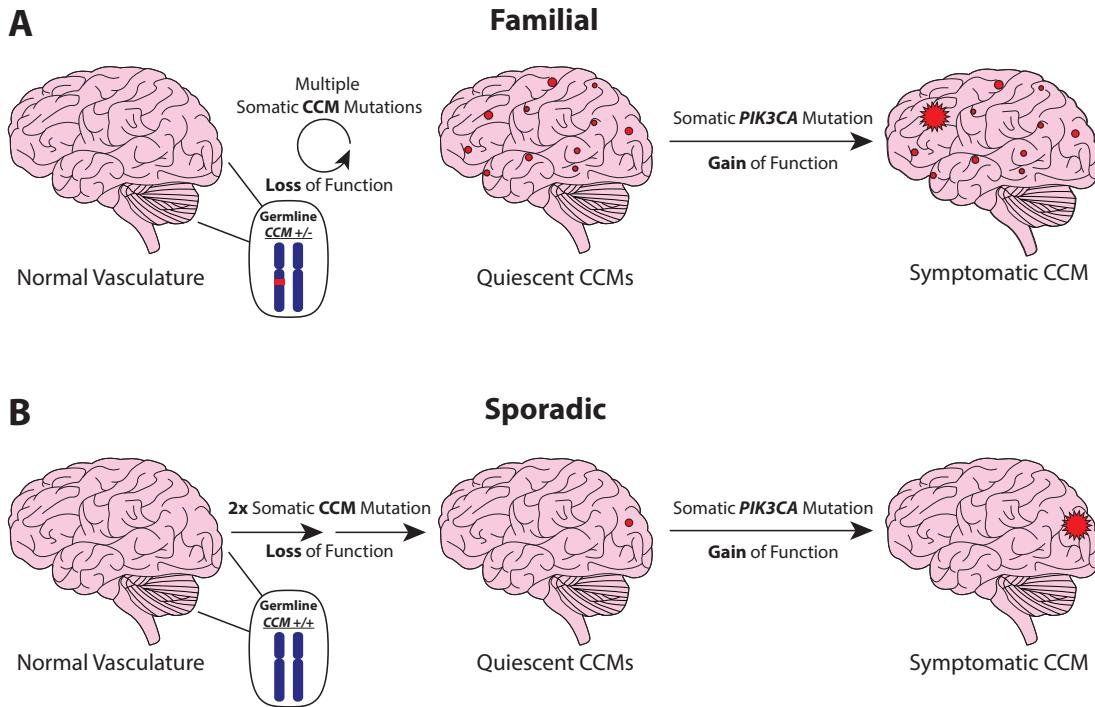


FIGURE 3.5: Three-Hit Model of CCM Pathogenesis.

Predicted progression of CCM development based on the finding of *PIK3CA* and CCM gene mutations in the same lesion. **A**, Model of familial CCM pathogenesis where single somatic mutations in the CCM gene are relatively common and lead to the formation of multiple lesions. A subset of those initial lesion may subsequently acquire a somatic mutation in *PIK3CA* that fuels lesion growth. **B**, Sporadic CCMs likely develop via two somatic mutations in a CCM gene resulting in an immature CCM that may or may not subsequently acquire a mutation in *PIK3CA*. Notably the single-nucleus DNA sequencing data presented in this chapter cannot resolve the temporal organization of mutations, however based on these data this is perhaps the most parsimonious model.

*Pik3ca<sup>H1047R</sup>* animals due to Cre leak that is exclusively endothelial, identifying the endothelial cell as the target cell type and, like the human genetic sequencing data, suggestive of a clonal mechanism in which emergence of a single compound mutant endothelial cell is sufficient for lesion formation.

### 3.3.3 Therapeutic Implications

Our identification of *PIK3CA* activation in CCM immediately suggests that PI3K inhibitors may be an effective therapeutic for CCM. *PIK3CA* mutations are extremely common in cancers. As a result, there is decades of extensive research on the mechanism of *PIK3CA* activation and many efforts to develop cancer therapies targeting this signaling axis. Among these therapeutics is rapamycin—an off-patent small-molecule drug targeting mTOR (a downstream effector of *PIK3CA*)—which may be made cheaply available to CCM patients. To test the efficacy of rapamycin for preventing CCM, Aileen Ren developed a mouse model with both biallelic CCM loss of function, and *Pik3ca* gain of function. These mice develop aggressive lesions soon after induction. Aileen showed that rapamycin is extremely effective in preventing lesion formation in these mice (Ren et al., 2021). This preliminary data suggests that rapamycin may be a good candidate for pre-clinical trials, though the true test will be whether rapamycin is able to regress existing lesions. Lesion prevention is of course a step forward but given the non-trivial side effects of rapamycin, it is unlikely that long-term treatment will be a viable option. In contrast, if rapamycin is able to regress lesions, then a treatment program may consist of brief intermittent periods of dosing which would more amenable to the patient population.

### 3.3.4 DVA Predispose to CCM and Other PI3K-Related Diseases

A clinical clue to the pathogenesis of human CCM disease that may be explained by our findings is the observation that sporadic CCMs frequently arise at sites of pre-existing developmental venous anomalies (DVAs) (24–32% assessed by MRI, and up to 100% in one study that assessed this relationship at the time of surgical resection (Wurm et al., 2005; Porter et al., 1999; Abdulrauf et al., 1999)). Since DVAs are benign and do not undergo surgical resection, there are no data addressing their genetic basis. However, DVA is found in a majority of individuals with Cowden’s

syndrome, an inherited disease caused by germline heterozygous loss of function mutations in *PTEN* that is also the result of PI3K gain of function (Dhamija et al., 2018; Tan et al., 2007). A parsimonious explanation for these clinical observations is that endothelial cells with pre-existing *PIK3CA* mutations sufficient to confer DVA subsequently acquire CCM loss of function mutations and transform into more aggressive sporadic CCM lesions. While such a clinical pathogenesis remains entirely speculative until DVA genetic sequencing is performed, it would support a mechanism highly analogous to cancer in which accrued mutations convert a benign vascular abnormality into a more malignant one.

### 3.4 Methods

#### *CCM Collection*

Human CCM tissue specimens were obtained from surgically resected specimens from three sources including the Barrow Neurological Institute, Angioma Alliance biobank, and University of Chicago. This study was approved by each institutions respective Institutional Review Board.

#### *Brain AVM Collection*

Brain AVM tissue specimens were obtained from the nodal tissue of surgically resected brain AVMs (M.T.L). For non-vascular lesion controls (NVLCs), temporal lobe specimens were similarly acquired from subjects undergoing anterior temporal lobectomies for medically refractory epilepsy. All tissues were frozen at -80°C and stored in a biobank. This study was approved by the University of California San Francisco Institutional Review Board and performed in compliance with the Health Insurance Portability and Accountability Act regulations.

### *DNA Extraction*

DNA from human CCM samples was extracted using the DNeasy Blood and Tissue Kit (Qiagen). Extractions were done as per the manufacturers' directions excepting cases where less than 25mg of tissue was available. For samples <25mg the manufacturer recommended volumes were either halved or quartered—depending on the amount of tissue available—to optimize final DNA concentration and yield. Final DNA concentrations were quantified using Qubit dsDNA BR assay kit (Invitrogen cat. Q32850) according to manufacturer recommended protocol.

### *Droplet Digital PCR*

Identification of somatic *PIK3CA* mutations was performed using droplet digital PCR (ddPCR) according to manufacturer protocol for mutation detection assays (BioRad 10047489 Ver B). For each sample with sufficient DNA yield, 30–100ng of DNA was incorporated into droplets using the QX200 AutoDG system (BioRad). After PCR, fluorescence from the resulting droplets was read using the QX200 droplet reader (BioRad). These steps were repeated for three assays testing the presence of the three most common *PIK3CA* mutations: E542K, E545K, and H1047 (ThermoFisher assay IDs: Hs000000085\_rm, Hs000000086\_rm, Hs000000088\_rm, respectively). Each assay included a no-template control, a wild-type control, and a mutation-positive control for the mutation being assayed. The mutation-positive controls were DNA extracted from cell lines with known heterozygous mutation of E542K (T84), E545K (HCT15), or H1047R (HCT116). The output of the droplet reader was analyzed using the QuantaSoft software (BioRad). The gates for positive and negative mutation status were drawn with respect to the distribution of droplets in the mutation-positive controls and applied to all samples.

### *SNaPshot*

Human CCM samples with an E542K, E545K, or H1047R *PIK3CA* mutation identified by sequencing or ddPCR underwent tertiary confirmation of mutation status using SNaPshot (Applied Biosystems), a single-base extension sanger sequencing assay. An initial round of PCR amplified exons 9 and 20. In a second round of PCR, primers directly adjacent to the assayed nucleotide was extended with ddNTPs and sequenced on a 3130 Genetic Analyzer (Applied Biosystems). Sequences were examined using GeneMapper software (Applied Biosystems). The allele frequency of the mutation by dividing the area under the peak of the mutant allele by the total area under both allele peaks. The primers used in this analysis were synthesized according to designs in a previously published assay for *PIK3CA* mutations (Hurst et al., 2009).

### *Sequencing*

Previous studies identifying somatic mutations in CCMs and bAVMs have reported alternate allele frequencies less than 1%. To enable the detection of variants at such low frequencies we aimed to sequence samples to an average of 1000x (actual mean coverage 1102x) coverage in addition to leveraging 10bp unique molecular identifiers (UMIs) to mitigate the impact of PCR duplication. These conditions allow us to theoretically detect variants as low as 0.5% allele frequency.

Sequencing libraries for the human CCMs and bAVMs were prepared using the SureSelect XT HS target enrichment workflow (Agilent). The targeting panel used for sequencing the CCMs covers the following genes: *KRIT1*, *CCM2*, *PDCD10*, *PIK3CA*, *PTEN*, *AKT1*, *KRAS*, *RAF*, *NRAS*, *MAP2K1*, *RASA1*, *TEK*, *GNAQ*, *GNA11*, *MAP2K2*, *PPP2R5D*, *ACVRL1*, *ENG*, *SMAD4*, *AKT2*, *AKT3*, *CCBE1*, *CDKN1C*, *FLT1*, *FLT4*, *FOXC2*, *GATA2*, *GDF2*, *GJC2*, *GLMN*, *KIF11*, *MTOR*, *PIK3R2*, *PTPN14*, *SOX18*, *STAMBP*, *VEGFC*, *MAP2K4*, *MAP3K1*, *MAPK1*, *JAK1*,

*JAK2, JAK3, KDR, NOTCH1, PDGFRA, PDGFRB, RET, HRAS, TP53, MSH2, MYB, MYCN, MYC, ERBB2, EGFR, NTRK2, ODC1, SLC25A21, PTTG1, TSC1, TSC2, EPHB2, TGFBR1, TGFBR2, TGFBR3.*

The bAVM samples were sequenced using a customized Agilent Comprehensive Cancer panel which covers 175 genes including *PIK3CA*. After library preparation, CCM samples were pooled and sequenced across 1 lane of a HiSeq4000 (illumina) with paired-end 150bp reads. bAVM samples were pooled and sequenced across a NovaSeq6000 SP flow cell (illumina) with paired-end 150bp reads.

### *Sequence Analysis*

Sequencing data was processed according to the GATK (Broad Institute) best practices for somatic short variant discovery with slight modifications for “tumor-only” sequencing data. As a secondary method for variant discovery we developed custom software designed specifically for the detection of somatic mutations in sequencing data from samples with no available normal tissue. This software was implemented as part of Gonomics, an ongoing effort to develop an open-source genomics platform in the Go programming language. Gonomics can be accessed at [github.com/vertgenlab/gonomics](https://github.com/vertgenlab/gonomics).

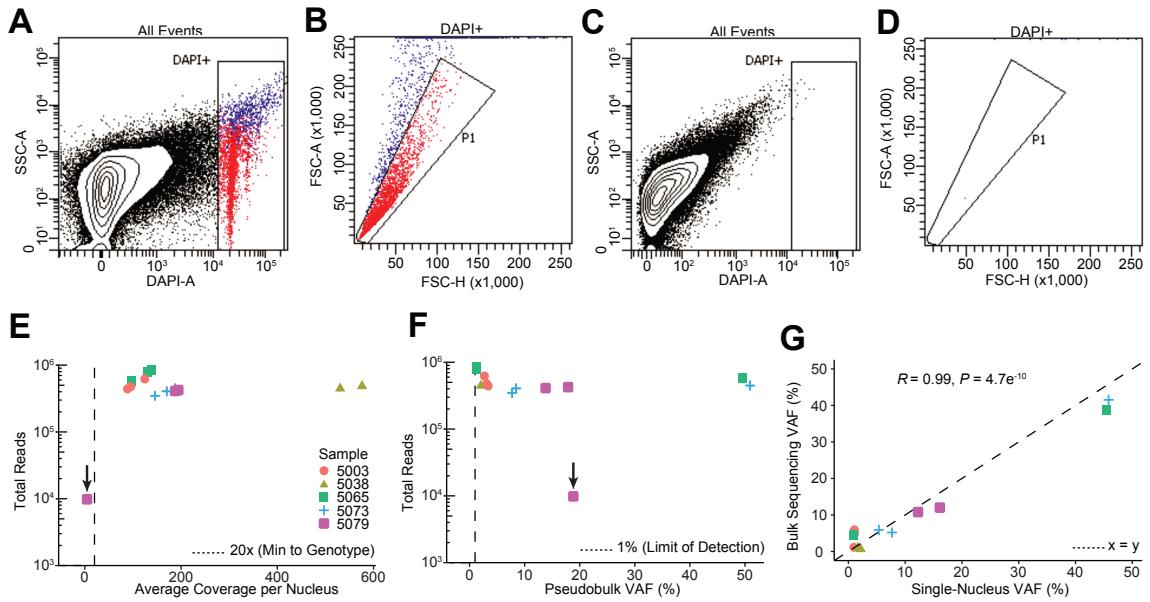
After variant calling the resulting variants were functionally using snpEff. In this process each variant was annotated with: the protein-level consequence of coding mutations; the predicted impact of missense mutations according to SIFT, PolyPhen2, and PROVEAN; membership in several SNP databases including dbSNP, 1000 Genomes project, and ExAC; and membership in the Catalogue of Somatic Mutations in Cancer (COSMIC). We filtered the resulting list of variants using the following inclusion criteria: >100x total coverage; >5 supporting reads; <90% strand specificity; >0.5% alternate allele frequency; <1% population allele frequency according to the above mentioned SNP databases; and membership in the COSMIC

database.

#### *Single-Nucleus DNA Sequencing*

Frozen human CCM lesion tissue obtained from medically-indicated, surgical resection were prepared for single-nucleus DNA sequencing (snDNA-seq) following a nuclei isolation protocol by Martelotto L. ([dx.doi.org/10.17504/protocols.io.3fkgjkw](https://dx.doi.org/10.17504/protocols.io.3fkgjkw)). All steps prior to loading on the Tapestri platform were performed in <3 hours. Nuclei were maintained at 4°C throughout the protocol. Frozen tissue was homogenized by Dounce in Nuclei EZ Lysis Buffer (Sigma-Aldrich), briefly washed, filtered through a 70 $\mu$ m mesh, stained with DAPI, and filtered through a 35 $\mu$ m mesh. The CCM homogenate was sorted using a FACS AriaII (BD) (70 $\mu$ m nozzle, 70psi, 4-Way Purity, chiller) gating to retain singlet DAPI-positive events (Figure 3.6). Up to 400,000 sorted nuclei were collected in 1ml of the following buffer prepared with ultrapure nuclease-free water: Na<sub>2</sub>SO<sub>4</sub> 82mM, K<sub>2</sub>SO<sub>4</sub> 30mM, glucose 10mM, HEPES 10mM, MgCl<sub>2</sub> · 6H<sub>2</sub>O 5mM, BSA 2%. Sorted nuclei were pelleted by centrifugation at 4°C (500rcf, 10min), supernatant discarded, and resuspended in 36 $\mu$ L of MissionBio Cell Buffer. The concentration of nuclei was determined by counting DAPI-positive nuclei with a hemocytometer on an EVOS FL (fluorescence) microscope (Thermo Fisher) while confirming that nuclei aggregates comprised <5% of total nuclei. Samples with <5% aggregate nuclei and a concentration within 2000–4000 nuclei/ $\mu$ L (diluting with additional MissionBio Cell Buffer where necessary) were used for snDNA-seq.

Library preparation was performed using the Tapestri platform (MissionBio) according to the manufacturers protocol (PN3354). Libraries were generated with a custom amplicon panel synthesized by MissionBio covering all exons of *KRIT1*, *CCM2*, *PDCD10*, and 7 amplicons covering somatic mutation hotspots in *PIK3CA*, per the COSMIC database. Up to three libraries were pooled and sequenced with a NextSeq Mid-Output 2 x 150bp kit (illumina). Data processing and QC was



**FIGURE 3.6: Correspondence of Bulk and Single-Nucleus Sequencing.**

**A–D**, Representative FANS plots of DAPI stained (A–B) and unstained (C–D) CCM homogenate samples. Doublet discrimination by forward scatter profile for DAPI stained and unstained samples are shown in B and D respectively. **E**, Total reads and average coverage per nucleus from snDNAseq for each mutation detected by bulk sequencing. Dotted line shows 20x coverage, the minimum cutoff used for establishing genotype. **F**, Pseudobulk allele frequency from snDNA-seq for each mutation detected by bulk sequencing. Dotted line shows 1% allele frequency. Note the data point with \* in E–F shows a mutation in sample 5079 detected in bulk sequencing which, due to poor amplification during snDNA-seq, received insufficient coverage per nucleus (4.5x) to establish nuclear genotypes however is clearly present in pseudobulk reads (1849/9814). **G**, Comparison of mutation allele frequency as detected by bulk and snDNA-seq. As nuclei are diploid for the relevant autosomes, the x-axis is equal to the fraction of mutant nuclei divided by two. Dotted line shows perfect correlation at  $x=y$ .  $R$  and  $P$  were calculated by Pearson's correlation coefficient.

performed by the MissionBio cloud-based analysis pipeline. Data quality for each nuclei barcode was determined using MissionBio recommended filtering settings. Data from low quality nuclei barcodes were removed prior to mutation analysis.

To determine the cellular phase of somatic mutations detected in bulk sequencing, nuclei barcodes were selected that had a minimum of 20x coverage across all mutant

regions to ensure that all nuclei included in the analysis have appropriate sensitivity to detect a mutation. For each nuclei barcode and each mutant position, reads containing the ref and alt alleles were counted. Mutant regions that had both a minimum of 10 alt reads and 10% allele frequency were marked as mutation positive. The number of nuclei barcodes with each possible genotype are recorded in Figure 3.4C.

The p value for each sample was determined by Chi-squared test comparing the observed number of triple mutant nuclei barcodes (or double mutant if only two mutations were identified) to the expected number of triple mutant nuclei barcodes if the null hypothesis is true. In this test the null hypothesis is that the observed somatic mutations do not occur in the same cell and instead exist in two clonal populations where the proportion of each population =  $2 \cdot \text{variant allele frequency}$  (assuming cells are diploid and heterozygous for the mutation). In samples where three somatic mutations were identified an alternative null hypothesis may be to use three clonal populations, however assuming that the three somatic mutations are partitioned into two clonal populations as asserted above is a more conservative test and therefore used here. To determine the expected number of triple mutant nuclei barcodes we first determined the expected number of droplets with two (or more) nuclei by approximating a Poisson distribution. The expected number of triple mutant nuclei barcodes is the product of the number of droplets with two nuclei, the proportion of mutant clone 1, and the proportion of mutant clone 2.

Here we report the Poisson estimation p value as this is the accepted method for estimating the rates of false-positive droplets when dealing with data generated using microfluidics. We can also consider a more intuitive upper bound for the p value by considering the extreme case where each droplet contains two nuclei. In this case the expected number of triple mutant nuclei barcodes is the product of the total number of nuclei barcodes, the proportion of mutant clone 1, and the proportion of

mutant clone 2. Even in this extreme case the highest p value among our data is  $<1 \times 10^{-17}$ .

### 3.5 Contributions and Acknowledgements

This chapter is adapted from a study published in *Nature* (Ren & Snellings et al., 2021) with the following authors: Aileen A. Ren, Daniel A. Snellings, Sophie Y. Su, Courtney C. Hong, Marco Castro, Alan T. Tang, Matthew R. Detter, Nicholas Hobson, Romuald Girard, Sharbel Romanos, Rhonda Lightle, Thomas Moore, Robert Shenkar, Christian Benavides, M. Makenzie Beaman, Helge Mueller-Fielitz, Mei Chen, Patricia Mericko, Jisheng Yang, Derek C. Sung, Michael T. Lawton, Michael Ruppert, Markus Schwaninger, Jakob Körbelin, Michael Potente, Issam A. Awad, Douglas A. Marchuk and Mark L. Kahn. AAR and DAS contributed equally to this study. AAR designed and performed most of the mouse and tissue culture experiments and wrote the manuscript. DAS performed the genetic studies of human CCM lesions and wrote the manuscript. SYS created and performed the adult cranial window assays in mice. CCH, ATT and MRD contributed to mouse genetic studies. MC performed in vitro studies. NH, RG, SR, RL, TM, RS and IAA performed microCT CCM lesion imaging and quantification in a blinded manner. MC and PM assisted with mouse genetic studies. JY and DCS performed histologic studies. MTL provided surgically excised human CCM samples. MS and JK provided critical reagents. MP, IAA, DAM and MK designed experiments and wrote the manuscript.

We thank the members of the Kahn lab for their thoughtful comments and advice during this work. We thank Angioma Alliance for patient enrollment, and the University of Chicago PaleoCT core facility for its expertise in imaging and image quantitation. We thank the Penn CDB Microscopy Core for support with microscopy. Flow cytometry was performed in the Duke Human Vaccine Institute Research Flow Cytometry Shared Resource Facility. We thank Duke University School of

Medicine for use of the Sequencing and Genomic Technologies Shared Resource for library preparation and sequencing. These studies were supported by National Institute of Health grants R01HL094326 and R01NS100949 (MK), P01NS092521 (MK, DM, IA) the Leducq Foundation (MK, MP), the AHA-Allen foundation (MK), T32 HL007150 (AR), F31HL152738 (DS), F31NS115256 (CH), F30NS100252 (AT), European Research Council (ERC) Synergy Grant-2019-WATCH-810331 (MS) and ERC Consolidator Grant EMERGE-773047 (MP).

# 4

## Developmental Venous Anomalies Predispose to Malformation

---

This chapter is adapted from a study in press at *Nat. Cardiovasc. Res.* (Snellings et al., 2022)

## 4.1 Premise

Cerebral cavernous malformations (CCMs) are hemorrhagic neurovascular malformations that may lead to stroke, seizures and other clinical sequelae. CCM disease may be inherited by an autosomal dominant loss of function (LOF) mutation in the genes encoding components of the CCM signaling complex: *KRIT1* (Labergele Couteulx et al., 1999b; Sahoo et al., 1999), *CCM2* (Liquori et al., 2003), or *PDCD10* (Bergametti et al., 2005). Solitary sporadic CCM lesions may also occur in the absence of inherited germline mutations in CCM complex genes. Previous studies have established that CCM pathogenesis follows a genetic two-hit model where somatic mutations cause biallelic LOF in *KRIT1*, *CCM2*, or *PDCD10* to initiate lesion formation (Akers et al., 2009; Gault et al., 2009, 2005; McDonald et al., 2014). Recent studies have found that somatic mutations in *PIK3CA* and *MAP3K3* also contribute to CCM pathogenesis (Hong et al., 2021; Ren et al., 2021; Weng et al., 2021). These discoveries opened new avenues for research and therapeutic development, but they also raised new questions about the roles of somatic mutations in familial versus sporadic lesions.

The presence of multiple somatic mutations in CCMs also helps explain a long-standing mystery: the association between sporadic CCM lesions and developmental venous anomalies (DVA). DVA are the most common vascular malformation present in 6–14% of the adult population (Brinjikji et al., 2017; Gokce et al., 2014; Jones et al., 2015; Linscott et al., 2016) with the majority developing prior to the age of 20 (Brinjikji et al., 2017). When assessed by magnetic resonance imaging, an adjacent DVA is identified in 24–32% of sporadic CCM cases (Abdulrauf et al., 1999; Porter et al., 1997; Wurm et al., 2005), and an even greater fraction of sporadic CCMs are found to be associated with a DVA at surgery (Abdulrauf et al., 1999; Porter et al., 1997). One study focused on DVA reported an adjacent sporadic CCM in 6.9% of

all DVAs in a general population (116 of 1689) (Brinjikji et al., 2017). These studies highlight the association between DVA and sporadic CCM. By contrast, familial CCM lesions have not been associated with DVA (Dammann et al., 2017). These combined data suggest that a DVA is not required for CCM formation but may be a predisposing factor in sporadic cases. We hypothesize that DVA are caused by somatic mutations, and that the genetic origin of DVA overlaps that of CCM. In this study we probe the relationship between mutations in *KRIT1*, *CCM2*, *PDCD10*, *PIK3CA*, and *MAP3K3* and explore a mechanism by which DVA act as a genetic primer for the genesis of sporadic CCM lesions.

## 4.2 Results

### 4.2.1 Mutations in *MAP3K3* and CCM Genes are Mutually Exclusive

To evaluate whether sporadic and familial CCMs have distinct somatic mutation spectra we identified somatic mutations present in 71 CCMs (20 familial CCMs and 51 sporadic/presumed sporadic CCMs). Mutations in *KRIT1*, *CCM2*, *PDCD10*, and *PIK3CA* were detected by targeted sequencing and/or droplet digital PCR (ddPCR) as previously described (Ren et al., 2021). The common gain of function mutation in *MAP3K3* (hg38 chr17:63691212, NM\_002401.3, c.1323C>G; NP\_002392, p.I441M) was detected by ddPCR using a previously published probe set (Couto et al., 2015).

The p.I441M mutation in *MAP3K3* was identified in 15/51 sporadic CCMs and 0/20 familial CCMs (Figure 4.1A). We also screened for *MAP3K3* p.I441M in 8 blood samples for which we were previously unable to identify a germline mutation in *KRIT1*, *CCM2*, *PDCD10*. None of the 8 blood samples harbored *MAP3K3* p.I441M. Notably 11/51 sporadic CCMs harbored at least 1 somatic mutation in *KRIT1*, *CCM2*, or *PDCD10*, however none of these CCMs also had a mutation in *MAP3K3* indicating that a mutual loss of the CCM complex and gain of function in MEKK3 (the protein product of *MAP3K3*) are not both required for CCM formation. As

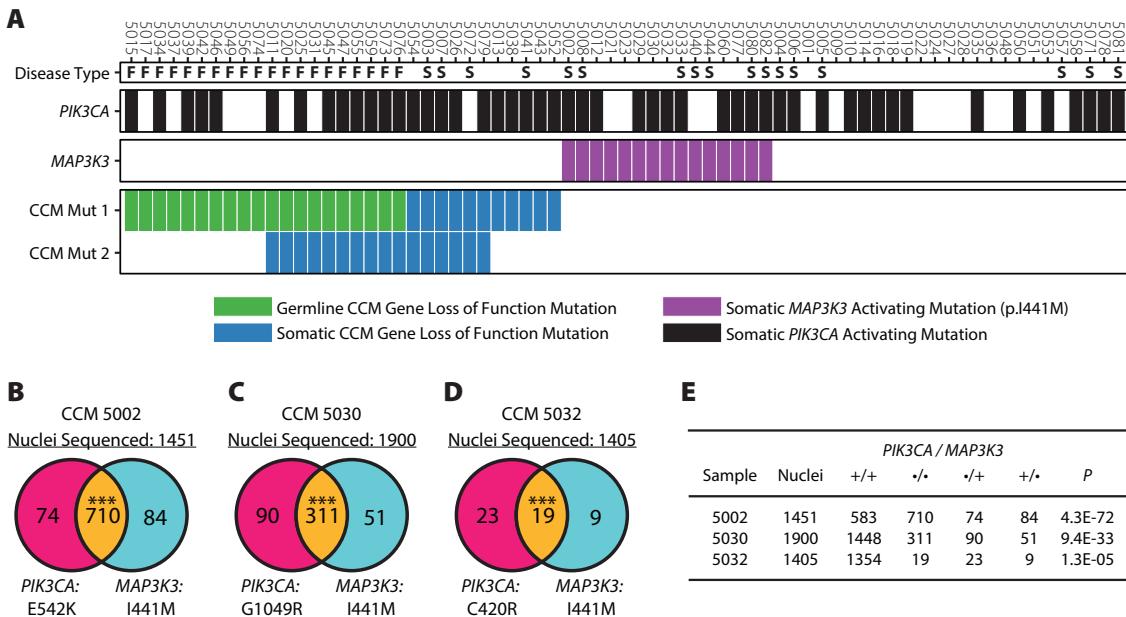


FIGURE 4.1: Mutations in *MAP3K3* are Mutually Exclusive with CCM Mutations and Occur in the Same Cells as *PIK3CA* Mutations.

**A.** Mutations present in 71 CCM samples. Disease type denotes whether the sample was familial (F), sporadic (S), or unknown (blank). The presence of somatic mutations in *PIK3CA* and *MAP3K3* are denoted by black and purple bars respectively. Germline and somatic mutations (green and blue respectively) in *KRIT1*, *CCM2*, or *PDCD10*, are shown in CCM Mut 1 with the second-hit mutation shown in CCM Mut 2 if present. **B–D.** Nuclei genotypes determined by snDNA-seq. The left and right circles in each Venn diagram shows the number of nuclei with the *PIK3CA* or *MAP3K3* mutations where the overlap shows nuclei harboring both mutations. \*\*\* P < 0.0001. **E.** Summary of data presented in B–D including P values determined by comparing the observed number of double mutant nuclei to the expected value derived from a Poisson distribution as done previously (Ren et al., 2021).

the CCM complex is known to be a direct inhibitor of MEKK3 activity (Zhou et al., 2015, 2016), these data strongly suggest identical functional consequences of these mutations.

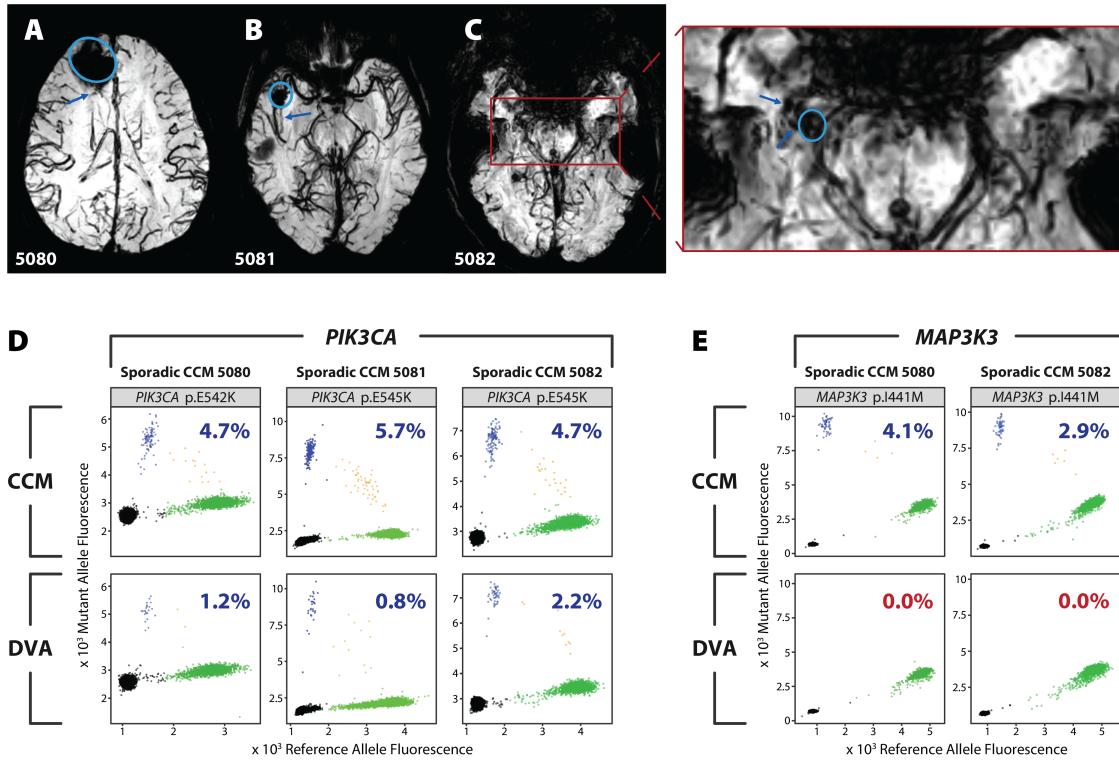
The majority of CCM and verrucous venous malformations with a mutation in *MAP3K3* harbor the p.I441M variant (Couto et al., 2015; Hong et al., 2021; Weng et al., 2021), however an alternative variant p.Y544H has also been identified in a

venous malformation (Al-Qattan et al., 2020). While ddPCR provides superior sensitivity and specificity compared to targeted sequencing, it is restricted to detecting a single mutation per assay. To determine whether other mutations that contribute to CCM pathogenesis—either *MAP3K3* mutations besides p.I441M, or mutations in yet undiscovered genes—we performed whole-exome sequencing (mean depth 133x) on 8 sporadic CCMs for which no somatic mutations in *KRIT1*, *CCM2*, *PDCD10*, or *MAP3K3* were found. No additional mutations in *MAP3K3* were identified and no candidate variants in other genes passed QC filters (see Methods).

#### 4.2.2 Cellular Phase of Somatic *MAP3K3* and *PIK3CA* Mutations

While somatic mutations in *KRIT1*, *CCM2*, *PDCD10*, and *MAP3K3* are mutually exclusive, somatic gain of function mutations in *PIK3CA* may co-occur with any other mutation (Figure 4.1A). We have previously shown that co-occurring mutations in *KRIT1/CCM2* and *PIK3CA* occur in the same clonal population of cells (Ren et al., 2021). To determine whether *MAP3K3* and *PIK3CA* mutations co-exist in the same cells we performed single-nucleus DNA-sequencing (snDNA-seq) on frozen tissue from three surgically resected CCMs determined to harbor both mutations (Figure 4.1B–D).

In CCMs 5002 and 5030, the vast majority of mutant nuclei harbor both mutations in *MAP3K3* and in *PIK3CA* indicating that these mutations co-exist in the same cells. In CCM 5032, 37% (19/51) of mutant nuclei harbor both mutations. While this is a far lower fraction compared to other samples, it is significantly higher than may be expected by chance when sampling from 1405 total nuclei ( $P = 1.3 \times 10^{-5}$ , Figure 4.1E). In bulk genetic analysis, the allele frequencies of *PIK3CA* and *MAP3K3* mutations detected in CCM 5002 were 19% and 13% respectively. In snDNA-seq the allele frequencies of these mutations increased to 54% and 55% respectively. This difference likely reflects the mosaic nature of CCMs. As snDNA-



**FIGURE 4.2: CCM and Adjacent DVA Have Identical *PIK3CA* Mutations.**

**A–C.** Axial magnetic resonance (MR) susceptibility weighted images acquired at 3 Tesla showing CCM (circle) and associated DVA branches sampled during surgery (arrow) in individuals with CCM 5080 (A) or 5081 (B) or 5082(C). The inset red box in C shows the region expanded to the right with the CCM and DVA marked. **D–E.** Somatic mutations in *PIK3CA* (D) and *MAP3K3* (E) in CCM (top panels) and the associated DVA (bottom panels) from samples 5081, 5082, and 5083. Mutations were detected by droplet digital PCR (ddPCR) and shown as the fluorescence of the reference probe on the x-axis, and the mutant probe on the y-axis. Droplets containing the reference allele, mutant allele, both, or neither, are colored in green, blue, orange, and black respectively. Percentage inset into each graph shows the variant allele frequency for the displayed mutation. If the mutation was determined to be present, the percentage is blue, else the percentage is red.

seq requires nuclei harvested from frozen tissue, we must sample a new area of the frozen lesion than was sampled for bulk sequencing. Sampling from different sites of the same lesion often results in minor changes in allele frequency, however the drastic change in allele frequency we find in CCM 5002 suggests either that our initial sample of the lesion for bulk sequencing contained largely non-lesion tissue, or an uneven

distribution of mutant cells in the lesion.

#### 4.2.3 CCMs and DVA Harbor a Shared Mutation in *PIK3CA*

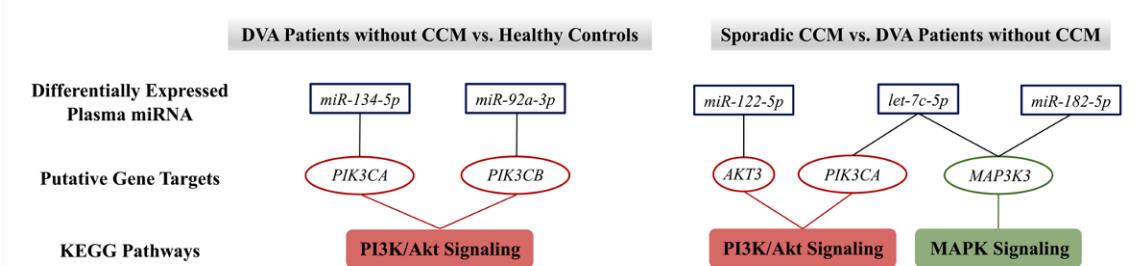
Many sporadic CCMs are found in the vicinity of a developmental venous anomaly (DVA). We hypothesized that CCM and DVA may have a common genetic origin, specifically that DVA may be a genetic precursor to CCM. To determine whether DVA and CCMs originate from a shared mutation, we collected three sporadic CCMs and sampled a portion of the associated DVA obtained during surgery (Figure 4.2A–C). Assays for mutations via ddPCR revealed that all three CCMs have a somatic activating mutation in *PIK3CA* and that the same mutation is present within the paired DVA samples at lower frequency (Figure 4.2D). Furthermore, ddPCR revealed that two of the CCMs harbored a mutation in *MAP3K3* in addition to the previously noted mutation in *PIK3CA*. However, unlike the *PIK3CA* mutation, the *MAP3K3* mutation was entirely absent from both DVA samples (Figure 4.2E). The presence of the *PIK3CA*, but not the *MAP3K3*, mutation in the DVA confirms that the *PIK3CA* mutation in the DVA did not arise via cross-sample contamination. The presence of multiple somatic mutations in these CCMs allows us to infer the developmental history of the lesion. The cancer field commonly uses the presence or absence of somatic mutations in clonal populations to track the evolutionary history of a tumor (Jiao et al., 2014; Jolly & Van Loo, 2018). Recent studies have expanded on this approach to use somatic mutations as endogenous barcodes to track embryonic development (Bizzotto et al., 2021). Using this same approach, we infer that the DVA was the first lesion to develop and that the associated CCM is derived from cells of the DVA following a somatic mutation in *MAP3K3*.

#### 4.2.4 Plasma miRNAs Reflect PI3K Activation in DVA Patients

In addition to assaying the presence of *PIK3CA* mutations in DVA associated with CCM, we would ideally also assay DVA that are not associated with CCM. Unfortunately, DVA are benign malformations and are not resected unless associated with an additional pathology. This has precluded the direct assessment of *PIK3CA* mutations in DVA without a CCM. To address this limitation, we sought another source of tissue that could be assayed for indirect evidence of *PIK3CA* activation. Thus, we collected plasma from individuals with DVA without a CCM and measured circulating miRNAs that might serve as biomarkers reflecting *PIK3CA* activity (Mori et al., 2019).

We sequenced the plasma miRNomes of 12 individuals with a sporadic CCM associated with a DVA (CCM + DVA), 6 individuals with a DVA without a CCM (DVA only), and 7 healthy controls. Three plasma miRNAs were DE in the DVA only group when compared to healthy controls ( $P < 0.05$ ; false discovery rate (FDR) corrected). Two of these DE miRNAs, *miR-134-5p* (FC = 0.10) and *miR-92a-3p* (FC = 3.10), putatively target *PIK3CA* and *PIK3CB* respectively, which are both components of the PI3K/AKT/mTOR pathway (Figure 4.3).

In addition, 18 plasma miRNAs were DE in CCM + DVA when compared to DVA only ( $P < 0.05$ ; FDR corrected). Two of the 18 DE miRNAs, *miR-122-5p* (FC = 5.25) and *miR-182-5p* (FC = 0.21), target *AKT3* and *MAP3K3* respectively, linking them to both the PI3K/AKT/mTOR and MAPK/ERK pathways. Additionally, *let-7c-5p* (FC = 12.64) targets both *PIK3CA* and *MAP3K3*, consistent with the role of these two pathways in sporadic CCM pathogenesis (Figure 4.3). Of interest, *let-7c-5p* also targets *COL1A1*, a DEG within the transcriptome of human sporadic CCM lesions (see Supplementary Data). This gene is associated with PI3K/AKT signaling, platelet activation, and ECM-receptor interaction KEGG pathways (Kanehisa et al.,



**FIGURE 4.3: Differentially Expressed micro-RNAs in CCM and DVA.**

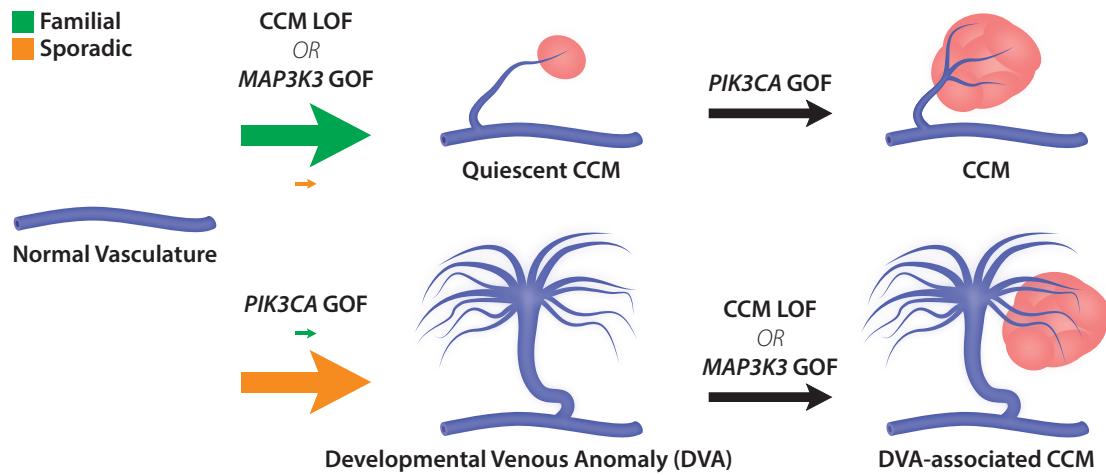
Two miRNAs differentially expressed (DE) in the plasma of individuals with DVA without a CCM, *miR-134-5p* and *miR-92a-3p*, putatively target genes in the Kyoto Encyclopedia of Genes and Genomes (KEGG) PI3K/AKT signaling pathway. In the plasma of individuals with sporadic CCMs with an associated DVA, *miR-122-5p*, *miR-182-5p*, and *let-7c-5p* were DE and found to target genes within both PI3K/AKT and MAPK signaling KEGG pathways. All results were  $P < 0.05$ ; false discovery rate corrected.

2021; Kanehisa & Goto, 2000), all of which have previously been implicated in CCM disease (Faurobert et al., 2013; Hong et al., 2021; Lopez-Ramirez et al., 2019; Ren et al., 2021).

Additionally, 28 DE plasma miRNAs were identified between CCM + DVA and healthy controls ( $P < 0.05$ ; FDR corrected). Four of these miRNAs putatively target *PIK3CA*: *miR-148a-3p* ( $FC = 3.27$ ), *miR-148b-3p* ( $FC = 2.64$ ), *miR-128-3p* ( $FC = 2.55$ ) and *let-7c-5p* ( $FC = 4.20$ ), which also targets *MAP3K3*.

#### 4.2.5 Relative Probability of Multiple Somatic Mutations

The association between DVA and sporadic—but not familial—CCMs suggests that there are at least two genetic trajectories by which a CCM may develop. The first trajectory is via a quiescent CCM caused by an initial mutation in *KRIT1*, *CCM2*, *PDCD10*, or *MAP3K3*. The second trajectory is via a DVA caused by an initial mutation in *PIK3CA* with subsequent mutations in *KRIT1*, *CCM2*, *PDCD10*, or *MAP3K3* leading to CCM formation. To understand whether one trajectory is favored in familial vs sporadic CCMs we use a simplified model to estimate the



**FIGURE 4.4: Genetic Model of CCM Pathogenesis.**

The genetic trajectories that underly familial and sporadic CCM pathogenesis. The initial path of familial and sporadic CCMs are denoted by arrow colors (green and orange respectively) with the relative probability of each path denoted by the size of the arrow. Familial CCM preferentially develop via a LOF mutation in the CCM complex (*KRIT1*, *CCM2*, *PDCD10*) to develop a quiescent CCM that may acquire an additional mutation in *PIK3CA* that drives lesion growth. Sporadic CCM are more likely to develop via a DVA caused by a *PIK3CA* mutation creating a clonal population of cells that may subsequently acquire a CCM complex LOF or *MAP3K3* GOF.

relative probability of CCM complex LOF (*KRIT1*, *CCM2*, *PDCD10*), *MAP3K3* GOF, and *PIK3CA* GOF.

Somatic mutagenesis is a complex process that depends on many variables including cell turnover rates, age, exposure to mutagens, and numerous genetic factors which facilitate DNA synthesis and repair. As a result, estimating absolute mutation rate is not feasible. However, by assuming a uniform mutation rate across these genes, we can determine the relative rate of CCM complex LOF, *MAP3K3* GOF, and *PIK3CA* GOF as simply the number of mutations that result in LOF or GOF. Determining this value for *MAP3K3* and *PIK3CA* is straightforward. To date there is only a single mutation in *MAP3K3* (p.I441M) that has been reported in CCMs. The spectrum of mutations in *PIK3CA* has been well documented in the catalogue

of somatic mutations in cancer (COSMIC). In the absence of functional assays for each mutation, we define a GOF mutation as any mutation that accounts for >1% of all *PIK3CA* mutations COSMIC which was determined to be 10 mutations. The number of CCM complex LOF mutations was determined by considering all possible single nucleotide variants that would result in a premature stop codon or disrupt a canonical splice site (see Methods). The average of these values for all three genes is 430 mutations (*KRIT1* = 718, *CCM2* = 353, *PDCD10* = 220). This is a very conservative estimate and should be considered a lower bound, since the majority of CCM complex LOF mutations are the result of frameshift mutations which are not accounted for here.

From the relative probability of each mutation, it is clear that familial CCMs will almost always develop via CCM complex LOF (Figure 4.4 top), consistent with the lack of association between familial CCM and DVA. The probability of two CCM complex LOF mutations in trans and in the same cell (as required for CCM complex LOF in sporadic CCMs) is far lower than for a single mutation (as required for CCM complex LOF in familial CCMs). In this study we identified 6 sporadic CCMs with biallelic somatic mutations in a CCM complex gene, therefore probability of this event is likely of similar magnitude to *MAP3K3* GOF which we identified in 15 sporadic CCMs. Assuming the same relative probability for CCM complex LOF as *MAP3K3* GOF in sporadic CCMs, we determine that sporadic CCM are at least 5x more likely to develop via *PIK3CA* GOF than CCM complex LOF or *MAP3K3* GOF (Figure 4.4 bottom). This finding is consistent with the strong association between DVA and sporadic CCM.

## 4.3 Discussion

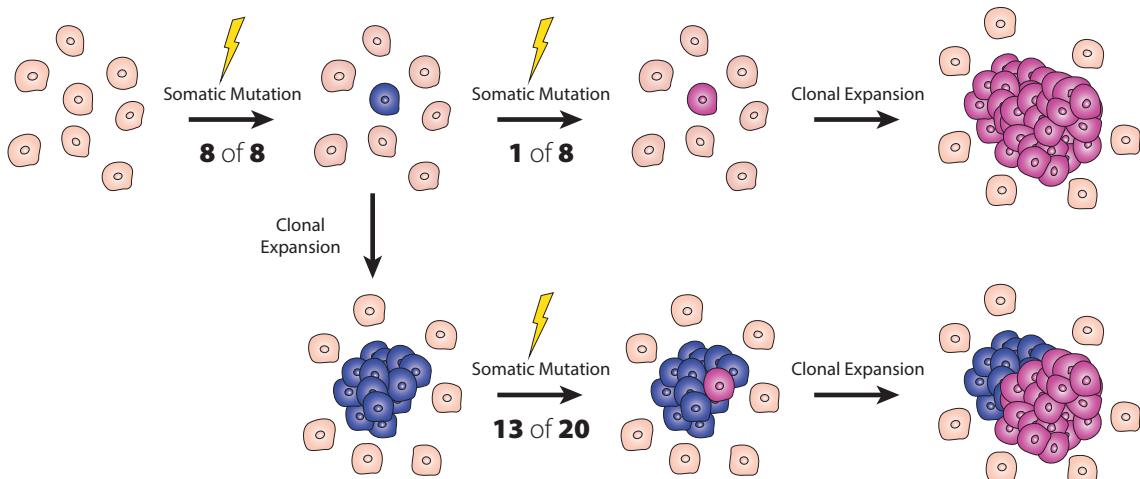
### 4.3.1 *MAP3K3 GOF and CCM LOF Have Similar Molecular Effects*

In this study we have further interrogated the relationship between somatic mutations in *KRIT1*, *CCM2*, *PDCD10*, *MAP3K3*, and *PIK3CA* which contribute to the pathogenesis of CCM. We find that somatic mutations in *MAP3K3* are not present in CCMs from individuals with familial CCM, consistent with a recent study (Weng et al., 2021). We find that sporadic CCMs may harbor mutations in *MAP3K3*, *KRIT1*, *CCM2*, or *PDCD10*, but that the lesion will only have mutations in one of these genes. This implies that mutations in any of *MAP3K3*, *KRIT1*, *CCM2*, or *PDCD10* are sufficient for CCM formation, without the need for mutations in a second gene. As *KRIT1*, *CCM2*, and *PDCD10* are all members of the CCM signaling complex, this reflects the fact that the loss of any component of the complex prevents normal signal transduction. Similarly, as the CCM complex is a direct inhibitor of *MAP3K3* activity (Zhou et al., 2015), this pathway may be activated by either CCM complex LOF or by *MAP3K3* GOF, but the mutual exclusivity of mutations in these genes suggests that only one of these events is necessary for lesion formation.

### 4.3.2 *DVA Clonal Expansion Promotes Additional Mutations*

CCMs often develop as the result of multiple somatic mutations that co-exist within the same cells as we show with snDNA-seq. Although several somatic mutations occur in every cell division (Rodin et al., 2021), the specificity of the mutations in CCM translates to a very low chance of acquiring these mutations within a single cell. This is especially true of somatic mutations in *MAP3K3* and *PIK3CA*, both of which have very narrow spectra of activating mutations. Despite this improbability, the accumulation of these mutations in CCM seems to occur frequently. One possible explanation for this phenomenon is that CCMs have an increased rate of somatic

mutations. There is little evidence supporting this theory and such a mechanism is difficult to conceive in cases of sporadic CCM where individuals have no known genetic predisposition to CCM. An alternative explanation is that after an initial somatic mutation, the singly-mutated cell undergoes clonal expansion to form an intermediate lesion. In this study we identify 7 CCMs with either biallelic LOF in a CCM complex gene or *MAP3K3* GOF in the absence a *PIK3CA* mutation, suggesting that *PIK3CA* activation is not required for CCM formation. Furthermore, previous work in mouse models has shown that loss of a CCM complex gene (with WT *Pik3ca*) leads to clonal expansion of the mutant cells and in later stages of growth, incorporation of WT endothelial cells (Detter et al., 2018; Malinverno et al., 2019). As a result of this clonal expansion, the probability of creating a double-mutant cell increases by a factor of the clonal population size as there are more cells in which the second mutation may occur (Figure 4.5). Somatic mutations in this clonal population may be acquired by genome replication during expansion or after



**FIGURE 4.5: Clonal Expansion Predisposes to Additional Mutations.**  
A schematic showing the probability of acquiring two somatic mutations with or without an intermediate clonal expansion step. In this example a single cell with an initial somatic mutation either remains quiescent (top) or clonally expands (bottom). The probability that a subsequent mutation occurs in a cell with the initial mutation increases by a factor of the mutant population (1 cell in top, 13 cells in bottom).

expansion as mutations accumulate via DNA damage and error-prone repair as has been reported in other quiescent cell types (Lodato et al., 2018).

The data presented in this study suggest that DVA are an intermediate lesion as posited by the latter explanation. Genetic analysis of DVA and paired CCM show that at least some DVA develop following a somatic activating mutation in *PIK3CA*. Furthermore, plasma miRNA analysis of individuals with DVA-associated CCM revealed differentially expressed miRNAs that putatively target *PIK3CA* and *MAP3K3*, the same genes that are mutated in these lesions. By contrast, individuals with DVA but no CCM differentially expressed miRNAs that putatively target *PIK3CA*, but not *MAP3K3*. These data are consistent with the distribution of mutations we observe in CCM and DVA, and the differential regulation of these miRNAs may reflect an attempt to compensate for dysregulation of PI3K and MAPK signaling due to the somatic mutations.

The presence of *PIK3CA* mutations in DVA suggests that DVA act as a genetic precursor to CCM, which would account for the strong association between sporadic CCM and DVA. Likewise, DVA are not associated with familial CCM because the presence of an inherited germline mutation in a CCM gene strongly biases probability towards a CCM gene somatic mutation occurring first, as there exist many different mutations that may cause LOF, but far fewer that would cause GOF in *PIK3CA*.

This study is the first to collect CCM and associated DVA for genetic analysis. Collecting tissue from CCM-associated DVA is challenging; however, collecting tissue from DVA not associated with CCM is yet more challenging as DVA are considered benign and are therefore not resected. We have attempted to address this limitation by studying biomarkers of PI3K activity which can be assayed noninvasively in blood plasma. Assaying the presence of *PIK3CA* mutations in DVA not associated with CCM will be the domain of future studies, but the data we present here demonstrate a clear link between DVA and *PIK3CA*, and suggest a model that explains the long

recognized—but poorly understood—association between CCM and DVA.

#### *4.3.3 DVA May Contribute to Additional Phenotypes*

While we are unable to address the presence of *PIK3CA* mutations in DVA not associated with CCM, it is worth noting that DVAs have been associated with other PI3K-related disorders(Brinjikji et al., 2018; Dhamija et al., 2018; Rasalkar & Paunipagar, 2010; Roux et al., 2020; Santucci et al., 2008; Tan et al., 2007; Ucler, 2019) including some cancers and neurological malformations, suggesting that DVA may have a role, possibly even as a genetic primer, in these other diseases.

### **4.4 Methods**

#### *Sample Collection*

Surgically resected CCMs were obtained from the University of Chicago, the Barrow Neurological Institute, and the Angioma Alliance biobank. Additional DVA tissue was discretely dissected from the lesion during surgical resection of the associated CCM at the University of Chicago. This study was approved by each institution’s respective Institutional Review Board.

#### *DNA Extraction*

DNA from CCM and DVA samples was extracted using the DNeasy blood and tissue kit (QIAGEN, catalog number 69504) per the manufacturers protocol. DNA purity was determined by Nanodrop and concentration was determined using the Qubit dsDNA BR assay kit (Invitrogen, catalog number Q32850) per the manufacturers protocol.

#### *Droplet Digital PCR*

Detection of *MAP3K3* p.I441M was performed via ddPCR using a previously described probe set(Couto et al., 2015). Assays were performed using 30–100ng of

DNA with the QX200 AutoDG system (BioRad) and quantified with the QX200 droplet reader (BioRad). Analysis was performed with the QuantaSoft software (BioRad).

### *Sequencing*

A total of 8 sporadic CCMs with no identified mutation in *KRIT1*, *CCM2*, *PDCD10*, or *MAP3K3* (5001, 5005, 5006, 5022, 5024, 5036, 5078, and 5081) were used for whole-exome sequencing prepared using the SureSelect Human All Exon V7 probe set (Agilent, Design ID S31285117) per the manufacturers protocol. Prepared libraries were sequenced on one lane of a NovaSeq 6000 S4 flow cell for a mean depth of 133x.

### *Sequence Analysis*

Sequencing data was processed using the Gene Analysis Toolkit (GATK, Broad Institute) while following the GATK best practices for somatic short variant discovery using Mutect2. Secondary variant detection was performed using gonomics (<https://github.com/vertgenlab/gonomics>) and bcftools mpileup to manually examine *KRIT1*, *CCM2*, *PDCD10*, and *MAP3K3* for somatic variants. Putative variants were annotated using Funcotator (GATK), the catalog of somatic mutation in cancer (COSMIC), and the genome aggregation database (gnomAD). Putative variants were filtered according to the following criteria: greater than 50x total coverage, less than 90% strand specificity, greater than 5 reads supporting the alternate allele, greater than 1% alternate allele frequency, less than 1% population allele frequency, and predicted protein/splicing change.

### *Single-Nucleus DNA Sequencing*

Nuclei isolation, snDNA-seq, and analysis were performed as previously described (Ren et al., 2021).

### *miRNA Extraction and Sequencing*

Total plasma RNA was extracted from the plasma of 12 individuals with a sporadic CCM and an associated DVA (CCM + DVA), 6 individuals with DVA and without a CCM (DVA only), and 7 healthy controls using the miRNeasy Serum/Plasma Kit (Qiagen, Hilden, Germany) following the manufacturer isolation protocol. Diagnosis of CCM with an associated DVA, as well as DVA without a CCM lesion was confirmed on susceptibility weighted MR imaging. Illumina small RNA-Seq kits (Clontech, Mountain View, CA, USA) were then used to generate cDNA libraries, and sequencing was completed with the Illumina HiSeq 4000 platform (Illumina, San Diego, CA, USA), with single-end 50bp reads, at the University of Chicago Genomics Core. Differential miRNA analyses were completed between (1) CCM + DVA to DVA only and then (2) DVA only to healthy controls. The differentially expressed miRNAs were identified having  $P < 0.05$ , FDR-corrected. All analyses were completed using the sRNAToolbox and DESeq2 R packages(Love et al., 2014; Rueda et al., 2015).

### *Identification of Putative Targets*

miRWalk 3.0 was used to identify the putative gene targets of each of the DE miRNAs, using a random forest tree algorithm with a bonding prediction probability higher than 95% on the 3 different gene locations (3' UTR, 5' UTR, and CDS)(Sticht et al., 2018). Putative gene targets of the DE miRNAs were identified in at least 2 of the 3 databases. DE miRNAs between (1) CCM + DVA and DVA only as well as (2) DVA only and healthy controls were then analyzed for potential targeting of the PI3K, AKT, and MAPK gene families.

### *Estimating Possible LOF Mutations in KRIT1, CCM2, and PDCD10*

The majority of LOF mutations in *KRIT1*, *CCM2*, and *PDCD10* result in either: creation of a premature stop codon (nonsense); disruption of a canonical splice site; or a frameshift. The first two of these are typically caused by a single nucleotide variant (SNV) and can be determined from the gene sequence. However, frameshift variants resulting from insertions and deletions may occur in any exon regardless of sequence context. This prevents a meaningful comparison between indel and SNV events, without making assumptions about the rate of somatic SNV and indel events. To determine a conservative lower bound for the number of LOF mutations in the CCM complex genes, we only consider nonsense and splice site mutations. The number of nonsense and splice site mutations were determined from the sequences of *KRIT1* (ENST00000394507.5), *CCM2* (ENST00000258781.11), and *PDCD10* (ENST00000392750.7). Nonsense mutations were considered any SNV in the coding sequence that would result in an in frame stop codon prior to the stop codon in the reference sequence. Splice site mutations were considered any SNV in the nearly invariant sequences at each exon-intron boundary that mediate canonical splicing.

### *Sporadic CCM Transcriptome*

Laser micro-dissected neurovascular units (NVUs) from sporadic CCM lesions were sequenced for transcriptomic analyses in comparison to micro-dissected NVUs from human brain microvasculature. The differential analyses identified 426 DE genes (DEGs) ( $P < 0.05$ ; FDR corrected;  $FC > |1.5|$ ). Additionally, 8 pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) were enriched ( $P < 0.05$ , FDR corrected).

### *Sporadic Transcriptome and KEGG Pathway Analyses*

Six human sporadic CCM lesions were surgically resected prior to embedding in optimal cutting temperature, snap-freezing, and storing at -80°C. Control brain tissue was collected during autopsy from three subjects lacking neurological disease, fixed in formalin, and embedded in paraffin blocks. Five- $\mu$ m tissue sections were mounted on Leica glass slides (Leica Biosystems Inc) and were stained, in accordance with the manufacturer's protocols, with HistoGene (Applied Biosystems) for frozen tissue and Paradise stain (Applied Biosystems) for paraffin-embedded tissue. The neurovascular units (NVUs) from sporadic CCM lesions and normal brain capillaries were then collected using laser capture microdissection and stored at -80°C. RNA was isolated using an RNA extraction kit (RNeasy Micro Kit, Qiagen). cDNA libraries were then generated with low-input strand-specific RNA-Seq kits (Clontech) and sequenced using the Illumina HiSeq 4000 platform with single-end 50bp reads. Differentially expressed genes were defined as  $P < 0.05$ , FDR corrected.

Enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways(Kanehisa et al., 2021; Kanehisa & Goto, 2000) were obtained for the DE genes (DEGs) in the sporadic transcriptome with fold change values greater than 1.5 ( $P < 0.05$ , FDR corrected). The sporadic transcriptome was compared to the putative targets of the DE miRNAs between (1) CCM + DVA and DVA only as well as (2) DVA only and healthy controls to obtain a set of overlapping genes. Enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were obtained for the overlapping genes, using a database and knowledge extraction engine with a Bayes factor greater than 3.

## 4.5 Contributions and Acknowledgements

This chapter is adapted from a study in press at *Nature Cardiovascular Research* (Snellings et al., 2022) with the following authors: Daniel A. Snellings, Romuald Girard, Rhonda Lightle, Abhinav Srinath, Sharbel Romanos, Ying Li, Chang Chen, Aileen A. Ren, Mark L. Kahn, Issam A. Awad, Douglas A. Marchuk. DAS designed and performed the genetic studies of human CCM lesions and wrote the manuscript. IAA performed surgical resection of CCM and DVA samples used in this study. RG, RL, AS, SR, YL, CC and IAA performed plasma miRNA sequencing and analysis. AAR and MLK assisted with experimental design. RG, IAA, and DAM designed experiments and wrote the manuscript.

We thank the patients who donated tissue for this study. We thank Angioma Alliance, the Barrow Neurological Institute, and the University of Chicago for patient enrollment and sample collection. Nucleus sorting was performed in the Duke Human Vaccine Institute Research Flow Cytometry Shared Resource Facility. We thank Duke University School of Medicine for use of the Sequencing and Genomic Technologies Shared Resource for library preparation and sequencing. These studies were supported by National Institute of Health grants, P01NS092521 (DM, IA, MK), F31HL152738 (DS).

# 5

## Conclusions & Musings

Vascular malformations are a class of lesion that are both phenotypically and genetically heterogeneous. Despite this heterogeneity, they share an overarching pathogenic mechanism whereby the initial formation of a lesion is seeded by a somatic mutation. These mutations may be in genes involved in canonical vascular development and maintenance or may be in well characterized oncogenes; nonetheless, somatic mutations play an integral role in the pathogenesis of many—if not all—vascular malformations. The studies described herein were performed with the goal of broadening our understanding of how somatic mutations may contribute to the pathogenesis of vascular malformations. Specifically, I aimed to show that somatic mutations may drive three distinct phases in the pathogenesis of vascular malformations: initiation (Chapter 2), progression (Chapter 3), and predisposition (Chapter 4).

The role of somatic mutations in the initiation of vascular malformations is widely accepted and relatively well studied, however the vascular malformations associated with hereditary hemorrhagic telangiectasia (HHT) remained one of the few VM disorders for which the contribution of somatic mutations remained unclear. Theories persisted for decades that telangiectasia and visceral AVMs in HHT were due to somatic mutations in a Knutsonian two-hit mechanism; however, due to a combination of tissue rarity and the limitations of sequencing technology, the theory remained untested. We were able to acquire punch biopsies of telangiectasia from individuals with HHT to test this hypothesis using next-generation sequencing and I was successfully able to show that the telangiectasia in HHT follow a two-hit mechanism. Notably, I only found two-hit somatic mutations in 9 of 19 telangiectasia suggesting there may be additional mutations that we missed in this study which will be further discussed in the following section. This finding fills a critical gap in our understanding of the genetics of HHT and adds to the growing list of VM disorders that are caused by somatic mutations.

The somatic mutations identified in vascular malformations to date have been monogenic. In my studies of cerebral cavernous malformations (CCMs) I found that CCMs often develop via digenic somatic mutations. Similar to what I showed for HHT, CCMs have been known to develop via a two-hit mechanism caused by somatic mutations in either *KRIT1*, *CCM2*, or *PDCD10*. In Chapter 3, I found that many CCMs harbored a somatic gain of function mutation in *PIK3CA* in addition to the previously described somatic mutations. In these CCMs, we have found as many as 3 distinct somatic mutations in a single lesion and I further showed that these mutations are all present in the same population of cells. Work from our collaborators, Aileen Ren and Mark Kahn, showed a potent synergistic effect of biallelic LOF in *KRIT1* and GOF in *PIK3CA* where either genotype alone only confers a modest vascular phenotype. These data suggest that CCMs that formed via biallelic LOF in a CCM gene may acquire a somatic mutation in *PIK3CA* that may induce lesion progression years after the initial lesion forms. This finding accounts for the clinical observations of rapid growth of CCMs that have been quiescent for years and also account for the significant intra-individual variability of lesions in familial CCM. This finding constitutes the first evidence of a digenic, three-hit mechanism, and suggests that secondary somatic mutations in vascular malformations may drive lesion progression after their initial formation. To date, this remains the only study where secondary somatic mutations have been identified in a vascular malformation, however future studies may find that this is a general mechanism of VM pathogenesis.

My finding that as many as 3 different somatic mutations could co-exist in a single sporadic CCM came as a surprise. There are numerous clinical cases of CCM of sporadic CCMs developing later in adulthood; a period where our mouse models suggest that CCMs cannot form from biallelic LOF in a CCM or *PIK3CA* GOF alone. This suggests that these cases must acquire 3 specific somatic mutations without forming an intermediate CCM—i.e. all 3 mutations must occur in an individual cell.

This series of events is extremely unlikely so I explored alternate hypotheses that may predispose to sporadic CCM formation. Previous studies of genetic predisposition to CCM have focused on germline variation, however I hypothesized that a benign somatic mosaic structure in the brain may predispose to sporadic CCM formation; specifically, developmental venous anomalies (DVA). DVA are the most common vascular malformation present in up to 16% of the general population, and are generally considered to be benign. Nearly all sporadic CCM directly abut a DVA, however the cause of this association has remained a mystery. I showed that DVA and the adjacent CCM harbor a shared somatic GOF mutation in *PIK3CA*, however the DVA lacks the *MAP3K3* mutation present in the adjacent CCM. This shows that DVA function as an intermediate lesion, effectively predisposing to sporadic CCM formation. This highlights DVA as a potential risk factor in CCM development and future studies may find that the *PIK3CA* mutation in DVA may predispose to the formation of other *PIK3CA*-mutated diseases.

Together, the studies presented herein suggest that somatic mutations may play a wider role in the pathogenesis of vascular malformations than previously appreciated. My research into the novel roles of somatic mutations has been focused on CCMs, however future studies of other vascular malformations may find additional somatic mutations that drive various aspects of their pathogenesis.

In the following sections I will discuss several open lines of inquiry, potential future directions, and assorted hypotheses related to the pathogenesis of vascular malformations.

## 5.1 Missing Mutations

One important open question across my studies is the mutation status of lesions where I did not identify somatic mutations. In Chapter 2, I did not find a somatic mutation in 10 of 19 telangiectasia. In Chapters 3 and 4, I did not find causal

somatic mutations (CCM genes or *MAP3K3*) in 25 of 51 sporadic CCMs and 10 of 20 familial CCMs. It is possible that these lesions have mutations in other genes that we did not sequence; however, I believe it is more likely that these lesions harbor mutations that cannot be detected by my sequencing method. Throughout these studies I used targeted short-read sequencing for mutation discovery. This method is extremely powerful for detecting single nucleotide variants and relatively small insertions or deletions, but is unable to detect all possible types of somatic loss of function mutations. In this section I will discuss two types of mutation that may result in loss of function, but are not detectable by traditional short-read sequencing approaches: somatic hypermethylation, and somatic loss of heterozygosity.

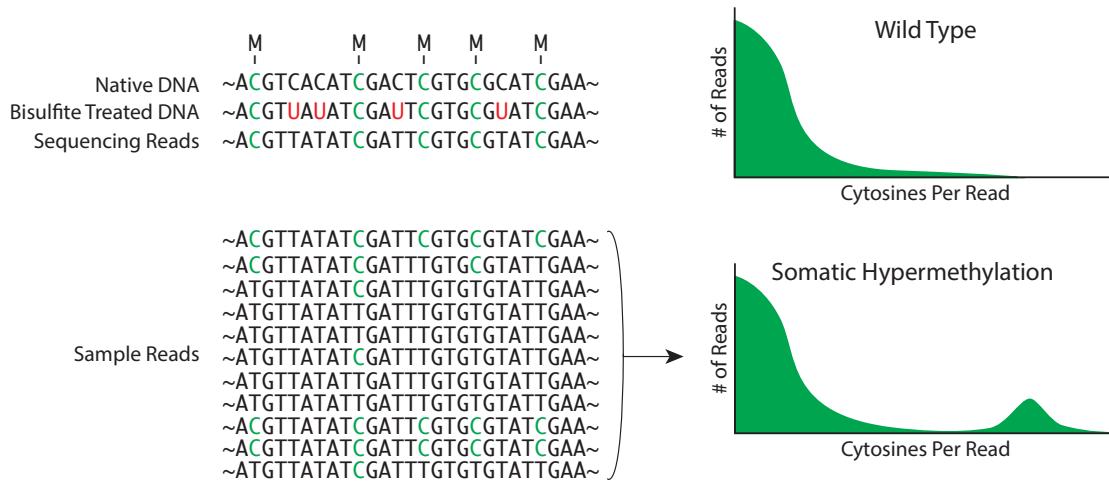
### 5.1.1 Somatic Hypermethylation

Cytosine methylation occurs at CpG dinucleotides and results in the addition of a methyl group *ortho* to the amine of cytosine. Methylation of cytosines in the promoter region of a gene is a well established regulatory mark that results in decreased expression. In cancer, increased methylation (or hypermethylation) in the promoter of a tumor suppressor gene is a common and effective method of gene silencing as reviewed here (Baylin, 2005). Similar to cancer, it is possible that somatic hypermethylation in the promoters of causal genes in HHT or CCM may result in silencing—and therefore LOF—which would be invisible to traditional sequencing.

I attempted to detect somatic hypermethylation events in mutation-negative telangiectasia using targeted bisulfite sequencing. Bisulfite treatment of DNA converts any unmethylated cytosines to uracil via deamination such that any cytosines in the sequencing reads can be identified as methylated (Figure 5.1). I first identified CpG islands in the promoters of *ENG*, *ACVRL1*, and *SMAD4* (Figure 5.2) and designed primers to target the islands in *ENG* and *ACVRL1* (Figure 5.3). Though *SMAD4* causes HHT, we did not design primers for this gene as we did not have any

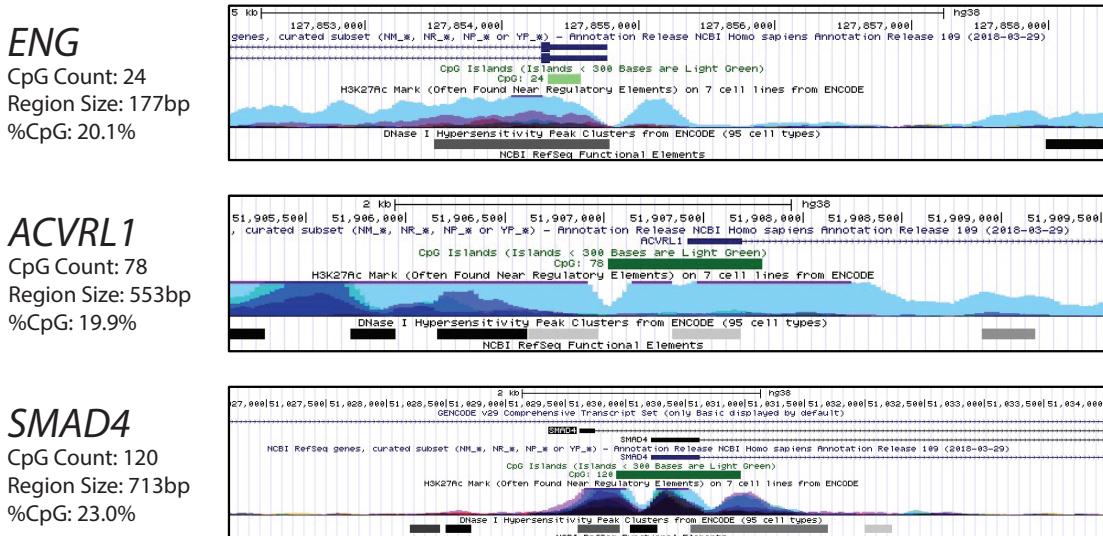
telangiectasia from individuals with a *SMAD4* germline mutation. As a control for these experiments we considered telangiectasia with a known somatic mutation as a negative control for somatic hypermethylation and compared them to mutation-negative telangiectasia which may have somatic hypermethylation as a means of LOF.

These experiments were ultimately uninformative as we observed high variability in the methylation rate in the *ENG* promoter and we only had a single sample with a germline mutation in *ACVRL1* that did not have a known somatic mutation (Figure 5.4). These results are likely confounded by the limited efficiency of bisulfite conversion as well a very limited sample size. In addition any somatic hypermethylation would be expected to occur mainly in endothelial cells; however, these crude punch biopsies contain a substantial amount of unaffected tissue around the telangiectasia, and endothelial cells are likely only a fraction of all cells in the



**FIGURE 5.1: Predicted Somatic Hypermethylation Profile.**

Bisulfite treatment converts any unmethylated cytosines to uracil which is read as a thymine on the sequencing. Wild type DNA should have low levels of methylation and a long tail representing inefficiencies in the bisulfite conversion and sequencing errors (top right). In contrast, a sample with somatic hypermethylation should have a similar distribution to wild type DNA, but is expected to have a secondary peak representing reads with heavy methylation (bottom right).



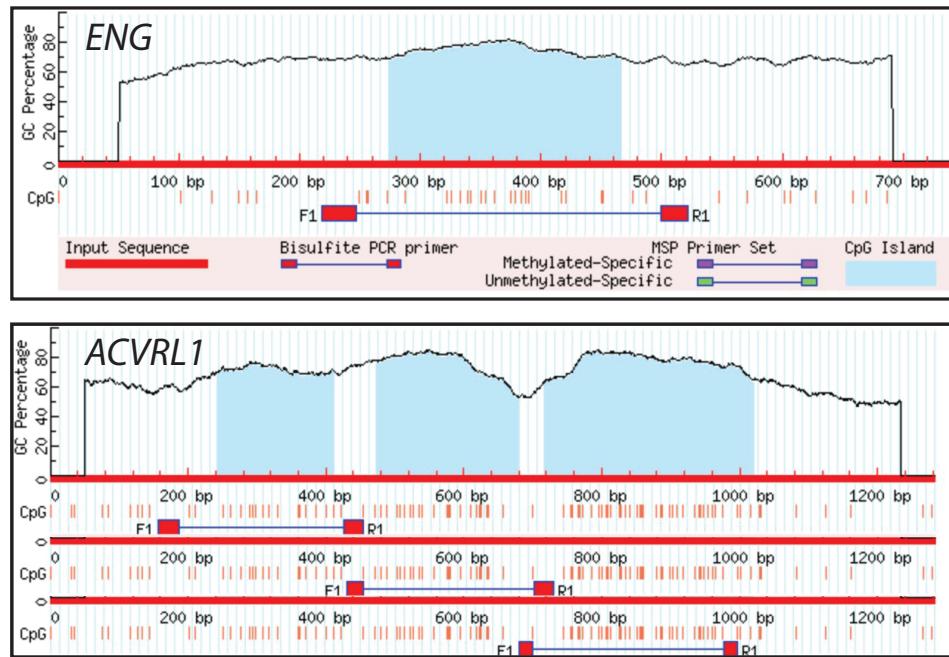
**FIGURE 5.2: CpG Islands in HHT Gene Promoters.**

UCSC genome browser images of the promoter regions of *ENG*, *ACVRL1*, and *SMAD4*. CpG islands are marked in light green. (<300bp) or dark green (>300bp) along with H3K27Ac marks over the region.

biopsy. More samples will be required to determine if somatic hypermethylation of endothelial cells could result in LOF via silencing, though we will also need to overcome the technical limitations we encountered in these pilot experiments.

Firstly, bisulfite conversion is not 100% efficient. This is not a concern when evaluating germline methylation levels, however high conversion efficiency will be required to pick up low frequency somatic events in a highly heterogeneous tissue. As an alternative to bisulfite conversion, several long-read sequencing technologies (e.g. Pacbio and Oxford Nanopore) are capable of detecting DNA base modifications such as methylation during sequencing. This would be a more direct measure of methylation and would bypass the need for bisulfite conversion.

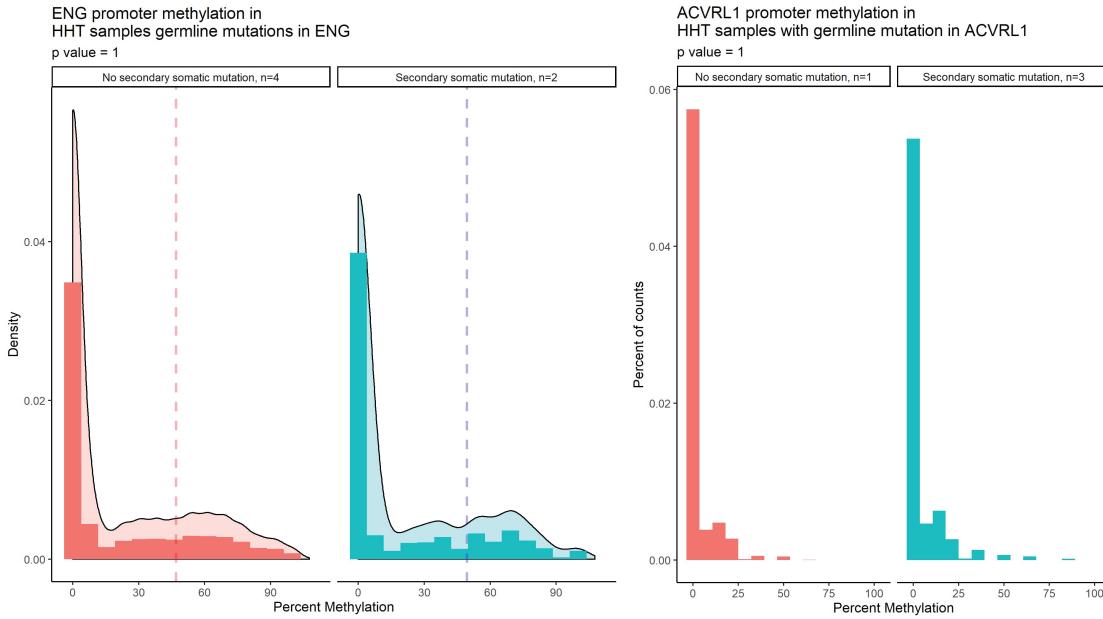
The second major limitation in these experiments is the heterogeneous tissue. It would be very beneficial to isolate endothelial cells prior to DNA extraction to increase the allele frequency of any potential somatic event. Traditionally this would



**FIGURE 5.3: Primers Targeting CpG Islands in *ENG* and *ACVRL1*.**

Primer design targeting dense regions of CpG dinucleotides (light blue) in *ENG* and *ACVRL1*. The CpG island in *ACVRL1* was too large to cover in a single amplicon, therefore I designed three primers to cover the regions of highest density, while designing primers to avoid CpGs which may impact annealing efficiency.

be done by generating a single-cell suspension and staining for an endothelial cell-specific protein such as PECAM and sorting based on that marker. As we receive telangiectasia primarily from Canada, it is necessary that they be frozen prior to shipment which disrupts the cell membrane thereby precluding the generation of a single-cell suspension. However, I have had some success generating single-*nucleus* suspensions from telangiectasia, similar to what I have done for CCMs in Chapters 3 and 4. The nuclei yield from telangiectasia is low relative to CCMs (around 20,000 per telangiectasia), however it is feasible to stain these nuclei. The common endothelial cell stain PECAM is not present on the nuclear membrane; however the transcription factor ERG is present inside nuclei, and is highly specific to endothelial cells. ERG is a commonly used stain and I have attempted to stain nuclei with it, however my



**FIGURE 5.4: Promoter Methylation in *ENG* and *ACVRL1*.**

Methylation of telangiectasia with (green) or without (red) a previously identified somatic mutation for *ENG* (left) or *ACVRL1* (right). The *ENG* promoter shows high levels of somatic methylation in both the experimental samples and the controls. The *ACVRL1* promoter shows rather low level of basal methylation, but shows no significant secondary peak as might be expected by somatic hypermethylation.

initial experiments were unsuccessful and I eventually abandoned the project in favor of bulk sequencing. This would be a powerful tool if a robust staining protocol could be developed.

### 5.1.2 Somatic Loss of Heterozygosity

The second type of event I will discuss is somatic loss of heterozygosity (LOH) events. LOH is a general term that refers to loss of an allele resulting in a previously heterozygous site becoming homozygous. For example, loss of the wild type *ENG* allele in an individual with a germline *ENG* mutation would result in biallelic LOF and could theoretically be a source of the missing mutations. LOH can result from a variety of different genomic events such as a large deletion or mitotic recombination.

Like somatic hypermethylation, LOH is a common mechanism in cancers to disrupt tumor suppressor genes. While targeted short-read sequencing is great for detecting small variants, it is blind to these chromosome-scale events which occur in a small number of somatic cells—i.e. we miss the forest for the trees.

LOH in cancer samples is often detected via a SNP array such that an LOH event of reasonable frequency (often  $\geq 25\%$ ) can be visualized as a shift in copy number for a subset of markers—e.g. heterozygous markers in an unaffected region of the chromosome are present at 1 : 1 ratio, whereas markers over an LOH event at 25% AF are present at a 1.25 : 0.75 ratio. This method works well for events with high allele frequencies, however the somatic variants I find in CCM and HHT are far lower, with most  $< 5\%$ . This makes detection by SNP array challenging as we must identify a ratio of 1.05 : 0.95 in a fairly noisy assay. Reaching statistical significance with such a minor shift would require a dense SNP array and a large LOH event making this methodology infeasible.

While detecting these events is nigh impossible with bulk sequencing, it is theoretically trivial with single-nucleus DNA sequencing (snDNA-seq). As an initial experiment, I performed snDNA-seq on familial CCM samples for which I did not find a somatic mutation. To detect somatic LOH in these samples, I needed only to sequence the pathogenic germline mutation and determine whether that heterozygous mutation was homozygous in some of the cells (Figure 5.5). The problem with this approach, as discussed in Chapter 3, is that snDNA-seq is confounded by stochastic allelic dropout (ADO) at a rate of 5–10% per allele. As a result, if 1000 nuclei are sequenced, we find that around 800 have the expected heterozygous genotype, 100 will be homozygous for the reference allele, and 100 will be homozygous for the alternate allele. As ADO often occurs at similar or higher frequencies than the expected somatic LOH events, this targeted method is infeasible.

In an attempt to overcome ADO I designed a panel for snDNA-seq that targets

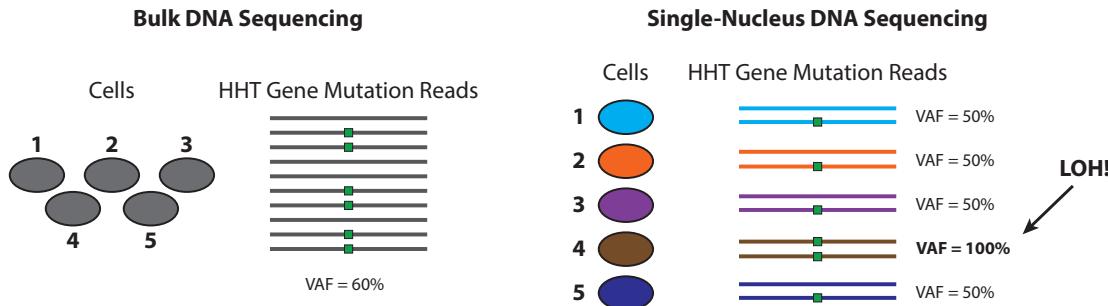


FIGURE 5.5: Somatic LOH Detection with snDNA-seq.

Reads from bulk sequencing show a pathogenic mutation at an allele frequency of 60%. From in reads from single-nucleus sequencing it is clear that four cells have the expected heterozygous genotype, but one cell is homozygous for the pathogenic mutation, indication a somatic LOH event.

common SNPs such that we capture a high density of SNPs in the region of the genes of interest, with progressively lower density of markers towards the distal ends of the chromosome. This pattern was selected because there is a limit to how many markers we can use on a single panel (about 400 amplicons) and this pattern maximizes our chance of finding informative heterozygous SNPs in the region of the gene, while still maintaining some sensitivity for larger events that may impact a large portion of the chromosome. Any LOH event should convert multiple contiguous heterozygous markers to homozygosity and would theoretically be robust against random ADO of individual amplicons. Using this method, I again found that ADO was a major confounder. I initially expected that ADO would be largely stochastic such that each marker would be independent of the state of other markers. However, I found that ADO events were in fact ‘linked’, such that a set of markers that was physically close on one allele would drop out as a set (Figure 5.6). In effect, I was seeing dropout of entire haplotypes rather than individual markers. This likely reflects the mechanism by which ADO occurs. Mostly likely ADO results from incomplete lysis of nuclei which leaves some regions of chromosomes tightly bound in chromatin and inaccessible to primers. As the chromatin would be bound over a contiguous

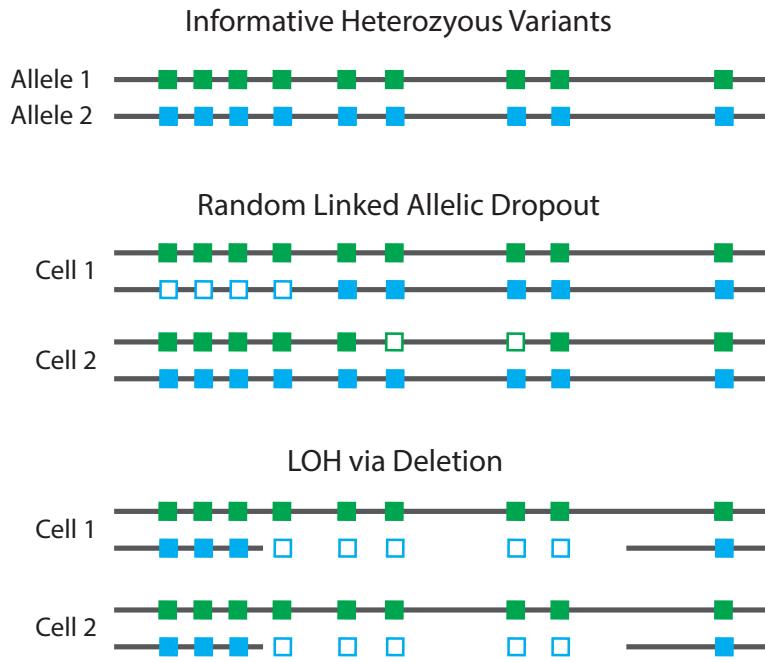
		Likely ADO Products (alt alleles on the same haplotype)				
		Clone 1	Clone 2	Clone 3	Clone 4	Clone 5
50 kb Region	<u>Germline Het SNP</u>					
	chr3:167402916 G/C	Het	WT	Hom	Het	WT
	chr3:167426660 C/T	Het	WT	Hom	Het	WT
	chr3:167451117 A/G	Het	WT	Hom	WT	<b>Het</b>
	chr3:167451230 C/T	Het	WT	Hom	WT	<b>Het</b>
Total Nuclei		1520	113	95	34	29
		(85%)	(6.3%)	(5.3%)	(1.9%)	(1.6%)
Linked ADO Across Haplotypes						

FIGURE 5.6: Linked Allelic Dropout.

Results of snDNA-seq of four SNPs, not involved in disease pathogenesis, that span a 50kb region and are all heterozygous in the germline. 85% of the sequenced nuclei have the expected heterozygous genotype (Clone 1). However, Clones 2 and 3 show linked dropout of the entire alternate or reference allele at 6.3% and 5.3% respectively. Similarly, Clones 4 and 5 show linked dropout of the alternate allele in a smaller region of the last 2 or first 2 variants that are physically close.

genomic region, we therefore get dropout of all markers that fall within the region of undigested chromatin. This phenomenon unfortunately looks almost exactly like what we would expect to see from a genuine somatic LOH event, however there are some important differences that I attempted to leverage to discriminate between linked ADO and LOH.

One critical difference between linked ADO and LOH is that linked ADO appears to occur randomly throughout the genome such that each nuclei should have a unique set of linked ADO products. In contrast, a biological LOH event should have identical breakpoints (Figure 5.7). The breakpoints generated by linked ADO and LOH can be determined in individual nuclei by identifying runs of homozygosity which can then be used to identify cells with identical breakpoints (My programs for this are available at <https://github.com/dasnellings/weaver>). In addition to differences

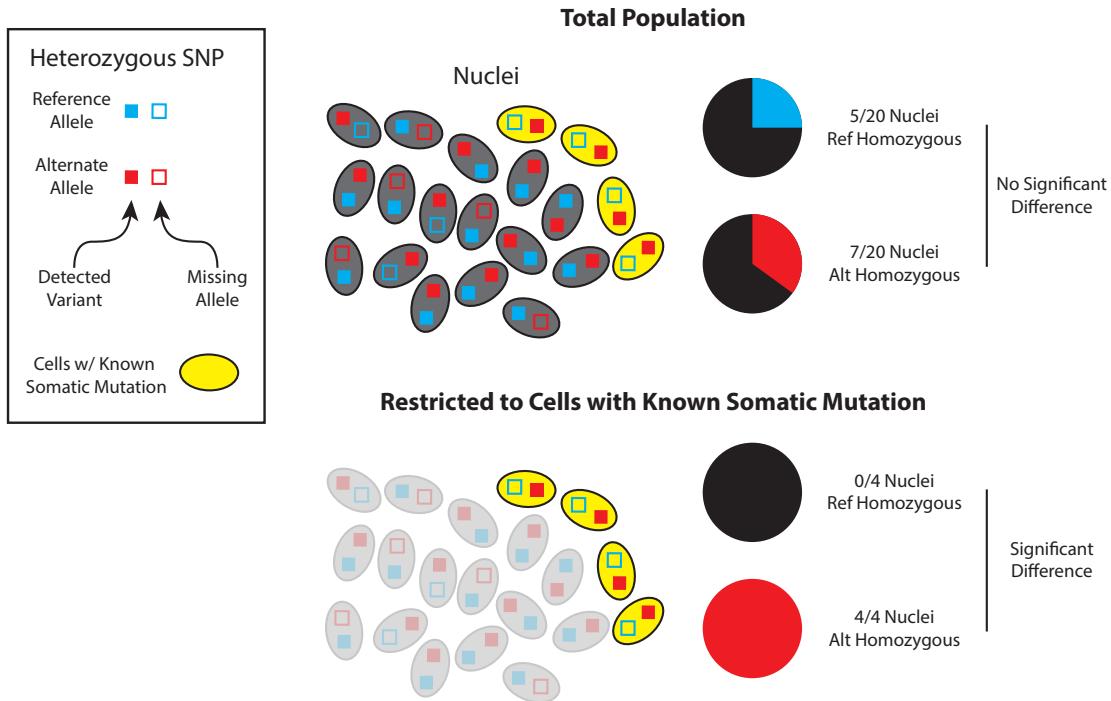


**FIGURE 5.7: Differences Between Linked ADO and LOH.**

Schematic illustrating differences between linked ADO and LOH by observing their effects on informative heterozygous variants on two alleles (green and blue squares). Linked ADO occurs randomly and affects both alleles equally as shown by dropped alleles (outlined squares). In contrast, LOH (caused by a deletion in this example) has identical breakpoints in all cells and is only present on a single allele.

in the breakpoints, linked ADO is expected to affect both alleles symmetrically, whereas LOH is asymmetric affecting only one allele.

In addition to leveraging the differences between linked ADO and LOH I have used one additional strategy to detect LOH events. Linked ADO occurs at a rate of 5–10% and LOH events are expected to occur as a similar or lower frequency, however pathogenic LOH events are expected to result in biallelic LOF. In a sample where one somatic mutation is already known, we expect that the LOH event should be present in the same population of cells. Therefore, by excluding any cells that do not contain the known somatic mutation, we can effectively enrich for cells likely to have an LOH event and thereby increase the allele frequency of the LOH event above the



**FIGURE 5.8: Leveraging Known Somatic Mutation for LOH Detection.**  
A known somatic mutation allows for enrichment of cells likely to have an LOH event. When considering the total population, the high rate of ADO confounds the detection of LOH events that occur at a similar frequency. However, when cells without a known somatic mutation are excluded, the LOH event may be enriched, increasing the allele frequency above the level of ADO noise.

ADO signal in this restricted population (Figure 5.8). The major limitation of this method is that it is only possible with samples that have at least 1 known somatic mutation, making it unusable for HHT and restricted to a subset of CCMs.

I have used this panel design and analysis method on 15 CCMs, 10 of which had at least 1 somatic mutation that could be used to restrict the population. From these experiments, I have not identified any somatic LOH events that are likely to be pathogenic; however, in one sample I have identified a putative somatic LOH event on chromosome 12 that does not overlap *KRIT1*, *CCM2*, or *PDCD10* (Figure 5.9). This event is most likely an incidental finding, however this event should be confirmed by a secondary method as a validation of the methodology.

Variant ID & Genotype					
Chr	Start	End	486–487–489–491–492–493–499	Nuclei Count	
chr12	62007874	82145194	Het–Het–Ref–Ref–Ref–Alt–Ref	188	
chr12	60968344	82145194	Het–Ref–Ref–Ref–Ref–Alt–Ref	26	
chr12	60968344	62007972	Het–Ref–Ref–Ref–Ref–Alt–Het	21	
chr12	60968342	62007972	Ref–Ref–Ref–Ref–Ref–Alt–Het	10	
chr12	60968342	82145194	Ref–Ref–Ref–Ref–Ref–Alt–Ref	9	
chr12	60968342	62007947	Ref–Ref–Ref–Ref–Ref–Het–Het	7	

Variant ID & Genotype					
457–459–460–461–462–463					
Chr	Start	End	Alt–Ref–Alt–Ref–Ref–Alt	Nuclei Count	
chr12	52711374	53129118	Alt–Ref–Alt–Ref–Ref–Alt	26	
chr12	52711374	53129118	Ref–Alt–Ref–Alt–Alt–Ref	25	

**FIGURE 5.9: Putative Somatic LOH Event in a CCM.**

Runs of homozygosity present in CCM 5056 derived from snDNA-seq data from 2500 nuclei. **A**, 6 different runs spanning a 21Mb region. All 6 runs are likely generated from the same LOH event, however ADO on the edges of the event are likely generating false breakpoints. Note that there were no run of homozygosity for the reciprocal allele that consisted of at least 5 contiguous variants (the threshold for this analysis). **B**, An example of linked ADO from the same sample and chromosome showing symmetric dropout of both haplotypes at similar frequencies.

It is a near certainty that somatic LOH events may drive the formation of CCMs and telangiectasia in HHT, however their detection has thus far been elusive. It is currently unclear if this is due to the rarity of such events or if there remain flaws in my methodology that preclude their detection.

## 5.2 Connection Between Sporadic Brain AVMs and HHT

Brain AVMs are common in individuals with HHT, however the majority of brain AVMs in the population occur sporadically in individuals with no known genetic predisposition. This paradigm of similar pathologies in sporadic lesions and lesions from a familial genetic disorder is very similar to CCM. CCMs also have a familial and sporadic form and research from myself and others has shown that both forms share many aspects of pathogenesis and are ultimately caused by identical mutations.

I hypothesized that this was also the case for sporadic and HHT-associated brain

AVMs.

In Chapter 2 I showed that HHT-associated telangiectasia are caused by a two-hit mechanism. This study focused on telangiectasia as HHT-associated visceral AVMs are difficult to study since they are rarely resected in favor of embolization and radiosurgery. Unlike brain AVMs in HHT, sporadic brain AVMs are commonly resected and are readily available to researchers. Previous studies found that many sporadic brain AVMs have a somatic mutation in *KRAS* (Nikolaev et al., 2018), however it remains unclear if these samples may have additional mutations. To test whether sporadic brain AVMs harbor somatic mutations in *ENG*, *ACVRL1*, or *SMAD4*, I sequenced 68 presumed sporadic brain AVMs that were surgically resected at the Barrow Neurological Institute.

As reported by others I found that 24 of the 68 had a gain of function mutation in *KRAS*. I also identified one sample with a gain of function mutation in *TEK* (p.L194F) which is known to cause venous malformations (Limaye et al., 2009). This sample may represent a novel mutation for AVMs, but more likely is the result of a misclassification of a venous malformation as an AVM. One additional sample has a mutation in *VHL* (p.N78S) which has been shown to be moderately activating in some assays (Rechsteiner et al., 2011), however it remains unclear if this is a causal mutation for AVMs. In one sample I found a germline missense mutation in *ENG* (p.R406C, 550/1178 reads, 47% VAF) which has not been previously reported to cause HHT, however given the context, it is highly likely to be pathogenic. I reviewed the case notes from the Barrow Neurological Institute for this individual and there was no information that could either support or reject an HHT diagnosis (telangiectasia, family history, epistaxis). Notably, this sample did not have a GOF mutation *KRAS* and I did not identify any somatic mutations in *ENG*.

In 3 of the 68 brain AVM samples I found a mutation in *ASXL1* (p.S1344F, p.W591\*, p.Q512\*), the latter found in a sample with a GOF *KRAS* mutation.

These mutations are all likely to cause LOF. LOF mutations in *ASXL1* have never been reported in a vascular malformation, however they are commonly found in cases of clonal hematopoiesis—a phenomenon where hematopoietic stem cells responsible for forming blood cells acquire mutations that give them a survival advantage and become the dominant clone in the blood as an individual ages (Fujino & Kitamura, 2020). This finding was initially very exciting as the mutation was found in multiple brain AVM samples; however, upon reevaluation it is most likely that these individuals had some degree of clonal hematopoiesis and we picked up the *ASXL1* mutation from residual blood present in the AVM sample. I encountered a similar issue in an HHT-associated pulmonary AVM where I identified a well-studied GOF mutation in *DNMT3A* (p.R882H, 57/1029 reads, 5.5% VAF). Similar to *ASXL1*, mutations in *DNMT3A* have not been reported in vascular malformations, but they are common in cases of clonal hematopoiesis (Buscarlet et al., 2017). This finding suggests mutations in blood cells could be a major confounder when searching for mutation in vascular malformations, and highlights the importance of paired blood sequencing.

These initial data suggest that while HHT-associated brain AVMs and sporadic brain AVMs have similar pathologies, they are genetically distinct. However, additional studies—especially of HHT-associated AVMs—will be required to fully understand their genetic underpinnings.

### 5.3 Syndromic Brain AVMs in the Amish

In 1997, members of the Marchuk lab collected blood samples from a large 12-generation kindred of the Amish of Lancaster Pennsylvania who develop syndromic brain AVMs. This kindred includes 6 individuals (2 female, 4 male, none are first-degree relatives) with a confirmed AVM in the brain or spinal cord; however, only 5 were available for sample collection. As the Amish are a small insular community,

it is possible that this AVM syndrome is caused by a rare ancestral mutation that has reached autozygosity in the affected individuals. As the Amish of Lancaster keep meticulous genealogical records, previous members of the lab were able to determine that the 6 affected individuals were born from pairings with kinship coefficients between 0.009 and 0.032. The inheritance pattern strongly suggests autosomal recessive inheritance; however, as with other genetic VM disorders it is likely that there is reduced penetrance. During sample collection, individuals filled out a survey with symptoms and several of the ‘unaffected’ individuals reported seizures and frequent headaches which are common symptoms of brain AVMs. As the Amish eschew health insurance, the only MRI available for these individuals are of low quality and were ultimately inconclusive.

Several previous members of the Marchuk lab have attempted to find the causal mutations in this kindred and there have been several mapping studies performed using SNP arrays. These mapping efforts revealed several putative regions of autozygosity on chromosomes 9 and 15 that were promising candidates, though a causal mutation was not found. To follow up on these previous efforts, I performed whole-exome sequencing on 4 of the affected individuals and whole-genome sequencing on 1 affected individual to validate the previously mapped regions of autozygosity and to identify a pathogenic variant. Due to the reduced penetrance, I only sequenced affected individuals as it was not possible to determine who is not affected. From these analyses, I determined 9 regions of autozygosity shared between all 5 individuals across chromosomes 2, 3, 11, 14, 19, and 20 (Figure 5.10). Notably, none of the previously identified regions were validated. I manually examined the previously determined regions on chromosomes 9 and 15 and found that, while the markers used in the initial study were accurate, there were heterozygous variants in between markers that disrupted the putative regions of autozygosity.

I did not identify any putative pathogenic variants within the regions of au-



**FIGURE 5.10: Autozygosity in Amish with Syndromic Brain AVMs.**  
 Regions of presumed autozygosity (blue bars) in the autosomes and ChrX in 5 individuals with syndromic brain AVMs. The red \* shows regions of autozygosity that are present in all 5 individuals. Note that the data from sample 2585 is derived from whole-genome sequencing data and therefore has finer resolution than the other samples which are derived from whole-exome sequencing data. Samples 2597, 2585, 2575 and 2557 are from males and are presumed to be hemizygous across ChrX. Sites on ChrX that were determined to be heterozygous in these samples are likely the result of either sequencing errors or misalignments.

tozygosity that I identified. I also identified regions that were autozygous in only 4 of the 5 individuals and again failed to find any putative pathogenic variants in these regions. I broadened my search to no longer be restricted by regions of autozygosity, and again failed to find any likely pathogenic homozygous variants in these individuals. I also manually examined variants in the previously identified regions on chromosomes 9 and 15 and failed to find any likely pathogenic homozygous variants. To explore the possibility that there were two different variants in the same gene, I expanded my search to heterozygous variants, present at <1% population allele frequency in gnomAD, and filtered for nonsynonymous variants. I did not find any recurrently mutated genes shared between these individuals. Note that I did not find any mutations in *ENG*, *ACVRL1*, or *SMAD4*.

The variants I identified to this point have been single nucleotide variants, small insertions, or small deletions. These variants are readily detectable with traditional variant callers, however these variant callers are not able to detect structural variation. Calling structural variants from exome sequencing data is challenging due to sparse coverage, however I was able to identify structural variants in sample 2585 for which I performed whole-genome sequencing. This analysis was performed with Manta using whole-genome sequencing data from blood outgrowth endothelial cells from an individual with HHT as a negative control (Chen et al., 2016). From this analysis I identified 601 putative structural variants that were homozygous in sample 2585 and absent in the negative control. Resources for programmatic adjudication of structural variants are sparse, specifically aggregate population allele frequency data, therefore I manually analyzed each putative variant while discarding any putative variant that was present in the NCBI Curated Common Structural Variants study (NCBI study accession no: nstd186) or the Database of Genomic Variants (MacDonald et al., 2014). The majority of putative variants were either of poor quality or were present in the aforementioned databases; even so, there remain many high quality,

undocumented variants. Several of the remaining variants fall in regions that overlap known ENCODE elements, the boundaries of topologically associated domains, and non-coding RNAs, however the impact of these variants remains unclear. Future studies should focus on performing whole-genome sequencing of additional samples to narrow the list of putative variants, and perform functional validation of any promising candidates.

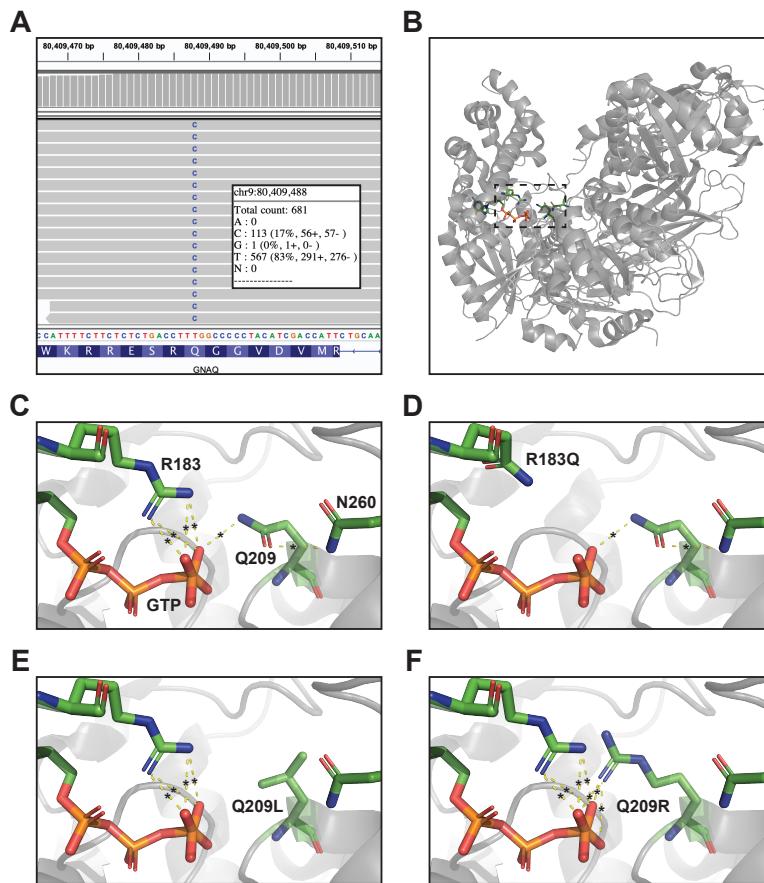
#### 5.4 Sturge-Weber Syndrome and Somatic Mutations in *GNAQ*

Sturge-Weber syndrome (SWS) is a sporadic vascular disease characterized by a mosaic capillary malformation (aka port-wine stain) present at birth and a leptomeningeal angioma in the brain. Unlike CCM and HHT, SWS is an exclusively sporadic disease, with no heritable component. A genetic study of SWS found that the capillary malformation and the leptomeningeal angioma are caused by a somatic GOF mutation in *GNAQ* (Shirley et al., 2013). Notably, this study also found that facial capillary malformations not associated with SWS are caused by the same mutation. The observation that SWS is due to a genetic mutation, yet is not heritable, suggests that GOF in *GNAQ* would not be viable when transmitted through the germline, and therefore SWS can only exist when the mutation occurs during postzygotic development—as was predicted by Rudolf Happle (Happle, 1987).

The genetics of *GNAQ* has captured my interest as it turns out that activating mutations are not only found in SWS, but also in circumscribed choroidal hemangioma (Le Guin et al., 2019), and uveal melanoma (a bona fide cancer) (Shoushtari & Carvajal, 2014). I found it curious that activation of the same gene could lead to these disparate phenotypes. More interesting yet, these diseases all have highly specific and distinct *GNAQ* mutational spectra: SWS is dominated by p.R183Q (although we recently found 1 case of a non-SWS capillary malformation with p.Q209R (Figure 5.11A)), circumscribed choroidal hemangioma is highly specific for p.Q209R,

whereas the vast majority of uveal melanomas have p.Q209L or p.Q209P with a minority having p.R183Q or p.Q209R. Based on this rigid genotype-phenotype correlation I hypothesized that—while all of these mutations activate *GNAQ*—they promote distinct modes of activation that result in different phenotypes.

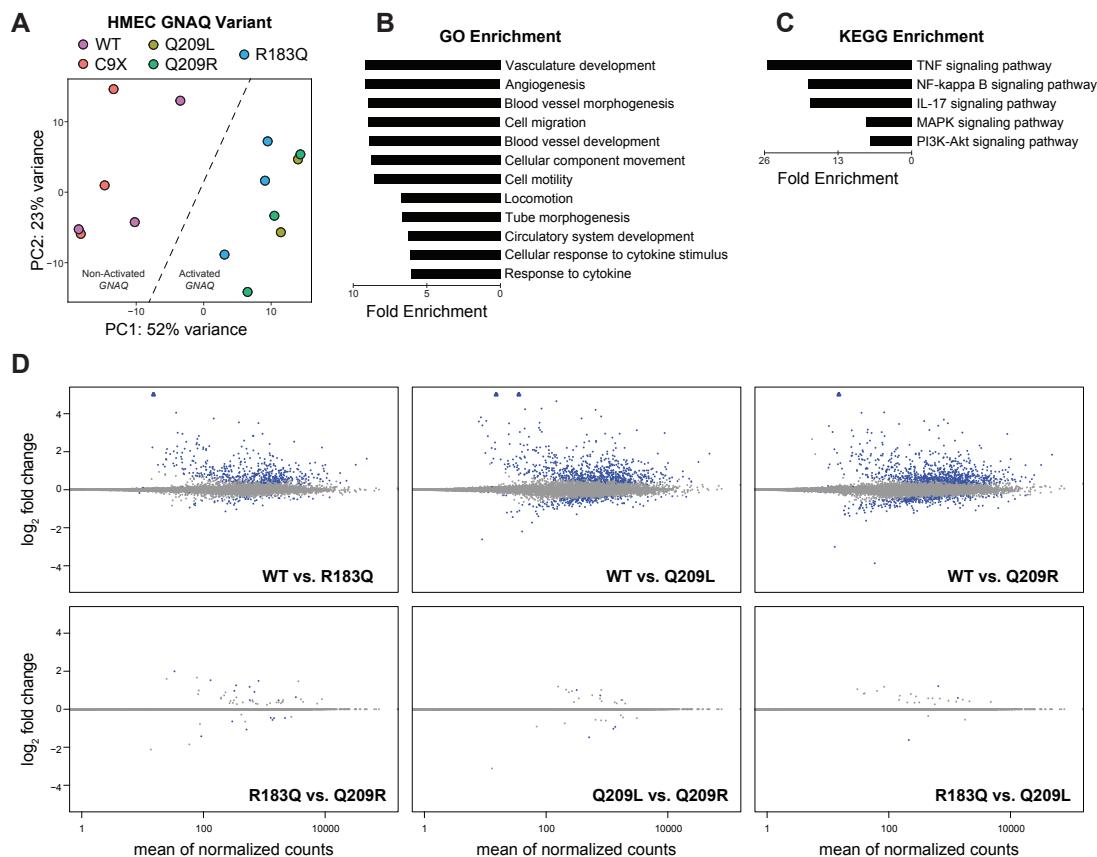
To test this hypothesis, Francesca Galeffi and I transfected human microvascular



**FIGURE 5.11: Structural Evaluation of Activating *GNAQ* Mutations.**

**A**, DNA sequencing reads from capillary malformation sample 3225 containing the p.Q209R mutation. **B**, Crystal structure of GNAQ (PDB: 3OHM). The catalytic site is by the dashed box. **C-F**, Zoom of the catalytic site with WT GNAQ or computational predictions of p.R183Q, p.Q209L, and p.Q209R. The yellow dashed lines with asterisks show predicted polar interactions with GTP. GNAQ was crystallized with GDP and a tetrafluoroaluminato ion ( $\text{AlF}_4^-$ ) in place of GTP. For the purposes of this analysis GDP +  $\text{AlF}_4^-$  was computationally mutated to GTP to determine polar interactions with GNAQ.

endothelial cells in culture with a plasmid containing 1 of 5 variants of *GNAQ*: wild type, p.C9X, p.R183Q, p.Q209R, or p.Q209L. We performed RNA-seq on these cultures in triplicate to determine if the different mutations in *GNAQ* were activating distinct transcriptional programs in endothelial cells. We found that each of the activated *GNAQ* plasmids produced extensive transcriptional changes relative



**FIGURE 5.12: Transcriptional Effects of Activating *GNAQ* Mutations.**

**A**, Principal component analysis of RNA-seq data produced from human microvascular endothelial cells transfected with WT or mutant *GNAQ* plasmids. **B-C**, Top enriched terms for gene ontology (B) and KEGG signaling pathway (C) using the top 100 differentially expressed genes between WT and p.Q209R groups. Enriched terms are sorted by fold enrichment and all terms shown have  $P < 1 \times 10^{-5}$ . **D**, MA plots of RNA-seq data for pairwise comparisons between WT and mutant *GNAQ* (top) and comparisons between mutations (bottom). Blue circles are statistically significant differentially expressed genes, blue triangles are significant differentially expressed genes off the scale, grey circles are not significant.

to the non-activated *GNAQ* plasmids (WT and p.C9X); however, there were few differentially expressed genes between any pairwise comparison of activated *GNAQ* plasmids (Figure 5.12).

This finding suggests that the different activating mutations in *GNAQ* produce very similar effects. Though each mutation has a similar transcriptional consequences, Francesca Galeffi has performed luciferase assays in cultured cells and determined that the strength of these mutations vary: p.R183Q is weakly activating, p.Q209R is moderately activating, and p.Q209L is highly activating. Notably, this finding is consistent with the predicted number of polar contacts with GTP in the catalytic site of *GNAQ* upon each mutation (Figure 5.11C-F). Why different activating mutations in *GNAQ* are highly specific in SWS, circumscribed choroidal hemangioma, and uveal melanoma remains an open question; however, each phenotype is likely defined by a combination of factors including the strength of the mutation, the timing of the mutation, and the cell type that acquires the mutation.

## 5.5 Somatic Mutations in Infantile Hemangioma

Infantile hemangioma (IH) are one of the most common vascular malformations. They occur in children and are typically present at birth as a red spot flush with the surrounding skin. Soon after birth the hemangioma rapidly grows and becomes raised from the skin. They are generally benign and are typically left alone unless they cover the child's mouth, nose, or eyes. What makes IH so interesting compared to other types of vascular malformations is that they almost always completely regress within the first few years of the child's life. While other types of vascular malformations may spontaneously regress (telangiectasia and AVMs), none do so with the consistency of IH. This phenomenon has been of great interest, not for the purpose of developing therapeutics for IH (propranolol is an extremely effective treatment for IH) but for uncovering the mechanism of regression in the hopes that what we learn can be

applied to regress other, more nefarious, vascular malformations.

### 5.5.1 GLUT1 in IH endothelium

Perhaps one of the most provocative discoveries into the mechanism of IH pathogenesis is the fact that endothelial cells from IHs highly express GLUT1 (North et al., 2000, 2001). GLUT1 is a glucose transporter that has remarkable specificity for the placental endothelium. This finding suggested that the IH may be comprised of cells that dislodged from the maternal placenta, then became hyper-proliferative in a post-fetal environment. If this hypothesis is correct, one would expect to find that the IH is a genetically chimeric growth between fetal and maternal cells. This hypothesis was put to the test using fluorescence *in situ* hybridization to assay the presence of XX cells in IH from a male infant with confirmation by sequencing microsatellites and SNPs that were divergent between mother and child. This analysis found no evidence for maternal-fetal chimerism in IH (Pittman et al., 2006). Despite this counter-evidence, the presence of GLUT1 in IH is strongly indicative of some link with the placenta though unfortunately this link currently remains elusive. The current literature has very clearly shown the presence of GLUT1 in IH endothelial cells, however the extent of the placental transcriptional program is unclear. Epigenomic profiling of paired IH and maternal placental samples may give valuable insights into the mechanism of IH.

### 5.5.2 Efforts to find somatic mutations in IH

As it is quickly becoming clear that the vast majority of vascular malformations are the result of somatic mutations—many occurring in known oncogenes—I thought that somatic mutations may also underlie IH. To test this, I sequenced 61 IH lesions on an ‘oncopanel’ covering many genes that are highly mutated in cancers as well as several genes previously implicated in vascular malformations (*KRIT1*, *CCM2*,

*PDCD10*, *ACVRL1*, *ENG*, *SMAD4*, etc.) Unfortunately after filtering putative variants, no there were no variants with likely functional significance and that occurred in more than a single sample. I am aware of at least 1 other group that has attempted to identify somatic mutations in IH via whole-exome sequencing, however to date there are no known somatic mutations in IH. One important aspect of these studies that must be noted is that they are invariably focused on coding regions of the genome. Non-coding variants are more than capable of causing disease however discovery-focused sequencing studies often ignore non-coding regions both because of the cost of sequencing the entire genome to a depth sufficient to detect somatic mutations, and the challenges associated with functional analysis of non-coding variants. Further studies may find that somatic mutations do cause IH, but they occur in a region of the genome that is missed by the majority of sequencing studies.

## 5.6 CCM & Meningioma

Although the vascular lesions are the primary sequelae of familial CCM, many groups have noted an increased prevalence of meningioma in individuals with familial CCM—especially those with a mutation in *PDCD10* (Labauge et al., 2009; Riant et al., 2013; Garaci et al., 2015). In addition, we have previously been contacted by an individual whose child had a sporadic CCM that regrew into a meningioma. Unfortunately we were unable to acquire a tissue sample for genetic analysis, however this case along with the strong link between familial CCM and meningioma has fueled my interest in understanding the link between CCM and meningioma.

### 5.6.1 Kruppel-like factor 4 (*KLF4*)

*KLF4* is a transcription factor with key roles in the pathogenesis of both CCM and meningioma. In CCM, *KLF4* (along with *KLF2*) is a key player in CCM signaling that is upregulated by loss of the CCM complex or gain of function in

*MAP3K3* (Cuttano et al., 2016; Zhou et al., 2016). Indeed, overexpression of *KLF4* in mouse models leads to an aggressive CCM-like phenotype (Ren et al., 2021). In meningioma, mutations in *KLF4* are common and often co-occur with mutations in *TRAF7* (Reuss et al., 2013). In addition, *KLF4* mutated meningioma have been shown to be responsive to treatment with temsirolimus (von Spreckelsen et al., 2020)—a derivative of rapamycin (sirolimus) which has been shown to be effective in treating CCM in mice (Ren et al., 2021). The importance of *KLF4* in both CCM and meningioma suggests that dysregulation of *KLF4* may underlie the development of meningioma in individuals with familial CCM.

The majority of *KLF4* mutations found in the catalog of somatic mutations in cancer (COSMIC) result in p.K409Q (Figure 5.13A). This narrow spectrum of mutations suggests that the p.K409Q variant results in KLF4 gain of function. Notably, the p.K409Q mutation is highly specific to meningioma, occurring in very few other cancer types (Figure 5.13B). The KLF4 protein contains 3 DNA-binding zinc-finger domains that contribute to its activity as a transcription factor. The p.K409Q mutation occurs in the first of the three zinc-finger domains (Figure 5.13C). A structural study of KLF4 determined that the second two zinc-finger domains consistently bind DNA, however the first zinc-finger only sometimes participates in DNA binding (Schuetz et al., 2011). They determined that the second and third domains are required for site specificity, but the third "inhibits cryptic self-renewal and block of differentiation activity". This immediately suggests a mechanism by which the p.K409Q mutation may disrupt the inhibitory capacity of KLF4 while retaining its functions that are independent of the first zinc-finger domain.

I thought that since overexpression of *KLF4* is sufficient to cause CCM in mice, the p.K409Q mutation may cause some sporadic CCMs in humans. To test this, I developed a ddPCR assay to test for this mutation in 30 mutation-negative sporadic CCM, however I did not find evidence for this mutation in any of these samples. This

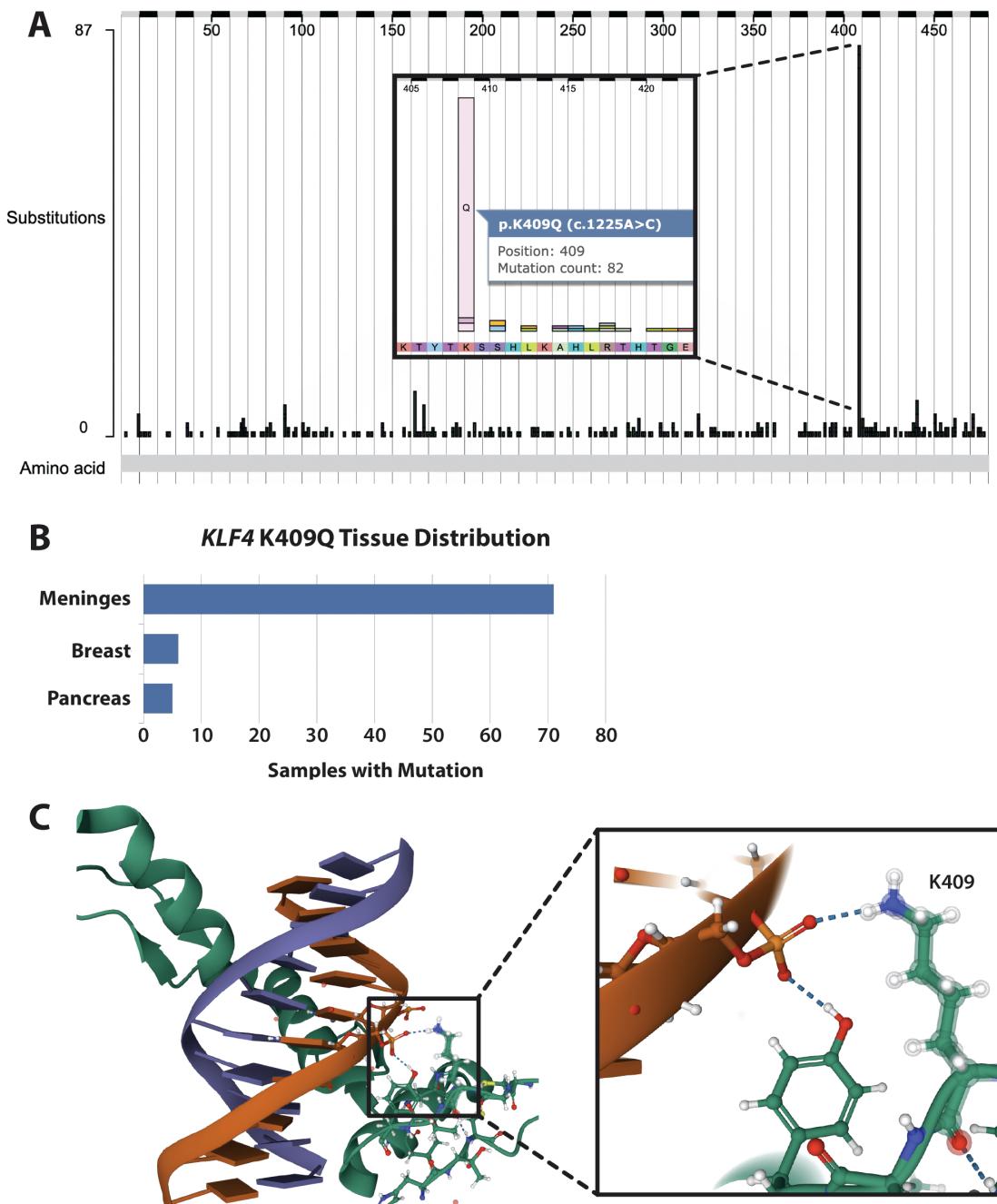


FIGURE 5.13: ***KLF4* Mutations in COSMIC.**

**A**, Distribution of somatic mutations in *KLF4* that are present in the catalog of somatic mutations in cancer (COSMIC). The location of the most frequent mutation (p.K409Q) is expanded in the inset. **B**, Distribution of *KLF4* p.K409Q in tissue types present in COSMIC. **C**, Structure of *KLF4* zinc-finger domains bound to DNA (PDB: 2WBU). Border box denotes the expanded region to the right showing the interaction between K409 and the DNA backbone.

may be a reflection of the highly specific functions of the first zinc-finger domain in *KLF4* as discussed above. This function may be critical for meningioma development, but less important for CCM development. However, the inverse is not necessarily true—i.e. while the p.K409Q mutation may not cause CCMs, overexpression of *KLF4* may cause meningioma. This is currently unknown, though if true, may account for the presence of meningioma in familial CCM. One straightforward way to test this hypothesis would be to sequence tissue from meningioma in individuals with familial CCM. I suspect that these meningioma harbor a somatic loss of function mutation in *KRIT1/CCM2/PDCD10* leading to biallelic loss of a CCM gene resulting in upregulation of *KLF4*. These meningioma may also harbor a second mutation in *TRAF7* which have often been found in non-CCM meningioma. Unfortunately, tissue samples of meningioma from individuals with familial CCM have been difficult to acquire, thus I have been unable to test this hypothesis. However, I expect that future studies will find that *KLF4* is a critical link between CCM and meningioma.

# **Appendix A**

Probability of Somatic Mutations

When describing my results to colleagues I am often met with the questions: “How can so many somatic mutations be occurring? Could these lesions have an elevated mutation rate?”. Though our intuition says it cannot be possible, I maintain that the somatic mutations I find in telangiectasia are the result of random chance via normal mutagenic processes. Unfortunately, too little is known about the rates of mutagenesis across the genome and in different tissues to accurately quantify the probability of these events. Despite this limitation, in this Appendix I attempt to conservatively estimate the probability of the events I describe in Chapter 2. The goal of this exercise is solely to highlight the disconnect between intuition and reality by showing that the empirically determined rate of disease is of similar magnitude to my conservative estimations.

## A.1 Two-Hit Mutations in HHT

In this section we consider the probability of biallelic loss of function (LOF) occurring in a single cell of an individual with HHT. For this exercise we will assume the individual has a germline heterozygous mutation in *ENG*. The probability of a somatic mutation resulting in biallelic LOF in a single cell ( $ENG_{sc}^{-/-}$ ) is a function of the somatic mutation rate ( $\mu_{som}$ ), the probability a mutation falls in the CDS of *ENG* ( $ENG_{CDS}$ ), results in LOF, and is in *trans* with the germline mutation (0.5 for diploid organisms) per somatic single-nucleotide variant (sSNV) such that

$$P(ENG_{sc}^{-/-}) = \mu_{som} \cdot \frac{P(ENG_{CDS}) \cdot P(LOF) \cdot 0.5}{sSNV}$$

Empirical data for  $\mu_{som}$  is not available for endothelial cells, however  $\mu_{som}$  has been evaluated for neurons via single-cell whole-genome sequencing (Lodato et al., 2018) and was determined to be roughly 40 sSNV per year of life ( $sSNV/year$ ). As mutation rates are tightly linked to DNA synthesis, the values from neurons should suffice as a conservative estimate of the  $\mu_{som}$  for endothelial cells.

Assuming that the rate of somatic mutations in *ENG* matches the genome-wide average, then

$$P(ENG_{CDS}) = \frac{ENG\ CDS\ Length}{Human\ Genome\ Size} = \frac{1977bp}{3.23 \times 10^9bp} = 6.1 \times 10^{-7}$$

Determining whether a given mutation may result in LOF is difficult as the effects of missense and silent variants are hard to predict. Therefore here we will only consider nonsense mutations, which will almost always result in LOF. Of the 5931 possible sSNVs in the CDS of *ENG*, 507 of them result in the creation of a premature stop codon (see <https://github.com/dasnellings/lofprob> for code). Therefore,

$$P(LOF) = \frac{507}{5931} = 0.085$$

Replacing these values in the original equation yields

$$P(ENG_{sc}^{-/-}) = \frac{40\ sSNV}{year} \cdot \frac{6.1 \times 10^{-7} \cdot 0.085 \cdot 0.5}{sSNV} = \frac{1.0 \times 10^{-6}}{year}$$

So every year, there is a 1 in 1,000,000 chance of biallelic LOF in *ENG* per cell. While this is a very low probability, consider that there are an estimated  $1 \times 10^{12}$  endothelial cells in an adult human (Jaffe, 1987). This suggests that 1,000,000 endothelial cells in an adult with HHT acquire biallelic LOF in *ENG* every year. Although the values I used for mutation rate and probability of LOF are almost certainly underestimates, the result is still many orders of magnitude higher than what we may expect. The most aggressive cases of HHT have hundreds of telangiectasia—not millions. This discrepancy likely reflects the fact that not all endothelial cells have the capacity to form an AVM, and likely only when conditions are right. Nonetheless, this demonstrates that the mutation itself is not as unlikely as it may initially appear.

# Bibliography

- Abdalla, S. A., Cymerman, U., Johnson, R. M., Deber, C. M., & Letarte, M. (2003). Disease-associated mutations in conserved residues of ALK-1 kinase domain. *Eur J Hum Genet*, 11(4), 279–87.  
PMID 12700602
- Abdalla, S. A., Pece-Barbara, N., Vera, S., Tapia, E., Paez, E., Bernabeu, C., & Letarte, M. (2000). Analysis of ALK-1 and endoglin in newborns from families with hereditary hemorrhagic telangiectasia type 2. *Hum Mol Genet*, 9(8), 1227–37.  
PMID 10767348
- Abdulrauf, S. I., Kaynar, M. Y., & Awad, I. A. (1999). A comparison of the clinical profile of cavernous malformations with and without associated venous malformations. *Neurosurgery*, 44(1), 41–6; discussion 46–7.  
PMID 9894962
- Akers, A., Al-Shahi Salman, R., I, A. A., Dahlem, K., Flemming, K., Hart, B., Kim, H., Jusue-Torres, I., Kondziolka, D., Lee, C., Morrison, L., Rigamonti, D., Rebeiz, T., Tournier-Lasserve, E., Waggoner, D., & Whitehead, K. (2017). Synopsis of Guidelines for the Clinical Management of Cerebral Cavernous Malformations: Consensus Recommendations Based on Systematic Literature Review by the Angioma Alliance Scientific Advisory Board Clinical Experts Panel. *Neurosurgery*, 80(5), 665–680.  
PMID 28387823
- Akers, A. L., Johnson, E., Steinberg, G. K., Zabramski, J. M., & Marchuk, D. A. (2009). Biallelic somatic and germline mutations in cerebral cavernous malformations (CCMs): evidence for a two-hit mechanism of CCM pathogenesis. *Hum Mol Genet*, 18(5), 919–30.  
PMID 19088123
- Al-Olabi, L., Polubothu, S., Dowsett, K., Andrews, K. A., Stadnik, P., Joseph, A. P., Knox, R., Pittman, A., Clark, G., Baird, W., Bulstrode, N., Glover, M., Gordon, K., Hargrave, D., Huson, S. M., Jacques, T. S., James, G., Kondolf, H., Kangesu, L., Keppler-Noreuil, K. M., Khan, A., Lindhurst, M. J., Lipson, M., Mansour,

S., O'Hara, J., Mahon, C., Mosica, A., Moss, C., Murthy, A., Ong, J., Parker, V. E., Riviere, J. B., Sapp, J. C., Sebire, N. J., Shah, R., Sivakumar, B., Thomas, A., Virasami, A., Waelchli, R., Zeng, Z., Biesecker, L. G., Barnacle, A., Topf, M., Semple, R. K., Patton, E. E., & Kinsler, V. A. (2018). Mosaic RAS/MAPK variants cause sporadic vascular malformations which respond to targeted therapy. *J Clin Invest*, *128*(4), 1496–1508.

PMID 29461977

Al-Qattan, M. M., Al-Balwi, M. A., Al-Zayed, E. M., Al-Sohaibani, M., Gelidan, A. G., & Alsheiban, S. (2020). Late-onset multiple venous malformations confined to the upper limb: link to somatic MAP3K3 mutations. *J Hand Surg Eur Vol*, *45*(10), 1023–1027.

PMID 32380920

Al-Shahi Salman, R., Hall, J. M., Horne, M. A., Moultrie, F., Josephson, C. B., Bhattacharya, J. J., Counsell, C. E., Murray, G. D., Papanastassiou, V., Ritchie, V., Roberts, R. C., Sellar, R. J., Warlow, C. P., & Scottish Audit of Intracranial Vascular Malformations, c. (2012). Untreated clinical course of cerebral cavernous malformations: a prospective, population-based cohort study. *Lancet Neurol*, *11*(3), 217–24.

PMID 22297119

Allinson, K. R., Carvalho, R. L., van den Brink, S., Mummary, C. L., & Arthur, H. M. (2007). Generation of a floxed allele of the mouse Endoglin gene. *Genesis*, *45*(6), 391–5.

PMID 17506087

Argyriou, L., Twelkemeyer, S., Panchulidze, I., Wehner, L. E., Teske, U., Engel, W., & Nayernia, K. (2006). Novel mutations in the ENG and ACVRL1 genes causing hereditary hemorrhagic telangiectasia. *Int J Mol Med*, *17*(4), 655–9.

PMID 16525724

Arthur, H. M., Ure, J., Smith, A. J., Renforth, G., Wilson, D. I., Torsney, E., Charlton, R., Parums, D. V., Jowett, T., Marchuk, D. A., Burn, J., & Diamond, A. G. (2000). Endoglin, an ancillary TGF $\beta$  receptor, is required for extraembryonic angiogenesis and plays a key role in heart development. *Dev Biol*, *217*(1), 42–53.

PMID 10625534

Awad, I. A., & Polster, S. P. (2019). Cavernous angiomas: deconstructing a neurosurgical disease. *J Neurosurg*, *131*(1), 1–13.

PMID 31261134

Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M. C., Kim, J., Reardon, B., Ng, P. K.,

Jeong, K. J., Cao, S., Wang, Z., Gao, J., Gao, Q., Wang, F., Liu, E. M., Mularoni, L., Rubio-Perez, C., Nagarajan, N., Cortes-Ciriano, I., Zhou, D. C., Liang, W. W., Hess, J. M., Yellapantula, V. D., Tamborero, D., Gonzalez-Perez, A., Suphavilai, C., Ko, J. Y., Khurana, E., Park, P. J., Van Allen, E. M., Liang, H., Group, M. C. W., Cancer Genome Atlas Research, N., Lawrence, M. S., Godzik, A., Lopez-Bigas, N., Stuart, J., Wheeler, D., Getz, G., Chen, K., Lazar, A. J., Mills, G. B., Karchin, R., & Ding, L. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*, 173(2), 371–385 e18.

PMID 29625053

Baumgartner, J. E., Ater, J. L., Ha, C. S., Kuttesch, J. F., Leeds, N. E., Fuller, G. N., & Wilson, R. J. (2003). Pathologically proven cavernous angiomas of the brain following radiation therapy for pediatric brain tumors. *Pediatr Neurosurg*, 39(4), 201–7.

PMID 12944701

Baylin, S. B. (2005). DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol*, 2 Suppl 1, S4–11.

PMID 16341240

Bayrak-Toydemir, P., McDonald, J., Akarsu, N., Toydemir, R. M., Calderon, F., Tuncali, T., Tang, W., Miller, F., & Mao, R. (2006). A fourth locus for hereditary hemorrhagic telangiectasia maps to chromosome 7. *Am J Med Genet A*, 140(20), 2155–62.

PMID 16969873

Bergametti, F., Denier, C., Labauge, P., Arnoult, M., Boetto, S., Clanet, M., Coubes, P., Echenne, B., Ibrahim, R., Irthum, B., Jacquet, G., Lonjon, M., Moreau, J. J., Neau, J. P., Parker, F., Tremoulet, M., Tournier-Lasserve, E., & Societe Francaise de, N. (2005). Mutations within the programmed cell death 10 gene cause cerebral cavernous malformations. *Am J Hum Genet*, 76(1), 42–51.

PMID 15543491

Bicknell, J. M., Carlow, T. J., Kornfeld, M., Stovring, J., & Turner, P. (1978). Familial cavernous angiomas. *Arch Neurol*, 35(11), 746–9.

PMID 718473

Bizzotto, S., Dou, Y., Ganz, J., Doan, R. N., Kwon, M., Bohrson, C. L., Kim, S. N., Bae, T., Abyzov, A., Network, N. B. S. M., Park, P. J., & Walsh, C. A. (2021). Landmarks of human embryonic development inscribed in somatic mutations. *Science*, 371(6535), 1249–1253.

PMID 33737485

Bossler, A. D., Richards, J., George, C., Godmilow, L., & Ganguly, A. (2006). Novel mutations in ENG and ACVRL1 identified in a series of 200 individuals

undergoing clinical genetic testing for hereditary hemorrhagic telangiectasia (HHT): correlation of genotype with phenotype. *Hum Mutat*, 27(7), 667–75.  
PMID 16752392

Boulday, G., Blecon, A., Petit, N., Chareyre, F., Garcia, L. A., Niwa-Kawakita, M., Giovannini, M., & Tournier-Lasserve, E. (2009). Tissue-specific conditional CCM2 knockout mice establish the essential role of endothelial CCM2 in angiogenesis: implications for human cerebral cavernous malformations. *Dis Model Mech*, 2(3-4), 168–77.  
PMID 19259391

Boulday, G., Rudini, N., Maddaluno, L., Blecon, A., Arnould, M., Gaudric, A., Chapon, F., Adams, R. H., Dejana, E., & Tournier-Lasserve, E. (2011). Developmental timing of CCM2 loss influences cerebral cavernous malformations in mice. *J Exp Med*, 208(9), 1835–47.  
PMID 21859843

Bourdeau, A., Cymerman, U., Paquet, M. E., Meschino, W., McKinnon, W. C., Guttmacher, A. E., Becker, L., & Letarte, M. (2000). Endoglin expression is reduced in normal vessels but still detectable in arteriovenous malformations of patients with hereditary hemorrhagic telangiectasia type 1. *Am J Pathol*, 156(3), 911–23.  
PMID 10702408

Bourdeau, A., Dumont, D. J., & Letarte, M. (1999). A murine model of hereditary hemorrhagic telangiectasia. *J Clin Invest*, 104(10), 1343–51.  
PMID 10562296

Brinjikji, W., El-Rida El-Masri, A., Wald, J. T., & Lanzino, G. (2017). Prevalence of Developmental Venous Anomalies Increases With Age. *Stroke*, 48(7), 1997–1999.  
PMID 28536179

Brinjikji, W., Hilditch, C. A., Tsang, A. C., Nicholson, P. J., Krings, T., & Agid, R. (2018). Facial Venous Malformations Are Associated with Cerebral Developmental Venous Anomalies. *AJNR Am J Neuroradiol*, 39(11), 2103–2107.  
PMID 30237297

Burn, S., Gunny, R., Phipps, K., Gaze, M., & Hayward, R. (2007). Incidence of cavernoma development in children after radiotherapy for brain tumors. *J Neurosurg*, 106(5 Suppl), 379–83.  
PMID 17566205

Buscarlet, M., Provost, S., Zada, Y. F., Barhdadi, A., Bourgoin, V., Lepine, G., Mollica, L., Szuber, N., Dube, M. P., & Busque, L. (2017). DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different

genetic predispositions. *Blood*, 130(6), 753–762.  
PMID 28655780

Castel, P., Carmona, F. J., Grego-Bessa, J., Berger, M. F., Viale, A., Anderson, K. V., Bague, S., Scaltriti, M., Antonescu, C. R., Baselga, E., & Baselga, J. (2016). Somatic PIK3CA mutations as a driver of sporadic venous malformations. *Sci Transl Med*, 8(332), 332ra42.  
PMID 27030594

Castillo, S. D., Baselga, E., & Graupera, M. (2019). PIK3CA mutations in vascular malformations. *Curr Opin Hematol*, 26(3), 170–178.  
PMID 30855339

Castillo, S. D., Tzouanacou, E., Zaw-Thin, M., Berenjeno, I. M., Parker, V. E., Chivite, I., Mila-Guasch, M., Pearce, W., Solomon, I., Angulo-Urarte, A., Figueiredo, A. M., Dewhurst, R. E., Knox, R. G., Clark, G. R., Scudamore, C. L., Badar, A., Kalber, T. L., Foster, J., Stuckey, D. J., David, A. L., Phillips, W. A., Lythgoe, M. F., Wilson, V., Semple, R. K., Sebire, N. J., Kinsler, V. A., Graupera, M., & Vanhaesebroeck, B. (2016). Somatic activating mutations in Pik3ca cause sporadic venous malformations in mice and humans. *Sci Transl Med*, 8(332), 332ra43.

PMID 27030595

Cau, M., Loi, M., Melis, M., Congiu, R., Loi, A., Meloni, C., Serrenti, M., Addis, M., & Melis, M. A. (2009). C329X in KRIT1 is a founder mutation among CCM patients in Sardinia. *Eur J Med Genet*, 52(5), 344–8.  
PMID 19454328

Cavalcanti, D. D., Kalani, M. Y., Martirosyan, N. L., Eales, J., Spetzler, R. F., & Preul, M. C. (2012). Cerebral cavernous malformations: from genes to proteins to disease. *J Neurosurg*, 116(1), 122–32.  
PMID 21962164

Chen, W., Guo, Y., Walker, E. J., Shen, F., Jun, K., Oh, S. P., Degos, V., Lawton, M. T., Tihan, T., Davalos, D., Akassoglou, K., Nelson, J., Pile-Spellman, J., Su, H., & Young, W. L. (2013). Reduced mural cell coverage and impaired vessel integrity after angiogenic stimulation in the Alk1-deficient brain. *Arterioscler Thromb Vasc Biol*, 33(2), 305–10.  
PMID 23241407

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Kallberg, M., Cox, A. J., Kruglyak, S., & Saunders, C. T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, 32(8), 1220–2.  
PMID 26647377

- Choi, E. J., Chen, W., Jun, K., Arthur, H. M., Young, W. L., & Su, H. (2014). Novel brain arteriovenous malformation mouse models for type 1 hereditary hemorrhagic telangiectasia. *PLoS One*, 9(2), e88511.  
PMID 24520391
- Choi, E. J., Walker, E. J., Shen, F., Oh, S. P., Arthur, H. M., Young, W. L., & Su, H. (2012). Minimal homozygous endothelial deletion of Eng with VEGF stimulation is sufficient to cause cerebrovascular dysplasia in the adult mouse. *Cerebrovasc Dis*, 33(6), 540–7.  
PMID 22571958
- Choquet, H., Pawlikowska, L., Nelson, J., McCulloch, C. E., Akers, A., Baca, B., Khan, Y., Hart, B., Morrison, L., Kim, H., & Brain Vascular Malformation Consortium, S. (2014). Polymorphisms in inflammatory and immune response genes associated with cerebral cavernous malformation type 1 severity. *Cerebrovasc Dis*, 38(6), 433–40.  
PMID 25472749
- Church, D. M., Stotler, C. J., Rutter, J. L., Murrell, J. R., Trofatter, J. A., & Buckler, A. J. (1994). Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nat Genet*, 6(1), 98–105.  
PMID 8136842
- Clark, J. V. (1970). Familial occurrence of cavernous angioma of the brain. *J Neurol Neurosurg Psychiatry*, 33(6), 871–6.  
PMID 5531907
- Cole, S. G., Begbie, M. E., Wallace, G. M., & Shovlin, C. L. (2005). A new locus for hereditary haemorrhagic telangiectasia (HHT3) maps to chromosome 5. *J Med Genet*, 42(7), 577–82.  
PMID 15994879
- Couto, J. A., Huang, A. Y., Konczyk, D. J., Goss, J. A., Fishman, S. J., Mulliken, J. B., Warman, M. L., & Greene, A. K. (2017). Somatic MAP2K1 Mutations Are Associated with Extracranial Arteriovenous Malformation. *Am J Hum Genet*, 100(3), 546–554.  
PMID 28190454
- Couto, J. A., Vivero, M. P., Kozakewich, H. P., Taghinia, A. H., Mulliken, J. B., Warman, M. L., & Greene, A. K. (2015). A somatic MAP3K3 mutation is associated with verrucous venous malformation. *Am J Hum Genet*, 96(3), 480–6.  
PMID 25728774
- Cutsforth-Gregory, J. K., Lanzino, G., Link, M. J., Brown, J., R. D., & Flemming, K. D. (2015). Characterization of radiation-induced cavernous malformations and

comparison with a nonradiation cavernous malformation cohort. *J Neurosurg*, 122(5), 1214–22.  
PMID 25699412

Cuttano, R., Rudini, N., Bravi, L., Corada, M., Giampietro, C., Papa, E., Morini, M. F., Maddaluno, L., Baeyens, N., Adams, R. H., Jain, M. K., Owens, G. K., Schwartz, M., Lampugnani, M. G., & Dejana, E. (2016). KLF4 is a key determinant in the development and progression of cerebral cavernous malformations. *EMBO Mol Med*, 8(1), 6–24.  
PMID 26612856

Dammann, P., Wrede, K., Zhu, Y., Matsushige, T., Maderwald, S., Umutlu, L., Quick, H. H., Hehr, U., Rath, M., Ladd, M. E., Felbor, U., & Sure, U. (2017). Correlation of the venous angioarchitecture of multiple cerebral cavernous malformations with familial or sporadic disease: a susceptibility-weighted imaging study with 7-Tesla MRI. *J Neurosurg*, 126(2), 570–577.

PMID 27153162

Davis, S., Ware, M. A., Zeiger, J., Deardorff, M. A., Grand, K., Grimberg, A., Hsu, S., Kelsey, M., Majidi, S., Matthew, R. P., Napier, M., Nokoff, N., Prasad, C., Riggs, A. C., McKinnon, M. L., & Mirzaa, G. (2020). Growth hormone deficiency in megalencephaly-capillary malformation syndrome: An association with activating mutations in PIK3CA. *Am J Med Genet A*, 182(1), 162–168.  
PMID 31729162

Denier, C., Labauge, P., Bergametti, F., Marchelli, F., Riant, F., Arnoult, M., Maciazeck, J., Vicaut, E., Brunereau, L., Tournier-Lasserve, E., & Societe Francaise de, N. (2006). Genotype-phenotype correlations in cerebral cavernous malformations patients. *Ann Neurol*, 60(5), 550–6.  
PMID 17041941

Denier, C., Labauge, P., Brunereau, L., Cave-Riant, F., Marchelli, F., Arnoult, M., Cecillon, M., Maciazeck, J., Joutel, A., Tournier-Lasserve, E., Societe Francaise de, N., & Societe de Neurochirurgie de Langue, F. (2004). Clinical features of cerebral cavernous malformations patients with KRIT1 mutations. *Ann Neurol*, 55(2), 213–20.  
PMID 14755725

Desmet, F. O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., & Beroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*, 37(9), e67.  
PMID 19339519

Detter, M. R., Shenkar, R., Benavides, C. R., Neilson, C. A., Moore, T., Lightle, R., Hobson, N., Shen, L., Cao, Y., Girard, R., Zhang, D., Griffin, E., Gallione,

C. J., Awad, I. A., & Marchuk, D. A. (2020). Novel Murine Models of Cerebral Cavernous Malformations. *Angiogenesis*, 23(4), 651–666.  
PMID 32710309

Detter, M. R., Snellings, D. A., & Marchuk, D. A. (2018). Cerebral Cavernous Malformations Develop Through Clonal Expansion of Mutant Endothelial Cells. *Circ Res*, 123(10), 1143–1151.  
PMID 30359189

Dhamija, R., Weindling, S. M., Porter, A. B., Hu, L. S., Wood, C. P., & Hoxworth, J. M. (2018). Neuroimaging abnormalities in patients with Cowden syndrome: Retrospective single-center study. *Neurol Clin Pract*, 8(3), 207–213.  
PMID 30105160

Dogruluk, T., Tsang, Y. H., Espitia, M., Chen, F., Chen, T., Chong, Z., Appadurai, V., Dogruluk, A., Eterovic, A. K., Bonnen, P. E., Creighton, C. J., Chen, K., Mills, G. B., & Scott, K. L. (2015). Identification of Variant-Specific Functions of PIK3CA by Rapid Phenotyping of Rare Mutations. *Cancer Res*, 75(24), 5341–54.  
PMID 26627007

Eerola, I., Plate, K. H., Spiegel, R., Boon, L. M., Mulliken, J. B., & Vakkula, M. (2000). KRIT1 is mutated in hyperkeratotic cutaneous capillary-venous malformation associated with cerebral capillary malformation. *Hum Mol Genet*, 9(9), 1351–5.  
PMID 10814716

Eleftheriou, N. M., Sjolund, J., Bocci, M., Cortez, E., Lee, S. J., Cunha, S. I., & Pietras, K. (2016). Compound genetically engineered mouse models of cancer reveal dual targeting of ALK1 and endoglin as a synergistic opportunity to impinge on angiogenic TGF-beta signaling. *Oncotarget*, 7(51), 84314–84325.  
PMID 27741515

Faurobert, E., Rome, C., Lisowska, J., Manet-Dupe, S., Boulday, G., Malbouyres, M., Balland, M., Bouin, A. P., Keramidas, M., Bouvard, D., Coll, J. L., Ruggiero, F., Tournier-Lasserve, E., & Albiges-Rizo, C. (2013). CCM1-ICAP-1 complex controls beta1 integrin-dependent endothelial contractility and fibronectin remodeling. *J Cell Biol*, 202(3), 545–61.  
PMID 23918940

Fauth, C., Rostasy, K., Rath, M., Gizewski, E., Lederer, A. G., Sure, U., Zschocke, J., & Felbor, U. (2015). Highly variable intrafamilial manifestations of a CCM3 mutation ranging from acute childhood cerebral haemorrhage to late-onset meningiomas. *Clin Neurol Neurosurg*, 128, 41–3.  
PMID 25462093

- Fischer, A., Zalvide, J., Faurobert, E., Albiges-Rizo, C., & Tournier-Lasserve, E. (2013). Cerebral cavernous malformations: from CCM genes to endothelial cell homeostasis. *Trends Mol Med*, 19(5), 302–8.  
PMID 23506982
- Fisher, O. S., & Boggon, T. J. (2014). Signaling pathways and the cerebral cavernous malformations proteins: lessons from structural biology. *Cell Mol Life Sci*, 71(10), 1881–92.  
PMID 24287896
- Francis, J. H., Milman, T., Grossniklaus, H., Albert, D., Folberg, R., Levitin, G., Coupland, S., Catalanotti, F., Rabady, D., Kandoth, C., Busam, K., & Abramson, D. (2019). GNAQ Mutations in Diffuse and Solitary Choroidal Hemangiomas. *Ophthalmology*, 126(5), 759–763.  
PMID 30537484
- Fujino, T., & Kitamura, T. (2020). ASXL1 mutation in clonal hematopoiesis. *Exp Hematol*, 83, 74–84.  
PMID 31945396
- Gallione, C. J., Detter, M. R., Christmas, H. M., & Marchuk, D. A. (2022). Genetic Genealogy Uncovers a Founder Deletion Mutation in the Cerebral Cavernous Malformations 2 Gene. *Research Square*
- Gallione, C. J., Klaus, D. J., Yeh, E. Y., Stenzel, T. T., Xue, Y., Anthony, K. B., McAllister, K. A., Baldwin, M. A., Berg, J. N., Lux, A., Smith, J. D., Vary, C. P., Craigen, W. J., Westermann, C. J., Warner, M. L., Miller, Y. E., Jackson, C. E., Guttmacher, A. E., & Marchuk, D. A. (1998). Mutation and expression analysis of the endoglin gene in hereditary hemorrhagic telangiectasia reveals null alleles. *Hum Mutat*, 11(4), 286–94.  
PMID 9554745
- Gallione, C. J., Repetto, G. M., Legius, E., Rustgi, A. K., Schelley, S. L., Tejpar, S., Mitchell, G., Drouin, E., Westermann, C. J., & Marchuk, D. A. (2004). A combined syndrome of juvenile polyposis and hereditary haemorrhagic telangiectasia associated with mutations in MADH4 (SMAD4). *Lancet*, 363(9412), 852–9.  
PMID 15031030
- Gallione, C. J., Solatycki, A., Awad, I. A., Weber, J. L., & Marchuk, D. A. (2011). A founder mutation in the Ashkenazi Jewish population affecting messenger RNA splicing of the CCM2 gene causes cerebral cavernous malformations. *Genet Med*, 13(7), 662–6.  
PMID 21543988

- Garaci, F., Marsili, L., Riant, F., Marziali, S., Cecillon, M., Pasquarelli, R., Sangiuolo, F., Floris, R., Novelli, G., Tournier-Lasserve, E., & Brancati, F. (2015). Cerebral cavernous malformations associated to meningioma: High penetrance in a novel family mutated in the *PDCD10* gene. *Neuroradiol J*, 28(3), 289–93. PMID 26246098
- Garrido-Martin, E. M., Nguyen, H. L., Cunningham, T. A., Choe, S. W., Jiang, Z., Arthur, H. M., Lee, Y. J., & Oh, S. P. (2014). Common and distinctive pathogenetic features of arteriovenous malformations in hereditary hemorrhagic telangiectasia 1 and hereditary hemorrhagic telangiectasia 2 animal models—brief report. *Arterioscler Thromb Vasc Biol*, 34(10), 2232–6. PMID 25082229
- Gault, J., Awad, I. A., Recksiek, P., Shenkar, R., Breeze, R., Handler, M., & Kleinschmidt-DeMasters, B. K. (2009). Cerebral cavernous malformations: somatic mutations in vascular endothelial cells. *Neurosurgery*, 65(1), 138–44; discussion 144–5. PMID 19574835
- Gault, J., Sain, S., Hu, L. J., & Awad, I. A. (2006). Spectrum of genotype and clinical manifestations in cerebral cavernous malformations. *Neurosurgery*, 59(6), 1278–84; discussion 1284–5. PMID 17277691
- Gault, J., Shenkar, R., Recksiek, P., & Awad, I. A. (2005). Biallelic somatic and germ line CCM1 truncating mutations in a cerebral cavernous malformation lesion. *Stroke*, 36(4), 872–4. PMID 15718512
- Gianfrancesco, F., Cannella, M., Martino, T., Maglione, V., Esposito, T., Innocenzi, G., Vitale, E., Liquori, C. L., Marchuk, D. A., & Squitieri, F. (2007). Highly variable penetrance in subjects affected with cavernous cerebral angiomas (CCM) carrying novel CCM1 and CCM2 mutations. *Am J Med Genet B Neuropsychiatr Genet*, 144B(5), 691–5. PMID 17440989
- Girard, R., Zeineddine, H. A., Koskimaki, J., Fam, M. D., Cao, Y., Shi, C., Moore, T., Lightle, R., Stadnik, A., Chaudagar, K., Polster, S., Shenkar, R., Duggan, R., Leclerc, D., Whitehead, K. J., Li, D. Y., & Awad, I. A. (2018). Plasma Biomarkers of Inflammation and Angiogenesis Predict Cerebral Cavernous Malformation Symptomatic Hemorrhage or Lesional Growth. *Circ Res*, 122(12), 1716–1721. PMID 29720384
- Gokce, E., Acu, B., Beyhan, M., Celikyay, F., & Celikyay, R. (2014). Magnetic resonance imaging findings of developmental venous anomalies. *Clin Neuroradiol*,

24(2), 135–43.  
PMID 24240482

Gonzalez, C. D., Cipriano, S. D., Topham, C. A., Stevenson, D. A., Whitehead, K. J., Vanderhooft, S., Presson, A. P., & McDonald, J. (2019). Localization and age distribution of telangiectases in children and adolescents with hereditary hemorrhagic telangiectasia: A retrospective cohort study. *J Am Acad Dermatol.* PMID 30819528

Goss, J. A., Huang, A. Y., Smith, E., Konczyk, D. J., Smits, P. J., Sudduth, C. L., Stapleton, C., Patel, A., Alexandrescu, S., Warman, M. L., & Greene, A. K. (2019). Somatic mutations in intracranial arteriovenous malformations. *PLoS One*, 14(12), e0226852.  
PMID 31891627

Grippaudo, F. R., Piane, M., Amoroso, M., Longo, B., Penco, S., Chessa, L., Giubettini, M., & Santanelli, F. (2013). Cutaneous venous malformations related to KRIT1 mutation: case report and literature review. *J Mol Neurosci*, 51(2), 442–5.  
PMID 23828392

Grosse, S. D., Boulet, S. L., Grant, A. M., Hulihan, M. M., & Faughnan, M. E. (2014). The use of US health insurance data for surveillance of rare disorders: hereditary hemorrhagic telangiectasia. *Genet Med*, 16(1), 33–9.  
PMID 23703685

Gunel, M., Awad, I. A., Finberg, K., Anson, J. A., Steinberg, G. K., Batjer, H. H., Kopitnik, T. A., Morrison, L., Giannotta, S. L., Nelson-Williams, C., & Lifton, R. P. (1996). A founder mutation as a cause of cerebral cavernous malformation in Hispanic Americans. *N Engl J Med*, 334(15), 946–51.  
PMID 8596595

Happle, R. (1987). Lethal genes surviving by mosaicism: a possible explanation for sporadic birth defects involving the skin. *J Am Acad Dermatol*, 16(4), 899–906.  
PMID 3033033

Heckl, S., Aschoff, A., & Kunze, S. (2002). Radiation-induced cavernous hemangiomas of the brain: a late effect predominantly in children. *Cancer*, 94(12), 3285–91.  
PMID 12115362

Heiskanen, O. (1993). Treatment of spontaneous intracerebral and intracerebellar hemorrhages. *Stroke*, 24(12 Suppl), I94–5; discussion I107–8.  
PMID 8249028

Hernandez, F., Huether, R., Carter, L., Johnston, T., Thompson, J., Gossage, J. R., Chao, E., & Elliott, A. M. (2015). Mutations in RASA1 and GDF2 identified in patients with clinical features of hereditary hemorrhagic telangiectasia. *Hum Genome Var*, 2, 15040.

PMID 27081547

Hong, T., Xiao, X., Ren, J., Cui, B., Zong, Y., Zou, J., Kou, Z., Jiang, N., Meng, G., Zeng, G., Shan, Y., Wu, H., Chen, Z., Liang, J., Xiao, X., Tang, J., Wei, Y., Ye, M., Sun, L., Li, G., Hu, P., Hui, R., Zhang, H., & Wang, Y. (2021). Somatic MAP3K3 and PIK3CA mutations in sporadic cerebral and spinal cord cavernous malformations. *Brain*.

PMID 33729480

Horne, M. A., Flemming, K. D., Su, I. C., Stapf, C., Jeon, J. P., Li, D., Maxwell, S. S., White, P., Christianson, T. J., Agid, R., Cho, W. S., Oh, C. W., Wu, Z., Zhang, J. T., Kim, J. E., Ter Brugge, K., Willinsky, R., Brown, J., R. D., Murray, G. D., Al-Shahi Salman, R., & Cerebral Cavernous Malformations Individual Patient Data Meta-analysis, C. (2016). Clinical course of untreated cerebral cavernous malformations: a meta-analysis of individual patient data. *Lancet Neurol*, 15(2), 166–173.

PMID 26654287

Hurst, C. D., Zuiverloon, T. C., Hafner, C., Zwarthoff, E. C., & Knowles, M. A. (2009). A SNaPshot assay for the rapid and simple detection of four common hotspot codon mutations in the PIK3CA gene. *BMC Res Notes*, 2, 66.

PMID 19402901

Jaffe, E. A. (1987). Cell biology of endothelial cells. *Hum Pathol*, 18(3), 234–9.

PMID 3546072

Jain, R., Robertson, P. L., Gandhi, D., Gujar, S. K., Muraszko, K. M., & Gebarski, S. (2005). Radiation-induced cavernomas of the brain. *AJNR Am J Neuroradiol*, 26(5), 1158–62.

PMID 15891176

Jiao, W., Vembu, S., Deshwar, A. G., Stein, L., & Morris, Q. (2014). Inferring clonal evolution of tumors from single nucleotide somatic mutations. *BMC Bioinformatics*, 15, 35.

PMID 24484323

Jin, Y., Muhl, L., Burmakin, M., Wang, Y., Duchez, A. C., Betsholtz, C., Arthur, H. M., & Jakobsson, L. (2017). Endoglin prevents vascular malformation by regulating flow-induced cell migration and specification through VEGFR2 signalling. *Nat Cell Biol*, 19(6), 639–652.

PMID 28530660

- Johnson, D. W., Berg, J. N., Baldwin, M. A., Gallione, C. J., Marondel, I., Yoon, S. J., Stenzel, T. T., Speer, M., Pericak-Vance, M. A., Diamond, A., Guttmacher, A. E., Jackson, C. E., Attisano, L., Kucherlapati, R., Porteous, M. E., & Marchuk, D. A. (1996). Mutations in the activin receptor-like kinase 1 gene in hereditary haemorrhagic telangiectasia type 2. *Nat Genet*, 13(2), 189–95.  
PMID 8640225
- Johnson, D. W., Berg, J. N., Gallione, C. J., McAllister, K. A., Warner, J. P., Helmbold, E. A., Markel, D. S., Jackson, C. E., Porteous, M. E., & Marchuk, D. A. (1995). A second locus for hereditary hemorrhagic telangiectasia maps to chromosome 12. *Genome Res*, 5(1), 21–8.  
PMID 8717052
- Jolly, C., & Van Loo, P. (2018). Timing somatic events in the evolution of cancer. *Genome Biol*, 19(1), 95.  
PMID 30041675
- Jones, B. V., Linscott, L., Koberlein, G., Hummel, T. R., & Leach, J. L. (2015). Increased Prevalence of Developmental Venous Anomalies in Children with Intracranial Neoplasms. *AJNR Am J Neuroradiol*, 36(9), 1782–5.  
PMID 26021620
- Kamezawa, T., Hamada, J., Niiro, M., Kai, Y., Ishimaru, K., & Kuratsu, J. (2005). Clinical implications of associated venous drainage in patients with cavernous malformation. *J Neurosurg*, 102(1), 24–8.  
PMID 15658092
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res*, 49(D1), D545–D551.  
PMID 33125081
- Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1), 27–30.  
PMID 10592173
- Kidd, H. A., & Cumings, J. N. (1947). Cerebral angioma in an Icelandic family. *Lancet*, 1(6457), 747.  
PMID 20241165
- Kim, Y. H., Choe, S. W., Chae, M. Y., Hong, S., & Oh, S. P. (2018). SMAD4 Deficiency Leads to Development of Arteriovenous Malformations in Neonatal and Adult Mice. *J Am Heart Assoc*, 7(21), e009514.  
PMID 30571376

Koike, T., Yanagimachi, N., Ishiguro, H., Yabe, H., Yabe, M., Morimoto, T., Shimizu, T., Takakura, H., & Kato, S. (2012). High incidence of radiation-induced cavernous hemangioma in long-term survivors who underwent hematopoietic stem cell transplantation with radiation therapy during childhood or adolescence. *Biol Blood Marrow Transplant*, 18(7), 1090–8.

PMID 22198541

Koren, S., Reavie, L., Couto, J. P., De Silva, D., Stadler, M. B., Roloff, T., Britschgi, A., Eichlisberger, T., Kohler, H., Aina, O., Cardiff, R. D., & Bentires-Alj, M. (2015). PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. *Nature*, 525(7567), 114–8.

PMID 26266975

Koskimaki, J., Zhang, D., Li, Y., Saadat, L., Moore, T., Lightle, R., Polster, S. P., Carrion-Penagos, J., Lyne, S. B., Zeineddine, H. A., Shi, C., Shenkar, R., Romanos, S., Avner, K., Srinath, A., Shen, L., Detter, M. R., Snellings, D., Cao, Y., Lopez-Ramirez, M. A., Fonseca, G., Tang, A. T., Faber, P., Andrade, J., Ginsberg, M., Kahn, M. L., Marchuk, D. A., Girard, R., & Awad, I. A. (2019). Transcriptome clarifies mechanisms of lesion genesis versus progression in models of Ccm3 cerebral cavernous malformations. *Acta Neuropathol Commun*, 7(1), 132.

PMID 31426861

Labauge, P., Enjolras, O., Bonerandi, J. J., Laberge, S., Dandurand, M., Joujoux, J. M., & Tournier-Lasserve, E. (1999). An association between autosomal dominant cerebral cavernomas and a distinctive hyperkeratotic cutaneous vascular malformation in 4 families. *Ann Neurol*, 45(2), 250–4.

PMID 9989629

Labauge, P., Fontaine, B., Neau, J. P., Bergametti, F., Riant, F., Blecon, A., Marchelli, F., Arnoult, M., Lannuzel, A., Clanet, M., Olschwang, S., Denier, C., & Tournier-Lasserve, E. (2009). Multiple dural lesions mimicking meningiomas in patients with CCM3/PDCD10 mutations. *Neurology*, 72(23), 2044–6.

PMID 19506228

Laberge-le Couteulx, S., Jung, H. H., Labauge, P., Houtteville, J. P., Lescoat, C., Cecillon, M., Marechal, E., Joutel, A., Bach, J. F., & Tournier-Lasserve, E. (1999a). Truncating mutations in CCM1, encoding KRIT1, cause hereditary cavernous angiomas. *Nat Genet*, 23(2), 189–93.

PMID 10508515

Laberge-le Couteulx, S., Jung, H. H., Labauge, P., Houtteville, J. P., Lescoat, C., Cecillon, M., Marechal, E., Joutel, A., Bach, J. F., & Tournier-Lasserve, E. (1999b). Truncating mutations in CCM1, encoding KRIT1, cause hereditary cavernous angiomas. *Nat Genet*, 23(2), 189–93.

PMID 10508515

Lapinski, P. E., Doosti, A., Salato, V., North, P., Burrows, P. E., & King, P. D. (2018). Somatic second hit mutation of RASA1 in vascular endothelial cells in capillary malformation-arteriovenous malformation. *Eur J Med Genet*, 61(1), 11–16.

PMID 29024832

Laurans, M. S., DiLuna, M. L., Shin, D., Niazi, F., Voorhees, J. R., Nelson-Williams, C., Johnson, E. W., Siegel, A. M., Steinberg, G. K., Berg, M. J., Scott, R. M., Tedeschi, G., Enevoldson, T. P., Anson, J., Rouleau, G. A., Ogilvy, C., Awad, I. A., Lifton, R. P., & Gunel, M. (2003). Mutational analysis of 206 families with cavernous malformations. *J Neurosurg*, 99(1), 38–43.

PMID 12854741

Laver, T. W., Caswell, R. C., Moore, K. A., Poschmann, J., Johnson, M. B., Owens, M. M., Ellard, S., Paszkiewicz, K. H., & Weedon, M. N. (2016). Pitfalls of haplotype phasing from amplicon-based long-read sequencing. *Sci Rep*, 6, 21746.

PMID 26883533

Le Guin, C. H. D., Metz, K. A., Kreis, S. H., Bechrakis, N. E., Bornfeld, N., Zeschnigk, M., & Lohmann, D. R. (2019). GNAQ Q209R Mutations Are Highly Specific for Circumscribed Choroidal Hemangioma. *Cancers (Basel)*, 11(7).

PMID 31336681

Letteboer, T. G., Mager, H. J., Snijder, R. J., Lindhout, D., Ploos van Amstel, H. K., Zanen, P., & Westermann, K. J. (2008). Genotype-phenotype relationship for localization and age distribution of telangiectases in hereditary hemorrhagic telangiectasia. *Am J Med Genet A*, 146A(21), 2733–9.

PMID 18831062

Li, D. Y., Sorensen, L. K., Brooke, B. S., Urness, L. D., Davis, E. C., Taylor, D. G., Boak, B. B., & Wendel, D. P. (1999). Defective angiogenesis in mice lacking endoglin. *Science*, 284(5419), 1534–7.

PMID 10348742

Limaye, N., Kangas, J., Mendola, A., Godfraind, C., Schlogel, M. J., Helaers, R., Eklund, L., Boon, L. M., & Viikkula, M. (2015). Somatic Activating PIK3CA Mutations Cause Venous Malformation. *Am J Hum Genet*, 97(6), 914–21.

PMID 26637981

Limaye, N., Wouters, V., Uebelhoer, M., Tuominen, M., Wirkkala, R., Mulliken, J. B., Eklund, L., Boon, L. M., & Viikkula, M. (2009). Somatic mutations in angiopoietin receptor gene TEK cause solitary and multiple sporadic venous malformations. *Nat Genet*, 41(1), 118–24.

PMID 19079259

Linscott, L. L., Leach, J. L., Jones, B. V., & Abruzzo, T. A. (2016). Developmental venous anomalies of the brain in children – imaging spectrum and update. *Pediatr Radiol*, *46*(3), 394–406; quiz 391–3.  
PMID 26795616

Liquori, C. L., Berg, M. J., Siegel, A. M., Huang, E., Zawistowski, J. S., Stoffer, T., Verlaan, D., Balogun, F., Hughes, L., Leedom, T. P., Plummer, N. W., Cannella, M., Maglione, V., Squitieri, F., Johnson, E. W., Rouleau, G. A., Ptacek, L., & Marchuk, D. A. (2003). Mutations in a gene encoding a novel protein containing a phosphotyrosine-binding domain cause type 2 cerebral cavernous malformations. *Am J Hum Genet*, *73*(6), 1459–64.  
PMID 14624391

Liquori, C. L., Berg, M. J., Squitieri, F., Leedom, T. P., Ptacek, L., Johnson, E. W., & Marchuk, D. A. (2007). Deletions in CCM2 are a common cause of cerebral cavernous malformations. *Am J Hum Genet*, *80*(1), 69–75.  
PMID 17160895

Little, J. B. (1998). Radiation-induced genomic instability. *Int J Radiat Biol*, *74*(6), 663–71.  
PMID 9881710

Lo Sardo, V., Ferguson, W., Erikson, G. A., Topol, E. J., Baldwin, K. K., & Torkamani, A. (2017). Influence of donor age on induced pluripotent stem cells. *Nat Biotechnol*, *35*(1), 69–74.  
PMID 27941802

Lodato, M. A., Rodin, R. E., Bohrson, C. L., Coulter, M. E., Barton, A. R., Kwon, M., Sherman, M. A., Vitzthum, C. M., Luquette, L. J., Yandava, C. N., Yang, P., Chittenden, T. W., Hatem, N. E., Ryu, S. C., Woodworth, M. B., Park, P. J., & Walsh, C. A. (2018). Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science*, *359*(6375), 555–559.  
PMID 29217584

Lodato, M. A., Woodworth, M. B., Lee, S., Evrony, G. D., Mehta, B. K., Karger, A., Lee, S., Chittenden, T. W., D'Gama, A. M., Cai, X., Luquette, L. J., Lee, E., Park, P. J., & Walsh, C. A. (2015). Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science*, *350*(6256), 94–98.  
PMID 26430121

Lopez-Ramirez, M. A., Pham, A., Girard, R., Wyseure, T., Hale, P., Yamashita, A., Koskimaki, J., Polster, S., Saadat, L., Romero, I. A., Esmon, C. T., Lagarrigue, F., Awad, I. A., Mosnier, L. O., & Ginsberg, M. H. (2019). Cerebral cavernous malformations form an anticoagulant vascular domain in humans and mice. *Blood*, *133*(3), 193–204.  
PMID 30442679

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12), 550. PMID 25516281

Luks, V. L., Kamitaki, N., Vivero, M. P., Uller, W., Rab, R., Bovee, J. V., Rialon, K. L., Guevara, C. J., Alomari, A. I., Greene, A. K., Fishman, S. J., Kozakewich, H. P., Maclellan, R. A., Mulliken, J. B., Rahbar, R., Spencer, S. A., Trenor, r., C. C., Upton, J., Zurakowski, D., Perkins, J. A., Kirsh, A., Bennett, J. T., Dobyns, W. B., Kurek, K. C., Warman, M. L., McCarroll, S. A., & Murillo, R. (2015). Lymphatic and other vascular malformative/overgrowth disorders are caused by somatic mutations in PIK3CA. *J Pediatr*, 166(4), 1048–54 e1–5. PMID 25681199

Lyne, S. B., Girard, R., Koskimaki, J., Zeineddine, H. A., Zhang, D., Cao, Y., Li, Y., Stadnik, A., Moore, T., Lightle, R., Shi, C., Shenkar, R., Carrion-Penagos, J., Polster, S. P., Romanos, S., Akers, A., Lopez-Ramirez, M., Whitehead, K. J., Kahn, M. L., Ginsberg, M. H., Marchuk, D. A., & Awad, I. A. (2019). Biomarkers of cavernous angioma with symptomatic hemorrhage. *JCI Insight*, 4(12). PMID 31217347

MacDonald, J. R., Ziman, R., Yuen, R. K., Feuk, L., & Scherer, S. W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res*, 42(Database issue), D986–92. PMID 24174537

Macmurdo, C. F., Woorderchak-Donahue, W., Bayrak-Toydemir, P., Le, J., Wallenstein, M. B., Milla, C., Teng, J. M., Bernstein, J. A., & Stevenson, D. A. (2016). RASA1 somatic mutation and variable expressivity in capillary malformation/arteriovenous malformation (CM/AVM) syndrome. *Am J Med Genet A*, 170(6), 1450–4. PMID 26969842

Mahmoud, M., Allinson, K. R., Zhai, Z., Oakenfull, R., Ghandi, P., Adams, R. H., Fruttiger, M., & Arthur, H. M. (2010). Pathogenesis of arteriovenous malformations in the absence of endoglin. *Circ Res*, 106(8), 1425–33. PMID 20224041

Malinverno, M., Maderna, C., Abu Taha, A., Corada, M., Orsenigo, F., Valentino, M., Pisati, F., Fusco, C., Graziano, P., Giannotta, M., Yu, Q. C., Zeng, Y. A., Lampugnani, M. G., Magnusson, P. U., & Dejana, E. (2019). Endothelial cell clonal expansion in the development of cerebral cavernous malformations. *Nat Commun*, 10(1), 2761. PMID 31235698

Martinez-Lage, J. F., de la Fuente, I., Ros de San Pedro, J., Fuster, J. L., Perez-Espejo, M. A., & Herrero, M. T. (2008). Cavernomas in children with brain tumors:

a late complication of radiotherapy. *Neurocirugia (Astur)*, 19(1), 50–4.  
PMID 18335155

McAllister, K. A., Grogg, K. M., Johnson, D. W., Gallione, C. J., Baldwin, M. A., Jackson, C. E., Helmbold, E. A., Markel, D. S., McKinnon, W. C., Murrell, J., & et al. (1994). Endoglin, a TGF-beta binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1. *Nat Genet*, 8(4), 345–51.  
PMID 7894484

McDonald, D. A., Shi, C., Shenkar, R., Gallione, C. J., Akers, A. L., Li, S., De Castro, N., Berg, M. J., Corcoran, D. L., Awad, I. A., & Marchuk, D. A. (2014). Lesions from patients with sporadic cerebral cavernous malformations harbor somatic mutations in the CCM genes: evidence for a common biochemical pathway for CCM pathogenesis. *Hum Mol Genet*, 23(16), 4357–70.  
PMID 24698976

McDonald, J., Bayrak-Toydemir, P., DeMille, D., Woorderchak-Donahue, W., & Whitehead, K. (2020). Curacao diagnostic criteria for hereditary hemorrhagic telangiectasia is highly predictive of a pathogenic variant in ENG or ACVRL1 (HHT1 and HHT2). *Genet Med*, 22(7), 1201–1205.  
PMID 32300199

McGranahan, N., Favero, F., de Bruin, E. C., Birkbak, N. J., Szallasi, Z., & Swanton, C. (2015). Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med*, 7(283), 283ra54.  
PMID 25877892

Milholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y., & Vijg, J. (2017). Differences between germline and somatic mutation rates in humans and mice. *Nat Commun*, 8, 15183.  
PMID 28485371

Mori, M. A., Ludwig, R. G., Garcia-Martin, R., Brando, B. B., & Kahn, C. R. (2019). Extracellular miRNAs: From Biomarkers to Mediators of Physiology and Disease. *Cell Metab*, 30(4), 656–673.  
PMID 31447320

Musunuru, K., Hillard, V. H., & Murali, R. (2003). Widespread central nervous system cavernous malformations associated with cafe-au-lait skin lesions. Case report. *J Neurosurg*, 99(2), 412–5.  
PMID 12924719

Nikolaev, S. I., Vetiska, S., Bonilla, X., Boudreau, E., Jauhainen, S., Rezai Jahromi, B., Khyzha, N., DiStefano, P. V., Suutarinen, S., Kiehl, T. R., Mendes Pereira, V., Herman, A. M., Krings, T., Andrade-Barazarte, H., Tung, T., Valiante, T., Zadeh,

G., Tymianski, M., Rauramaa, T., Yla-Herttuala, S., Wythe, J. D., Antonarakis, S. E., Frosen, J., Fish, J. E., & Radovanovic, I. (2018). Somatic Activating KRAS Mutations in Arteriovenous Malformations of the Brain. *N Engl J Med*, 378(3), 250–261.

PMID 29298116

North, P. E., Waner, M., Mizeracki, A., & Mihm, J., M. C. (2000). GLUT1: a newly discovered immunohistochemical marker for juvenile hemangiomas. *Hum Pathol*, 31(1), 11–22.

PMID 10665907

North, P. E., Waner, M., Mizeracki, A., Mrak, R. E., Nicholas, R., Kincannon, J., Suen, J. Y., & Mihm, J., M. C. (2001). A unique microvascular phenotype shared by juvenile hemangiomas and human placenta. *Arch Dermatol*, 137(5), 559–70.

PMID 11346333

Novelli, P. M., Reigel, D. H., Langham Gleason, P., & Yunis, E. (1997). Multiple cavernous angiomas after high-dose whole-brain radiation therapy. *Pediatr Neurosurg*, 26(6), 322–5.

PMID 9485161

Oh, S. P., Seki, T., Goss, K. A., Imamura, T., Yi, Y., Donahoe, P. K., Li, L., Miyazono, K., ten Dijke, P., Kim, S., & Li, E. (2000). Activin receptor-like kinase 1 modulates transforming growth factor-beta 1 signaling in the regulation of angiogenesis. *Proc Natl Acad Sci U S A*, 97(6), 2626–31.

PMID 10716993

Ola, R., Kunzel, S. H., Zhang, F., Genet, G., Chakraborty, R., Pibouin-Fragner, L., Martin, K., Sessa, W., Dubrac, A., & Eichmann, A. (2018). SMAD4 Prevents Flow Induced Arteriovenous Malformations by Inhibiting Casein Kinase 2. *Circulation*, 138(21), 2379–2394.

PMID 29976569

Olivieri, C., Pagella, F., Semino, L., Lanzarini, L., Valacca, C., Pilotto, A., Corno, S., Scappaticci, S., Manfredi, G., Buscarini, E., & Danesino, C. (2007). Analysis of ENG and ACVRL1 genes in 137 HHT Italian families identifies 76 different mutations (24 novel). Comparison with other European studies. *J Hum Genet*, 52(10), 820–9.

PMID 17786384

Pagenstecher, A., Stahl, S., Sure, U., & Felbor, U. (2009). A two-hit mechanism causes cerebral cavernous malformations: complete inactivation of CCM1, CCM2 or CCM3 in affected endothelial cells. *Hum Mol Genet*, 18(5), 911–8.

PMID 19088124

Park, S. O., Lee, Y. J., Seki, T., Hong, K. H., Fliess, N., Jiang, Z., Park, A., Wu, X., Kaartinen, V., Roman, B. L., & Oh, S. P. (2008). ALK5- and TGFBR2-independent role of ALK1 in the pathogenesis of hereditary hemorrhagic telangiectasia type 2. *Blood*, 111(2), 633–42.

PMID 17911384

Pece, N., Vera, S., Cymerman, U., White, J., R. I., Wrana, J. L., & Letarte, M. (1997). Mutant endoglin in hereditary hemorrhagic telangiectasia type 1 is transiently expressed intracellularly and is not a dominant negative. *J Clin Invest*, 100(10), 2568–79.

PMID 9366572

Pece-Barbara, N., Cymerman, U., Vera, S., Marchuk, D. A., & Letarte, M. (1999). Expression analysis of four endoglin missense mutations suggests that haploinsufficiency is the predominant mechanism for hereditary hemorrhagic telangiectasia type 1. *Hum Mol Genet*, 8(12), 2171–81.

PMID 10545596

Petersen, T. A., Morrison, L. A., Schrader, R. M., & Hart, B. L. (2010). Familial versus sporadic cavernous malformations: differences in developmental venous anomaly association and lesion phenotype. *AJNR Am J Neuroradiol*, 31(2), 377–82.

PMID 19833796

Pittman, K. M., Losken, H. W., Kleinman, M. E., Marcus, J. R., Blei, F., Gurtner, G. C., & Marchuk, D. A. (2006). No evidence for maternal-fetal microchimerism in infantile hemangioma: a molecular genetic investigation. *J Invest Dermatol*, 126(11), 2533–8.

PMID 16902414

Plauchu, H., de Chadarevian, J. P., Bideau, A., & Robert, J. M. (1989). Age-related clinical profile of hereditary hemorrhagic telangiectasia in an epidemiologically recruited population. *Am J Med Genet*, 32(3), 291–7.

PMID 2729347

Plummer, N. W., Zawistowski, J. S., & Marchuk, D. A. (2005). Genetics of cerebral cavernous malformations. *Curr Neurol Neurosci Rep*, 5(5), 391–6.

PMID 16131422

Polster, S. P., Cao, Y., Carroll, T., Flemming, K., Girard, R., Hanley, D., Hobson, N., Kim, H., Koenig, J., Koskimaki, J., Lane, K., Majersik, J. J., McBee, N., Morrison, L., Shenkar, R., Stadnik, A., Thompson, R. E., Zabramski, J., Zeineddine, H. A., & Awad, I. A. (2019). Trial Readiness in Cavernous Angiomas With Symptomatic Hemorrhage (CASH). *Neurosurgery*, 84(4), 954–964.

PMID 29660039

Polster, S. P., Sharma, A., Tanes, C., Tang, A. T., Mericko, P., Cao, Y., Carrion-Penagos, J., Girard, R., Koskimaki, J., Zhang, D., Stadnik, A., Romanos, S. G., Lyne, S. B., Shenkar, R., Yan, K., Lee, C., Akers, A., Morrison, L., Robinson, M., Zafar, A., Bittinger, K., Kim, H., Gilbert, J. A., Kahn, M. L., Shen, L., & Awad, I. A. (2020). Permissive microbiome characterizes human subjects with a neurovascular disease cavernous angioma. *Nat Commun*, 11(1), 2659.

PMID 32461638

Porter, P. J., Willinsky, R. A., Harper, W., & Wallace, M. C. (1997). Cerebral cavernous malformations: natural history and prognosis after clinical deterioration with or without hemorrhage. *J Neurosurg*, 87(2), 190–7.

PMID 9254081

Porter, R. W., Detwiler, P. W., Spetzler, R. F., Lawton, M. T., Baskin, J. J., Derksen, P. T., & Zabramski, J. M. (1999). Cavernous malformations of the brainstem: experience with 100 patients. *J Neurosurg*, 90(1), 50–8.

PMID 10413155

Queisser, A., Boon, L. M., & Vikkula, M. (2018). Etiology and Genetics of Congenital Vascular Lesions. *Otolaryngol Clin North Am*, 51(1), 41–53.

PMID 29217067

Rasalkar, D. D., & Paunipagar, B. K. (2010). Developmental venous anomaly associated with cortical dysplasia. *Pediatr Radiol*, 40 Suppl 1, S165.

PMID 20429001

Rath, M., Pagenstecher, A., Hoischen, A., & Felbor, U. (2020). Postzygotic mosaicism in cerebral cavernous malformation. *J Med Genet*, 57(3), 212–216.

PMID 31446422

Rechsteiner, M. P., von Teichman, A., Nowicka, A., Sulser, T., Schraml, P., & Moch, H. (2011). VHL gene mutations and their effects on hypoxia inducible factor HIFalpha: identification of potential driver and passenger mutations. *Cancer Res*, 71(16), 5500–11.

PMID 21715564

Ren, A. A., Snellings, D. A., Su, Y. S., Hong, C. C., Castro, M., Tang, A. T., Detter, M. R., Hobson, N., Girard, R., Romanos, S., Lightle, R., Moore, T., Shenkar, R., Benavides, C., Beaman, M. M., Mueller-Fielitz, H., Chen, M., Mericko, P., Yang, J., Sung, D. C., Lawton, M. T., Ruppert, M., Schwaninger, M., Korbelin, J., Potente, M., Awad, I. A., Marchuk, D. A., & Kahn, M. L. (2021). PIK3CA and CCM mutations fuel cavernomas through a cancer-like mechanism. *Nature*. PMID 33910229

Reuss, D. E., Piro, R. M., Jones, D. T., Simon, M., Ketter, R., Kool, M., Becker, A., Sahm, F., Pusch, S., Meyer, J., Hagenlocher, C., Schweizer, L., Capper, D.,

Kickingederer, P., Mucha, J., Koelsche, C., Jager, N., Santarius, T., Tarpey, P. S., Stephens, P. J., Andrew Futreal, P., Wellenreuther, R., Kraus, J., Lenartz, D., Herold-Mende, C., Hartmann, C., Mawrin, C., Giese, N., Eils, R., Collins, V. P., Konig, R., Wiestler, O. D., Pfister, S. M., & von Deimling, A. (2013). Secretory meningiomas are defined by combined *KLF4* K409Q and *TRAF7* mutations. *Acta Neuropathol*, 125(3), 351–8.

PMID 23404370

Riant, F., Bergametti, F., Fournier, H. D., Chapon, F., Michalak-Provost, S., Cecillon, M., Lejeune, P., Hosseini, H., Choe, C., Orth, M., Bernreuther, C., Boulday, G., Denier, C., Labauge, P., & Tournier-Lasserve, E. (2013). CCM3 Mutations Are Associated with Early-Onset Cerebral Hemorrhage and Multiple Meningiomas. *Mol Syndromol*, 4(4), 165–72.

PMID 23801932

Ricard, N., Bidart, M., Mallet, C., Lesca, G., Giraud, S., Prudent, R., Feige, J. J., & Bailly, S. (2010). Functional analysis of the BMP9 response of ALK1 mutants from HHT2 patients: a diagnostic tool for novel ACVRL1 mutations. *Blood*, 116(9), 1604–12.

PMID 20501893

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L., & Committee, A. L. Q. A. (2015). 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*, 17(5), 405–24.

PMID 25741868

Rodin, R. E., Dou, Y., Kwon, M., Sherman, M. A., D'Gama, A. M., Doan, R. N., Rento, L. M., Girsakis, K. M., Bohrson, C. L., Kim, S. N., Nadig, A., Luquette, L. J., Gulhan, D. C., Brain Somatic Mosaicism, N., Park, P. J., & Walsh, C. A. (2021). The landscape of somatic mutation in cerebral cortex of autistic and neurotypical individuals revealed by ultra-deep whole-genome sequencing. *Nat Neurosci*, 24(2), 176–185.

PMID 33432195

Rodriguez-Laguna, L., Agra, N., Ibanez, K., Oliva-Molina, G., Gordo, G., Khurana, N., Hominick, D., Beato, M., Colmenero, I., Herranz, G., Torres Canizalez, J. M., Rodriguez Pena, R., Vallespin, E., Martin-Arenas, R., Del Pozo, A., Villaverde, C., Bustamante, A., Ayuso, C., Lapunzina, P., Lopez-Gutierrez, J. C., Dellinger, M. T., & Martinez-Glez, V. (2019). Somatic activating mutations in PIK3CA cause generalized lymphatic anomaly. *J Exp Med*, 216(2), 407–418.

PMID 30591517

- Roux, A., Boddaert, N., Grill, J., Castel, D., Zanello, M., Zah-Bi, G., Chretien, F., Lefevre, E., Ros, V. D., Zerah, M., Puget, S., Pallud, J., & Varlet, P. (2020). High Prevalence of Developmental Venous Anomaly in Diffuse Intrinsic Pontine Gliomas: A Pediatric Control Study. *Neurosurgery*, 86(4), 517–523.  
PMID 31342064
- Rueda, A., Barturen, G., Lebron, R., Gomez-Martin, C., Alganza, A., Oliver, J. L., & Hackenberg, M. (2015). sRNAToolbox: an integrated collection of small RNA research tools. *Nucleic Acids Res*, 43(W1), W467–73.  
PMID 26019179
- Ruiz-Llorente, L., Gallardo-Vara, E., Rossi, E., Smadja, D. M., Botella, L. M., & Bernabeu, C. (2017). Endoglin and alk1 as therapeutic targets for hereditary hemorrhagic telangiectasia. *Expert Opin Ther Targets*, 21(10), 933–947.  
PMID 28796572
- Sahoo, T., Johnson, E. W., Thomas, J. W., Kuehl, P. M., Jones, T. L., Dokken, C. G., Touchman, J. W., Gallione, C. J., Lee-Lin, S. Q., Kosofsky, B., Kurth, J. H., Louis, D. N., Mettler, G., Morrison, L., Gil-Nagel, A., Rich, S. S., Zabramski, J. M., Boguski, M. S., Green, E. D., & Marchuk, D. A. (1999). Mutations in the gene encoding KRIT1, a Krev-1/rap1a binding protein, cause cerebral cavernous malformations (CCM1). *Hum Mol Genet*, 8(12), 2325–33.  
PMID 10545614
- Samuels, Y., Diaz, J., L. A., Schmidt-Kittler, O., Cummins, J. M., Delong, L., Cheong, I., Rago, C., Huso, D. L., Lengauer, C., Kinzler, K. W., Vogelstein, B., & Velculescu, V. E. (2005). Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell*, 7(6), 561–73.  
PMID 15950905
- Santucci, G. M., Leach, J. L., Ying, J., Leach, S. D., & Tomsick, T. A. (2008). Brain parenchymal signal abnormalities associated with developmental venous anomalies: detailed MR imaging assessment. *AJNR Am J Neuroradiol*, 29(7), 1317–23.  
PMID 18417603
- Satas, G., & Raphael, B. J. (2018). Haplotype phasing in single-cell DNA-sequencing data. *Bioinformatics*, 34(13), i211–i217.  
PMID 29950014
- Schuetz, A., Nana, D., Rose, C., Zocher, G., Milanovic, M., Koenigsmann, J., Blasig, R., Heinemann, U., & Carstanjen, D. (2011). The structure of the Klf4 DNA-binding domain links to self-renewal and macrophage differentiation. *Cell Mol Life Sci*, 68(18), 3121–31.  
PMID 21290164

- Seki, T., Yun, J., & Oh, S. P. (2003). Arterial endothelium-specific activin receptor-like kinase 1 expression suggests its role in arterialization and vascular remodeling. *Circ Res*, 93(7), 682–9.  
PMID 12970115
- Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol*, 14(8), e1002533.  
PMID 27541692
- Shenkar, R., Shi, C., Rebeiz, T., Stockton, R. A., McDonald, D. A., Mikati, A. G., Zhang, L., Austin, C., Akers, A. L., Gallione, C. J., Rorrer, A., Gunel, M., Min, W., De Souza, J. M., Lee, C., Marchuk, D. A., & Awad, I. A. (2015). Exceptional aggressiveness of cerebral cavernous malformation disease associated with PDCD10 mutations. *Genet Med*, 17(3), 188–196.  
PMID 25122144
- Shirley, M. D., Tang, H., Gallione, C. J., Baugher, J. D., Frelin, L. P., Cohen, B., North, P. E., Marchuk, D. A., Comi, A. M., & Pevsner, J. (2013). Sturge-Weber syndrome and port-wine stains caused by somatic mutation in GNAQ. *N Engl J Med*, 368(21), 1971–9.  
PMID 23656586
- Shoushtari, A. N., & Carvajal, R. D. (2014). GNAQ and GNA11 mutations in uveal melanoma. *Melanoma Res*, 24(6), 525–34.  
PMID 25304237
- Shovlin, C. L., Guttmacher, A. E., Buscarini, E., Faughnan, M. E., Hyland, R. H., Westermann, C. J., Kjeldsen, A. D., & Plauchu, H. (2000). Diagnostic criteria for hereditary hemorrhagic telangiectasia (Rendu-Osler-Weber syndrome). *Am J Med Genet*, 91(1), 66–7.  
PMID 10751092
- Shovlin, C. L., Hughes, J. M., Scott, J., Seidman, C. E., & Seidman, J. G. (1997). Characterization of endoglin and identification of novel mutations in hereditary hemorrhagic telangiectasia. *Am J Hum Genet*, 61(1), 68–79.  
PMID 9245986
- Sirvente, J., Enjolras, O., Wassef, M., Tournier-Lasserve, E., & Labauge, P. (2009). Frequency and phenotypes of cutaneous vascular malformations in a consecutive series of 417 patients with familial cerebral cavernous malformations. *J Eur Acad Dermatol Venereol*, 23(9), 1066–72.  
PMID 19453802
- Snellings, D. A., Gallione, C. J., Clark, D. S., Vozoris, N. T., Faughnan, M. E., & Marchuk, D. A. (2019). Somatic Mutations in Vascular Malformations

of Hereditary Hemorrhagic Telangiectasia Result in Bi-allelic Loss of ENG or ACVRL1. *Am J Hum Genet*, 105(5), 894–906.  
PMID 31630786

Snellings, D. A., Hong, C. C., Ren, A. A., Lopez-Ramirez, M. A., Girard\*, R., Srinath, A., Marchuk, D. A., Ginsberg, M. H., Awad, I. A., & Kahn, M. L. (2021). Cerebral Cavernous Malformation: From Mechanism to Therapy. *Circ Res*, 129(1), 195–215.

PMID 34166073

Soblet, J., Kangas, J., Natynki, M., Mendola, A., Helaers, R., Uebelhoer, M., Kaakinen, M., Cordisco, M., Dompmartin, A., Enjolras, O., Holden, S., Irvine, A. D., Kangesu, L., Leaute-Labreze, C., Lanoel, A., Lokmic, Z., Maas, S., McAleer, M. A., Penington, A., Rieu, P., Syed, S., van der Vleuten, C., Watson, R., Fishman, S. J., Mulliken, J. B., Eklund, L., Limaye, N., Boon, L. M., & Vakkula, M. (2017). Blue Rubber Bleb Nevus (BRBN) Syndrome Is Caused by Somatic TEK (TIE2) Mutations. *J Invest Dermatol*, 137(1), 207–216.

PMID 27519652

Srinivasan, S., Hanes, M. A., Dickens, T., Porteous, M. E., Oh, S. P., Hale, L. P., & Marchuk, D. A. (2003). A mouse model for hereditary hemorrhagic telangiectasia (HHT) type 2. *Hum Mol Genet*, 12(5), 473–82.  
PMID 12588795

Starke, R. M., McCarthy, D., Komotar, R. J., & Connolly, E. S. (2018). Somatic KRAS Mutation Found in Sporadic Arteriovenous Malformations. *Neurosurgery*, 83(1), E14–E15.  
PMID 29917132

Sticht, C., De La Torre, C., Parveen, A., & Gretz, N. (2018). miRWALK: An online resource for prediction of microRNA binding sites. *PLoS One*, 13(10), e0206239.  
PMID 30335862

Strenger, V., Sovinz, P., Lackner, H., Dornbusch, H. J., Lingitz, H., Eder, H. G., Moser, A., & Urban, C. (2008). Intracerebral cavernous hemangioma after cranial irradiation in childhood. Incidence and risk factors. *Strahlenther Onkol*, 184(5), 276–80.  
PMID 18427759

Szulwach, K. E., Chen, P., Wang, X., Wang, J., Weaver, L. S., Gonzales, M. L., Sun, G., Unger, M. A., & Ramakrishnan, R. (2015). Single-Cell Genetic Analysis Using Automated Microfluidics to Resolve Somatic Mosaicism. *PLoS One*, 10(8), e0135007.  
PMID 26302375

Tan, W. H., Baris, H. N., Burrows, P. E., Robson, C. D., Alomari, A. I., Mulliken, J. B., Fishman, S. J., & Irons, M. B. (2007). The spectrum of vascular anomalies in patients with PTEN mutations: implications for diagnosis and management. *J Med Genet*, 44(9), 594–602.

PMID 17526801

Tang, A. T., Choi, J. P., Kotzin, J. J., Yang, Y., Hong, C. C., Hobson, N., Girard, R., Zeineddine, H. A., Lightle, R., Moore, T., Cao, Y., Shenkar, R., Chen, M., Mericko, P., Yang, J., Li, L., Tanes, C., Kobuley, D., Vosa, U., Whitehead, K. J., Li, D. Y., Franke, L., Hart, B., Schwaninger, M., Henao-Mejia, J., Morrison, L., Kim, H., Awad, I. A., Zheng, X., & Kahn, M. L. (2017). Endothelial TLR4 and the microbiome drive cerebral cavernous malformations. *Nature*, 545(7654), 305–310.

PMID 28489816

Tang, A. T., Sullivan, K. R., Hong, C. C., Goddard, L. M., Mahadevan, A., Ren, A., Pardo, H., Peiper, A., Griffin, E., Tanes, C., Mattei, L. M., Yang, J., Li, L., Mericko-Ishizuka, P., Shen, L., Hobson, N., Girard, R., Lightle, R., Moore, T., Shenkar, R., Polster, S. P., Roedel, C. J., Li, N., Zhu, Q., Whitehead, K. J., Zheng, X., Akers, A., Morrison, L., Kim, H., Bittinger, K., Lengner, C. J., Schwaninger, M., Velcich, A., Augenlicht, L., Abdelilah-Seyfried, S., Min, W., Marchuk, D. A., Awad, I. A., & Kahn, M. L. (2019). Distinct cellular roles for PDCD10 define a gut-brain axis in cerebral cavernous malformation. *Sci Transl Med*, 11(520).

PMID 31776290

Ten Broek, R. W., Eijkelenboom, A., van der Vleuten, C. J. M., Kamping, E. J., Kets, M., Verhoeven, B. H., Grunberg, K., Schultze Kool, L. J., Tops, B. B. J., Ligtenberg, M. J. L., & Flucke, U. (2019). Comprehensive molecular and clinicopathological analysis of vascular malformations: A study of 319 cases. *Genes Chromosomes Cancer*, 58(8), 541–550.

PMID 30677207

Torsney, E., Charlton, R., Diamond, A. G., Burn, J., Soames, J. V., & Arthur, H. M. (2003). Mouse model for hereditary hemorrhagic telangiectasia has a generalized vascular abnormality. *Circulation*, 107(12), 1653–7.

PMID 12668501

Tual-Chalot, S., Oh, S. P., & Arthur, H. M. (2015). Mouse models of hereditary hemorrhagic telangiectasia: recent advances and future challenges. *Front Genet*, 6, 25.

PMID 25741358

Ucler, N. (2019). An intracranial developmental venous anomaly presenting with seizure. *Neurol India*, 67(2), 604–605.

PMID 31085894

Vinchon, M., Leblond, P., Caron, S., Delestret, I., Baroncini, M., & Coche, B. (2011). Radiation-induced tumors in children irradiated for brain tumor: a longitudinal study. *Childs Nerv Syst*, 27(3), 445–53.  
PMID 21234575

von Spreckelsen, N., Waldt, N., Poetschke, R., Kesseler, C., Dohmen, H., Jiao, H. K., Nemeth, A., Schob, S., Scherlach, C., Sandalcioglu, I. E., Deckert, M., Angenstein, F., Krischek, B., Stavrinou, P., Timmer, M., Remke, M., Kirches, E., Goldbrunner, R., Chiocca, E. A., Huettelmaier, S., Acker, T., & Mawrin, C. (2020). KLF4(K409Q)-mutated meningiomas show enhanced hypoxia signaling and respond to mTORC1 inhibitor treatment. *Acta Neuropathol Commun*, 8(1), 41.

PMID 32245394

Walker, E. J., Su, H., Shen, F., Choi, E. J., Oh, S. P., Chen, G., Lawton, M. T., Kim, H., Chen, Y., Chen, W., & Young, W. L. (2011). Arteriovenous malformation in the adult mouse brain resembling the human disease. *Ann Neurol*, 69(6), 954–62.  
PMID 21437931

Wang, C., Schultzhaus, J. N., Taitt, C. R., Leary, D. H., Shriver-Lake, L. C., Snellings, D., Sturiale, S., North, S. H., Orihuela, B., Rittschof, D., Wahl, K. J., & Spillmann, C. M. (2018). Characterization of longitudinal canal tissue in the acorn barnacle *Amphibalanus amphitrite*. *PLoS One*, 13(12), e0208352.  
PMID 30532169

Wang, X., Liu, X. W., Lee, N., Liu, Q. J., Li, W. N., Han, T., Wei, K. K., Qiao, S., & Chi, Z. F. (2013). Features of a Chinese family with cerebral cavernous malformation induced by a novel CCM1 gene mutation. *Chin Med J (Engl)*, 126(18), 3427–32.  
PMID 24034083

Weng, J., Yang, Y., Song, D., Huo, R., Li, H., Chen, Y., Nam, Y., Zhou, Q., Jiao, Y., Fu, W., Yan, Z., Wang, J., Xu, H., Di, L., Li, J., Wang, S., Zhao, J., Wang, J., & Cao, Y. (2021). Somatic MAP3K3 mutation defines a subclass of cerebral cavernous malformation. *Am J Hum Genet*, 108(5), 942–950.  
PMID 33891857

Wetzel-Strong, S. E., Detter, M. R., & Marchuk, D. A. (2017). The pathobiology of vascular malformations: insights from human and model organism genetics. *J Pathol*, 241(2), 281–293.  
PMID 27859310

Wooderchak-Donahue, W. L., McDonald, J., Farrell, A., Akay, G., Velinder, M., Johnson, P., VanSant-Webb, C., Margraf, R., Briggs, E., Whitehead, K. J., Thomson, J., Lin, A. E., Pyeritz, R. E., Marth, G., & Bayrak-Toydemir, P. (2018).

Genome sequencing reveals a deep intronic splicing ACVRL1 mutation hotspot in Hereditary Haemorrhagic Telangiectasia. *J Med Genet*, 55(12), 824–830.  
PMID 30244195

Wooderchak-Donahue, W. L., McDonald, J., O'Fallon, B., Upton, P. D., Li, W., Roman, B. L., Young, S., Plant, P., Fulop, G. T., Langa, C., Morrell, N. W., Botella, L. M., Bernabeu, C., Stevenson, D. A., Runo, J. R., & Bayrak-Toydemir, P. (2013). BMP9 mutations cause a vascular-anomaly syndrome with phenotypic overlap with hereditary hemorrhagic telangiectasia. *Am J Hum Genet*, 93(3), 530–7.

PMID 23972370

Wurm, G., Schnizer, M., & Fellner, F. A. (2005). Cerebral cavernous malformations associated with venous anomalies: surgical considerations. *Neurosurgery*, 57(1 Suppl), 42–58; discussion 42–58.  
PMID 15987569

Xu, L., Durruthy-Durruthy, R., Eastburn, D. J., Pellegrino, M., Shah, O., Meyer, E., & Zehnder, J. (2019). Clonal Evolution and Changes in Two AML Patients Detected with A Novel Single-Cell DNA Sequencing Platform. *Sci Rep*, 9(1), 11119.

PMID 31366893

Zhou, Z., Rawnsley, D. R., Goddard, L. M., Pan, W., Cao, X. J., Jakus, Z., Zheng, H., Yang, J., Arthur, J. S., Whitehead, K. J., Li, D., Zhou, B., Garcia, B. A., Zheng, X., & Kahn, M. L. (2015). The cerebral cavernous malformation pathway controls cardiac development via regulation of endocardial MEKK3 signaling and KLF expression. *Dev Cell*, 32(2), 168–80.

PMID 25625206

Zhou, Z., Tang, A. T., Wong, W. Y., Bamezai, S., Goddard, L. M., Shenkar, R., Zhou, S., Yang, J., Wright, A. C., Foley, M., Arthur, J. S., Whitehead, K. J., Awad, I. A., Li, D. Y., Zheng, X., & Kahn, M. L. (2016). Cerebral cavernous malformations arise from endothelial gain of MEKK3-KLF2/4 signalling. *Nature*, 532(7597), 122–6.

PMID 27027284

# Biography

Daniel Aaron Snellings graduated from the Pennsylvania State University in 2017 with a Bachelor of Science in Biochemistry and Molecular Biology. Daniel entered Duke in 2017 through the Cell and Molecular Biology graduate program and joined the lab of Douglas A. Marchuk in 2018. Daniel was awarded a Ruth L. Kirschstein predoctoral fellowship from the NHLBI in 2020. In 2021, Daniel was selected as a featured plenary speaker for the annual meeting of the American Society of Human Genetics and there was selected as a semifinalist for the Charles J. Epstein award. Daniel received a Ph.D. in Molecular Genetics and Microbiology from Duke University in 2022. In April 2022, Daniel began a postdoctoral fellowship in the lab of Christopher A. Walsh at Boston Children's Hospital.

# Publications

- D. A. Snellings**, et al. (2022) *Nat. Cardiovasc. Res.*, in press
- D. A. Snellings**, et al. (2021) *Circ Res*
- A. A. Ren & **D. A. Snellings**, et al. (2021) *Nature*
- D. A. Snellings**, et al. (2019) *Am J Hum Genet*
- J. Koskimaki, et al. (2019) *Acta Neuropathol Commun*
- M. R. Detter, **D. A. Snellings**, D. A. Marchuk. (2018) *Circ Res*
- C. Wang, et al. (2018) *PLoS One*