UNIVERSITY OF TURIN

Doctoral thesis in Complex Systems for Life Sciences

DOCTORAL THESIS

Genetic basis of Cerebral Cavernous Malformation and genetic modifiers of disease severity

---

Supervisor:

Chiar.mo Prof. Saverio Francesco Retta

Author:

Valerio Benedetti

XXXIII Cycle

Date

Declaration of Authorship

I, Valerio Benedetti, declare that this thesis titled, "***Genetic basis of Cerebral Cavernous Malformation and genetic modifiers of disease severity***" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Date Signature

UNIVERSITY OF TURIN

Abstract

Doctoral school in Complex Systems for Life Sciences

Department of Biological and Clinical Sciences

Genetic basis of Cerebral Cavernous Malformation

and genetic modifiers of disease severity

by Valerio Benedetti

Loss-of-function (LOF) mutations of KRIT1/CCM1, CCM2 or CCM3 genes have been associated to Cerebral Cavernous Malformation (CCM), a major cerebrovascular disease that may occur either sporadically (sCCM) or as an inherited (fCCM) autosomal-dominant condition with incomplete penetrance and highly variable expressivity (OMIM 116860). This disease is characterized by the formation of abnormally dilated capillaries, which are referred to as CCM lesions and occur mainly in the central nervous system, showing a characteristic appearance on MRI and causing various clinical symptoms, including seizures, focal neurological deficits and intracerebral hemorrhage (ICH). However, there also clear evidence that CCM lesions can remain clinically quiescent in up to ~70% of the cases.

Along with the original identification of the genetic basis of CCM disease, the recent discovery of putative genetic modifiers of its severity has opened a new perspective on the comprehensive understanding of pathogenetic mechanisms and the development of preventive and therapeutic strategies. In this light, the fast-growing next generation sequencing (NGS) techniques represent very useful tools for a more rapid, comprehensive and effective identification of both causative mutations and genetic modifiers in the familial form of CCM disease. Indeed, exome sequencing, which sequences the protein-coding region of the genome, has been rapidly applied to variant discovery in research settings, whereas recent coverage and accuracy improvements have accelerated the development of clinical exome sequencing (CES) platforms

targeting disease-related genes and enabling variant identification in patients with suspected genetic diseases. Consistently, CES is rapidly becoming the diagnostic test of choice in patients with suspected Mendelian diseases, especially for those with heterogeneous etiology and clinical presentation. Furthermore, while genetic heterogeneity, phenotypic variability and disease rarity are all factors that make the traditional diagnostic approach to genetic disorders very challenging and obsolete, reporting large CES series can improve genetic variant identification and interpretation through clinically-oriented data sharing, and may also inform guidelines on best practices for test utilization and clinical management procedures.

In this thesis, I describe the application of NGS protocols for the identification and analysis of causative genes and genetic modifiers in patients affected by CCM disease, as well as their advantages over conventional approaches, such as those based on Sanger sequencing of candidate genes.

The experimental outcomes demonstrated that the application of optimized NGS workflows based on clinical gene panels, including genes and genetic variants involved in distinct cerebrovascular diseases, made it possible to identify novel causative variants in known CCM genes, as well as other potential causative genes and genetic risk factors associated to CCM disease onset and severity, thus providing useful insights into the improvement of current diagnostic procedures for a more effective diagnosis, prognosis and clinical management of CCM disease.

Acknowledgements

I would like to express my sincere gratitude to my tutor Prof. Saverio Francesco Retta for the support and encouragement that he has given to me during the past 5 year. He has given to me the opportunity to join the Applied Biology Group of the University of Turin and to learn how to do a serious research work. He has provided me the opportunity to work on challenging and very interesting topics and he has always given to me all the needed technical assistance.

I would also like to thank Prof. Roberto Piva, Prof. Alfredo Brusco, Prof.ssa Francesca Cordero, Prof. Antonio D'Avolio, Prof. Luigi Battaglia, Dott.ssa Elisa Giorgio, Dott.ssa Laura Follia, Dott.ss Patrizia Dentelli, Dott. Giulio Ferrero, Dott.ssa Jessica Cusato and Dott.ssa Elisa Pellegrino for their careful and essential support of this thesis and for all the work done together.

Finally, I would like to express my gratitude to Dott. Carlo Arduino for conveying to me his passion and dedication in the dedication to better understand genetic diseases, and to Dott. Eliana Trapani for her professionalism and dedication to teamwork and friendship.

During the years that I spent doing research, I met many special people and I really enjoyed working with them.

I would like also to acknowledge all the Groups: the "XXXIII cycle Group", the DSCBs, Genetics, Computer Science, Amedeo di Savoia, "Terroni e Polenta Group" and Unito ladies for the help and friendship that they gave to me.

Of course, I would like to thank them also for the pleasant time spent together.

I would like to thank my parents and my sister for supporting me and giving me the possibility to choose what I would prefer to do when I grew up.

I would like to acknowledge the Lagrange Project - CRT Foundation/ISI Foundation for providing me with a grant to study rare cerebrovascular diseases of genetic origin.

Contents

- Conclusion
- References

List of Figures

Figure 10: CCM proteins in signal transduction. Signal transduction from the CCM complex signaling platform is shown and the roles of CCM proteins outside of the CCM complex signaling platform are also indicated. The CCM proteins play roles in cell adhesion complexes and integrin signaling, in kinase signaling cascades, and in degradation/regulation of the Rho family of small GTPases (Fisher O S., 2013).

List of Tables

List of Abbreviations

WHO - World Health Organization, **CVDs** - Cardiovascular diseases, **CCM** - Cerebral Cavernous Malformation, **NGS** - Next generation sequencing, **CES** – Clinical exome sequencing, **WGS** - Whole Genome Sequencing, **WES** - Whole Exome Sequencing, **SG** - SOPHiATm Genetics, **CNS** – Central Nervous System, **MRI** - Magnetic resonance imaging, **KRIT1**/**CCM1** - Krev interaction trapped 1/Cerebral Cavernous Malformation 1(gene), **CCM2** - Cerebral Cavernous Malformation 2(gene), **CCM3**/**PDCD10** - Cerebral Cavernous

Malformation 3/ Programmed cell death 10 (gene), **fCCM** - Familial cerebral cavernous malformation, **sCCM** – Sporadic cerebral cavernous malformation, **ICH** - intra-cerebral hemorrhages, **LOF** - Loss of function, **SNV** – Small Nucleotide Variant, **CNV** - Copy Number Variant, **CADASIL** - Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy, **UPL** - Universal ProbeLibrary, **ACMG** - American College of Medical Genetics , **GACI2** - Pseudoxanthoma Elasticum and Arterial Calcification Generalized Of Infancy 2, **MSMDS** - Multisystemic Smooth Muscle Dysfunction Syndrome, **AAT6** - Aortic Aneurysm Familial Thoracic 6, **HHT2** - Telangiectasia Hereditary Hemorrhagic Type 2, **PPH1** - Pulmonary Hypertension Primary 1, **CAA-APP** - Cerebral Amyloid Angiopathy App-Related, **AD** - Alzheimer Disease, **MYMY4** - Moyamoya disease 4 with short stature hypergonadotropic hypogonadism and facial dysmorphism, **T-PLL** - T-Cell Prolymphocytic Leukemia, **EDSVASC** - Ehlers-Danlos syndrome vascular type, **PMGEDSV** - Polymicrogyria with or without vascular-type Ehlers-Danlos syndrome, **RATOR** - Retinal Arteries Tortuosity Of, **HANAC** - Angiopathy Hereditary With Nephropathy Aneurysms and Muscle Cramps, **BSVD2 -** Brain small vessel disease 2, **HHT1** - Telangiectasia hereditary hemorrhagic type , **HHT** - Hereditary hemorrhagic telangiectasia, **MFS** - Marfan syndrome, **SSKS** - Stiff skin syndrome, **FD** - Fabry disease, **CARASIL** - Cerebral arteriopathy autosomal recessive with subcortical infarcts and leukoencephalopathy, **CADASIL2** - Cerebral arteriopathy autosomal dominant with subcortical infarcts and leukoencephalopathy type 2, **HDBSCC** - Hemorrhagic destruction of the brain subependymal calcification and cataracts, **JBS** - Jacobsen syndrome, **AAT4** - Aortic aneurysm familial thoracic 4, **FAA** - familial thoracic aortic aneurysm and aortic dissection, **AAT4** - aortic aneurysm familial thoracic 4, **FAA** - Familial thoracic aortic aneurysm and aortic dissection, **AAT7** - Aortic aneurysm familial thoracic 7, **MMIHS** - Megacystis-microcolon-intestinal hypoperistalsis syndrome, **LMNS** - Lateral meningocele syndrome, **AGS5** - Aicardi-Goutieres syndrome 5, **CHBL2** - Chilblain Lupus 2 , **MSSE** - Multiple Self-Healing Squamous Epithelioma, **LDS1** – Loeys-Dietz Syndrome 1, **LDS2** - Loeys-Dietz syndrome 2, **HNPCC6** –Colorectal cancer hereditary nonpolyposis type 6, **RVCL** - Vasculopathy retinal with cerebral leukodystrophy, **AGS1** - Aicardi-Goutieres syndrome 1,

Introduction

Cardiovascular diseases (CVDs) are a group of disorders of the heart and blood vessels, including congenital heart disease, coronary heart disease, peripheral arterial disease, and cerebrovascular disease. According to the World Health Organization (WHO), CVDs are the first cause of death globally, taking an estimated 17.9 million lives each year (WHO, 2016). Among CVDs, cerebrovascular diseases are both leading causes of mortality and the main cause of adult long-term disability in developed countries. They encompass a wide range of disorders of the brain vasculature, including ischemic and hemorrhagic strokes, aneurysms, and vascular malformations, and may also indirectly affect the nervous system leading to a large variety of transient or permanent neurological disorders. A few types of cerebrovascular diseases are known to have a genetic basis and thus can be inherited. These include Cerebral Cavernous Malformation (CCM), a major cerebrovascular disorder that affects capillaries in the central nervous system and may cause severe neurological deficits and life-threatening hemorrhagic stroke. The identification of causative genes and genetic modifiers of inherited cerebrovascular diseases has provided useful insights into the understanding of pathogenetic mechanisms and the development of novel therapeutic approaches. In particular, the advent of the next generation sequencing (NGS) technique has enabled a more rapid and effective identification of both causative mutations and genetic modifiers in patients with suspected genetic diseases, including familial forms of cerebrovascular diseases, thus having a great impact on the study of such complex diseases as well as in their management either by available treatments or genetic counseling. Indeed, genetic heterogeneity, phenotypic variability, and disease rarity are all factors that make the traditional diagnostic approach to genetic disorders, whereby a specific gene is selected for sequencing based on the clinical phenotype, very challenging and obsolete. On the other hand, the NGS technology allows high speed and throughput DNA sequencing for identifying causal gene variants in the rare diseases via several different ways, including whole-genome sequencing (WGS), whole-exome sequencing (WES), clinical exome sequencing (CES), transcriptome sequencing, and other sequencing approaches. WES, which sequences the protein-coding region of the

genome, has been rapidly applied to variant discovery in research settings. Moreover, recent coverage and accuracy improvements have accelerated the development of CES platforms targeting disease-related genes and enabling variant identification in patients with suspected genetic diseases. Accordingly, CES is rapidly becoming the diagnostic test of choice in patients with suspected Mendelian diseases, especially for those with heterogeneous etiology and clinical presentation. Reporting large CES series can improve guidelines on best practices for test utilization, and a better variant interpretation through clinically oriented data sharing.

In my doctoral research work, I focused on genetic studies of CCM disease using both Sanger sequencing and NGS approaches, including the development and application of optimized NGS workflows based on panels of genes involved in distinct cerebrovascular diseases. Besides further highlighting the advantages of NGS technologies over conventional approaches, the outcomes of these studies allowed the identification of novel causative gene variants and putative genetic modifiers in patients affected by CCM disease, thus providing a useful framework for the improvement of current diagnostic and clinical management procedures.

Cerebral Cavernous Malformation

Cerebral cavernous malformation (CCM) disease is a major cerebrovascular disorder of genetic origin (OMIM 116860). It is characterized by the formation of clusters of abnormally dilated and fragile capillaries, which are referred to as CCM lesions and are known to occur in either sporadic (sCCM) or familial (fCCM) forms with a prevalence of 0.3-0.5% in the general population. CCM lesions occur mainly in the central nervous system (CNS), and may remain clinically silent or cause various symptoms, ranging from headache to focal neurological signs, seizures and severe intra-cerebral hemorrhages (ICH). Magnetic resonance imaging (MRI) is the current diagnostic modality of choice, allowing the detection of CCM lesions of various sizes. Generally, the sCCM and fCCM forms are characterized by the presence of single or multiple lesions, respectively, the latter form being also inherited as an autosomal dominant condition with incomplete penetrance and highly variable expressivity. Familial cases have been linked to mutations in three genes, *KRIT1* (*CCM1*), *mgc4607* (CCM2) and *pdcd10* (CCM3). Herein, we report the outcomes of genetic studies in patients and families carrying multiple CCM lesions and showing a wide range of clinical expressions, including the identification of new mutations in *KRIT1*/*CCM1* as well as in other genes not yet associated to CCM disease.

Genetic aspects

The genetic nature of the disease is already known. Genetic screening in sporadic cases with multiple CCM lesions will identify a mutation in 60% of the cases, in this study I identified a causative mutation in 80% of analyzed samples. This information will not change the patient's clinical care but may be useful for genetic counseling. However, the patient has to be aware that a negative test does not exclude a genetic cause (Cavernous Malformations of the Nervous System – book - by Daniele Rigamonti, 2011).

Commentato [MD1]: 60 o 80? dai dati sotto, sembra 80

To date, mutations in three genes, KRIT1/CCM1, CCM2 and CCM3, have been demonstrated to cause CCM disease, and the following mutation rates have been reported: 50% on KRIT1/CCM1, 20% on CCM2 and 10% on CCM3. However, in 20% of cases no mutations have been identified in the three known CCM genes, suggesting that

other genes could be involved in pathological mechanisms. The three known CCM genes are localized on chromosome 7 (KRIT1/CCM1 and CCM2) and chromosome 3 (CCM3) (Figure 1).

CCM is an autosomal dominant disease caused mainly by loss-of-function (LOF) mutations of CCM genes. Indeed, LOF mutations of any of the three known CCM genes have been clearly associated to various pathological phenotypes, such as an increase in intracellular ROS levels and ROS-mediated cellular dysfunctions, including a reduced ability to maintain a quiescent state, linked to KRIT1 LOF (Goitre L., 2010); endothelial cell dysfunction and loss of vascular integrity linked to CCM2 LOF (Boulday G, 2009); and alteration of the actin and microtubule cytoskeleton affecting neuronal morphology and laminar positioning of primarily late-born neurons linked to CCM3 LOF (Angeliki Louvi, 2014).

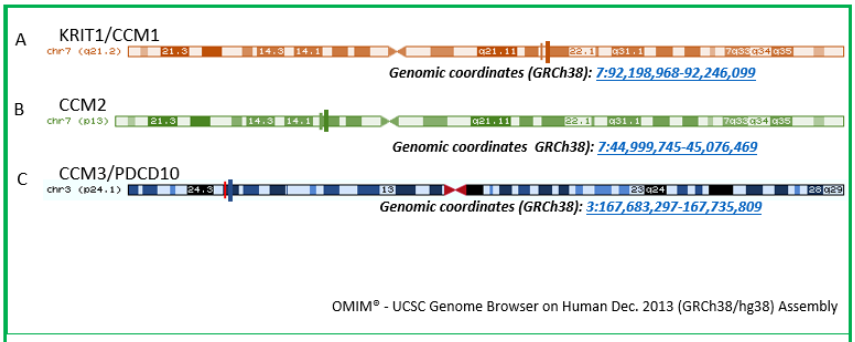

Figure 1: CCM genes: A) KRIT1/CCM1 - Chromosomal location: 7q21.2 – Genomic coordinates (GRCh38): 7:92,198,968-92,246,099; B) CCM2 - Chromosomal location: 7p13 - Genomic coordinates (GRCh38): 7:44,999,745-45,076,469; C) CCM3/PDCD10 - Chromosomal location: 3q26.1 - Genomic coordinates (GRCh38): 3:167,683,297-167,735,809.

Clinical Manifestations

CCM disease is the second most common incidental vascular finding – after aneurysms – on brain MRI, with a prevalence of 1 in 625 neurologically asymptomatic people (Horne MA, 2015). The incidence of incidentally detected CCM has increased substantially thanks to the widespread use of MRI (Moore SA, 2014 May). As I mentioned above, symptomatic patients could manifest a wide range of symptoms due to lesion positioning into the CNS. Location is the most important factor that reflects the clinical manifestation of CCM lesion (Ellis JA, 2017). Indeed, clinical course of CCM is highly variable and location-dependent. Clinical manifestations management and treatment are tightly correlated to lesion position; indeed, lesions can cause symptoms ranging from headache to focal neurological signs, seizures and severe intra-cerebral hemorrhages. Surgical resection is the preferred treatment option for symptomatic CCM. In select cases, targeted radiotherapy is used to treat lesions that are surgically inaccessible. There is no role for endovascular therapy in CCM. Most relevant parameters are symptomatic hemorrhage and hemorrhagic risk.

Hereditary forms of CCM (fCCM) are caused by LOF variants in one of three known genes (KRIT1/CCM1, CCM2 and PDCD10/CCM3). However, CCM patients carrying the same mutation can manifest different phenotypes and symptoms. Such variable expressivity has been attributed to the emerged pleiotropic redox-sensitive mechanisms of CCM disease pathogenesis (Goitre L., 2010; Guazzi, 2012; Goitre L. E., 2014; Marchi S. E., 2016; Marchi S. S., 2016; Marchi S. S., 2016; Retta, 2016), as well as to the inter-individual heterogeneity of susceptibility to local oxidative-stress and inflammatory events (Trapani, 2015; Choquet, 2016). Indeed, distinct genetic modifiers of endothelial cell responses to such microenvironmental stress factors, including single-nucleotide polymorphisms (SNPs) of distinct members of the CYP and MMP gene families, have been reported to impact the severity of CCM disease (Trapani, 2015; Choquet, 2016). Remarkably, whereas these evidences prompt complementary diagnostic approaches for the identification of genetic risk factors that may allow an effective risk stratification of patients with enhanced susceptibility to develop the most severe phenotypes of CCM disease, including ICH (Trapani, 2015; Choquet, 2016), they are also providing a

fundamental framework for the development of preventive and therapeutic approaches (Marchi S. E., 2016; Retta, 2016; Goitre L. P., 2017; Gibson CC, 2015; Moglia A, 2015; Moglianetti M D. L., 2016; De Luca E, 2018; Perrelli A, 2018; Moglianetti M P. D., 2020).

Genetic diagnosis of familial CCM disease

The sporadic form of CCM disease (sCCM) is characterized by a lack of family history of the disease and usually patients present a single lesion on MRI. In contrast, familial cases (fCCM) mostly exhibit multiple lesions that show progression in both number and size over time. Genetic studies conducted over the past 20 years have identified a wide spectrum of pathogenic mutation types and distribution within the three known CCM genes, including Small Nucleotide Variants (SNVs): nonsense, frameshift, missense, and splice site mutations as well as complex mutations as Copy number Variants (CNVs): large deletions and insertions (Choquet H, 2015; Spiegler S N. J.-S., 2014; Spiegler S R. M., 2018). Moreover, there is also evidence for some alternative splicing variants that may eventually generate protein isoforms with specific subcellular compartmentation and distinct expression patterns among various tissues and cells, suggesting that they abnormal expression may also contribute to CCM disease pathogenesis (Retta SF, 2004; Jiang X, 2019). On the other hand, although most familial cases of CCM have been clearly linked to heterozygous LOF mutations in any of the three known CCM genes, no mutation in either of these genes were found in 5–15% of all CCM cases with a positive family history, suggesting the existence of additional CCM genes (Choquet H, 2015; Spiegler S R. M., 2018). Furthermore, whereas an incomplete penetrance and no clear genotype-phenotype correlations have been observed so far, the wide variability in phenotypes seen among carriers of the same CCM gene mutation suggests the influence of additional genetic and/or environmental modifiers (Choquet H, 2015; Trapani, 2015; Choquet, 2016). Indeed, the clinical behavior in individual patients, including the development of numerous and large lesions, and the risk of serious complications, such as ICH, remain highly unpredictable even among family members of similar ages carrying the same disease-associated genetic defect.

CCM patients

CCM Families

Family 1.

Commentato [MD2]: 1 o 2?

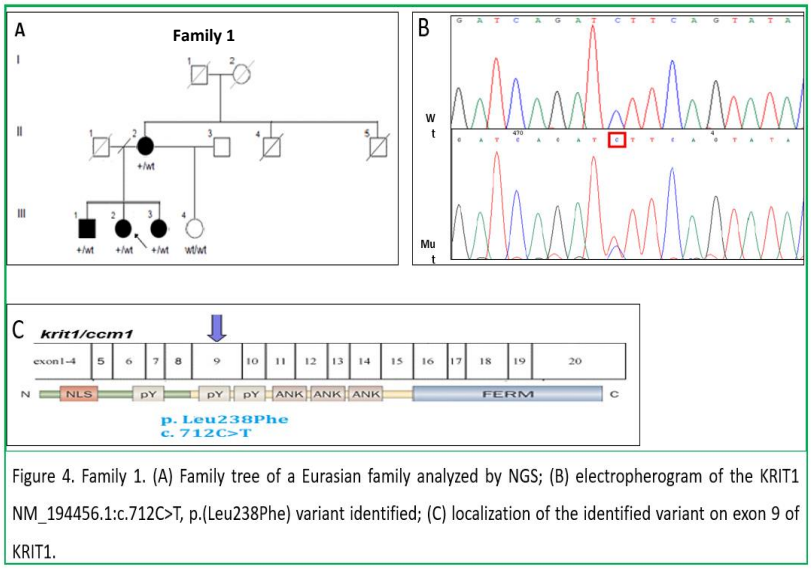

Figure 4. Family 1. (A) Family tree of a Eurasian family analyzed by NGS; (B) electropherogram of the KRIT1 NM_194456.1:c.712C>T, p.(Leu238Phe) variant identified; (C) localization of the identified variant on exon 9 of KRIT1.

Proband (III.2 – Figure 4 A) has been identified through MRI after a seizure event in 2012; she is a young girl whit Eastern Asia origin. MRI was positive for multiple lesions: the main in the left frontal hemisphere, the others in the right frontal hemisphere and homolateral occipital hemisphere. She suffers from seizures, generally in the stage of falling asleep or in awakening stage, and is treated with Clobazam and Levitiracetam, and during the seizure crisis Diazepam. No bleeding events were reported and she has not undergone surgery. The family B is composed by four brothers, three of whom sons (III.1-2-3 – Figure 4 A) of the same father's proband (II.1 – Figure 4 A). Mother's proband reports the death of her two brothers, one of them (II.4 – Figure 4 A) at 12 years old had undergo neurosurgery (he was suffering of cefalea) and he died ten years later; the second (II.5 A) died due to a leukemia. Four family members (III.1-2-3 - Figure 4 A) were affected by CCMs disease and evaluation of non-carriers (subjects III.4, II.3 – Figure 4 A) indicated no signs or symptoms of disease. CCM mutations were inspected in proband III.2 (Family 1 – Figure 4 A) through Sanger sequencing.

Family 2.

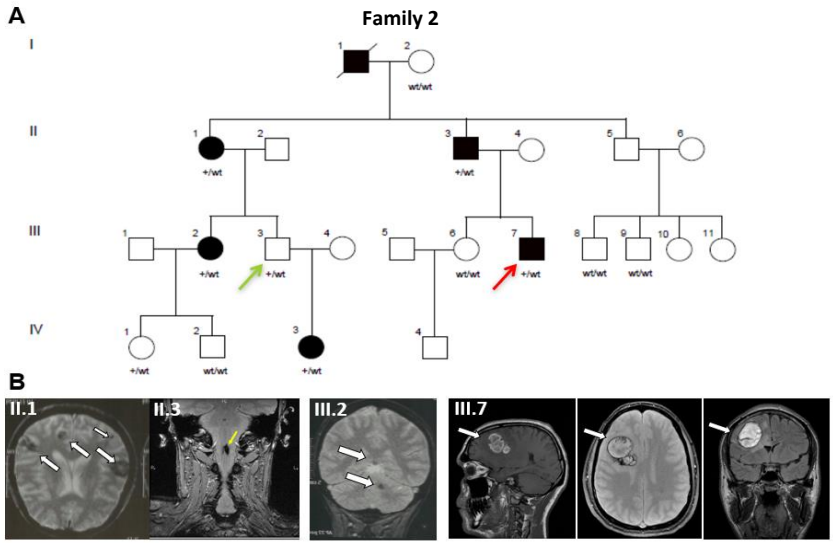

Figure 5. Family tree (A) and MRI (B) of first CCM family. Members carrying the same mutation with a similar genetic background could different disease manifestation.

The second case is a north Italian family (Figure 5). The proband (III.7 – Figure 5) is a male patient that suffered from tonic-clonic seizures and was treated with valproate and carbamazepine. His MRI showed multiple gigantic cavernomas, located in the left periventricular region and left frontal lobe, with signs of mild recent bleeding (III.7 – Figure 5B). No alterations were seen in the spinal cord. Our proband's father (II.3 – Figure 5) also showed neurological symptoms. He presented with a spinal radiculopathy, with sensory symptoms in the upper thigh, and suffered from epilepsy during childhood. Both cerebral and spinal MRI were positive for cavernous malformations. His cerebral MRI showed multiple small cavernomas, up to 8 mm, in the white matter of the right cerebellar hemisphere, and in the right frontal and parietal lobes. Signs of recent micro-bleeding in the parietal lesion were found at gradient echo images. The spinal MRI showed multiple spinal cavernomas localized in the cervical, thoracic and lumbar spinal cord. In his family medical history, our proband reported that his grandfather (I.1 –

Figure 5 A) deceased from a cerebrovascular accident, but no further medical information about him was available. Evaluation of non-carriers (subjects I.2, III.6, III.8, III.9, IV.1 – Figure 5 A) indicated no signs or symptoms of disease. To date, mutations in three genes have been demonstrated to cause familial CCM; KRIT1, CCM2 and PDCD10, then we performed NGS analysis in proband (III.7 – Figure 5 A).

CCM Patients

Here I describe a 48-year-old woman affected by CCM with a known PTEN mutation, manifest: intestinal polyposis, Cowden syndrome, meningioma and multiple CCM lesions. In addition, I report a case of CNV in chromosome 7 and involved KRIT1 gene. Other patients were sequenced during my PhD project and we decided to report the most important. All the mutation identified were reported in table 4.

Commentato [MD3]: Non capisco se questi sono risultati oppure sono una descrizione delle famiglie sulla base dei dati già disponibili. Se sono una descrizione basata su dati già disponibili questa marte va spostata nell'introduzione ad illustrare le due famiglie che poi andrai a caratterizzare sequenziando

Commentato [MD4]: Valerio, questa è introduzione, quindi qui riporti una descrizione delle analisi molecolari con i loro pro ed i loro contro, poi riporti il sequenziamento dell'esoma e dell'esoma clinico, spieghi le differenze e qui chiudi l'introduzione. Quello che descrivi sotto sono Mat&Met e risultati, ridisctribuiscili nelle apposite sezioni e cancella i doppioni

Molecular analyses

Molecular analyses are described in "*Next Generation Sequencing (NGS) Strategies for Genetic Testing of Cerebral Cavernous Malformation (CCM) Disease*" a detailed *Methods in Molecular Biology* book chapter edited by Valerio Benedetti, Elisa Pellegrino, Alfredo Brusco, Roberto Piva, and Saverio Francesco Retta (V Benedetti, 2020).

Clinical exome sequencing (CES) is a new state-of-the-art molecular diagnostic genetic test. It has become one of the most powerful genetic method in clinical practice. In particular, as I mentioned above, a detailed clinical diagnosis and the identification of causative mutations are crucial for genetic counseling of affected patients and their families, for understanding genotype-phenotype correlations and developing therapeutic approaches.

NGS Data analyses

Raw sequence ∗.bcl files generated by the NextSeq 500 instrument were converted to ∗.fastQ files using the bcl2fastq module provided by Illumina.

Sequencing alignment and variant analysis were performed using Sophia DDM® version 5.2 software (Sophia Genetics, Saint-Sulpice, Switzerland). All variants with an alternative variant frequency ≤5%, minor allele frequency (MAF) >1% and/or intronic variants at more than 12bp away from exon-intron boundaries were excluded. Classify variants based on the American College of Medical Genetics (ACMG) Pathogenicity Guidelines (Richards, 2015), assigning each variant a classification as one of the following: 1) pathogenic; 2) likely pathogenic; 3) variant of uncertain significance; 4) likely benign; or 5) benign. The pathogenicity of the mutation identified by CES was evaluated based on the findings in the Database of Genomic Variants (DGV, http://projects.tcag.ca/variation), the ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/), Alamut® Visual (https://www.interactive-biosoftware.com/alamut-visual/), and Human Splice Finder tool (http://www.umd.be/HSF/).

Material and Methods

Nucleic acid extraction

For any NGS platform, accuracy of sequencing depends largely on using high-quality starting DNA that is quantified accurately. The total yield of genomic DNA from whole blood samples depends on the sample volume and number of white blood cells/mL. Whole blood samples collected in EDTA, ACD or heparin can be used. These samples may be either fresh or frozen. Frozen samples should be thawed and mixed before processing. Samples with little available DNA must be extracted using a method that maximizes the amount of DNA recovered but elutes in a small volume. The Maxwell® 16 LEV method has been optimized to process samples in a low elution volume (LEV) automated format to address these concerns. The magnetic particle-based methodology avoids common problems such as clogged tips or partial reagent transfers that result in suboptimal purification processing by other commonly used automated systems. A Maxwell® 16 Blood DNA Purification Kit is sufficient for 48 automated isolations from whole blood samples and must be stored at 15–30°C. For more detailed information is available on Promega Technical Manual at: www.promega.com/protocols.

A flow chart of DNA extraction is report in the following figure (Figure 3)

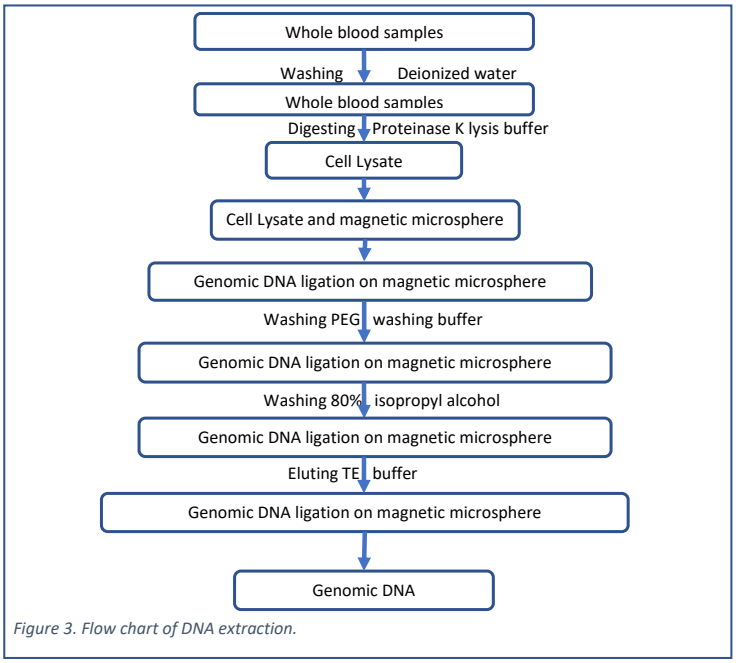

Figure 3. Flow chart of DNA extraction.

Genomic DNA (gDNA) extraction

Genomic DNA was extracted from whole blood samples using the automated Maxwell® 16 DNA Purification Kits set to the LEV (low elution volume) configuration. DNA samples were purified using paramagnetic particles (PMPs), which provide a mobile solid phase that optimizes gDNA capture, washing and elution.

NGS Library Preparation, Hybridization Capture, and Sequencing

Library preparation

Individual libraries were prepared using the DNA Library Prep Kit by SOPHiA^{Tm} GENETICS® (Saint-Sulpice, Switzerland) as well as AMPure XP® beads and individual adapters provided as part of the Custom Solution kit by SOPHiA^{Tm} recommendations. DNA was quantified by Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and Qubit fluorometer (Thermo Fisher Scientific).

Specifically, 200 ng gDNA for sample was enzymatically digested in a total volume of 50 µL for 5 min at 32°C and followed by end repair and A-tailing for 30 min at 65°C and was then immediately placed on ice. Illumina-compatible index adapters were then ligated for 15 min at 20°C in a total volume of 100 µL. Enzymes, buffer, and PCR primers were removed from the ligation reactions using 80 µL AMPure XP beads followed by clean-up with 80% ethanol and elution with 105 µL of IDTE buffer (using DNase/RNase-Free Water). After magnetic separation, 100 µL of the post-ligation clean-up reaction was transferred to a new tube. Dual size selection of 300–700 bp DNA fragments was performed in two steps. To remove fragments larger than 700 bp, 60 µL AMPure XP beads were added to 100 µL of post-ligation clean-up reaction, and 140 µL supernatant was retained after magnetic separation. Then, to remove fragments shorter than 300 bp, this supernatant was mixed with 20 µL fresh AMPure XP beads, cleaned with 80% ethanol and eluted with 23.5 µL of IDTE buffer (using DNase/RNase-Free Water). The adapter-ligated and size-selected libraries were amplified in a volume of 50 µL for 8 PCR cycles: 98°C for 20 s, 60°C for 30 sec, and 72°C for 30 sec. Finally, the libraries were purified using 50 µL AMPure XP beads followed by 80% ethanol (using DNase/RNase-Free Water)) clean-up and elution to 20 µL with nuclease-free water. Individual libraries were quantified using a Qubit dsDNA HS Assay Kit on a Qubit 2.0 fluorometer (Invitrogen, Thermo Fisher Scientific). Library size was verified using capillary electrophoresis (2200 TapeStation, Agilent Technologies, Santa Clara, USA) according to the manufacturer's guidelines. Before capture enrichment, 150 ng of each individual library was pooled to obtain 8 tubes, each containing 12 individual libraries (1.8 µg DNA per tube). To prevent nonspecific binding of the probes, Cot-1 DNA and blocking oligos were also added to the libraries, and the solution was dried using a vacuum DNA concentrator until mix is completely lyophilized (45–50°C). Safe stopping point at -20 °C.

Commentato [MOU5]: Duplicate sentence

Hybridization Capture Using xGenR LockdownR Probes

The procedure for hybridization capture using the xGenR LockdownR Probes starts with dried pooled libraries resuspended in 13 µL hybridization buffer and denatured at 95°C for 10 min. The libraries were then hybridized at 65°C for 4 h by the addition of 4 µL HPD_v1 biotinylated probes (SOPHiA GENETICS). The probes target genomic regions

associated to human diseases with genetic origin or influence. 17 µL of each pool of libraries hybridized to biotin probes was bound for 45 min at 65°C with 100 µL cleaned streptavidin beads. After the binding step, unbound and nonspecific DNA was removed using two stringent washes at 65°C followed by several washes at room temperature, and the beads were resuspended in 20 µL nuclease-free water in accordance with the manufacturer's recommendations. The post-capture pooled libraries were amplified for 14 cycles (98 °C for 15 s, 60°C for 30 s, and 72°C for 30 s) using KAPA HiFi HotStart ReadyMix and Illumina-compatible library amplification primer mix (Kapa/Roche, Basel, Switzerland) in a total volume of 50 µL. After PCR amplification, the pooled libraries were purified using 50 µL AMPure XP beads (Beckman Coulter) and washed using 200 µL 80% ethanol (using nuclease-free water). The libraries were then quantified using the Qubit dsDNA HS Assay Kit on the Qubit 2.0 fluorometer (Invitrogen, Thermo Fisher Scientific). The size of the captured libraries was analyzed using capillary electrophoresis (2200 TapeStation, Agilent Technologies, Santa Clara, USA) according to the manufacturer's guidelines.

Library DNA fragments should have a size distribution between 300 bp and 700 bp.

NGS Sequencing

Each pooled library obtained was diluted to a final concentration of 4 nM and denatured with 0.2 N NaOH (using nuclease-free water). Equimolar concentrations of the 8 pools of the 12 individual libraries were diluted and mixed to a final concentration of 10 pM (using nuclease-free water). The pooled libraries were combined with 1% PhiX Control before loading onto a NextSeq 500 instrument (Illumina, San Diego, USA). Sequencing was performed with a 300 cycle mid-output kit v2. Sequencing of all the protein-coding genes Genomic DNA from the affected individual was sequenced to Clinical Exome Sequencing (CES).

Clinical Exome sequencing protocol is described elsewhere by SG (Sophia Genetics, Saint-Sulpice, Switzerland).

SOPHiA^{Tm} Clinical Exome Sequencing cover the coding regions (±5bp of exon-flanking regions) of 4,490 genes associated with the vast majority of inherited diseases. CES - SG

is characterized by probe design optimized to provide high coverage uniformity throughout the entire target region, resulting in valuable data quality. Indeed, achieve very high on-target rates, ensure reliable coverage uniformity values across all the target regions, even in GC-rich regions. Equal read coverage is of crucial importance for the precise identification of multiple types of variations (mainly for small indels), including CNVs. Data obtained from clinical exome sequencing could be analyzed using SOPHiA^{Tm} DDM® software, the platform of choice for performing genomic testing. The interface and advanced features facilitate the visualization and interpretation of genomic variants. Data is kept safe by applying the highest industrial standards of encryption. SOPHiA^{Tm} DDM® software offers several dedicated features as reduces the complexity of determining the significance of genomic variants and facilitates the interpretation process, thus reducing turnaround time; a dual variant pre-classification: Improve assessment of variants pathogenicity with the pre-classification of both ACMG guidelines and SOPHiA's prediction and the use of Virtual Panels that restricts the interpretation of CES genes to sub-panels of genes of interest.

Sanger Sequencing

The pathogenic mutations were sequenced and confirmed by both PCR amplification and Sanger sequencing and was analyzed on an ABI PRISM 3130xl instrument (Applied Biosystems, Thermo Fisher, Waltham, USA). In particular, Sanger sequencing was performed to confirm CES data for proband (III.7 – Figure 5 A) of family 2. We designed a set of primers specific for exon 16 of KRIT1 (NM_194456.1) using the KAPA 2G Fast HS PCR kit (Kapa Biosystems). Amplification was performed in a total volume of 12 µl containing a mix of: KAPA buffer A (5X with Mg), KAPA Fast polymerase, dNTP mix and 1 µL (100 pM/µL) of forward (5'- gcttttcccatattaagttgttcatt - 3') and reverse (5'- ttggttaaacagaatcttaagcatagc - 3') primers and DNA samples (50ng/µL). The PCR was performed on a 7500 fast apparatus (Life Technologies) following touchdown protocol: 2 min at 94°C, 20 cycles 15 sec 94°C, 15 sec 64° C and 1 min 72° C, followed by 14 cycles 15 sec 94°C, 15 sec 64°C to 57°C (-0.5°C for each cycle) and 1 min 72°C, ending with extension temperature 72°C for 7 min (protocols.io: https://dx.doi.org/10.17504/protocols.io.bd55i886). Real time PCR was performed to

confirm exon 16 of KRIT1 exclusion. In second family, CCM lesions were confirmed by MRI and in a second time, CCM genes were analyzed by Sanger sequencing in proband (III.2 Figure 4 A). In particular, amplification of exon 9 mutated sequence of KRIT1 (NM_194456.1) was performed in a total volume of 12µl containing a mix of: KAPA buffer A (5X with Mg), KAPA Fast polymerase, dNTP mix and 1 µL (100 pM/µL) of forward (5'- atgaacatattatcacatgtgc - 3') and reverse (5'- agccactgtaccaggccttcatg - 3') primers and DNA samples (50ng/µL). Afterward, mutation was confirmed in all affected family members (Figure 4 A). The genetic exams were performed at the Medical Genetics Unit of the "Città della salute e della Scienza di Torino". Notch3 NM_000435.2 variant was confirmed by Sanger sequencing on exon 18 using common PCR protocol; amplification was in a total volume of 12µl containing a mix of: KAPA buffer A (5X with Mg), KAPA Fast polymerase, dNTP mix and 1 µL (100 pM/µL) of forward (5'- agcttcttgccgagataaggg - 3') and reverse (5'- cgacccccgttttgacaagg - 3') primers and DNA samples (50ng/µL).

CNV

To date, the best practice for CNV detection often require using NGS analysis for detecting large CNVs and the bioinformatic analysis applied to sequencing data. Therefore, CES was performed to discover a case of a CCM disease-associated CNV in a symptomatic 46-year-old patient, and to verify CNV following by real-time qRT-PCR validation using Applied Biosystems® TaqMan® RNase P Detection Reagents kit and Universal ProbeLibrary (UPL) probes (Roche Molecular Systems®) and relative recommended protocols (www.universalprobelibrary.com). In particular, I performed qRT-PCR in PCR Reactions Mix in total volume of 11µl containing: Master Mix TaqMan®, forward and reverse primers for each gene and corresponding UPL probe as show in table 1 or RNase P Primer-Probe Mix, Nuclease-free Water and cDNA Human Genomic DNA (30ng/µL). UPL probes were drawn using Universal ProbeLibrary Assay Design Center® software.

Table 1.

Gene	KRIT1	NM_194456.1			
Ex-Int	Int4				
Use probe #72 (cat. no. 04688953001)					
Primer	**Length**	**Position**	**Tm**	**%GC**	**Sequence**
Left	25	220-244	59	36	tgaacagtaaagatgcagtcaaagt
Right	27	320-346	59	22	ttgaaaactttatttccacataatca
Amplicon (127 nt)					
tgaacagtaaagatgcagtcaaagtcattcctggcttgccttttaggcactgtatattctaggatagagaacattttatttttttttaatttttcgacatgattatgtggaaataaaagttttcaa					

Gene	KRIT1	NM_194456.1			
Ex-Int	Ex13				
Use probe #84 (cat. no. 04689089001)					
Primer	**Length**	**Position**	**Tm**	**%GC**	**Sequence**
Left	21	371 - 391	59	38	tctgttgaattgaagcatgga
Right	21	422 - 442	59	48	ttcttgagagagacgcattcc
Amplicon (72 nt)					
tctgttgaattgaagcatggaaataataccacagtgcagcagataatggaaggaatgcgtctctctcaagaa					

Gene	KRIT1	NM_194456.1			
Ex-Int	Int19				
Use probe #54 (cat. no. 04688511001)					
Primer	**Length**	**Position**	**Tm**	**%GC**	**Sequence**
Left	23	531-553	60	39	cccaatgtcatgaatgtcccttta
Right	27	614-640	59	33	ccatttagcttcattaacagttttacc
Amplicon (110 nt)					
cccaatgtcatgaatgtcctttaaagattcttacatgttttatatttattattgacatgttttactttgtaggctggtctcgtggtaaaactgttaatgaagctaaatgg					

Gene	AKAP9	NG_011623.1			
Ex-Int	Int16				
Use probe #18 (cat. no. 04686918001)					
Primer	**Length**	**Position**	**Tm**	**%GC**	**Sequence**
Left	23	46-68	59	39	tctttttgtttagaaccgtgagg
Right	27	144-170	59	22	aaagaaataaaagcaaacaataaagga
Amplicon (125 nt)					
tctttttgtttagaaccgtgaggccatgctttttttctgtttatacaattttgagtactaatttactttttaaattcctgctgatatktctaaatctaatcctttattgtttgctttttatttcttt					

Gene	AKAP9	NG_011623.1			
Ex-Int	Int38				
Use probe #7 (cat. no. 04685059001)					
Primer	**Length**	**Position**	**Tm**	**%GC**	**Sequence**
Left	24	630-653	60	46	gaatatagccagtatcccaacagc
Right	27	689-715	59	30	agaaaaatcagaacaatgattacctct
Amplicon (86 nt)					
gaatatagccagtatcccaacagccccttctcccaagcacatgctcctctgagtaacttagaggtaatcattgttctgatttttct					

Gene	CYP51A1	NM_000786.3			
Use probe #52 (cat. no. 04688490001)					
Primer	**Length**	**Position**	**Tm**	**%GC**	**Sequence**
Left	20	84-103	59	55	accatctgccagcttctctc
Right	20	144-163	60	50	ttcactccgtcggaaacact
Amplicon (80nt)					
accatctgccagcttctctcgttccgtcgattgggaggagcggtggcgacctcggccttcagtgtttccgacggagtgaa					

Gene	LRRD1	NM_001161528.1			
Use probe #6 (cat. no. 04685032001)					
Primer	**Length**	**Position**	**Tm**	**%GC**	**Sequence**
Left	26	233-258	59	38	ccagatttatgaaacacatcctagac
Right	27	291-317	59	33	ttgttcttcatttctcttagacttcct
Amplicon (85nt)					
ccagatttatgaaacacatcctagacagaatacattagagtcaacatcttcctctggaaggaagtctaagagaaatgaagaacaa					

Gene	ANKIB1	NM_019004.1			
Use probe #22 (cat. no. 04686969001)					
Primer	**Length**	**Position**	**Tm**	**%GC**	**Sequence**
Left	20	1219-1238	59	45	ccaacgatcaggtgttcaaa
Right	20	1261-1280	60	55	gcgtgtcccaggcattatac
Amplicon (62 nt)					
ccaacgatcaggtgttcaaatgccaactccaccaccaagtgggtataatgcctgggacacgc					
Use probe #1 (cat. no. 04684974001)					
Primer	**Length**	**Position**	**Tm**	**%GC**	**Sequence**
Left	20	4280-4299	59	50	caaggcatttaggtgtgcag
Right	20	4324-4343	60	55	ttactcccgagaacgctgac
Amplicon (64 nt)					
caaggcatttaggtgtgcagcatcctagagtgctccagggcagtgtcagcgttctcgggagtaa					
Gene	TMBIM7P	NR_145992.1			
Use probe #47 (cat. no. 04688074001)					
Primer	**Length**	**Position**	**Tm**	**%GC**	**Sequence**
Left	18	228-245	60	61	gcctctgccccttctgat
Right	21	285-305	59	57	ctgagaggcactagaggatgc
Amplicon (78 nt)					
gcctctgccccttctgatttgcaagacactggagaggaaatcaacttggttatctgagcatcctctagtgcctctcag					
Use probe #12 (cat. no. 04685113001)					
Primer	**Length**	**Position**	**Tm**	**%GC**	**Sequence**
Left	19	1291-1309	59	58	ggtagtcacgcaaccaagg
Right	19	1359-1377	59	58	gacttctgcatgcctgagc
Amplicon (87 nt)					
ggtagtcacgcaaccaaggctggctcgtgcggagaagaggaaggaggggtgctgtggagtgttgccgggctcaggcatgcagaagtc					
Gene	GATAD1	NM_021167.4			
Use probe #2 (cat. no. 04684982001)					
Primer	**Length**	**Position**	**Tm**	**%GC**	**Sequence**
Left	20	3960-3979	60	50	gtgtgatggttcccattgct
Right	20	4005-4024	59	55	acagaccagggccatatcag
Amplicon (65 nt)					
gtgtgatggttcccattgcttatatttctcctgctgaggccggacctgatatggccctggtctgt					

Table 1. Validation of CNV: primers and UPL probes used for qRT-PCR

The complete procedure is performed in a Medical Genetics Unit, AOU Città della Salute e della Scienza di Torino, Turin, Italy.

Commentato [MD6]: Questa non è introduzione, ma Mat&Met che dovrebbero essere organizzati meglio e soprattutto espansi, a meno che UniTO non preveda Mat&Met così stringati che comunque per me non sono sufficienti. Dire che sequenzi con un Nextseq non vuol dire niente

Results

Commentato [MOU7]: I would suggest you start with more systematic overview of your results. This section is somewhat confusing.

During my PhD course, I studied the CCM manifestations in patients with similar or heterogeneous genetic background finalized to understand the mechanisms underlying the high variable expressivity of CCM disease. Clinical information from all patients involved in the study was collected by experienced neurologists and clinical geneticists, and at the same time I collected all the DNA samples to create a cohort of patients. Informed consent was obtained from patients or their legal representative.

I Started sequencing CCM patients using Sanger sequencing, as I will describe in family two (Family 2 - Figure 5), and I have been continuing my PhD project supported by a powerful diagnostic genetic method (CES). Therefore, I took care of patients classification, quantification samples, following NGS sequencing the raw data analyses and genetic confirmation of causative disease variants.

In particular, I analyzed a cohort of 25 patients affected by CCM disease, some of these previously sequenced with Sanger sequencing technique (Family 2 - Figure 5) and without a causative variant of disease.

Sanger sequencing was used for SNVs detection, and we decided to test negative samples for Sanger sequencing using NGS to discover possible undetected SNVs or complex structural variant as CNVs.

The experimental work allows to identify causative disease variant (SNVs) or CNVs in the 80% of analyzed cases as I report in table 4. In particular, in this thesis I report:

a novel variant c.1664C>T on exon 16 of KRIT1 gene (NM_194456.1) in six family members (family 2 – Figure 5), a known causative variant: the KRIT1 NM_194456.1:c.712C>T in four family members of first family (family 1 – Figure 4); a deletion of 47.13 Kb on chromosome 7: the (NC_000007.14:g.(?_ 92198969)_(92246100 _?)del), involving two neighboring genes, KRIT1 and AKAP9, including heterozygous deletion of all the 16 coding exons of KRIT1 and the last 13 coding exons (from 38 to 50) of AKAP9, this patient was resulted negative for Sanger sequencing as I will explain subsequently and in figure 6; variant with “uncertain significance”: the NM_000435.2(NOTCH3):c.2960G>C correlated to CADASIL-1 disease in patient affected

by CCM disease (Figure 8); a pathological variant: c.959dupT of exon 8 of PTEN (NM_ 000314.6).

Overall, the experimental outcomes demonstrated that the application of optimized NGS workflows based on clinical gene panels, including genes and genetic variants involved in distinct cerebrovascular diseases, made it possible to identify novel causative variants in known CCM genes, as well as other potential causative genes and genetic risk factors associated to CCM disease onset and severity, thus providing useful insights into the improvement of current diagnostic procedures for a more effective diagnosis, prognosis and clinical management of CCM disease.

Genetic analyses

DNA extraction was performed by Valerio Benedetti at Medical Genetics Unit, Città della Salute e della Scienza University Hospital, Turin.

NGS Library Preparation, Hybridization Capture, and Sequencing were performed by Dott. Elisa Pellegrino and Valerio Benedetti at Center for Experimental Research and Medicine Studies (CERMS), Azienda Ospedaliera Città della Salute e della Scienza di Torino, Turin, and Molecular Biotechnology Center (MBC), University of Turin, Turin.

Sanger Sequencing and Real time qPCR were performed by Valerio Benedetti at Medical Genetics Unit, Città della Salute e della Scienza University Hospital, Turin.

NGS data were analyzed by Valerio Benedetti with the support of Dott. Elisa Giorgio, Prof. Roberto Piva and Porf. Alfredo Brusco.

In this thesis, I performed a sanger analysis and identified a causative mutation in family 1. In a second time, NGS was introduced for discovery of pathogenetic variants. In particular, NGS data analysis using SOPHiA^{Tm} DDM® version 5.2 software, was performed at three different depth levels: 1- restricted to CCM genes, 2- including cerebrovascular malformations genes, 3- extended to all genetic variants:

1- It was limited to the identification of possible mutation occurring on the three causative genes: Krit1/CCM1, mgc4607/CCM2 and pdcd10/CCM3.

2- Consequently, in case of negative outcome, the second level took advantage of NGS virtual panel: "*Cerebro*" containing the 23 genes associated to all cerebrovascular malformations (ABCC6, ACTA2, ACVRL1, APP, BRCC3, CCM2, COL3A1, COL4A1, COL4A2, ENG, FBN1, GLA, HTRA1, JAM3, KRIT1, MYH11, MYLK, NOTCH3, PDCD10, SAMHD1, TGFBR1, TGFBR2, TREX1). Genetic variants were selected through a virtual panel that filtered the data.

3- In case that no mutations were detected, the last level was a broader analysis of all variants obtained from clinical exome sequencing (CES) who analyzed more than 4400 genes.

In particular, genetic variants were subdivided by pathogenic classes: highly pathogenic (A), potentially pathogenic (B), unknown significance (C) and likely benign (D).

First level was limited to CCM-related genes to discover possible causative or uncommon variant; the second level ("*Cerebro*" gene panel) refers 23 cerebrovascular-related genes associated to cerebrovascular genetic disease and it is composed by followed genes reported in next table (Table 2):

Gene	Omim ID	Disease
ABCC6	603234	Pseudoxanthoma elasticum and arterial calcification generalized of infancy 2 – GACI2
ACTA2	102620	Multisystemic smooth muscle dysfunction syndrome - MSMDS and aortic aneurysm familial thoracic 6 - AAT6
ACVRL1	601284	Telangiectasia hereditary hemorrhagic Type 2 - HHT2 and pulmonary hypertension primary 1 - PPH1
APP	104760	Cerebral amyloid angiopathy app-related - CAA-APP and Alzheimer disease - AD
BRCC3	300617	Moyamoya disease 4 with short stature hypergonadotropic hypogonadism and facial dysmorphism - MYMY4 and T-Cell prolymphocytic leukemia - T-PLL
CCM2	607929	Cerebral cavernous malformations 2 - CCM2 and cerebral cavernous malformations - CCM
COL3A1	120180	Ehlers-Danlos syndrome vascular type - EDSVASC and polymicrogyria with or without vascular-type Ehlers-Danlos syndrome - PMGEDSV
COL4A1	120130	Retinal arteries tortuosity of - RATOR and angiopathy hereditary with nephropathy aneurysms and muscle cramps - HANAC
COL4A2	120090	Brain small vessel disease 2 - BSVD2 and intracerebral hemorrhage - ICH
ENG	131195	Telangiectasia hereditary hemorrhagic type 1 - HHT1 and hereditary hemorrhagic telangiectasia - HHT
FBN1	134797	Marfan syndrome - MFS and stiff skin syndrome - SSKS
GLA	300644	Fabry disease - FD and rare cardiomyopathy
HTRA1	602194	Cerebral arteriopathy autosomal recessive with subcortical infarcts and leukoencephalopathy - CARASIL and cerebral arteriopathy autosomal dominant with subcortical infarcts and leukoencephalopathy type 2 - CADASIL2
JAM3	606871	Hemorrhagic destruction of the brain subependymal calcification and cataracts - HDBSCC and Jacobsen syndrome - JBS
KRIT1	604214	Cerebral Cavernous Malformations and Cerebral Cavernous Malformations type 1- CCM1
MYH11	160745	Aortic aneurysm familial thoracic 4 - AAT4 and familial thoracic aortic aneurysm and aortic dissection - FAA
MYLK	600922	Aortic aneurysm familial thoracic 7 - AAT7 and megacystis-microcolon-intestinal hypoperistalsis syndrome - MMIHS
NOTCH3	600276	Cerebral arteriopathy autosomal dominant with subcortical infarcts and leukoencephalopathy type 1 - CADASIL1 and lateral meningocele syndrome - LMNS
PDCD10	609118	Cerebral Cavernous Malformations 3 – CCM3 and Cerebral Cavernous Malformations - CCM
SAMHD1	606754	Aicardi-Goutieres syndrome 5 - AGS5 and Chilblain Lupus 2 - CHBL2
TGFBR1	190181	Multiple Self-Healing Squamous Epithelioma – MSSE and Loeys-Dietz Syndrome 1 - LDS1
TGFBR2	190182	Loeys-Dietz syndrome 2 – LDS2 and colorectal cancer hereditary nonpolyposis type 6 - HNPCC6
TREX1	606609	Vasculopathy retinal with cerebral leukodystrophy - RVCL and Aicardi-Goutieres syndrome 1 – AGS1

Table 2. Cerebro gene panel: composed by all genes associated to cerebrovascular genetic diseases.

The last level includes all the 4490 CES genes (selected by SOPHiA^{Tm} Genetics ®) associated with the vast majority of inherited diseases. Genetic raw data (∗.bcl) obtained from CES analysis were converted to ∗.fastQ files using the bcl2fastq module provided by Illumina, sequencing alignment and variant analysis were performed using SOPHiA^{Tm} DDM® software, data filtering was performed by software that report a vast among of variants and applied supplementary filters as: variant with frequency rate less or equal to 0.1% in ExAc database (Exome Aggregation Consortium, it is a coalition of investigators seeking to aggregate and harmonize exome sequencing data from a variety of large-scale sequencing projects) (Karczewski KJ, 2017), less or equal to 0.01% in esp5400 database (Exome Variant Server, NHLBI GO Exome Sequencing Project http://evs.gs.washington.edu/EVS/2017), less or equal to 0.01% in the 1000 genomes database (1000 Genomes Project European Cohort - 1KGP, a database of genomes 26 of different populations) (Auton A, 2015), less or equal to 0.01 in GnomAD database (The Genome Aggregation Database, a data set contains data from 125,748 exomes and 15,708 whole genomes, all mapped to the GRCh37/hg19 reference sequence) (Karczewski, 2020), for frequency in the pool samples less or equal to 4, and last one parameter is frequency in SOPHiA^{Tm} DDM® software database less or equal to 100.

CES generates data on large numbers of genetic variants and these unfiltered data are in the average of 30500 variants for each sample (Table 4), and retained variant result from software are in the average of 6600, many of which, in the context of a rare genetic disease study, are not relevant since they are very unlikely to have a functional effect at the protein and/or systemic level. Hard variant filtering, as I said in supplementary filters, consists of choosing specific thresholds for one or more annotations and throwing out any variants that have annotation values above or below the set thresholds. Filtering strategies inevitably starts with multiple logical assumptions regarding the properties of causative variants, which when combined with more stringent-parameters on automation and predictive tools may either discard related-disease variant too early or produce false positives. A multilevel analysis allows to reduce the false negative risks and to identify with a high probability the possible pathological cause and the relative genetic modifiers (Choquet, 2016). In particular, virtual panels and stringent-frequency parameters can provide a selection of candidate

Commentato [MD8]: Nessuna referenza? Sulla base di cosa sono stati selezionati? Sono tutti quelli che sono nell'esoma clinico?

Commentato [MD9]: G1000?

Commentato [MD10]: Riporta I valori medi, oppure min max

Commentato [MD11]: idem

variants as for CCM samples after appropriate filtering: 30500 variants (average value) obtained from CES analysis to 130 variants for Cerebro virtual panel and 8 variants related to CCM genes (Table 4). Notably, read depth (or coverage, conventionally a number followed by "×") is the number of independent reads with overlapping alignment at a locus of interest. This is often expressed as an average or percentage exceeding a cutoff over a set of intervals (such as exons, genes, or panels). For example, in my case a clinical report might say that a clinical exome or virtual panel 150x for median coverage with max depth of 1,500×, minimal coverage of 5x and a coverage uniformity of 98% for targeted bases covered >100×. SOPHiA^{Tm} Genetics ® data details recommended a sample coverage per run >50x coverage depth, and in this work, I consider 25x as "reliable coverage", because sequencing depth of 15-20x as an appropriate compromise of singleton detection power and sample size for studies of rare variants in complex disease (Rashkin S, 2017).

Commentato [MD12]: idem

All causative SNVs and CNVs identified with NGS were confirmed by means of a traditional Sanger method or real-time PCR on a second genomic DNA extraction. The analytical approach will be considered satisfactory if the new generation sequencing will demonstrate the presence of mutations in 100% of the positive controls (or at least in 100% of the cases with point mutations).

Genetic variants identification

The novel variant c.1664C>T on exon 16 of KRIT1 gene (NM_194456.1), which was absent in ClinVar database, has been identified in family 2 through CES analysis. The cytosine for guanine substitution in position 1664 is predicted to results in two potential effects, as follow:

1) Alanine substitution with Valine at codon 555: p.Ala555Val. The alanine in position 555 is a highly conserved amino acid residue in the evolution of proteins. Conservative replacement of a non-polar with a polar amino acid is considered with low relevance score (Blosum62 score 0). Bioinformatics predictions supported the likely pathogenic effect (damaging substitution for PROVEAN, SIFT, FATHMM, DANN,

DEOGEN2, MVP, PrimateAI, MetaSVM and MetaLR; deleterious for LRT; disease causing for MutationTaster).

2) Four different in silico prediction described the replacement c.1664C>T as new exonic donor site in exon 16, with the consequent loss of 68 amino acids and the generation of a new stop codon in exon 17. This event could be involved in partially exclusion of exon 16, leading to a premature stop-codon.

The Human Genome Variation Society nomenclature classified the mutation as: cDNA level: NM_194456.1: c.1664C>T; gDNA level: Chr7(GRCh37): g.91843991G>A; protein level: p.(Ala555Val). Pathogenicity clues was defined form several databases: highly conserved nucleotide (phyloP: 6.18 [-14.1;6.4]); highly conserved amino acid, up to C. elegans (considering 12 species); small physicochemical difference between Ala and Val (Grantham dist.: 64 [0-215]); this variant is in protein domains: FERM central domain, FERM domain and Band 4.1 domain; align GVGD: C0 (GV: 208.40 - GD: 65.28); SIFT: Deleterious (score: 0, median: 3.44); MutationTaster: disease causing (p-value: 1). In addition, substitution of C with T induces an activation of a new splice site in position c.1662, which was confirmed using Human Splicing Finder tool http://www.umd.be/ (Desmet, 2009), with a splicing variation of +40.4%.

Proband of family 2 (III.7 – Figure 5) carried c.1644C>T variant on KRIT1, was analyzed with a CES analysis, NGS data show a max depth of 2,503×, minimal coverage of 5x, target region observed coverage of 75x with a correlate 25x coverage border of 99,52% and a coverage uniformity of 98% for targeted bases covered >100× (Figure 7). NGS data analyses identified 30260 variants on clinical exome sequencing test, 6632 retained variants from SOPHiA^{Tm} DDM® database, 336 variants associated to Cerebro virtual panel and 61 retained variants, using additional filters, as I mentioned above, only one variant results: c.1644C>T variant on KRIT1 (or c.1520C>T NM_001013406.1, p.(Ala507Val)). CCM virtual panel discovers 10 variants, 4 of these retained and one using hard-filtering (c.1644C>T variant on KRIT1).

Commentato [V13]: Supplementary in quanto vorrei chiederle se secondo lei può essere utile inserire questa figura.

Commentato [MD14R13]: ok

Commentato [V15]: Viene identificata dal software mediante un altro nm e di conseguenza il nome della variante è differente rispetto all'nm utilizzato da noi per disegnare i primers e per confermare tale mutazione.

Commentato [MD16R15]: Spiegalo nel testo

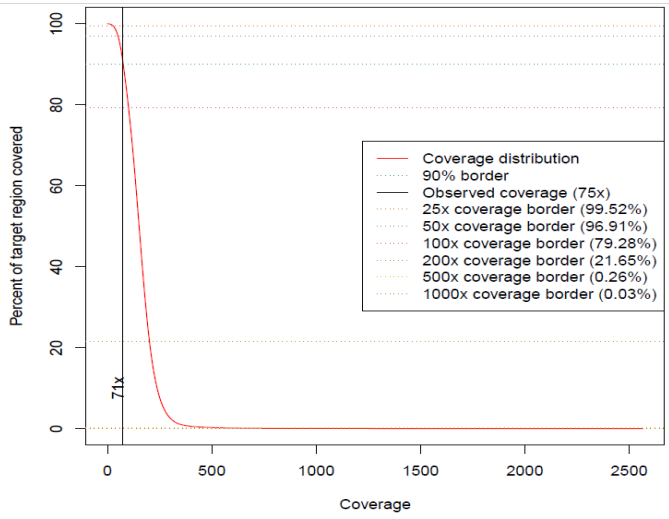

Figure 1.8: Target region coverage.

Table 1.4: Percentage of target regions with certain coverage.

Sample ID	25×	50×	100×	200×	500×	1000×	Coverage 10% quantile	Coverage Heterogeneity
210030-30-S12	99.52%	96.91%	79.28%	21.65%	0.26%	0.03%	75×	0.83%

Figure 7: Proband of family 2, Sequencing and coverage information. Percent of target region covered: 71x and Percentage of target regions with certain coverage: 25x →99.52%

Genetic variants were automatically filtered through SOPHiA^{Tm} DDM® software to increase the detection rate and minimize the rate of false positives only for a pre-classification of significant variants identified as "retained variants".

Several oxidative stress and inflammation related genetic markers could be associated to CCM1 disease, such as SNPs of CYP and MMP families that may impact the severity of CCM1 disease, including the development of numerous and large CCM lesions and ICH. Indeed, two variants were discovered on CYPs genes after hard-filtering: c.475A>T on CYP4F12 (NM_023944.3, exon 5, fraction variant rate of 39%) p.(Ile159Phe); c.108T>A on CYP7A1 (NM_000780, exon 2, fraction variant rate of 50%) p.(Asn36Lys).

The causative mutation in family 2 was identified as the c.712C>T on exon 9 of KRIT1 gene (NM_194456.1). This mutation has also been reported previously (Fisher OS, 2014; Flávio Domingues, 2008). The Human Genome Variation Society nomenclature classified the mutation as: cDNA level: NM_194456.1:c.712C>T; gDNA level: Chr7(GRCh37):g.91864734G>A; Protein level: p.(Leu238Phe). Pathogenicity clues was defined form several databases: highly conserved nucleotide (phyloP: 5.53 [-14.1;6.4]); highly conserved amino acid, up to Zebrafish (considering 12 species); small physicochemical difference between Leu and Phe (Grantham dist.: 22 [0-215]); Align GVGD: C0 (GV: 21.82 - GD: 0.00); SIFT: Tolerated (score: 0.39, median: 3.44); MutationTaster: disease causing (p-value: 1).

Another Italian CCM patient (referring to CNV - Figure 6) was analyzed using Sanger sequencing and after using CES and results negative for causative SNVs. NGS data show a high number of variants: 29638 with a max depth of 1105x, a observed coverage of 82x and percentage of target regions with 25x at 99.44%, 6684 variant of total were retained from database, applying Cerebro and CCM virtual panels and stringent filtering were resulted respectively only one variant associated with first virtual panel: FBN1 c.8176C>T (NM_000138, exon 65, fraction variant rate of 50%). Additional variants was identified on CYP2A6(NM_000762): c.1312A>T p.(Asn438Tyr) with variant rate fraction of 39%. No pathological variant on CCM genes were found using CES and gene panels, but a CNV case was identified on chromosome 7: a deletion of 47.13 Kb (NC_000007.14:g.(?_ 92198969)_(92246100 _?)del). Structural alteration on long arm of chromosome 7 (Ch.7q) was previously identified by Muscarella et al, (Muscarella LA, 2010) in a similar patient with a complete loss of 7q21.2 locus. The structural alteration includes: CYP51A1 (found in a CCM family (Monika Lewinska, 2014) and could be effects

Commentato [MOU17]: Which family?

on disease severity (Choquet, 2016)), LRRD1 (the codified protein is a positive regulation of Ras protein signal transduction and could play an important role in correlated signaling as actin cytoskeletal integrity, cell proliferation, cell differentiation, cell adhesion, apoptosis, and cell migration), AKAP9 and KRIT1.

The c.2960G>C variant in the NOTCH3 gene (NM_000435.2) with "uncertain significance" was identified using Cerebro virtual panel on CES data (Figure 8) after CCM genes analysis where no variants were resulted after hard-filtering. Variant fraction results in 44% with a coverage of 193 depth reads. Sample CES variants were 28971 and 6606 were retained, through Cerebro virtual panel analysis results three NOTCH3 variants: c.4333G>T exon 24 of NOTCH3 with a variant fraction of 44% and 214 of reads depth, missense variant causes an amino acid substitution p.(Ala1445Ser); a synonymous c.771C>A exon 5 of NOTCH3, variant fraction of 57% with 162 reads depth; the last one is the c.2960C>G exon 18 of NOTCH3, variant fraction of 44% and 193 reads depth. This variant was confirmed using Sanger sequencing (as mentioned above and show in follow Figure 8).

Figure 8.

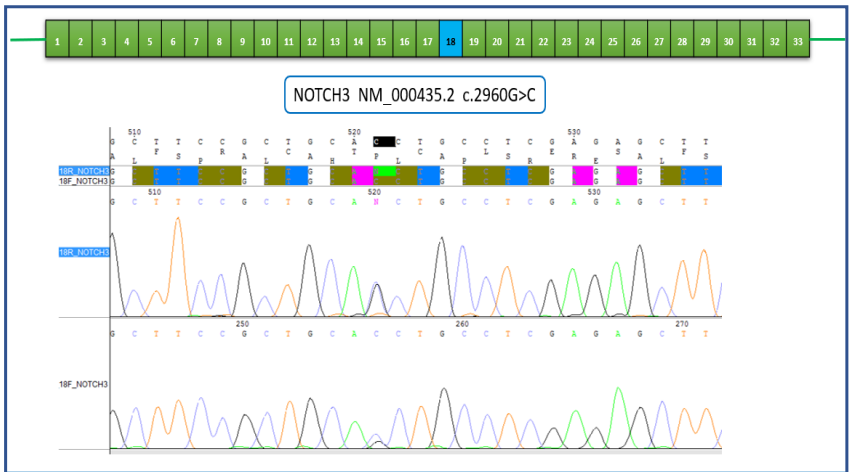

Figure 8. NM_000435.2(NOTCH3): c.2960G>C, a novel variant with "uncertain significance" identified on exon 18 of the NOTCH3 gene using virtual panel on CES data and confirmed by Sanger sequencing (electropherogram).

PTEN (NM_ 000314.6) pathological variant: c.959dupT was discovered by Sanger sequencing and confirmed using CES technology in patient with complex pathological situation due to the concomitant occurrence of multiple cavernomas, Cowden Syndrome, and meningioma. Variant was found to be a heterozygote with 41% of variant fraction, a depth reads of 119 and was classified by SOPHiA^{Tm} DDM® software as potentially pathogenetic; PTEN was not included in Cerebro virtual panel. A deepen analysis not only limited on cerebrovascular diseases genes and hard-filtering didn't show PTEN pathological variant, consequently it is very important discover possible incidental findings following the standard guidelines (Kalia SS, 2017). 30271 variants were obtained from CES and no CCM-associated variants were identified in this samples.

Notably, variants obtained from DNA samples sequenced by CES technology were considered attenable with a sequencing coverage upper than 20X (Rashkin S, 2017).

Family 1.

As I mentioned above, this family was not sequenced with NGS. Sanger sequencing identified a mutation on exon 9 of KRIT1/CCM1 NM_194456.1 gene, afterwards sequenced in all available family members. Missense mutation was transmitted from the mother (II.2 – Figure 4 A) to the offspring (III.1, III.2, III.3- Figure 4 A).

Family histories were remarkable for cerebral cavernous malformation disease type 1 (fCCM1)

Family 2.

CES analysis identified a missense mutation on exon 16 of KRIT1/CCM1 NM_194456.1 gene (NP_919438.1) in several family members. To validate the mutation identified by CES, we performed Sanger sequencing in all available subjects of the family. Missense mutation was inherited from the grandfather (I.1 – Figure 5 A), and was present in the offspring (II.1, II.3 – Figure 5 A) and in the nephews (III.2, III.3, III.7, IV.2 and IV.33 – Figure 5 A).

Other CCM cases.

Clinical Exome Sequencing method provides the ability to detect and report both SNVs and structural variants as the copy number variants (CNVs). Evidence is accumulating that CNVs play important roles in human disease and I report a case of a CCM disease-associated CNV in a symptomatic 46-year-old patient. Using the experimental procedure described in chapter of Methods in Molecular Biology (V Benedetti, 2020), we identified distinct germline genetic variations of pathogenic significance in DNA samples from Italian patients with multiple CCM lesions on MRI examination and/or a positive family history of CCM disease. Herein, I briefly report a result of NGS analysis carried out in a 46-year-old female patient with symptomatic CCM disease and earlier negative Sanger sequencing results for pathogenic mutations. The DNA sample from the case study displayed a mean coverage of 82X, with 99.44% of the target regions covered at least 25X (Rashkin S, 2017). As I mentioned above, after performing variant calling and annotation with the custom bioinformatics analysis workflow, the patient was found to harbor a CNV of a genomic segment on chromosome 7 containing the KRIT1 gene. Specifically, the identified CNV consisted of a large heterozygous deletion involving the entire KRIT1 gene and part of the neighboring gene AKAP9, including coding exons from 38 to 50 (Figure 6). The heterozygous CNV mutation was confirmed by real-time qRT-PCR (figure 6 B), defined a deletion in AKAP9, KRIT1 and CYP51A1 genes, control 2 for GATA2 gene results partially reduced in real time qRT-PCR (figure 5 B) and not associated with CNV validation. Genes analyzed were shown in table 3.

Figure 6 A.

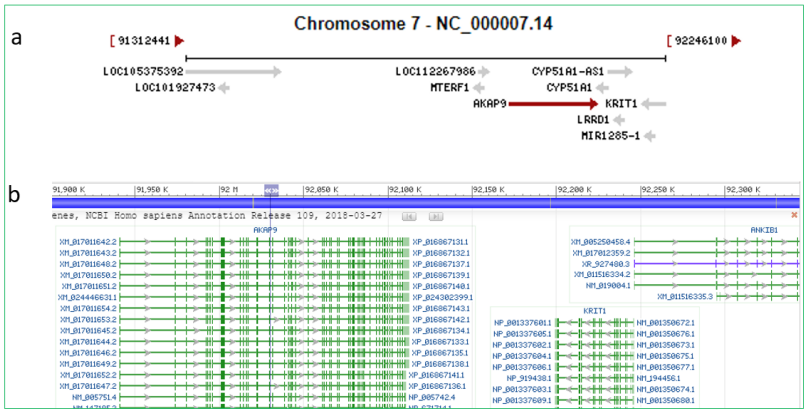

Figure 6 A.

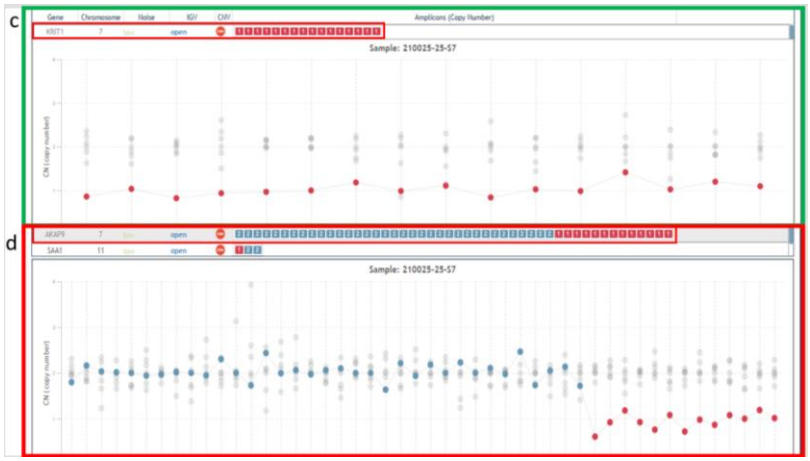

Figure 6 B.

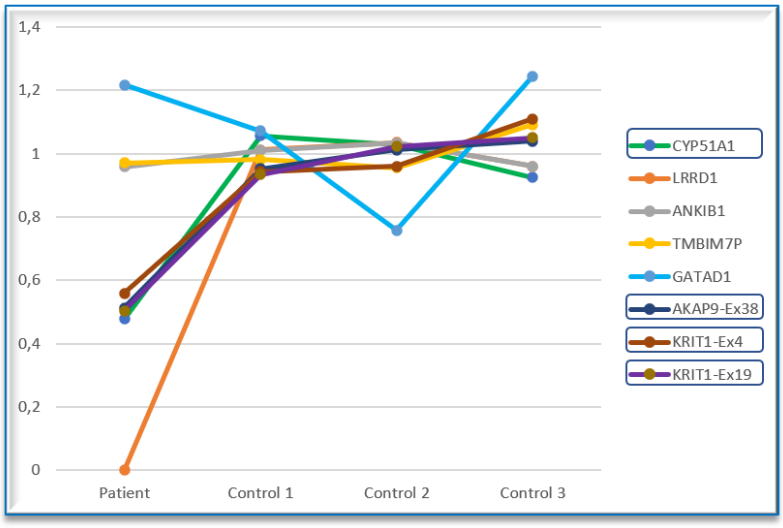

Figure 6: **A** - CNV report from the CCM case study generated by SOPHiA DDM® software. The outputs (panels a-c) show a CNV of a genomic segment on chromosome 7 involving two neighboring genes, KRIT1 and AKAP9, including heterozygous deletion of all the 16 coding exons of KRIT1 (b) and the last 13 coding exons (from 38 to 50) of AKAP9 (c) (V Benedetti, 2020). **B** – Graph represents quantitative data from qRT-PCR analysis of CNV patient and relative controls.

Gene	Exon	Probe	NM/NR/NG
KRIT1	Int 4	#72	NM_194456.1
KRIT1	Ex 13	#84	NM_194456.1
KRIT1	Int 19	#54	NM_194456.1
AKAP9	Int 16	#18	NG_011623.1
AKAP9	Int38	#7	NG_011623.1
ANKIB1	Ex 6	#22	NM_019004
ANKIB1	Ex 20	#1	NM_019004
CYP51A1	Ex 1	#52	NM_000786.3
GATAD1	Ex 5	#2	NM_021167.4
LRRD1	Ex 1	#6	NM_001161528.1
TMBIMP7	Ex 10	#12	NR_145992.1
TMBIMP7	Ex 1	#47	NR_145992.1

Table 3. Table of genes, exons / introns and relatives UPL probes corresponding to NM, NR or NG number.

Finally, I report two other cases. The first is a 47 years-old women that presented a complex pathological situation due to the concomitant occurrence of multiple cavernomas, Cowden Syndrome, and meningioma. Remarkably, despite familial CCM (fCCM) disease is characterized by multiple lesions, the patient resulted negative for mutations in CCM genes, but resulted affected by a mutation in PTEN (NM_000314.6) gene, suggesting the potential involvement of PTEN in CCM disease pathogenesis. The second case is a patient affected by fCCM disease that was shown to carry a mutation of uncertain significance in the NOTCH3 gene (NM_000435.2:c.2960G>C) (Figure 8). Notably, such mutation could induce a new exonic cryptic acceptor site, but further analyses are required to address this possibility.

Here I report a table with all the patient sequenced using CES and relative variants identified (Table 4).

ID	Gender	CES variants	CES variants Retained	CES variants Low Freq	Cerebro variants	Cerebro variants Retained	Cerebro variants Low Freq	CCM variants	CCM variants Retained	CCM variants Low Freq	CNV	Cerebro variants Low Freq info
1	M	28971	6606	117	256	54	3	2	0	0	2	(NOTCH3 c.2960G>C)
2	M	28396	6627	136	304	46	1	4	0	0	2	(Notch3 c.1733G>T)
3	F	30271	6818 (PTEN c.959dupT)	169	277	54	3	8	2	0	3	(Myh11, Mylk, Myh11) AND and CES variant: PTEN c.959dupT)
4	M	29438	6623	127	301	49	0	15	3	0	0	
5	M	30052	6582	119	311	43	1	12	3	0	0	(Myyh11 c.2096C>T)
6	M	27812	6781	135	281	54	2	6	1	0	2	(APP c.2113C>G, Myh11)
7	F	29638	6684	127	272	51	1	6	3	0	8 (KRIT-AKAP9)	NC_000007.14:g.(?_ 92198969)_(92246100 _?)del)
8	F	27957	6614	136	286	54	1	4	1	1 (PDCD10 c.575dupT)	5	(PDCD10 c.575dupT)
9	F	27584	6692	126	335	59	1	6	1	0	3	(ABCC6 c.1263 C>T)
10	M	27917	6527	104	259	54	0	6	2	0	1	
11	F	28496	6648	132	260	47	0	5	2	0	2	
12	M	30260	6632	99	336	61	1	10	4	1 (Krit1 c.1520C>T or c.1644C>T)	6	(Krit1 c.1520C>T or c.1644C>T)
13	M	29350	6492	143	294	50	2	4	1	1 (Krit1 c.1520C>T or c.1644C>T)	2	(Krit1 c.1520C>T or c.1644C>T; ABCC6 c.96G>A)
14	F	29800	6674	141	295	53	3	14	3	1 (Krit1 c.1520C>T or c.1644C>T)	3	(Krit1 c.1520C>T or c.1644C>T; ABCC6 c.96C>A; Mylk c.1164C>T)
15	F	30104	6681	139	297	55	3	12	3	0	4	(ACVRL1 c.1443C>T, COL4A2 c.4929G>A, Mylk c.4893A>G)
16	F	29411	6521	135	310	61	2	7	4	1 (Krit1 c.1520C>T or Krit1 c.1644C>T)	3	(Krit1 c.1520C>T or c.1644C>T; ABCC6 c.96G>A)
17	F	28854	6505	132	301	51	0	7	2	0	3	
18	F	27494	6313	120	361	61	0	12	2	0	5	
19	F	35355	6660	134	368	51	0	8	2	0	2 (MFF, COL4A3)	
20	F	35953	6869	159	326	51	0	9	2	0	0	
21	M	32892	6728	127	315	47	2	14	4	1	4	(var KRIT1 c.302G>T, ENG c.792C>T)
22	M	35688	6649	117	363	60	2	7	2	1 (CCM2 indel and frameshift c.463del)	3	(CCM2 indel c.463del, FNB1 c.5442C>T)
23	F	33792	6650	109	349	50	1	11	4	1 (frameshift KRIT1 c.1057_1060del)	0	(frameshift KRIT1 c.1057_1060del)
24	F	36239	6782	161	320	50	4	5	1	1 (nosense PDCD10 c.103C>T)	4	(nosense PDCD10 c.103C>T, MYH11 c.348G>A, MYLK c.2910C>T, NOCH3 c.945C>T)
25	F	35267	6624	144	377	50	0	10	3	0	3	

Table 4: Table of patients and relatives CCES variants, CES retained-variant, Cerebro Variants, Cerebro retained-variant, CCM variants, CCM retained-variant, CES and virtual panel variants filtered by frequency rate: less or equal to 0.1% in ExAc database (Exome Aggregation Consortium, less or equal to 0.01% in esp5400 database (Exome Variant Server, NHLBI GO Exome Sequencing Project http://evs.gs.washington.edu/EVS/2017), less or equal to 0.01% in the g100 database (1000 Genomes Project European Cohort - 1KGP, less or equal to 0.01 in GnomAD database (The Genome Aggregation Database, a data set contains data from 125,748 exomes and 15,708 whole genomes, all mapped to the GRCh37/hg19 reference sequence), for frequency in the pool samples less or equal to 4, and last one parameter is frequency in SOPHiATm DDM® software database less or equal to 100.

Discussion

Cerebral Cavernous Malformation (CCM) is a major cerebrovascular disease of proven genetic origin (OMIM 116860). It has been estimated to affect between 0.3% and 0.5% of the human population, and can occur either sporadically (sCCM) or as an inherited (fCCM) autosomal-dominant condition with incomplete penetrance and highly variable expressivity. CCM lesions consist of dilated vascular channels that have a characteristic appearance on MRI. CCMs are usually found intracranially, although such lesions can also affect the spinal cord and the retina. Individuals with CCMs can present with epilepsy and focal neurological deficits or acute intracerebral hemorrhage (ICH). CCMs are often clinically quiescent lesions. Indeed, neuroradiological series revealed that up to ∼70% of patients with such malformations are asymptomatic (Batra, 2009; Rigamonti, 2011; Fontanella, 2015). Hereditary forms of CCM (fCCM) are caused by loss-of-function mutations in one of three genes: KRIT1 (CCM1), CCM2 and PDCD10 (CCM3).

We decide to introduce CES analysis to have a good perspective on all genes associated to genetic diseases. In particular, using virtual panel made up by cerebrovascular related-genes we discover possible association between the 20 causative genes for a wide spectrum of cerebrovascular diseases and CCM disease as for main mechanisms involved in: the ABCC6 associated to extracellular matrix deposition (Bergen, 2000); ACTA2 related to vascular smooth muscle cells (SMCs) and regulated blood pressure and flow, LOF mutations induced the decrease SMC contraction (Dong-Chuan Guo 1, 2007); association with ACVRL1 to cerebrovascular development (Urness, 2000); to smooth muscle degeneration and irregular thickening of the basement membrane in some vessels with APP mutations (Herzig, 2004); in defective angiogenesis related to Moyamoya disease, BRCC3 (Miskinyte, 2011) and COL4A1 genes (Maeshima, 2000); COL3A1 associated to adhesion receptor in the brain (Vandervore, 2017); with multifocal intracerebral hemorrhages for COL4A2 (Jeanne, 2012); ENG with human vascular endothelium (Marchuk, 2003), FBN1 for structural support (Zhang, 1995); enzymatic lysosome degradation for GLA gene; in degradation of several extracellular matrix components for HTRA1 gene (Grau, 2006); in junctional adhesion between KAM3 and beta-1 integrins (Santoso, 2002); elasticity reduction of smooth muscle cells in

maintaining the mechanical properties for MYH11 gene (Zhu, 2006), reduction of smooth muscle contraction in the brain caused by MYLK mutation (Potier, 1995); association of NOTCH3 with structure and function of human blood vessels (Wimmer R. A., 2019); SAMHD1 is an important player in the replication stress response (Coquel F, 2018); association to polarity complex, structural component of tight junctions and dissolution of tight junctions during epithelial-mesenchymal transitions by TGFBR1 (Barrios-Rodiles, 2005) and TGFBR2 genes (Ozdamar, 2005); response to superoxide generation and related redox homeostasis (Chowdhury, 2006). In addition, using Cerebro virtual panel we extended the analysis to patients affected by cerebrovascular diseases and not only to CCM patients.

Genetic modifiers of disease severity

Why could we consider Akap9, Notch3, CYP51A1, MMP and PTEN and cerebrovascular-related ("Cerebro" gene panel) genes as a possible genetic modifiers of CCM disease?

AKAP9

AKAP350 (AKAP450/CG-NAP/AKAP9) is a multiple spliced A-kinase anchoring protein (AKAP), which assembles protein complexes at the Golgi apparatus and the centrosomes (Schmidt PH, 1999; Takahashi M, 1999; Witczak O, 1999). Largely known as the main microtubule nucleating organelle, centrosomes have garnered interest during the last decade because of their central role in various signaling events, including those involved in the establishment of cell polarity (Sumiyoshi E, 2012), and the regulation of the cell cycle progression (Doxsey, 2001). AKAP9 gene (OMIM 604001) encodes for homologous protein that plays a role in regulating microtubule dynamics and enhancing endothelial barrier downstream of Epac activation. Whereas AKAP9 is required for persistent microtubule growth under steady-state conditions, it was found not to be essential for maintaining basal barrier properties. Instead, it was required when the cell was challenged to make new adhesive contacts, as is the case when Epac activation promotes adhesion at cell-cell contacts to enhance barrier function (Seema Sehrawat, 2011) and regulate vascular permeability (A. Glading, 2007). Therefore, AKAP9 could be

participate to CCM pathogenesis through the mechanisms mentioned above and show in figure 9 (Seema Sehrawat, 2011) and 9 (Fisher O S., 2013).

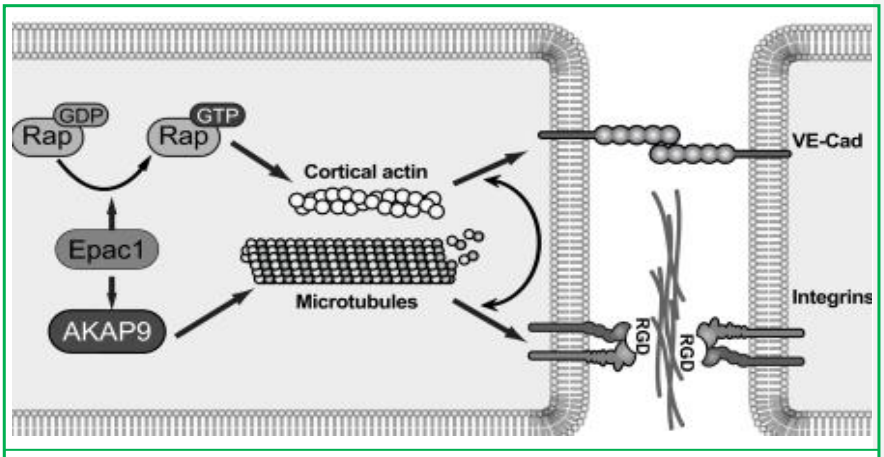

Figure 9. Model of cAMP/Epac1-induced endothelial barrier function requiring integrins. cAMP activation of Epac1 triggers 2 pathways: Epac mediates GTP-loading of Rap GTPases, which promotes cortical actin and VE-cadherin-induced cell-cell adhesion, and Epac activation of AKAP9 promotes microtubule growth. This pathway drives integrin adhesion at cell-cell borders. The coordinated function of VE-cadherin and integrins at junctional sites enhances endothelial cell barrier properties in response to cAMP-mediated activation of Epac. (Seema Sehrawat, 2011)

Figure 10: (Fisher O S., 2013).

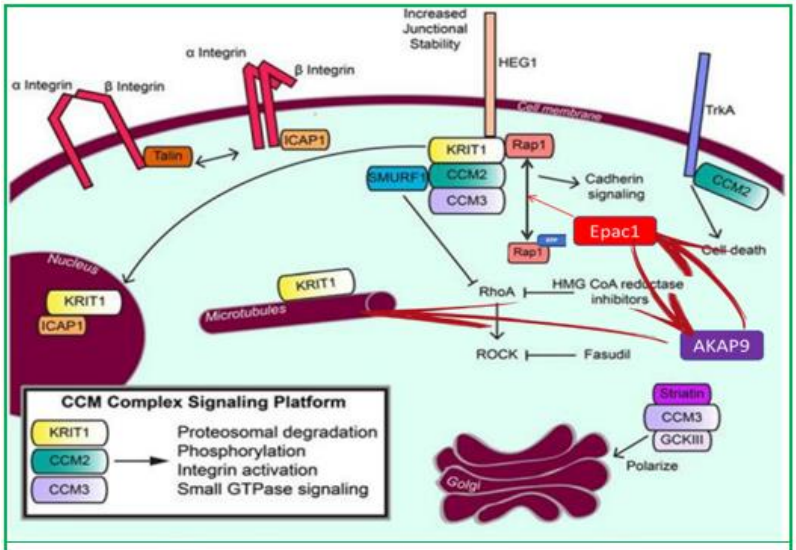

Figure 10. CCM proteins in signal transduction. Signal transduction from the CCM complex signaling platform is shown and the roles of CCM proteins outside of the CCM complex signaling platform are also indicated. The CCM proteins play roles in cell adhesion complexes and integrin signaling, in kinase signaling cascades, and in degradation/regulation of the Rho family of small GTPases. (Fisher O S., 2013)

Cyp51a1 and Mmp

Complex diseases are defined by an interaction of multiple genes and environmental factors, and genetic modifiers play a role in onset, severity and disease progression. Two gene families could play a role in variable expressivity of CCM disease: MMP and CYP gene families (Choquet, 2016). During my PhD project, I studied the CYPs family, in particular the CYP51A1 gene (OMIM 601637) codifies for lanosterol 14α-demethylase CYP51, a cytochrome P450 monooxygenases that catalyses many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. Therefore, CCM disease severity is associated to cholesterol metabolism (Romuald Girard, 2016) and suggest the pathological role of cytochromes in complex disease. CYP51A1 was found in a family with cerebral cavernous malformations, and the gene was proposed as a candidate for the cause of pediatric cataracts (Monika Lewinska, 2014). On the other side, MMP gene family codified for 28 different enzymes that require a zinc ion in their

active site for catalytic activity. MMPs are critical for maintaining tissue allostasis. Members of the MMP family include the "classical" MMPs, the membrane-bound MMPs (MT-MMPs) the ADAMs (a disintegrin and metalloproteinase; adamlysins) and the ADAMTS (a disintegrin and metalloproteinase with thrombospondin motif). Consequently, the endothelial expression of MMPs may affect the vascular matrix stability, and thus can contribute to severe phenotype of CCM (M Fujimura, 2007)

Notch3

In my thesis, I report a case of patients affected by CCM disease, negative for mutation in CCM genes but carrying a variant of uncertain significance in the NOTCH3 gene (OMIM 600276). CCM disease was confirmed by immunohistochemical analysis on lesion/pathological tissue. Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADSIL disease) is an autosomal dominant small-vessel disease of the brain caused by mutations in the NOTCH3 gene lead to arterial accumulation of Notch3 protein and granular osmiophilic material upon aging. CADASIL clinical symptoms are transient ischemic attacks or strokes, cognitive impairment, progressive dementia, pseudobulbar palsy, migraine with aura, and psychiatric disturbances (Chabriat H, 1995; Dichgans M, 1998), which are similar to CCM manifestation. The SNV in 15291515 position of chromosome 19 on NOTCH3 gene (NM_000435.2: c.2960G>C - p.Cys987Ser) is describe as “likely pathogenetic” and “pathogenic” in 83.7% of published cases (Christos Kopanos, 2018)

Pten

Phosphatase and tensin homolog deleted on chromosome 10, or PTEN (OMIM 601728), is a tumor suppressor involved in several disease due to loss-of-function mutations or environmental epigenetic modifications. PTEN gene is located on chromosome 10q23 and provides instructions for making an enzyme that is found in almost all tissues in the body. The enzyme has tumor suppressor function and modifies proteins and lipids by removing phosphate groups. Loss of PTEN function increases genetic instability and result correlated to many human cancers (LLOYD KM 2nd, 1963), in Cowden syndrome and in complex disease (Nicholas R. Leslie, 2016). Consequently, mouse models of CCMs lend some support to the two-hit hypothesis correlated to complex genetic diseases.

We and others have generated knockout alleles for both murine Ccm1 and Ccm2 (Plummer NW, 2006; Whitehead KJ, 2004; Goitre L., 2010). Moreover, the p53 null background was used to increase genetic and genomic instability and create a sensitized background for somatic mutation (Plummer N W., 2004). This sensitized animal model for CCM further supports the two-hit mechanism of pathogenesis. However, to date, somatic mutations have not been identified in lesions from the murine models of CCM. Clinical and histological evidence shows that CCM lesions exhibit characteristics of both vascular tumors and vascular malformations (Amy L. Akers, 2009). The potential involvement of PTEN promoter methylation in CCM disease particularly in familial CCM has been previously proposed, suggesting that this epigenetic alteration may contribute to the pathogenesis of CCM disease (Yuan Zhu, 2009). Insertion of thymine base (NM_000314.6: c.959dupT - p.L320Ffs*5) in 89720807 position of chromosome 10 in PTEN gene, resulted as "potentially pathogenic" or "pathogenic" with a variant fraction of 41%. The annotated criteria explain the variant as a null variant (frame-shift), which induce loss-of-function is a known mechanism of disease (gene has 433 pathogenic LOF variants and LOF Z-Score = 3.2 is greater than 0.7) (Christos Kopanos, 2018).

"Cerebro" genes

In this PhD project, we decide to put a light on all the genes associated with cerebrovascular disease for several reasons. The first one is the possibility to identified a correlation between CCM severity/variable expressivity and variants on cerebrovascular-related genes; the second one is the simplification of cerebrovascular genetic disease clinical analysis, not only focused on CCM disease and the last one: the implementation of NGS technologies into clinical diagnostics and research because is a promising approach for the optimization of diagnosis and treatment of all cerebrovascular genetic disease. Cerebro genes (Table 2) could play a role in phenotype variability or could be associated to CCM disease. In such scenarios, associated variants will accumulate in genes with a property that does not itself play a role in the phenotype, but is correlated with another gene property that does (de Leeuw CA, 2015). Moreover, cerebrovascular genetic variants could be useful as potential pharmacological target as gene-based therapeutic approaches, for example patients with CYP27A1 phatogentic

variants are treatment with chenodeoxycholic acid to prevent neurological damage (Smalley, 2015), and FBN1 pathogenetic variant who are treated with beta-blocker therapy (Milleron O, 2020).

Conclusion

Next generation sequencing (NGS) is the biggest advancement of technology that evolved during recent years. NGS provides high-throughput and base pair resolution data, providing the analysis of multiple genetic loci and samples from different couples simultaneously. The use of massively parallel sequencing has increasingly been the object of study in recent years. Whole Genome, Whole Exome and Clinical Exome Sequencing (WGS – WES - CES) clinical application are not limited to diagnosis, these are also widely used in identifying mutation targets for targeting therapy and in identifying a high-risk population for certain hereditary cancers, also become possible thanks to drastic cost reduction.

CCM causative mutations in the three CCM genes account for 85 to 95 percent of all cases of familial CCM. The remaining 5 to 15 percent of cases may be due to mutations in unidentified genes or to other unknown causes. Mutations in KRIT1, CCM2, and CCM3 genes are also been detected in sporadic CCMs. (Denier C, 2006; Liquori CL, 2007; Riant F, 2013; Spiegler S N. J.-S., 2014).

Notably, I report cases of CCM patients carrying possible causative disease variants in disease-not related genes, such as PTEN for cell proliferation and differentiation, migration and NOTCH3 for the stability of vascular smooth muscle cells in the brain; therefore, we can speculate a possible pathogenetic role of these new genes as possible genetic modifiers of CCM disease.

In particular, the development of an optimized NGS protocol made it possible to identify novel causative variants in known CCM genes as well as other genes and potential genetic risk factors associated to cerebrovascular disease, thus providing useful insights into the improvement of current diagnostic procedures for a more effective diagnosis,

prognosis and clinical management of CCM disease. Pathologic-driven virtual panel is an effective method to analyze CES data of cerebrovascular disease and additional CYP and MMP genes could help to understand the high variable expressivity in complex disease. Moreover, clinical application of CES who sequenced more than thousand genes not only 3 causative genes, we applied this technique for the identification of relevant genetic information in a cohort of patients affected by most of known genetic disease. CES application in diagnosis of CCM patients allowed a rapid detection and screening of pathological variants and genetic modifiers of disease severity and progression, as for case of CNV who involved genes related to fundamental function as: enhancing endothelial barrier (akap9 protein), role in actin cytoskeletal integrity, cell proliferation, cell differentiation, cell adhesion, apoptosis, and cell migration for LRRD1, and no direct effects on disease from cyp51a1 but seems that much of the genetic control of common diseases is due to rare variants with a strong impact on disease risk in individual patients (Cirulli ET, 2010) and could be effects on disease severity (Choquet, 2016)).

The heterozygous mutations leading to decreased protein levels or enzyme activity often result in mild phenotypes that can be easily trascured. Cumulative effects of single mutations, which may cause very mild phenotypes independently, together with the environmental factors, carry important information about causes of complex diseases.

PTEN mutation is associated to hamartoma tumor syndrome, abnormal skin growths, and intellectual disabilities, but could be involved in the same mechanisms at the base of CCM disease (Yuan Zhu, 2009).

NGS approaches allows to test a large number of genes simultaneously and to obtained a large number of variants, also considered “pathogenic” or “potentially pathogenic”. For example, an high number of variants were identified in my NGS analysis, as I report above (Table 4). However, the largest part of these thousands of genetic variants are defined as “common variants”, only a few common gene variants are causative or associated with genetic disease.

Commentato [V18]: *High number of genetic variants*

Computational algorithms, associated to common data-analysis software, are currently used to predict the potential functional impact and relevance of variants, but due to the sheer volume of data, extreme caution should be exercised when weighing the

importance of these predictions. Indeed, de novo or rare missense mutations found by genome sequencing, nearly half inferred as deleterious were in fact nearly neutral mutations, based on data from functional genomics studies (Miosge, 2015). On the other side, variants found by NGS considered as likely benign or benign, these can be expected to no-impact protein function, but a more detailed analysis could warrant as they may be novel disease-causing mutations.

The problem and the advantage of hard-variant filtering is that it is very limiting because it forces you to look a restrict group of retained-variant, advantage because I could focused on a small group of variant and could find a causative variant of genetic disease. On the other side, the limiting is an indirectly filtering out germline causative variants of complex disease as CCM.

My PhD thesis aim is not to define absolute rules for NGS data filtering but rather to provide guidelines that can be adjusted or replaced as genomic knowledge and software improve. It is important to note that these recommendations are for research use only and, therefore, do not necessarily conform to interpretation guidelines suggested by clinical sequencing regulatory and advisory bodies.

Every filtering strategy inevitably starts with multiple seemingly logical assumptions regarding the properties of causal variants, which when combined with over-reliance on automation and predictive tools may either discard relevant candidates too early or produce excessive false positives. It is also important to note that any given individual's exome contains a large number of protein sequence–altering inherited genetic variations and de novo mutations, but only a small fraction of these can be expected to impact protein function (e.g., nonsense mutations that result in a truncated protein). A smaller number of these mutations are likely to be deleterious or relevant to a disease study, even if they significantly alter protein function or expression levels, due to incomplete or non-penetrance.

Finally, it is important to note that false-positive variant calls are still a reality, and therefore any candidates should be independently verified using the current gold standard, Sanger sequencing, before proceeding with additional analyses. Indeed, as I mentioned in supplementary filters (for frequency in the pool samples less or equal to

4, and frequency in SOPHiA^{Tm} DDM® software database less or equal to 100), the presence of the same variant in multiple samples can provide some level of confidence, particularly if the data were generated from multiple sequencing runs, it recommend applying a combination of computational and manual assessment in the variant filtering process to avoid time-consuming prioritization and hypothesis generation based on false positives.

Return of “incidental” or “off-target” findings remains an area of debate in medical genetics (Kalia SS, 2017).

This study has a number of important limitations. For example, no trio-CES analysis were performed for a deeper comprehension of variable expressivity into the same family; diagnostic usage of CES is not able to detect all causal mutation types, and thus specific mutations will be unobserved by this test. CES does not detect pathogenic repeat expansions (for disorders such as spinocerebellar ataxia) or most CNV. On the other side, CNV is a common source of genomic variation and an important genetic cause of disease. At day, most diagnostic genetic studies are generally limited to the study of SNVs and other small variants. With the introduction of CES, it is now possible to identified both (SNVs and CNVs) with one single and sensible test (Pfundt R, 2017).

Notably, genetic variant interpretation required a high experience in clinical genetics and the only hard-filtering or bioinformatic approach is not sufficient to a better diagnosis of genetic disease.

Thanks to Next Generation Sequencing (NGS), it became possible sequencing of multiple DNA molecules in parallel, enabling hundreds of millions of DNA molecules to be sequenced at a time. This advantage allows high throughput sequencing to be used to create large data sets, generating more comprehensive insights into the cellular genomic and transcriptomic signatures of various diseases and developmental stages. The perspective of new genetic sequencing methods is increasing therapeutic opportunities and challenges for gene discovery of rare diseases and allows to look ahead to the future.

References

A. Glading, J. H. (2007). *KRIT-1/CCM1 is a Rap1 effector that regulates endothelial cell–cell junctions.* J. Cell Biol., 179 (2007), pp. 247-254.

Amy L. Akers, E. J. (2009). *Biallelic somatic and germline mutations in cerebral cavernous malformations (CCMs): evidence for a two-hit mechanism of CCM pathogenesis.* Human Molecular Genetics, Volume 18, Issue 5, 1 March 2009, Pages 919–930, https://doi.org/10.1093/hmg/ddn430.

Angeliki Louvi, S. N. (2014). *Ccm3, a gene associated with cerebral cavernous malformations, is required for neuronal migration.* Development 2014 141: 1404-1415; doi: 10.1242/dev.093526.

Antognelli, C. E. (2018). *KRIT1 loss-of-function induces a chronic Nrf2-mediated adaptive homeostasis that sensitizes cells to oxidative stress: Implication for Cerebral Cavernous Malformation disease. .* Free Radic Biol Med .

Antognelli, C. E.-o.-f. (2018). *Data in support of sustained upregulation of adaptive redox homeostasis mechanisms caused by KRIT1 loss-of-function. .* Data Brief 16, 929-938.

Batra, S. D. (2009). *Cavernous malformations: natural history, diagnosis and treatment.* Nat Rev Neurol 5, 659-670.

Bergametti, F. C.-L. (2005). *Mutations within the programmed cell death 10 gene cause cerebral cavernous malformations.* Neurochirurgie.

Boulday G, B. A.-K.-L. (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. 2009 Mar-Apr; 2(3-4):168-77.

Chabriat H, V. K.-Z. (1995). *Clinical spectrum of CADASIL: a study of 7 families. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy.* Lancet. 1995 Oct 7; 346(8980):934-9.

Choquet H, P. L. (2015). *Genetics of cerebral cavernous malformations: current status and future prospects. .* J Neurosurg Sci 59:211–220.

Choquet, H. E. (2016). *Cytochrome P450 and matrix metalloproteinase genetic modifiers of disease severity in Cerebral Cavernous Malformation type 1. .* Free Radic Biol Med 92, 100-109.

Christos Kopanos, V. T. (2018). *VarSome: The Human Genomic Variant Search Engine. .* Oxford Bioinformatics, bty897, 30 October 2018. doi: https://doi.org/10.1093/bioinformatics/bty897.

Cianfruglia, L. A.-O.-F. (2019). *KRIT1 Loss-Of-Function Associated with Cerebral Cavernous Malformation Disease Leads to Enhanced. .* Antioxidants (Basel) 8.

D'Angelo R, M. V. (2011). *Mutation analysis of CCM1, CCM2 and CCM3 genes in a cohort of Italian patients with cerebral cavernous malformation.* Brain Pathol. 2011;21:215–24.

De Luca E, P. D. (2018). *Multifunctional Platinum@BSA-Rapamycin nanocarriers for the combinatorial therapy of cerebral cavernous malformation.* ACS Om.

Denier C, L. P.-L. (2006). *Genotype-phenotype correlations in cerebral cavernous malformations patients.* . Ann Neurol. 2006;60:550–6.

Desmet, F. O.-B. (2009). *Human Splicing Finder: an online bioinformatics tool to predict splicing signals.* Nucleic Acids Res 37, e67.

Dichgans M, M. M.-H. (1998). *The phenotypic spectrum of CADASIL: clinical findings in 102 cases.* . Annals of Neurology, 01 Nov 1998, 44(5):731-739 DOI: 10.1002/ana.410440506 PMID: 9818928 .

Doxsey, S. (2001). *Re-evaluating centrosome function.* Nat Rev Mol Cell Biol 2001; 2:688-98; PMID:11533726; http://dx.doi.org/10.1038/35089575.

Ellis JA, B. D. (2017). *Supratentorial cavernous malformations.* . Handb Clin Neurol.

Faurobert, S. B.-D.-R. (2007). *Krit 1 interactions with microtubules and membranes are regulated by Rap1 and integrin cytoplasmic domain associated protein-1.* FEBS J. 2007 Nov; 274(21): 5518–5532.

Fisher O S., B. T. (2013). *Signaling pathways and the cerebral cavernous malformations proteins: lessons from structural biology.* Medicine, Biology ; Cellular and Molecular Life Sciences - DOI:10.1007/s00018-013-1532-9Corpus ID: 14836490.

Fisher OS, W. L. (2014). *Structural basis for the disruption of the Cerebral Cavernous Malformations 2 (CCM2) interaction with Krev Interaction Trapped 1 (KRIT1) by disease-associated mutations.* Journal of Biological Chemistry.

Flávio Domingues, E. L. (2008). *FAMILIAL CEREBRAL CAVERNOUS MALFORMATIONS. Rio de Janeiro study and review of the recommendations for management.* Arq Neuropsiquiatr.

Fontanella, M. (2015). *Cerebral Cavernous Malformations.* Edizioni Minerva Medica.

Gibson CC, Z. W.-K. (2015). *Strategy for identifying repurposed drugs for the treatment of cerebral cavernous malformation.* Circulation 131:289–299.

Goitre L., B. F. (2010). *KRIT1 regulates the homeostasis of intracellular reactive oxygen species.* PLoS One doi: 10.1371/journal.pone.0011786.

Goitre, L. E. (2014). *KRIT1 loss of function causes a ROS-dependent upregulation of c-Jun.* Free Radic Biol Med.

Goitre, L. P. (2017). *Up-regulation of NADPH oxidase-mediated redox signaling contributes to the loss of barrier function in KRIT1 deficient endothelium.* Scientific Reports.

Guazzi, P. L. (2012). *Identification of the Kelch family protein Nd1-L as a novel molecular interactor of KRIT1.* . PLoS One.

Horne MA, F. K. (2015). *Clinical course of untreated cerebral cavernous malformations: a meta-analysis of individual patient data.* The Lancet. Neurology.

ISTAT. (2017). *Istituto Nazionale di Statistica (ISTAT). Anni 2003-2014. L'evoluzione della mortalità per causa: le prime 25 cause di morte, https://www.istat.it/it/files/2017/05/Report-cause-di-morte-2003-14.pdf* . Istat.

Jiang X, P. A.-F. (2019). *Alternatively spliced isoforms reveal a novel type of PTB domain in CCM2 protein. .* Scientific Rep 9(1).

Laberge-le Couteulx, S. H.-L. (1999). *Truncating mutations in CCM1, encoding KRIT1, cause hereditary cavernous angiomas.* Nat Genet.

Liquori CL, B. M. (2007). *Deletions in CCM2 are a common cause of cerebral cavernous malformations.* Am J Hum Genet. 2007;80:69–75.

Liquori, C. L. (2003). *Mutations in a Gene Encoding a Novel Protein Containing a Phosphotyrosine-Binding Domain Cause Type 2 Cerebral Cavernous Malformations.* American Journal of Human Genetics.

LLOYD KM 2nd, D. M. (1963). *Cowden's disease. A possible new symptom complex with multiple system involvement.* Ann Intern Med. 1963 Jan; 58():136-42.

M Fujimura, M. W. (2007). *Expression of matrix metalloproteinases (MMPs) and tissue inhibitor of metalloproteinase (TIMP) in cerebral cavernous malformations: immunohistochemical analysis of MMP-2, -9 and TIMP-2.* Acta Neurochir - PMID: 17043747 DOI: 10.1007/s00701-006-0929-8.

Marchi, S. E. (2016). *Beyond multiple mechanisms and a unique drug: Defective autophagy as pivotal player in cerebral cavernous malformation pathogenesis and implications for targeted therapies.* Rare Diseases .

Marchi, S. M. (2015). *Defective autophagy is a key feature of cerebral cavernous malformations.* EMBO Mol Med.

Marchi, S. S. (2016). *Cellular processes underlying cerebral cavernous malformations: Autophagy as another point of view. .* Autophagy .

Moglia A, L. G. (2015). *Evaluation of the bioactive properties of avenanthramide analogs produced in recombinant yeast.* BioFactors.

Moglianetti M, D. L. (2016). *Platinum nanozymes recover cellular ROS homeostasis in an oxidative stress-mediated disease model.* Nanoscale.

Moglianetti M, P. D. (2020). *Intracellular antioxidant activity of biocompatible citrate-capped palladium nanozymes. .* Nanomaterials .

Monika Lewinska, P. J. (2014). *Small Deletion at the 7q21.2 Locus in a CCM Family Detected by Real-Time Quantitative PCR.* PLoS One. 2014; 9(11): e112787.

Moore SA, B. R. (2014 May). *Long-term natural history of incidentally discovered cavernous malformations in a single-center cohort.* J. Neurosurg.

Muscarella LA, V. G. (2010). *Small deletion at the 7q21.2 locus in a CCM family detected by real-time quantitative PCR.* Journal of biomedicine & biotechnology 2010: 854737.

Nicholas R. Leslie, M. L. (2016). *Inherited PTEN mutations and the prediction of phenotype.* Seminars in Cell & Developmental Biology. Volume 52, April 2016, Pages 30-38.

Perrelli A, G. L. (2018). *Biological activities, health benefits, and therapeutic properties of avenanthramides: from skin protection to prevention and treatment of cerebrovascular diseases.* Oxidative Med Cell Longev.

Plummer N W., C. J. (2004). *Loss of p53 Sensitizes Mice with a Mutation in Ccm1 (KRIT1) to Development of Cerebral Vascular Malformations.* American Journal of Pathology, Vol. 165, No. 5, November 2004.

Plummer NW, S. T. (2006). *Neuronal expression of the Ccm2 gene in a new mouse model of cerebral cavernous malformations.* Mamm Genome. 2006 Feb; 17(2):119-28.

Retta SF, A. M. (2004). *Identification of Krit1B: a novel alternative splicing isoform of cerebral cavernous malformation gene-1. .* Gene 325:63–78.

Retta, S. F. (2016). *Oxidative stress and inflammation in cerebral cavernous malformation disease pathogenesis: Two sides of the same coin. .* Int J Biochem Cell Biol.

Riant F, B. F.-P.-L. (2013). *CCM3 Mutations Are Associated with Early-Onset Cerebral Hemorrhage and Multiple Meningiomas. .* Mol Syndromol. 2013;2013;4:165–72.

Richards, S. N.-F. (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, 405-424.

Rigamonti, D. (2011). *Cavernous Malformations of the nervous system.*

Romuald Girard, O. K. (2016). *Peripheral plasma vitamin D and non-HDL cholesterol reflect the severity of cerebral cavernous malformation disease.* Biomark Med. 2016 Mar; 10(3): 255–264.

Schmidt PH, D. D. (1999). *AKAP350, a multiply spliced protein kinase A-anchoring protein associated with centrosomes.* J Biol Chem 1999; 274:3055-66; PMID:9915845; http://dx.doi.org/10.1074/jbc.274.5.3055.

Seema Sehrawat, T. E. (2011). *AKAP9 regulation of microtubule dynamics promotes Epac1-induced endothelial barrier properties.* Prepublished online 2010 Oct 15. Blood. 2011 Jan 13; 117(2): 708–718. doi: 10.1182/blood-2010-02-268870.

Spiegler S, N. J.-S. (2014). *High mutation detection rates in cerebral cavernous malformation upon stringent inclusion criteria: one-third of probands are minors. .* Mol Genet Genomic Med 2:176–185.

Spiegler S, R. M. (2018). *Cerebral cavernous malformations: an update on prevalence, molecular genetic analyses, and genetic counselling.* Mol Syndromol 9(2):60–69.

Sumiyoshi E, S. A. (2012). *Cell polarity: centrosomes release signals for polarization.* Curr Biol 2012; 22:R281-3; PMID:22537634; http://dx.doi. org/10.1016/j.cub.2012.03.009.

T Sahoo 1, E. W.-L.-N. (1999). *Mutations in the gene encoding KRIT1, a Krev-1/rap1a binding protein, cause cerebral cavernous malformations (CCM1).* Hum Mol Genet.

Takahashi M, S. H. (1999). *Characterization of a novel giant scaffolding protein, CG-NAP, that anchors multiple signaling enzymes to centrosome and the golgi apparatus. .* J Biol

Chem 1999; 274:17267- 74; PMID:10358086; http://dx.doi.org/10.1074/ jbc.274.24.17267.

Trapani, E. a. (2015). *Cerebral cavernous malformation (CCM) disease: from monogenic forms to genetic susceptibility factors. .* J Neurosurg Sci 59, 201-209.

V Benedetti, E. P. (2020). *Next Generation Sequencing (NGS) Strategies for Genetic Testing of Cerebral Cavernous Malformation (CCM) Disease.* Methods in Molecular Biology,.

Whitehead KJ, P. N. (2004). *Ccm1 is required for arterial morphogenesis: implications for the etiology of human cavernous malformations.* Development. 2004 Mar; 131(6):1437-48.

WHO. (2016). *Cardiovascular Diseases, https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1.* WHO - world health Organization.

Witczak O, S. B. (1999). *Cloning and characterization of a cDNA encoding an A-kinase anchoring protein located in the centrosome, AKAP450.* EMBO J 1999; 18:1858- 68; PMID:10202149; http://dx.doi.org/10.1093/ emboj/18.7.1858.

Yuan Zhu, A. W. (2009). *Involvement of PTEN Promoter Methylation in Cerebral Cavernous Malformations.* Stroke - https://doi.org/10.1161/STROKEAHA.108.526376 - 40:820–826.