Received: 8 March 2019 | Revised: 12 June 2019 | Accepted: 24 June 2019

DOI: 10.1002/humu.23851

DATA ARTICLE

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

Molecular diagnostic workflow, clinical interpretation of sequence variants, and data repository procedures in 140 individuals with familial cerebral cavernous malformations

Carmela Fusco[1] | Massimiliano Copetti[2] | Tommaso Mazza[3] | Luigi Amoruso[4] | Sandra Mastoianno[5] | Grazia Nardella[1,6] | Vito Guarnieri[1] | Lucia Micale[1] | Leonardo D'Agruma[1] | Marco Castori[1]

[1]Division of Medical Genetics, Fondazione IRCCS-Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy

[2]Unit of Biostatistics, Fondazione IRCCS-Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy

[3]Unit of Bioinformatics, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy

[4]Division of Neurology, Fondazione IRCCS-Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy

[5]Division of Cardiology, Fondazione IRCCS-Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy

[6]Department of Experimental Medicine, Sapienza University of Rome, Rome, Italy

Correspondence
Carmela Fusco, Division of Medical Genetics, Fondazione IRCCS-Casa Sollievo della Sofferenza, San Giovanni Rotondo, Viale Padre Pio, 771013 San Giovanni Rotondo, Foggia, Italy.
Email: c.fusco@operapadrepio.it

Funding information
Ministry of Health -Italy

Abstract

Familial cerebral cavernous malformation (FCCM) is an autosomal dominant vascular disorder caused by heterozygous deleterious variants in *KRIT1*, *CCM2* or *PDCD10*. In a previous study, we presented the clinical and molecular findings in 140 FCCM individuals. In the present work, we report supporting information on (a) applied diagnostic workflow; (b) clinical significance of molecular findings according to the American College of Medical Genetics and Genomics/Association for Molecular Pathology recommendations; (c) standardization of molecular and clinical data according to the Human Phenotype Ontology; (d) preliminary genotype-phenotype correlations on a subgroup of patients by considering sex, age at diagnosis, neurological symptoms, and number and anatomical site(s) of vascular anomalies; (e) datasets submitted to the Leiden Open Variation Database. An overview of the changes of our diagnostic approach before and after the transition to next-generation sequencing is also reported. This work presents the full procedure that we apply for molecular testing, data interpretation and storing in public databases in FCCM.

KEYWORDS
CCM2, human phenotype ontology, KRIT1, leiden open variation database, mutation, PDCD10, variant interpretation

1 | INTRODUCTION

Cerebral cavernous malformations (CCMs; MIM# 116860, 603284, 603285) are congenital vascular anomalies of the brain that may lead to hemorrhage, seizures, and neurologic deficits. The familial cerebral cavernous malformation (FCCM) is mostly linked to loss-of-function variants in one of the following genes: *KRIT1*, *CCM2*, and *PDCD10* (Scimone et al., 2018). Germline variants in these genes may occur in autosomal dominant families with CCM, as well as in sporadic cases with multiple CCM. Incomplete penetrance and variable expressivity are features of FCCM. Molecular pathogenesis and genotype-phenotype correlations of identified germline variants are still poorly understood in FCCM. In addition, a mutational database dedicated to FCCM is still lacking and the clinical description of published cases is regularly offered by a non-standardized nomenclature. Therefore,

Companion Article

Nardella G., Visci G., Guarnieri V., Castellana S., Biagini T., Bisceglia L., Palumbo O., Trivisano M., Vaira C., Scerrati M., Debrasi D., D'Angelo V., Carella M., Merla G., Mazza T., Castori M., D'Agruma L., Fusco C. A single-center study on 140 patients with cerebral cavernous malformations: 28 new pathogenic variants and functional characterization of a PDCD10 large deletion. Hum. Mutat. 2018; 39: 1885–1900. doi: 10.1002/humu.23629. (PMID: 30161288).

more homogeneity in molecular and clinical studies of FCCM is expected for improving the translational nature of ongoing research. We recently published the molecular findings and selected clinical features in a cohort of 140 individuals with FCCM, together with some insights on the potential role of PDCD10 into the autophagy process (Nardella et al., 2018).

2 | DATA SPECIFICATIONS

Data type	Tables, text, and graphs
Data acquisition method	Variant interpretation: ACMGG/AMP guidelines, multiple in silico tools.
	Phenotype classification: human phenotype ontology.
	Genotype–phenotype correlation: Kruskal–Wallis and Fisher exact tests.
	Data submission and storing: LOVD Data Base.
Data format	Raw and analyzed.
Experimental factors	None.
Experimental features	Mutational analysis with a multi-technique approach providing the highest variant call rate and reporting quality in FCCM. Systematic approach to variant interpretation, phenotype description, and data storing in public databases for FCCM.
Data source location	Division of medical genetics
	Fondazione IRCCS-Casa Sollievo della Sofferenza
	Poliambulatorio "Giovanni Paolo II"
	Viale Padre Pio, 7
	71013 San Giovanni Rotondo (Foggia), Italy
Data availability statement	https://databases.lovd.nl/shared/individuals/00165057
	https://databases.lovd.nl/shared/individuals/00165058
	https://databases.lovd.nl/shared/individuals/00165059
	https://databases.lovd.nl/shared/individuals/00165061
	https://databases.lovd.nl/shared/individuals/00165062
	https://databases.lovd.nl/shared/individuals/00165063
	https://databases.lovd.nl/shared/individuals/00165064
	https://databases.lovd.nl/shared/individuals/00165065
	https://databases.lovd.nl/shared/individuals/00165066
	https://databases.lovd.nl/shared/individuals/00165067
	https://databases.lovd.nl/shared/individuals/00165068
	https://databases.lovd.nl/shared/individuals/00165069
	https://databases.lovd.nl/shared/individuals/00165070
	https://databases.lovd.nl/shared/individuals/00165071
	https://databases.lovd.nl/shared/individuals/00165072
	https://databases.lovd.nl/shared/individuals/00165073
	https://databases.lovd.nl/shared/individuals/00165074
	https://databases.lovd.nl/shared/individuals/00165075
	https://databases.lovd.nl/shared/individuals/00165076
	https://databases.lovd.nl/shared/individuals/00165077
	https://databases.lovd.nl/shared/individuals/00165080
	https://databases.lovd.nl/shared/individuals/00165083
	https://databases.lovd.nl/shared/individuals/00165082
	https://databases.lovd.nl/shared/individuals/00165700
	https://databases.lovd.nl/shared/individuals/00165085
	https://databases.lovd.nl/shared/individuals/00165081
	https://databases.lovd.nl/shared/individuals/00165079
	https://databases.lovd.nl/shared/individuals/00165043
	https://databases.lovd.nl/shared/individuals/00165078

3 | IMPACT OF DATA

Here, we report supporting information concerning the diagnostic work-flow ongoing in our laboratory, clinical interpretation of molecular findings according to American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMGG/AMP; Richards et al., 2015), procedures of clinical and molecular data reporting in public databases, as well as some preliminary genotype-phenotype correlations. Our goal is to provide guidance and discuss best practices and standards for each stage tracked in the mutational analysis reported in (Nardella et al., 2018).

This article would stimulate research for integrating information on genotype–phenotype correlations into the future diagnostic work-flow of FCCM, as well as systematic data representation by using established nomenclatures, data models and ontologies in FCCM. This could open the path to the generation of a clinical-molecular database dedicated to FCCM with potential applications in the fields of molecular, epidemiologic, and basic research.

4 | EXPERIMENTAL DESIGN, MATERIALS, AND METHODS

4.1 | Diagnostic workflow

The Division of Medical Genetics of Foundation IRCCS-Casa Sollievo della Sofferenza is one of the two Italian reference laboratories for the molecular diagnosis of FCCM and offers full diagnostics facilities for this condition from 2004. Samples are sent, together with informed consent and key clinical/demographic features, from secondary (e.g., local medical genetics services) and tertiary clinical centers (e.g., divisions of neurology or neurosurgery of University Hospitals) of different Italian regions, particularly Southern Italy. From 2004–2017, samples were processed by Sanger sequencing. Molecular testing appropriateness is verified by reviewing available data compared to the criteria for a clinical suspect of FCCM, including at least one among (a) magnetic resonance imaging with evidence of two or more central nervous system cavernous malformations; (b) one or more relative with a clinical diagnosis of FCCM. Lack of appropriateness or adequate clinical information prompts the senior biologist to get in touch with the clinical provider for further information. Before 2017, Sanger sequencing testing started with *KRIT1* (NM_194456.1) analysis and then proceeded on *CCM2* (NM_031443.3) and *PDCD10* (NM_007217.3). Any variant of potential clinical interest was confirmed by Sanger sequencing on an independently extracted second DNA sample. In case of negative results, intragenic rearrangements and whole gene deletions were investigated with multiple ligation-dependent probe amplification (MLPA). Abnormal results for a single probe were confirmed by quantitative polymerase chain reaction (qPCR). Deletion of a single MLPA probe requested exclusion of an allelic drop-out by review of the sequencing data. Deletion/duplication of two or more MLPA probes was validated by a second MLPA experiment on an independently extracted DNA sample. Whole gene rearrangements or rearrangements involving the 5′ and 3′ extremes were confirmed and further refined by SNP-based array genome hybridization (Figure 1).

In 2017, molecular diagnostics of FCCM shifted on a next-generation sequencing (NGS) platform including all the three genes (*KRIT1*, *CCM2*, and *PDCD10*). Nonrecurrent/sample-specific gaps of coding and exonic–intronic junction sequences were covered by Sanger sequencing. This offered us the opportunity to significantly reduce the turnaround time of the analysis and cost. In case of negative results from the NGS and MLPA analyses, available clinical data were further reviewed for possible overlap with closely related disorders. The NGS panel (SureSelect Design: 3186631) was set up and the library of all coding regions of the following genes was obtained using the SureSelect target enrichment system (Agilent Technologies, Santa Clara, CA), according to the manufacturer's instructions. NGS was performed on a MiSeq sequencer (Illumina, San Diego, CA) using a MiSeq Reagent kit V3 300 cycles flow cell. Characteristics of target regions for *KRIT1*, *CCM2* and *PDCD10* are reported in Table 1.

Clinical providers were regularly interrogated for any hypothesis of partially overlapping genetic condition that can be verified by other available NGS diagnostic panels (e.g., hereditary connective tissue disorders, or von Hippel–Lindau syndrome and related disorders) or exome sequencing. In the case of a clear-cut clinical diagnosis of FCCM, biobanking for future genomic research was proposed (Figure 2).

All identified variants were interpreted according to the ACMGG/AMP guidelines (Richards et al., 2015). Only pathogenic and likely pathogenic variants and variants of unknown significance (VUS) were included in the clinical report. For selected variants of particular scientific interest, the clinical provider was asked to check the availability of the patient for further sampling (e.g., second blood sample, buccal mucosa, skin biopsy) to achieve additional studies.

4.2 | Variant interpretation

Variants identified before the publication of the ACMGG/AMP guidelines were reanalyzed and interpreted accordingly. Variants without clinical significance at the time of reporting (i.e., benign and likely benign) were excluded by the presence of (a) one (or more) stand-alone criteria for benignity, (b) two or more strong criteria for benignity, (c) one strong criterion plus one supporting criterion for benignity, or (d) two or more supporting criteria for benignity. Variants which passed this preliminary selection were candidate for reporting (pathogenic, likely pathogenic or VUS).

The following databases were used for population data of identified variants: dbSNP150 (https://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi), gnomAD (https://gnomad.broadinstitute.org), esp6500 (http://evs.gs.washington.edu/EVS/), and ExAC (http://exac.broadinstitute.org). To establish the

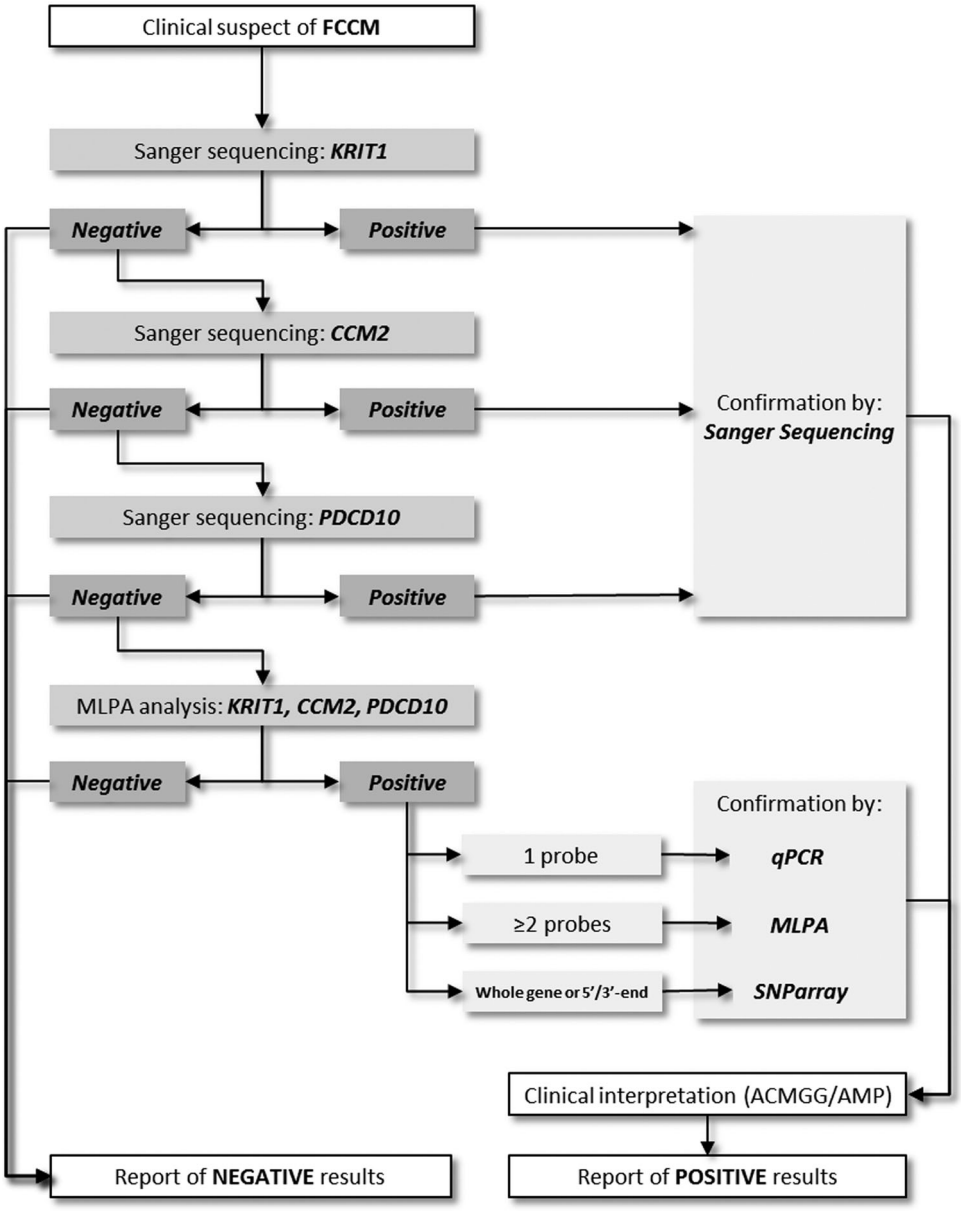


FIGURE 1 2004–2017 diagnostic workflow for FCCM. ACMGG/AMP, American College of Medical Genetics and Genomics/Association for Molecular Pathology; MLPA, multiplex ligation-dependent probe amplification; qPCR, quantitative polymerase chain reaction

likelihood of the association between the identified variant and the disease, we considered the highest minor allele frequency (MAF) score. Variants were also checked if previously reported in the literature, as well as in clinically relevant databases: Leiden Open Variation Database (LOVD, http://www.lovd.nl; Fokkema et al., 2011) and Clinical Variations database (ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/). Polyphen-2 (version 2.2.2, available at: http://genetics.bwh.harvard.edu/pph), Eigen (version 1.1.3, available at: https://omictools.com/eigen-tool), SIFT (version 1.03, available at: http://sift.jcvi.org/), MutationTaster 2 (Schwarz, Cooper, Schuelke, & Seelow, 2014), MetaSVM and MetaLRT (Dong et al., 2015), M-CAP (Jagadeesh et al., 2016), CADD (version 1.3,42), DANN (Quang, Chen, & Xie, 2015), FATHMMMKL (Shihab et al., 2015), and MutationAssessor (Reva, Antipin, & Sander, 2007) were used for in silico prediction of pathogenicity of missense variants. The majority (≥75%) of in silico predictors was arbitrarily used as the cut-off for the attribution of the corresponding supporting criterion of pathogenicity. These tools were selected because of their maintenance frequency, estimation congruency and/or superior classification records (Castellana & Mazza, 2013). The effects of splicing variants were predicted by using Human Splicing Finder (HSF 3.1, version 3.1, http://www.umd.be/HSF/), Net2Gene (http://www.cbs.dtu.dk/services/NetGene2), and Splice Site Prediction by Neural Network (BDGP, http://www.fruitfly.org/seq_tools/splice.html). In silico prediction, analyses followed the guidelines reported in (Vihinen, 2013).

TABLE 1 Characteristics of the target genes for next-generation sequencing analysis

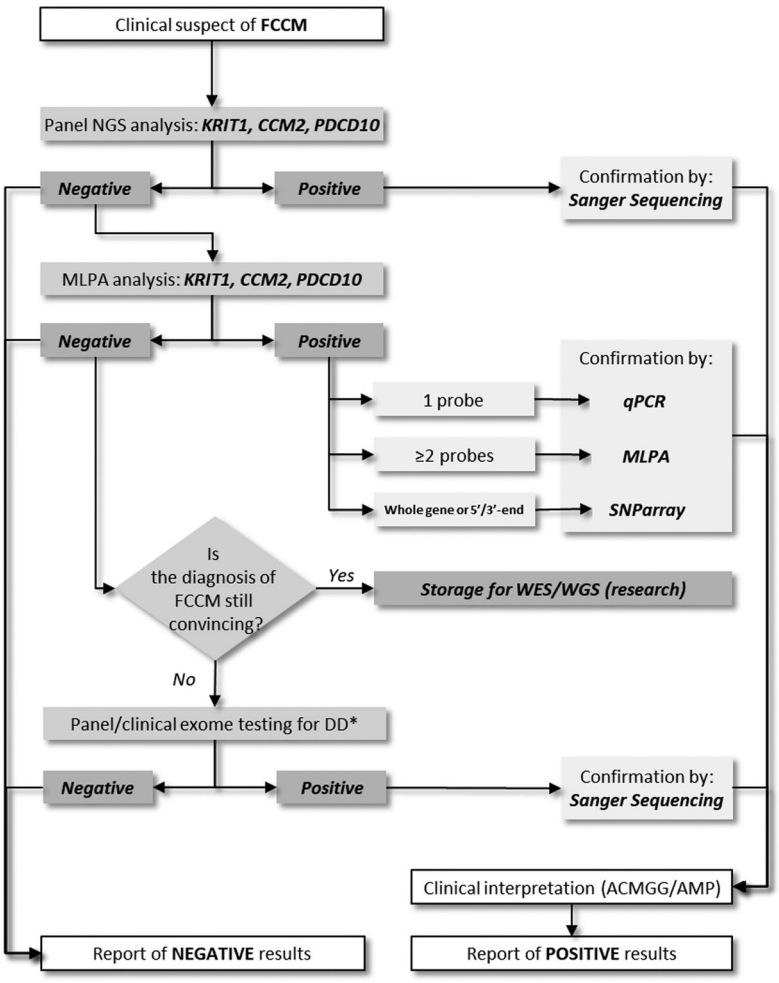
Target gene	Interval	Regions	Size	Reference sequence	Coverage
KRIT1	chr7:91830025-91871474	16	3011	NM_194456.1	100.0
CCM2	chr7:45039908-45115681	11	1978	NM_031443.3	100.0
PDCD10	chr3:167402071-167437970	7	989	NM_007217.3	100.0

4.3 | Phenotype description for storing of data

The Human Phenotype Ontology (HPO; https://hpo.jax.org/) has been developed with the goal to cover all phenotypic abnormalities that are commonly encountered in human monogenic diseases (Robinson et al., 2008). The structure of the HPO allows flexible searches for disease entities according to phenotypic abnormalities, with a broad or narrow focus. Accordingly, a list of selected clinical features and their relative HPO definitions were identified for each FCCM gene (see tables in Nardella et al., 2018) and this approach was followed for uploading clinical data on public databases.

4.4 | Statistical methods for genotype–phenotype correlations

Patients' characteristics were reported as means and standard deviations (or median and min–max) for continuous variables, and as frequencies and percentages for categorical variables. Group comparisons were performed using the Kruskall–Wallis test or Fisher exact test for continuous and categorical variables, respectively. A p value < .05 was considered statistically significant. All statistical analyses were performed using the R program (a language and environment for statistical computing).



*Genes to consider for further molecular testing

Alternative/additional CNS vascular anomaly	Candidate genes
Intracranial arteriovenous malformation	*ACVRL1, ARHGAP31, DLL4, DOCK6, ENG, EOGT, EPHB4, KRAS (somatic), GDF2, NOTCH1, PTEN, RASA1, RBPJ, SMAD4*
Aneurysm of the Galen vein	*EPHB4, RASA1*
Moyamoya disease	*ACTA2, GUCY1A3, JAG1, NF1, NOTCH2, PCNT, RNF213, SMARCAL1*
Intracranial arterial dissection	*COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, COL5A1, COL5A2, FBN1, PLOD1, SCL2A10, SMAD3, TGFB2, TGFB3, TGFBR1, TGFBR2*
Brain and/or spinal emangioblastomas	*VHL*

FIGURE 2 2017-present diagnostic workflow for familial cerebral cavernous malformation (FCCM). ACMGG/AMP, American College of Medical Genetics and Genomics/Association for Molecular Pathology; CNS, central nervous system; DD, differential diagnosis; MLPA, multiplex ligation-dependent probe amplification; qPCR, quantitative polymerase chain reaction; WES, whole exome sequencing; WGS, whole genome sequencing

TABLE 2 Clinical interpretation of 56 variants in *KRIT1*, *CCM2*, and *PDCD10*

Gene	Nucleotide variant	Predicted aminoacid change	Pathogenicity criteria										Clinical interpretation
			A	B	C	D	E	F	G	H	I	J	
KRIT1	c.151_154delAAAG	p.(Lys51Phefs*13)	+				+	+		+	+		Pathogenic
	c.196_197delCA	p.(Gln66Argfs*54)	+			+				+	+		Pathogenic
	c.206T>A	p.(Leu69*)	+			+				+	+		Pathogenic
	c.646_647delAA	p.(Lys216Glufs*2)	+			+				+	+		Pathogenic
	c.703G>A	p.(Gly235Arg)		+		+			+	+	+		Likely pathogenic
	c.729+1G>A	p.(Val244Glyfs*7)			+		+	+			+		Pathogenic
	c.746delT	p.(I249Kfs*7)	+			+				+	+		Pathogenic
	c.763delC	p.(Leu255*)	+			+					+		Pathogenic
	c.825delG	p.(Met275Ilefs*13)	+			+				+	+		Pathogenic
	c.880C>T	p.(Arg294*)	+				+	+			+		Pathogenic
	c.966G>A	p.(Trp322*)	+			+					+		Pathogenic
	c.(845+1_846−1)_(989+1_990−1)del	p.(?)	+			+				+	+		Pathogenic
	c.1048C>T	p.Leu350Phe		+		+			+	+	+		Likely pathogenic
	c.990_1146del	p.Trp330Cysfs*3	+			+					+		Pathogenic
	c.1249_1252delAAAC	p.Lys417Hisfs*19	+			+				+	+		Pathogenic
	c.(1146+1_1147−1)_(1254+1_1255−1)del	p.(?)	+			+					+		Pathogenic
	c.1255−1_1256del	p.(Tyr419Phefs*15)	+				+	+			+		Pathogenic
	c.1267C>T	p.(Arg423*)	+				+	+		+			Pathogenic
	c.1292_1293delAT	p.(Tyr431Serfs*4)	+			+				+	+		Pathogenic
	c.1362_1363delTC	p.(Gln455Argfs*24)	+				+	+		+	+		Pathogenic
	c.1363C>T	p.(Gln455*)	+				+	+		+	+		Pathogenic
	c.1444C>T	p.(Gln482*)	+			+				+	+		Pathogenic
	c.1505_1511delAAACACC	p.(Glu502Valfs*5)	+			+				+	+		Pathogenic
	c.1517T>G	p.(Leu506Arg)		+		+			+	+	+		Likely pathogenic
	c.1524_1528delAAGAA	p.(Arg510Cysfs*8)	+			+				+	+		Pathogenic
	c.1579_1588delGCTATTCTTA	p.(Ala527Phefs*11)	+			+					+		Pathogenic
	c.1688_1689delAT	p.(Tyr563Trpfs*4)	+			+					+		Pathogenic
	c.(1730+1_1731−1)_(1818+1_1819−1)del	p.(Asn577Lysfs*55)	+			+					+		Pathogenic
	c.1877T>A	p.(Leu626*)	+			+				+	+		Pathogenic
	c. 1979_1980dupTA	p.(Gly661*)	+			+				+	+		Pathogenic
	c.2020dup	p.(Thr674Asnfs*2)	+			+					+		Pathogenic
	c.2026−3C>A	p.(Ala676_Gln714del)				+				+	+	+	Likely pathogenic
	c.2026−12A>G	p.(Ala676_Gln714del)					+			+	+		Likely pathogenic
	c.2025+1G>A	p.(Asn607_Lys675del)			+	+					+		Pathogenic
	c.2029_2030insA	p.(Leu677Tyrfs*5)	+			+				+	+		Pathogenic
	c.(1818+1_1819−1)_(2142+1_2143−1)del	p.(Asn607_Gln714del)	+			+					+		Pathogenic
	c.(?_−1)_(*1_?)del	p.(?)	+			+				+	+		Pathogenic
CCM2	c.(?_−848)_(30+1_31−1)del	p.(?)	+			+				+	+		Pathogenic
	c.55C>T	p.(Arg19*)	+				+	+		+	+		Pathogenic
	c.134_135delTG	p. (Val45Glyfs*6)	+			+					+		Pathogenic
	c.205−1G>A	p.(?)			+	+					+		Pathogenic
	c.(30+1_31−1)_(204+1_205−1)del	p.(Pro11_Lys68del)	+			+					+		Pathogenic
	c.500_501delAG	p.(Leu169Valfs*66)	+			+					+		Pathogenic
	c.779delC	p.(Tyr261Thrfs*31)	+			+					+		Pathogenic

(Continues)

TABLE 2 (Continued)

Gene	Nucleotide variant	Predicted aminoacid change	Pathogenicity criteria										Clinical interpretation
			A	B	C	D	E	F	G	H	I	J	
	c.(915+1_916−1)_(1054+1_1055−1)del	p.(Leu306Valfs*2)	+			+					+		Pathogenic
PDCD10	g.167,734,826_167,734,214	p.(?)	+			+					+		Pathogenic
	c.(−117_−116)_(96+1_97−1)	p.(?)	+			+					+		Pathogenic
	c.103C>T	p.(Arg35*)	+			+				+	+		Pathogenic
	c.(96+1_97−1)_(268+1_269−1)del	p.(Leu33Serfs*3)	+			+					+		Pathogenic
	c.(150+1_151−1)_(268+1_269−1)del	p.(Ala51Serfs*3)	+			+					+		Pathogenic
	c.283C>T	p.(Arg95*)	+				+				+		Pathogenic
	c.586C>T	p.(Arg146*)	+			+				+	+		Pathogenic
	g.(167,256,020_167,256,021)_ (167,440,972_167,440,973) del*	p.(?)	+			+				+	+	+	Pathogenic
	c.(?_−1)_(*1_?)del	p.(?)	+			+							Pathogenic

Note: A, *null* (nonsense, frameshift) variant in a gene previously described as disease-causing by haploinsufficiency or loss-of-function.
B, missense variant located in a critical and well-established functional domain.
C, variant affecting canonical splicing sites (i.e., ±1 or ±2 positions).
D, variant absent in allele frequency population databases.
E, variant reported in allele frequency population databases but with a MAF significantly lower than the known disease frequency in the general population.
F, variant predicted as pathogenic/deleterious in ClinVar and/or LOVD.
G, missense variant defined pathogenic/deleterious in most (≥75%) of the selected in silico predictors (see text).
H, variant co-segregating in two or more affected relatives.
I, the reported phenotype (including laboratory and instrumental findings) is highly specific for the selected disease-gene.
J, the predicted pathogenic effect has been confirmed by appropriate functional study/ies.
KRIT1: NM_194456.1 sequencing and NM_19446.1 for MLPA analysis. *CCM2*: NM_031443.3. *PDCD10*: NM_007217.3. "g." indicates variant genomic coordinates identified by SNParray and according to the GRCh37/hg19 build of the Human Genome.

5 | DATA

5.1 | Variant interpretation

In our cohort of 140 FCCM individuals with germline variants in *KRIT1*, *CCM2*, and *PDCD10*, 56 different variants were identified (Nardella et al., 2018). Clinical interpretation of the 54 identified variants is reported in Table 2. Five (9%) were likely pathogenic and 51 (91%) pathogenic. No VUS were identified among the 56 variants, according to the ACMGG/AMP guidelines. Prediction data on the five splicing variants are reported in Table 3. The output of the analysis performed with three different online tools suggested that all identified splicing variants can be scored as likely pathogenic. Three of them affected canonical splice sites and, therefore, were conclusively interpreted as pathogenic according to (Richards et al. 2015).

5.2 | Human phenotype ontology

Clinical information available in the companion paper by Nardella et al. (2018) is in according to this nomenclature.

TABLE 3 In silico predictions for splicing variants

Gene	Exon	DNA variation	Predicted AA change	HSF 3.1	Net2Gene	BDGP	Clinical prediction
KRIT1	9	c.729+1G>A	p.(Val244Glyfs*7)	92.81 vs 65.97	67% vs 0%	0.99 vs 0.00	Likely pathogenic
KRIT1	17	c.2026−3C>A	p.(Ala676_Gln714del)	88.08 vs 78.69	15% vs 0%	0.69 vs 0.00	Likely pathogenic
KRIT1	17	c.2026−12A>G	p.(Ala676_Gln714del)	88.08 vs 88.27	15% vs 0%	0.69 vs 0.67	Likely pathogenic
KRIT1	18	c.2025+1G>A	p.(Asn607_Lys675del)	78.54 vs 66.76	15% vs 0%	1.00 vs 0.00	Likely pathogenic
CCM2	3	c.205−1G>A	p.(?)	87.81 vs 81.29	17% vs 0%	0.87 vs 0.00	Likely pathogenic

Note: HSF 3.1: Consensus values range from 0 to 100 while the threshold is defined at 65.
Net2Gene: The level of confidence is relative to the cutoff used to find nearly all true sites set up at 50%.
BDGP: Splice site predictions have a cutoff for donor and acceptor site score of 0.40.

TABLE 4 Demographic and clinical characteristics of available individuals by affected gene

Characteristic	*KRIT1*	*CCM2*	*PDCD10*	*p*
No. of individuals	28	3	6	
No. of pedigrees	13	2	2	
Age at diagnosis (median [range])	36 years (2–70 years)	61 years (55–76 years)	23.5 years (19–54 years)	.06
Females (%)	19 (67.9)	1 (33.3)	2 (33.3)	.176
Males (%)	9 (32.1)	2 (66.7)	4 (66.7)	
Familial (%)	21 (75.0)	3 (100.0)	4 (66.7)	.679
Sporadic (%)	7 (25.0)	0 (0.0)	2 (33.3)	
No symptom (%)	7 (25.0)	0 (0.0)	2 (33.3)	.679
Symptomatic (%)	21 (75.0)	3 (100.0)	4 (66.7)	
Symptom onset (median [range])	22 years (0.6–68 years)	39 years (16–44 years)	30.5 years (18–50)	.728
Headache migraine (%)	12 (42.9)	2 (66.7)	2 (33.3)	.727
Epilepsy (%)	6 (21.4)	1 (33.3)	0 (0.0)	.35
Intracranial haemorrhage (%)	5 (17.9)	0 (0.0)	2 (33.3)	**.609**
Dementia (%)	1 (3.6)	0 (0.0)	0 (0.0)	1
Behavioral/psychiatric abnormalities (%)	2 (7.1)	0 (0.0)	4 (66.7)	.006
Fatigability/weakness of skeletal muscles (%)	3 (10.7)	1 (33.3)	1 (16.7)	.352
Hemiparesis (%)	3 (10.7)	0 (0.0)	1 (16.7)	.69
Sovratentorial hemangiomas (%)	8 (28.6)	0 (0.0)	2 (33.3)	.686
Cerebellar hemangiomas (%)	18 (64.3)	3 (100.0)	2 (33.3)	.163
Spinal hemangiomas (%)	2 (7.1)	0 (0.0)	2 (33.3)	.141
Bilateral lesions (%)	6 (21.4)	1 (33.3)	1 (16.7)	.816
Intracranial meningiomas (%)	0 (0.0)	0 (0.0)	0 (0.0)	–
Hepatic hemangiomas (%)	1 (3.6)	0 (0.0)	0 (0.0)	1

Note: Significant *p* values are in bold.

5.3 | Genotype–phenotype correlations

Complete demographic and clinical data were available for 37 individuals out of 140 (26.4%) (see results in Nardella et al. (2018)). Distribution of characteristics by affected genes (*KRIT1*, *CCM2*, and *PDCD10*) is reported in Table 4 and Figure 3. The 28 individuals with *KRIT1* variants, the three with *CCM2* variants and the six with *PDCD10* variants belonged to 13, 2 and 2 pedigrees, respectively. Although available data are relatively scanty compared to the whole samples' cohort, our analysis showed a lack of any consistent genotype-phenotype correlation among patients with available clinical information. This is in line with the literature, which does not indicate the existence of any key clinical features predicting the involved gene. The unique significant association, but highly

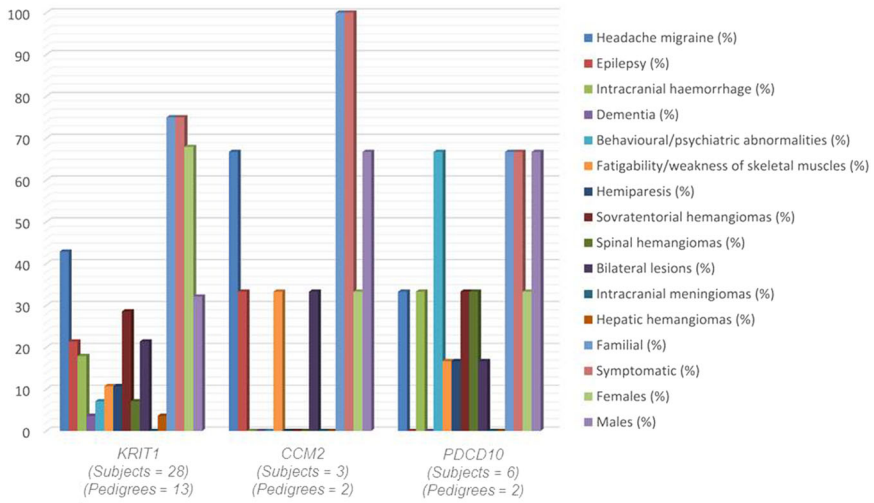


FIGURE 3 Relative frequencies of patients' demographical and clinical characteristics according to a mutated gene

TABLE 5 Data submitted to the Leiden open variation database

ID*	Ethnicity	Geo-graphic origin	# of carrying subjects	OMIM no.	Inheri-tance	Phenotype	Technique	Tissue	Allele**	Chr	Type	Genomic/DNA	Transcript	Effect id***	DNA	Protein	EXON
1	Caucasian	Italy	5	607929	Familial		MLPA	Blood	0	7	del	g.(?_45039085)_(45039963_45077851)del	NM_031443.3	7	c.(?_−848)_(30+1_31−1)del	p.(?)	5′ UTR
2	Caucasian	Italy	1	607929	Familial		MLPA	Blood	0	7	del	g.45113171_45115375del	NM_031443.3	7	c.(915+1_916−1)_(1054+1_1055−1)del	p.(Leu306-Valfs*2)	10
3	Caucasian	Italy	1	609118	Familial		MLPA	Blood	0	3	del	g.167413511_167422629del	NM_145860.1	7	c.(150+1_151−1)_(268+1_269−1)del	p.(Ala51-Serfs*3)	5
4	Caucasian	Italy	1	609118	Familial		MLPA	Blood	0	3	del	g.167413511_167437849del	NM_145860.1	7	c.(96+1_97−1)_(268+1_269−1)del	p.(Leu33-Serfs*3)	5
5	Caucasian	Italy	1	604214	Familial		MLPA	Blood	0	7	del	g.91842716_91843924del	NM_194456.1	7	c.(1730+1_1731−1)_(1818+1_1819−1)del	p.(Asn577-Lysfs*55)	17
6	Caucasian	Italy	2	604214	Familial	Cerebellar hemangiomas	Sanger	Blood	0	7	subst	g.91855938G>A	NM_194456.1	7	c.1048C>T	p.(Leu350-Phe)	11
7	Caucasian	Italy	2	604214	Familial		Sanger	Blood	0	7	subst	g.91851262A>C	NM_194456.1	7	c.1517T>G	p.(Leu506-Arg)	14
8	Caucasian	Italy	4	604214	Familial	Intracranial haemorrage, fatigable weakness of skeletal (HP:0030197), cerebellar hemangiomas, bilateral lesion	Sanger	Blood	0	7	subst	g.91864743C>T	NM_194456.1	7	c.703G>A	p.(Gly235-Arg)	8
9	Caucasian	Italy	1	609118	Familial	Headache (HP:0002076), sovratentorial hemangiomas, behavioral psychiatric abnormalities (HP:0000708)	Sanger	Blood	0	3	subst	g.167405441G>A	NM_145860.1	7	c.436C>T	p.(Arg95*)	7
10	Caucasian	Italy	2	604214	Familial		Sanger	Blood	0	7	dup	g.91842554_91842555dup	NM_194456.1	7	c.1979_1980dup-TA	p.(Gly661*)	17

(Continues)

TABLE 5 (Continued)

ID*	Ethnicity	Geo-graphic origin	# of carrying subjects	OMIM no.	Inheri-tance	Phenotype	Technique	Tissue	Allele**	Chr	Type	Genomic/DNA	Transcript	Effect id***	DNA	Protein	EXON
11	Caucasian	Italy	5	604214	Familial		Sanger	Blood	0	7	ins	g.91842604_91842605insA	NM_194456.1	7	c.2029_2030insA	p.(Leu677-Tyrfs*5)	18
12	Caucasian	Italy	4	609118	Familial	Headache (HP:0002076), intracranial hemorrhage, behavioral psychiatric abnormalities (HP:0000708), hemiparesis (HPO:0001269) sovratentorial hemangiomas, cerebellar hemangiomas	SNP-Array	Blood	0	3	del	g.167256020_167440972del	NM_145860.1	9	g.(167,256,020_167,256,021)_(167,440,972_167,440,973)del*	p.(?)	exo-n4--10
13	Caucasian	Italy	1	609118	Familial		MLPA	Blood	0	3	del	g.(167422684_167437849)_(167438062_167452001)del	NM_145860.1	7	c.(−117_−116)_(96+1_97−1)	p.(?)	3
14	Caucasian	Italy	1	604214	Familial		Sanger	Blood	0	7	del	g.91816381_91844076del	NM_194456.1	7	c.1579_1588del-GCTATTCTTA	p.(Ala527-Phefs*11)	15
15	Caucasian	Italy	3	604214	Familial		Sanger	Blood	0	7	del	g.91870372_91870373del	NM_194456.1	7	c.193_194delAC	p.(Gln66A-rgfs*54)	5
16	Caucasian	Italy	1	607929	Familial		MLPA	Blood	0	7	del	g.45039963_45103516del	NM_031443.3	7	c.(30+1_31−1)_(204+1_205−1)del	p.(Pro11_-Lys68del)	3
17	Caucasian	Italy	1	607929	Familial	Headache (HP:0002076), cerebellar hemangiomas	Sanger	Blood	0	7	del	g.45077955_45077956del	NM_031443.3	7	c.134_135delTG	p.(Val45-Glyfs*6)	2
18	Caucasian	Italy	1	604214	Familial		MLPA	Blood	0	7	del	g.91852293_91855839del	NM_194456.1	7	c.(1146+1_1147−1)_(1254+1_1255−1)del	p.(?)	13
19	Caucasian	Italy	2	604214	Familial		Sanger	Blood	0	7	del	g.91852184_91852185del	NM_194456.1	7	c.1356_1357del-TC	p.(Gln455-Argfs*24)	13
20	Caucasian	Italy	2	604214	Familial		Sanger	Blood	0	7	del	g.91851268_91851274del	NM_194456.1	7	c.1505_1511de-lAAACACC	p.(Glu502-Valfs*5)	14

(Continues)

TABLE 5 (Continued)

ID*	Ethnicity	Geo-graphic origin	# of carrying subjects	OMIM no.	Inheri-tance	Phenotype	Technique	Tissue	Allele**	Chr	Type	Genomic/DNA	Transcript	Effect id***	DNA	Protein	EXON
21	Caucasian	Italy	4	604214	Familial	Headache (HP:0002076), cerebellar hemangiomas, sovratentorial hemangiomas	Sanger	Blood	0	7	del	g.91851251_91851255del	NM_194456.1	7	c.1524_1528de-lAAGAA	p.(Arg510-Cysfs*8)	14
22	Caucasian	Italy	3	604214	Familial		Sanger	Blood	0	7	del	g.91870372_91870373del	NM_194456.1	7	c.196_197delCA	p.(Gln66A-rgfs*54)	5
23	Caucasian	Italy	3	604214	Familial		Sanger	Blood	0	7	subst	g.91870363A>T	NM_194456.1	7	c.206T>A	p.(Leu69*)	5
24	Caucasian	Italy	2	604214	Familial		Sanger	Blood	0	7	del	g.91864799_91864800del	NM_194456.1	7	c.646_647delAA	p.(Lys216-Glufs*2)	8
25	Caucasian	Italy	2	604214	Familial		Sanger	Blood	0	7	del	g.91864221del	NM_194456.1	7	c.746delT	p.(Ile249-Lysfs*7)	9
26	Caucasian	Italy	2	604214	Familial		Sanger	Blood	0	7	del	g.91864204del	NM_194456.1	7	c.763delC	p.(Leu255*)	9
27	Caucasian	Italy	5	604214	Familial		Sanger	Blood	0	7	del	g.91864142del	NM_194456.1	7	c.825delG	p.(Met275Ilefs*13)	9
28	Caucasian	Italy	1	604214	Familial		Sanger	Blood	0	7	subst	g.91863786C>T	NM_194456.1	7	c.966G>A	p.(Trp322*)	10
29	Caucasian	Italy	3	604214	Familial		MLPA	Blood	0	7	del	g.91855840_91855996del	NM_194456.1	7	c.990_1146del	p.(Trp330-Cysfs*3)	12

preliminary was with behavioral/psychiatric abnormalities (p =0 .006), as they occurred in two patients with *KRIT1* variants (7.1%), none patients with *CCM2* variants and four patients with *PDCD10* variants (66.7%). As six patients only reported behavioral/psychiatric abnormalities and four of them belonged to a single *PDCD10* family, this evidence might be artefactual and needs further investigations.

5.4 | Data submission to public databases

Laboratories are encouraged to contribute to public variant databases to share information relevant to both clinics and research. Accordingly, all novel variants reported in Nardella et al. (2018) were stored into LOVD. Each variant uploaded into the database was supported by information at the DNA, RNA, and protein levels, in accordance to the HGVS nomenclature (http://varnomen.hgvs.org/; den Dunnen et al., 2016). Genotype, variant effect, and pathogenicity were described as detailed as possible in combination with the HPO nomenclature, when possible. Additional inheritance information and data related to the type of technique used for diagnosis were also annotated. Data submitted to LOVD are summarized in Table 5.

6 | DISCUSSION

The above-reported procedures and data illustrate a highly heterogeneous genetic architecture, underlying still largely unraveled genotype-phenotype correlations in FCCM. Although most causative variants are null heterozygous alleles in FCCM, its marked allelic heterogeneity and the possible identification of missense variants prompt the need of applying a multidimensional evaluation to sequencing data for clinical reporting in many cases. The integration of (a) different molecular tools covering the widest spectrum of possible mutational events (e.g., NGS, MLPA, and genomic array), (b) population data for MAF estimation, (c) previously published information on the presumed clinical significance of known variants, (d) in silico prediction for missense and splicing variants, and, perhaps, (e) functional studies validated for clinical use, and (f) data on soft clinical features (e.g., meningiomas or cutaneous vascular malformations) in favor of the diagnosis (in the absence of an overt brain clinical picture), appear prerequisites for qualified activities in diagnostic laboratories. Translational research should be also in continuity with the clinical practice for optimal implementation of the diagnostic output, especially in the case of VUS. In this scenario, contributing to public variant databases (especially those which include clinical and functional data) is emerging as an essential activity in diagnostic services, for both clinical and research purposes. This is extremely valid in FCCM, in which a relatively well-known genetic background does not explain the observed clinical variability, and the presumed molecular pathogenesis may be considered a promising target for novel therapies. A disease-specific database including accurate radiological findings or, perhaps, images will also help in exploring genotype-phenotype correlations, that might support the clinical interpretation of selected variants and/or stimulate researchers in studying the natural history(ies) of FCCM.

ACKNOWLEDGMENTS

The authors thank all families that participated in the study. The authors are also grateful to the Genomic and Genetic Disorder Biobank, member of the Telethon Network of Genetic Biobanks (Telethon, Italy grant GTB12001) and the EuroBioBank for providing the specimens. This work was supported by the Ricerca Corrente 2018–2020 Program from the Italian Ministry of Health. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

ORCID

Carmela Fusco http://orcid.org/0000-0001-7794-6046
Lucia Micale http://orcid.org/0000-0003-3604-194X

REFERENCES

Castellana, S., & Mazza, T. (2013). Congruency in the prediction of pathogenic missense mutations: State-of-the-art web-based tools. *Briefings in Bioinformatics*, *14*(4), 448–459. https://doi.org/10.1093/bib/bbt013

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., et al. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human Molecular Genetics*, *24*(8), 2125–2137. https://doi.org/10.1093/hmg/ddu733

den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., et al. (2016). HGVS recommendations for the description of sequence variants: 2016 Update. *Human Mutation*, *37*(6), 564–569. https://doi.org/10.1002/humu.22981

Fokkema, I. F., Taschner, P. E., Schaafsma, G. C., Celli, J., Laros, J. F., & den Dunnen, J. T. (2011). LOVD v.2.0: The next generation in gene variant databases. *Human Mutation*, *32*(5), 557–563. https://doi.org/10.1002/humu.21438

Jagadeesh, K. A., Wenger, A. M., Berger, M. J., Guturu, H., Stenson, P. D., Cooper, D. N., et al. (2016). M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nature Genetics*, *48*(12), 1581–1586. https://doi.org/10.1038/ng.3703

Nardella, G., Visci, G., Guarnieri, V., Castellana, S., Biagini, T., Bisceglia, L., et al. (2018). A single-center study on 140 patients with cerebral cavernous malformations: 28 new pathogenic variants and functional characterization of a PDCD10 large deletion. *Human Mutation*, *39*(12), 1885–1900. https://doi.org/10.1002/humu.23629

Quang, D., Chen, Y., & Xie, X. (2015). DANN: A deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*, *31*(5), 761–763. https://doi.org/10.1093/bioinformatics/btu703

Reva, B., Antipin, Y., & Sander, C. (2007). Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biology*, *8*(11), R232. https://doi.org/10.1186/gb-2007-8-11-r232

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (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. *Genetics in Medicine*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

Robinson, P. N., Kohler, S., Bauer, S., Seelow, D., Horn, D., & Mundlos, S. (2008). The Human Phenotype Ontology: A tool for annotating and analyzing human hereditary disease. *American Journal of Human Genetics*, *83*(5), 610–615. https://doi.org/10.1016/j.ajhg.2008.09.017

Schwarz, J. M., Cooper, D. N., Schuelke, M., & Seelow, D. (2014). MutationTaster2: Mutation prediction for the deep-sequencing age. *Nature Methods*, *11*(4), 361–362. https://doi.org/10.1038/nmeth.2890

Scimone, C., Donato, L., Katsarou, Z., Bostantjopoulou, S., D'Angelo, R., & Sidoti, A. (2018). Two Novel KRIT1 and CCM2 mutations in patients affected by cerebral cavernous malformations: New information on CCM2 penetrance. *Frontiers in Neurology*, *9*, 953. https://doi.org/10.3389/fneur.2018.00953

Shihab, H. A., Rogers, M. F., Gough, J., Mort, M., Cooper, D. N., Day, I. N., et al. (2015). An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics*, *31*(10), 1536–1543. https://doi.org/10.1093/bioinformatics/btv009

Vihinen, M. (2013). Guidelines for reporting and using prediction tools for genetic variation analysis. *Human Mutation*, *34*(2), 275–282. https://doi.org/10.1002/humu.22253

How to cite this article: Fusco C, Copetti M, Mazza T, et al. Molecular diagnostic workflow, clinical interpretation of sequence variants and data repository procedures in 140 individuals with familial cerebral cavernous malformations. *Human Mutation*. 2019;40:e24–e36. https://doi.org/10.1002/humu.23851