Supplemental Material

Comprehensive Transcriptome Analysis of Cerebral Cavernous Malformation Across Multiple Species and Genotypes

Janne Koskimäki [*,1]; Romuald Girard[*,1]; Yan Li[2]; Laleh Saadat[1]; Hussein A. Zeineddine[1]; Rhonda Lightle[1]; Thomas Moore[1]; Seán Lyne[1]; Kenneth Avner[1]; Robert Shenkar[1]; Ying Cao[1]; Changbin Shi[1]; Sean P. Polster[1]; Dongdong Zhang[1]; Julián Carrión-Penagos[1]; Sharbel Romanos[1]; Gregory Fonseca[3]; Miguel A. Lopez-Ramirez[4]; Eric M. Chapman[5]; Evelyn Popiel[5]; Alan T. Tang[6]; Amy Akers[7]; Pieter Faber[8]; Jorge Andrade[2]; Mark Ginsberg[4]; W. Brent Derry[5,9]; Mark L. Kahn[6]; Douglas A. Marchuk[10]; Issam A. Awad[1]

From:
[1]Neurovascular Surgery Program, Section of Neurosurgery, The University of Chicago Medicine and Biological Sciences, Chicago, IL
[2]Center for Research Informatics, The University of Chicago, Chicago, IL
[3]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA
[4]Department of Medicine, University of California San Diego, La Jolla, CA
[5]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
[6]Department of Medicine and Cardiovascular Institute, University of Pennsylvania, Philadelphia, PA
[7]Angioma Alliance Norfolk, Virginia
[8]University of Chicago Genomics Facility, The University of Chicago, Chicago, IL
[9]Program in Developmental and Stem Cell Biology, Hospital for Sick Children, Toronto, ON, Canada.
[10]The Molecular Genetics and Microbiology Department, Duke University Medical Center, Durham, NC
*Equal Contribution
Correspondence to:
Issam A. Awad, MD, MSc, FACS, MA (hon), Section of Neurosurgery, University of Chicago Medicine, 5841 S. Maryland, MC3026/Neurosurgery J341, Chicago, IL 60637.
Telephone +1 773-702-2123; Fax +1 773-702-3518. E-mail iawad@uchicago.edu

Supplemental Material and Methods

Patient demographics

The descriptive data of the subjects for whom 5 human cerebral cavernous malformation (CCM) lesions were surgically resected are presented in **Table S1**. Neurovascular unit (NVU) controls were collected from the brains of autopsy subjects free of neurological diseases.

Laser capture microdissection

NVUs from CCMs and normal capillaries were collected using laser capture microdissection at 40x magnification, and cut NVUs were subsequently stored at -80°C (Leica LMD 6500, Leica Biosystems Inc.) (1).

Bioinformatics analyses

The quality of raw sequencing reads was assessed by FastQC (v0.11.2; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The post-alignment quality control was evaluated using RSeQC(2), and Picard tools (v1.117; http://broadinstitute.github.io/picard/). Reads were mapped to University of California Santa Cruz (UCSC) human genome model (hg19), GENCODE mouse model (GRCm38), Ensembl *Caenorhabditis elegans* genome model (WBcel235) separately with respect to different species with STAR v2.5.3a (3). Gene transcripts were assembled and quantified using featureCounts (4) with respect to corresponding reference genome gene annotations.

Differentially expressed genes (DEGs) analyses were conducted using DESeq2(5), with an additive model for batch effect correction if necessary. The human homologous genes of DEGs from mouse and *C. elegans* animal models were annotated based on the National Center for Biotechnology Information (NCBI) HomoloGene database. Out of 18,038 corresponding human genes in the database, 16,200 (around 90%) genes from mouse and 3013 (~17%) genes from *C. elegans* have corresponding human homologous genes. Because the homologous genes limitation from *C. elegans*, we implemented different threshold to DEGs from *C. elegans* ($p \leq 0.25$, false discovery rate (FDR) corrected; and |fold change (FC)|≥1.2) compared to human lesional ECs and the 2 murine models ($p \leq 0.05$, FDR corrected; and |FC|≥1.2) when comparing all the models together.

Flow chart of the study is presented in **Figure S1**. We implemented 2 approaches to study the transcriptome of CCM models using sub-network analysis and gene ontology (GO) enrichment analyses. In the first, DEGs of engineered animal models were compared with the DEGs from human lesional NVUs to identify common DEGs across species (*Ccm1/Krit1* or *Ccm3/Pdcd10*). In the second approach, genotype specific associated DEGs from mouse and *C. elegans* were compared separately.

Both approaches were followed with GO enrichment analyses, network essential genes identification, network clustering analysis, and where sub-network analysis, network essential genes identification, and network clustering analysis. GO enrichment analysis was conducted with R bioconductor package clusterProfiler. Network analyses were performed with ReactomeFIViz (6) in Cytoscape (http://www.cytoscape.org/) based on a highly reliable Reactome function interaction network (7) (version 2016). The entire FI network was constructed by merging interactions extracted from human curated pathways with interactions predicted using a machine learning approach.

The individual *CCM* gene expression profiles for each human patient have been generated following trimmed mean of M-values (TMM) normalization implemented into the raw count reads after low expressed gene features removal. The TMM normalized expression values can be compared directly across different patients and genes.

Supplemental Results

RNA quality

For each sample of laser-captured NVUs from human surgically resected CCM lesion and normal capillaries the concentration of RNA was between 145 and 3360 pg/µl (RNA integrity numbers = [1.8 - 5.1]). For each sample of BMEC models and their respective controls, the concentration of RNA was between 322 and 89 ng/µl (RNA integrity numbers = [6.0 - 8.7]). Similarly, for *C. elegans* models and their respective controls the concentration of RNA was between 55 and 309 ng/µl (RNA integrity numbers = [9.8 - 10]).

Definition of neurovascular unit and restrictions of laser capture microdissection in humans

NVUs in capillaries are composed of endothelial cells, pericytes, astrocytes and neurons (8). As we acknowledge the technical limitations of laser capture microdissection to isolate pure capillary endothelial cells **(Figure S7, video)**, we determined the proportion of DEGs that may be attributed to other cell types than endothelial cells.

We first defined the top 500 human homologous DEGs derived for each of the following analyses: neurons *versus* endothelial cell (ECs), astrocytes *versus* ECs, new oligodendrocytes *versus* ECs, myelinating oligodendrocytes *versus* ECs and microglia *versus* ECs (https://web.stanford.edu/group/barres_lab/brain_rnaseq.html) (9). We then cross-compared each top 500 DEGs to the 915 DEGs from human lesional NVUs. The results showed that 655 of the 915 DEGs (72%) in the human lesional NVUs represented EC DEGs versus other non-EC cell types. Among the 260 non-EC DEGs, 108 were common DEGs with neurons, 50 with astrocytes, 70 with new oligodendrocytes, 72 with myelinating and 108 with microglia.

Supplementary References

1. Espina V, Milia J, Wu G, Cowherd S, Liotta LA. Laser capture microdissection. *Methods Mol Biol.* 2006;319:213-29.
2. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*. 2012;28(16):2184-5.
3. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
4. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-30.
5. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
6. Wu G, Dawson E, Duong A, Haw R, Stein L. ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis. *F1000Res*. 2014;3:146.
7. Wu G, Feng X, Stein L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol*. 2010;11(5):R53.
8. Iadecola C. The Neurovascular Unit Coming of Age: A Journey through Neurovascular Coupling in Health and Disease. *Neuron*. 2017;96(1):17-42.
9. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci*. 2014;34(36):11929-47.
10. Berman JR, Kenyon C. Germ-cell loss extends C. elegans life span through regulation of DAF-16 by kri-1 and lipophilic-hormone signaling. *Cell*. 2006;124(5):1055-68.
11. Ito S, Greiss S, Gartner A, Derry WB. Cell-nonautonomous regulation of C. elegans germ cell death by kri-1. *Curr Biol*. 2010;20(4):333-8.

Supplementary Figures and Tables

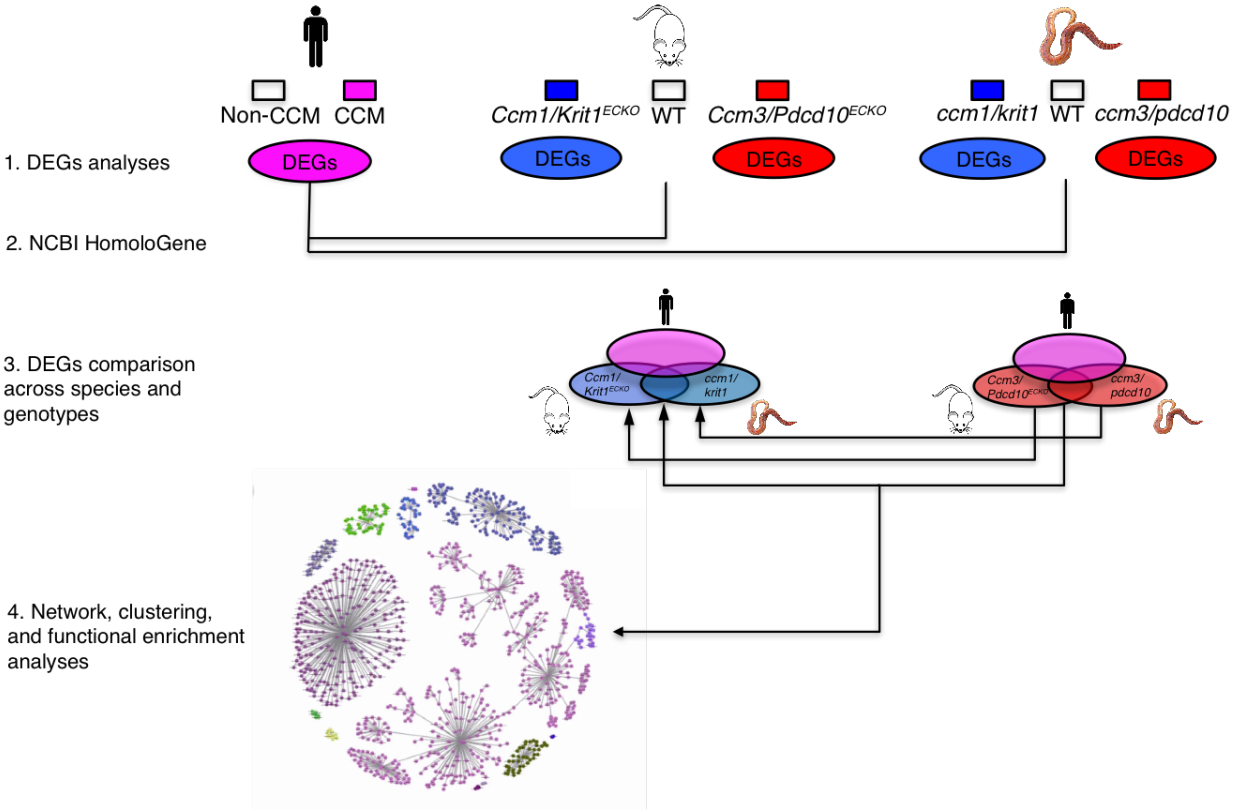

Figure S1. Flowchart of the study. 1. We analyzed differentially expressed genes (DEGs) from human lesional neurovascular units, mouse brain microvascular endothelial cell *Ccm1/Krit1*^{ECKO} and *Ccm3/Pdcd10* ^{ECKO}, as well as *C. elegans ccm1/kri-1* and *ccm3/pdcd10* compared with the appropriate controls for each model. 2. Genes from the animal models were annotated based on the National Center for Biotechnology Information (NCBI) HomoloGene database to allow direct comparison with human genes. 3. We compared DEGs across species and genotypes. 4. We finally performed gene set ontology enrichment, clustering, and network analyses.

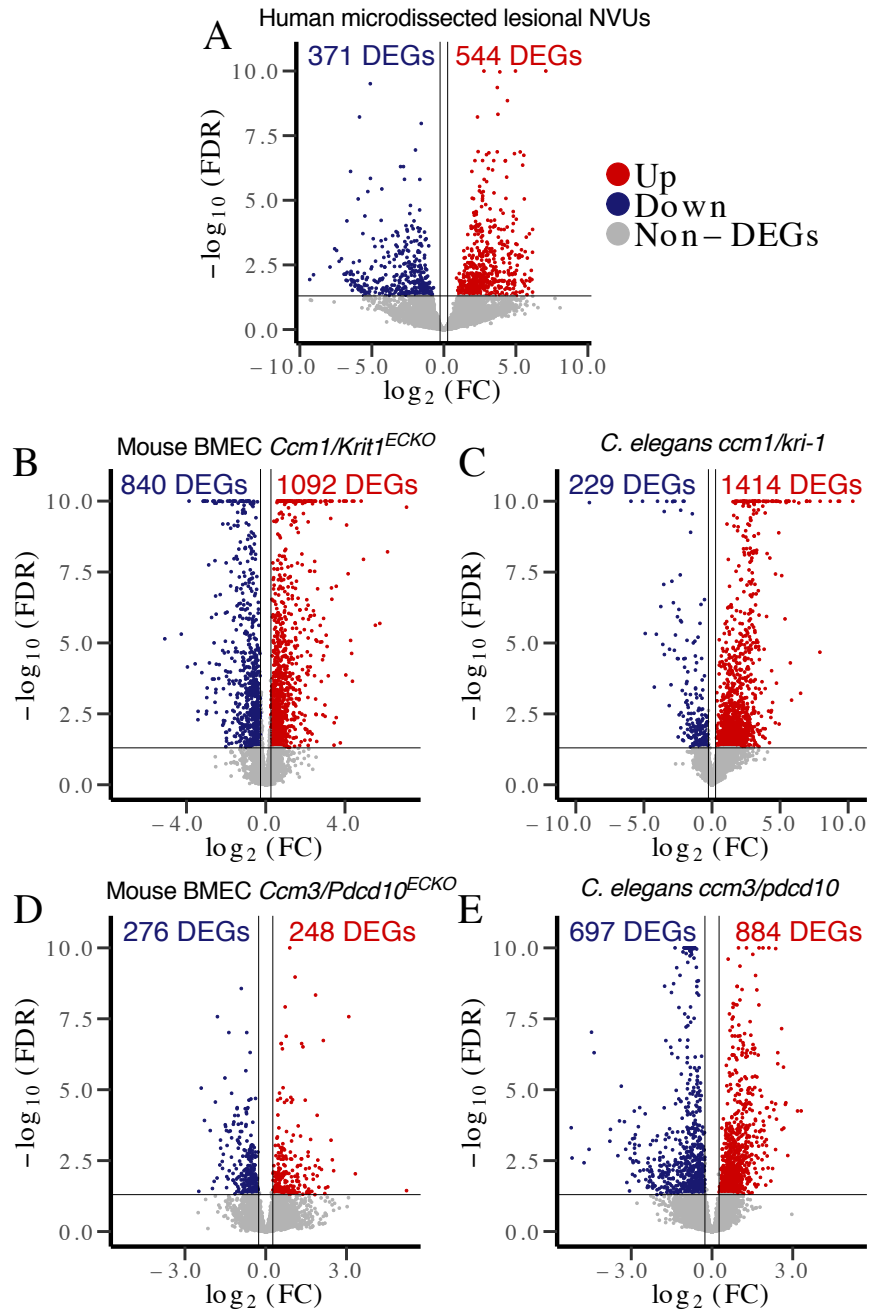

Figure S2. Volcano plots of differentially expressed genes (DEGs) in human lesional neurovascular units (NVUs), mouse brain microvascular endothelial cell (BMEC) *Ccm1/Krit1* and *Ccm3/Pdcd10*, as well as *C. elegans ccm1/kri-1* and *ccm3/pdcd10* models. The transcriptome analyses identified (**A**) 915 DEGs in human lesional NVUs, (**B**) 1932 DEGs in BMEC *Ccm1/Krit1*^{ECKO}, (**C**) 1643 DEGs in *C. elegans ccm1/kri-1* (**D**) 524 DEGs in BMEC *Ccm3/Pdcd10*^{ECKO}, and (**E**) 1581 DEGs in *C. elegans ccm3/pdcd10*. All DEGs were considered at a fold change (FC)≥1.2; significance level p<0.05, false discovery rate (FDR) corrected. Each dot represents an individual DEG with respect to the control. Red dots represent up-regulated DEGs and blue dots are down-regulated DEGs. Results are presented as x-axis $\log_2$ FC and y-axis $\log_{10}$ FDR corrected p-value.

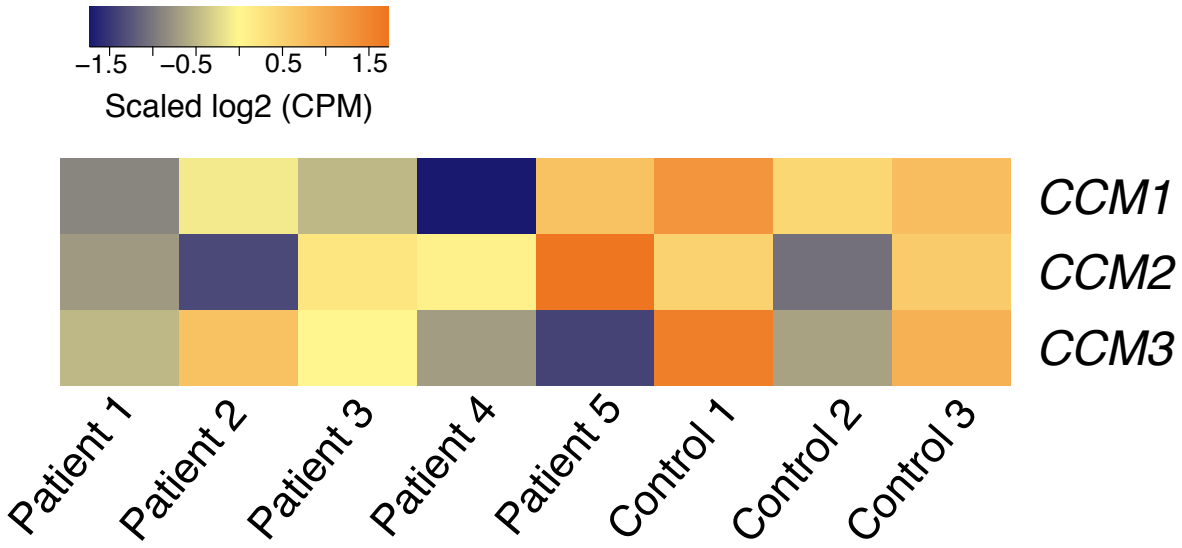

Figure S3. Individual gene expression profiles for *CCM1, CCM2 and CCM3* of each patients with cerebral cavernous malformation (CCM) and autopsy controls. The results suggest that at least one of the CCM genes was down-regulated in the human microdissected lesional neurovascular units (NVUs). Particularly, patient 4 has been clinically genotyped and identified with a familial *CCM1* mutation. Her gene expression profile supports this genotype by showing a down-regulation of CCM1 (darkest blue box). CPM = counts per million.

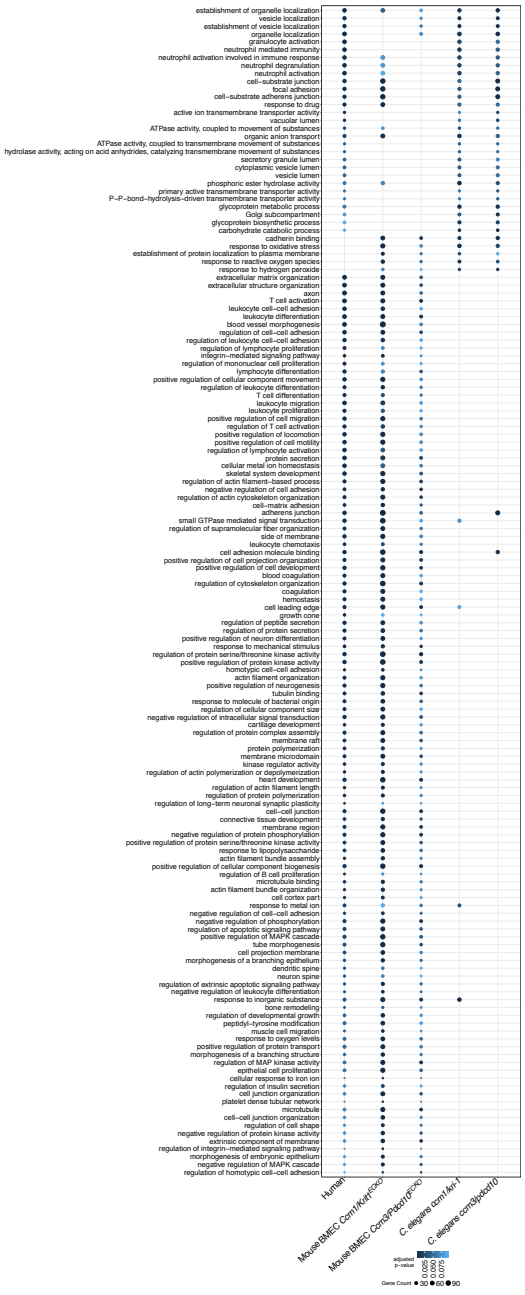

Figure S4. Common gene ontology (GO) terms across species. One GO term (*GO:0051656 establishment of organelle localization*) was common across the 5 models studied. In addition, 24 enriched GO terms were present in 4 of the 5 models; and 149 GOs were commonly identified in at least 3 ($p<0.1$, false discovery rate corrected for all comparisons). These identified GO terms were related to cell-cell adhesion, vesicle transportation, cell organelle localization and movement, immunity and inflammation, ion transmembrane transporter activity, MAPK signaling and responses to oxidative stress. A GO term was defined as common between species if it was identified in both genotypes of the same species in order to exclude GOs that may be specific for either genotype.

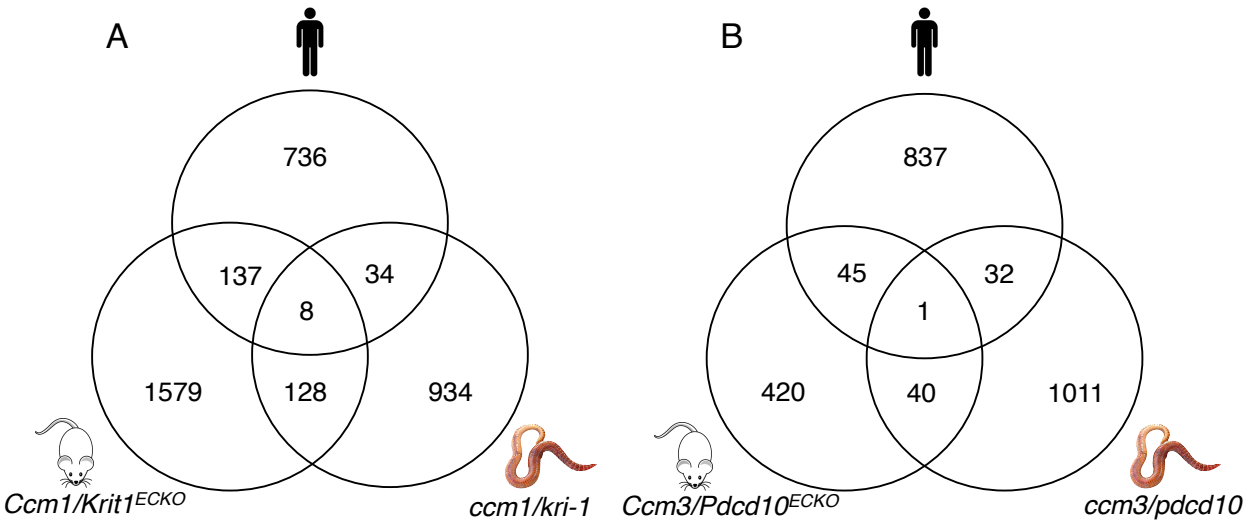

Figure S5. Venn diagrams of human homologous differentially expressed genes (DEGs) in *Ccm1/Krit1* and *Ccm3/Pdcd10* genotype models. A total number of 915 DEGs were identified in human lesional neurovascular units, 1852 and 506 in respectively brain microvascular endothelial cell (BMEC) *Ccm1/Krit* ^{ECKO} and *Ccm3/Pdcd10* ^{ECKO}, as well as 1104 and 1084 in *C. elegans ccm1/kri-1* and *ccm3/pdcd10* respectively (for human and mouse BMEC FC≥1.2; $p<0.05$ false discovery rate (FDR) corrected, for *C. elegans* FC>1.0; $p<0.25$ FDR corrected). (**A**) Eight DEGs were common across species in *Ccm1/Krit1* models: *SPARCL1*, *PDGRFA*, *FAXC*, *UNC13A*, *FAT1*, *PLXDC2*, *PLCD3,* and *GNAO1*. (**B**) One DEG was common across species in *Ccm3/Pdcd10* models: *FAT1*.

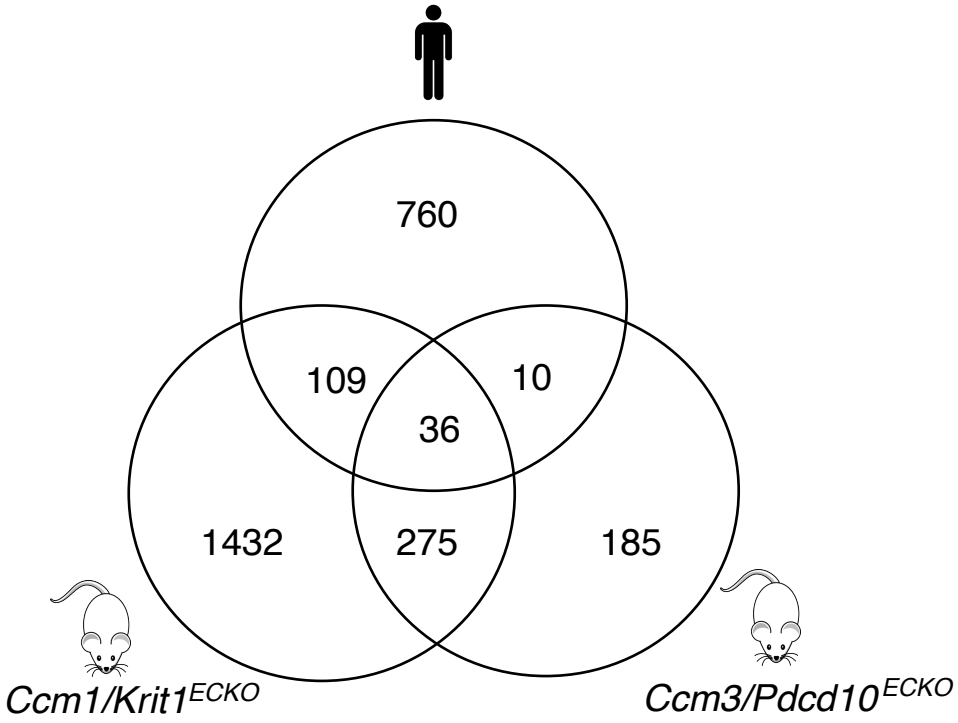

Figure S6. Venn diagram of human homologous differentially expressed genes (DEGs) between human micro-dissected lesional neurovascular units (NVUs) and mouse brain microvascular endothelial cell (BMEC) *Ccm1/Krit1*^{ECKO} and *Ccm3/Pdcd10*^{ECKO} models. The transcriptomes analyses identified 36 common DEGs between the human microdissected lesional neurovascular units (NVUs), mouse BMECs *Ccm1/Krit1*^{ECKO} and *Ccm3/Pdcd10*^{ECKO} models. In addition, further analyses showed that 109 and 10 DEGs were common between human NVUs and respectively *Ccm1/Krit1*^{ECKO} and *Ccm3/Pdcd10*^{ECKO} (FC≥1.2, $p<0.05$; FDR corrected).

Figure. S7. Laser capture microdissection of a cerebral cavernous malformation *in vivo (video)*.

Table S1: Characteristics of CCM subjects used in neurovascular unit RNAseq.

Patient	Age	Sex	Sporadic/Familial	Indication of resection
1	13	Female	Sporadic	Hemorrhagic expansion
2	31	Male	Sporadic	Seizures
3	34	Female	Sporadic	Size of the lesion
4	4	Female	Familial (*CCM1*)	Hemorrhagic expansion
5	40	Female	Sporadic	Hemorrhagic expansion/Seizures

Table S2. List of identified differentially expressed genes (DEGs) in all models. (For human and mouse BMEC FC≥1.2; p<0.05 false discovery rate corrected (FDR), for *C. elegans* FC>1.0; p<0.25 FDR corrected).

Table S3. List of identified GO terms in all models (p<0.10, false discovery rate corrected).

Table S4. List of 95 enriched gene ontology terms using the 16 genes with more than 20 edges in the analysis (p<0.01, false discovery rate corrected).

Table S5. Identified differentially expressed genes between human microdissected lesional neurovascular units and mouse BMEC *Ccm1/Krit1*^{ECKO} and *Ccm3/Pdcd10*^{ECKO} models.

Table S6. List of 71 differently expressed genes (DEGs) common across mouse brain microvascular endothelial cells (BMECs) and *C. elegans* that were identified only in the *Ccm1/Krit1* mutant models regardless of the direction of the fold change. Mouse BMEC DEGs were considered at a fold change (FC)≥1.2; significance level p<0.05, false discovery rate (FDR) corrected. *C. elegans* DEGs were considered at a FC≥1.2; significance level p<0.25, FDR corrected.

Table S7. List of 11 differently expressed genes (DEGs) common across mouse brain microvascular endothelial cells (BMECs) and *C. elegans* that were identified only in the *Ccm3/Pdcd10* mutant models regardless of the direction of the fold change. Mouse BMEC DEGs were considered at a fold change (FC)≥1.2; significance level p<0.05, false discovery rate (FDR) corrected. *C. elegans* DEGs were considered at a FC≥1.2; significance level p<0.25, FDR corrected.

Table S8. List of genes with a documented role in CCM disease for each model. We identified a list of 54 genes with a documented role in CCM disease using DisGeNet (disease ID C2919945, Cavernous Hemangioma of Brain). Among the 54 documented genes, 42 were found in the human lesional neurovascular units transcriptome, 32 human homologues in mouse BMEC *Ccm1/Krit1*^{ECKO}, 32 in mouse BMEC *Ccm3/Pdcd10*^{ECKO}, 6 in *C. elegans ccm1/kri1* and 6 in *C. elegans ccm3/pdcd10*. *C. elegans ccm1/kri1* mRNA levels are upregulated but the transcripts have been shown to be non-functional(10, 11).

Table S9. Full list of enriched gene ontology terms present only within the *Ccm1/Krit1* genotype models (p<0.1, false discovery rate (FDR) corrected).

Table S10. Full list of enriched gene ontology terms present only within the *Ccm3/Pdcd10* genotype models (p<0.1, false discovery rate (FDR) corrected).