Supplementary Appendix 1

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Nikolaev SI, Vetiska S, Bonilla X, et al. Somatic activating *KRAS* mutations in arteriovenous malformations of the brain. N Engl J Med 2018;378:250-61. DOI: 10.1056/NEJMoa1709449

Supplementary Appendix:

Somatic Activating *KRAS* Variants in Brain Arteriovenous Malformations

List of investigators:

Sergey I Nikolaev, Ph.D [1,2,†,*], Sandra Vetiska, Ph.D [3,†], Ximena Bonilla, M.D., Ph.D [1], Emilie Boudreau, M.Sc [4,5,6], Suvi Jauhiainen, Ph.D. [7,8], Behnam Rezai Jahromi, M.B. [7,8], Nadiya Khyzha, B.S. [4,5,6], Peter V. DiStefano, Ph.D. [4,5,6], Santeri Suutarinen, M.B. [7,8], Tim-Rasmus Kiehl, M.D. [5,9], Vitor Mendes Pereira, M.D. [10,11], Alexander M. Herman, Ph.D. [12,13], Timo Krings, M.D. [10, 11], Hugo Andrade-Barazarte, M.D., Ph.D [10], Takyee Tung, B.Sc. [10], Taufik Valiante, M.D., Ph.D. [10], Gelareh Zadeh, M.D., Ph.D. [10], Mike Tymianski, M.D., Ph.D. [3,10], Tuomas Rauramaa, M.D., Ph.D [8,14], Seppo Ylä-Herttuala, M.D., Ph.D. [7], Joshua D. Wythe, Ph.D. [12,13], Stylianos E. Antonarakis, M.D. [1,2,15], Juhana Frösen, M.D., Ph.D [8,*], Jason E. Fish, Ph.D.[4,5,6,§,*], Ivan Radovanovic, M.D. , Ph.D [3,10,§,*].

[1] Department of Genetic Medicine and Development, University of Geneva Medical School, Switzerland
[2] Service of Genetic Medicine, University Hospitals of Geneva, Switzerland
[3] Department of Fundamental Neurobiology, Krembil Research Institute, University Health Network, Toronto, Canada
[4] Toronto General Hospital Research Institute, University Health Network, Toronto, Canada
[5] Department of Laboratory Medicine & Pathology, University of Toronto, Canada
[6] Heart & Stroke Richard Lewar Centre of Excellence in Cardiovascular Research, Toronto, Canada
[7] Department of Molecular Medicine, AIV-Institute, University of Eastern Finland
[8] Hemorrhagic Brain Pathology Research Group, Department of Neurosurgery/NeuroCenter, Kuopio University Hospital and University of Eastern Finland
[9] Department of Pathology, University Health Network, Toronto, Canada
[10] Division of Neurosurgery, Department of Surgery, University Health Network, Toronto, Canada
[11] Joint Division of Medical Imaging, Department of Medical Imaging, Toronto Western Hospital, University Health Network, Toronto, Ontario, Canada
[12] Cardiovascular Research Institute, Baylor College of Medicine, Houston, Texas, USA
[13] Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, Texas, USA
[14] Department of Pathology, Kuopio University Hospital and University of Eastern Finland
[15] iGE3, Institute of Genetics and Genomics of Geneva, Switzerland

Table of Contents

Supplemental Methods

Clinical diagnosis and treatment of BAVMs

The clinical diagnosis of brain arteriovenous malformation for all patients included in this study was initially performed as standard clinical practice and was part of clinical care documented in the patient chart. Diagnosis was made at clinical presentation and before treatment by a multidisciplinary group consisting of surgeons, interventional neuroradiologists and neurosurgeons practicing or familiar with radiosurgery of AVMs. All patients were discussed formally in a multidisciplinary AVM conference and a consensus treatment decision was made (i.e. surgical resection of the BAVM). For the Toronto cohort, treatment (surgical resection) was done by IR and MT and based on the above mentioned diagnostic process after written informed consent for the treatment and for tissue storage and use for research purposes was obtained. After surgical resection, a staff neuropathologist confirmed the diagnosis as per standard clinical care protocol with a standard pathology report included as part of the patient chart. For the purpose of this study specifically, the diagnosis was re-reviewed and confirmed by IR and HAB (surgeons), VMP and TK (interventional neuroradiologists) and TRK (pathologist). Clinical data was collected by IR, HAB and VMP. For the Kuopio cohort, AVM diagnosis was confirmed by the treating radiologist, operating surgeon and neuropathologist. JFr reviewed the reports of all three and personally assessed the histopathology and the available imaging studies.

Sample preparation and exome sequencing:

Frozen BAVM samples and paired blood (leukocyte-enriched fraction) were obtained from the BAVM tissue bank at the Division of Neurosurgery at the University Health Network (UHN) in Toronto. All patients signed an informed consent agreement and the study was approved by the Research Ethics Board of UHN. BAVM samples were ground in liquid nitrogen using a mortar and pestle. Some of this ground tissue was used for DNA extraction and, when sufficient tissue was obtained, some was frozen for further experimental studies. Commercially available kits (e.g. QIAGEN Gentra Puregene) with established protocols were used to isolate genomic DNA for whole exome sequencing and downstream analyses.

Exome sequencing was performed by the Princess Margaret Genomics Centre, Toronto,

Canada, using a SureSelect V5 kit (Agilent). For library preparation, the Agilent SureSelect XT target enrichment kit was used with 200 ng of DNA from BAVMs or blood samples. Libraries were sequenced on an Illumina HiSeq2000 instrument with 100 bp paired-end sequencing, with a median coverage of 190x ± 43 for BAVMs and 80x ± 14 for blood samples. Average Illumina error rate for sequenced samples was 0.38 +/- 0.07% for read 1 and 0.4 +/- 0.09% for read 2, with a cluster density of ~930 K/mm^2. The mean of >= Q30 bases was 91.83%. Sequencing reads were mapped to the reference genome with the Burrows-Wheeler Aligner (BWA) [1] and then processed with the Genome Analysis Toolkit (GATK version 3.3.0) [2] following best practices for exome sequencing. Sequencing quality and target enrichment were verified with Picard tools metrics (http://broadinstitute.github.io/picard/). For the BAVMs with a matching germline sample, MuTect version 1.1.4 [3] with standard parameters was used to identify somatic single nucleotide variants (SNVs) and the GATK's HaplotypeCaller was used to call somatic insertions or deletions (e.g. "indels"). For the 9 samples for which a matched germline control sample was not available, a Panel of Normals (PON) was constructed from 50 germ-line samples. Functional variants including stop gain, splice site, or damaging (as predicted by PolyPhen2 [4]) and non-synonymous mutations were considered as putative drivers. Additionally, variants were considered putative drivers if they were reported in the COSMIC v76 database and were absent in ExAC Version 0.3 (or present with less than 0.00001 MAF). Due to high tissue heterogeneity of the BAVM tissue samples, the fraction of the cells with the causative mutation could be small, therefore the Variant Allele Frequency (VAF) of putative somatic variants is expected to be low. In order to detect candidate variants with low VAF we applied a multistep approach. Primarily, we analyzed 17 BAVMs for which a matching blood sample was available and selected somatic variants with the "KEEP" flag in the repeat masked target regions using MuTect2. Additionally, we screened for the candidate variants with the "REJECT" flag, which were rejected due to their VAF being <3% (default setting of MuTect2). Furthermore, we performed another high sensitivity and low specificity analysis on *KRAS* using a homemade algorithm, "VariantChecker", that screens all sites known to be recurrently mutated in cancer and reports VAF of non-reference nucleotides. This analysis reported all *KRAS* missense variants with a VAF >0.5% which were present on forward and reverse strand reads, or at least present in 5 reads on one strand (similar to [5]).

MuTect2 analysis of 17 BAVM–blood pairs rendered 22,351 variants. Of these variants, only 475 passed the internal MuTect2 quality filter (judgement==TRUE). All frequent variants, present in ExAC 0.3.1 database with a frequency more than 0.00002, were further filtered out, which reduced the number of candidate somatic functional variants to 164. Lastly, we applied stringent filtering criteria based on the coverage, VAFs, variant reads per strand, and quality of the predictions (Performed with the homemade tools (Variant Checker); Table S1):

1/ number of alternative reads in BAVM more than 6: "t_alt_count > 6";

2/ fraction of alternative reads in BAVM more than 2%: "tumor_f > 0.02";

3/ alternative reads on the forward strand should be present: "i_t_ALT_F1R2 > 0";

4/ alternative reads on the reverse strand should be present: "i_t_ALT_F2R1 > 0";

5/ alternative reads in the control sample should be absent: "n_alt_count == 0";

6/ the coverage in the control sample should be more than 60 fold: "n_ref_count + n_alt_count > 60";

7/ the coverage in the BAVM sample should be more than 150 fold "t_ref_count + t_alt_count >= 150";

8/ The tumor LOD score should be more than 10: "t_lod_fstar__INFO__ >10"

Application of such criteria resulted in a set of 9 SNVs which includes 6 missense variants; 4 *KRAS* variants at position chr12:25398284 and 2 in other genes (*TP53BP1* p.Leu55Phe, *PCSK5* p.Ala1783Val)

Digital Droplet PCR (ddPCR) - Toronto cohort:

Detection of rare variants in *KRAS* was performed on the QX200 Droplet Digital PCR system (Bio-Rad Laboratories, Inc., Hercules, CA, USA) at The Centre for Applied Genomics (TCAG) at The Hospital for Sick Children (Toronto, Canada) using Taqman hydrolysis probe chemistry (Life Technologies, Carlsbad, CA, USA). A 40X custom genotyping primer and probe mix was designed for each variant. The 20 μL reaction mix consisted of 10 μL of 2x ddPCR SuperMix for Probes (Bio-Rad Laboratories), 0.5 μL of the 40X SNP genotyping assay, 8.5 μL water and 1 μL of 50 ng/μL genomic DNA. Cycling conditions for the reaction were 95°C for 10 min, followed by 45 cycles of 94°C for 30 sec and 60°C for 1 min, then 98°C for 10 minutes, and finally a 10°C hold on a Life Technologies Veriti thermal cycler. All assays were validated by temperature gradient to ensure optimal separation of reference and variant signals. Data was analyzed using QuantaSoft v1.4 (Bio-Rad Laboratories). Positive control DNAs for each assay included NCI-

H441 (ATCC® HTB-174™) for *KRAS* c.35G>T;p.(Gly12Val) and synthetic construct gBlocks Gene Fragment (IDT, Coralville, Iowa, USA) for *KRAS* c.38G>T;p.(Gly13Val), *KRAS* c.436G>A;p.(Ala146Thr), *KRAS* c.183A>T;p.(Gln61His) and *HRAS* c.35G>A;p.(Gly12Asp). A positive control was not available for *KRAS* c.35G>A;p.(Gly12Asp). In addition, human genomic DNA (HuRef) plus one non-template control were genotyped with the study samples. Primers for ddPCR are shown in Table S4. We set a minimum of 0.5% fractional abundance to call a sample positive (similar to [6]).

DNA isolation and *KRAS* ddPCR from paraffin embedded and frozen samples in the Kuopio University Hospital replication cohort:

The replication cohort from Kuopio University Hospital (KUH) consisted of paraffin embedded BAVM samples (n=54), frozen BAVM samples (n=4), as well as paraffin embedded Circle of Willis samples (n=54) and cavernoma samples (n=8) that were used as controls. Thin (10-30 μm) sections were cut from paraffin embedded tissue samples using a microtome. The sections were deparaffinized with xylene for 3x15 min in 65^{o}C, rehydrated with a graded ethanol series and excessive ethanol removed by incubation at 40^{o}C. The tissue sections were then resuspended in 650 μL of DNA extraction buffer [5 mM EDTA, 250 mM NaCl, 1% SDS, 50 mM Tris, pH 7.6, and 0.2 mg/mL Proteinase K (Thermo Scientific, Waltham, MA)] and homogenized using Precellys Homogenizer (Bertin Instruments, Montigny-le-Bretonneux, France) (5500 rpm 2x30 sec) and soft tissue homogenizing CK14 tubes containing 1.4 mm ceramic (zirconium oxide) beads (Bertin Instruments). After incubation for 18 h at 50^{o}C under continuous agitation, the DNA was extracted with phenol:chloroform:isoamyl alcohol 25:24:1 (saturated with 10 mM Tris, pH 8.0; Amresco-inc, Solon, OH), followed by further purification with chloroform and then isopropanol precipitation. After two washes with 70% ethanol, the DNA was dissolved in nuclease-free H_2O and the DNA concentration measured using a NanoDrop spectrophotometer (NanoDrop products, Wilmington, DE). DNA isolation from KUH-derived frozen samples was performed similarly as described above for paraffin embedded samples, with the exception that no xylene or ethanol rehydration was needed and the samples were sectioned using a cryotome.

ddPCR for *KRAS* c.35G>A;p.(Gly12Asp) and c.35G>T;p.(Gly12Val) variants was performed from the DNA extracted from KUH-derived paraffin embedded and frozen samples as described previously for the original discovery cohort, with DNA concentrations of 50 or 250 ng of genomic DNA isolated from frozen samples or 250, 500, or 1000 ng of genomic DNA when

using DNA isolated from paraffin embedded samples. All the ddPCRs for c.35G>A;p.(Gly12Asp) and c.35G>T;p.(Gly12Val) were replicated at least two times, with the whole protocol repeated starting from the isolation of the genomic DNA. For most of the samples the c.35G>A;p.(Gly12Asp) ddPCRs were performed at least three times (Table S7). The ddPCR results of samples that did not have ≥ 500 droplets for *KRAS* wild-type in at least two replicates were considered unreliable and were excluded from further analysis (21/54 AVM samples). For larger AVM samples divided in several blocks, the DNA isolation and ddPCR was performed separately for each of the tissue blocks.

Since formalin-fixation can introduce false positives in *KRAS* c.35G>A;p.(Gly12Asp) ddPCR (as discovered also in our control samples from normal Circle of Willis and cavernomas), the c.35G>A;p.(Gly12Asp) ddPCR of all those samples that showed fractional abundance for c.35G>A;p.(Gly12Asp) above background level, were repeated using DNA that underwent uracil DNA glycosylase pretreatment, as described previously [7,8]. In brief, genomic DNA was incubated for 3 h at 37^{o}C with 2.5 μL (12.5 U) of uracil DNA glycosylase (New England BioLabs, Ipswich, MA) and 1 μL of 10x Reaction Buffer in a final volume of 10 μL. This pretreatment significantly reduced the background level (Table S7, Figure S3).

Cell culture of AVM and normal brain vascular endothelial cells:
Briefly, tissue was washed with PBS, sectioned into ~3 mm^2 fragments, and incubated with 0.1% collagenase (Sigma) at 37°C, for 20 min. Pre-digested tissue was triturated with a 2 mL pipette and filtered through a 100 μm cell strainer (BD Biosciences). The cell suspension was centrifuged at 2000 RPM for 5 min, and cells were washed and resuspended in EBM-2 media (Lonza), supplemented with EGM-2 MV SingleQuots (5% FBS, hydrocortisone, hFGF, VEGF, IGF-1, ascorbic acid, hEGF, GA-1000; Lonza).

ECs were isolated using anti-CD31 Dynabeads® (Life Technologies) according to the manufacturer's recommendations. CD31-positive and CD31-negative cell fractions were cultured in endothelial (EGM-2, Lonza) and smooth muscle (SmGM-2, Lonza) media, respectively, and used for further analyses. For immunocytochemistry, cells were fixed using 4% paraformaldehyde, permeabilized using 0.5% TX-100, and blocked using 5% normal goat serum (NGS). The rabbit polyclonal CD31 antibody (dilution, 1:20, cat. No. ab28364; Abcam, Cambridge, MA, USA) and the mouse monoclonal alpha SMA antibody (dilution, 1:400, cat. No. ab7817; Abcam) were added to the cells for 1.5 h at room temperature. Cells were washed 3

times, followed by 1 h incubation using goat anti-rabbit IgG, Alexa Fluor® 568 secondary antibody (dilution 1:500; cat. No. A11011; Thermofisher) or goat anti-mouse IgG, AlexaFluor® 488 (dilution 1:400; cat. No. A11029; Thermofisher) at room temperature. After further washing, coverslips were mounted in mounting media containing DAPI. Images were taken on a Zeiss AxioImager.M2 microscope, using a Hamamatsu digital camera.

Cell culture and treatment of human umbilical vein endothelial cells (HUVECs):
Primary endothelial cells were purchased from ScienCell and cultured according to manufacturer's instructions. Inhibitors of MEK (U0126, 20 μM, InvivoGen) or PI3K (LY294002, 10 μM, Cell Signaling) were added upon serum-starvation [0.1% FBS in basal media (ScienCell) with no additional growth factors] for 6 h. All drugs were dissolved in DMSO, and comparison was made to vehicle-treated (i.e. DMSO, 0.1%) controls. For experiments assessing the ability of MEK inhibition to rescue the adherens junction phenotype, confluent cells were treated with 20 μM U0126 for 18 h (in serum-starvation media).

Cell culture of telomerize-immortalized human aortic endothelial cells (TeloHAEC):
Cells were purchased from ATCC and grown in complete endothelial cell medium (Promocell).

Plasmids:
The following plasmids from Daniel Haber's lab were purchased from Addgene: #35634, pLenti-PGK-KRAS4A(G12V) and #35633, pLenti-PGK-KRAS4B(G12V). The open reading frames were amplified using primers that included an added BamHI site and a Kozak sequence on the 5' end and a XhoI site on the 3' end. PCR amplicons were then subcloned into $pCS2^{+}$ using the BamHI/XhoI sites. The LifeAct-GFP construct was from Ibidi. To assess electroporation efficiency (typically ~70%), pmaxGFP (Lonza) was co-electroporated in some experiments.

Electroporation:
~0.5 x 10^6 HUVECs were electroporated with 2.5 μg of pCS2, pCS2-$KRAS4A^{G12V}$ or pCS2-$KRAS4B^{G12V}$ constructs, together with 0.2 μg of pmaxGFP or 0.5 μg LifeAct-GFP using the P5 Primary Cell Kit (Lonza) and a 4D-nucleofector (Lonza). Cells were plated on one well of a 6-well dish or two wells of a 12-well dish, which were pre-coated with attachment factor (Thermo Scientific). Media was changed after 1 h.

Transfection of TeloHAEC:

Cells were transfected in OptiMEM medium at 60% confluency in 6-well plates using 2.5 μg of pCS2 or pCS2-KRAS4A^{G12V}, together with 0.2 μg of LifeAct-GFP and 10 μL of Lipofectamine 2000 (Invitrogen), according to the manufacturers' recommendations. After 4 h, transfection complexes were removed and replaced with complete medium (without antibiotics). After 24 h, cells were plated in 2-well LabTek coverglass plates (ThermoFisher, 155380) coated with attachment factor and incubated for 24 h prior to live imaging. 3 h prior to live imaging, cells were serum-starved in 0.1% FBS-containing medium without growth factors. Images were acquired on a Yokogawa Spinning Disk Confocal with a UPlanSApo 20x/0.75 NA objective. Cells were imaged every 3 minutes (10 μm stacks) for 3 hours.

Gene expression and cell biology experiments:

For gene expression studies, electroporated HUVECs were serum-starved the following day for 6 h, followed by RNA and protein isolation. In some experiments, small molecule inhibitors of signaling pathways were added. For cell biology experiments, cells were plated onto Permanox chamber slides the day after electroporation. Cells were serum-starved overnight prior to imaging. To assess cell migration, HUVEC were plated in 2-well culture inserts (Ibidi, Catalog #80209) on attachment factor-coated cover glass slides in complete media. Inserts were removed after 18 h and serum-starvation medium was added during 'scratch' repair. Cells were either fixed for static confocal imaging or were imaged using time-lapse confocal microscopy. To calculate cell migration rate, individual cells were tracked from time-lapse videos (5 h duration, 12 images/hour) taken during migration into the cell-free area of a 'scratch' assay. The distance between the starting and ending position of the cell was determined using the Manual Cell Tracking feature of Volocity software, and migration rate was determined by dividing this distance by the duration of migration. To assess cell proliferation, cells were serum-starved for 6 h, followed by addition of 10 μM BrdU (Sigma, Catalog #B5002) for 1 h at 37^{o}C. Cells were washed with PBS, fixed in 4% paraformaldehyde and permeabilized with 0.5% Triton X-100 in 10% DMSO, and quenched with 3% H_2O_2. Cells were then treated with DNase I (1:50 dilution, 4u/μL final concentration, 100 μL total volume, Invitrogen) at 37^{o}C for 1.5 h. Anti-BrdU-biotin (1:50 in blocking buffer, Biolegend, Catalog # 339810) was added overnight at 4^{o}C, followed by streptavidin-HRP (1:200 in blocking buffer, Perkin Elmer, Catalog #NEL750001EA) at room

temperature for 30 min. Tyramide-FITC (1:50, Perkin Elmer, Catalog #SAT701001EA) was added for 7 min at room temperature. After washing, slides were mounted in Vectashield containing DAPI (Vectorlabs, Catalog #H-100). The number of BrdU-positive nuclei was quantified using ImageJ in multiple random fields (40x objective, NA 0.8) and expressed as a percentage of total cell number. To assess cell survival, cells were cultured in complete or serum-stave medium for 24 h, followed by fixation in 4% paraformaldehyde and permeabilization in 0.5% Triton X-100 in 10% DMSO. TUNEL staining was performed using the In Situ Cell Death Detection kit (Roche, Catalog #11684795910), according to the manufacturer's recommendations. Slides were mounted in Vectashield containing DAPI, and the percentage of TUNEL-positive cells was determined as above.

Confocal microscopy:
Static imaging was performed using an Olympus FV1000 Confocal microscope using UplanSApo 20X/0.75NA, LumPlanFI40X/0.8NA, or Oil PlanApo 60X/1.42NA objectives. Time-lapse microscopy was performed using the WaveFX Yokogawa Spinning Disk Confocal with integrated environmental chamber, using UPlanSApo 20X/0.75NA or UPlanFL 10X/0.3NA objectives at a rate of 12 or 15 images per hour.

RNA isolation, and qRT-PCR:
RNA was isolated using Trizol.
Quantitative reverse-transcription PCR (qRT-PCR) - Following DNase I treatment (Invitrogen), reverse transcription was performed on 1 μg of RNA using the High Capacity cDNA Synthesis Kit (Thermo Scientific). qRT-PCR was performed using a Roche Lightcycler 480® with LC 480 SYBR Green I Master Mix (Roche). Data were normalized to Tata Box Binding Protein (TBP) using the Δ - Δ Ct method. Primer sequences are listed in Table S10.

Immunofluorescence:
HUVEC were grown on attachment factor-coated Permanox or glass chamber slides (Nunc-LabTek). Cells were fixed with 4% paraformaldehyde followed by permeabilization with 0.25% Triton X-100. Staining with anti-pERK (rabbit polyclonal, Cell Signaling, #9101, 1:500) or anti-VE-Cadherin (R&D Systems, #MAB9381) was performed overnight at 4°C, followed by secondary antibody addition (anti-rabbit-AlexaFluor647, Thermo Scientific, #A21235). Slides

were mounted using Vectashield mounting medium with DAPI (Vector Labs, #H-1200) and imaged using an Olympus FV1000 Confocal microscope.

SDS-PAGE and Immunoblotting:
Protein lysates from cultured cells were homogenized in lysis buffer (250 mM NaCl, 20 mM HEPES, 30 mM $MgCl_2$, 0.5 mM EDTA, 0.1 mM EGTA, 20% glycerol, 1% NP-40) supplemented with 0.25 mM DTT, 0.1 mg/mL AEBSF, phosphatase- and protease-inhibitor cocktails, and centrifuged at 14,000 xg, for 10 min at 4°C. Protein concentration was determined using the BioRad protein assay. 40 μg of protein was separated by SDS-PAGE, followed by transfer to PVDF membrane. The following antibodies were used: rabbit anti-AKT (Cell Signaling, Cat. #9272), rabbit anti-phospho-AKT(S473) (SAB, Cat. #11054), rabbit anti-ERK1/2 (Cell Signaling, Cat. #9102), rabbit anti-phospho-ERK(T202/T204) (Cell Signaling, Cat. #9101), rabbit anti-p38 (Cell Signaling, Cat. #9212), rabbit anti-phospho-p38(T180/T182) (Cell Signaling, Cat. #9211), rabbit anti-B-actin (Cell Signaling, Cat. #5125).

Immunohistochemistry:

4 μm FFPE sections of 25 BAVMs included in this study (with and without *KRAS* mutations detected by ddPCR) were stained with phospho-ERK. Paraffin-embedded sections were dewaxed in 5 changes of xylene and rehydrated to water through graded alcohol washes. Heat induced epitope retrieval in 10 mM citrate buffer (pH 6.0) was used and endogenous peroxidase blocked with 3% hydrogen peroxide. The antibody to phosho-ERK (Cell Signaling; Cat. #9101) was used at a dilution of 1:800 overnight. The detection system used was species-specific ImmPRESS polymer system Anti-Rabbit IgG Cat# MP-7401 (Vector Labs). After following kit instructions, color development was performed with freshly prepared DAB (DAKO Cat# K3468). Finally, sections were counterstained lightly with Mayer's Hematoxylin, dehydrated in alcohols, cleared in xylene and mounted with Permount mounting medium (Fisher, cat# SP15-500).

RNA sequencing:
RNA sequencing was done at the Princess Margaret Genomics Center, Toronto, Canada. Total RNA was extracted from HUVEC electroporated with pCS2, pCS2-KRAS4A^{G12V} and pCS2-KRAS4B^{G12}. 200 ng of RNA was used to synthesize cDNA. RNA samples were quantified by Qubit (Life Technologies) and by Bioanalyzer (Agilent). All samples had RIN above 8. Libraries

were prepared using TruSeq Stranded Total RNA kit (Illumina). 200 ng from RNA samples were ribosomal RNA depleted using Ribo-zero Gold rRNA beads, and following purification the RNA was fragmented. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers. This was followed by second strand cDNA synthesis using RNase H and DNA Polymerase I. A single "A" base was added and adapter ligated followed by purification and enrichment with PCR to create cDNA libraries. Final cDNA libraries were size validated by Bioanalyzer and concentration was validated by qPCR. All uniquely bar-coded cDNA libraries were normalized to 10 nM and pooled together. 10 pM of pooled libraries were loaded onto an Illumina cBot for cluster generation. The clustered flow cell was then sequenced (paired-end, 100 cycles V3) using an Illumina HighSeq 2000 to achieve a minimum of ~35 million reads per sample. Overall read quality was checked using FASTQC v.0.11.2 and for RNA QC the tool RNA-SeQC (v1.1.7) was used. Raw sequence data, in the form of FASTQ files, was aligned to the human genome (hg19, iGenome GTF definition file) using the BOWTIE/TOPHAT pipeline (BOWTIE v2.2.6, TOPHAT 2.1.0). Accessory programs for the alignment stage include SAMTOOLS (v1.2) and Flexbar (2.5.0). Transcript assembly, abundance estimation, and tests for differential regulation were done using CUFFLINKS (v2.2.1)

GSEA analysis

Gene Set Enrichment analysis was performed using the GSEA software [9] (corrected p-value was obtained after 10000 permutations of the gene sets). Comparison was made between RNA-sequencing data from HUVEC over-expressing KRAS4A^{G12V} or KRAS4B^{G12V} and data from HUVEC stimulated with VEGF for 1 h (GEO dataset: GSE41166) [10].

References

1. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26:589-95.

2. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.

3. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31:213-9.

4. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013;Chapter 7:Unit7 20.

5. Limaye N, Kangas J, Mendola A, et al. Somatic Activating PIK3CA Mutations Cause Venous Malformation. Am J Hum Genet 2015;97:914-21.

6. Luks VL, Kamitaki N, Vivero MP, et al. Lymphatic and other vascular malformative/overgrowth disorders are caused by somatic mutations in PIK3CA. J Pediatr 2015;166:1048-54 e1-5.

7. Do H, Dobrovic A. Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil- DNA glycosylase. Oncotarget 2012;3:546-58.

8. Serizawa M, Yokota T, Hosokawa A, et al. The efficacy of uracil DNA glycosylase pretreatment in amplicon-based massively parallel sequencing with DNA extracted from archived formalin-fixed paraffin-embedded esophageal cancer tissues. Cancer Genet 2015;208:415-27.

9. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.

10. Zhang B, Day DS, Ho JW, et al. A dynamic H3K27ac signature identifies VEGFA-stimulated endothelial enhancers and requires EP300 activity. Genome Res 2013;23:917-27.

Supplemental Tables

Table S1: Excel spreadsheet showing MuTec2 results for the 6 coding missense variants including 4 variants causing a *KRAS* c.35G>A;p.(Gly12Asp) mutation in 4 patients [Included as a separate Excel file]

Table S2: Summary of *KRAS* variants found with exome sequencing of the Toronto cohort using two algorithms, MuTecT2 and Variant Checker

* variants reported by MuTect2 with a "Rejected" flag if their VAFs were below 3%. Nidus: samples taken from the BAVM nidus. DV: sample taken from the draining vein. ()= number of reads

Patient	Tissue	CONTROL	MUTECT2_DR100	Variant Checker_KRAS
1	Nidus	yes	KRAS_c.35G>A;p.(Gly12Asp)_2.4%_(7)	KRAS_c.35G>A;p.(Gly12Asp)_2%_(6/0)
1	DV	yes	KRAS_c.35G>A;p.(Gly12Asp)_1%_(4)	KRAS_c.35G>A;p.(Gly12Asp)_1.3%_(3/3)
2	Nidus	yes	KRAS_c.35G>T;p.(Gly12Val)_3%_(9)	KRAS_c.35G>T;p.(Gly12Val)_3.4%_(7/2)
3	Nidus	yes		KRAS_c.35G>A;p.(Gly12Asp)_0.9%_(1/2)
4	Nidus	yes	KRAS_c.35G>A;p.(Gly12Asp)_4%_(13)	KRAS_c.35G>A;p.(Gly12Asp)_4.2%_(9/4)
5	Nidus	no	KRAS_c.35G>T;p.(Gly12Val)_3.2%_(8)	KRAS_c.35G>T;p.(Gly12Val)_3.2%_(8/1)
6	Nidus	yes	KRAS_c.35G>A;p.(Gly12Asp)_1.7%_(5)*	KRAS_c.35G>A;p.(Gly12Asp)_1.5%_(4/1)
7	Nidus	yes	KRAS_c.35G>T;p.(Gly12Val)_2%_(4)*	KRAS_c.35G>T;p.(Gly12Val)_1.8%_(3/1)
8	Nidus	yes		
9	Nidus	no		KRAS_c.38G>T;p.(Gly13Val)_0.9%_(1/1)
9	Nidus	no		KRAS_c.35G>T;p.(Gly12Val)_1.3%_(1/2)
10	Nidus	yes	KRAS_c.35G>A;p.(Gly12Asp)_3.3%_(13)	KRAS_c.35G>A;p.(Gly12Asp)_3.1%_(7/6)
11	Nidus	yes		
12	Nidus	no		
13	Nidus	no		
14	Nidus	yes		KRAS_c.35G>A;p.(Gly12Asp)_1.4%_(3/1)
15	Nidus	no		KRAS_c.35G>A;p.(Gly12Asp)_0.9%_(1/2)
16	Nidus	yes		
17	Nidus	yes		
18	Nidus	no		KRAS_c.35G>A;p.(Gly12Asp)_1.3%_(2/1)
19	Nidus	yes		
20	Nidus	no		KRAS_ c.183A>T;p.(Gln61His)_4.1%_(1/1)
21	Nidus	yes		
22	Nidus	yes		
23	Nidus	yes		
24	Nidus	no		KRAS_p.Gly13Cys_0.6%_(1/1)
25	Nidus	yes		
26	Nidus	no		

Table S3: Summary of somatic variants of *KRAS* in brain arteriovenous malformation tissue samples from Toronto and Kuopio.

FFPE: formalin fixed paraffin embedded, NV: normal vessels (NV1, NV2, NV3 were used for cell cultures shown in Figure 3A and S4. For NV1 and NV2, no whole tissue was available for analysis). AVF: arteriovenous fistula, CCM: cerebral cavernous malformation

Toronto Cohort (fresh frozen samples)						**Finnish Cohort (FFPE)**			
Patient	**Sample**	**Whole Exome Sequencing**		**droplet digital PCR**		**Patient**	**Blocs**	**droplet digital PCR**	
		allelic frequency (%)	***KRAS* variant**	**fractional abundance (%)**	***KRAS* variant**		+/total	**fractional abundance (%)**	***KRAS* variant**
1	**nidus**	2	c.35G>A; p.(Gly12Asp)	2.88	c.35G>A; p.(Gly12Asp)	**1**	0/1	-	-
1	**Vein**	1.3	c.35G>A; p.(Gly12Asp)	1.67	c.35G>A; p.(Gly12Asp)	**2**	0/1	-	-
1	**blood**	-	-	-	-	**3**	0/2	-	-
2	**nidus**	3.4	c.35G>T; p.(Gly12Val)	3.17	c.35G>T; p.(Gly12Val)	**4**	1/1	1.4	c.35G>A; p.(Gly12Asp)
2	**blood**	-	-	-	-	**5**	1/1	3.6	c.35G>A; p.(Gly12Asp)
3	**nidus**	0.9	c.35G>T; p.(Gly12Val)	1.43	c.35G>A; p.(Gly12Asp)	**6**	0/6	-	-
3	**blood**	-	-	-	-	**7**	0/4	-	-
4	**nidus**	4.2	c.35G>A; p.(Gly12Asp)	4.37	c.35G>A; p.(Gly12Asp)	**8**	0/3	-	-
4	**blood**	-	-	-	-	**9**	2/2	2.15	c.35G>A; p.(Gly12Asp)
5	**nidus**	3.2	c.35G>T; p.(Gly12Val)	3	c.35G>T; p.(Gly12Val)	**10**	1/1	3.19	c.35G>A; p.(Gly12Asp)
6	**nidus**	1.5	c.35G>A; p.(Gly12Asp)	2.98	c.35G>A; p.(Gly12Asp)	**11**	1/1	1.35	c.35G>T; p.(Gly12Val)
6	**blood**	-	-	-	-	**12**	0/5	-	-
7	**nidus**	1.8	c.35G>T; p.(Gly12Val)	2.46	c.35G>T; p.(Gly12Val)	**13**	1/1	3.45	c.35G>A; p.(Gly12Asp)
7	**blood**	-	-	-	-	**14**	1/2	1.28	c.35G>T; p.(Gly12Val)
8	**nidus**	-	-	2.41	c.183A>T; p.(Gln61His)	**15**	0/1	-	-
8	**blood**	-	-	-	-	**16**	1/2	0.7	c.35G>T; p.(Gly12Val)
9	**nidus**	1.3	c.35G>T; p.(Gly12Val)	1.56	c.35G>T; p.(Gly12Val)	**17**	1/1	1.05	c.35G>A; p.(Gly12Asp)
9	**blood**	-	-	-	-	**18**	1/1	1.31	c.35G>A; p.(Gly12Asp)
10	**nidus**	3.1	c.35G>A; p.(Gly12Asp)	2.4	c.35G>A; p.(Gly12Asp)	**19**	0/1	-	-

11	**nidus**	-	-	0.5	c.35G>T; p.(Gly12Val)	**20**	1/1	1.80	c.35G>T; p.(Gly12Val)
11	**blood**	-	-	-	-	**21**	2/2	2.08	c.35G>A; p.(Gly12Asp)
12	**nidus**	-	-	1.44	c.35G>T; p.(Gly12Val)	**22**	2/4	1.68	c.35G>A; p.(Gly12Asp)
13	**nidus**	-	-	1.16	c.35G>T; p.(Gly12Val)	**23**	0/1	-	-
14	**nidus**	1.4	c.35G>A; p.(Gly12Asp)	0.62	c.35G>A; p.(Gly12Asp)	**24**	0/1	-	-
14	**blood**	-	-	-	-	**25**	0/1	-	-
15	**nidus**	0.9	c.35G>A; p.(Gly12Asp)	1.66	c.35G>A; p.(Gly12Asp)	**26**	0/3	-	-
16	**nidus**	-	-	0.64	c.35G>A; p.(Gly12Asp)	**27**	1/1	3.00	c.35G>A; p.(Gly12Asp)
17	**nidus**	-	-	2.19	c.35G>A; p.(Gly12Asp)	**28**	0/1	-	-
17	**blood**	-	-	-	-	**29**	1/1	1.24	c.35G>A; p.(Gly12Asp)
18	**nidus**	1.3	c.35G>A; p.(Gly12Asp)	0.76	c.35G>A; p.(Gly12Asp)	**30**	0/1	-	-
18	**blood**	-	-	-	-	**31**	0/3	-	-
19	**nidus**	-	-	-	-	**32**	1/1	2.20	c.35G>A; p.(Gly12Asp)
19	**blood**	-	-	-	-	**33**	0/1	-	-
20	**nidus**	n/a	n/a	-	-			**Finnish Cohort**	**(Fresh Frozen)**
21	**nidus**	-	-	-	-	Patient	Blocs	FA	*KRAS* variant
22	**nidus**	-	-	-	-	2	0/1	-	-
22	**blood**	-	-	-	-	3	0/1	-	-
23	**nidus**	-	-	-	-	10	1/1	0.87	c.35G>A; p.G12D
23	**blood**	-	-	-	-				
24	**nidus**	-	-	-	-				
24	**blood**	-	-	-	-				
25	**nidus**	-	-	-	-				
26	**nidus**	-	-	-	-				
26	**blood**	-	-	-	-				
27	**nidus**	-	-	-	-				
28	**nidus**	n/a	n/a	6.72	c.35G>A; p.(Gly12Asp)				
29	**nidus**	n/a	n/a	0.763	c.35G>A; p.(Gly12Asp)				
30	**nidus**	n/a	n/a	1.6	c.35G>T; p.(Gly12Val)				

31	**nidus**	n/a	n/a	0.753	c.35G>A; p.(Gly12Asp)				
32	**nidus**	n/a	n/a	n/a	n/a				
33	**nidus**	n/a	n/a	1.4	c.35G>A; p.(Gly12Asp)				
33	**blood**	n/a	n/a	-	-				
34	**nidus**	n/a	n/a	0.43	c.35G>A; p.(Gly12Asp)				
34	**blood**	n/a	n/a	-	-				
35	**nidus**	n/a	n/a	1.7	c.35G>T; p.(Gly12Val)				
35	**blood**	n/a	n/a	-	-				
36	**nidus**	n/a	n/a	1.09	c.35G>A; p.(Gly12Asp)				
36	**blood**	n/a	n/a	-	-				
37	**nidus**	n/a	n/a	1.34	c.35G>A; p.(Gly12Asp)				
38	**nidus**	n/a	n/a	-	-				
38	**blood**	n/a	n/a	n/a	n/a				
39	**nidus**	n/a	n/a	2.34	c.35G>A; p.(Gly12Asp)				
39	**vein**	n/a	n/a	0.283	c.35G>A; p.(Gly12Asp)				
39	**blood**	n/a	n/a	n/a	n/a				
Normal Vessels -Toronto Cohort (fresh frozen samples)						**Normal Vessels-Finnish Cohort (FFPE)**			
Patient	**Sample**	**Whole Exome Sequencing**		**droplet digital PCR**		**Patient**	**Blocs**	**droplet digital PCR**	
		allelic frequency (%)	***KRAS* variant**	**fractional abundance (%)**	***KRAS* variant**		+/total	**fractional abundance (%)**	***KRAS* variant**
1	NV1	n/a	n/a	n/a	n/a	**All patients (1 to 54)**		-	-
2	NV2	n/a	n/a	n/a	n/a				
3	NV3	n/a	n/a	-	-				
4	NV4	n/a	n/a	-	-				
Non-AVM Disease -Toronto Cohort (fresh frozen samples)						**Non-AVM Disease-Finnish Cohort (FFPE)**			
Patient	**Sample**	**Whole Exome**		**droplet digital PCR**		**Patient**	**Blocs**	**droplet digital PCR**	

		Sequencing							
		allelic frequency (%)	***KRAS* variant**	**fractional abundance (%)**	***KRAS* variant**		+/total	**fractional abundance (%)**	***KRAS* variant**
1	**AVF**	n/a	n/a	-	-	**1 (AVF)**	0/1	-	-
2	**CCM**	n/a	n/a	-	-	**2 (AVF)**	0/3	-	-
3	**CCM**	n/a	n/a	-	-	**3 (CCM)**	0/1	-	-
						4 (CCM)	0/1	-	-
						5 (CCM)	0/1	-	-
						6 (CCM)	0/1	-	-
						7 (CCM)	0/1	-	-
						9 (CCM)	0/4	-	-
						10 (CCM)	0/3	-	-
						11 (CCM)	0/1	-	-
						12 (CCM)	0/1	-	-
						13 (CCM)	0/2	-	-
						15 (CCM)	0/1	-	-
						16 (CCM)	0/1	-	-
						18 (CCM)	0/1	-	-
						19 (CCM)	0/1	-	-
						20 (CCM)	0/4	-	-

Table S4: Primers used for ddPCR analyses

KRAS c.35G>A;p.(Gly12Asp)	
Forward Primer	AGGCCTGCTGAAAATGACTGAATAT
Reverse Primer	GCTGTATCGTCAAGGCACTCTT
Reporter 1 (VIC) sequence	TTGGAGCTG**G**TGGCGTA
Reporter 2 (FAM) sequence	TTGGAGCTG**A**TGGCGTA
KRAS c.35G>T;p.(Gly12Val)	
Forward Primer	AGGCCTGCTGAAAATGACTGAATAT
Reverse Primer	GCTGTATCGTCAAGGCACTCTT
reporter 1 (VIC) sequence	TTGGAGCTG**G**TGGCGTA
reporter 2 (FAM) sequence	TTGGAGCTG**T**TGGCGTA
KRAS c.38G>T;p.(Gly13Val)	
Forward Primer	AGCTGTATCGTCAAGGCACTCTT
Reverse Primer	GGCCTGCTGAAAATGACTGAA
Reporter 1 (VIC) sequence	CCTACG**C**CACCAGCT
Reporter 2 (FAM) sequence	CTACG**A**CACCAGCT
HRAS c.35G>A;p.(Gly12Asp)	
Forward Primer	CACAAAATGGTTCTGGATCAGC
Reverse Primer	CGATGACGGAATATAAGCTGGTG
Reporter 1 (VIC) sequence	TTGCCCACACCG**C**CG
Reporter 2 (FAM) sequence	CTTGCCCACACCG**T**CG

KRAS c.183A>T;p.(Gln61His) and *KRAS* c.436G>A;p.(Ala146Thr) validated TaqMan® SNP Genotyping Assays for digital PCR were obtained from Life Technologies (Part 4332077, Assay IDs: AHFA90M and AHWSLAG, respectively).

Table S5: Excel spreadsheet showing the number of positive events and the number of wild-type and total events for every Toronto BAVM sample tested by *KRAS* c.35G>T;p.(Gly12Val)- or c.35G>A;p.(Gly12Asp)-specific ddPCR [Included as a separate Excel file]

Table S6: Summary of clinical data and *KRAS* status of Fresh Frozen Toronto patients
Summary of the clinical presentation and *KRAS* c.35G>A;p.(Gly12Asp), c.35G>T;p.(Gly12Val) and *KRAS* c.183A>T;p.(Gln61His) status of Toronto patients and their BAVM samples.

PATIENT	age	Gender	KRAS mutation	Rupture	Location	size 1 (0-3cm), 2(3-6cm), 3 (>6cm)	venous drainage	Prior Radiation
1	26	M	G12D	yes	cingulate	2	deep	no
2	23	M	G12D	yes	temporal	1	deep	yes
3	21	M	G12V	yes	cerebellar	1	both	yes
4	18	F	G12D	no	parieto-occipital	1	superficial	no
5	42	F	G12V	no	frontal	2	superficial	no
6	18	M	G12D	no	frontal	1	superficial	no
7	53	F	G12V	no	temporal	2	superficial	no
8	39	F	Q61H	no	occipital	2	both	no
9	26	M	G12V	yes	parietal	1	superficial	no
10	26	M	G12D	no	frontal	2	deep	no
11	42	F	G12V	yes	occipital	1	superficial	no
12	66	M	G12V	yes	temporal	1	superficial	no
13	59	F	G12V	no	temporal	2	superficial	no
14	17	F	G12D	no	temporal	1	deep	no
15	29	M	G12D	yes	frontal	1	superficial	no
16	59	M	G12D	yes	frontal	1	superficial	no
17	34	F	G12D	no	frontal	1	superficial	no
18	19	F	G12D	yes	cerebellar	2	superficial	no
19	26	M	neg	yes	occipital	1	deep	no
20	52	F	neg	yes	parietal	3.0	both	no
21	29	M	neg	yes	temporal	1	superficial	no
22	32	F	neg	no	frontal	1	both	no
23	27	F	neg	no	frontal	2	superficial	no
24	59	F	neg	yes	parietal	1	superficial	yes
25	56	M	neg	no	temporal	1	superficial	no
25	25	M	neg	yes	temporal	1	superficial	yes
26	48	F	neg	no	parietal	2	superficial	no
28	32	F	G12D	no	insular	1	superficial	no
29	25	F	G12D	no	frontal	2	both	no
30	75	M	G12V	yes	cerebellar	2	both	yes
31	40	M	G12D	no	parieto-occipital	1	deep	no
32	46	M		yes	frontal	1	superficial	no
33	48	F	G12D	yes	frontal	1	both	no
34	27	M	G12D	yes	frontal	1.0	superficial	no
35	25	F	G12V	no	occipital	1.0	superficial	yes
36	46	M	G12D	yes	parieto-occipital	1.0	superficial	no
37	58	F	G12D	no	occipital	2.0	superficial	no
38	53	F	neg	yes	frontal	2	both	no
39	58	M	G12D	yes	parietal	2.0	superficial	no

Table S7: Excel spreadsheet showing the number of positive events and the number of wild-type events for every Kuopio sample tested by ddPCR, including UDC treatment of FFPE BAVM and control vascular samples [Included as a separate Excel file]
The amount of genomic DNA in the ddPCR reactions was 500 ng unless otherwise indicated in *italic*, in which case 250 ng was used.

Table S8: Summary of clinical data and ddPCR results of Finnish FFPE BAVM samples
Summary of the clinical presentation and *KRAS* c.35G>A;p.(Gly12Asp) and c.35G>T;p.(Gly12Val) ddPCR results for the KUH replication cohort. * Fractional abundance is presented as mean and range (min-max) of all those PCRs performed on the sample that gave a WT replication of >500 copies

Patient	Age	Sex	Rupture	Epilepsy at presentation	Location	Size 1=<3cm, 2=3-6cm, 3>6cm	Venous drainage	Prior radiation	KRAS variant	Fractional abundance G12D	G12D ddPCRs	Fractional abundance G12V	G12V ddPCRs	Mutation + /all blocks
1	42	F	yes	no	Occipito-parietal	1	superficial	no	neg	0.20 (0.04-0.32)	4	0.00 (0.00-0.00)	2	0/1
2	60	M	yes	no	Temporal	1	superficial	no	neg	0.12 (0.09-0.14)	3	0.04 (0.02-0.06)	2	0/1
3	39	F	no	no	Occipital	1	superficial	no	neg	0.27 (0.12-0.50)	3	0.04 (0.00-0.11)	2	0/2
4	32	M	yes	no	Occipital	1	superficial	no	**G12D**	**1.40 (1.20-1.60)**	3	0.04 (0.00-0.07)	2	**1/1**
5	31	F	yes	no	Frontal	1	superficial	no	**G12D**	**3.60 (3.40-3.70)**	3	0.07 (0.00-0.13)	2	**1/1**
6	41	F	no	no	Occipital	2	superficial	yes	neg	0.24 (0.00-0.80)	2-4	0.02 (0.00-0.15)	2	0/6
7	57	M	no	no	Occipital	1	deep	no	neg	0.26 (0.00-0.60)	3-6	0.00 (0.00-0.00)	2	0/4
8	13	F	yes	yes	Frontal	2	superficial	no	neg	0.06 (0.00-0.11)	4	0.00 (0.00-0.00)	2	0/3
9	40	F	yes	no	Frontal	1	deep	no	**G12D**	**2.15 (1.73-2.70)**	4	0.00 (0.00-0.00)	2	**2/2**
10	62	M	yes	yes	Fronto-parietal	1	superficial	no	**G12D**	**3.19 (2.70-3.50)**	4	0.00 (0.00-0.01)	2	**1/1**
11	52	M	yes	no	Parietal	3	superficial	no	**G12V**	0.25 (0.19-0.30)	2	**1.35 (0.94-1.90)**	3	**1/1**
12	21	F	yes	no	Frontal	1	superficial	no	neg	0.12 (0.00-0.20)	3-4	0.00 (0.00-0.00)	2	0/5
13	42	M	yes	no	Occipital	1	superficial	no	**G12D**	**3.45 (1.90-4.42)**	3	0.00 (0.00-0.00)	2	**1/1**
14	48	M	yes	yes	Parietal	1	superficial	no	**G12V**	0.50 (0.31-0.87)	2	**1.28 (0.70-2.10)**	4	**1/2**
15	36	M	yes	no	Temporal	1	superficial	no	neg	0.17 (0.00-0.33)	2	0.00 (0.00-0.00)	2	0/1
16	45	M	yes	no	Fronto-temporo-parietal	1	superficial	no	**G12V**	0.32 (0.20-0.46)	2-3	**0.70 (0.41-1.00)**	4	**1/2**
17	38	F	yes	no	Occipital	1	superficial	no	**G12D**	**1.05 (0.86-1.40)**	3	0.02 (0.00-0.04)	2	**1/1**
18	38	M	yes	no	Parietal	1	not known	no	**G12D**	**1.31 (1.28-1.33)**	3	0.01 (0.00-0.02)	2	**1/1**
19	45	F	no	no	Parietal	1	not known	no	neg	0.31 (0.14-0.48)	2	0.12 (0.00-0.23)	3	0/1
20	36	F	yes	no	Sylvian fissure /temporal	1	deep & superficial	no	**G12V**	0.49 (0.19-0.69)	2	**1.80 (1.10-1.80)**	3	**1/1**
21	21	F	no	yes	Frontal	2	superficial	no	**G12D**	**2.08 (1.20-3.40)**	3	0.00 (0.00-0.00)	2	**2/2**
22	12	F	yes	no	Frontal	uncertain	deep & superficial	no	**G12D**	**1.68 (0.80-2.40)**	2-3	0.00 (0.00-0.00)	2	**2/4**
23	14	F	yes	no	Frontal	1	superficial	no	neg	0.37 (0.26-0.48)	2	0.05 (0.00-0.11)	2	0/1
24	29	F	no	yes	Temporobasal	3	deep & superficial	no	neg	0.23 (0.14-0.31)	2	0.02 (0.00-0.04)	2	0/1
25	60	M	no	yes	Frontal	1	superficial	no	neg	0.13 (0.00-0.20)	2	0.07 (0.00-0.29)	2-3	0/1
26	52	F	yes	no	Frontoparietal	3	superficial	yes	neg	0.39 (0.17-0.80)	2	0.00 (0.00-0.00)	2	0/3
27	4	F	yes	no	Parietal	1	superficial	no	**G12D**	**3.00 (2.90-3.10)**	3	0.00 (0.00-0.00)	2	**1/1**
28	64	F	yes	yes	Occipital	1	superficial	no	neg	0.31 (0.26-0.36)	2	0.01 (0.00-0.03)	2	0/1
29	33	F	no	no	Sylvian fissure	1	superficial	no	**G12D**	**1.24 (1.06-1.45)**	3	0.00 (0.00-0.00)	2	**1/1**
30	22	M	yes	no	Thalamic	2	deep	no	neg	0.43 (0.38-0.48)	2	0.03 (0.01-0.05)	2	0/1
31	50	M	yes	no	Sylvian fissure / parietal	2	superficial	no	neg	0.25 (0.07-0.60)	2-3	0.01 (0.00-0.08)	2	0/3
32	65	M	no	no	Sylvian fissure / frontal	1	superficial	no	**G12D**	**2.20 (1.90-2.70)**	3	0.04 (0.00-0.07)	2	**1/1**
33	25	M	yes	no	Frontal / basal ganglia	uncertain	deep	no	neg	0.29 (0.28-0.29)	2	0.00 (0.00-0.00)	2	0/1

Table S9: ddPCR results and fractional abundance of *KRAS* variants in MAC-sorted CD31$^+$ and CD31$^-$ fractions of primary vascular cell cultures of BAVMs and normal brain vessels.

Note that all material from AVM tissue from Patient #32 was used for MAC-sorting, so analysis of the AVM itself was not possible.

Sample	Fractional Abundance (%)	KRAS variant
AVM 1 tissue (Patient #29)	0.763	c.35G>A; p.(Gly12Asp)
AVM 2 tissue (Patient #30)	1.6	c.35G>T; p.(Gly12Val)
AVM 3 tissue (Patient #31)	0.753	c.35G>A; p.(Gly12Asp)
AVM 4 tissue (Patient #32)	n/a	n/a
AVM 5 tissue (Patient #38)	0.0	0.0
AVM 6 tissue (Patient #39)	Vein: 0.283 Nidus: 2.34	c.35G<T; p.(Gly12Asp)
AVM 1 CD31+ (Patient #29)	4.38	c.35G>A; p.(Gly12Asp)
AVM 2 CD31+ (Patient #30)	14.9	c.35G>T; p.(Gly12Val)
AVM 3 CD31+ (Patient #31)	2.13	c.35G>A; p.(Gly12Asp)
AVM 4 CD31+ (Patient #32	45.8	c.35G>A; p.(Gly12Asp)
AVM 5 CD31+ (Patient #38)	0.0	0.0
AVM 6 CD31+ (Patient #39)	Vein: 30.5 Nidus: 52.1	c.35G>A; p.(Gly12Asp)
AVM 1 CD31- (Patient #29)	0.0	0.0
AVM 2 CD31- (Patient #30)	0.0	0.0
AVM 3 CD31- (Patient #31)	0.0	0.0
AVM 4 CD31- (Patient #32)	0.0	0.0
AVM 5 CD31- (Patient #38)	0.0	0.0
AVM 6 CD31- (Patient #39)	0.0	0.0
NV 1 CD31+	0.0	0.0
NV 2 CD31+	0.0	0.0
NV 3 CD31+	0.0	0.0
NV 1 CD31-	0.0	0.0
NV 2 CD31-	0.0	0.0
NV 3 CD31-	0.0	0.0

KRAS variants c.35G>A;p.(Gly12Asp) or c.35G>T;p.(Gly12Val) are found in 5 of 6 endothelium enriched, CD31$^+$ cultures, but are absent in endothelium-depleted, CD31$^-$ BAVM fractions, or in either fractions of normal brain vessel cultures.

Table S10: Primers used for qRT-PCR analyses

Gene	Forward Primer	Reverse Primer
BHLHE40	CCT TGA AGC ATG TGA AAG CA	CAT GTC TGG AAA CCT GAG CA
DUSP1	CCA ACC ATT TTG AGG GTC AC	ACC CTT CCT CCA GCA TTC TT
DUSP5	ATG GAT CCC TGT GGA AGA CA	TCA CAG TGG ACC AGG ACC TT
DUSP6	CTG TCG ATG AAC GAT GCC TA	AGC AGC TGA CCC ATG AAG TT
EGR1	CAG CAC CTT CAA CCC TCA G	TAA CTG GTC TCC ACC AGC AC
HES1	CGG ACA TTC TGG AAA TGA CA	GTC ACC TCG TTC ATG CAC TC
IL1A	TGT GAC TGC CCA AGA TGA AG	CCG TGA GTT TCC CAG AAG AA
SIK1	CAG TAG GCA CCC GAG CAG	CGC TGA ACT CCG ACA TGA TA
SPRY4	TCC GTA CAG TCC AGG ACC TC	GGC TGG ACC ATG ACT GAG TT
TBP	CTC CAC GGA GGA GAG AAC C	AAG CAT CAT CTT CCC CCT TC
TRIB1	TCG GAG AGT TCT GGG ATT GT	CAC GAA GTG CAA TGG TCT TT

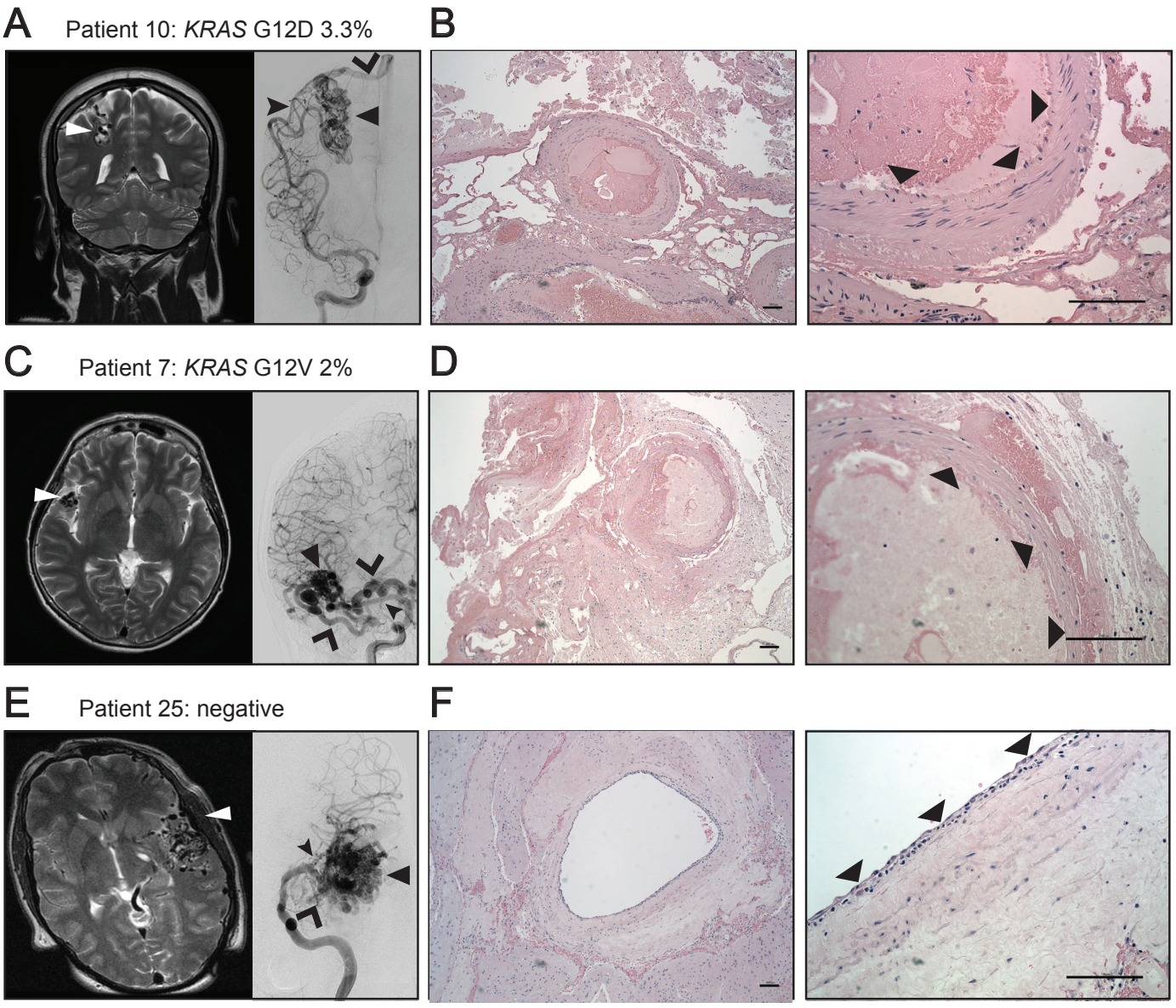
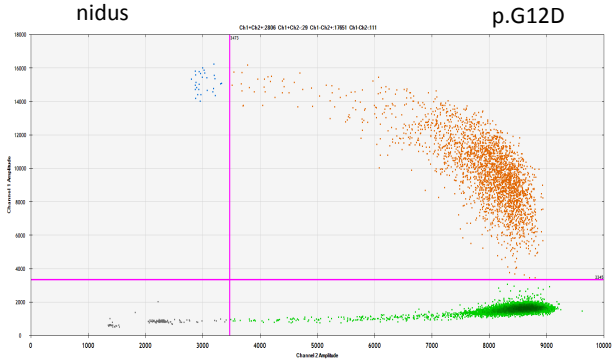


Figure S1: Examples of brain arteriovenous malformations (BAVMs) in this study. (A, C, E) Right: MRI T2 sequence (right) showing BAVMs. Characteristic dark flow voids delineate the nidus (white arrowheads) and surrounding feeding arteries and draining veins. Left: angiographic appearance of BAVMs, demonstrating BAVM angioarchitecture: feeding arteries (black arrowhead) directly connect to the nidus (black pointer) and shunt arterialized blood in early filling draining vein(s) (open arrowhead). **(B, D, F)** Histopathological appearance of BAVMs shown in A, C, and E respectively, composed of abnormal vessels of varied size and thickness. The vessel wall consists of a single endothelial cell lining and a multilayer thick wall of smooth muscle cells. Brain or loose connective tissue is found between the vessels. Left panels are magnified 10x, right panels 40X. **(A, B)** A right parietal BAVM in Patient 10 carries a *KRAS* c.35G>A;p.(Gly12Asp) variant with an abundance of 3.3% which represents 6.6% of cells in the specimen, assuming that the variant is heterozygous. **(C, D)** A right temporal BAVM in Patient 7 carries a variant in c.35G>A;p.(Gly12Asp) with an abundance of 2% (4% affected cells) and **(E, F)** A left temporal BAVM in Patient 25, which was negative for *KRAS* variants. Scale bars = 100 μm.

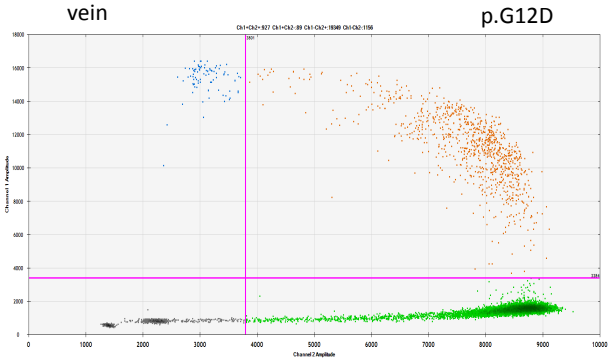
Patient 1
nidus

p.G12D



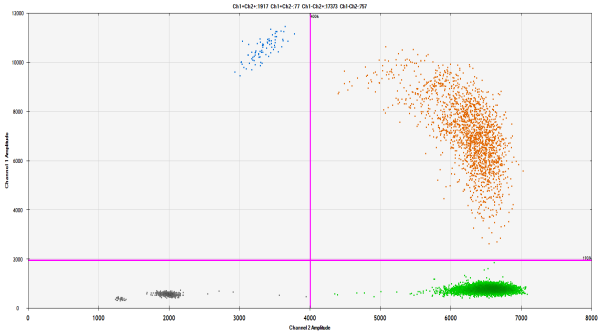
Patient 1
vein

p.G12D



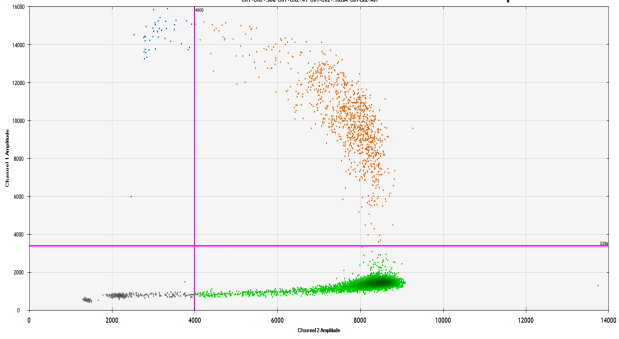
Patient 2

p.G12V



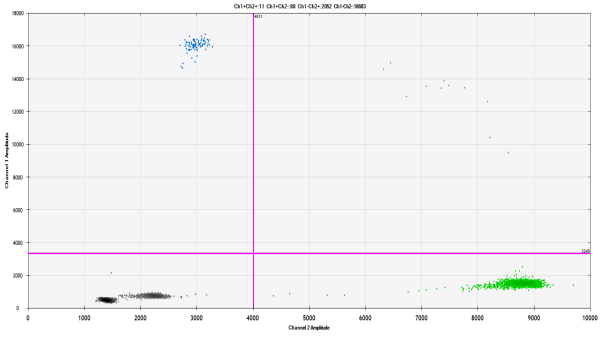
Patient 3

p.G12D



Patient 4

p.G12D

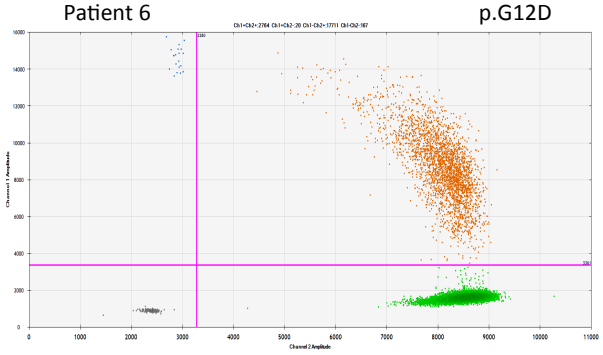


Patient 5

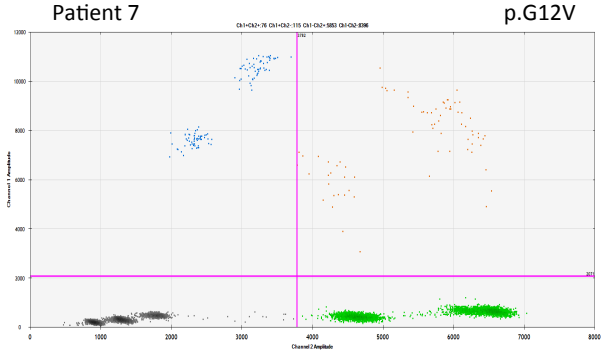
p.G12V

Figure S2

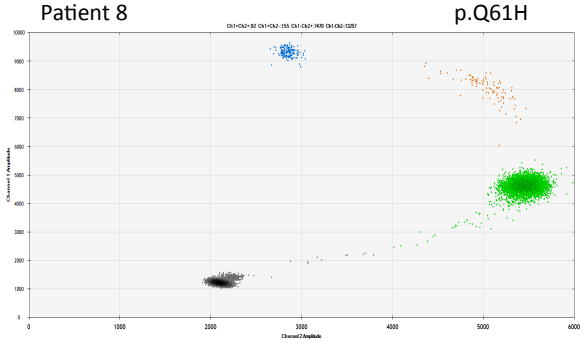
Patient 6 p.G12D



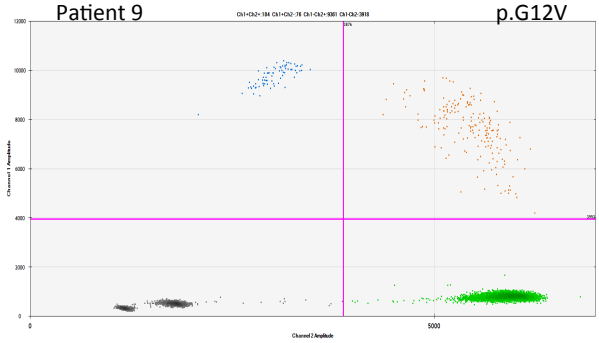
Patient 7 p.G12V



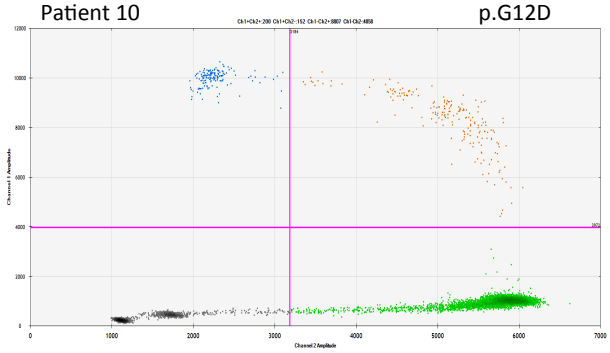
Patient 8 p.Q61H



Patient 9 p.G12V



Patient 10 p.G12D



Patient 11 p.G12V

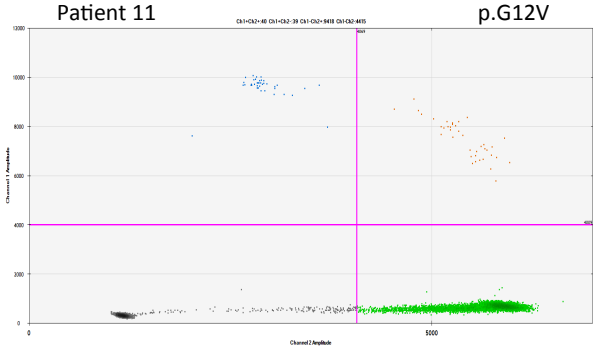
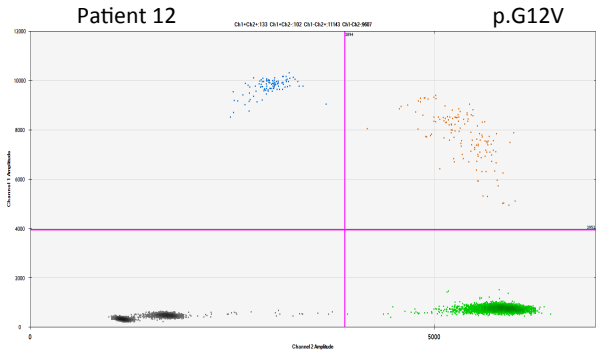
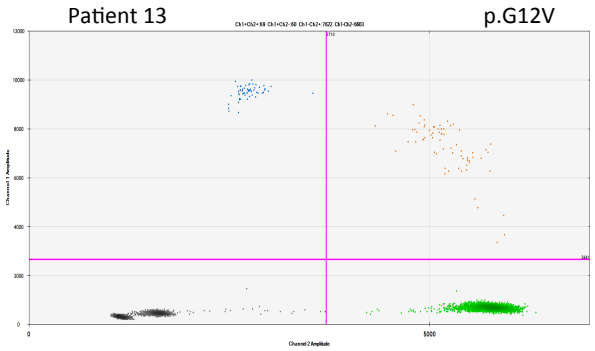
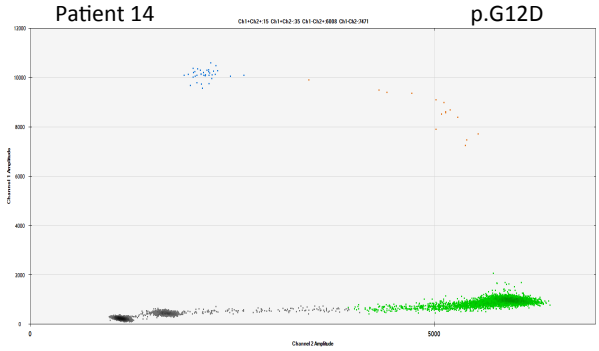
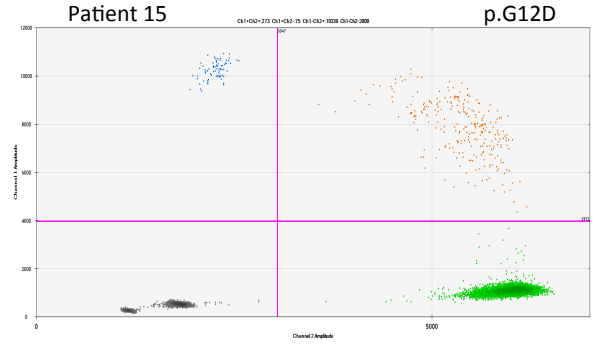


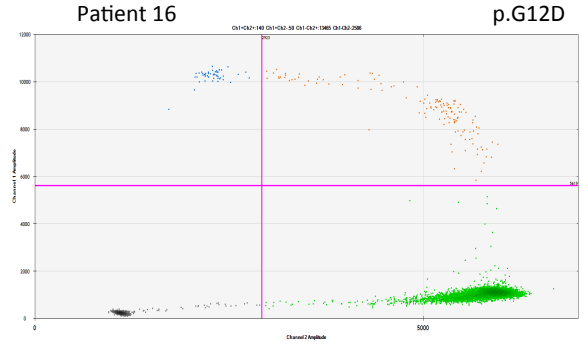
Figure S2 - continued











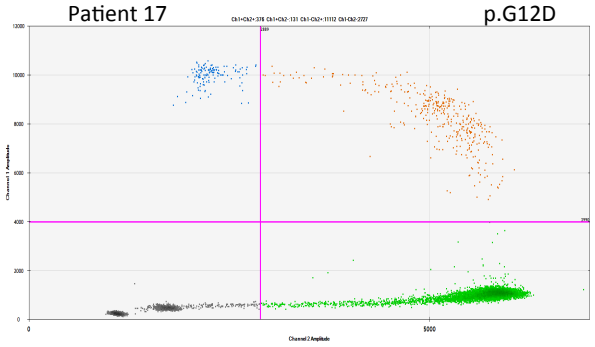
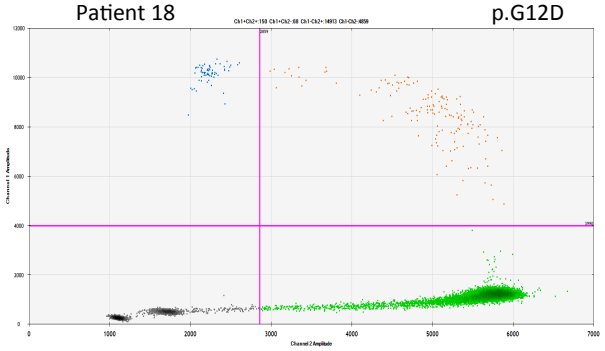
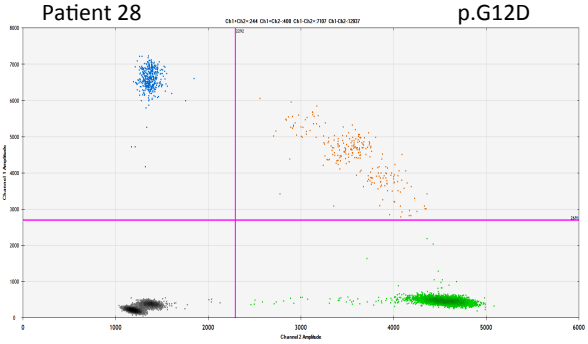


Figure S2 - continued

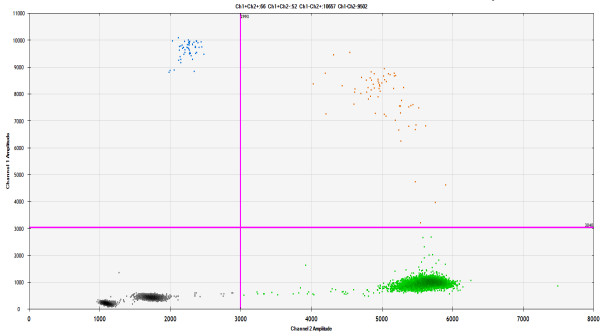
Patient 18 p.G12D



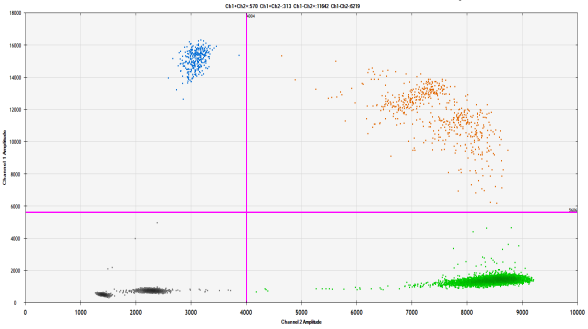
Patient 28 p.G12D



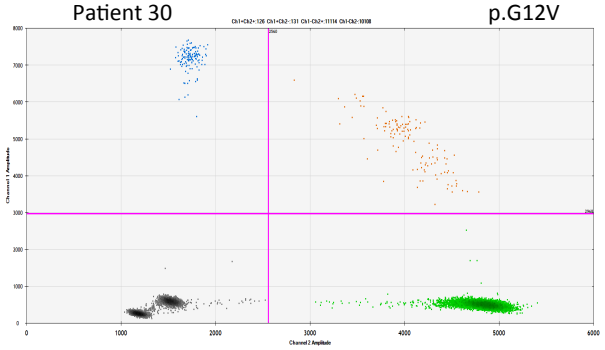
Patient 29 p.G12D



Patient 29, AVM1, EC-enriched p.G12D



Patient 30 p.G12V



Patient 30, AVM2, EC-enriched p.G12V

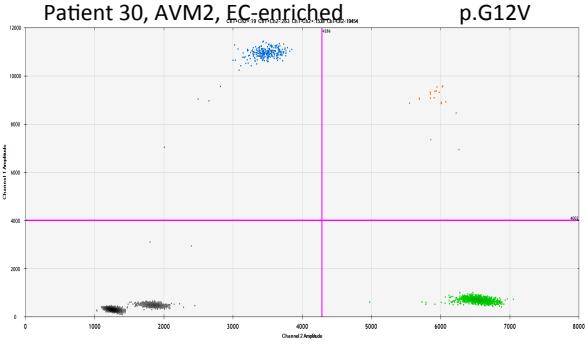
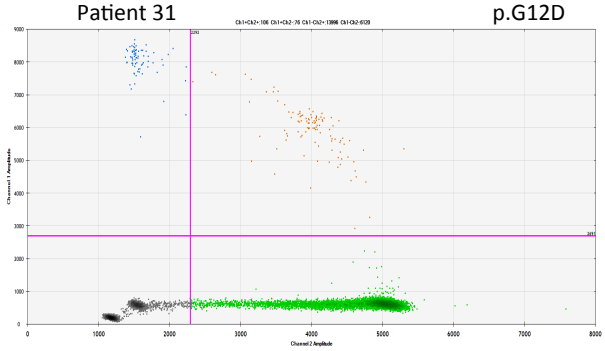
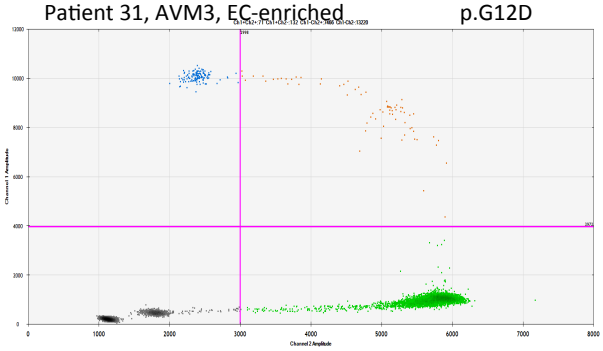


Figure S2 - continued

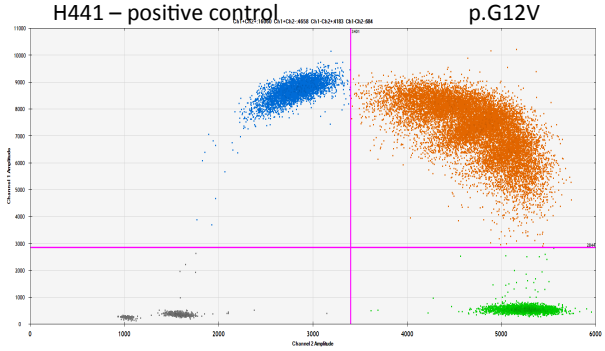
Patient 31 p.G12D



Patient 31, AVM3, EC-enriched p.G12D



H441 – positive control p.G12V



Patient 32, AVM4, EC-enriched p.G12D

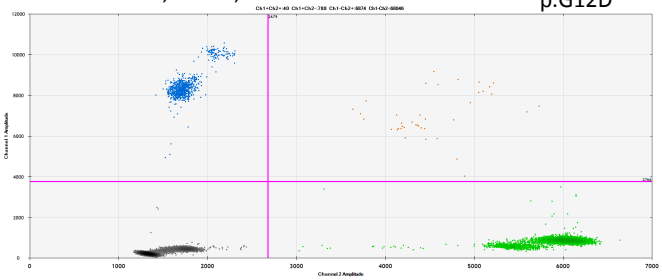
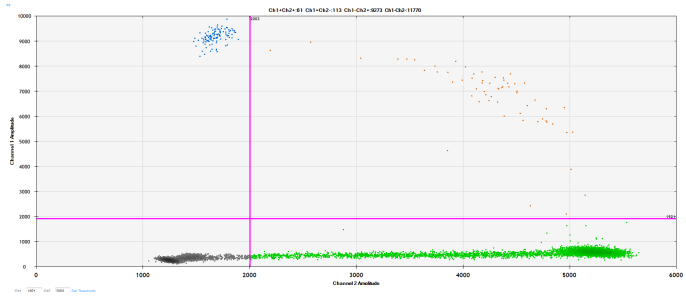
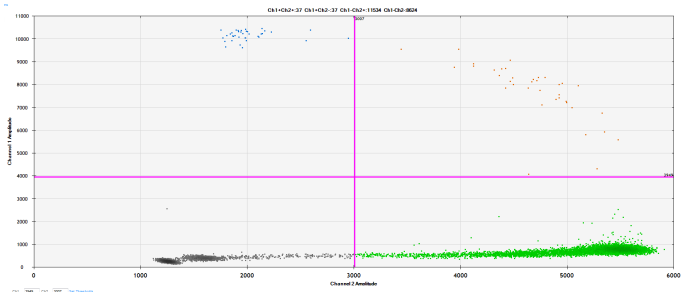


Figure S2 - continued

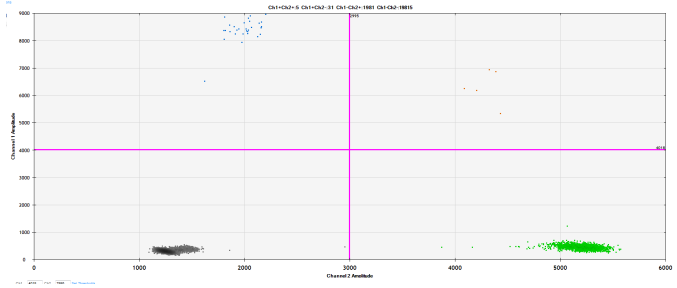
Patient 33 p.G12D



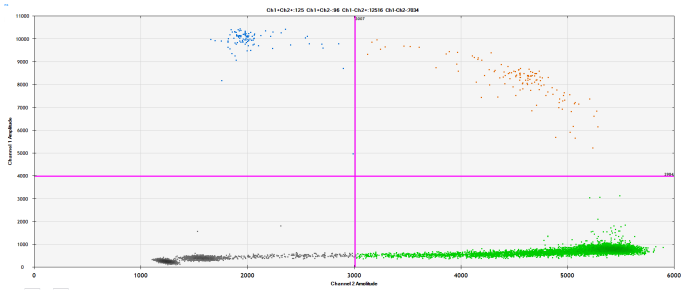
Patient 34 p.G12D



Patient 35 p.G12V



Patient 36 p.G12D



Patient 37 p.G12D

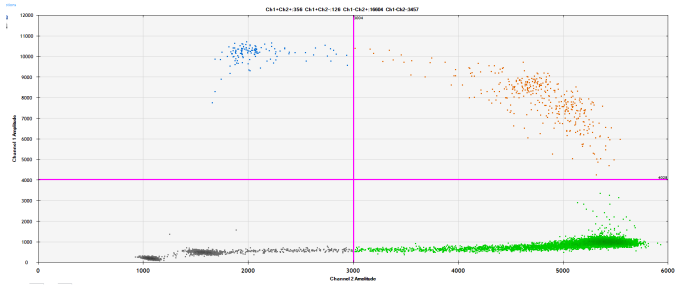
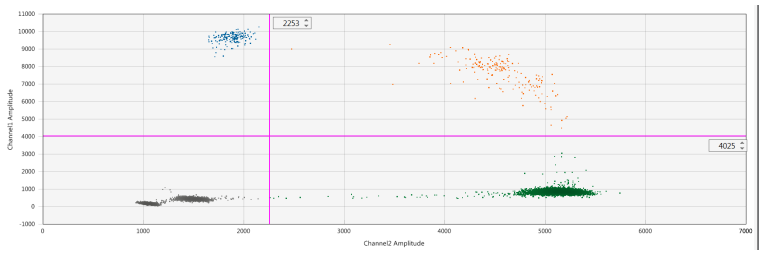


Figure S2 - continued

Patient 39 p.G12D
Nidus



Patient 39 p.G12D
Vein

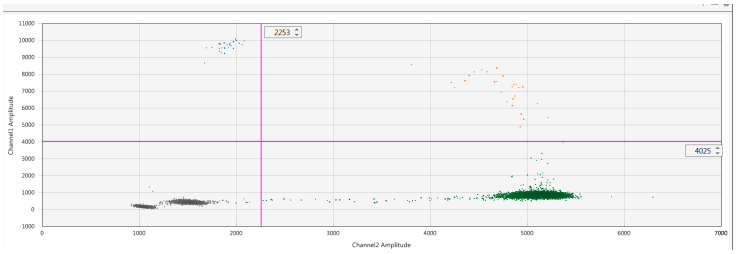


Figure S2: Amplitude scatter plots of *KRAS* c.35G>T;p.(Gly12Val) and c. 35G>A;p.(Gly12Asp) ddPCRs of frozen samples (Toronto cohort). All BAVM samples (including EC-enriched samples) with their respective anonymized patient numbers are shown. Dots represent single well data. Chanel 1 is the variant assay (FAM) duplexed with the wild-type reference assay (HEX, Chanel 2). Positive events are in the upper quadrant of each plot.

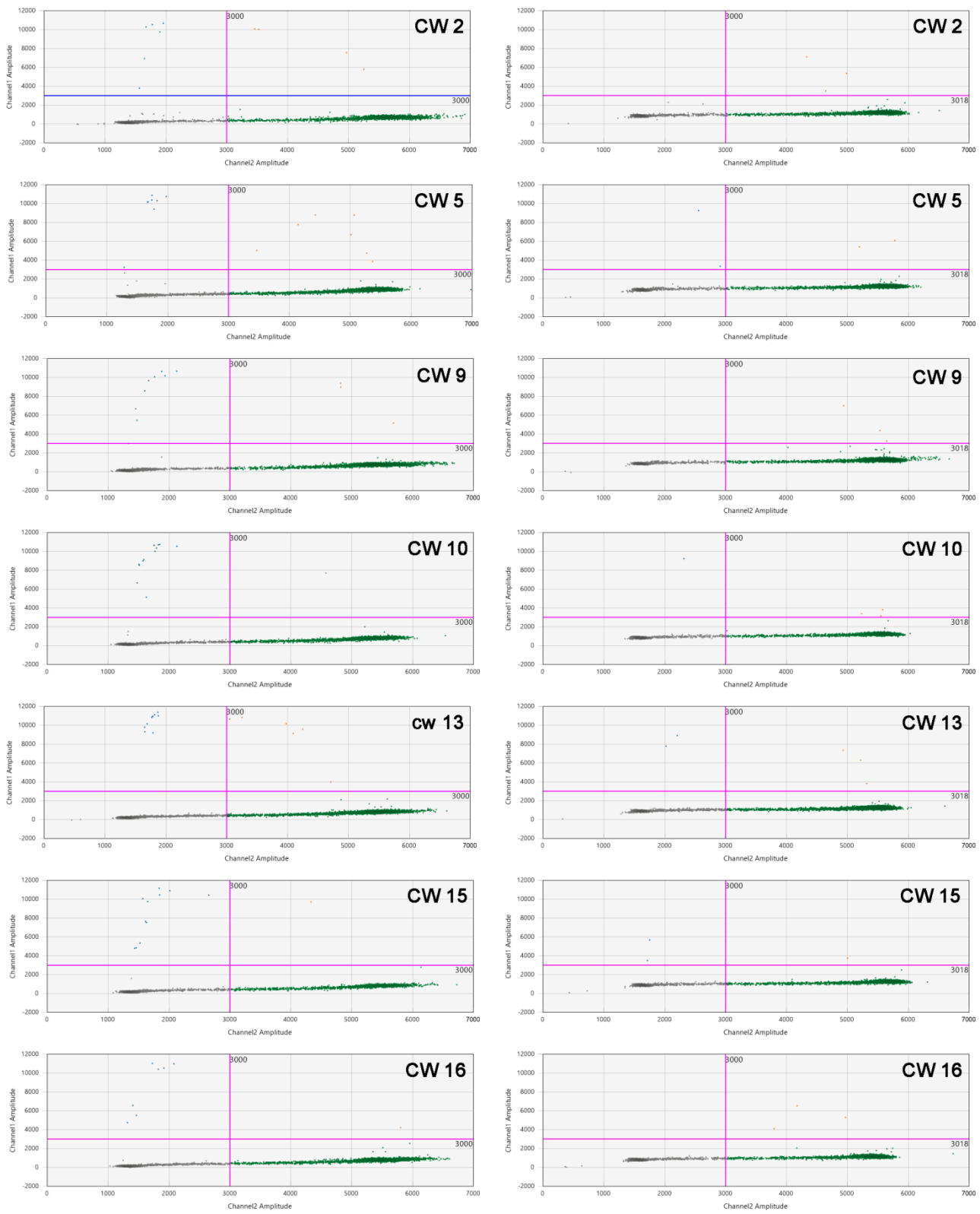
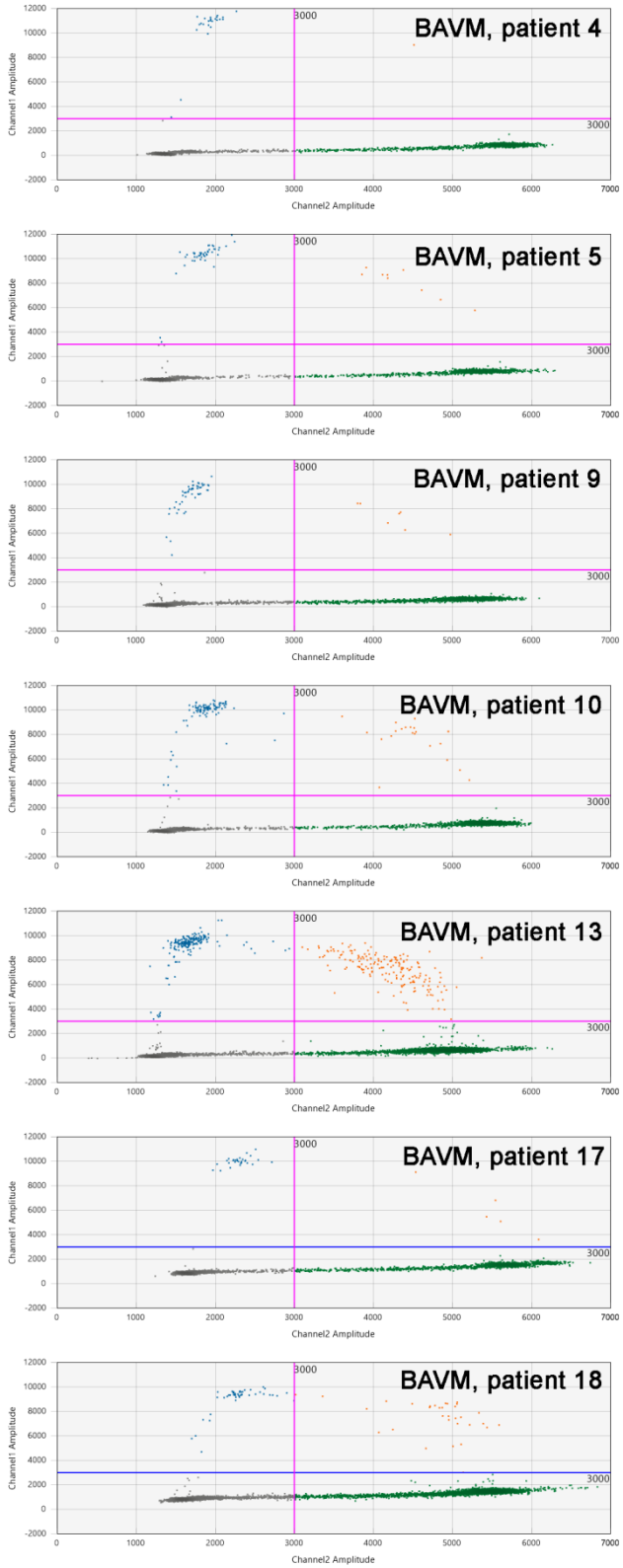

KRAS G12D, non-treated gDNA:
KRAS G12D, UDG-treated gDNA:
CW 2
CW 5
CW 9
CW 10
cw 13
CW 13
CW 15
CW 16
Channel1 Amplitude
Channel2 Amplitude
3000
3018

Figure S3

KRAS G12D, non-treated gDNA:



KRAS G12D, UDG-treated gDNA:

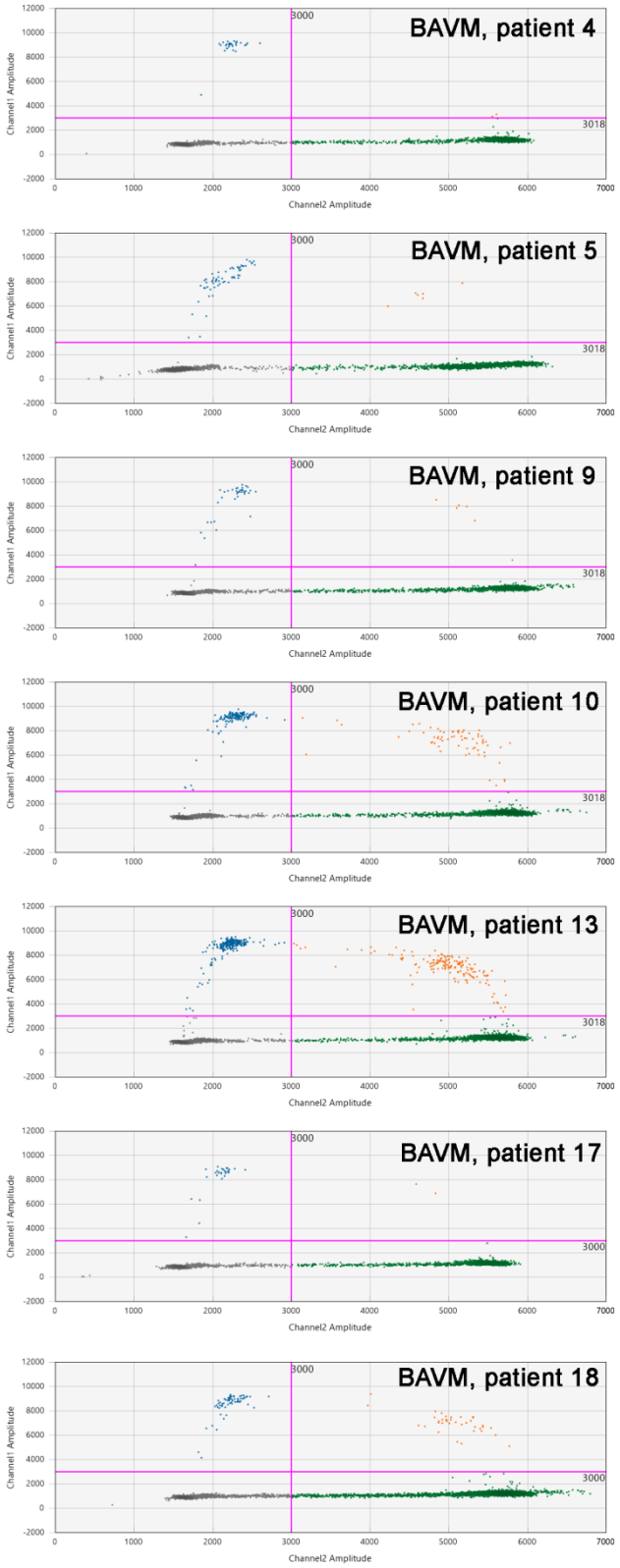
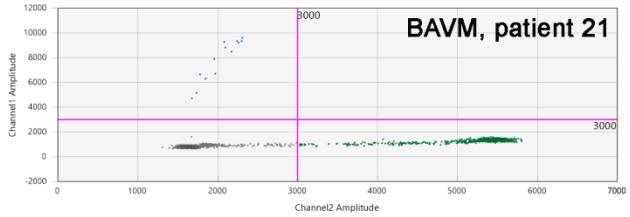
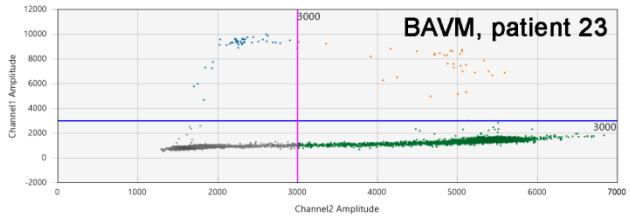
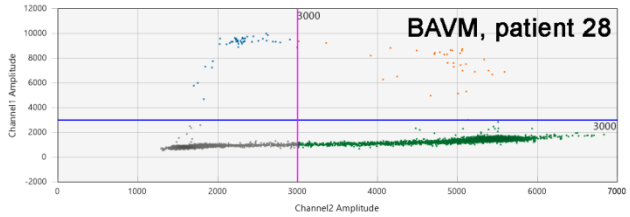


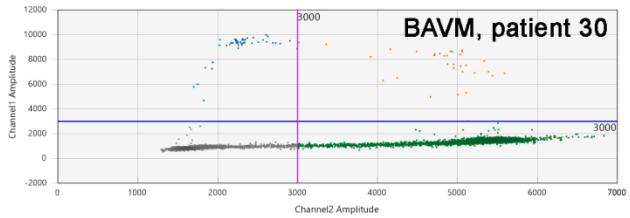
Figure S3 - continued

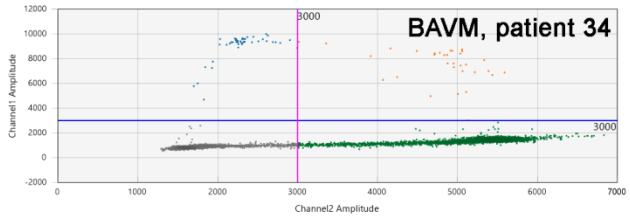
KRAS G12D, non-treated gDNA:



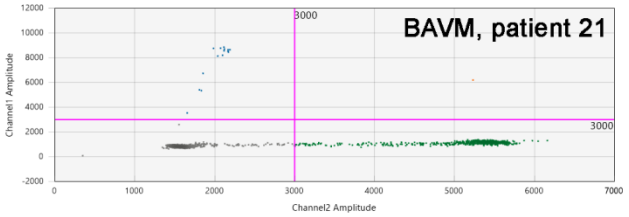


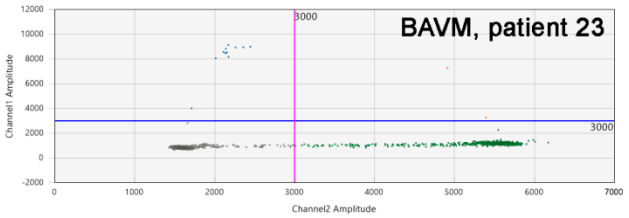


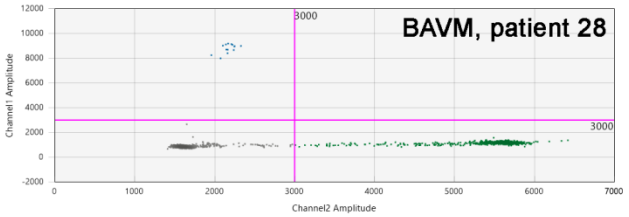


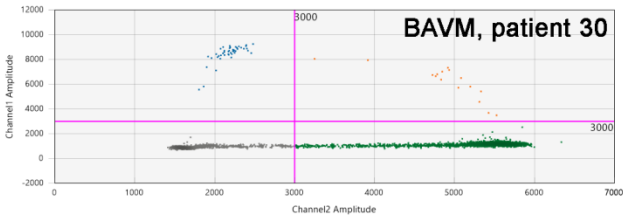


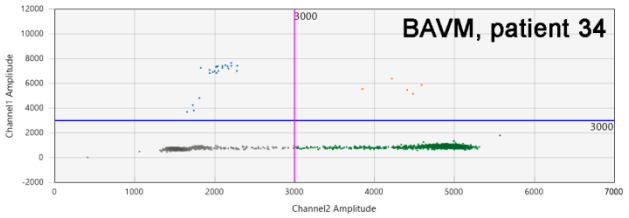
KRAS G12D, UDG-treated gDNA:



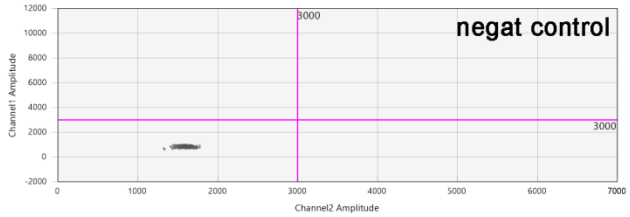


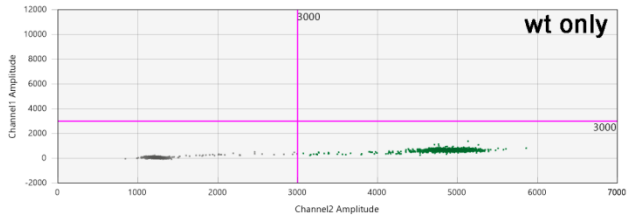






KRAS G12D, control DNA sequences:





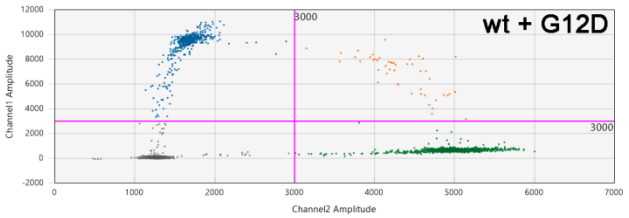


Figure S3 - continued

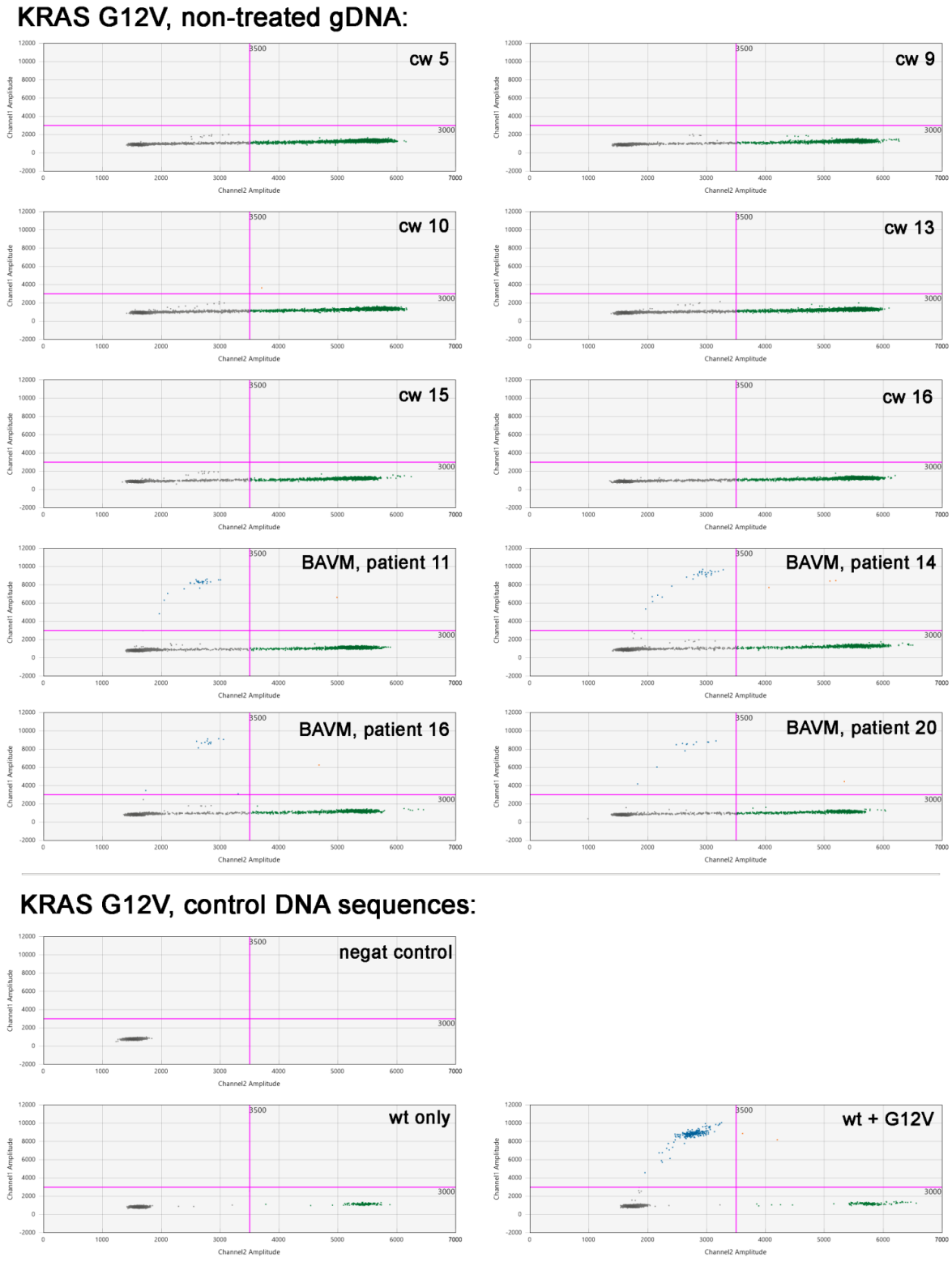


Figure S3: Amplitude scatter plots of *KRAS* c.35G>T;p.(Gly12Val) and c.35G>A;p.(Gly12Asp) ddPCRs of FFPE sample (Finnish cohort) with and without UDG treatment. ddPCR plots with and without uracil DNA glycosylase treatment for a representative subgroup of Circle of Willis control samples from the KUH cohort is shown. All variant-positive BAVM samples from the KUH cohort with their respective anonymized patient numbers are shown with and without UDG treatment. Dots represent single well data. Chanel 1 is the variant assay (FAM) duplexed with the wild-type reference assay (HEX, Chanel 2). Positive events are in the upper quadrant of each plot.

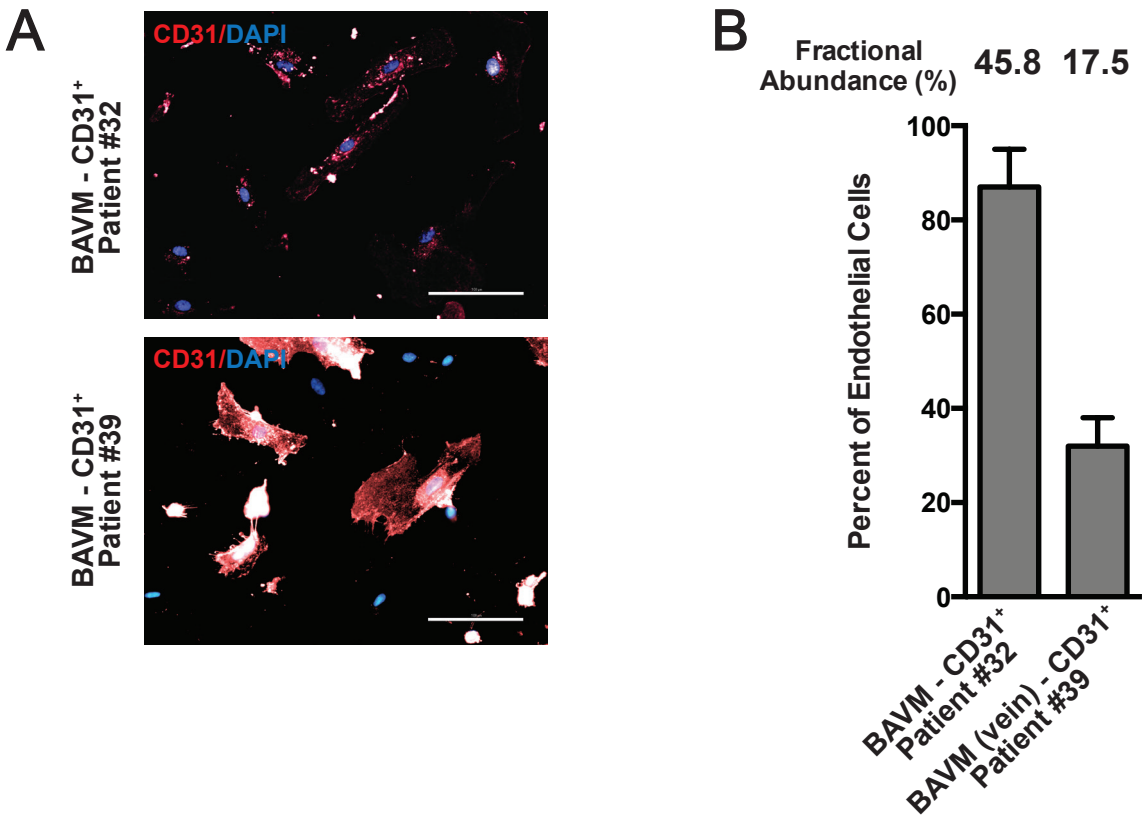


Figure S4: Endothelial cell enrichment of *KRAS* variants in BAVMs. (A) CD31 immunostaining of MACS-sorted CD31⁺ fractions of BAVMs from Patients #32 and #39 used for quantification of endothelial cells. Scale bars=100 µm. **(B)** Quantification of CD31⁺ cells in CD31 MAC-sorted fractions of BAVMs from Patients #32 and #39 compared to their respective allele frequency of *KRAS* variants.

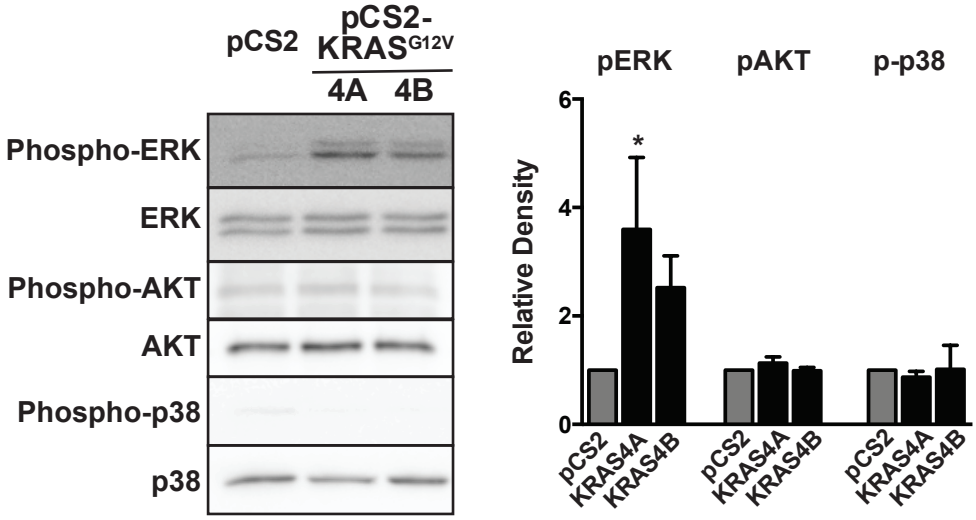


Figure S5: Active KRAS specifically induces ERK1/2 phosphorylation in endothelial cells. HUVEC expressing c.35G>T;p.(Gly12Val) KRAS isoforms 4A or 4B show elevated phospho-ERK levels, whereas no increases in phospho-Akt or phospho-p38 are detected.

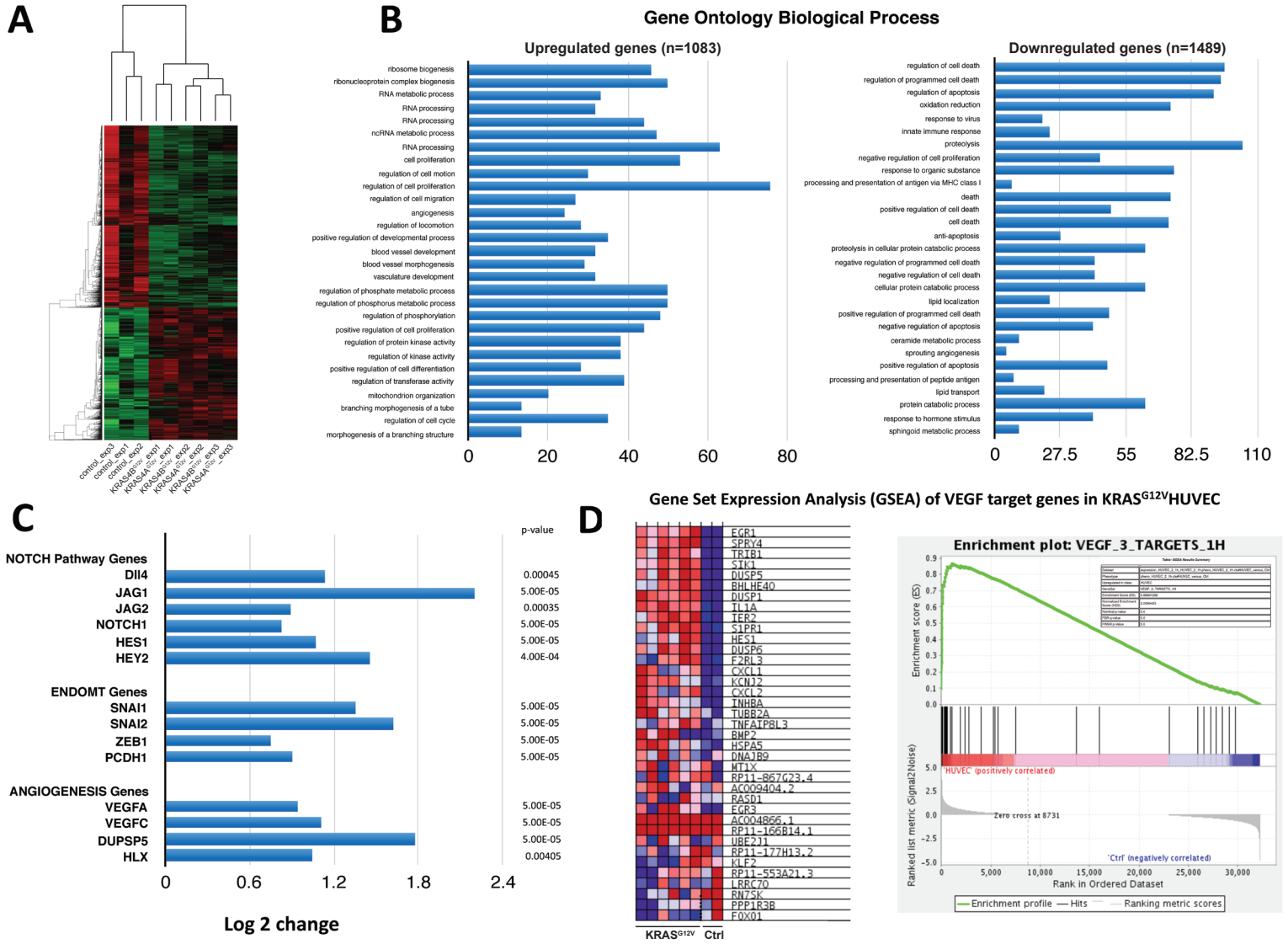


Figure S6: Expression of activated KRAS induces pro-angiogenic and Notch pathway genes in endothelial cells. (A) Heatmap showing general clustering of control Human Umbilical Endothelial Cells (HUVECs) versus KRAS4A^{G12V}- and KRAS4B^{G12V}-expressing HUVECs. **(B)** Gene Ontology (GO) term biological process analysis of upregulated and downregulated genes in KRAS^{G12V}-expressing HUVECs showing the 30 most significant biological processes. Bars show the number of genes with significant differential expression for each biological process. **(C)** Over-expression of KRAS^{G12V} induced the expression of several Notch signaling components (i.e. *DLL4, JAG1, JAG2, NOTCH1, HEY2, HES1*), Endothelial to Mesenchymal Transition (EndoMT) genes (i.e *SNAI1, SNAI2, ZEB1, PCDH1*) and induced angiogenic genes (*VEGFA, VEGFC, HLX*) and a MAPK/ERK-responsive gene (*DUSP5*). **(D)** Heatmap showing differential expression of a VEGF induced gene set in HUVECs and Gene Set Expression Analysis showing enrichment of VEGF induced genes in KRAS^{G12V}-expressing HUVECs.

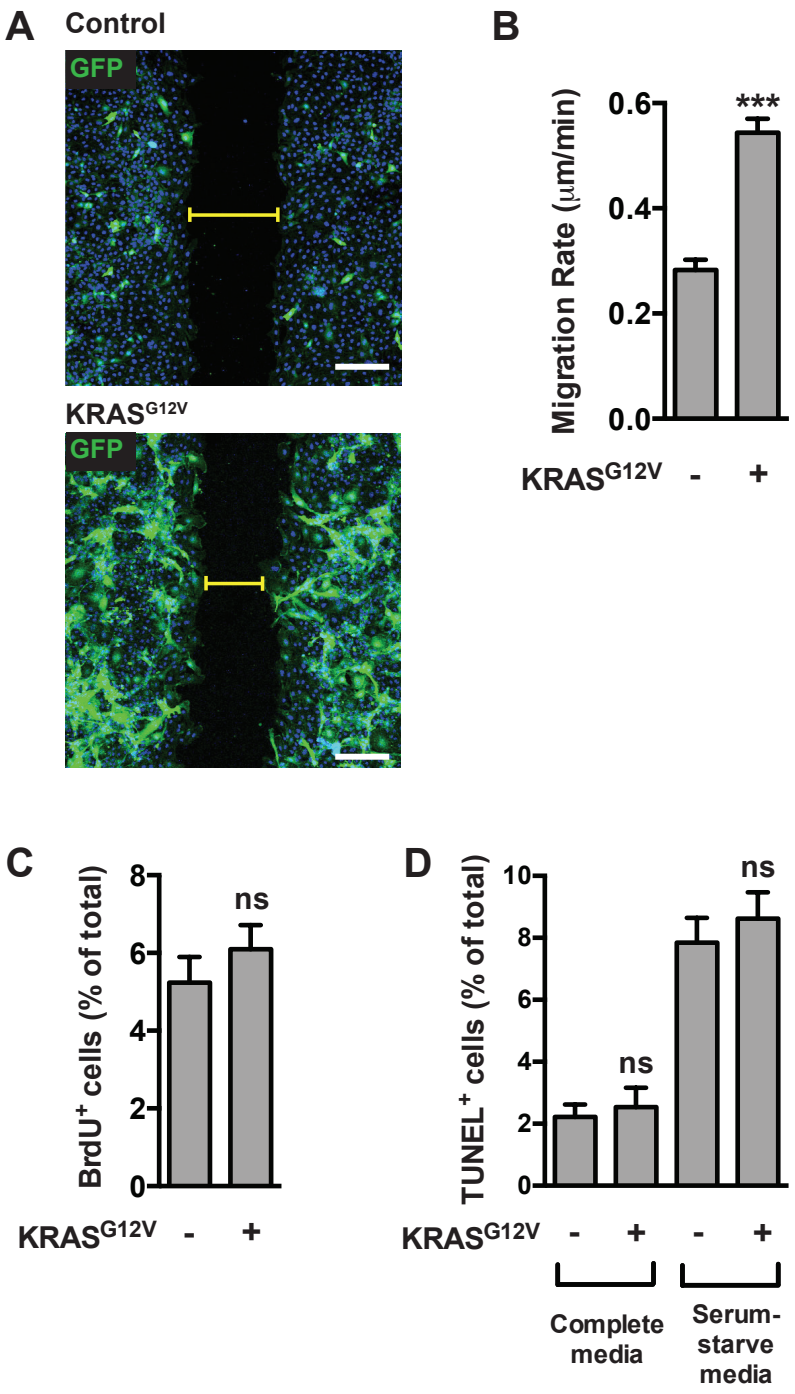


Figure S7: Expression of activated KRAS induces a pro-angiogenic phenotype in HUVECs. **(A)** Expression of KRAS4A^{G12V} resulted in altered cell morphology in an 8 h 'scratch' assay performed in the absence of growth factors or FBS. Scale bar = 200 μm. **(B)** Migration rate in serum-starve media was quantified by tracking multiple individual cells during time-lapse confocal microscopy in a 'scratch' assay. n=37 cells (from 14 videos) for control and n=41 (from 15 videos) for KRAS4A^{G12V} from a representative experiment. See Supplemental Video 1 for a representative time-lapse. **(C)** Cell proliferation in serum-starve medium was measured by BrdU incorporation. n = 23-24 non-overlapping fields of view from 3 independent experiments. **(D)** TUNEL staining in serum-starve or complete media revealed no differences in apoptosis in cells expressing KRAS4A^{G12V}. n = 17-18 fields of view from 3 independent experiments.

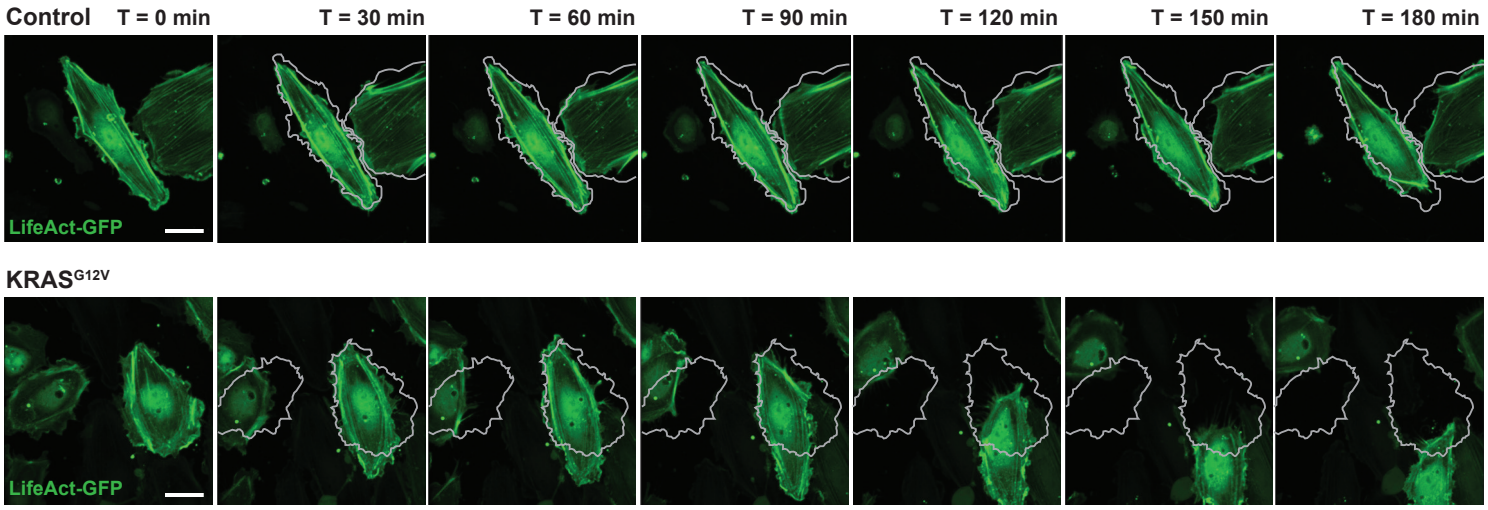


Figure S8: Expression of KRAS^{G12V} enhances migratory behaviour in a human arterial endothelial cell line (Telo-HAEC). Actin dynamics (LifeAct-GFP) were altered upon expression of KRAS4A^{G12V} in the absence of a migratory cue. Actin turnover was more rapid and cells displayed increased motility (an outline of the position of the cells at the beginning of the time-lapse is indicated). Static images from representative videos (Supplemental Video 3) are shown. Scale bar = 20 μm.