Heterozygous transcriptional and nonsense decay signatures in blood outgrowth endothelial cells from patients with hereditary haemorrhagic telangiectasia

Maria E Bernabeu-Herrero[1], Dilip Patel[1], Adrianna Bielowka[1], Sindu Srikaran[2], Patricia Chaves Guerrero[1], Fatima S Govani[1], Inês G Mollet[1], Michela Noseda[1], Micheala A. Aldred[3], Claire L Shovlin[1, 4]

[1]National Heart and Lung Institute, Imperial College London, UK; [2] London North West University Healthcare NHS Trust, London, UK; [3] Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, USA; [4]Respiratory and Specialist Medicine, Imperial College Healthcare NHS Trust, London, UK

Running head: RNASeq of HHT BOECs with heterozygous nonsense variants

Corresponding authors:

Claire L. Shovlin PhD FRCP, National Heart and Lung Institute, Imperial Centre for Translational and Experimental Medicine, Imperial College London, Hammersmith Campus, Du Cane Road, London W12 0NN, UK. Email c.shovlin@imperial.ac.uk

Micheala A Aldred PhD, Division of Pulmonary, Critical Care, Sleep and Occupational Medicine, Department of Medicine, 980W Walnut St, Indiana University School of Medicine, Indianapolis, IN, USA, Email maaldred@iu.edu

Key Words: expression QTL; frameshift; Gini coefficient; GnomAD; Genotype Tissue Expression (GTEx) Project; iron; nonsense mediated decay; ND index; random dataset clustering; premature termination codon (PTC); pulse chase; RNASeq; single cell qRT-PCR

ABSTRACT

In order to identify cellular phenotypes resulting from nonsense (gain of stop/premature termination codon) variants, we devised a framework of analytic methods that minimised confounder contributions, and applied to blood outgrowth endothelial cells (BOECs) derived from controls and patients with heterozygous nonsense variants in *ACVRL1*, *ENG* or *SMAD4* causing hereditary haemorrhagic telangiectasia (HHT). Following validation of 48 pre-selected genes by single cell qRT-PCR, discovery RNASeq ranked HHT-differential alignments of 16,807 Ensembl transcripts. Consistent gene ontology (GO) processes enriched compared to randomly-selected gene lists included bone morphogenetic protein, transforming growth factor-β and angiogenesis GO processes already implicated in HHT, further validating methodologies. Additional terms/genes including for endoplasmic reticulum stress could be attributed to a generic process of inefficient nonsense mediated decay (NMD). NMD efficiency ranged from 78-92% (mean 87%) in different BOEC cultures, with misprocessed mutant protein production confirmed by pulse chase experiments. Genes in HHT-specific and generic nonsense decay (ND) lists displayed differing expression profiles in normal endothelial cells exposed to an additional stress of exogenous 10μmol/L iron which acutely upregulates multiple mRNAs: Despite differing donors and endothelial cell types, >50% of iron-induced variability could be explained by the magnitude of transcript downregulation in HHT BOECs with less efficient NMD. The Genotype Tissue Expression (GTEx) Project indicated ND list genes were usually most highly expressed in non-endothelial tissues. However, across 5 major tissues, although 18/486 nonsense and frameshift variants in highly expressed genes were captured in GTEx, none were sufficiently prevalent to obtain genome-wide significant p values for expression quantitative trait loci (GnomAD allele frequencies <0.0005). In conclusion, RNASeq analytics of rare genotype-selected, patient-derived endothelial cells facilitated identification of natural disease-specific and more generic transcriptional signatures. Future studies should evaluate wider relevance and whether injury from external agents is augmented in cells with already high burdens of defective protein production.

INTRODUCTION

Amongst tens of thousands of changes in cells resulting from genomic or environmental stimuli, it can be difficult to disentangle damaging responses that would be better corrected, from necessary cellular adaptations, or chance observations. Both mechanistic understanding, and efficacious preclinical testing of potentially new therapeutic agents require robust laboratory assays, particularly for rare diseases, and genome-personalised medicine when sample size is limiting. The current decade benefits from global 'omics approaches that provide prior evidence in normal states,[1] and avoid over-targeted experimental readouts that can miss unanticipated adaptations employed by cells in order to survive, including therapy-nullifying responses.[2] Evaluation platforms now include patient-derived cells for bulk, single cell, and organ-on-a-CHIP analyses.

Our goals were to deliver robust methodologies for pre-clinical studies in patient-derived blood outgrowth endothelial cells (BOECs), and to test the extent to which transcripts harbouring premature termination codons (PTCs) undergo nonsense mediated decay (NMD)[3,4] in BOECs. We focussed on patients with the vascular dysplasia hereditary hemorrhagic telangiectasia (HHT) where new treatments are required to further reduce morbidity from hemorrhage caused by abnormal blood vessels.[5-14] HHT is inherited as an autosomal dominant trait, and typically caused by heterozygous, loss-of-function variants in *ACVRL1, ENG* and *SMAD4*.[15-17] As recently reviewed,[16,17] these genes encode endothelial-cell expressed proteins that transmit or modify signalling by the BMP/TGF-β superfamily, with HHT causality confirmed by null, heterozygous and endothelial-specific disease gene models.[18-22] Signalling defects are usually studied in conditional homozygous null models which generate pronounced and reproducible phenotypes in the context of trauma[19] or angiogenesis,[23,24] and have provided important insights into HHT pathophysiology: *ACVRL1, ENG* and *SMAD4* deficiency results in reduced endothelial integrity, red cell extravasation and increased endothelial proliferation following ischemia-induced angiogenesis.[20,22-24] Modified BMP/TGF-β signalling by lower ALK5 and ALK1 receptor expression/signalling has been interpreted as an adaptive response that enabled more *ENG* deficient cells to survive.[24] Other potentially relevant BMP/TGF-β independent roles of endoglin and ALK1

were summarised recently[16] and include actin cytoskeleton organization, modulation of integrin-dependent endothelium-leukocyte adhesion mediated by CXCL12, and regulation of endothelial nitric oxide synthase (eNOS) activation, with *ACVRL1*$^{+/-}$ and *ENG*$^{+/-}$ cells generating more eNOS-derived superoxide and reactive oxygen species.[25-27]

In HHT, early evidence demonstrated that the wildtype allele is still expressed in diseased tissues *in vivo*,[28,29] however, a relevant phenotype in heterozygous HHT cells has been described as "elusive". We hypothesised that the heterozygous HHT state would also display differences that would be important to complement the null models, and that these would be definable in modest sample size numbers. Blood outgrowth endothelial cells (BOECs) are an established platform for examining endothelial cells in a variety of disease states including pulmonary arterial hypertension which can also be caused by null alleles in the *ACVRL1* gene,[30] and HHT.[31] Here we report that employing stringent steps to minimise other biological and methodological variation, enabled identification of HHT heterozygous BOEC transcriptomic signatures, and a new transcriptional index related to the efficiency of nonsense mediated decay in specific cultures. Together the findings advance pathogenic understanding and preclinical testing opportunities.

METHODS *(Additional information in Supplemental Methods).*

Endothelial cell methods

This research was approved by national ethics committees, and all human participants provided written informed consent. Primary studies were performed in blood outgrowth endothelial cells (BOECs) established from healthy volunteers and pre-genotyped HHT patients. BOEC methods were previously optimised for similar studies in patient with pulmonary arterial hypertension.[30] Comparator assays were performed in normal human umbilical vein endothelial cells (HUVEC), and human dermal microvascular endothelial cells (HDMEC) as previously described.[30,32-34] All analyses were performed

in confluent passage 3 vials.

Following confirmation of endothelial status by bulk flow cytometry (Supplemental Figure 1), for RNASeq library preparations, RNA was extracted from 16 separate cultures of confluent BOECs from 4 separate donors, cultured in the presence and absence of BMP9 at 10ng/ml for 1 hr. Library preparations, alignments to *Homo sapiens* GRCh38, and initial variant calls[35] were performed by Genewiz (Leipzig, Germany).

To facilitate quantitative identification of subtle readouts, novel housekeeping genes were sought from genes with low Gini Coefficients (GC) within multiple cell lines.[36,37] These included *RBM45* (catalogue number 4351372, assay number HS00912280_M1), and *SF3B2* (catalogue number 4331182, assay number HS00199190_M1). These were evaluated using bulk quantitative reverse transcriptase PCR (qRT-PCR) as described[33] using pre-designed TaqMan qRT-PCR assays (Applied Biosystems, Thermo Fisher, Waltham, MA, Supplemental Table 1).

To provide numeric validations of a larger subset of genes, single cell qRT-PCR was performed as detailed further in the Data Supplement:[38,39] For each culture, 40 viable (DRAQ7 negative) single-cells were sorted directly into 96-well-plates containing pre-amplification mix for 48 genes selected as endothelial gene markers as validation for endothelial differentiation, likely relevance to the HHT cellular phenotype, or important negative controls. Specific target amplification was performed using CellDirect One-Step qRT-PCR Kit (Invitrogen, Thermo Fisher, Waltham, MA) as described.[38,39] The 48 genes examined by sc qRT-PCR were normalized to *UBC* expression, analysed, and plotted using an R script developed in house.

For further endothelial validations, we compared RNASeq findings in the HHT patient and control-derived BOECs to earlier RNASeq studies in normal primary HDMEC.[33] For the current manuscript,

comparisons focussed on alignment changes after 1 hour treatment with TGF-β1, BMP9 or 10μmol/L ferric citrate, as described.[33]

RNA Sequencing Analyses

Genewiz analytic pipelines aligned sequenced reads to *Homo sapiens* GRCh38 using STAR aligner v2.5.2b, counted unique gene reads that fell within exon regions using Subread package v1.5.2, and normalised to total alignment counts per library (Genewiz, Leipzig, Germany). Genewiz also assessed sample-to-sample similarities by Euclidean distance and principal component analyses (PCA).

In-house, the intra-assay coefficient of variation (CV) was calculated for replicate pairs as *100 x standard deviation (SD)/mean*, and for each transcript, alignment pairs were categorised by whether they did or did not meet a CV≤ 10% (CV10). Associations between RNASeq low Gini[36,37]-normalised alignments with donor age, sex, experimental variability, and HHT disease status were examined by Spearman rank correlation, and multivariate linear regression (STATA IC v 15.0., Statacorp, College Station, TX).

Quantitative RNASeq validations were performed in endothelial cells from different HHT and control donors, using bulk qRT-PCR, and single cell qRT-PCR.[38,39] The 48 genes examined by scPCR were categorised by quartiles of ranking in control BOECs (*number of expressing single BOECs*relative expression*) and compared to alignment values in RNASeq.

To focus on the biological discovery question 'how do HHT BOECs differ to control BOECs?', each gene was ranked by differential alignments in untreated HHT and control BOECs, using the maximum alignment per pair. The *x* axis of a scatter plot plotted the log of alignment fold differences described further in the Data Supplement:

$$Log\left(\frac{Lowest\ RBM45\_normalised\ ENG, ACVRL\ or\ SMAD4\ \ BOEC\ maximum\ alignment}{Highest\ RBM45\_normalised\ control\ BOEC\ alignment}\right)$$

The *y* axis plotted a simple measure of statistical confidence[40]:

$$-Log\left(\frac{1}{alignment\ standard\ deviation\ across\ all\ untreated\ BOECs}\right)$$

The product was squared, and the square root used to generate a single rank score:

$$\sqrt{\left\{\text{Log}\left(\frac{Lowest\ RBM45\_normalised\ HHT\ BOEC\ maximum\ alignment}{Highest\ RBM45\_normalised\ control\ BOEC\ alignment}\right)x\ -Log\left(\frac{1}{standard\ deviation}\right)\right\}^2}$$

Quantitative RNASeq validations of these ranking methods were performed using the HHT differential expression by scPCR of 48 genes which were categorised by no change ($p \geq 0.05$); increased expression ($p < 0.05$) and decreased expression ($p < 0.05$).

Nonsense mediated decay

BOECs undergoing RNASeq had been established from HHT patients heterozygous for nonsense ("stop gain") pathogenic[15,41,42] DNA variants in *ENG*, *ACVRL1* or *SMAD4*. Binary sequence alignment (bam) and vcf files were analysed in Galaxy Version 2.4.1[43] and uploaded to the UCSC Genome Browser,[44,45] and Integrated Genome Browser (IGV) 9.1.8[46] for further analyses. Percentage loss of the nonsense allele in *ACVRL1*$^{+/-}$, *ENG*$^{+/-}$ and *SMAD4*$^{+/-}$ BOECs was calculated assuming that without nonsense mediated decay (NMD)[3], there would have been an approximately equal number of alignments to wildtype and nonsense alleles in each HHT BOEC library. BOECs were categorised according to this % loss of the nonsense pathogenic variant to generate an "ND" index of NMD inefficiency in each HHT BOEC dataset. This was validated by comparison to ratios of common single nucleotide variant (SNV) alignments within the untranslated regions of *ACVRL1*, *ENG* and *SMAD4* called by Genewiz using

VarScan2 set to standard thresholds.

To validate persistence of RNA transcripts and protein synthesis, metabolic labelling was performed in BOEC protein extracts using ENG mAb Sn6H (Dako Denmark A/S, Denmark). Modifying methods of Pece et al,[47] a 1 hour "pulse" with 50μCi/ml of ^{35}S methionine, preceded a 0, 1, 2 or 3.5hr "chase" in media containing unlabelled methionine, immunoprecipitation and SDS-PAGE gel fractionation.

Gene ontology enrichment analysis

Gene Ontology[48] (GO) enrichment analysis was performed by Functional Annotation Clustering using the Database for Annotation, Visualization and Integrated Discovery (DAVID)[49,50] v6.8. This was performed initially for the top 100, 200, 300, 400, 500 750, 1,000, 1,250, 1,500 and 2,000 differentially expressed genes ranked as above, and subsequently for 10 random sets of 1000 genes derived from the complete gene dataset,[33] and genes where the ND index was significantly associated with *RBM4*5-normalised alignments in *ACVRL1*$^{+/-}$ and *SMAD4*$^{+/-}$ BOECs. GO terms meeting cluster enrichment of ≥1.2 fold and GO term $p<0.05$ were compared between experimentally-ranked, randomly-generated and ND index-flagged genes. Data were visualised using GraphPad Prism 9 (GraphPad Software, San Diego, CA), and STATA IC v 15.0 (Statacorp, College Station, TX).

Broader tissue and variant examinations

To examine if findings may be of broader relevance, gene expression was examined across RNASeq data from the 54 tissues representing 17,382 samples, 948 donors (V8, Aug 2019).[1,51] To explore whether there was statistical evidence that other nonsense variants were associated with modified expression of genes identified in the current study, we examined expression quantitative trait loci (eQTLs) within GTEx,[1,51] focussing on all nonsense and frameshift variants listed in GnomAD v3.1.2[52] that were present in the 100 most highly expressed transcripts in at least two of 5 tissues of disease-specific relevance (colon, heart, lung, liver, and skeletal muscle). The tissues were selected to span different categories observed for eQTLs, allele specific expression (aseQTLs) and transQTLs in the full

GTEX databases.[1,51] Following cross check on the UCSC Genome Browser,[44,45] 11 common variants (MAF >0.0005) that were only present in minor alternate transcripts (*HLA-B, RPL29)*, or had been misassigned to the categories (*TPM2)*, were removed.

Data Sharing Statement

Non-sensitive data underlying this article are available at 10.5281/zenodo.5201823 and can be used under the Creative Commons Attribution licence. Primary sequence data used in this research was collected subject to the informed consent of the participants. Access to these data will only be granted in line with that consent, subject to approval by the project ethics board and under a formal Data Sharing Agreement.

RESULTS

Endothelial cell phenotypes

BOECs were confirmed as $CD31^+$, $CD44^+$ $CD105^+$, $CD90^-$ by bulk flow cytometric evaluations, and displayed consistently high expression of a panel of endothelial marker genes (Supplemental Figure 1). From each of the 16 BOEC cultures, RNASeq generated between 34,400,000 and 45,100,000 (median 39,800,000) alignments to human genome build GRCh38. A median of 35,400,000 alignments per sample (88.0%) mapped uniquely to GRCh38, representing alignments to 16,807 Ensembl transcripts. Mean alignments across the 16 samples corresponded better with endothelial expressed genes from our earlier data in primary human dermal microvascular endothelial cells (HDMEC Figure 1Ai) than pulmonary artery endothelial cells (HPAEC, Figure 1Aii).

Two genes with low Gini coefficients of variation[36,37] were defined as improved housekeeper genes compared to those previously used for reference (Figure 1B). BOEC *RBM45* alignments were similar to the lowest quartile of endothelial transcripts (Figure 1Ci), including many HHT target genes examined in wider studies (data not shown). *SF3B2* alignments were similar to endothelial-expressed vascular malformation disease genes including for HHT (Figure 1Cii).

RNASeq detects the 3 HHT genotypes

VarScan2 did not detect the nonsense substitution in HHT BOECs. Instead, BOECs were confirmed as retaining the source HHT genotype by examining raw alignments to the nucleotides corresponding to the HHT donors' variants: *ACVRL1* exon 8 (c.1171G>T, p.Glu391X), *ENG* exon 3 (c.277C>T, p.Arg93X), and *SMAD4* exon 9 (c.1096C>T, p.Gln366X, (Supplemental Figure 2). In keeping with expectation from nonsense mediated decay[3,4] and earlier studies in HHT patient lymphoblastoid lines,[53] none of the nonsense variants (at chr12:51,916,157, chr9:127,829,769 or chr18:51065562) were present at a 1:1 ratio as would otherwise be expected for germline heterozygous alleles (Figure 1D). Instead, the allele alignment ratios implied substantial loss of the mutant allele (mean 89%, range 78-92% loss), and this was reflected in reduced total alignments to the respective gene (Figure 1E). Micro RNA alignments were also in keeping with published data of lower miR-10a and miR-370 in *ENG*$^{+/-}$ patients compared to *ACVRL1*$^{+/-}$ (Supplemental Figure 3).

Further confirmation was obtained in *ACVRL1* BOECs that by chance, were from a donor heterozygous for a common variant in the 3' untranslated region (UTR) where the wildtype allele was *in cis* with the exon 8 pathogenic nonsense substitution, enabling variant allele detection by VarScan2: In each of the *ACVRL1*$^{+/-}$ BOECs, loss of the exon 8 pathogenic nonsense substitution closely paralleled loss of the exon 10 proximal 3'UTR wildtype allele (r^2 0.92, p=0.011, Supplementary Figure 4).

Global RNASeq Analyses

Having confirmed that BOECs had maintained their endothelial cell phenotype and HHT genotype, global comparative evaluations were undertaken. Initial Genewiz principal component (Figure 2A) and

Euclidean distance (Figure 2B) analyses clustered *ENG*$^{+/-}$ and *SMAD4*$^{+/-}$ BOECs more closely than *ACVRL1*$^{+/-}$ BOECs. Since control BOECs were placed between HHT BOECs (Figure 2B), it was concluded that these methodologies were not identifying shared heterozygous HHT BOEC phenotypes.

Further, although the greatest similarity was seen between the two members of each pair of untreated BOECs (Figure 2B), and there was high sequencing fidelity (Supplemental Figures 5,6), only 5,014/ 16,807 (29.8%) Ensembl transcripts met an intra-assay coefficient of variation of 10% when normalised to total alignments across all libraries. Notably, there was marked variability in *RBM45* and *SF3B2* alignments across sample pairs, despite very low known biological variability,[36,37] confirmed biological replicates, and standardized culture conditions (Figure 2C).

To minimise this element of variability, for further analyses, all RNASeq alignments were normalised to *SFB2* or *RBM45*. The maximal value for each transcript in each donor replicate pair was used to minimise differential technical losses e.g. during sample/library generation (Supplemental Figure 7). We then examined whether an HHT disease contribution could be identified in the RNASeq datasets, once adjusted for donor age and sex. In all tested multivariate models of donor demographics and sample variability, the "HHT disease" term explained an additional 0.3% of the r-squared value (Supplemental Tables 2/3). This relatively modest value was ~30-fold higher than the equivalent proportion for sex, and ~3.5-fold higher than the equivalent proportion for age. It was concluded that there was likely to be a discernible difference in RNASeq alignments according to whether BOECs were control or HHT patient-derived.

Validations were then performed for *RBM45*-normalised RNASeq alignments. Across all 48 genes examined by sc qRT-PCR in a different HHT donor, sc qRT-PCR expression mirrored RNASeq alignments, as indicated by quartile rankings of the sc qRT-PCR expression in 20 control BOECs (Figure 3A). In total, 59% of the variability in the mean RNASeq alignments in control BOECs was accounted for by sc qRT-PCR ranking expression in controls ($p<0.0001$, Figure SA2). Single cell qRT-PCR results also mirrored those obtained by RNASeq when examining the genes that met Dunn's $p<0.05$ threshold for differential expression in 20 untreated HHT BOECs compared to 20 untreated control BOECs (Figure 3Bii, 3Biii): In RNASeq distribution plots, these 8 genes tended to be

positioned towards the top of the *y* axis representing significance (Figure 3Biii). The magnitude of change was mirrored by differential placements on the *x* axis particularly for *VWF* and *POSTN*, with the remaining 40 genes sited more centrally or absent (Figure 3Bii). Overall, sc qRT-PCR differential expressions, either combined (Mann Whitney p=0.0011), or when subcategorised by magnitude differences (Figure 3Bii) were associated with significantly greater RNASeq alignment differentials between HHT and control BOECs.

Discovery RNASeq

Biological processes differing between HHT and control BOECs were sought by identifying gene ontology (GO) terms generated by genes exhibiting the most differential alignments between HHT and control BOECs. The total number of clusters increased by clustering a larger slice of the top-ranked genes between 100-2,000 of the 16,807 genes (Figure 4A). GO clusters meeting significance also initially increased before plateauing at 1,000 genes (Figure 4A). Fewer GO processes were obtained by clustering the 10 randomly-derived datasets of 1,000 genes, compared to HHT differential-ranked genes (Mann Whitney p=0.0001, Figure 4B).

Overall, 377 of the top-ranked 1,000 genes were clustered, including 156 in GO terms meeting significance (Supplemental Table 4). However, 38 (19%) of GO terms that met significance thresholds in the top-ranked 100, 200, 300, 400, 500, 750 and 1,000 HHT/control differential genes, were also "significantly-enriched" in at least one random dataset, while a further 79 (40%) of GO terms were present in random gene-derived clusters not meeting significance (Figure 4C). Most GO terms reaching significance in random datasets referred to common cellular processes (Figure 4C). In contrast, the 61 GO terms relevant to endothelial cells that were clustered only in the HHT-enriched datasets referred to more specific processes (Figure 4D), with the highest cluster enrichment score inversely related to the lowest p value of a GO term in the cluster (Figure 4E, Supplemental Table 5).

The landscape of genes displaying differential alignments in HHT BOECs

In the remaining GO terms potentially relevant to an HHT transcriptome phenotype, the highest fold enrichments related to basement membrane, extracellular region/space, sequestration of TGF-β in the extracellular matrix, and microfibril binding (Figure 4E, Supplemental Table 4, Supplemental Figure 8). There were also significant enrichments for GO terms corresponding to BMP/TGF-β/SMAD responses; epithelial-mesenchymal transition (which is TGF-β dependent[54]); and endothelial proliferation, migration and vascular endothelial cell growth factor responses. Selected examples of genes with consistent changes across all HHT genotypes are illustrated in Figure 4F. This highlights lower alignments in HHT BOECs to *SMAD3* and *FOS* that encode proteins transmitting or regulate TGF-β signalling; to *BMP6* and *HES1* (induced by BMP9[55]); and to genes encoding extracellular matrix proteins now recognised to have roles in BMP crosslinking (*POSTN, LOXL2),* BMP antagonism (*SMOC2),* and BMP ligand presentation (*FBN1*).[56-65] We concluded that the data were in-keeping with expectations[24] for consequences of perturbed BMP/TGF-β signalling resulting from heterozygous loss of ENG, ACVRL1 or SMAD4 function.

Also illustrated in Figure 4F are examples of HHT differential alignments to genes encoding metallopeptidases (*ADAMTS12*, *MMP24*); components of the pentose phosphate shunt (*TKT*, *G6PD),* and nuclear encoded mitochondrial tetrahydrofolate reductases ('folic acid metabolic process' GO terms, *ALDH1L2*, *MTHFD2*). Additional enriched GO-terms spanned a number of further biological and molecular functions including RNA polymerase II and primary microRNA (pri-miRNA) transcription; NMDA/AMPA glutamate receptor clustering, and responses to endoplasmic reticulum stress (Figure 4E, Supplemental Table 4, Supplemental Figure 8).

Pulse chase highlights transient retention of immature protein from nonsense alleles

Enrichment of genes in the GO term responses for endoplasmic reticulum stress (GO:1990440) overlapping with the unfolded protein response GO term (GO:0036499) was unexpected, but recognised to be potentially relevant for the HHT BOECs that harboured heterozygous nonsense alleles in highly expressed endothelial genes (Figure 1Cii).

As noted in Figure 1, nonsense mediated decay was not 100% efficient, with nonsense allele alignment loss differing between the 3 genotypes (Figure 5A). To test if this required prior knowledge of the causal genotype, allele ratios for heterozygous common variants in the relevant 3'untranslated region (UTR) were examined. (Figure 5B). These also deviated from expected 1:1 ratios, supporting nonsense mediated decay (NMD) of mutant transcripts but indicated that mutant transcripts could escape NMD.

With evidence this was most marked in the *ENG* BOECs (Figure 5A), pulse-chase experiments tested if nonsense *ENG* alleles produced a defective protein. Commencing the chase when only newly-synthesised, immature endoglin protein was present (Figure 5C), mature protein became the more pronounced species within 1hr in control (Figure 5Ci,ii), *SMAD4*$^{+/-}$ (Figure 5Ciii), and *ACVRL1*$^{+/-}$ BOECs (Figure 5Civ). Mature endoglin protein was the only species visible by 2hr in these BOECs (Figure 5Ci-iv). In contrast, for BOECs heterozygous for *ENG* c.1306C>T, (p.Gln436X) or *ENG* c.277C>G, (p.Arg93X), immature protein remained the more pronounced species at 1hr, with traces still present at 2hr (Figure 5Cv,vi). We concluded that partially decayed nonsense alleles were generating mutant protein and aberrant protein maturation kinetics.

Endoplasmic reticulum stress genes

The HHT BOECs differed in the efficiency of nonsense mediated decay as judged by persistence of the nonsense allele (Figure 1, Figure 5A). We therefore examined if the genes enriched in the GO terms for cellular responses in relation to endoplasmic reticulum stress (GO:1990440: *ATF3, ATF4, HSPA5* and *XBP1*) were similarly changed across the 3 HHT genotypes. As shown in Figure 6A, the patterns for all 4 genes were highest in controls, with a 40.4-57.6% (mean 56.3%) reduction across all HHT genotypes. The lowest values were always seen in the *ENG*$^{+/-}$ BOECs, the cultures with the least efficient nonsense mediated decay as judged by persistence of the nonsense allele (Figure 1, Figure 5A).

Transcriptional associations with inefficient nonsense mediated decay

To test whether interculture differences in nonsense mediated decay efficiency were associated with any further differential gene alignments, the pathogenic nonsense allele loss in *ACVRL1*$^{+/-}$, *ENG*$^{+/-}$ or *SMAD4*$^{+/-}$ BOECs was expressed as an "ND" index for each individual BOEC culture, and compared to differential alignments in the same BOECs across a wider group of genes.

Fifty-four genes that were in significant GO processes for differential alignments in HHT BOECs also displayed different alignments that paralleled the efficiency of NMD in the particular BOEC line, as quantified by the ND index. We confirmed that the relationships were seen across the 8 *ACVRL1* and *SMAD4* BOECs, in addition to across the 12 HHT BOECs adjusting for the *ENG* genotype with its lower NMD efficiency (Figure 6B, Supplemental Figure 9), and termed these the endothelial "ND List" genes (Figure 7A).

Distinguishing the HHT and ND gene lists

Genes in the HHT and ND lists displayed different patterns of alignment. When plotted on the RNASeq distribution graph, the 54 ND-List genes predominantly distributed to the extreme left, representing lower expression with the most extreme named in Figure 6C. This was confirmed both limiting to the *ACVRL1* and *SMAD4* datasets (Figure 6D), and examining all 12 BOEC cultures (data not shown).

For GO terms originally identified by HHT differential expression rankings, the percentage of genes that were in the endothelial ND list ranged from 0-75% (median 22%). The 17 "HHT-significant" GO terms that did not contain more than 1 gene in the ND list were also the most "HHT-expected," including 7 for BMP/TGF-β/SMAD pathway signalling; 4 for vascular endothelial growth factor responses, and two for endothelial cell migration. In contrast, the *GO* term originally identified as "HHT-significant" that contained the highest number of genes in the ND list was GO:1990440 (response to endoplasmic reticulum stress genes), where 3 of 4 genes were in the ND list (*ASNS*, *ATF3*, *XBP2*, Figure 6A). This led us to test whether there was evidence that the ND list genes may play further roles in cellular stress phenotypes.

First we tested alignments to genes in the HHT green list in HDMEC following 1hr TGF-β1 or BMP9 treatment, and showed that they did not differ in these short term settings of stimuli relevant to HHT (Figure 6E).

Next, we focussed on iron treatment which is an injurious stimulus particularly relevant to HHT patients,[5-9] that we had also previously examined using RNASeq in normal endothelial cells (HDMECs), demonstrating upregulation of multiple mRNAs within 1 hour.[33] We tested if iron- induced differential alignments in normal primary HDMECs differed according to whether the gene was in the HHT list (Figure 6E), ND list (Figure 6Fi) or a random gene list (Figure 6Fi), and demonstrated that the ND list genes displayed higher alignments in HDMEC following 1hr treatment with 10μM iron (Figure 6Ei). The higher levels of expression became more obvious restricting to genes meeting HDMEC thresholds for post-iron alignment differences of $p<0.05$, $p<0.10$ or $p<0.20$ (Figure 6Eii). Post-iron HDMEC alignments were normally distributed (Figure 6Eiii), and when tested by linear regression, the magnitude of the higher alignments in iron-treated HDMEC was inversely related to the magnitude of alignments in HHT BOECs compared to control BOECs (Figure 6Eiv). Again, this became more pronounced restricting to genes meeting the threshold significance of $p<0.20$ (Figure 6Ev). Despite the different donors and different endothelial cell types, 25.4% of variability in iron-induced changes of all ND list genes ($p=0.001$), and 50.6% of variability across 13 genes meeting $p<0.20$ ($p=0.006$), was explained by the magnitude of differential alignments in HHT BOECs compared to controls. In other words, in the setting of inefficient NMD, and in proportion to the inefficiency of NMD, the HHT BOECs displayed reduced expression of genes that would normally be up-regulated following a short injury from 10μmol iron.

Extending the relevance to further tissues

To examine whether the ND List genes generated in endothelial cells harbouring highly expressed nonsense-containing transcripts may have broader relevance to additional tissues and disease states, we first tested whether the 54 genes were expressed exclusively in endothelial cells. Gene expression in 54 tissues from GTEx RNASeq data from 17,382 samples, 948 donors (V8, Aug 2019)[1,51] indicated that

most genes were more broadly expressed: 10/54 (18.5%) of the ND list genes were expressed most highly in blood vessels (aorta or coronary arteries), but none were exclusively expressed in endothelial cells. Furthermore, seven were expressed most highly in cultured cells (fibroblasts or EBV-transformed lymphoblastoid cell lines), and other sites of maximal expression included the adrenal gland, brain, gastrointestinal tract, lung, ovary, pancreas and thyroid (Figure 7B).

We sought statistical evidence that other nonsense variants could be associated with modified expression of genes identified in the current study. In the 93 genes comprising the "top 100" of two or more selected GTEx tissues (excluding *HLA-B*), GnomAD 3.1.2[52] listed 486 frameshift and nonsense (stop-gain) alleles (Figure 7D). By chance, 19 were present in at least one of the 948 GTEx donors, but all were very rare with GnomAD 3.2.1 minor allele frequencies (MAF) between 6.6 E^{-6} and 0.0026, and no significant eQTLs were found.

DISCUSSION

Using patient-derived blood outgrowth endothelial cells, we increased RNASeq discovery yields by considering unpredictable generic cellular processes that have the potential to confound stringently designed and executed experiments. Most of the inter-BOEC variability in transcript expressions was not HHT specific. In the first arm of the work, we used stringent methods to identify consistent transcriptome signatures with GO term responses fitting predicted elements of the HHT phenotype, in keeping with knowledge of BMP/TGF-β roles. Crucially, the methods and data also identified a new index relevant to generic aspects of cellular perturbations in the settings of environmental stressors and endogenous protein degradation burdens.

What would normally be considered a weakness of a study (small sample numbers in the initial discovery phase) was in fact a strength in terms of optimising analytic approaches. In experimental

systems it is common to overcome the noise of experimental variability by using increased samples sizes, but this luxury is not so feasible for rare diseases, particularly for personalised medicine with individual unique genomic repertoires. By restricting donor selection and standardising experimental culture conditions to minimise genomic, environmental and tissue expression differences, the current study enhanced the proportion of variability due to biological differences in transcript expression within the BOECs. Our primary goal was to capture endothelial cell phenotypes that would be suitable for use in high-throughput assay screens, noting the GO terms and genes were not meant to be exhaustive. Having identified the new ND list, it became imperative to offer the data early in order to facilitate rapid recognition and development of what appears to be an important and previously unsuspected consequence for frameshift and nonsense variants that in aggregation, are commonly present in highly expressed genes.

The study methodologies provided discovery methods to identify uncharacterised loss of function alleles by pulse chase and particularly 3'UTR common SNV analyses, noting that the latter need to account for alternate splicing and polyadenylations sites that may not be evident in current data repositories.

The methods also identify two cautionary sets of RNASeq genes: The first are designated as a "black/red list" for BOECs where GO terms were significantly enriched in clusters from randomly-selected transcripts (Figure 4Ci, Supplemental Table 5A). These included general cellular processes such as transcription, translation and phosphorylation, and additionally adherens/cellular junction terms that might have been considered to be relevant to HHT. The amber list (Figure 4Cii, Supplemental Table 5) consisted of GO terms where the majority of genes were clustered in random data sets without meeting enrichment thresholds. Excluding amber-list GO terms may have been overly stringent, for instance the amber list designation of terms for the actin cytoskeleton and integrin-mediated signalling when endoglin is known to be involved in these processes,[15] and nuclear transcribed mRNA catabolic

processes when NMD is operational in HHT patient-derived cells. We considered this preferable to retention of GO terms more likely related to chance. Further, the fact that disease and variant-relevant terms were identified in the excluded amber list datasets further enhanced confidence in the findings of the final selected terms.

We then focused on two lists of specific and more general value. The third set of genes (Figure 4D, Figure 4E, Supplemental Table 6) represents a green list priority for an HHT BOEC phenotype. As for all observational studies, it is not appropriate to infer disease-causality to any differences between HHT and control samples, although the gene lists included the expected genes relevant to BMP/TGF-β signalling, endothelial migration/proliferation and responses to vascular endothelial cell growth factor (VEGF). Genes within the basement membrane and extracellular matrix GO terms were the most differentially aligned in HHT BOECs, and theoretically attractive as HHT cellular biomarkers: Basement membrane genes tether many BMPs and other TGF-β ligands; have angiogenic roles, and disruption can result in haemorrhage.[57,58] Furthermore, experimental studies link *ENG*, *ACVRL1* and *SMAD4* deficiency to reductions in endothelial integrity,[21,23,24] relevant to the core clinical HHT phenotype of hemorrhage.[5-9] Amongst the top 50 HHT-differentially aligned genes, many encode extracellular proteins that regulate TGF-β/BMP signalling including periostin[56] (*POSTN)*, fibulin-6/hemicentin 1[59,60] (*HMCN1)*, lysyl oxidase-like 2[61] (*LOXL2*), SPARC-related modular calcium binding protein 2[62] (*SMOC2)*, laminin receptor 1 (*RSPA*),[63] and fibrillin 1 (*FBN1*, Figure 4F, Supplementary Figure 7)[58] Likely relevance was further enhanced because BMP/TGF-β ligand release from extracellular pools is by matrix metallopeptidases,[64,65] a further GO term enriched in the HHT BOEC datasets (through 8 genes, most significantly *ADAMTS18*, and *MMP24*, Figure 4F, Supplementary Figure 7). Our initial expectation was that their modulation may be further cellular adaptive responses that enable more *ACVRL1*$^{+/-}$, *ENG*$^{+/-}$ and *SMAD4*$^{+/-}$ BOECs to survive, analogous to lower ALK5 and ALK1 receptor expression/signalling.[24]

However, we then demonstrated that a subset of the genes which were the most differentially aligned in HHT BOECs were potentially generically relevant to any cell with high expression of nonsense substitutions. We provided evidence indicating that restricting to the 12 HHT BOEC datasets, nonsense mediated decay inefficiency was associated with substantial changes in alignments to a subset of genes. NMD efficiency was lowest in BOECs heterozygous for an *ENG* exon 3 nonsense variant, and by pulse chase, appeared similar in a more distal variant. While we cannot rule out an *ENG* reason why these BOECs had less efficient NMD, data from ourselves (Figure 6A-C) and others indicate that in HHT patient-derived cells, mutant proteins destined for degradation can saturate endoplasmic reticulum, nonsense mediated decay, and other clearance mechanisms.[53,66-70] We then demonstrated that collectively, the genes in the ND list appear important for early stress responses in relation to iron, a reactive oxygen species donor.[33] Our data therefore suggest that failure to clear such proteins at a sufficiently high level may have deleterious cellular consequences that can be specifically tested: Endothelial cell injury from oral iron tablets can be demonstrated in healthy controls,[71] at least 1 in 20 HHT patients report hemorrhage following iron treatments,[71,72] and there are apparent positive feedback loops in a proportion of patients with HHT who require intravenous iron.[73] It has long been a puzzle why iron treatments are tolerated less well by the HHT population than reported in general medicine,[71-73] and given at least 67% of pathogenic variant alleles require degradation by nonsense mediated decay,[32] the current findings provide a plausible explanation. This is supported by the GO terms for the pentose phosphate shunt/pathway (e.g. via the rate-limiting enzyme of the PPP, thiamine-dependent enzyme encoded by *TKT*, and glucose-6-phosphate dehydrogenase (*G6PD*)), and folic acid metabolic process (e.g. via 10-formyltetrahydrofolate dehydrogenase (*ALDH1L2*), and bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase (*MTHFD2*, Figure 4F, Supplementary Figure 7)): These remind of the heightened requirements for reactive oxygen species (ROS) defences predicted for HHT genotypes due to reduced ENG expression (Figure 1A). This is because ENG and ALK1 regulate endothelial nitric oxide synthase (eNOS) activation, and $ENG^{+/-}$ and $ACVRL1^{+/-}$ cells generate more eNOS-derived superoxide and ROS.[25-27] In multiple cell types, including endothelial cells, the pentose phosphate pathway plays a key role in protecting cells against oxidative stress by

providing reduced glutathione to reduce ROS,[74,75]while folic acid recouples eNOS, improving endothelial function.[76,77] The findings therefore suggest that ROS-related phenotypes may be more pronounced in *ENG*$^{+/-}$ and *ACVRL1*$^{+/-}$ HHT: whether attendant consequences cause or augment the HHT phenotype requires further study, although it is already known that *ENG*, *ACVRL1*, and *SMAD4* deficiency results in reduced endothelial integrity and red cell extravasation.[21,23,24] We note that irrespective of mechanisms, in appropriately selected cells, differential secretion of extracellular proteins may be well-suited to high throughput screens for preclinical testing of potentially new therapeutic agents.

Beyond HHT, we consider that the current and prior studies provide sufficient evidence for impacts on specific cellular and tissue adaptive phenotypes to be further explored. Recent studies emphasise how commonly deleterious rare variants occur in individuals when considered in an aggregate across relevant genes.[78,79] In particular, whether inefficient NMD in the setting of persistent, highly expressed nonsense-derived proteins compromises cellular ability to respond to exogenous stresses will need to be the subject of further experimental studies based on preselected genotypes since in the non-selected GTEx tissues, MAFs were too low to obtain genome-wide significant p values for expression QTLs. Notably, the findings are in keeping with longstanding[80] and more recent[81-83]evidence that cellular resource-favourable adaptations of downregulation are more likely to be selected over lifetimes than upregulations which commonly characterise acute responses to maintain homeostasis, as seen for iron in the current and earlier studies.[33] The findings also highlight that for inherited diseases, cellular transcriptional profiles (and responses to exogenous stressors), are likely to differ if deleterious variants generate proteins that are present but require degradation in the endoplasmic reticulum (e.g. nonsense, frameshift, and the majority of splice site consensus variants[32,42]), as opposed to being absent due to deletion alleles or start codon loss, or appropriately positioned but non-functional due to active site substitutions or inframe indels.

In conclusion, we have provided systems and data that help select appropriate and robust readouts from RNASeq, both for scientific evaluations, and to guide therapeutic approaches. Although initially intended to address the clinical problems of patients with HHT, the finding of a group of iron injury-response genes that are downregulated in cells where nonsense mediated decay is inefficient, broadens the relevance of the findings and targets for future studies.

ACKNOWLEDGMENTS

The project received specific funding from The US Department of Defense Discovery Award W81XWH-16-1-0607 (PI Aldred, CoI Shovlin, *Stopping the Stops: A Novel Therapeutic Approach for Hemorrhage from Vascular Malformations*); The National Institute for Health Research Imperial BRC Institute for Translational Medicine and Therapeutics (PI Shovlin, CoI Aldred; *Nonsense read-through topical therapies for patients with inherited genetic diseases: Proof of concept studies in hereditary haemorrhagic telangiectasia*); and Imperial College Healthcare NHS Trust (PI Shovlin, CoI Noseda; *Towards NHS Laboratory functional assays of genomic variants of uncertain significance –defining shared monocyte/target endothelial signatures in inherited vasculopathies*). The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were from the V8, Aug 2019 release accessed in October and November 2021. Early HDMEC studies[32] were funded by the British Heart Foundation, the National Institute for Health Research Imperial Biomedical Research Centre, and the Averil MacDonald Memorial Fund. We thank the study participants for their willing cooperation in these studies, and Professor Stefan Marciniak for helpful discussions on ER stress. The views expressed are those of the authors, and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care

AUTHORSHIP CONTRIBUTIONS AND CONFLICT OF INTEREST DISCLOSURES

Initials	Contribution	Conflicts of Interest
MEBH	Derived and maintained blood outgrowth endothelial cells (BOECs), co-designed experimental approaches, contributed to data discussions, and reviewed and approved the final manuscript.	MBH has no competing financial interests
AB	Cultured BOECs and HUVECs, generated RNAs for RNASeq libraries; codesigned single cell experiment; contributed to optimisation of housekeeper genes; derived RNASeq coefficients of variation, contributed to data discussions, and reviewed and approved the final manuscript.	AB has no competing financial interests
DP	Derived and maintained blood outgrowth endothelial cells (BOECs); selected and performed housekeeper optimisations; generated RNAs for RNASeq libraries. Designed and performed pulse chase experiments, contributed to data discussions, and reviewed and approved the final manuscript	DP has no competing financial interests
SS	Contributed to project set up and patient recruitment. Contributed to data discussions, and reviewed and approved the final manuscript	SS has no competing financial interests
PCG	Designed and performed scPCR analyses. Reviewed and approved the final manuscript	PCG has no competing financial interests
FG	Designed and performed RNASeq library generation of HDMEC and HPAEC datasets	FG has no competing financial interests
IGM	Designed and performed analysis of HDMEC and HPAEC datasets	IGM has no competing financial interests
MN	Designed and performed scPCR analyses. Reviewed and approved the final manuscript	MN has no competing financial interests
MAA	MAA assisted in establishing BOEC isolation, provided critical review of manuscript drafts and approved the final version	MAA has no competing financial interests
CLS	CLS obtained ethics approvals, contributed to recruitments, codesigned single cell and RNASeq experiments; devised and performed BOEC primary data analyses; devised and performed the secondary HDMEC and GTEX analyses presented in the manuscript; generated the Figures, wrote the first manuscript draft, generated Supplementary Data, and revised and approved the final manuscript.	CLS has no competing financial interests.

EXCLUDED REVIEWERS

None

REFERENCES

1 GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020 Sep 11;369(6509):1318-1330

2 Myers SH, Ortega JA, Cavalli A. Synthetic Lethality through the Lens of Medicinal Chemistry. J Med Chem. 2020;63(23):14151-14183.

3 Kurosaki T, Popp MW, Maquat LE. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. Nat Rev Mol Cell Biol. 2019 Jul;20(7):406-420

4 Maquat LE. When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. RNA. 1995 Jul;1(5):453-65

5 Shovlin CL, Buscarini E, Sabbà C, et al. The European Rare Disease Network for HHT frameworks for management of hereditary haemorrhagic telangiectasia in general and speciality care. 2021 Nov 1;104370. doi: 10.1016/j.ejmg.2021.104370. Online ahead of print.

6 Faughnan ME, Mager JJ, Hetts SW, et al. Second International Guidelines for the Diagnosis and Management of Hereditary Hemorrhagic Telangiectasia. Ann Intern Med. 2020;173(12):989-1001

7 Finnamore H, Le Couteur J, Hickson M, Busbridge M, Whelan K, Shovlin CL. Hemorrhage-adjusted iron requirements, hematinics and hepcidin define hereditary hemorrhagic telangiectasia as a model of hemorrhagic iron deficiency. PLoS One. 2013;8(10):e76516.

8 Al-Samkari H, Kasthuri RS, Parambil JG, et al. An international, multicenter study of intravenous bevacizumab for bleeding in hereditary hemorrhagic telangiectasia: the InHIBIT-Bleed study. Haematologica. 2021; 106(8):2161-2169.

9 Al-Samkari H. Hereditary hemorrhagic telangiectasia: systemic therapies, guidelines, and an evolving standard of care. Blood. 2021;137(7):888-895.

10 Robert F, Desroches-Castan A, Bailly S, Dupuis-Girod S, Feige JJ. Future treatments for hereditary hemorrhagic telangiectasia. Orphanet J Rare Dis. 2020;15(1):4.

11 Eker OF, Boccardi E, Sure U, et al. European Reference Network for Rare Vascular Diseases (VASCERN) position statement on cerebral screening in adults and children with hereditary haemorrhagic telangiectasia (HHT). Orphanet J Rare Dis. 2020;15(1):165.

12 Shovlin CL, Condliffe R, Donaldson JW, Kiely DG, Wort SJ; British Thoracic Society. British Thoracic Society Clinical Statement on Pulmonary Arteriovenous Malformations. Thorax. 2017;72(12):1154-1163.

13 Revuz S, Decullier E, Ginon I, et al. Pulmonary hypertension subtypes associated with hereditary haemorrhagic telangiectasia: Haemodynamic profiles and survival probability. PLoS One. 2017;12(10):e0184227

14 Vorselaars V, Velthuis S, van Gent M, et al. Pulmonary Hypertension in a Large Cohort with Hereditary Hemorrhagic Telangiectasia. Respiration. 2017;94(3):242-250.

15 Shovlin CL, Simeoni I, Downes K, et al. Mutational and phenotypic characterization of hereditary hemorrhagic telangiectasia. Blood. 2020;136(17):1907-1918.

16 Bernabeu C, Bayrak-Toydemir P, McDonald J, Letarte M. Potential Second-Hits in Hereditary Hemorrhagic Telangiectasia. J Clin Med. 2020;9(11):3571.

17 Snodgrass RO, Chico TJA, Arthur HM. Hereditary Haemorrhagic Telangiectasia, an Inherited Vascular Disorder in Need of Improved Evidence-Based Pharmaceutical Interventions. Genes (Basel). 2021;12(2):174.

18 Bourdeau A, Dumont DJ, Letarte M. A murine model of hereditary hemorrhagic telangiectasia. J Clin Invest. 1999;104:1343–1351

19 Park SO, Wankhede M, Lee YJ, et al. Real-time imaging of de novo arteriovenous malformation in a mouse model of hereditary hemorrhagic telangiectasia. J Clin Invest. 2009;119(11):3487-96.

20 Jin Y, Muhl L, Burmakin M, Wang Y, et al. Endoglin prevents vascular malformation by regulating flow-induced cell migration and specification through VEGFR2 signalling. Nat. Cell Biol. 2017;19:639–652.

21 Ola R, Künzel SH, Zhang F, et al. SMAD4 Prevents Flow Induced Arteriovenous Malformations by Inhibiting Casein Kinase 2. Circulation. 2018;138:2379–2394.

22 Garrido-Martín EM, Blanco FJ, Roque M, et al. Vascular injury triggers Krüppel-like factor 6 mobilization and cooperation with specificity protein 1 to promote endothelial activation through upregulation of the activin receptor-like kinase 1 gene. Circ Res. 2013;112:113–127

23 Seghers L, de Vries MR, Pardali E, et al. Shear induced collateral artery growth modulated by endoglin but not by ALK1. J Cell Mol Med. 2012;16:2440–2450.

24 Lebrin F, Goumans MJ, Jonker L, et al. Endoglin promotes endothelial cell proliferation and TGF-beta/ALK1 signal transduction. EMBO J. 2004;23:4018–4028.

25 Toporsian M, Gros R, Kabir MG, et al. A role for endoglin in coupling eNOS activity and regulating vascular tone revealed in hereditary hemorrhagic telangiectasia. Circ Res. 2005;96:684–692.

26 Jerkic M, Sotov V, Letarte M. Oxidative stress contributes to endothelial dysfunction in mouse models of hereditary hemorrhagic telangiectasia. Oxid Med Cell Longev. 2012;2012:686972.

27 Jerkic M, Letarte M. Contribution of oxidative stress to endothelial dysfunction in hereditary hemorrhagic telangiectasia. Front Genet. 2015;6:34.

28 Bourdeau A, Cymerman U, Paquet ME, Meschino W, McKinnon WC, Guttmacher AE, Becker L, Letarte M. Endoglin expression is reduced in normal vessels but still detectable in arteriovenous malformations of patients with hereditary hemorrhagic telangiectasia type 1. Am J Pathol. 2000 Mar;156(3):911-23.

29 Matsubara S, Bourdeau A, terBrugge KG, Wallace C, Letarte M. Analysis of endoglin expression in normal brain tissue and in cerebral arteriovenous malformations. Stroke. 2000 Nov;31(11):2653-60

30 Drake KM, Dunmore BJ, McNelly LN, Morrell NW, Aldred MA. Correction of nonsense BMPR2 and SMAD9 mutations by ataluren in pulmonary arterial hypertension. Am J Respir Cell Mol Biol. 2013;49(3):403-9.

31 Fernandez-L A, Sanz-Rodriguez F, Zarrabeitia R, Pérez-Molino A, Hebbel RP, Nguyen J, Bernabéu C, Botella LM. Blood outgrowth endothelial cells from Hereditary Haemorrhagic

Telangiectasia patients reveal abnormalities compatible with vascular lesions. Cardiovasc Res. 2005 Nov 1;68(2):235-48

32 Govani FS, Giess A, Mollet IG, et al. Directional next-generation RNA sequencing and examination of premature termination codon mutations in endoglin/hereditary haemorrhagic telangiectasia. Mol Syndromol. 2013;4(4):184-96.

33 Mollet IG, Patel D, Govani FS, et al. Low Dose Iron Treatments Induce a DNA Damage Response in Human Endothelial Cells within Minutes. PLoS One. 2016;11(2):e0147990.

34 Shovlin CL, Angus G, Manning RA, et al. Endothelial cell processing and alternatively spliced transcripts of factor VIII: potential implications for coagulation cascades and pulmonary hypertension. PLoS One. 2010;5(2):e9154.

35 Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22(3):568-76.

36 Wright Muelas M, Mughal F, O'Hagan S, Day PJ, Kell DB. The role and robustness of the Gini coefficient as an unbiased tool for the selection of Gini genes for normalising expression profiling data. Sci Rep. 2019;9(1):17960.

37 O'Hagan S, Wright Muelas M, Day PJ, Lundberg E, Kell DB. GeneGini: Assessment via the Gini Coefficient of Reference "Housekeeping" Genes and Diverse Human Transporter Expression Profiles. Cell Syst. 2018;6(2):230-244.e1

38 Noseda M, Harada M, McSweeney S, et al. PDGFRα demarcates the cardiogenic clonogenic Sca1+ stem/progenitor cell in adult murine myocardium. Nat Commun. 2015;6:6930

39 Massaia A, Chaves P, Samari S, et al. Single Cell Gene Expression to Understand the Dynamic Architecture of the Heart. Front Cardiovasc Med. 2018;5:167.

40 Judd CM, McClelland GH, Ryan CS. Data Analysis. A Model Comparison Approach. Second Edn. New York : Routledge, New York; 2009.

41 Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-2

42 Davieson CD, Joyce KE, Sharma L, Shovlin CL. DNA variant classification–reconsidering 'allele rarity' and 'phenotype' criteria in ACMG/AMP guidelines. Eur J Med Genet 2021 Oct;64(10):104312

43 Afgan E, Baker D, Batut B, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update, Nucleic Acids Res. 2018:46;W1: W537–W544.

44 Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. Genome Res. 2002; 12(6):996-1006

45 Lee BT, Barber GP, Benet-Pagès A, et al. The UCSC Genome Browser database: 2022 update. Nucleic Acids Res. 2021 Oct 28:gkab959. doi: 10.1093/nar/gkab959. Epub ahead of print.1057.

46 Freese NH, Norris DC, Loraine AE. Integrated genome browser: visual analytics platform for genomics. Bioinformatics. 2016;32(14):2089-95.

47 Pece N, Vera S, Cymerman U, White RI Jr, Wrana JL, Letarte M. Mutant endoglin in hereditary hemorrhagic telangiectasia type 1 is transiently expressed intracellularly and is not a dominant negative. J Clin Invest. 1997;100(10):2568-79

48 Dolan M, Drabkin H, Hill DP, et al. Gene Ontology annotations and resources. Gene Ontology Consortium, Blake JA. Nucleic Acids Res. 2013;41(Database issue):D530-5.

49 Dennis G Jr, Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003;4(5):P3

50 Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 2009;4(1):44-57.

51 GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013 Jun;45(6):580-5.

52 Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020 May;581(7809):434-443

53 Shovlin CL, Hughes JM, Scott J, Seidman CE, Seidman JG. Characterization of endoglin and identification of novel mutations in hereditary hemorrhagic telangiectasia. Am J Hum Genet. 1997;61(1):68-79.

54 Batlle E, Massagué J. Transforming Growth Factor-β Signaling in Immunity and Cancer. Immunity. 2019 Apr 16;50(4):924-940.

55 Seong CH, Chiba N, Kusuyama J, Subhan Amir M, Eiraku N, Yamashita S, Ohnishi T, Nakamura N, Matsuguchi T. Bone morphogenetic protein 9 (BMP9) directly induces Notch effector molecule Hes1 through the SMAD signaling pathway in osteoblasts. FEBS Lett. 2021 Feb;595(3):389-403

56 Kii I. Periostin Functions as a Scaffold for Assembly of Extracellular Proteins. Adv Exp Med Biol. 2019;1132:23-32.

57 Pozzi A, Yurchenco PD, Iozzo RV. The nature and biology of basement membranes. Matrix Biol. 2017 Jan;57-58:1-11

58 Zimmermann LA, Correns A, Furlan AG, Spanou CES, Sengle G. Controlling BMP growth factor bioavailability: The extracellular matrix as multi skilled platform. Cell Signal. 2021 Sep;85:110071..

59 Lin MH, Pope BD 3rd, Sasaki T, Keeley DP, Sherwood DR, Miner JH. Mammalian hemicentin 1 is assembled into tracks in the extracellular matrix of multiple tissues. Dev Dyn. 2020 Jun;249(6):775-788

60 Chowdhury A, Hasselbach L, Echtermeyer F, Jyotsana N, Theilmeier G, Herzog C. Fibulin-6 regulates pro-fibrotic TGF-β responses in neonatal mouse ventricular cardiac fibroblasts. Sci Rep. 2017 Feb 17;7:42725. doi: 10.1038/srep42725

61 Schmelzer CEH, Heinz A, Troilo H, Lockhart-Cairns MP, Jowitt TA, Marchand MF, Bidault L, Bignon M, Hedtke T, Barret A, McConnell JC, Sherratt MJ, Germain S, Hulmes DJS, Baldock C, Muller L. Lysyl oxidase-like 2 (LOXL2)-mediated cross-linking of tropoelastin. FASEB J. 2019 Apr;33(4):5468-5481

62 Long F, Shi H, Li P, Guo S, Ma Y, Wei S, Li Y, Gao F, Gao S, Wang M, Duan R, Wang X, Yang K, Sun W, Li X, Li J, Liu Q. A SMOC2 variant inhibits BMP signaling by competitively binding to BMPR1B and causes growth plate defects. Bone. 2021 Jan;142:115686.

63 Dolez M, Nicolas JF, Hirsinger E. Laminins, via heparan sulfate proteoglycans, participate in zebrafish myotome morphogenesis by modulating the pattern of Bmp responsiveness. Development. 2011 Jan;138(1):97-106.

64 Shi M, Zhu J, Wang R, et al. Latent TGF-beta structure and activation. Nature. 2011;474:343-349.

65 Furlan AG, Spanou CES, Godwin ARF, et al. A new MMP-mediated prodomain cleavage mechanism to activate bone morphogenetic proteins from the extracellular matrix. FASEB J. 2021 Mar;35(3):e21353.

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

67 Cymerman U, Vera S, Pece-Barbara N, et al. Identification of hereditary hemorrhagic telangiectasia type 1 in newborns by protein expression and mutation analysis of endoglin. Pediatr Res. 2000;47(1):24-35.

68 Paquet ME, Pece-Barbara N, Vera S, et al. Analysis of several endoglin mutants reveals no endogenous mature or secreted protein capable of interfering with normal endoglin function. Hum Mol Genet. 2001;10(13):1347-57.

69 Ali BR, Ben-Rebeh I, John A, et al. Endoplasmic reticulum quality control is involved in the mechanism of endoglin-mediated hereditary haemorrhagic telangiectasia. PLoS One. 2011;6(10):e26206.

70 Hume AN, John A, Akawi NA, et al. Retention in the endoplasmic reticulum is the underlying mechanism of some hereditary haemorrhagic telangiectasia type 2 ALK1 missense mutations. Mol Cell Biochem. 2013;373(1-2):247-57.

71 Shovlin CL, Gilson C, Busbridge M, Patel D, Shi C, Dina R, Abdulla FN, Awan I. Can Iron Treatments Aggravate Epistaxis in Some Patients With Hereditary Hemorrhagic Telangiectasia? Laryngoscope. 2016 Nov;126(11):2468-2474

72 Shovlin CL, Patel T, Jackson JE. Embolisation of PAVMs reported to improve nosebleeds by a subgroup of patients with hereditary haemorrhagic telangiectasia. ERJ Open Res. 2016 Apr 29;2(2):00035-2016.

73 Thielemans L, Layton DM, Shovlin CL. Low serum haptoglobin and blood films suggest intravascular hemolysis contributes to severe anemia in hereditary hemorrhagic telangiectasia. Haematologica. 2019 Apr;104(4):e127-e130.

74 Sharkey TD. Pentose Phosphate Pathway Reactions in Photosynthesizing Cells. Cells. 2021;10(6):1547.

75 Francis RO, D'Alessandro A, Eisenberger A, et al. Donor glucose-6-phosphate dehydrogenase deficiency decreases blood quality for transfusion. J Clin Invest. 2020;130(5):2270-2285.

76 Li H, Forstermann U. Pharmacological prevention of eNOS uncoupling. Curr Pharm Des. 2014;20(22):3595-606.

77 Huang K, Wang Y, Siu KL, Zhang Y, Cai H. Targeting feed-forward signaling of TGFβ/NOX4/DHFR/eNOS uncoupling/TGFβ axis with anti-TGFβ and folic acid attenuates formation of aortic aneurysms: Novel mechanisms and therapeutics. Redox Biol. 2021;38:101757.

78 Ferraro NM, Strober BJ, Einson J, Abell NS, Aguet F, Barbeira AN, Brandt M, Bucan M, Castel SE, Davis JR, Greenwald E, Hess GT, Hilliard AT, Kember RL, Kotis B, Park Y, Peloso G, Ramdas S, Scott AJ, Smail C, Tsang EK, Zekavat SM, Ziosi M, Aradhana; TOPMed Lipids Working Group, Ardlie KG, Assimes TL, Bassik MC, Brown CD, Correa A, Hall I, Im HK, Li X, Natarajan P; GTEx Consortium, Lappalainen T, Mohammadi P, Montgomery SB, Battle A. Transcriptomic signatures across human tissues identify functional rare genetic variation. Science. 2020 Sep 11;369(6509):eaaz5900

79 Joyce KE, Onabanjo E, Brownlow S, Nur F, Olupona K, Fakayode K, Sroya M, Thomas G, Ferguson T, Redhead J, Millar CM, Cooper N, Layton DM, Boardman-Pretty F, Caulfield MJ, Genomics England Research Consortium, Shovlin CL. High definition analyses of single cohort, whole genome sequencing data provides a direct route to defining sub-phenotypes and personalising medicine. medRxiv 2021 Oct 1. https://www.medrxiv.org/content/10.1101/2021.08.28.21262560v1

80 Fisher RA. The nature of adaptation. Oxford: OUP; 1930. Reprinted in Ridley M Ed. Evolution (second edition). Oxford: Oxford University Press; 2004.

81 Agozzino L, Balázsi G, Wang J, Dill KA. How Do Cells Adapt? Stories Told in Landscapes. Annu Rev Chem Biomol Eng. 2020;11:155-182.

82 Gill G, Ptashne M. Negative effect of the transcriptional activator GAL4. Nature. 1988; 334:721–24

83 Silveira MAD, Bilodeau S. Defining the Transcriptional Ecosystem. Mol Cell. 2018 Dec 20;72(6):920-924.

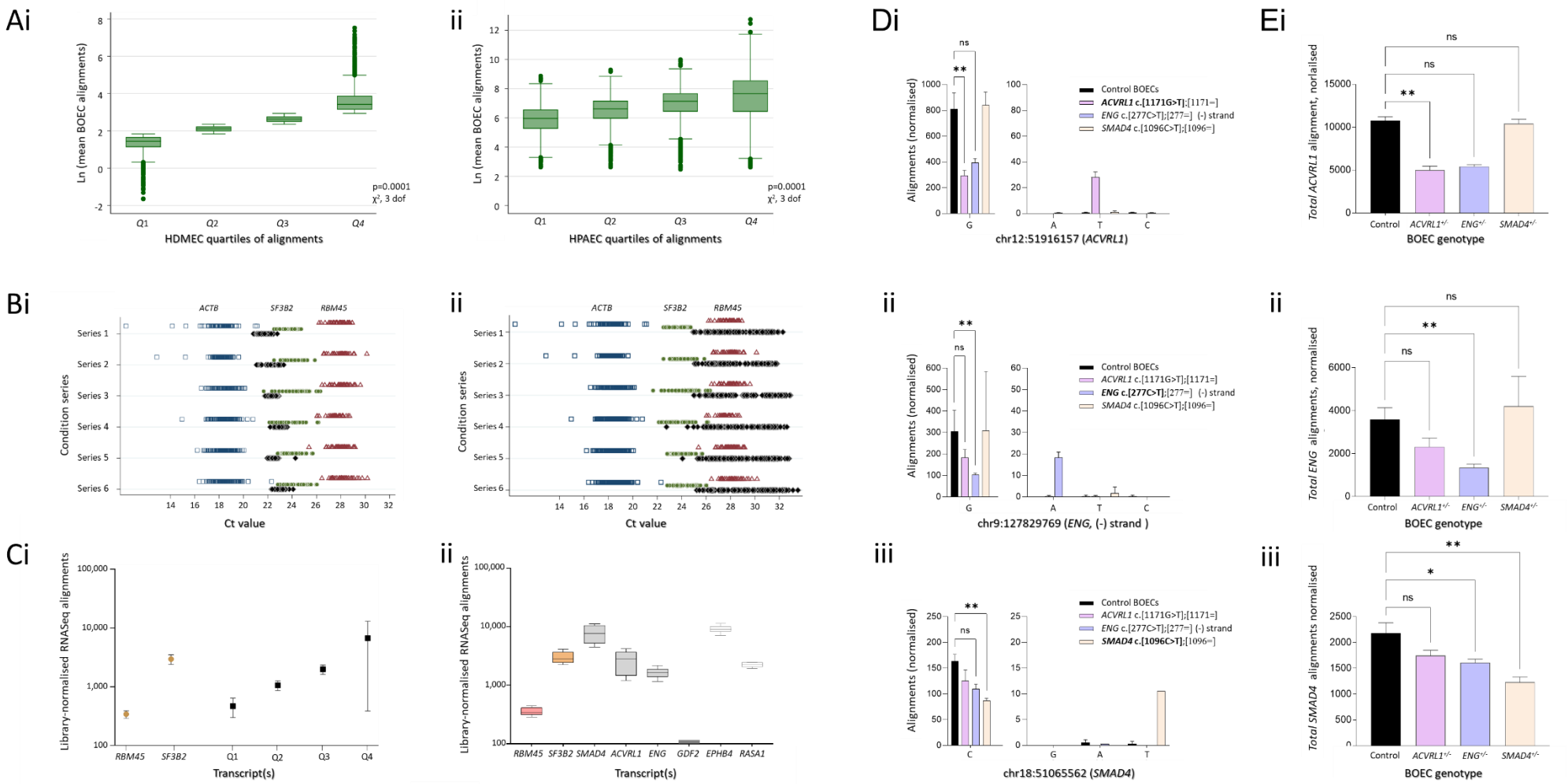


Figure 1: Overview of RNASeq findings in human blood outgrowth endothelial cells (BOECs).

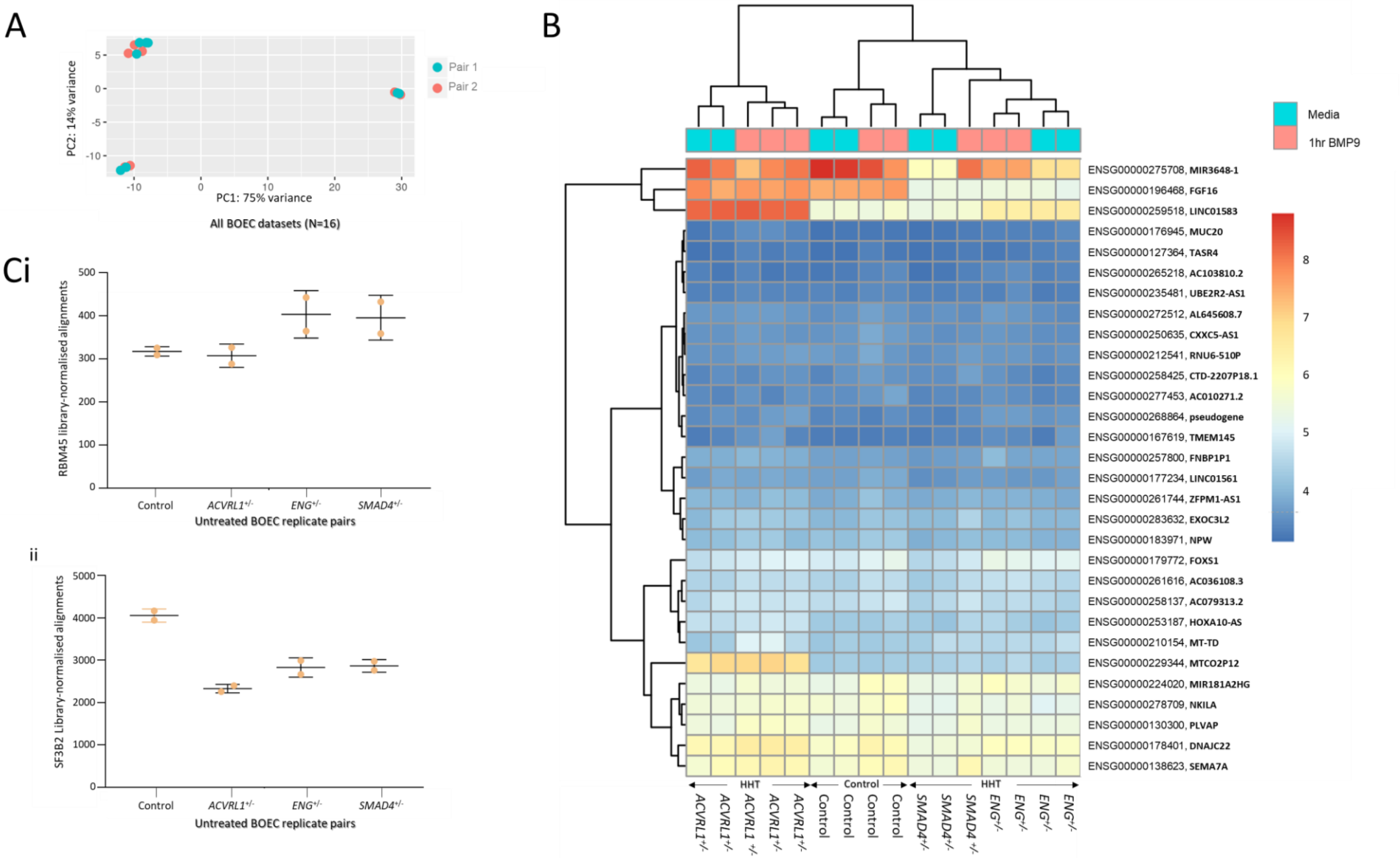
A) Mean alignments across the 16 BOEC samples to Ensembl transcripts by quartiles of alignments in **i)** primary human dermal vascular endothelial cells (HDMEC), and **ii)** primary human pulmonary artery endothelial cells, as published.[33]

B) Comparison of qRT-PCR Ct values across 6 condition series in human umbilical vein endothelial cells (HUVEC) for housekeeper genes *ACTB* (beta actin, blue squares), *SF3B2* (green circles), *RBM45* (red triangles, standard curves in Supplemental Figure SA1), and two selected target genes (black diamonds): **i)** *ID1*, **ii)** *MAP3K4*.

C) BOEC RNASeq library alignment-normalised counts for *RBM45* and *SF3B2* compared to **i)** transcripts in the four quartiles of alignment counts meeting a coefficient of variation of <10% (CV10; error bars indicate mean and standard deviation), and **ii)** causal genes for HHT (*ENG*, *ACVRL1*, and *SMAD4*) and capillary malformation-arteriovenous malformation syndromes CM-AVM1 (*RASA1*) and CM-AVM2 (*EPHB4*): Median, interquartile range, and minimum to maximum values indicated.

D) Quantitative metrics for BOEC RNASeq alignments to nonsense alleles in *ACVRL1*, *ENG* and *SMAD4* genomic loci corresponding to known heterozygous pathogenic variants *ENG* c.277C>T (p.Arg93X), *ACVRL1* c.1171G>T, (p.Glu391X) and *SMAD4* c.1096C>T, (p.Gln366X) in source BOECs. Left graph: wildtype allele at **i)** chr12:51,916,157, **ii)** chr9:127,829,769, **iii)** chr18:51065562 (all *Homo sapiens* GRCh38 coordinates). **p<0.005 Dunn's test post Kruskal Wallis (overall p-values for wildtype allele alignments: *ACVRL1* p<0.0001, *ENG* p=0.0002, *SMAD4* p=0.003). Right graph: alternate alleles at same genomic position at 10X scale. Colour-coded are defined by the key in which both alleles are described for each donor with the relevant nonsense donor allele indicated in bold. All error bars represent mean and standard deviation. Note no nonsense variant was detected by VarScan2[35] set to usual minimum variant allele frequency of 25%. Representative raw traces are provided in Supplemental Figure 2.

E) Total alignments to HHT gene transcripts *ENG* (ENSG00000106991), *ACVRL1* (ENSG00000139567), and *SMAD4* (ENSG00000141646) in the control and HHT BOECs. Data are from all samples per donor, with error bars indicating mean and standard deviation, colour key as in **D**. **p<0.005 by Dunn's test post Kruskal Wallis (p values for overall alignments: *ACVRL1* p<0.003, *ENG* p<0.0001, *SMAD4* p=0.003). The trends for lower *ENG* alignments in *ACVRL1*$^{+/-}$ BOECs have been previously observed[31], and are discussed further in Supplemental Figure 1.



BERNABEU-HERRERO ME, PATEL D, BIELOWKA A, SRIKARAN S, CHAVES GUERRERO P, GOVANI FS, MOLLET IG, NOSEDA M, ALDRED MA, SHOVLIN CL. HETEROZYGOUS TRANSCRIPTIONAL AND NONSENSE DECAY SIGNATURES IN BLOOD OUTGROWTH ENDOTHELIAL CELLS FROM PATIENTS WITH HEREDITARY HAEMORRHAGIC TELANGIECTASIA. *UPLOADED 05.12.2021 AS BIORXIV/2021/421269*

Figure 2: Alignments across all samples in control, *ACVRL1*$^{+/-}$, *ENG*$^{+/-}$ and *SMAD4*$^{+/-}$ BOECs.
A) Principal component (PC) analysis across the 16 RNASeq datasets using Genewiz analytic pipelines of sample-to-sample similarities using alignments normalised to total alignment counts per library. Note the 3 clusters with overlapping coordinates for all treated and untreated samples within the respective clusters.

B) Euclidian distance examining alignments to the top 30 transcripts ranked by differential alignments between untreated and treated BOECs. Note that *ENG*$^{+/-}$ and *SMAD4*$^{+/-}$ BOECs are adjacent, but control BOECs are positioned between the *ENG*$^{+/-}$ and *ACVRL1*$^{+/-}$ BOECs. Only two transcripts (to one mRNA, and one long non coding RNA) met the Genewiz-predesignated threshold for significant differential alignments between treated and untreated BOEC (adjusted p <0.05; absolute log2fold change >1).

C) Intra-duplicate variation in BOEC donor replicate pairs for **i)** *RBM45* and **ii)** *SF3B2*. Intra-replicate variation represented 10-50% of the total range of RBM45 alignments, and 7-17% of the total range of SF3B2 alignments across the 8 samples. Mean and standard deviation indicated.

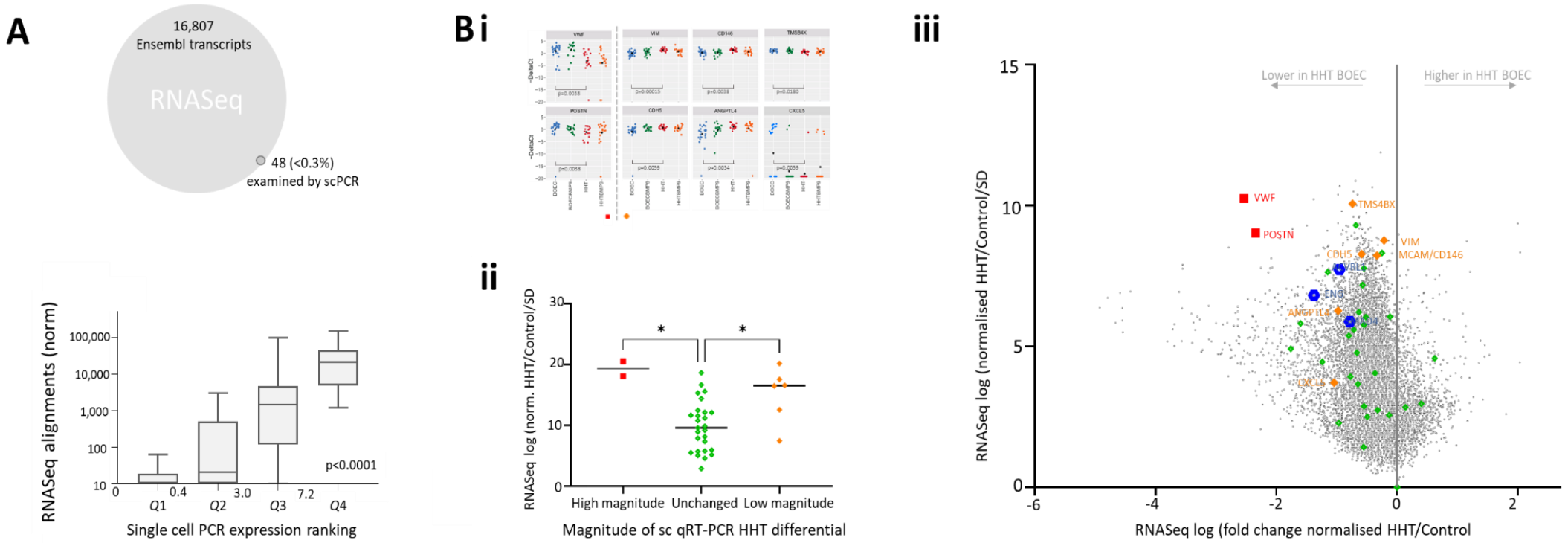


Figure 3: Validations of RNASeq findings by single cell (sc) qRT-PCR.

A) BOEC RNASeq and single cell qRT-PCR transcript evaluations in untreated control BOECs: Compared to 16,807 transcripts evaluated by RNASeq, 48 genes were ranked by expression in single untreated control BOECs. The graph plots RBM45-normalised RNASeq alignments in control BOECs on the y axis, and the single cell qRT-PCR expression for the same gene, categorised by quartiles of expression, on the x axis which indicates the single cell qRT-PCR expression scale from 0-10. P value (p<0.0001) calculated by Kruskal Wallis.

B) HHT differential RNASeq alignment validations by single cell qRT-PCR:
i) Dot plot illustration of the 8 genes meeting Dunn's p<0.05 for untreated control versus untreated HHT BOECs after Kruskal Wallis across all 4 datasets. Each box represents the results of 4 x20 viable BOECs, colour coded per donor (blue/green control; red/orange HHT (*SMAD4*$^{+/-}$), with mean values for each culture indicated by a black circle. Due to the higher magnitude changes for VWF and POSTN, these were subcategorised separately from the other 6 prior to RNAseq comparisons.
ii) The significance index for RNASeq differences in gene alignments between HHT and control BOECs (corresponding to the *y* axis in Figure 4Biii), plotting the 48 sc qRT-PCR evaluated genes categorised by sc qRT-PCR expression as in **Bi)**. P values were calculated by Dunn's test after Kruskal Wallis across all 48 genes (p<0.0054).
iii) Visual representation of all RBM45-normalised alignments in HHT BOECs compared to control BOECs. The *x* axis quantifies lower and higher expression using the maximal value per pair from the untreated datasets, with the standard deviation across all untreated BOECs used for the *y* axis (for details, see Methods and Supplemental data). Blue symbols highlight the positions of HHT genes *ENG, ACVRL1* and *SMAD4*, green, orange and red symbols, the transcripts evaluated in single BOECs. Green symbols indicate the 40 transcripts with p>0.05, red and orange symbols the transcripts differentially expressed using the colour coding as in **Bi/ii.**

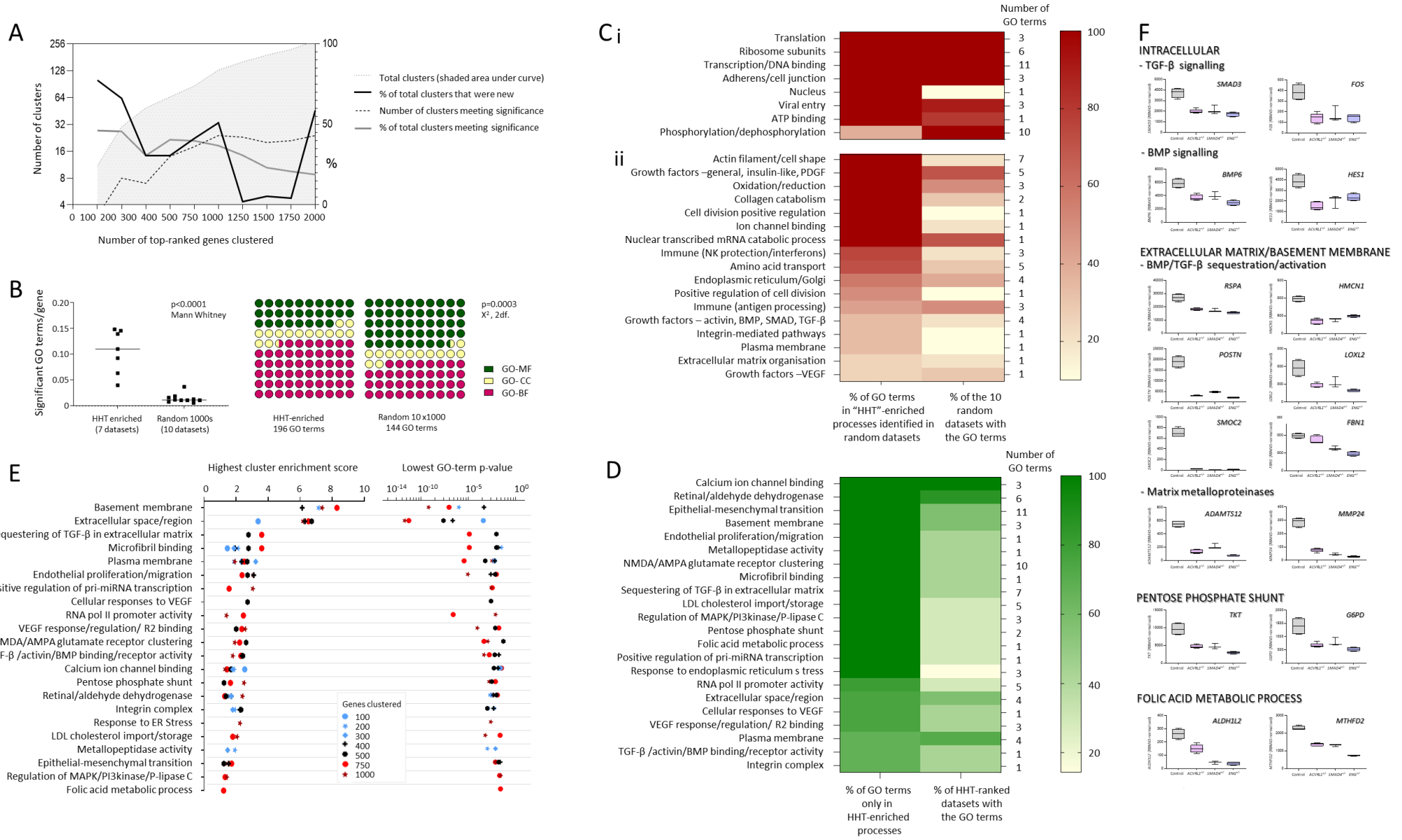


Figure 4: Gene ontology term clustering of genes displaying differential alignments in HHT BOECs compared to controls.

A) Overview of gene ontology (GO) term[48] clusters obtained in DAVID[49,50] restricting to BP-DIRECT (biological processes), CC-DIRECT (cellular compartment) and MF-DIRECT (molecular function) terms, for the top 100, 200, 300, 400, 500, 750, 1,000, 1,250, 1,500, 1,750 and 2,000 ranked genes based on differential alignments between HHT and control BOECs. Total number of clusters indicates all clusters (shaded area under the curve). Clusters met significance if enriched >1.2 fold and contained a GO term meeting $p<0.05$. 'New' clusters were clusters not identified in preceding datasets.

B) Comparison of the number of GO terms[48] in significant clusters[49,50] derived from the 7 HHT datasets used in final analyses (top 100, 200, 300, 400, 500, 750, and 1,000 ranked genes based on differential alignments between HHT and control BOECs), and 10 sets of 1,000 randomly selected genes. The graph indicates the number of GO terms meeting significance and the dot blots, the relative proportion of GO molecular function (MF), cellular composition (CC) and biological function (BF) terms.

C) Generating black and amber lists of processes where clusters[49,50] and GO terms[48] were present in randomly-selected gene datasets (from the 10 x 1,000 randomly selected genes). The first column indicates the proportion of GO terms in the full cluster that were present in the random datasets, and the second column, the percentage of the random datasets that contained the GO terms, with the % represented as per heatmap scale. The upper graph displays the 8 process types for 38 GO terms that were placed on the black list because they reached significance in a random dataset. The lower graph displays the 17 process types for 44 GO terms that were placed on the amber list due to presence in a random dataset, though without reaching significance.

D) Generating a green list of processes where GO terms[48] were not present in randomly-selected gene datasets. The first column indicates the proportion of GO terms in the full cluster that were only present in the HHT-enriched datasets, and the second column, the % of HHT-ranked datasets (top 100, 200, 300, 400, 500, 750, and 1,000 ranked genes) that contained the GO terms, with % represented as per heatmap scale. The graph displays 22 process types (for 38 GO terms) that were considered for the green list.

E) For the processes/GO terms[48] on the potential green list (Figure 5D), highest cluster enrichment scores,[49,50] and lowest GO term p-value for GO term within the cluster[49,50] are displayed for each of the HHT-ranked datasets (top 100, 200, 300, 400, 500, 750, and 1,000 ranked genes) in which the cluster(s) were identified, using the key indicated. For the list of genes presented in Supplementary Table 5, four processes were removed as they were less relevant to HHT/endothelial cells, or contained a very large number of genes in clusters that were similar to amber list processes (RNA polymerase II promoter activity; regulation of MAPK4/PI3 kinase/P-lipase C).

F) Inter-genotypic comparisons for 16 representative genes displaying differential alignments between control and HHT BOECs, categorised by TGF-β signalling (*SMAD3, FOS*); BMP signalling (*BMP6, HES1*); extracellular matrix: BMP/TGF-β sequestration/activation (*RSPA, HMCN1, POSTN, LOXL2, SMOC2, FBN1*); matrix metalloproteinases (*ADAMTS12, MMP24)*; pentose phosphate shunt (*TKT, G6PD*) and nuclear-encoded mitochondrial tetrahydrofolate reductases (*ALDH1L2, MTHFD2*). Further details on the respective protein functions are provided in Supplemental Figure 8.

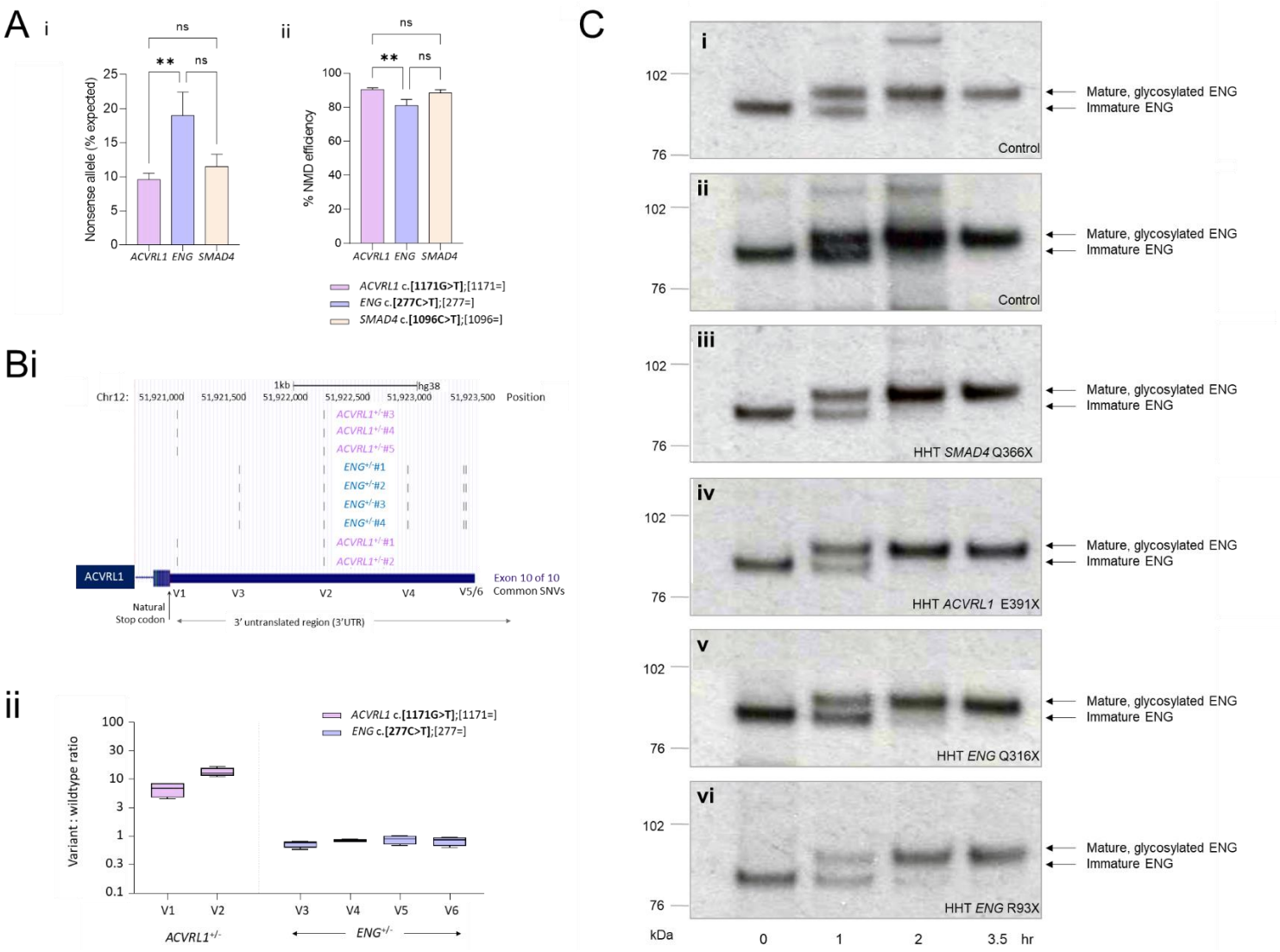


Figure 5: Expression of nonsense allele in HHT BOECs.

A) Alignments to nonsense, pathogenic allele in the HHT BOECs by genotype, colour coded as in key. **i)** As percentage of total expected if equal to wildtype alignment; **ii)** As percentage "loss" of total expected allele alignments.

B) Alignments to common single nucleotide variants (SNVs) in the *ACVRL1* 3'untranslated region (UTR). Note these were not known prior to identification following the BOEC RNASeq and alignment analyses.

i) Genomic (GRCh38) positions of SNVs V1-V6 incidentally identified in BOECs from donors heterozygous for *ACVRL1* c.1171G>T, (p.Glu391X), or *ENG* c.277C>G, (p.Arg93X), i.e. *ACVRL1* $^{wildtype+/+}$. Positions are illustrated by custom tracks uploaded to the University of California Santa Cruz (UCSC) Genome Browser.[44,45]
ii) Variant:wildtype ratios in the *ACVRL1*$^{+/-}$ and *ACVRL1* $^{wildtype+/+}$ (*ENG*$^{+/-}$) BOECs, indicating that the *ACVRL1* wildtype alleles that were detected at <100% (for homozygotes), but >50% (for heterozygotes) were in *cis* with *ACVRL1* c.1171G>T. The alternate allele in *trans,* present at <<50% was not detected by VarScan2 set to Genewiz standard minimum detection threshold of 25%.

C) Pulse chase experiments, modified from methods described in reference [66], demonstrating generation of immature endoglin protein in *ENG*$^{+/-}$ nonsense-containing BOECs. Data are from BOEC lysates following a 1 hr pulse with ^{35}S-methionine and chase with unlabelled methionine for 0, 1, 2 or 3.5hr (see methods for further details).
i-iv) *ENG*$^{+/+\,(wildtype)}$ 'controls': **i)** Control A; **ii)** Control B; **iii)** *SMAD4* c.1096C>T, (p.Gln366X); **iv)** *ACVRL1* c.1171G>T, (p.Glu391X).
v-vi) *ENG*$^{+/-}$ nonsense BOECs: **v)** Exon 10 *ENG* c.1306C>T, (p.Gln436X); **vi)** Exon 3 *ENG* c.277C>G, (p.Arg93X).

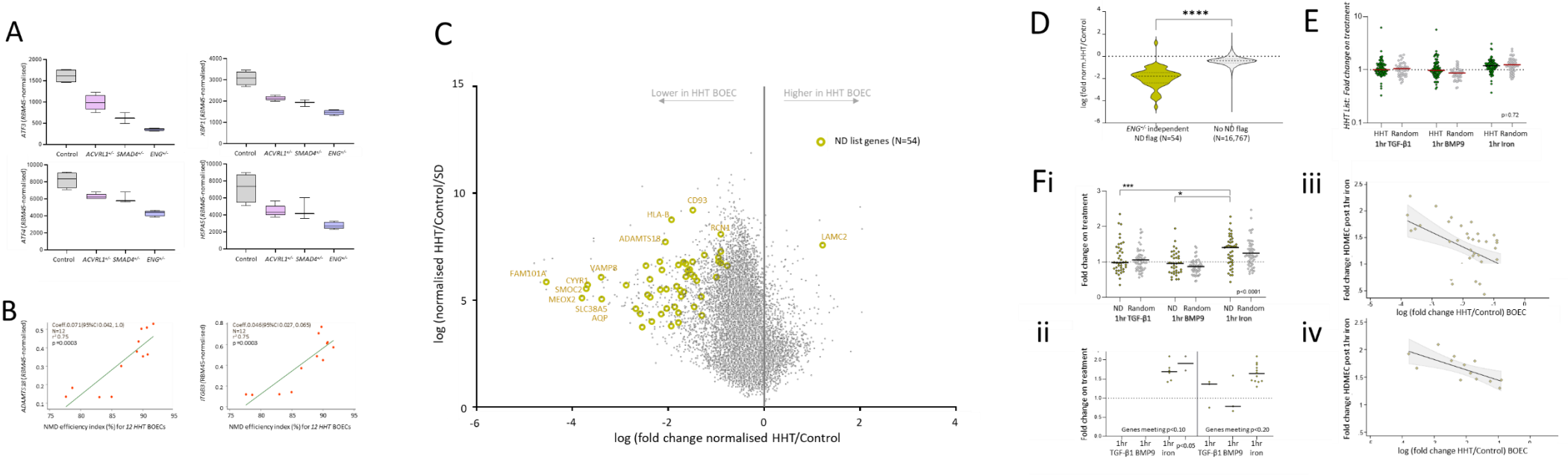


Figure 6: Expression of nonsense-containing alleles, and relationships with differential transcript expressions in HHT BOECs and normal HDMEC.

A) Comparative *RBM45*-normalised alignment to the 4 genes clustered as part of the GO:1990440 term for endoplasmic reticulum stress. Note the similar patterns across the donor BOECs.

B) Two representative genes in GO processes that met significance thresholds for RNASeq differential HHT alignments, but where RNASeq alignments also displayed significant relationships to the ND index (efficiency of nonsense allele loss) both including and excluding the *ENG*$^{+/-}$ BOECs. The graphs show alignments in each HHT BOEC culture compared to the mean of the 4 control BOEC cultures across all 12 cultures of HHT BOECs, with metrics calculated by linear regression. Further examples are provided in Supplemental Figure 9.

C) Visual representation of the 54 ND-index flagged genes ('ND list', khaki symbols) on the graph of *RBM45*-normalised alignments for HHT BOECs compared to control BOECs (as in Figure 4B).

D) Magnitude of difference between HHT and control BOECs (corresponding to the *x* axis in Figure 6C), for genes meeting ND list significance excluding the *ENG* BOECs, and the genes with no ND flag (Mann Whitney p value <0.0001).

E) Genes on the HHT green list, examined for differential expression in normal primary human dermal microvascular endothelial cells (HDMEC). 87 were identified in HDMEC. Following 1hr treatment with TGF-β, BMP9 or 10µmol iron[33], there were no significant differences between the datasets, with and without incorporation of randomly selected genes from the dataset.

F) Genes on the ND list, examined for differential expression examined in normal primary HDMEC. **i)** 41 transcripts were identified in HDMEC. Note there was a significant (p<0.0001) increase in alignments to these genes in BOECs treated for 1hr with 10µmol iron[33] which was not seen for randomly selected genes, or following 1hr treatment with TGF-β or BMP9. **ii)** Fold change difference in ND list genes meeting designated levels of significance following indicated treatments (Rx), at p<0.10 and p<0.05 [left side]; or p<0.20 [right side], as detailed in reference[33]: HDMEC were treated with TGF-β, BMP9 or iron for 1hr as described [33] **iii)** The inverse relationship for all 41 ND list genes identified in HDMEC (regression coefficient -0.27 (95% CI -0.43, -0.12), r^2 =0.25, p=0.001). i**v)** The inverse relationship for the 13 genes on the ND list meeting post iron [HDMEC] p<0.20[33] (regression coefficient -0.19 (95% CI -0.31, -0.065), r^2 =0.51, p=0.006).

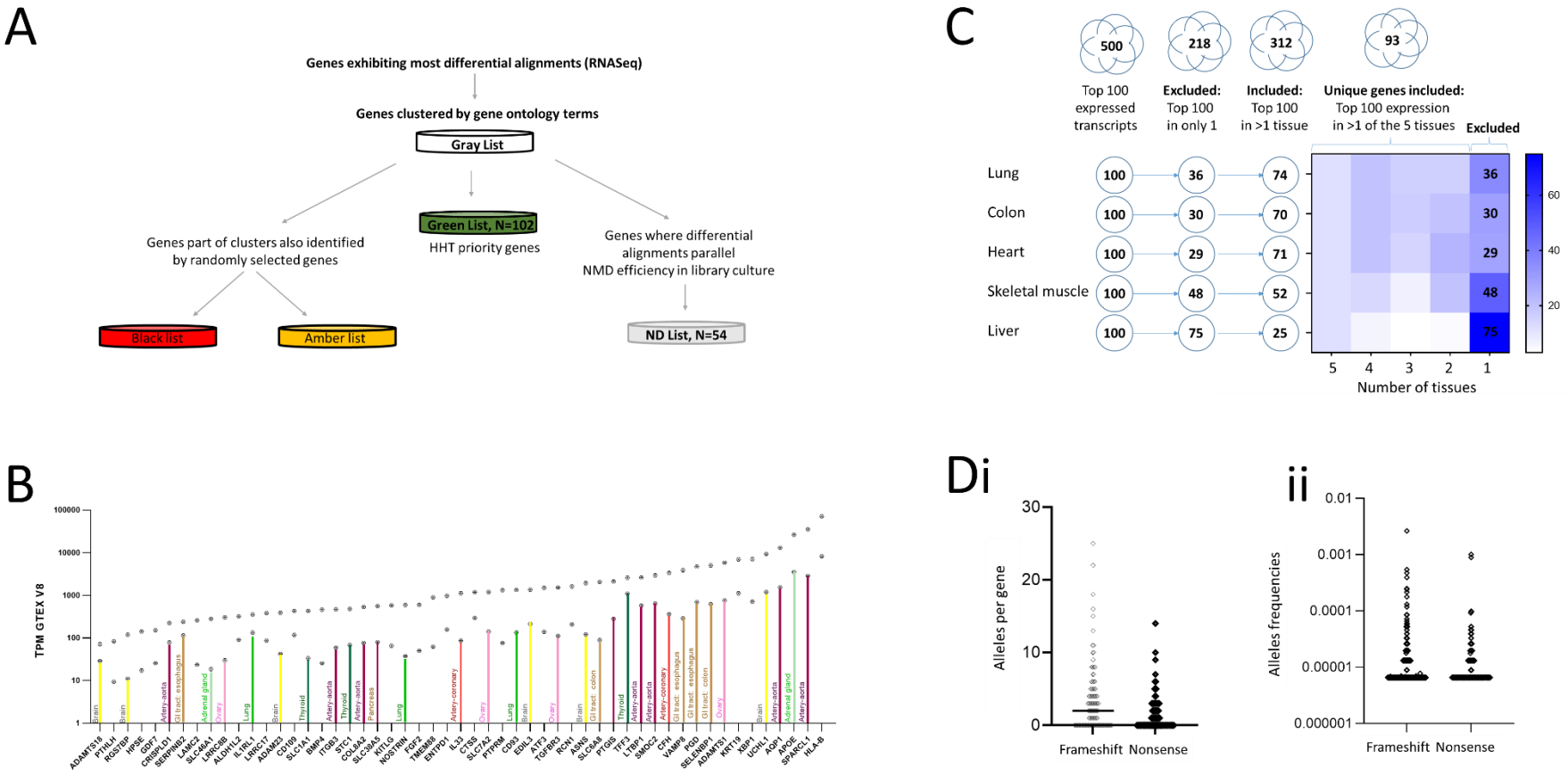


Figure 7: Extending the implications beyond HHT and endothelial cells:

A) The 4 lists of genes identified, emphasising random and nonsense-mediated decay (ND) associated genes with less specific disease relevance: The 'red/black" list represents genes in GO terms that reached significant thresholds in random datasets (and were also present in most or all random datasets (Figure 4Ci)). The amber list represents genes for GO terms that were present in a smaller number of random datasets and did not reach significant thresholds in any (Figure 4Cii). For further details of the **102** and **54** genes on the green and ND lists, see text, Figure 4 (green list), and Figure 6 (ND list). For the full lists of genes in these categories, please see Supplementary Table 6.

B) Tissue expression of the ND list genes in GTEx[1,51]: Upper symbols represent total Transcripts Per Million (TPM) per gene, and lower symbols the highest single tissue contributor, with major tissues colour coded. Note that most genes on the ND list have their highest tissue expression in non endothelial tissues.

C) Breakdown of the 100 most expressed genes in 5 major tissues in GTEx.[1,51] 218/500 (44%) of listings were in the top 100 for only one tissue, and were excluded in order to limit potential biases introduced by highly tissue specific genes. The remaining 282 listings represented 93 unique genes. The heat map categorises proportions for each tissue that were excluded because present in the 100 most expressed genes in 1 tissue only, and the 93 included genes which were in the top 100 transcripts in 2,3,4, or all 5 tissues.

D) GnomAD 3.2.1[52]-listed nonsense and frameshift alleles in the 93 genes that encoded the "top 100" expressed transcripts in at least two of the 5 selected non-endothelial GTEx tissues. **i)** Number of variants per gene by molecular mechanism that generates a premature termination codon (PTC) Nonsense ≡ gain of stop. Consensus splice site variants were not included because up to one third may cause inframe indels which would not result in a PTC to trigger NMD.[32] **ii)** Allele frequencies. Horizontal bars at median.