Heterozygous transcriptional signatures unmask variable premature termination codon (PTC) burden alongside pathway-specific adaptations in blood outgrowth endothelial cells from patients with nonsense DNA variants causing hereditary hemorrhagic telangiectasia.

Maria E Bernabeu-Herrero[1,2], Dilip Patel[1,2], Adrianna Bielowka[1,2], Patricia Chaves Guerrero[1], Stefan J Marciniak[3], Michela Noseda[1], Micheala A. Aldred[4], Claire L Shovlin[1, 2,5]

[1] National Heart and Lung Institute, Imperial College London; London, UK; [2] NIHR Imperial Biomedical Research Centre; London, UK; [3]CIMR, University of Cambridge, Cambridge, UK; [4]Indiana University School of Medicine; Indianapolis, USA; [5]Imperial College Healthcare NHS Trust; London, UK

Running head: RNASeq of HHT BOECs

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

Abstract 250

Word count 3822

ABSTRACT

Frameshift and nonsense DNA variants represent the commonest causes of monogenic inherited diseases. They usually generate premature termination codon (PTC)-containing RNA transcripts that produce truncated proteins in recombinant systems, but endogenously are subject to nonsense mediated decay. To examine native consequences of these variants, we derived cells from pre-genotyped patients. Blood outgrowth endothelial cells (BOECs) were established from individuals with hereditary hemorrhagic telangiectasia (HHT) due to a heterozygous nonsense variant in *ACVRL1*, *ENG* or *SMAD4* that each encode an endothelial cell-expressed protein mediating bone morphogenetic protein (BMP)/ transforming growth factor (TGF)-β signalling. RNA sequencing alignments to PTC alleles varied from 8-23% of expected, and differed between same-donor replicates. Differential gene expression analyses were validated by single cell qRT-PCR, and identification of changes in wider, disease-specific BMP/TGF-β pathway components. However, the most differentially expressed genes clustered to disease-independent terms for post translational protein modification (isopeptide bond; ubiquitin-like conjugation). They were the only terms meeting Benjamini significance after clustering Bonferroni-ranked, differentially expressed genes from the 5,013 meeting 10% intraassay coefficients of variation, and significance was robust to normalisation methods. Optimised pulse chase experiments supported perturbed wildtype protein maturation, but no PTC-truncated protein was identified. Unexpectedly, BOEC cultures with highest PTC persistence were discriminated in unsupervised hierarchical clustering of low GINI coefficient 'invariant' housekeeper genes, and patterns were compatible with higher cellular stress. The findings support a model whereby PTCs are more of a burden in stressed cells, and lead us to conclude that overlooked and varying PTC burdens contribute to biological variability.

INTRODUCTION

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,[1,2] and avoid over-targeted experimental readouts that can miss unanticipated adaptations employed by cells in order to survive, including therapy-nullifying responses.[3] Evaluation platforms now include patient-derived cells for bulk, single cell, and organ-on-a-CHIP analyses.

Our goals were to deliver methodologies for pre-clinical studies in patient-derived blood outgrowth endothelial cells (BOECs), incorporating understanding of the extent to which disease-causing RNA transcripts harbouring premature termination codons (PTCs) are lost in normal cellular safeguarding processes such as nonsense mediated decay (NMD).[4-6] Currently, PubMed searches for "nonsense mediated decay" retrieves <3% of the number for "truncated protein" search terms, while "truncating variant" (not a recommended descriptor[7,8]), is still widely used to describe DNA variants in manuscripts and clinical reports that imply truncated proteins are generated in endogenous tissues, as in recombinant *in vitro* studies.

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.[9,10] HHT is inherited as an autosomal dominant trait, and typically caused by heterozygous, loss-of-function variants in *ACVRL1*, *ENG* and *SMAD4*.[11-13] The most common variants are PTC-generating such as frameshift and nonsense ('stop gain'), with hundreds of unique pathogenic PTC variants together responsible for 42-67% of HHT cases, according to causal gene.[14] While a small number of HHT cells may lose the second wildtype allele,[15] in the majority, it is retained, even in abnormal vessels where approximately half normal ENG protein expression was demonstrated in *ENG*$^{+/-}$ HHT patients.[16-18] For the disease-causing allele, there is evidence in HHT patient-derived cells that the transcript may be lost at the RNA level (*ENG* nonsense variants in EBV-transformed lymphoblastoid cell lines[19]), or generate only short-lived, endoplasmic reticulum-retained proteins (*ENG* and *ACVRL1* multiexon deletion and missense

alleles[16-18,20-23]). There has been limited attention on how these processes operate alongside pathway-specific perturbations in the heterozygous state.

Pathway-specific perturbations in HHT have been reviewed recently.[12,13] *ACVRL1*, *ENG* and *SMAD4* encode endothelial-cell expressed proteins that transmit or modify signaling by the bone morphogenetic protein (BMP)/ transforming growth factor (TGF)-β superfamily, with HHT causality confirmed by null, heterozygous and endothelial-specific disease gene models generated using recombinant technology.[24-28] Early heterozygous models were recognised as excellent simulations of the human condition[24,29] but the same variability and often mild phenotypes emphasised the importance of a second and even third "hit" in addition to the heterozygous state.[30] Signaling defects are usually studied in conditional homozygous null models which generate pronounced and reproducible phenotypes, for example in the context of trauma[25] or angiogenesis.[31,32]These have provided important insights, including modified BMP/TGF-β receptor expression interpreted as an adaptive response that enabled more *ENG* deficient cells to survive.[32]

To examine native consequences of heterozygous alleles that generate PTCs, we generated blood outgrowth endothelial cells (BOECs) from pre-genotyped HHT donors and controls. Here we report marked variability in the persistence of nonsense pathogenic variants in replicate, resting state cultures from human heterozygous HHT donors; associated changes in protein maturation kinetics; and independent, unsupervised RNAseq discrimination of BOECs cultures with higher PTC burdens.

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 BOECs established from healthy volunteers and pre-genotyped HHT patients, using previously optimised methods.[33-35] All analyses were performed in confluent passage 3 vials as described further in the Data Supplement.

For RNASeq library preparations, RNA was extracted from 16 separate cultures of confluent BOECs from 4 separate control and HHT donors, cultured in the presence and absence of BMP9 at 10ng/ml for 1 hr. Library preparations, alignments to *Homo sapiens* GRCh38 (STAR aligner v2.5.2b), counts of unique gene reads that fell within exon regions (Subread package v1.5.2), normalisation to total alignment counts per library, and initial variant calls[36] were performed by Genewiz (Leipzig, Germany).

To provide numeric validations of a subset of genes, single cell (sc) qRT-PCR was performed as detailed further in the Data Supplement:[37,38] For each culture, 40 viable (DRAQ7 negative) single cells were sorted directly into 96-well-plates containing pre-amplification mix for 48 genes selected either as endothelial gene markers as validation for endothelial differentiation, or likely relevance to the HHT cellular phenotype, or as important negative controls (Supplementary Table 1). Specific target amplification was performed using CellDirect One-Step qRT-PCR Kit (Invitrogen, Thermo Fisher, Waltham, MA) as described.[37,38] The 48 genes examined by sc qRT-PCR were normalized to *UBC* expression, analysed, and plotted using an R script developed in house. Novel housekeeping genes were sought from genes with low Gini Coefficients (GC) within multiple cell lines.[39,40]

To enhance the technological feasibility of detecting endogenous low level, truncated proteins, metabolic labelling was performed in BOEC protein extracts using ENG mAb SN6h (Dako Denmark A/S, Denmark). Modifying methods of Pece et al,[18] 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.

RNA Sequencing Analyses

Percentage loss of the nonsense allele in *ACVRL1*$^{+/-}$, *ENG*$^{+/-}$ and *SMAD4*$^{+/-}$ BOECs was calculated assuming that without nonsense mediated decay (NMD),[4-6] there would have been an approximately equal number of alignments to wildtype and nonsense alleles in each HHT BOEC library.

Across all 16,807 genes, the intra-assay coefficient of variation (CV),[41] *100 x standard deviation (SD)/mean*, was calculated for replicate pairs, adjusting the alignment per gene for total library read counts, and restricting to genes where all four untreated replicate pairs had an intra-assay coefficient of variation <10% ('met CV10'). The 25 genes with GINI Coefficients (GCs)<0.15 in diverse cells, i.e. representing the least variable of human transcripts,[39,40] were tested for BOEC expression and used as housekeepers for DeSeq2 normalization.[42] This calculated the ratio of alignment counts for each selected GINI housekeeper gene in each BOEC dataset to the geometric mean of that GINI gene across all 16 BOEC datasets. The median value of GINI gene ratios in each BOEC library was then used to generate the 'size factor' to scale that library's alignments.[42,43]

Gene Ontology[44,45](GO) and pathway enrichment analysis of HHT differentially expressed genes was performed using Functional Annotation Clustering through the Database for Annotation, Visualization and Integrated Discovery (DAVID)[46]v6.8. Primary clustering analyses were on the most differentially expressed genes meeting CV10[41] in all four untreated replicate pairs of BOECs, and where the gene's differential expression between control and HHT BOECs met either Bonferroni $p<0.05$, or Bonferroni $p<1.00$. GO and pathway terms clustered by these genes were considered significant if the individual pathway terms met Benjamini $p<0.05$. To test for robustness of normalization methods and specificity, findings were compared to all gene ontology clusters generated by genes ranked without GINI Coefficient housekeeper gene normalization, and to all clusters generated by 10 sets of 1,000 genes randomly derived from the complete dataset of gene alignments in the BOECs.

Data Analysis

STATA IC v 15.0 (Statacorp, College Station, TX) was used for data handling including calculation of summary statistics, comparison of groups using Kruskal Wallis when Dunn's test was used to derive pairwise comparisons; linear regression, unsupervised heirachical clustering, and principal component analyses (PCA). Data were visualised using GraphPad Prism 9 (GraphPad Software, San Diego, CA), and STATA IC v 15.0 (StataCorp., College Station, TX).

RESULTS

Establishing BOEC platforms

To examine native consequences of PTC-generating variants, we cultured primary blood outgrowth endothelial cells (BOECs) from pre-genotyped individuals with known disease-causal nonsense variants (HHT patients) and healthy volunteers. BOEC establishment rates were independent of the number of peripheral blood mononuclear cells plated; time to first colony appearance; HHT versus control status; and *ACVRL1*$^{+/PTC}$ versus *ENG*$^{+/PTC}$ donor (all p-values >0.22, Supplementary Fig. 1). BOECs were confirmed as CD31^{+}, CD44^{+}, CD105^{+}, CD90^{-} by bulk flow cytometric evaluations (Supplementary Fig. 2A) and displayed consistently high expression of a panel of endothelial marker genes evaluated by single cell (sc) qRT-PCR (Supplementary Fig. 2B, Supplementary Table 1).

BOECs were used to derive libraries for RNA sequencing (RNASeq), yielding reads of consistently high quality (Supplementary Fig. 3). RNASeq methodology was validated by same-donor control BOECs of identical passage but defrosted and cultured separately for single cell (sc) qRT-PCR experiments: For 48 genes examined using both methodologies, mean expression across 20 single control BOECs mirrored mean bulk RNASeq alignments in replicate control BOEC cultures (Supplementary Fig. 2C). Overall, 59% of the variability in the mean control BOEC RNASeq alignments normalized to total library count was accounted for by sc qRT-PCR ranking expression ($p<0.0001$).

BOEC identity supported by pathway-specific gene analyses

Initial global unsupervised analytics were performed by Genewiz blinded to RNA origin, and based on alignments normalized to total library reads. Euclidean distances (Fig. 1A) and principal component analyses (PCA, Fig 1B) distinguished clusters of BOECs, and on unblinding these were seen to be distinguished by genotype. In keeping with the expected loss due to nonsense mediated decay,[4-6] HHT BOECs demonstrated reduced overall alignments to the gene harbouring the mutant PTC allele (Fig. 1C). There were also trends for *ACVRL1* alignments to be low in *ENG*$^{+/PTC}$ BOECs, and for *ENG* alignments to be low in *ACVRL1*$^{+/PTC}$ BOECs, although these did not reach statistical significance.

Individual genes encoding other BMP and TGF-β signalling pathway components ranked highly in the genes differentially expressed between HHT and control BOECs, whether normalized to total library reads, or using more stringent analyses that restricted to 5,013 genes meeting an intra-assay coefficient of variation of 10% (CV10),[41] and employing DeSeq2 normalization[42] using low Gini Coefficient (GC<0.15[39,40]) housekeeper genes. Examples of differentially expressed genes are provided in Fig. 1D, including those encoding superfamily ligands (*TGFB1* Fig. 1Di, *BMP6* Fig. 1Dii), and extracellular matrix proteins involved in TGF-β/BMP ligand activation (latent transforming growth factor beta binding protein 1 (*LTBP1*, Fig. 1Diii), fibrillin 1 (*FBN1* Fig. 1Div), and leucine rich repeat containing 32 (*LRRC32*, Fig. 1Dv) which were all lower in the HHT BOECs than controls.

Taken together, the findings supported the HHT genotypic BOEC origin, with evidence suggesting wider changes in transcripts for the BMP/TGF-β signaling pathways following heterozygous loss of ACVRL1, ENG or SMAD4 function.

Gene ontology and pathway term clustering identifies differences in protein processing pathways in HHT BOECs

To explore differences beyond the specific pathways impacted by loss-of-function variants, unbiased gene ontology clustering[44-46] was performed on differentially expressed genes in the 5,013 meeting CV10.[41] To examine at the highest stringency ('Stringency Level 1'), we restricted analyses to the 30 genes (0.6% of the 5,013 CV10 genes) that met Bonferroni p<0.05 for differences between untreated control and HHT BOECs (Supplementary Table 3A). Only one cluster contained terms meeting predesignated significance. These terms were Uniprot Keywords[47] (KW)-1017 isopeptide-bond (9 genes, Benjamini p=0.034), and KW-0832 Ubl conjugation (10 genes involved in ubiquitin-like (Ubl) conjugation, Benjamini p=0.034, Fig. 2Ai). Ten of the 30 (33%) most differentially expressed genes meeting Bonferroni p<0.05 mapped to these terms, reflecting the overlapping biology (ubiquitin and ubiquitin-like (Ubl) peptides bind to protein substrates via isopeptide bonds to target for lysosomal degradation). For Stringency Level 2, the 161 genes that met Bonferroni p<1.00 were clustered (Supplementary Table 3B). These generated 2 clusters meeting predesignated significance, with terms again including KW-1017~isopeptide-bond (29 genes, Benjamini

p=0.015), and KW-0832~Ubl conjugation (36 genes, Benjamini p=0.026, Fig. 2Aii). With no apparent difference between the BOEC HHT datasets (Fig. 2B), we concluded that the data were consistent first, with differences in post translational protein processes between the PTC (HHT) and non-PTC (control) datasets, and second, with recent experimental data demonstrating a ubiquitin pathway is used for nascent protein degradation from PTC-containing mRNAs.[48]

For a broader overview, lower stringency analyses were performed examining HHT-differentially expressed gene normalized to total read counts per library (Supplementary Methods; Fig. 2Ci). As shown in Fig. 2Cii and Supplementary Table 4, in step-wise increases from the top 100 (~2%) to top 1,000 (~20%) genes, each increment increased the total number of clusters and GO terms meeting significance, with significant terms emerging repeatedly. There was a marked contrast between the number of significant terms emerging from the clustering of genes differentially expressed in HHT versus control BOECs, compared to 10 sets of randomly selected 1,000 genes from the same BOEC datasets (Fig. 2D). Notably, the most significant terms were not specifically related to endothelial function or BMP/TGF-β signaling predicted by the loss-of-function alleles, but again Uniprot key words KW-1017 (isopeptide-bond) and KW-0832 (Ubiquitin-like conjugation) (Fig. 2E).

PTC-containing transcript retention in BOECs

The differential expression of genes clustering to protein terms would be most easily explained if HHT BOECs were generating mutant protein from the PTC allele. Initial unsupervised VarScan2[36] variant call format (vcf) files did not detect the nonsense variants using standard parameters (minor allele frequency >25%). To test whether PTC-containing RNAs were persisting at lower levels in HHT BOECs, supervised examination of RNASeq binary sequence alignment map (bam) files was performed. In 12 of the 16 HHT and control BOECs, alignments to GRCh38/hg38[49] demonstrated significant mismatches at chr12:51,916,158, chr9:127,829,769 or chr18:51,065,563 corresponding to the HHT donors' genotypes *ACVRL1* c.1171G>T, p.(Glu391X), *ENG* c.277C>T, p.(Arg93X), and *SMAD4* c.1096C>T, p.(Gln368X) (Fig. 3A, Supplementary Fig. 5). In all cases, HHT nonsense allele alignments were within the expected BOECs, and aligned to spliced exons (almost all alignments sharply defined the exon-intron boundaries,

Supplementary Fig. 6). All ratios were skewed from the 1:1 expected for germline heterozygous alleles with a mean wildtype: nonsense allele ratio of 9:1 (range 5:1-13:1, Fig. 3B).

Retention of PTC-containing transcripts at levels below those detected by standard variant call file metrics would enable common heterozygous variants *in cis* with the wildtype allele to emerge as over-represented rather than the sole sequence, and this was the case: By chance, both *ENG* $^{+/PTC}$ and *ACVRL1*$^{+/PTC}$ donors were heterozygous for common variants in the *ACVRL1* 3'untranslated region (UTR) which was examined in the UCSC Genome Browser[50,51] (Fig 3Ci). VarScan2 indicated the expected 1:1 ratio for four different *ACVRL1* 3'UTR variants in *ENG* $^{+/PTC}$ BOECs (i.e. *ACVRL1*$^{+/+}$), but skewed allele ratios for the two *ACVRL1* 3'UTR variants in *ACVRL1* $^{+/PTC}$ BOECs (Fig. 3C). The findings were consistent with persistence of a proportion of *ACVRL1* PTC-containing transcripts to the 3'UTR, which would potentially permit in-frame protein translation.

Pulse chase examination of ENG protein in HHT and control BOECs

To test aberrant production of mutant protein, we focused on endoglin which had the highest retention of PTC-containing transcripts in BOECs (Fig. 3B), and well-established monoclonal antibodies for pulse chase experiments (mAbs PD31, P4A4 and SN6h[52,53]). These were optimised for use in BOECs as detailed in the Supplementary Methods. Although most HHT-causal *ENG* nonsense variants are too proximal (5') for predicted proteins to contain both of the ENG orphan (OR) domains that span the mAb epitopes (Fig. 4A), we were able to establish BOECs from an HHT patient heterozygous for a rare nonsense substitution (*ENG* c.1306C>T) generating a PTC at Q436, at the end of the zona pellucida (ZP)-N domain. This predicted that the complete OR1/2 domains and wildtype tertiary structure of all SN6h epitope residues (A119-G230)[53] would be present in a protein truncated at the PTC (Fig. 4A). Even so, Western blot and pulse chase analyses displayed no size differences in endoglin SN6h-reacting protein extracts from *ENG* Q436X BOECs compared to control BOECs (Fig. 4B).

The maturation of wildtype ENG protein was examined across further HHT and control BOECs. Commencing the chase when only newly-synthesised, immature endoglin protein was present (Fig. 4C),

mature protein became the more pronounced species within 1hr in control BOECs, and was the only species visible by 2hr (Fig. 4Ci, 4Cii). For *SMAD4*$^{+/PTC}$ (Fig. 4Ciii), and *ACVRL1*$^{+/PTC}$ BOECs (Fig. 4Civ), a similar pattern was observed. However, as seen in Fig. 4B, in BOECs heterozygous for *ENG* c.1306C>T (Q436X) or the more proximal *ENG* c.277C>G (R93X), immature *ENG* protein remained the more pronounced species at 1hr, with traces still present at 2hr (Fig. 4Cv,vi).

Potential origins of aberrant protein

Two alternate mechanisms generate aberrant proteins that could be indistinguishable by pulse chase from the wildtype protein – PTC read-through (generating full length protein with a missense substitution), and ribosomal re-initiation (generating shorter, amino-truncated proteins that would also escape detection in BOECs using available methods).[5,6] As detailed in Supplementary Table 7, no PTC had features to suggest strong PTC read-through enhancement, though all PTC contexts were in support of ribosomal re-initiation. Further, since PTC readthrough is a therapeutic strategy[33,54] and no HHT pathogenic missense substitutions are described at these variants (Supplementary Table 7), if high level natural read-through were the mechanism, the *ENG*$^{+/PTC}$ BOECs with highest PTC persistence would be predicted to resemble control BOECs more than other HHT BOECs. Instead, as shown by the global analyses in Fig. 1A, *ENG*$^{+/PTC}$ BOECs resembled control BOECs less than other HHT BOECs (Fig. 1A). Thus the BOECS with the least efficient degradation of PTC transcripts were the most distinct (Fig. 5A).

GINI genes also distinguish ENG$^{+/PTC}$ BOECS

Our final analyses were intended simply to explore further evidence for differences between *ENG*$^{+/PTC}$ and other HHT BOECs. We have previously shown that for human monocyte rRNA-depleted RNASeq libraries, housekeeper normalisation using the 25 genes with GINI coefficients (GC) <0.15[39,40] reduced control variability and that the relative ranking of these genes was similar across 24 donor-treatment combinations,[55] in keeping with data from cell lines indicating these are the least variant of all human transcripts.[39,40] In contrast, our preliminary studies using the least variant genes in the BOECs had identified apparent discrepancies.[56]

We therefore performed unsupervised hierarchical clustering for all 25 genes with GC<0.15 (Fig. 5B). Unsupervised multivariate analyses of these 'housekeeper' genes ranked by relative expression across the 16 BOEC cultures again demarcated the 4 cultures of *ENG*$^{+/-}$ BOECs (Fig. 5C). The unsupervised analyses were performed in the opposite direction, and expected to confirm the observations observed in HHT-derived monocytes without PTCs.[55] Instead, across the HHT BOEC cultures harboring PTCs, there was marked variability in GINI gene respective rankings (Fig. 5D).

PTC associations with cellular stress phenotypes.

Within BOECs, for the least variant (GC 0.11-0.12) genes, the respective rankings of *RBM45* and *NXF1* were seen to switch between the BOECs with 14% and 11% PTC persistence (Fig. 5D). *RBM45* encodes RNA binding protein 45 (RBM45) that binds to m^{6}A-modified mRNAs[57] and forms reversible nuclear inclusions during stress,[58] while *NXF1* encodes the nuclear RNA export receptor for bulk (homeostatic) export of mRNAs, dissociated from mRNAs under stress.[59,60] Thus, the RNASeq findings for these previously considered near-invariant genes would be consistent with the highest PTC-expressing BOEC cultures exhibiting a stressed phenotype (Fig. 6).

DISCUSSION

Frameshift and nonsense mutations are common causes of genetic disease, and their PTC-containing transcripts are often thought either to be fully degraded by NMD, or to persist to generate proteins truncated at the PTC. Using patient-derived blood outgrowth endothelial cells, we showed neither assumption to be correct. Specifically, we demonstrated that a fraction of PTC-containing transcripts are retained in spliced RNA transcripts at differing proportions between replicate cultures from the same donor. While genes differentially expressed between patient and control endothelial cells were consistent with pathway-specific changes due to these loss-of-function variants, RNASeq profiles also indicated a second, pathway-independent set of differentially expressed genes related to ubiquitin protein degradation pathways. BOECs with greatest PTC persistence were discriminated in unsupervised hierarchical clustering, and in a state

displaying a stress-compatible switch in normally invariant low GINI coefficient genes. In other words, PTC-generating mutations were associated with a second phenotype beyond the conventional pathway-specific perturbations that have been the focus of models for human genetic disease.

Study strengths include the integration of recent developments in RNA biology with molecular medicine, and study design elements that focused on variant examination in endogenous cells from carefully selected donors. Confidence in discovery analyses was enhanced by single cell qRT-PCR validations; restricting analyses to genes with low intra-replicate variability; focusing on findings robust to different analytic methods; and using exacting pulse chase characterization of primary, non-transformed endothelial cells in as close to the natural state as possible. The main weakness is the relatively small number of BOEC datasets examined, although this has been a more rigorous evaluation across all 3 major HHT genotypes in a way that has not been previously performed due to the rarity of the *SMAD4*$^{+/PTC}$ genotype in HHT.[11,66] Another study limitation is that it does not distinguish processes generically associated with a PTC, from processes that are specific to HHT cells. Although unlikely given the nature of the RNA, protein and pathway findings, we cannot exclude the changes in some way being augmented in HHT BOECs.

Detailed mechanistic dissection needs to be the subject of future work, incorporating the latest experimental data where a reporter system identified a ubiquitin pathway for nascent protein degradation from PTC containing mRNAs.[48] That said, the current data already support testable models that may solve and facilitate mitigation of previously unexplained phenotypic differences:

- First, while all HHT BOECs displayed differential gene expression profiles supporting pathway-specific changes associated with disease causation[11-28] and isopeptide bond/ubiquitin-like conjugation relevant to post translational protein degradation, it was *ENG*$^{+/PTC}$ BOECs with the highest PTC burdens that displayed the most distinctive transcriptomes.
- Second, there was marked variability in persistence of PTCs between replicate cultures from the same donor. The cellular state may not have been examined previously as shown by the distinction using normally stably expressed genes, and points to a potential cut off at between 11-14% PTC

persistence when two of the most invariant (GC<0.13) housekeeper genes switched expression to a pattern more in-keeping with stressed cells.

- The data point to the *ENG*$^{+/PTC}$ variants being more detrimental in BOECs. It is not possible to say whether *ENG*$^{+/PTC}$ mutant monomers or heterodimers are more challenging for protein degradation pathways than other HHT mutant proteins (it has long been known that many HHT mutant, non-functional proteins are retained in the endoplasmic reticulum for degradation.[18-23,67]) Homeostatic mechanisms may simply be saturated by excessive production of mutant proteins, irrespective of their nature. Consistent with this, the reversal of *RBM45* and *NXF1* expression patterns was observed with PTC persistence of ≥14% in one of *SMAD4*$^{+/PTC}$ BOEC cultures in addition to all 4 *ENG*$^{+/PTC}$ BOEC cultures.

While it is plausible that varying PTC retention rates of unknown origin were the cause of divergent cellular phenotypes in the BOECs, we favor an alternate, testable interpretation. It is recognized that through the integrated stress response and phosphorylation of eIF2α, a variety of cellular stresses suppress protein translation and NMD and for transient stress, that this suppression is reversible.[61-64] Thus, stress inhibition of NMD would increase the number of PTC-harboring transcripts, and cell burden[48] consequences when protein translation resumed.

In summary, we provide evidence that in blood outgrowth endothelial cells (BOECs), operating alongside pathway-specific consequences due to their loss-of-function nature, PTC-generating variants provide a parallel and novel genotypic cause of cellular variability. Precise mechanisms, contributions from environmental stressors, and relevance to other PTCs will need to be the subject of future work.

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*). MAA was supported by the National Institutes of Health (grant R35HL140019). 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. We thank the study participants for their willing cooperation in these studies. 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

Author contributions:

MBH, MAA and CLS devised the BOEC strategy. MBH, DP and MAA performed BOEC cultures. DP prepared RNA, co-designed pulse chase experiments, performed pulse chase experiments, and provided blots for Fig. 4B/C. AB performed early RNASeq analyses, devised the CV10 approach and contributed to sc qRT-PCR design. PCG and MN designed and performed sc qRT-PCR, bulk flow cytometric analyses, and provided data for Supplementary Table 1 and Supplementary Fig. 2A/B. SJM advised on the integrated stress response. CLS reviewed and phenotyped patients, recruited patients, contributed to sc qRT-PCR and pulse chase experimental design, performed data analyses generating all other Figures/Tables, and wrote the manuscript. All authors have reviewed and approved the final manuscript.

Disclosure and Competing interests: Authors declare that they have no competing interests.

Data availability:

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 participants' informed consent. Access to these data will only be granted in line with that consent, subject to approval by the project ethics board and under a Data Sharing Agreement. Blood outgrowth endothelial cells used in this research were collected subject to the informed consent of the participants. Access will only be granted in line with that consent, subject to approval by the project

REFERENCES

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

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

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

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

5 Hoek TA, Khuperkar D, Lindeboom RGH, Sonneveld S, Verhagen BMP, Boersma S, Vermeulen M, Tanenbaum ME. Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay. Mol Cell 2019;75(2): 324-339.

6 Embree CM, Abu-Alhasan R, Singh G. Features and factors that dictate if terminating ribosomes cause or counteract nonsense-mediated mRNA decay. J Biol Chem 2022;298(11):102592.

7 Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. 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

8 Human Genome Variation Society. Sequence variant nomenclature. Current recommendations. Available at https://varnomen.hgvs.org/

9 Shovlin CL, Buscarini E, Sabbà C, Mager HJ, Kjeldsen AD, Pagella F, Sure U, Ugolini S, Torring PM, Suppressa P, Rennie C, Post MC, Patel MC, Nielsen TH, Manfredi G, Lenato GM, Lefroy D, Kariholu U, Jones B, Fialla AD, Eker OF, Dupuis O, Droege F, Coote N, Boccardi E, Alsafi A, Alicante S, Dupuis-Girod S. The European Rare Disease Network for HHT frameworks for management of hereditary haemorrhagic telangiectasia in general and speciality care. Eur J Med Genet 2022;65(1):104370.

10 Faughnan ME, Mager JJ, Hetts SW, Palda VA, Lang-Robertson K, Buscarini E, Deslandres E, Kasthuri RS, Lausman A, Poetker D, Ratjen F, Chesnutt MS, Clancy M, Whitehead KJ, Al-Samkari H, Chakinala M, Conrad M, Cortes D, Crocione C, Darling J, de Gussem E, Derksen C, Dupuis-Girod S, Foy P, Geisthoff U, Gossage JR, Hammill A, Heimdal K, Henderson K, Iyer VN, Kjeldsen AD, Komiyama M, Korenblatt K, McDonald J, McMahon J, McWilliams J, Meek ME, Mei-Zahav M, Olitsky S, Palmer S, Pantalone R, Piccirillo JF, Plahn B, Porteous MEM, Post MC, Radovanovic I, Rochon PJ, Rodriguez-Lopez

J, Sabba C, Serra M, Shovlin C, Sprecher D, White AJ, Winship I, Zarrabeitia R. Second International Guidelines for the Diagnosis and Management of Hereditary Hemorrhagic Telangiectasia. Ann Intern Med. 2020;173(12):989-1001.

11 Shovlin CL, Simeoni I, Downes K, Frazer ZC, Megy K, Bernabeu-Herrero ME, Shurr A, Brimley J, Patel D, Kell L, Stephens J, Turbin IG, Aldred MA, Penkett CJ, Ouwehand WH, Jovine L, Turro E. Mutational and phenotypic characterization of hereditary hemorrhagic telangiectasia. Blood. 2020;136(17):1907-1918.

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

13 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.

14 Govani FS, Giess A, Mollet IG, Begbie ME, Jones MD, Game L, Shovlin CL.. Directional next-generation RNA sequencing and examination of premature termination codon mutations in endoglin/hereditary haemorrhagic telangiectasia. Mol Syndromol. 2013;4(4):184-96.

15 Snellings DA, Gallione CJ, Clark DS, Vozoris NT, Faughnan ME, Marchuk DA. Somatic Mutations in Vascular Malformations of Hereditary Hemorrhagic Telangiectasia Result in Bi-allelic Loss of ENG or ACVRL1. Am J Hum Genet. 2019;105(5):894-906.

16 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;156(3):911-23.

17 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;31(11):2653-60.

18 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.

19 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.

20 Paquet ME, Pece-Barbara N, Vera S, Cymerman U, Karabegovic A, Shovlin C, Letarte M. 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.

21 Ali BR, Ben-Rebeh I, John A, Akawi NA, Milhem RM, Al-Shehhi NA, Al-Ameri MM, Al-Shamisi SA, Al-Gazali L. Endoplasmic reticulum quality control is involved in the mechanism of endoglin-mediated hereditary haemorrhagic telangiectasia. PLoS One. 2011;6(10):e26206.

22 Hume AN, John A, Akawi NA, Al-Awadhi AM, Al-Suwaidi SS, Al-Gazali L, Ali BR. Retention in the endoplasmic reticulum is the underlying mechanism of some hereditary haemorrhagic telangiectasia type 2 ALK1 missense mutations. Mol Cell Biochem. 2013 Jan;373(1-2):247-57.

23 Alaa El Din F, Patri S, Thoreau V, Rodriguez-Ballesteros M, Hamade E, Bailly S, Gilbert-Dussardier B, Abou Merhi R, Kitzis A. Functional and splicing defect analysis of 23 ACVRL1 mutations in a cohort of patients affected by Hereditary Hemorrhagic Telangiectasia. PLoS One. 2015;10(7):e013211124

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

25 Park SO, Wankhede M, Lee YJ, Choi EJ, Fliess N, Choe SW, Oh SH, Walter G, Raizada MK, Sorg BS, Oh SP. Real-time imaging of de novo arteriovenous malformation in a mouse model of hereditary hemorrhagic telangiectasia. J Clin Invest. 2009;119(11):3487-96.

26 Jin Y, Muhl L, Burmakin M, Wang Y, Duchez AC, Betsholtz C, Arthur HM, Jakobsson L. Endoglin prevents vascular malformation by regulating flow-induced cell migration and specification through VEGFR2 signalling. Nat Cell Biol;2017;19:639–652.

27 Ola R, Künzel SH, Zhang F, Genet G, Chakraborty R, Pibouin-Fragner L, Martin K, Sessa W, Dubrac A, Eichmann A. SMAD4 Prevents Flow Induced Arteriovenous Malformations by Inhibiting Casein Kinase 2. Circulation. 2018;138:2379–2394.

28 Garrido-Martín EM, Blanco FJ, Roquè M, Novensà L, Tarocchi M, Lang UE, Suzuki T, Friedman SL, Botella LM, Bernabéu C. 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

29 Shovlin CL. Supermodels and disease: insights from the HHT mice. J Clin Invest. 1999;104(10):1335-6.

30 Tual-Chalot S, Oh SP, Arthur HM. Mouse models of hereditary hemorrhagic telangiectasia: recent advances and future challenges. Front Genet. 2015;6:25.

31 Seghers L, de Vries MR, Pardali E, Hoefer IE, Hierck BP, ten Dijke P, Goumans MJ, Quax PH. Shear induced collateral artery growth modulated by endoglin but not by ALK1. J Cell Mol Med. 2012;16:2440–2450.

32 Lebrin F, Goumans MJ, Jonker L, Carvalho RL, Valdimarsdottir G, Thorikay M, Mummery C, Arthur HM, ten Dijke P. Endoglin promotes endothelial cell proliferation and TGF-beta/ALK1 signal transduction. EMBO J. 2004;23:4018–4028.

33 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.

34 Mollet IG, Patel D, Govani FS, Giess A, Paschalaki K, Periyasamy M, Lidington EC, Mason JC, Jones MD, Game L, Ali S, Shovlin CL. Low Dose Iron Treatments Induce a DNA Damage Response in Human Endothelial Cells within Minutes. PLoS One. 2016;11(2):e0147990.

35 Shovlin CL, Angus G, Manning RA, Okoli GN, Govani FS, Elderfield K, Birdsey GM, Mollet IG, Laffan MA, Mauri FA. Endothelial cell processing and alternatively spliced transcripts of factor VIII: potential implications for coagulation cascades and pulmonary hypertension. PLoS One. 2010;5(2):e9154.

36 Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22(3):568-76.

37 Noseda M, Harada M, McSweeney S, Leja T, Belian E, Stuckey DJ, Abreu Paiva MS, Habib J, Macaulay I, de Smith AJ, al-Beidh F, Sampson R, Lumbers RT, Rao P, Harding SE, Blakemore AI, Jacobsen SE, Barahona M, Schneider MD. PDGFRα demarcates the cardiogenic clonogenic Sca1+ stem/progenitor cell in adult murine myocardium. Nat Commun. 2015;6:6930

38 Massaia A, Chaves P, Samari S, Miragaia RJ, Meyer K, Teichmann SA, Noseda M. Single Cell Gene Expression to Understand the Dynamic Architecture of the Heart. Front Cardiovasc Med. 2018;5:167.

39 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.

40 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.

41 Reed GF, Lynn F, Meade BD. Use of coefficient of variation in assessing variability of quantitative assays. Clin Diagn Lab Immunol. 2002;9(6):1235-9.

42 Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11(10):R106.

43 Evans C, Hardin J, Stoebel DM. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. Brief Bioinform. 2018;19(5):776-792.

44 Gene Ontology Consortium; Blake JA, Dolan M, Drabkin H, Hill DP, Li N, Sitnikov D, Bridges S, Burgess S, Buza T, McCarthy F, Peddinti D, Pillai L, Carbon S, Dietze H, Ireland A, Lewis SE, Mungall CJ, Gaudet P, Chrisholm RL, Fey P, Kibbe WA, Basu S, Siegele DA, McIntosh BK, Renfro DP, Zweifel AE, Hu JC, Brown NH, Tweedie S, Alam-Faruque Y, Apweiler R, Auchinchloss A, Axelsen K, Bely B, Blatter M-, Bonilla C, Bouguerleret L, Boutet E, Breuza L, Bridge A, Chan WM, Chavali G, Coudert E, Dimmer E, Estreicher A, Famiglietti L, Feuermann M, Gos A, Gruaz-Gumowski N, Hieta R, Hinz C, Hulo C, Huntley R, James J, Jungo F, Keller G, Laiho K, Legge D, Lemercier P, Lieberherr D, Magrane M, Martin MJ, Masson P, Mutowo-Muellenet P, O'Donovan C, Pedruzzi I, Pichler K, Poggioli D, Porras Millán P, Poux S, Rivoire C, Roechert B, Sawford T, Schneider M, Stutz A, Sundaram S, Tognolli M, Xenarios I, Foulgar R, Lomax J, Roncaglia P, Khodiyar VK, Lovering RC, Talmud PJ, Chibucos M, Giglio MG, Chang H-, Hunter S, McAnulla C, Mitchell A, Sangrador A, Stephan R, Harris MA, Oliver SG, Rutherford K, Wood V, Bahler J, Lock A, Kersey PJ, McDowall DM, Staines DM, Dwinell M, Shimoyama M, Laulederkind S, Hayman T, Wang S-, Petri V, Lowry T, D'Eustachio P, Matthews L, Balakrishnan R, Binkley G, Cherry JM, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hitz BC, Hong EL, Karra K, Miyasato SR, Nash RS, Park J, Skrzypek MS, Weng S, Wong ED, Berardini TZ, Huala E, Mi H, Thomas PD, Chan J, Kishore R, Sternberg P, Van Auken K, Howe D, Westerfield M. Gene Ontology annotations and resources. Nucleic Acids Res. 2013 Jan;41(Database issue):D530-5.

45 Gene Ontology Consortium; Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL, Hill DP, Lee R, Mi H, Moxon S, Mungall CJ, Muruganugan A, Mushayahama T, Sternberg PW, Thomas PD, Van Auken K, Ramsey J, Siegele DA, Chisholm RL, Fey P, Aspromonte MC, Nugnes MV, Quaglia F, Tosatto S, Giglio M, Nadendla S, Antonazzo G, Attrill H, Dos Santos G, Marygold S, Strelets V, Tabone CJ, Thurmond J, Zhou P, Ahmed SH, Asanitthong P, Luna Buitrago D, Erdol MN, Gage MC, Ali Kadhum M, Li KYC, Long M, Michalak A, Pesala A, Pritazahra A, Saverimuttu SCC, Su R, Thurlow KE, Lovering RC, Logie C, Oliferenko S, Blake J, Christie K, Corbani L, Dolan ME, Drabkin HJ, Hill DP, Ni L, Sitnikov D, Smith C, Cuzick A, Seager J, Cooper L, Elser J, Jaiswal P, Gupta P, Jaiswal P, Naithani S, Lera-Ramirez M, Rutherford K, Wood V, De Pons JL, Dwinell MR, Hayman GT, Kaldunski ML, Kwitek AE, Laulederkind SJF, Tutaj MA, Vedi M, Wang SJ, D'Eustachio P, Aimo L, Axelsen K, Bridge A, Hyka-Nouspikel N, Morgat A, Aleksander SA, Cherry JM, Engel SR, Karra K, Miyasato SR, Nash RS, Skrzypek MS, Weng S, Wong ED, Bakker E, Berardini TZ, Reiser L, Auchincloss A, Axelsen K, Argoud-Puy G, Blatter MC, Boutet E, Breuza L, Bridge A, Casals-Casas C, Coudert E, Estreicher A, Livia Famiglietti M, Feuermann M, Gos A, Gruaz-Gumowski N, Hulo

C, Hyka-Nouspikel N, Jungo F, Le Mercier P, Lieberherr D, Masson P, Morgat A, Pedruzzi I, Pourcel L, Poux S, Rivoire C, Sundaram S, Bateman A, Bowler-Barnett E, Bye-A-Jee H, Denny P, Ignatchenko A, Ishtiaq R, Lock A, Lussi Y, Magrane M, Martin MJ, Orchard S, Raposo P, Speretta E, Tyagi N, Warner K, Zaru R, Diehl AD, Lee R, Chan J, Diamantakis S, Raciti D, Zarowiecki M, Fisher M, James-Zorn C, Ponferrada V, Zorn A, Ramachandran S, Ruzicka L, Westerfield M. The Gene Ontology knowledgebase in 2023. Genetics. 2023;224(1):iyad031.

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

47 UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51(D1):D523-D531.

48 Inglis AJ, Guna A, Gálvez-Merchán Á, Pal A, Esantsi TK, Keys HR, Frenkel EM, Oania R, Weissman JS, Voorhees RM. Coupled protein quality control during nonsense-mediated mRNA decay. J Cell Sci. 2023;136(10):jcs261216.

49 National Center for Biotechnology Information, U.S. National Library of Medicine: Genome Reference Consortium Human Build 38, available at https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/.

50 Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002;12(6):996-1006.

51 Nassar LR, Barber GP, Benet-Pagès A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN, Hinrichs AS, Lee BT, Lee CM, Muthuraman P, Nguy B, Pereira T, Nejad P, Perez G, Raney BJ, Schmelter D, Speir ML, Wick BD, Zweig AS, Haussler D, Kuhn RM, Haeussler M, Kent WJ. The UCSC Genome Browser database: 2023 update. Nucleic Acids Res. 2023;51(D1):D1188-D1195.

52 Pichuantes S, Vera S, Bourdeau A, Pece N, Kumar S, Wayner EA, Letarte M. Mapping epitopes to distinct regions of the extracellular domain of endoglin using bacterially expressed recombinant fragments. Tissue Antigens. 1997;50(3):265-76.

53 Saito T, Bokhove M, Croci R, Zamora-Caballero S, Han L, Letarte M, de Sanctis D, Jovine L. Structural Basis of the Human Endoglin-BMP9 Interaction: Insights into BMP Signaling and HHT1. Cell Rep. 2017;19(9):1917-1928.

54 Dabrowski M, Bukowy-Bieryllo Z, Zietkiewicz E (2018) Advances in therapeutic use of a drug-stimulated translational readthrough of premature termination codons. Mol. Med. 2018;24, 25.

55 Xiao S, Kai Z, Murphy D, Li D, Murphy D, Li D, Patel D, Bielowka A, Bernabeu-Herrero ME, Abdulmogith A, Mumford AD, Westbury S, Aldred MA, Vargesson N, Caulfield MJ, Genomics England Research Consortium, Shovlin CL. Functional filter for whole genome sequence data identifies stress

impact, non-coding, alternate polyadenylation site variants >5kb from coding DNA. medRxiv 2023. www.medrxiv.org/content/10.1101/2023.05.10.23289736v1. 14 May 2023.

56 Bernabeu-Herrero ME, Patel D, Bielowka A, Srikaran S, Chaves Guerrero P, Govani FS, Mollet IG, Noseda M, Aldred MA, Shovlin CL. Transcriptional and nonsense decay signatures in heterozygous HHT patient-derived blood outgrowth endothelial cells. BioRxiv 2021, www.biorxiv.org/content/10.1101/2021.12.05.471269v1. 06 Dec 2021.

57 Choi SH, Flamand MN, Liu B, Zhu H, Hu M, Wang M, Sewell J, Holley CL, Al-Hashimi HM, Meyer KD. RBM45 is an m6A-binding protein that affects neuronal differentiation and the splicing of a subset of mRNAs. Cell Rep. 2022;40(9):111293.

58 Collins M, Li Y, Bowser R. RBM45 associates with nuclear stress bodies and forms nuclear inclusions during chronic cellular stress and in neurodegenerative diseases. Acta Neuropathol Commun. 2020;8(1):91.

59 Scott DD, Aguilar LC, Kramar M, Oeffinger M. It's Not the Destination, It's the Journey: Heterogeneity in mRNA Export Mechanisms. Adv Exp Med Biol. 2019;1203:33-81.

60 Wende W, Friedhoff P, Sträßer K. Mechanism and Regulation of Co-transcriptional mRNP Assembly and Nuclear mRNA Export. Adv Exp Med Biol. 2019. 1203:1-31.

61 Costa-Mattioli M, Walter P. The integrated stress response: From mechanism to disease. Science. 2020;368(6489):eaat5314.

62 Usuki F, Yamashita A, Fujimura M. Environmental stresses suppress nonsense-mediated mRNA decay (NMD) and affect cells by stabilizing NMD-targeted gene expression. Sci Rep. 2019;9(1):1279.

63 Tweedie A, Nissan T. Hiding in Plain Sight: Formation and Function of Stress Granules During Microbial Infection of Mammalian Cells. Front Mol Biosci. 2021;8:647884

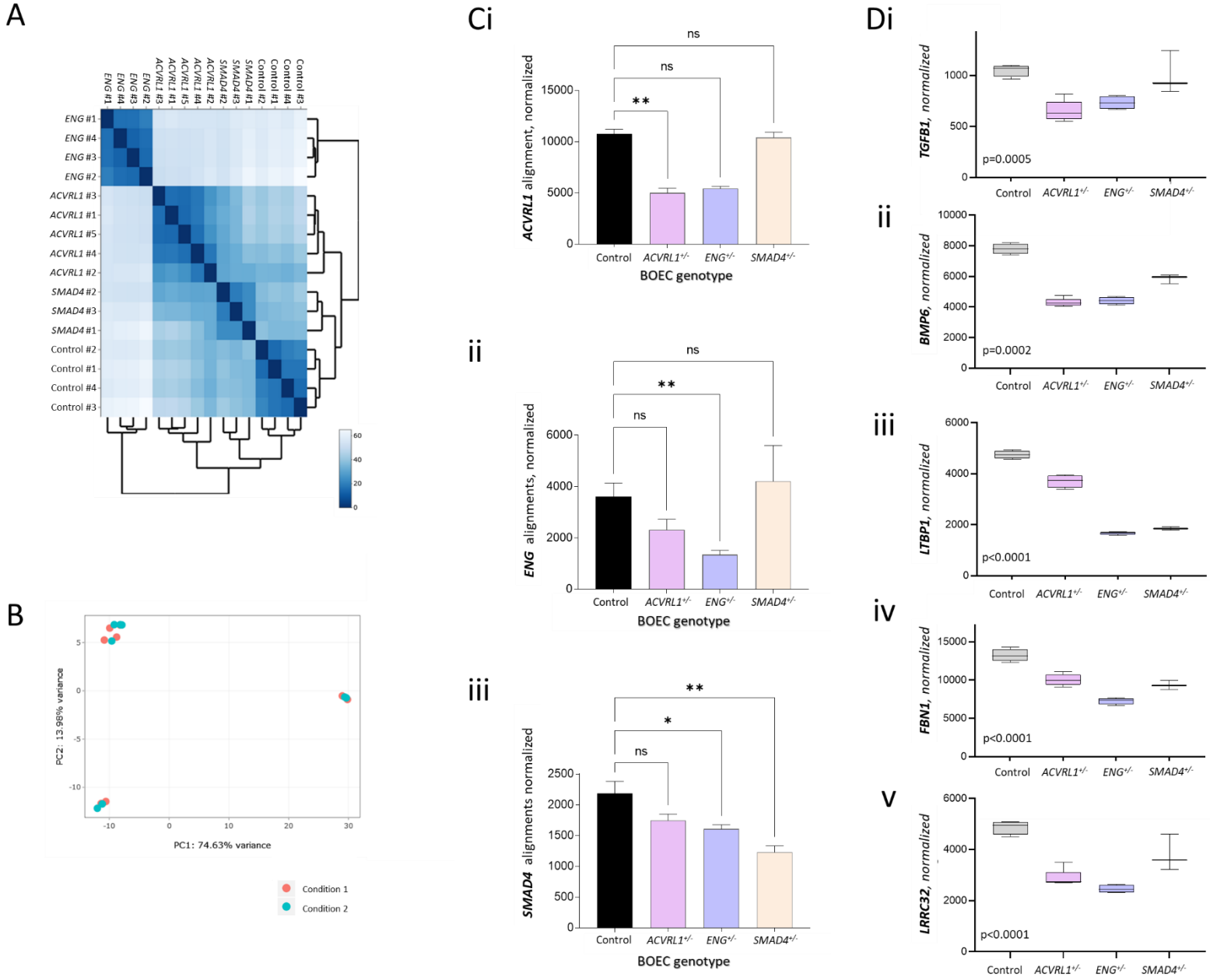
64 Pakos-Zebrucka K, Koryga I, Mnich K, Ljujic M, Samali A, Gorman AM. The integrated stress response. EMBO Rep. 2016;17(10):1374-1395.

65 Shovlin CL, Vizcaychipi MP. Vascular inflammation and endothelial injury in SARS-CoV-2 infection: The overlooked regulatory cascades implicated by the ACE2 gene cluster. QJM. 2020 Aug 10:hcaa241. doi: 10.1093/qjmed/hcaa241. Epub ahead of print.

66 ARUP Laboratories. Mutation Databases, March 2021 update available at https://arup.utah.edu/database.

67 Ricard N, Bidart M, Mallet C, Lesca G, Giraud S, Prudent R, Feige JJ, Bailly S. Functional analysis of the BMP9 response of ALK1 mutants from HHT2 patients: a diagnostic tool for novel ACVRL1 mutations. Blood 2010;116(9):1604–1612.

Fig 1.


A
ENG #1
ENG #4
ENG #3
ENG #2
ACVRL1 #3
ACVRL1 #1
ACVRL1 #5
ACVRL1 #4
ACVRL1 #2
SMAD4 #2
SMAD4 #3
SMAD4 #1
Control #2
Control #1
Control #4
Control #3
B
PC2: 13.98% variance
PC1: 74.63% variance
Condition 1
Condition 2
Ci
ACVRL1 alignment, normalized
ns
ns
**
Control
ACVRL1+/-
ENG+/-
SMAD4+/-
BOEC genotype
ii
ENG alignments, normalized
ns
**
ns
BOEC genotype
iii
SMAD4 alignments normalized
**
*
ns
BOEC genotype
Di
TGFB1, normalized
p=0.0005
ii
BMP6, normalized
p=0.0002
iii
LTBP1, normalized
p<0.0001
iv
FBN1, normalized
p<0.0001
v
LRRC32, normalized
p<0.0001

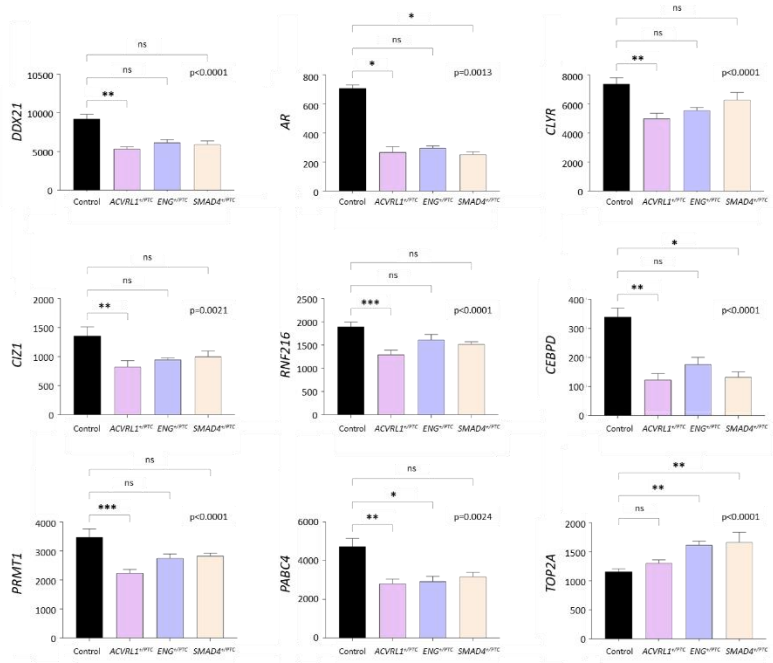
Global and disease pathway RNASeq findings in BOECs. A) Unsupervised Euclidean distance analyses for the 16 BOEC RNASeq datasets. When unblinded, donor replicates had clustered together. Note the cluster of 4 BOECs distinguished by greatest Euclidian distances were from genotype *ENG*$^{+/PTC}$. **B)** Unsupervised principal component analysis of the 16 BOEC RNASeq datasets based on the distance matrix where samples were projected to a 2D plane spanned by their first two principal components. An experimental covariate (10ng/ml BMP9 for 1 hour) was discernible only in control BOECs (top left). **C)** Normalized alignments to HHT gene transcripts ENG (ENSG00000106991), ACVRL1 (ENSG00000139567), and SMAD4 (ENSG00000141646) in control and HHT BOECs. Data from all samples per donor, error bars indicate mean and standard deviation,**$p<0.005$ by Dunn's test post Kruskal Wallis. **D)** Expression of genes encoding TGF-β/BMP ligands [**i)** *TGF*-β1, **ii)** *BMP6*], and extracellular matrix BMP/TGF-β sequestration and activation proteins [**iii)** *LTBP1*, **iv)** *FBN1*, **iv)** *POSTN* and **v)** *LRRC32*]. Expression was DeSeq2-normalized[42] using low GINI coefficient genes[39,40] that displayed least variability in BOECs (Supplementary Table 2, Supplementary Figure 4). Box plots show interquartile range, error bars minimum to maximum. P-values were calculated by Kruskal Wallis.

Fig 2.

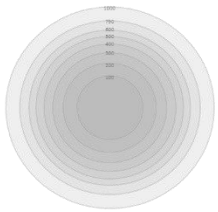
A

	ENRICHMENT		P VALUES		GENES IN CLUSTER
	CLUSTER	TERM	BASE	BENJAMINI	
i) STRINGENCY LEVEL 1: BONFERRONI p<0.05					
KW-1017~Isopeptide bond	1.05	3.17	0.0032	0.034	*TOP2A, AR, ACLY, CIZ1, RNF216, CEBPD, PRMT1, PABPC4, DDX21*
KW-0832~Ubl conjugation	1.05	2.48	0.0008	0.034	*TOP2A, AR, ACLY, CIZ1, RNF216, CEBPD, PRMT1, PABPC4, DDX21, MARK4*
ii) STRINGENCY LEVEL 2: BONFERRONI p<1.00:					
GO:0005739~mitochondrion	3.40	2.70	2.13E-06	6.10E-04	*CLPB, TIMM9, AKAP8, DDX21, TBRG4, GRPEL1, TOP1MT, BSG, LONP1, ME1, SLC25A23, DCPS, MDH2, UQCC1, IDH1, GBF1, PRDX6, LRPPRC, TNFRSF1A, VWA8, BCKDK, SLC25A37, SMDT1, NDUFS8, RPUSD3, ATG4B, PEMT, UBA1, UQCRC2*
KW-0809~Transit peptide	3.40	3.27	5.55E-04	0.010	*CLPB, MDH2, LRPPRC, TBRG4, VWA8, BCKDK, SMDT1, NDUFS8, GRPEL1, TOP1MT, RPUSD3, LONP1, UQCRC2*
KW-0496~Mitochondrion	3.40	2.15	0.0012	0.018	*CLPB, MDH2, UQCC1, TIMM9, DDX21, CISD2, LRPPRC, TBRG4, VWA8, BCKDK, SLC25A37, SMDT1, NDUFS8, GRPEL1, TOP1MT, RPUSD3, LONP1, PEMT, ATG4B, UBA1, UQCRC2, SLC25A23*
KW-1017~Isopeptide bond	2.37	1.79	0.0022	0.015	*TOP2A, CEBPD, DHX9, POGK, SRSF1, DCUN1D4, AKAP8, DDX21, KIF11, ACTG1, RNF216, CUX1, PSMD3, SPIN1, EIF4EBP1, RPL18, CIZ1, ZNF142, UBE2I, PRMT1, PABPC4, CBX1, ACLY, AR, ZEB2, BHLHE40, SNAI1, ATG4B, EIF3H*
KW-0832~Ubl conjugation	2.37	1.56	0.0050	0.026	*TOP2A, CEBPD, DHX9, POGK, SRSF1, DCUN1D4, AKAP8, DDX21, KIF11, RNF216, CUX1, PSMD3, ZC3H12A, SPIN1, EIF4EBP1, MARK4, RPL18, N4BP2, MAP4K2, CIZ1, ZNF142, UBE2I, PRMT1, PABPC4, CBX1, ACLY, AR, ZEB2, PER3, AXL, BHLHE40, SNAI1, ATG4B, EIF3H, UBA1, FAN1*

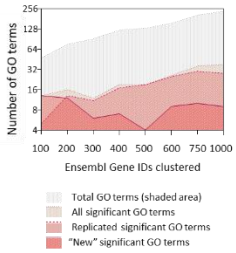
B



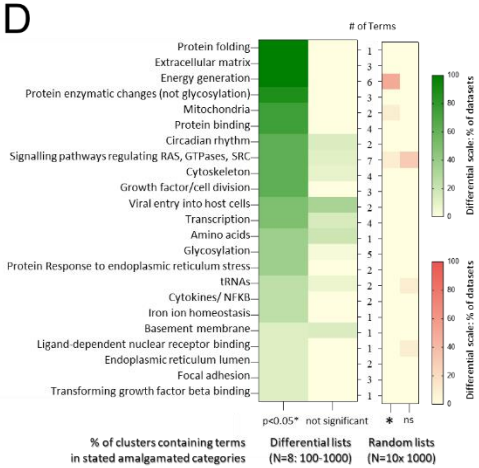
Ci



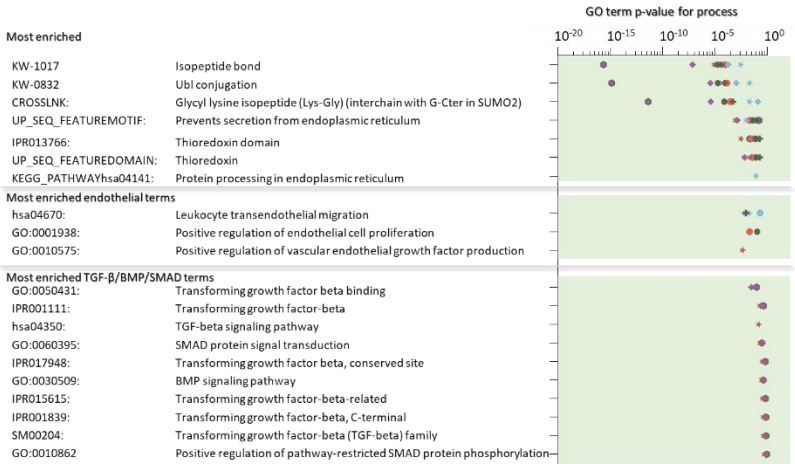
ii



D



E



Gene Ontology clustering of genes differentially expressed between HHT and control BOECs

A) All significant terms enriched by clustering genes differentially expressed to **i)** Bonferroni $p<0.05$ (N=30); **ii)** Bonferroni $p<1.00$ (N=161).

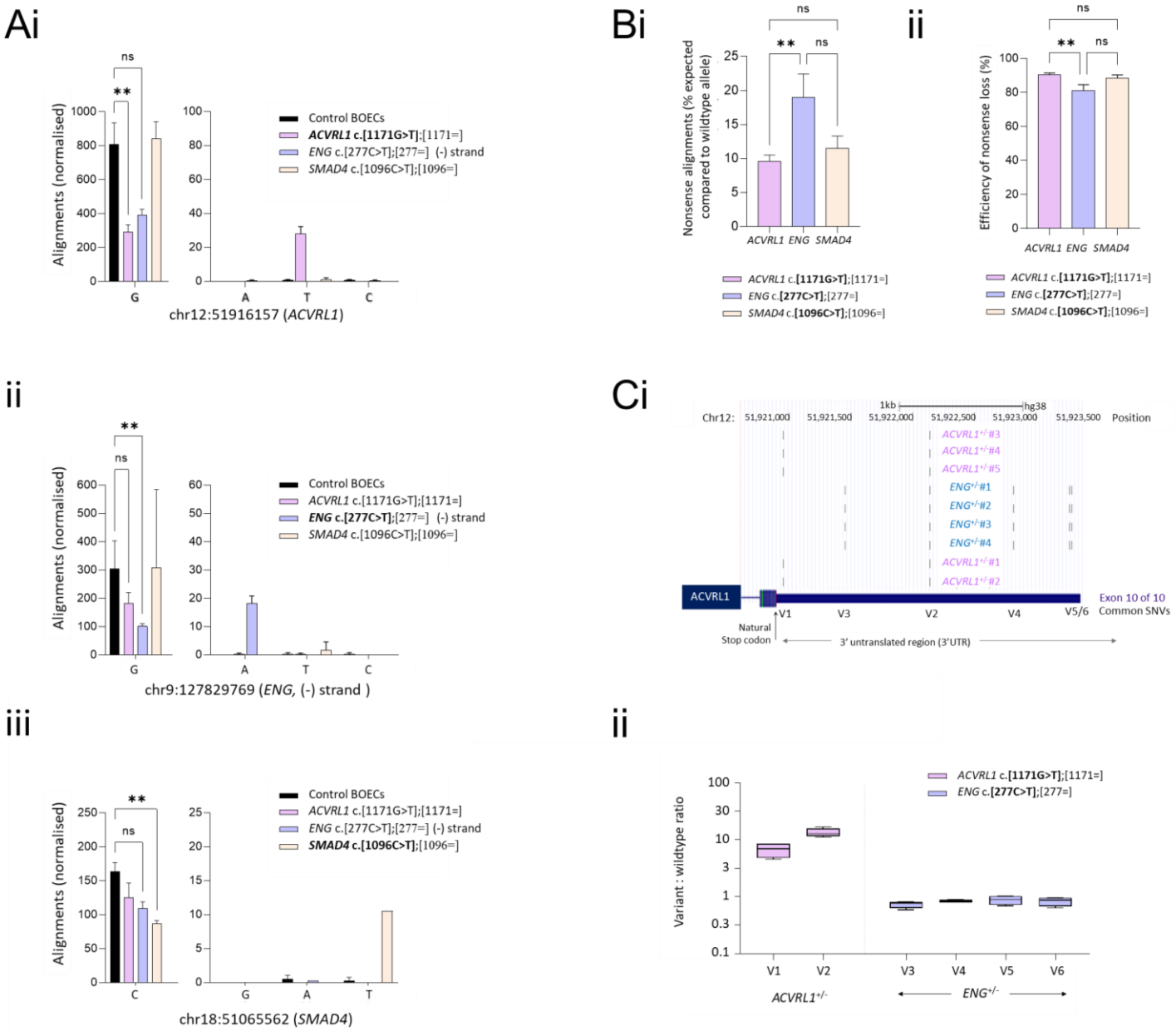
B) Intergenotype comparisons of expression for the 9 'Bonferroni $p<0.05$' genes that clustered to both isopeptide-bond and Ubl conjugation. Mean, standard devation and Kruskal Wallis p-values displayed; post-test pairwise p values by Dunn's test: *$p<0.05$, **$p<0.01$, ***$p<0.005$, ns not significant, $p\geq0.05$.

C) Broader, lower stringency clustering from the top ranked 100-1,000 genes by HHT differential expression (all in the 5,013 Ensembl gene IDs meeting an intraassay CV10.[41]). **i)** Overview–each of the top-sliced 100-1,000 genes (listed in Supplementary Table 4) are contained within the larger datasets. **ii)** Comparison of gene ontology (GO) term numbers[44] clustered significantly by the top 100–1,000 genes ranked by differential expression between HHT and control BOECs (details in Supplementary Table 5). The y axis indicates total GO-Direct terms; GO-Direct terms meeting significance; significant terms replicated in ≥1 other gene set, and 'new' clusters meeting significance that had not been identified in preceding, smaller datasets**.**

D) Heatmaps of amalgamated GO-term processes across the top 100-1,000 ranked HHT differential genes from the 5,013 Ensembl gene IDs meeting CV10 (green), and 10 random datasets of 1,000 genes (red, Supplementary Table 6). Left column: % of datasets containing GO terms meeting significance ($p<0.05$); right column: proportion of datasets where only a GO term not meeting significance had been clustered.

E) Specific terms enriched by clustering differentially expressed genes in each dataset. The lowest p-value for each cluster is displayed. Note consistency with the most enriched terms displayed in **C.**

Fig.3



Evidence for loss of nonsense alleles in human blood outgrowth endothelial cells (BOECs).
A) Quantitative metrics for BOEC RNASeq alignments to nonsense alleles in *ACVRL1*, *ENG* and *SMAD4* genomic loci in GRCh38/hg38[49] corresponding to known heterozygous pathogenic variants in source BOECs. Left: wildtype allele at **i)** chr12:51,916,158, **ii)** chr9:127,829,769, **iii)** chr18:51065563. Right: alternate alleles at genomic position at 10X scale. Colour-code key describes both alleles for each donor with the relevant nonsense donor allele indicated in bold. Mean and standard deviation displayed. Individual pairwise p-values (** p<0.005) calculated by Dunn's test post Kruskal Wallis (overall p-values p<0.0001 [*ACVRL1*], p=0.0002 [*ENG*], p=0.003 [*SMAD4*]).
B) Alignments to PTC allele in the HHT BOECs by genotype, colour coded as in **A)**. **i)** Percentage of total expected if equal to wildtype allele alignment expected for heterozygotes. **ii)** Percentage "loss" of total expected allele alignments**.**
C) Alignments to common single nucleotide variants (SNVs) in *ACVRL1* 3' UTR quantified by VarScan2.[36]
i) Genomic (GRCh38/hg38[49]) positions of six common single nucleotide variants (SNVs V1-V6) in BOECs from donors heterozygous for *ACVRL1* c.1171G>T, (p.Glu391X), or *ENG* c.277C>G, (p.Arg93X). Positions are illustrated by custom tracks uploaded to the University of California Santa Cruz (UCSC) Genome Browser.[50,51] **ii)** Variant: wildtype ratios in *ACVRL1*$^{+/PTC}$ and wildtype *ACVRL1*$^{+/+}$; *ENG*$^{+/PTC}$ BOECs. The higher ratios for V1 and V2 in *ACVRL1*$^{+/PTC}$ BOECs can be attributed to presence *in* cis with wildtype c.1171G, loss of the c.1171T inframe transcripts generating the PTC, and presence of the SNV in all (V1), or only longer (V2) *ACVRL1* 3'UTR transcripts. (See also Supplementary Fig. 7).

Fig. 4

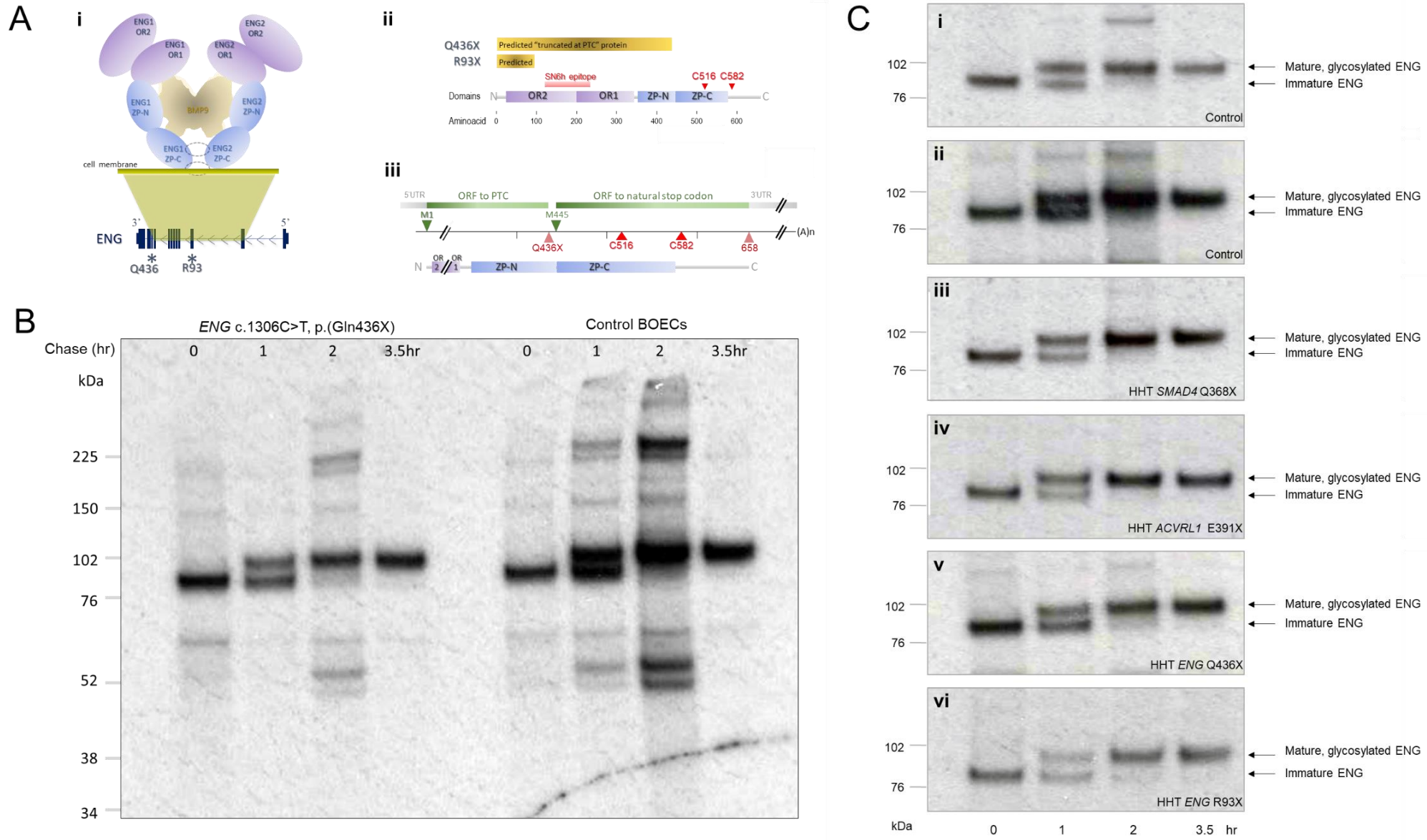

A
i
ENG1 OR2
ENG1 OR1
ENG2 OR2
ENG2 OR1
ENG1 ZP-N
ENG2 ZP-N
BMP9
ENG1 ZP-C
ENG2 ZP-C
cell membrane
ENG
3'
5'
Q436
R93
ii
Q436X
Predicted "truncated at PTC" protein
R93X
Predicted
SN6h epitope
C516 C582
Domains
N
OR2
OR1
ZP-N
ZP-C
C
Aminoacid
0
100
200
300
400
500
600
iii
5'UTR
ORF to PTC
ORF to natural stop codon
3'UTR
M1
M445
Q436X
C516
C582
658
(A)n
N
OR 2
OR 1
ZP-N
ZP-C
C
B
ENG c.1306C>T, p.(Gln436X)
Control BOECs
Chase (hr)
0
1
2
3.5hr
0
1
2
3.5hr
kDa
225
150
102
76
52
38
34
C
i
102
76
Control
Mature, glycosylated ENG
Immature ENG
ii
102
76
Control
Mature, glycosylated ENG
Immature ENG
iii
102
76
HHT SMAD4 Q368X
Mature, glycosylated ENG
Immature ENG
iv
102
76
HHT ACVRL1 E391X
Mature, glycosylated ENG
Immature ENG
v
102
76
HHT ENG Q436X
Mature, glycosylated ENG
Immature ENG
vi
102
76
HHT ENG R93X
Mature, glycosylated ENG
Immature ENG
kDa
0
1
2
3.5 hr

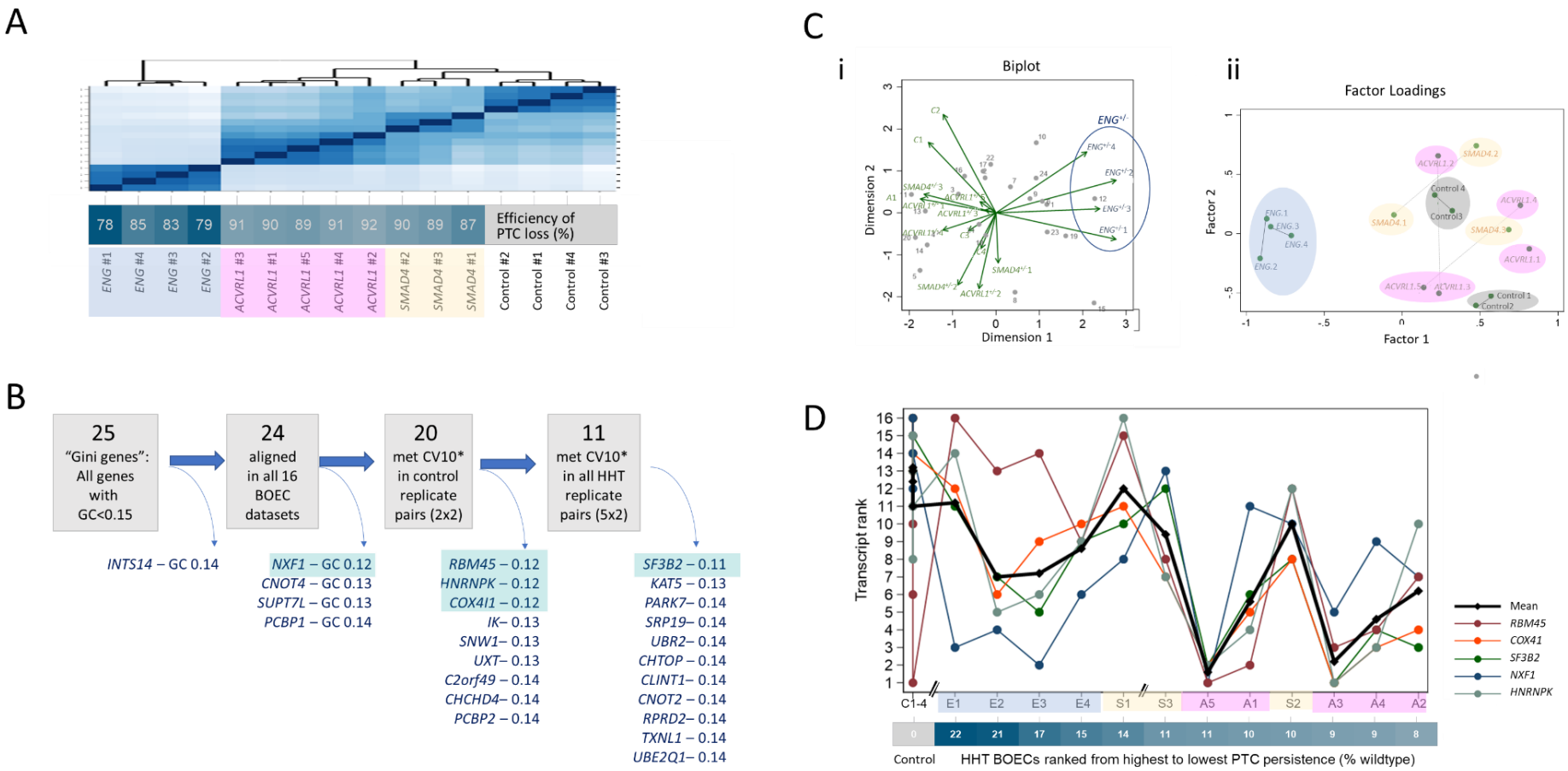
Fig. 4: Testing truncated ENG protein expression in normal and HHT endothelial cells

A) ENG cartoons: **i)** Extracellular ENG protein domains in BMP-binding heterodimeric conformation by genomic DNA origin, highlighting Q436 distal to orphan domains OR1 and OR2, but proximal to intermolecular dimerization cysteines C516 and C582 that are sited in and beyond the zona pellucida (ZP) domains.[53] **ii)** "Predicted" truncated ENG Q436X and R93X proteins by wildtype ENG domains and SN6h epitope[52,53]). **iii) ENG** Q436X open reading frames from methionine (M)1 to PTC, and M445 to natural stop codon.

B) Pulse chase experiments demonstrating generation of endoglin protein in control and *ENG* c.1306C>T, (p.Gln436X) BOEC lysates that were size fractionated on a reducing gel. Note no difference in patterns of sub full length proteins in HHT or control BOECs, and no evidence for a truncated protein product at an expected size based on truncation at the PTC.

C) Comparison of maturation patterns of full length endoglin protein following pulse chase in BOECs of **i-iv)** $ENG^{+/+}$ (wildtype) 'controls' and **v-vi)** $ENG^{+/PTC}$) BOECs: **i)** Control A; **ii)** Control B (same blot as Fig. 5B); **iii)** *SMAD4* c.1096C>T, (Q368X); **iv)** *ACVRL1* c.1171G>T, (E391X), and **v**) Exon 10 *ENG* c.1306C>T, (Q436X, same blot as Fig. 5B) ; **vi)** Exon 3 *ENG* c.277C>G, (R93X).

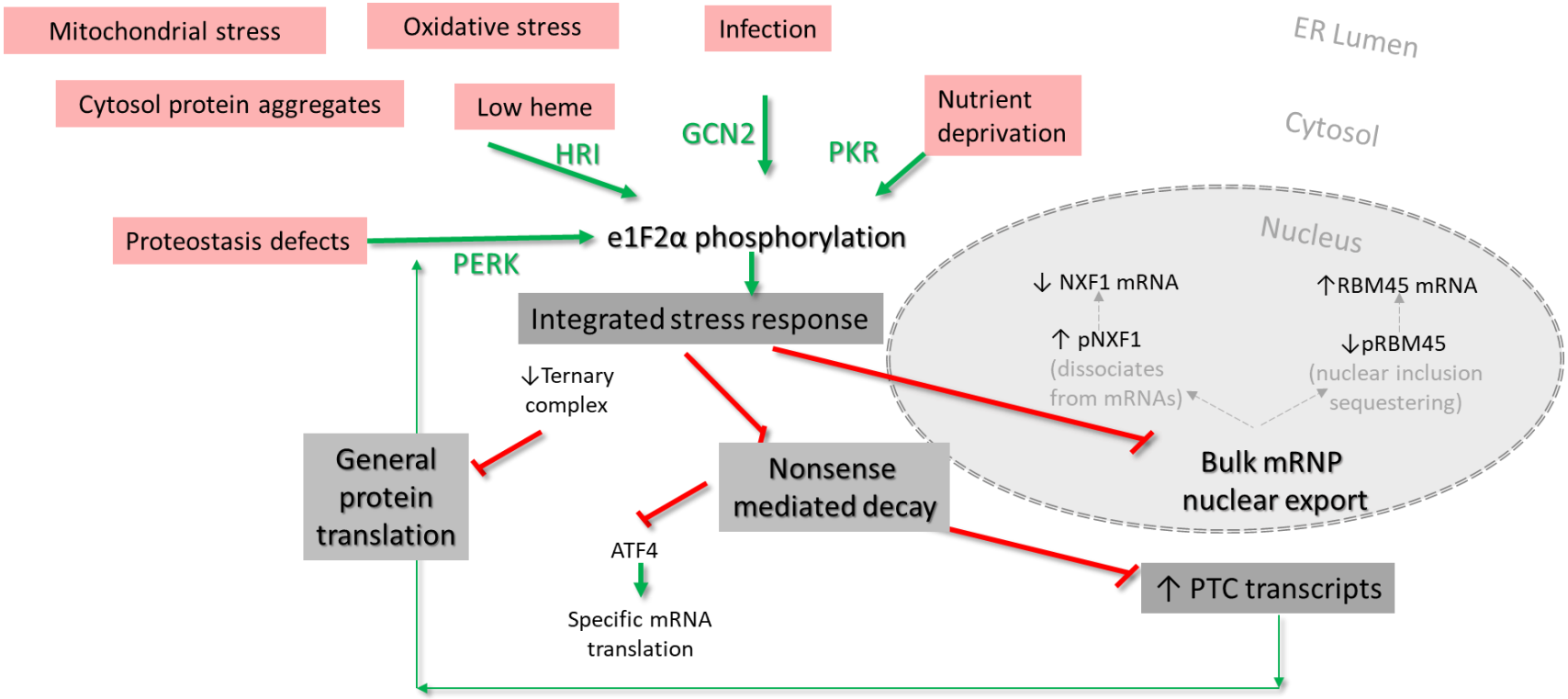
Fig. 5



PTC sequence and endothelial contexts

A) Unsupervised Euclidean distance analyses for the 16 BOEC RNASeq datasets: Note efficiency of PTC loss was not associated with the differences identified by the initial global data. **B)** BOEC expression of the 25 lowest Gini Coefficient (GC <0.15) genes[39,40] with GC<0.13 genes highlighted in blue (details in Supplementary Table 2). **C)** Unsupervised analyses of the 24 BOEC-expressed Gini genes ranked in 16 BOEC cultures. **i)** Multivariate analysis of genes (observations, grey dots) across the BOECs (variables, green arrows). Note separation of the 4 *ENG*$^{+/PTC}$ BOEC cultures**. ii)** Principal factor analysis on 16 BOEC datasets using ranks of the 24 GINI genes. The 15 retained factors and 120 evaluated parameters again distinguished the *ENG*$^{+/PTC}$ BOEC cultures (blue) from *ACVRL1*$^{+/PTC}$ (pink); *SMAD4*$^{+/PTC}$ (yellow) and control BOECs (grey). **D)** Respective ranks of the 5 lowest GC genes (highlighted blue in **B**) across BOEC cultures ranked by % PTC persistence and genotype (C: control, E: *ENG*$^{+/PTC}$, S: *SMAD4*$^{+/PTC}$, A: *ACVRL1*$^{+/PTC}$. Note switch between 14 and 11% persistence for *NXF1* (blue) and *RBM45* (brown).

Fig. 6



Extended model of stress and premature termination codon interaction

As described,[61-64] in the integrated stress response (ISR), diverse cytosolic stresses are sensed by four kinases, protein kinase RNA-like ER kinase (PERK), heme-regulated eIF2α (HRI) kinase, general control nonderepressible 2 (GCN2) kinase and protein kinase R (PKR). PERK, HRI, GCN2 and PKR can each phosphorylate the α subunit of eukaryotic initiation factor 2 (eIF2α) which inhibits assembly of the charged methionyl-initiator tRNA eIF2•GTP•methionyl-intiator tRNA ternary complex (TC) required for new protein synthesis. This inhibits translation-dependent nonsense mediating decay (NMD) which allows PTC-harbouring transcripts to persist. Genes regulated by upstream open reading frames use NMD suppression to increase full length protein translation by ribosomal reinitiation. Examples include *ATF4* encoding activating transcription factor 4 that acts as a master transcription factor during the ISR. Stress conditions also modify export factors preventing bulk mRNA export via nuclear RNA export factor 1 protein (pNXF1), and dissociating other RNA binding proteins such as the RNA binding motif protein 45 (pRBM45)[57,58] that aggregates in nuclear foci. Although GCN2 is classically considered to sense viral infection, a variety of bacterial and other microbial infections are now implicated.[63] Where the ISR is not transient and stress sustained, apoptosis results.[61-64] Note the only new data in the cartoon incorporate low GINI coefficient genes *NXF1*[59,60] and *RBM45*[57,58] where theoretical considerations[34,65] suggest mRNAs are likely to be higher to replace 'sequestered' RBM45 protein, and lower where free protein concentrations of NXF1 are increased.