Mutational and phenotypic characterisation of hereditary hemorrhagic telangiectasia

Claire L Shovlin[1,2]†, Ilenia Simeoni[3,4], Kate Downes[3,4], Zoe Frazer,[5] Karyn Megy[3,4], Maria Bernabeu Herrero[1], Abigail Shurr[1], Jennifer Brimley,[2] Dilip Patel,[1] Loren Kell[1,6], Jonathan Stephens[3,4], Isobel Turbin[1], Micheala A. Aldred[7], Christopher Penkett[3,4], Willem H Ouwehand[3,4], Luca Jovine[8] and Ernest Turro[3,4, 9]†

Affiliations: [1]NHLI Cardiovascular Sciences, Imperial College London, UK. [2] VASCERN HHT European Reference Centre and Respiratory Medicine, Imperial College Healthcare NHS Trust, London UK. [3]Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK; [4]NIHR BioResource - Rare Diseases, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge, UK; [5] Department of Surgery and Cancer, Imperial College London, UK; [6] University College London, UK [7] Indiana University School of Medicine, Indianapolis, IN, USA. [8]Department of Biosciences and Nutrition Karolinska Institutet, Huddinge, Sweden [9]Medical Research Council Biostatistics Unit, Cambridge Biomedical Campus, Cambridge CB2 0SR, UK.

Corresponding authors:

Claire L. Shovlin PhD FRCP, NHLI Cardiovascular Sciences, Imperial Centre for Translational and Experimental Medicine, Imperial College London, Hammersmith Campus, Du Cane Road, London W12 0NN, UK. Phone (44) 207 594 2725, Fax (44) 207 594 3654; email c.shovlin@imperial.ac.uk

Ernest Turro PhD, Department of Haematology, University of Cambridge, NHS Blood and Transplant, Long Rd, Cambridge CB4 3DF, UK. Phone (44) 1223588174; email et341@cam.ac.uk

ABSTRACT 255 words

TEXT 4103 words- required to address Reviewers' comments

Prior Abstract/Poster presentation: NONE

Prior Presubmission website listing: NONE

KEY POINTS

- We integrate systematic variant identification by deep high-throughput DNA sequencing with the application of the ACMG Guidelines for pathogenicity assignments.
- We show that detailed clinical information, statistical and structural modelling all contribute to a framework to better prognosticate and treat HHT.

KEY WORDS:

ACVRL1, activin receptor like kinase, bleeding scale, bone morphogenetic protein (BMP9), CD105, *ENG*, endoglin, *GDF2*, *SMAD4*

ABSTRACT

Hereditary hemorrhagic telangiectasia (HHT) is a vascular dysplasia inherited as an autosomal dominant trait. Care delivery is impeded by requirements for laborious, repeated phenotyping, and gaps in knowledge regarding the relationships between causal DNA variants in *ENG*, *ACVRL1*, *SMAD4* and *GDF2*, and clinical manifestations. To address this, we analysed DNA samples from 183 previously uncharacterised, unrelated HHT and suspected HHT cases using the ThromboGenomics high-throughput sequencing platform. We identified 168 heterozygous variants, 127 unique. Applying modified ACMG Guidelines, 106 were classified as pathogenic/likely pathogenic, 21 as non-pathogenic (variants of uncertain significance/benign). Unlike the protein products of *ACVRL1* and *SMAD4*, the extracellular ENG amino acids are not strongly conserved. Our inferences of the functional consequences of causal variants in *ENG* were therefore informed by the crystal structure of endoglin. We then compared the accuracy of predictions of the causal gene blinded to the genetic data using two approaches: subjective clinical predictions and statistical predictions based on eight Human Phenotype Ontology (HPO) terms. Both approaches had some predictive power but they were insufficiently accurate to be used clinically in isolation from genetic testing. The distributions of red cell indices from larger HHT and control populations differed by causal gene but not sufficiently for clinical use in isolation of genetic data. We conclude that parallel sequencing of the four known HHT genes, MDT review of variant calls sequencing results in the context of detailed clinical information, and statistical and structural modelling are all required to provide a framework to better prognosticate and treat HHT.

INTRODUCTION

Hereditary hemorrhagic telangiectasia (HHT, syn. Osler Weber Rendu syndrome)[1] is a vascular dysplasia affecting circa 1 in 6,000 people,[2-5] with clinical manifestations varying between affected individuals. Inherited as an autosomal dominant[6] disease, HHT leads to the development of large visceral arteriovenous malformations (AVMs), and smaller telangiectasia at characteristic mucocutaneous and gastrointestinal sites.[7-10] Affected individuals usually need specific management of symptoms including those from hemorrhage and anemia,[11-14] together with screening and multisystemic management programmes to prevent future complications, and to increase life expectancy.[1,4,15,16] Although clinical manifestations vary by genetic aetiology, such programmes currently do not distinguish between the major HHT genotypes, potentially exposing some individuals to excessive investigations and treatments, and others to insufficient management.

Molecular diagnostics to identify the single allele that is responsible for HHT in any given family, offers the clearest route to date to tailor care pathways to best suit the individual, by both defining the presence of HHT with attendant screening requirements, and sub-categorising according to emerging gene-specific risk profiles.[16-26] However, despite earlier recommendations,[14] serial data from international surveys indicate that between 2011 and 2018, the proportion of families tested for a familial HHT genetic diagnosis had only risen from 23.4%[27] to 49.4%[28] with 3-fold differences between countries.[28]

The majority of HHT patients with a molecular diagnosis have a pathogenic DNA sequence variant in *ENG*, encoding endoglin (ENG, HHT1),[29] or *ACVRL1*, encoding activin receptor-like kinase (ALK1, HHT2).[30] A smaller proportion harbour a pathogenic variant in *SMAD4*, which also causes other pathologies, including juvenile polyposis and aortopathy.[23-26] Essentially all known HHT pathogenic variants are null alleles. Separate pathophysiological considerations apply to why individual vascular abnormalities then develop at particular sites (for example, specific local triggers in the setting of germline haploinsufficiency,[31] potentially including somatic mosaic loss of the second allele),[32] but these considerations are not the focus of the current manuscript.

ENG, *ACVRL1* and *SMAD4* encode endothelial cell-expressed proteins that transmit or regulate signals by bone morphogenetic protein/transforming growth factor-β (BMP/TGF-ß) superfamily ligands, through heteromeric receptor serine-threonine kinase complexes, to canonical (SMAD-based) and non-canonical (non SMAD) pathways.[33] HHT was initially considered to result from aberrant TGF-ß signalling, but newer data implicate the specific ALK1 ligands BMP9 (encoded by *GDF2*) and BMP10.[34] Accordingly, BMP9/10

inhibition recapitulates features of HHT[35] and a direct interaction between BMP9 and ENG was detected biochemically[36] as well as captured by a crystal structure of the endoglin-BMP9 complex determined in 2017.[37] *GDF2* was considered a candidate gene for HHT because missense substitutions were identified in 3 of 191 unrelated individuals with nosebleeds and skin telangiectasia, who did not carryany pathogenic variants in *ENG, ACVRL1* or *SMAD4*.[38] Although unreplicated for many years, one *GDF2* family with HHT appears to have been identified through the 100,000 Genomes Project.[39]

Reaching a conclusive molecular diagnosis is complicated by the appreciable proportions of missense variants in the HHT genes to which pathogenicity cannot be assigned without protein or functional studies.[40] Early HHT causal gene-phenotype studies focussed on the primary structural abnormalities (AVMs/telangiectasia).[17-22] We hypothesised that a more systematic approach to HHT molecular diagnostics including evaluation of both primary and secondary HHT phenotypes, would provide opportunities for the genetic stratification of clinical care. Here, we report the identification and phenotypic integration of 168 variants (127 unique, of which 64 are novel) across 147 HHT families.

METHODS:

The research was approved by national ethics committees, and all human participants gave written informed consent (further details in eMethods).

Cohort 1 (n=183)

Cohort 1 comprised unrelated patients attending a single UK institution for management of a clinical diagnosis of known or suspected HHT[41,42] where the family's HHT pathogenic variant was not already known from research[43-46] or clinical genotyping. HHT symptoms and features were classified using the Curaçao Criteria of nosebleeds, mucocutaneous telangiectasia, visceral involvement, and an affected first degree relative.[47] Using a templated record, 301 Human Phenotype Ontology (HPO) terms,[48] and numerical indices, were recorded from primary notes, radiology systems and electronic blood records. We also recorded the number and proportion of relatives affected. Overall, Cohort 1 comprised 183 probands (113 females [61.7%]).

Cohort 2 (n=94)

Cohort 2 was from the same institution and comprised 94 further probands (62 [66.0%] female) with a prior clinical and molecular diagnosis, in order to provide a replicate dataset from the same geographical background. Sixty-two had *ENG* pathogenic variants, 27 *ACVRL1* and five *SMAD4*.

Bleeding, red cell, and oxygenation indices (cohorts 1 and 2)

With ethical approvals (LREC 00/5792), for both cohorts, all available HPO, bleeding phenotypes and quantitative red cell and oxygenation indices were collected on the probands, and on affected family members reviewed in the same clinical service. All were measured in the same laboratory –red cell indices on XE Series Analysers (Sysmex, UK), pulse oximetry on Ohmeda Biox 3900s (Boulder, Colorado). Data collections were performed in mid-2017 and validated in 2018. Red cell indices were measured by complete blood counts, and oxygen saturation (SaO_2, reduced proportional to the fraction of right ventricular output passing through pulmonary AVMs) by finger oximetry over 10 minutes standing.[49-51] Blinded to genotype, clinical descriptions of patients' bleeding status were assigned to our institutional 7 point HHT bleeding scale [41,52]: 0-4 is based on nosebleeds (0: none; 1: <1/year; 2: ≤1/month; 3: several a week, 4: daily). An extra point is added for states leading to additional iron losses[13], in this cohort menorrhagia (N=4), severe gastrointestinal bleeding (N=8), blood donors (N=1) or large volume near-daily nosebleeds (N=24). A"6" is reserved for patients who are blood transfusion- or intravenous-iron-dependent at the time, either with daily "extreme or torrential" nosebleeds, or severe gastrointestinal bleeding,[13,41,52] Comparative analyses were restricted to the first recorded full dataset per patient, irrespective of whether this preceded therapeutic corrections. For binary analyses, severe bleeders were categorised by a bleeding score >4.

Sequencing and bioinformatics

The ThromboGenomics diagnostic HTS platform[54] version 2 was used to identify variants in the exonic fraction of targeted HHT causal genes (*ENG, ACVRL1* and *SMAD4),* and the candidate gene *GDF*2 using the following transcripts: *ENG*: LRG_589, NM_000118.3, and NM_001114753.2; *ACVRL1* LRG_543, NM_000020.2; *SMAD4*: LRG_318, NM_005359.5. Relevant intronic regions, untranslated regions (UTRs) and 1,000 bp promoter regions were included. The reads in the de-multiplexed paired-end FASTQ files were processed as described previously[54] to call single nucleotide variants (SNVs), indels and large deletions.

Variant interpretation

Variants were annotated and filtered in an automated fashion using the rules described previously[54]. Candidate variants were assessed during multi-disciplinary team meetings (MDTs) to assign their pathogenicity status for this autosomal dominant disease caused by a single pathogenic null allele, recognising the possibility that second variants may influence function sufficiently to modify disease presentation. Where probands exhibited more than one variant in the HHT genes, evidence for (likely) pathogenic status was sought for all of them. ACMG/AMP criteria were the primary source for classifying evidence for pathogenicity in terms of disease causation: For pathogenic, very strong [PVS]; strong [PS]; moderate [PM] and supporting [PP], and for benign impact, standalone [BA], strong [BS] and supporting [BP] criteria.[55] The letter code is followed by a 1-7 according to the precise criterion.[55] Pathogenicity assignments for some known variants are inconsistent within existing databases (*Supplemental Figures i,ii*), and we made minor modifications to improve assignments for this rare disease (*Supplemental Figures iii-vi*) . The key steps followed by the MDT, discussed further in the Supplemental data, were:

(1) Rare variants met the ACMG/AMP criterion [PVS1] for ***Pathogenic*** if they clearly resulted in null alleles (transcript ablation multiexon, pan exon, and frameshift insertions/deletions, stop codon gains, and disruption of start or splice site consensus sequences[55]).

(2) Rare inframe indels and missense substitutions were evaluated conservatively, mindful that only *ACVRL1* and *SMAD4* exhibit appreciable constraint against missense variation: In ExAC, the expected number of missense variants in *ACVRL1* was 199.6, but only 132 were observed (z-score=2.34); in *SMAD4* 179.6 expected, only 65 observed (z-score=4.18); in *ENG* 235.9 expected, 223 observed (z-score=0.41); in *GDF2* 165.7 expected, 158 observed (z-score=0.29).[56] In agreement with ExAC, in gnomAD[57] there is also significant constraint for *ACVRL1* and *SMAD4* but not for *ENG* and *GDF2*, showing that *ENG* and *GDF2* have a higher tolerance of missense substitutions than *ACVRL1* or *SMAD4* Even for *SMAD4*, amino acid conservation assessed using Consurf cautioned us against applying criteria across a full gene.[58] Thus rare inframe indels and missense substitutions met the rules for ***Likely Pathogenic*** if there was published cellular evidence that they generate null alleles [PS3], and/or they were sited within known functional protein domains of ALK1 (serine-threonine kinase catalytic domain, ligand binding site) or SMAD4 (MH2 domain) [PM1 and a sufficient number of other strong, moderate or supporting criteria] [55].

(3) Following publication of crystal structures of human endoglin and its complex with BMP9 in June 2017,[37] all missense and in-frame deletions in the endoglin gene were examined, blinded to evolving MDT assignments. Modelling of certain individual variants offered an additional criterion [PM1] to confirm, and/or to move a variant of uncertain significance (VUS) towards ***Likely Pathogenic*** status based on its predicted

effect on either the folding of endoglin or the protein's ability to interact with BMP9.[37] Three-dimensional clustering of the variants was also used to identify possible novel functional sites of endoglin.

(4) Any common alleles (allele frequency >0.05 in 1000 Genomes[59] meeting [BA1][55], had been removed by the ThromboGenomics pipeline. Remaining relatively common alleles (allele frequency >0.02 in 1000 Genomes[59]) met [BS1][55], and were designated as ***Likely Benign.*** The majority of these occurred in conjunction with a clear high impact pathogenic allele[5].

Causal Gene Predictions

Prior to DNA sequencing, based on the sites and morphology of any telangiectasia, AVMs, and family history, an experienced HHT clinician predicted whether the disease was likely to be HHT, a different inherited vasculopathy masquerading as HHT,[60,61] or an isolated AVM (commonly referred to as "sporadic"[1]). HPO terms for the proband and family were then used to predict which causal gene defect was most likely, and to provide a subjective confidence in these predictions. Key assignment terms were juvenile polyposis (*SMAD4* predicted), severe hepatic AVMs or pulmonary arterial hypertension (*ACVRL1* predicted) and multiple pulmonary or cerebral AVMs (*ENG* predicted). In some cases, it was considered possible to exclude one gene but not two, for example, a family with multiple pulmonary or cerebral AVMs and either colorectal polyposis or unusual cancers would have "*ENG* or *SMAD4*" assigned.

Following sequencing, where a pathogenic variant was identified for a Cohort 1 DNA, the pre-MDT accuracy of the clinical prediction of the causal gene one year earlier was evaluated. Additionally, the positive predictive value (PPV), negative predictive value (NPV), sensitivity and specificity of the subjective expert clinician's call at varying levels of confidence, was reviewed for the different molecular subtypes, particularly the more common subtypes of frameshift deletions, splice site, nonsense and missense substitutions.

To evaluate if a statistical approach could aid in predicting the causal gene, phenotypes that were thought to discriminate between the three established HHT genes were used in a predictive model. For each of *ENG*, *ACVRL1* and *SMAD4*, prior frequencies of the discriminatory HPO-coded abnormalities due to defects in that gene were encoded using beta distributions, parameterised according to data from earlier studies[17-26] and prior clinical observations - for example, the smaller number of *SMAD4* families evaluated previously appear to more resemble *ENG* than *ACVRL1* in terms of visceral AVMs.[23-26] The prevalence of *ACVRL1*, *ENG* and *SMAD4*-related causes of HHT were assumed to be 27%, 67% and 6%, respectively, based on the genotypes in the Cohort 2 families.

The above gene-specific prior distributions of trait frequencies were used to predict the causal gene for each individual. We computed a posterior distribution of the causal gene, independently for each individual. For a given individual, let g be a random variable indicating which of the three genes is causal, let $p_j$be the probability that the individual manifests the jth HPO term, and let $y_j$be a binary variable representing whether or not the individual has the jth HPO term. We modelled the HPO data as follows:

$$P(g) = Multinomial(q)$$

$$P(p|g) = \prod_j \quad Beta(\alpha_{gj}, \beta_{gj})$$

$$P(y|p) = \prod_j \quad Bernoulli(p_j)$$

We set $q = (0.27, 0.67, 0.06)$ based on the expected relative frequencies of *ACVRL1*, *ENG* and *SMAD4* as the cause of HHT in the genotyped population. The α_{gj} and β_{gj} shape parameters were set as described in the Data Supplement. It can be shown that:

$$P(g = k|y) \propto \prod_j \quad \frac{\alpha_{kj}^{y_j} \beta_{kj}^{(1-y_j)}}{\alpha_{kj} + \beta_{kj}}$$

The gene indexed by $\max_k P(g = k|y)$, which is the gene with the highest posterior probability, was predicted to be the causal gene.

After variant pathogenicity assignment, the correspondence between the gene harbouring a likely pathogenic or pathogenic variant as decided at the MDTs, and the predicted gene (both from the manual prediction, and the statistical prediction), was assessed.

Data Sharing Statement

Original data may be found in a data supplement available with the online version of this article. The variants are being submitted to ClinVar[62] and will also be available on the HHT Mutation Database[40] after publication.

RESULTS:

Coverage profile of the ThromboGenomics platform

The mean coverage across the targeted regions of the 4 genes examined (*ENG*, *ACVRL1*, *SMAD4* and *GDF2*, comprising 17,878bp) was on average 832.6X (range, 637.0 to 1395.0) *(Figure 1A)*. The mean fraction of exonic bases covered at 20X and 50X was 0.9999994 and 0.9999404 respectively *(Figure 1B)*. Individual coverage profiles for each gene on the platform showed that virtually all exonic bases of the ThromboGenomics transcripts for these four genes are covered sufficiently for sensitive variant calling (*Supplemental Figure 1*), with the profile of *GDF2* shown *in Figure 1C* as a specific example.

Variant calling

Automated variant filtering yielded 168 heterozygous candidate variants within the 4 examined genes in 147 of the 183 patients in Cohort 1. The remaining thirty-six patients had no candidate variants identified. Among the 168 candidate variants, 127 were unique variants of which 64 (50%) had not been previously identified in HHT cases. The full list of the filtered variants is provided in *Supplemental Table 1.*

Variants were brought to MDT meetings for pathogenicity assignment. Among the 127 unique variants, 106 were labelled as pathogenic or likely pathogenic, providing a molecular diagnosis in 140/183 (74.86%) patients. The 68 rare variants assigned as ***Pathogenic***, and 38 rare variants as ***Likely Pathogenic*** spanned genomic structural variants, indels, consensus splice variants, and single nucleotide variants affecting amino acid residues within critical functional domains of the ALK-1 or SMAD4 proteins, noting these were not always mutually exclusive (*Supplemental Table 2, Supplemental Figure 2*).

In contrast to ALK1 and SMAD4 proteins, extracellular ENG amino acids displayed highly variable normalised conservation scores *(Figure 2)* implying that domain-based, secondary structural predictions were unlikely to be feasible. The three-dimensional mapping of likely pathogenic *ENG* missense and in-frame indels to the crystallographic model of extracellular endoglin[37] suggested that previously unsuspected regions of the protein as important for its function *(Figure 3),* thus implicating non-contiguous amino acids as being of particular functional importance. Specifically, a mutational "hot spot" for pathogenic variants was identified

in α-helix 2 and β-strand 16 of orphan domain 1 (OR1), adjacent to the region that directly contacts BMP9 (*Figure 3Ai, 3Aii*).

Of the 147 patients where a filtered variant had been identified, 20 (13.6%) patients had two or more variants within the 4 analysed genes (*Supplemental Table 3*). In 19 of these patients, one of the variants was a high impact allele in *ACVRL1, ENG,* or *SMAD4* that explained the HHT phenotype, while the remaining one was a relatively common (allele frequency >0.02[59]) missense variant and for this reason labelled as benign. Five further patients carried missense variants of uncertain significance. Focussing on *ENG* missense variants designated likely pathogenic (N=24), and benign (N=14), it was notable that only one of the 14 benign variants (7.1%) was located in a potentially critical region of the tertiary structure, compared to 18/24 (75%) of the pathogenic missense substitutions (*Supplemental Table 2C).* For 3 benign variants, the near-linear three-dimensional arrangement (*Figure 3Aiv*) suggested that amino acid substitutions in one of the faces of the ZP module of endoglin do not generate the null alleles required for the HHT phenotype.

As noted above, at least one variant was identified within the 4 genes in 147 of the 183 (80.3%) previously uncharacterised, unrelated HHT patients with definite or suspected HHT, but not all of these variants were pathogenic. Restricting to the 150 unrelated individuals in Cohort 1 with ≥3 Curacao Criteria (i.e a definite clinical diagnosis of HHT),[47,63] 125 (83.3%) had a pathogenic variant, providing a clinical diagnostic yield of 125/150 (83.3%).

Comparison with previously published HHT cohorts

Pathogenic variants in Cohorts 1 and 2 were compared to HHT Mutation Database[40] entries on 24.11.2018. This version differed from a previous version from 2012 due to the recent institution of stringent ACMG/AMP criteria[55] and, consequently, contained fewer pathogenic missense substitutions (*Supplemental Figure i*). The distribution of the molecular types of variants in Cohort 1 was similar to those reported on the 2018 HHT Mutation Database. This correspondence was greatest for the most common causal gene *ENG* (*Figure 3B*), despite 38 of the 74 (52.1%) Cohort 1 *ENG* pathogenic variants being newly-described.

Four candidate variants were identified in *GDF2* in four individuals: one was a synonymous splice site adjacent-variant in an individual with an *ENG* missense variant. Three were missense variants (including one previously reported as being pathogenic (*GDF2* c.997C>T (p.Arg333Trp)[38]), but all were identified in individuals with a rare likely pathogenic *ENG* variant and were therefore assigned benign status (*Supplemental Table 4*).

PREDICTED AND OBSERVED PHENOTYPES

Of the 301 distinct HPO terms assigned to cases in the current cohort, eight were predicted *a priori* to differ in prevalence between patients of different HHT genotypes.[17-26] Figure 4A compares the predicted frequency of these phenotypes in each genotype based on existing literature[17-26] (triangles), and the observed frequencies across Cohorts 1 and 2 (circles). Generally, these concurred, with notable exceptions where screening of asymptomatic patients was not performed in the cohort. Particularly, observed frequencies were lower than predicted for colorectal polyposis in *SMAD4*, and hepatic AVMs in *ACVRL1* (Figure 4A).

Prior to genotyping, suspected genotypes of the probands had been assigned based on the distribution of clinical phenotypes within their families. The clinician's prior suspected genes for probands subsequently demonstrated to have *ACVRL1*, *ENG* or *SMAD4* causal variants are illustrated in Figure 4B (upper panel). Overall, positive prediction values (PPVs) for the clinician's *ENG* calls were consistently higher than for *ACVRL1* (≥85.0% versus ≤70%, all p values <0.001), and negative prediction values (NPVs) for *SMAD4* were ≥99% (*Supplemental Table 5*). The lower panel shows the distribution of the suspected causal genes from statistical predictions using the eight HPO terms predicted to be discriminatory. The automated predictions provided very similar predictive accuracy compared to predictions by an expert clinician, particularly for *ENG* (Figure 4B), though *SMAD4* phenotypes were less well predicted by the statistical model.

Erythrocytic indices

To test whether quantitative measurements might enhance predictions of the causal gene, red cell indices were examined, since anemia is both a challenging problem in HHT,[8,13] and a driver of higher cardiac outputs, flow, and potentially AVM enlargement.[64-66] As expected, across Cohorts 1 and 2, HHT cases commonly demonstrated extremely low red cell values relative to 50,000 healthy controls (blood donors from the INTERVAL study,[69] Figure 5A). Median values were lower in females than males in all cases. Notably however, the majority of values were similar to the healthy controls, the median values were similar for the sex, and there was an excess of extremely high red cell values in the HHT patients (Figure 5A).

The causes are illustrated in Figure 5B: Male and female HHT patients with more severe bleeding (bleeding scale >4) tended to have lower red cell count, hematocrit and hemoglobin than individuals without severe bleeding (Mann Whitney p<0.0001 in all cases), in keeping with bleeding leading to iron-restricted erythropoiesis. However, there was a significant anti-correlation between these three red cell indices and SaO_2

(p<0.0001) (Figure 5B), expected since pulmonary AVM-induced hypoxemia leads to compensatory secondary erythrocytosis.[67,68]

Figure 5C compares these quantitative HHT phenotypes between HHT genotypes. There was a suggestive trend for bleeding scale status for *ACVRL1* to be higher than in *ENG* cases, and clear evidence that pulmonary AVM-induced hypoxemia was more severe in *ENG* patients (Figure 5C: p<0.001). However, there was no statistically significant difference in red cell traits between *ENG* and *ACVRL1* cases with or without adjustment for sex, SaO_2 and iron (data not shown). In other words, any associations between the HHT causal gene and red cell traits were mediated by bleeding severity and SaO_2, and did not provide opportunities for improving our predictive model.

DISCUSSION

In this manuscript, we applied standardised phenotyping, predictive modelling, and interpretation of high-throughput sequencing at several hundred-fold depth. By identifying 106 pathogenic variants, of which 55 (52%) were newly described, this study provided a confirmed molecular diagnosis for 140 HHT families, and focussed attention on the HHT genes and functional deductions from HHT-causal variants.

All HHT pathogenic variants were heterozygous changes in the established HHT genes *ENG*, *ACVRL1* and *SMAD4* but not all rare variants identified in those genes are pathogenic. Substantial work may be required to unravel which missense, in-frame indels, non consensus splice site and intronic variants affect protein function sufficiently to produce null, HHT-causing alleles. Hypomorphic, non-null variants may affect protein function and/or modify disease severity, but on current understanding, would not be the causal allele segregating in the family to cause this autosomal dominant disease.

The current study demonstrates that clinician predictions alone cannot indicate the HHT causal gene with sufficient confidence, and formal sequencing analysis of the four genes included in this study is indicated. While negative results would not currently change clinical management of the patient, a positive result which is more likely using presented methods (*Supplemental Figure vi*), allows distinction from other vasculopathies with lesser requirements for visceral surveillance[60,61], and leads to additional changes in HHT patient

management: *SMAD4* patients need to enter endoscopic and echocardiographic surveillance programmes;[24-26] the presence of a causal *ACVRL1* variant increases the perceived risk of hepatic AVMs leading to changes in care pathways, and there are early suggestions that drug adverse event profiles may differ according to the causal HHT gene, thus influencing prescription choices.[16,53] Differing natural history data according to the causal gene may permit tailored risk-benefit weightings for other management elements such as repeat imaging/radiation exposure[70,71] or HHT interventions with complications in addition to efficacy.[71-75] We speculate that the higher clinician accuracy for predicting *ENG* nonsense substitutions compared to frameshift indels and splice site variants (*Supplemental Table 5*) may reflect more complex and functional consequences in the latter two, although this needs to be examined in further studies. Our data confirm that the value of HHT gene testing extends beyond predictive testing of relatives, and we encourage funding within mainstream medicine.

The proportion of unsolved cases (36/183 patients (19.6%)) demonstrates that an inconclusive HHT gene test does not exclude a diagnosis of HHT, unless this has excluded the known HHT pathogenic variant in the family. Further molecular testing is then indicated in due course, while the patient continues to be managed for HHT. Statistical modelling from readily assessable HPO terms offers predictive value, and the approach is adaptable to particular characteristics of the population to account for local screening/referral patterns and detection rates. Future statistical models of the causal gene could incorporate quantitative traits that expose substantial processes influencing HHT phenotypes (*Figure 4*). As for other recently published clinical measurements (spirometry[76]; haptoglobin/hemolysis[41]), these allow evaluation of HHT pathophysiology beyond the championed cellular processes of angiogenesis, vascular repair, and endothelial-mural cell communications.[77-82] We cannot rule out hypomorphic alleles having pathogenic roles in specific settings, but suggest variant labelling as a "potential disease modifier" should also take into account the wider clinical phenotypic influences from large population-based studies. The latter include not only bleeding phenotypes and SaO_2 (as in the current manuscript), but also other hemodynamic factors, such as systemic vascular resistance, cardiac output, and cardiac function.

Our results lead us to propose the following systematic approaches to HHT care incorporating:

(1) Molecular confirmation/exclusion of HHT status to facilitate preventative strategies while limiting unnecessary investigations: parallel panel-based, high depth sequencing across HHT causal and candidate genes appears optimal, enabling non-HHT causal variants to be catalogued for clinical and research use.

(2) A systematic approach to application of the ACMG Guidelines for a standardisation of variant interpretation and databasing of all variant classifications by date.

(3) Systematic capture of clinical measures that influence the HHT phenotype: at a minimum, we suggest these include a broad, time-independent bleeding scale and SaO_2, which is not only a biomarker for severity and risk of pulmonary AVMs, but also a major influencer of hemodynamic and hematological phenotypes in HHT.

We conclude that high-throughput, high-depth sequencing, platforms of HHT causal genes, statistical and structural modelling of heterozygous, potentially null alleles, and standardised phenotyping are the methods of choice in the 21st century.

ACKNOWLEDGMENTS,

For patient recruitment we thank the Hammersmith Hospital, Imperial College Healthcare Trust and Imperial College sponsored programmes, ethically approved by the National Research Ethics Committee East of Scotland Multicentre Research Ethics Committee (MREC/98/0/42; LREC 99/5637M; and 07/MRE00/19). We gratefully acknowledge our colleagues and staff at Hammersmith Hospital, Imperial College Healthcare NHS Trust for assistance with patient phenotyping as part of standard NHS care, and the patients for their willing cooperation in these studies. The patient-facing research was co-funded by the NIHR Imperial Biomedical Research Centre, the Margaret Hayton Memorial Trust and the Averil Macdonald Memorial Fund.

For the ThromboGenomics platform, we thank the NIHR BioResource Scheme who funded and developed the ThromboGenomics platform (NIHR RG65966). We gratefully acknowledge the Cambridge NIHR BioResource Centre, for their contribution. We thank the National Institute for Health Research, and NHS Blood and Transplant.

For individual authors: CLS is supported by the Imperial College Biomedical Research Centre and for this project, via serial Wellcome Trust (TF/037257, AF/053286), British Heart Foundation (PG/2000067, FS/04/089, and PG/09/041/27515), National Institute for Health Research, and Imperial Biomedical Research Centre funding. IS, KM, CP, ET are supported by NIHR BioResource - Rare Diseases (NIHR RG65966). KD is supported as a NHS HSST trainee by Health Education England and NIHR BioResource - Rare Diseases (NIHR RG65966). WHO is supported by: RBAG/181, NIHR BioResource - Rare Diseases (NIHR RG65966), British Heart Foundation (RBAG/245, 208, 226), European Commission (RBAG/344), MRC (RBAG/285, 295), NHS Blood and Transplant (RBAG/142), and Wellcome Trust (RBAG/342). LJ is supported by Swedish Research Council (grant 2016-03999), Center for Innovative Medicine (grant 2-537/2014) and KI Research Foundations (grant 2018-02166).

The views expressed are those of the author(s) 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
CLS	CLS obtained ethics approvals, performed all clinical studies, consented families, extracted DNAs, predicted phenotypes, led the MDT, defined and applied the ACMG-based algorithm, wrote the first manuscript draft,	CLS has no competing financial interests. Potential non financial conflicts of interest include roles as Chair of VASCERN HHT; Lead Clinician for the VASCERN HHT Reference Centre at Hammersmith Hospital; Chair of Genomics England Respiratory GeCIP subdomains for HHT

	generated Supplementary Data, developed the text and Figures with ET and LJ, and revised and approved the final manuscript.	and PAVMs; and December 2018 acceptance of an invitation to sit on the Clinical Genome Resource (ClinGen) HHT Variant Curation Expert Panel, noting all manuscript text was completed prior to participating in any calls
IS	ThromboGenomics platform design, managed ThromboGenomics, processed samples, reviewed and approved the final manuscripts	IS has no competing financial interests
KD	Managed ThromboGenomics, reviewed and approved the final manuscripts	Kate Downes is supported as a NHS HSST trainee by Health Education England
KM	ThromboGenomics platform design, gene and transcript curation, variant submission to public repository, and reviewed and approved the final manuscript	KM has no competing financial interests
ZF	Extracted DNAs, contributed to data discussions, and reviewed and approved the final manuscript	ZF has no competing financial interests
MBH	Participated in MDT, contributed to data discussions, and approved the final manuscript	MBH has no competing financial interests
AS	Performed database comparisons presented in Supplemental Figure ii, *and approved the final manuscript*	AS has no competing financial interests
JB	Recruited study participants, contributed to data discussions, and reviewed and approved the final manuscript	JB has no competing financial interests
DP	Extracted DNAs, contributed to data discussions, and reviewed and approved the final manuscript	DP has no competing financial interests
LK	Assisted with study phenotyping, and reviewed and approved the final manuscript	LK has no competing financial interests
JS	Processed Thrombogenomics samples, performed deletion validations and approved the final manuscript.	JS has no competing financial interests
IT	Contributed to data discussions, and reviewed and approved the final manuscript	IT has no competing financial interests
MA	Participated in MDT, contributed to data discussions, and reviewed and approved the final manuscript	MA has no competing financial interests
CP	Contributed to data discussions, and reviewed and approved the final manuscript	CP has no competing financial interests
WHO	Contributed to data discussions, and reviewed and approved the final manuscript	WHO has no competing financial interests
LJ	Performed structural analysis, generated Figure 3A and associated text, contributed to data discussions, and reviewed and approved the final manuscript.	LJ has no competing financial interests
ET	Performed data analyses, developed the statistical models, contributed to data discussions, generated Figures, co-authored the text, and approved the final manuscript.	ET has no competing financial interests

EXCLUDED REVIEWERS None

REFERENCES

1 VASCERN-HHT European Reference Network For Vascular Diseases. Hereditary haemorrhagic telangiectasia, January 2019, available at www.orpha.net/consor/www/cgi-bin/OC_Exp.php?lng=EN&Expert=774 (accessed 24th November 2019)

2 Bideau A, Brunet G, Heyer E, Plauchu H, Robert J-M. An abnormal concentration of cases of Rendu-Osler disease in the Valserine valley of the French Jura: a geneological and demographic study. *Ann Hum Biol*. 1992;19:233-247

3 Dakeishi M, Shioya T, Wada Y, et al. Genetic epidemiology of hereditary hemorrhagic telangiectasia in a local community in the northern part of Japan. *Hum Mut*. 2002;19:140-148

4 Kjeldsen AD, Vase P, Green A. Hereditary haemorrhagic telangiectasia: a population-based study of prevalence and mortality in Danish patients. *J Intern Med*. 1999;245:31-39

5 Donaldson JW, McKeever TM, Hall IP, Hubbard RB, Fogarty AW. The UK prevalence of hereditary haemorrhagic telangiectasia and its association with sex, socioeconomic status and region of residence: a population-based study. *Thorax*. 2014 Feb;69(2):161-7

6. Mendel G. Versuche über Plflanzenhybriden. Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865, Abhandlungen, 3–47, available as Mendel G. Experiments in Plant Hybridization at www.esp.org/foundations/genetics/classical/gm-65.pdf (accessed 24th November2019).

7 Guttmacher AE, Marchuk DA, White RI Jr. Hereditary hemorrhagic telangiectasia. *N Engl J Med*. 1995 Oct 5;333(14):918-24

8 Shovlin CL. Hereditary haemorrhagic telangiectasia: Pathogenesis, diagnosis and treatment. *Blood Rev*. 2010; 24 :203-19

9 McDonald J, Pyeritz RE. Hereditary Hemorrhagic Telangiectasia. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, Amemiya A, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2019. 2000 Jun 26 [updated 2017 Feb 2].

10. Plauchu H, de Chadarévian J-P, Bideau A, Robert J-M. Age-related clinical profile of hereditary hemorrhagic telangiectasia in an epidemiologically recruited population. *Am J Med Genet.* 1989;32:291-297.

11 Shovlin CL, Buscarini E, Kjeldsen AD, et al. European Reference Network For Rare Vascular Diseases (VASCERN) Outcome Measures For Hereditary Haemorrhagic Telangiectasia (HHT). *Orphanet J Rare Dis*. 2018 Aug 15;13(1):136. doi: 10.1186/s13023-018-0850-2.

12 Hoag JB, Terry P, Mitchell S, Reh D, Merlo CA. An epistaxis severity score for hereditary hemorrhagic telangiectasia. *Laryngoscope.* 2010 Apr;120(4):838-43.

13 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 Oct 16;8(10):e76516.

14 Faughnan ME, Palda VA, Garcia-Tsao G, et al. International guidelines for the diagnosis and management of hereditary haemorrhagic telangiectasia. *J Med Genet*. 2011 Feb;48(2):73-87.

15 Kjeldsen A, Aagaard KS, Tørring PM, Möller S, Green A. 20-year follow-up study of Danish HHT patients-survival and causes of death. *Orphanet J Rare Dis*. 2016 Nov 22;11(1):157.

16 Buscarini E, Botella LM, Geisthoff U, et al. Safety of thalidomide and bevacizumab in patients with hereditary hemorrhagic telangiectasia. *Orphanet J Rare Dis*. 2019 Feb 4;14(1):28. doi: 10.1186/s13023-018-0982-4.https://ojrd.biomedcentral.com/articles/10.1186/s13023-018-0982-4

17 Kjeldsen AD, Møller TR, Brusgaard K, et al. Clinical symptoms according to genotype amongst patients with hereditary haemorrhagic telangiectasia. *J Int Med*. 2005;258:349-355

18 Letteboer TG, Mager JJ, Snijder RJ, et al. Genotype-phenotype relationship in hereditary haemorrhagic telangiectasia. *J Med Genet*. 2006 Apr;43(4):371-7

19 Bayrak-Toydemir P, McDonald J, Markewitz B, et al. Genotype-phenotype correlation in hereditary hemorrhagic telangiectasia. *Am J Med Genet A*. 2006;140:463-470

20 Bossler AD, Richards J, George C, Godmilow L, Ganguly A. Novel mutations in ENG and ACVRL1 identified in a series of 200 individuals undergoing clinical genetic testing for hereditary hemorrhagic telangiectasia (HHT): correlation of genotype with phenotype. *Hum Mutat*. 2006;27:667-675

21 Sabbà C, Pasculli G, Lenato GM, et al. Hereditary hemorrhagic telangiectasia: clinical features in ENG and ALK1 mutation carriers. *J Thromb Haemost*. 2007;5:1149-1157.

22 Letteboer TG, Mager HJ, Snijder RJ, et al. Genotype-phenotype relationship for localization and age distribution of telangiectases in hereditary hemorrhagic telangiectasia. *Am J Med Genet A*. 2008 Nov 1;146A(21):2733-9.

23 Gallione C, Repetto GM, Legius E, et al. A combined syndrome of juvenile polyposis and hereditary haemorrhagic telangiectasia is associated with mutations in MADH4 (SMAD4). *Lancet*. 2004;363:852-859.

24 Jelsig AM, Tørring PM, Kjeldsen AD, et al. JP-HHT phenotype in Danish patients with SMAD4 mutations. *Clin Genet*. 2016 Jul;90(1):55-62.

25 Heald B, Rigelsky C, Moran R, et al. Prevalence of thoracic aortopathy in patients with juvenile Polyposis Syndrome-Hereditary Hemorrhagic Telangiectasia due to SMAD4. *Am J Med Genet A*. 2015 Aug;167A(8):1758-62.

26 Vorselaars VMM, Diederik A, Prabhudesai V, et al. SMAD4 gene mutation increases the risk of aortic dilation in patients with hereditary haemorrhagic telangiectasia. *Int J Cardiol*. 2017 Oct 15;245:114-118.

27 Shovlin CL, Awan I, Cahilog Z, Abdulla FN, Guttmacher AE. Reported cardiac phenotypes in hereditary hemorrhagic telangiectasia emphasize burdens from arrhythmias, anemia and its treatments, but suggest reduced rates of myocardial infarction. *Int J Cardiol*. 2016 Jul 15;215:179-85.

28 Boother E, von Widekind S, Post M, et al. International similarities and differences in hereditary hemorrhagic telangiectasia (HHT) pathways reported by patients and clinicians. *Thorax* 2019;74 (Suppl_2).A156.

29 McAllister KA, Grogg KM, Johnson DW, et al. Endoglin, a TGF-beta binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1. *Natur Genet.* 1994;8:345-351

30 Johnson DW, Berg JN, Baldwin MA, et al. Mutations in the activin receptor-like kinase 1 gene in hereditary haemorrhagic telangiectasia type 2. *Natur Genet.* 1996;13:189-195.

31 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

32 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 Oct 9. pii: S0002-9297(19)30349-0. doi: 10.1016/j.ajhg.2019.09.010. [Epub ahead of print]

33 Massagué J. TGFbeta in Cancer. *Cell*. 2008;134:215–230.

34 David L, Mallet C, Mazerbourg S, Feige JJ, Bailly S. Identification of BMP9 and BMP10 as functional activators of the orphan activin receptor-like kinase 1 (ALK1) in endothelial cells. *Blood*. 2007;109:1953-1961

35 Ruiz S, Zhao H, Chandakkar P, Chatterjee PK, Papoin J, Blanc L, Metz CN, Campagne F, Marambaud P. A mouse model of hereditary hemorrhagic telangiectasia generated by transmammary-delivered immunoblocking of BMP9 and BMP10. *Sci Rep.* 2016 Nov 22;5:37366. doi: 10.1038/srep37366

36 Castonguay R, Werner ED, Matthews RG, et al. Soluble endoglin specifically binds bone morphogenetic proteins 9 and 10 via its orphan domain, inhibits blood vessel formation, and suppresses tumor growth. *J Biol Chem.* 2011 Aug 26;286(34):30034-46.

37 Saito T, Bokhove M, Croci R, et al. Structural Basis of the Human Endoglin-BMP9 Interaction: Insights into BMP Signaling and HHT1. *Cell Rep.* 2017 May 30;19(9):1917-1928.

38 Wooderchak-Donahue WL, McDonald J, O'Fallon B, et al. BMP9 mutations cause a vascular-anomaly syndrome with phenotypic overlap with hereditary hemorrhagic telangiectasia. *Am J Hum Genet.* 2013 Sep 5;93(3):530-7

39 Balachandar S, Shimonty A, Sroya M, et al. 100,000 Genomes Project identification and validation of a novel pathogenic variant in *GDF2* (BMP9) responsible for hereditary hemorrhagic telangiectasia and

pulmonary arteriovenous malformations. 69th Annual Meeting of The American Society of Human Genetics (October 2019, Houston, Texas).

40 HHT Mutation Database, hosted by the ARUP Laboratories, University of Utah. http://www.arup.utah.edu/database/HHT/. (accessed 24th November 2019)

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

42 Livesey JA, Manning RA, Meek JH, et al. Low serum iron levels are associated with elevated plasma levels of coagulation factor VIII and pulmonary emboli/deep venous thromboses in replicate cohorts of patients with hereditary haemorrhagic telangiectasia. *Thorax.* 2012 Apr;67(4):328-33.

43 Shovlin CL, Hughes JM, Tuddenham EG, et al. A gene for hereditary haemorrhagic telangiectasia maps to chromosome 9q3. *Nat Genet.* 1994 Feb;6(2):205-9.

44 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 Jul;61(1):68-79.

45 Govani FS. The molecular basis of hereditary haemorrhagic telangiectasia PhD Thesis, Imperial College London 2009

46 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 Apr;4(4):184-96.

47 Shovlin CL, Guttmacher AE, Buscarini E, et al. Diagnostic criteria for hereditary hemorrhagic telangiectasia (Rendu-Osler-Weber syndrome). *Am J Med Genet.* 2000 Mar 6;91(1):66-7

48 Köhler S, Doelken SC, Mungall CJ, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D966-74.

49 Boother EJ, Brownlow S, Tighe HC, Bamford KB, Jackson JE, Shovlin CL. Cerebral Abscess Associated With Odontogenic Bacteremias, Hypoxemia, and Iron Loading in Immunocompetent Patients With Right-to-Left Shunting Through Pulmonary Arteriovenous Malformations. *Clin Infect Dis.* 2017 Aug 15;65(4):595-603.

50 Shovlin CL, Buscarini E, Hughes JMB, Allison DJ, Jackson JE. Long-term outcomes of patients with pulmonary arteriovenous malformations considered for lung transplantation, compared with similarly hypoxaemic cohorts. *BMJ Open Respir Res.* 2017 Oct 13;4(1):e000198.

51 Shovlin CL, Chamali B, Santhirapala V, et al. Ischaemic strokes in patients with pulmonary arteriovenous malformations and hereditary hemorrhagic telangiectasia: associations with iron deficiency and platelets. *PLoS One.* 2014 Feb 19;9(2):e88812.

52 Soni A, Badiani N, Boother E, Shovlin CL. Haemorrhage adjusted iron-requirements and exercise capacity in hereditary haemorrhagic telangiectasia. *Thorax* 2019;74 (Suppl_2). A156

53 Shovlin CL, Millar CM, Droege F *et al*. Safety of direct oral anticoagulants in patients with hereditary hemorrhagic telangiectasia. Orphanet J Rare Dis. 2019 Aug 28;14(1):210

54 Simeoni I, Stephens JC, Hu F, et al. A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders. *Blood*. 2016 Jun 9;127(23):2791-803.

55 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 May;17(5):405-24.

56 Ruderfer DM, Hamamsy T, Lek M, et al. Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nat Genet*. 2016 Oct;48(10):1107-11.

57 Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–91.

58 Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016 Jul 8;44(W1):W344-50. doi: 10.1093/nar/gkw408, webserver http://consurf.tau.ac.il

59 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*. 491(7422):56-65.

60 Revencu N, Boon LM, Mendola A, et al. RASA1 mutations and associated phenotypes in 68 families with capillary malformation-arteriovenous malformation. *Hum Mutat*. 2013 Dec;34(12):1632-41.

61 Amyere M, Revencu N, Helaers R, et al. Germline Loss-of-Function Mutations in EPHB4 Cause a Second Form of Capillary Malformation-Arteriovenous Malformation (CM-AVM) Deregulating RAS-MAPK Signaling. *Circulation.* 2017 Sep 12;136(11):1037-1048

62 National Center for Biotechnology Information. ClinVar available at https://www.ncbi.nlm.nih.gov/clinvar/ (accessed 24th November 2019)

63 van Gent MW, Velthuis S, Post MC, et al. Hereditary hemorrhagic telangiectasia: how accurate are the clinical criteria? *Am J Med Genet A.* 2013 Mar;161A(3):461-6.

64 Buscarini E, Leandro G, Conte D, et al. Natural history and outcome of hepatic vascular malformations in a large cohort of patients with hereditary hemorrhagic teleangiectasia. *Dig Dis Sci* 2011;56:2166–2178

65 Shovlin CL. Circulatory contributors to the phenotype in hereditary hemorrhagic telangiectasia. *Front Genet*. 2015 Apr 9;6:101.

66 Dupuis-Girod S, Cottin V, Shovlin CL. The Lung in Hereditary Hemorrhagic Telangiectasia. *Respiration.* 2017;94(4):315-330.

67 Santhirapala V, Williams LC, Tighe HC, Jackson JE, Shovlin CL. Arterial oxygen content is precisely maintained by graded erythrocytotic responses in settings of high/normal serum iron levels, and predicts exercise capacity: an observational study of hypoxaemic patients with pulmonary arteriovenous malformations. *PLoS One*. 2014 Mar 17;9(3):e90777.

68 Rizvi A, Macedo P, Babawale L, Tighe HC, Hughes JMB, Jackson JE, Shovlin CL. Hemoglobin Is a Vital Determinant of Arterial Oxygen Content in Hypoxemic Patients with Pulmonary Arteriovenous Malformations. *Ann Am Thorac Soc.* 2017 Jun;14(6):903-911.

69 Moore C, Bolton T, Walker M, et al. Recruitment and representativeness of blood donors in the INTERVAL randomised trial assessing varying inter-donation intervals. *Trials*. 2016 Sep 20;17(1):458. doi: 10.1186/s13063-016-1579-7.

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

71 Hanneman K, Faughnan ME, Prabhudesai V. Cumulative radiation dose in patients with hereditary hemorrhagic telangiectasia and pulmonary arteriovenous malformation. *Can Assoc Radiol J*. 2014 May;65(2):135-40

72 Mohr JP, Parides MK, Stapf C, et al. Medical management with or without interventional therapy for unruptured brain arteriovenous malformations (ARUBA): a multicentre, non-blinded, randomised trial. *Lancet*. 2014 Feb 15;383(9917):614-21.

73 European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Vascular diseases of the liver. *J Hepatol.* 2016 Jan;64(1):179-202

74 Lund VJ, Darby Y, Rimmer J, Amin M, Husain S. Nasal closure for severe hereditary haemorrhagic telangiectasia in 100 patients. The Lund modification of the Young's procedure: a 22-year experience. *Rhinology*. 2017 Jun 1;55(2):135-141.

75 Fatania G, Gilson C, Glover A, et al Uptake and radiological findings of screening cerebral magnetic resonance scans in patients with hereditary haemorrhagic telangiectasia. *Intract. Rare Dis Res.* 2018 Nov;7(4):236-244

76 Gawecki F, Strangeways T, Amin A, et al. Exercise capacity reflects airflow limitation rather than hypoxaemia in patients with pulmonary arteriovenous malformations. *QJM*. 2019:112 (5); 335–342

77 Abdalla S, Letarte M. Hereditary haemorrhagic telangiectasia: current views on genetics and mechanisms of disease. *J Med Genet.* 2006;43:97-110

78 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;119:3487–3496

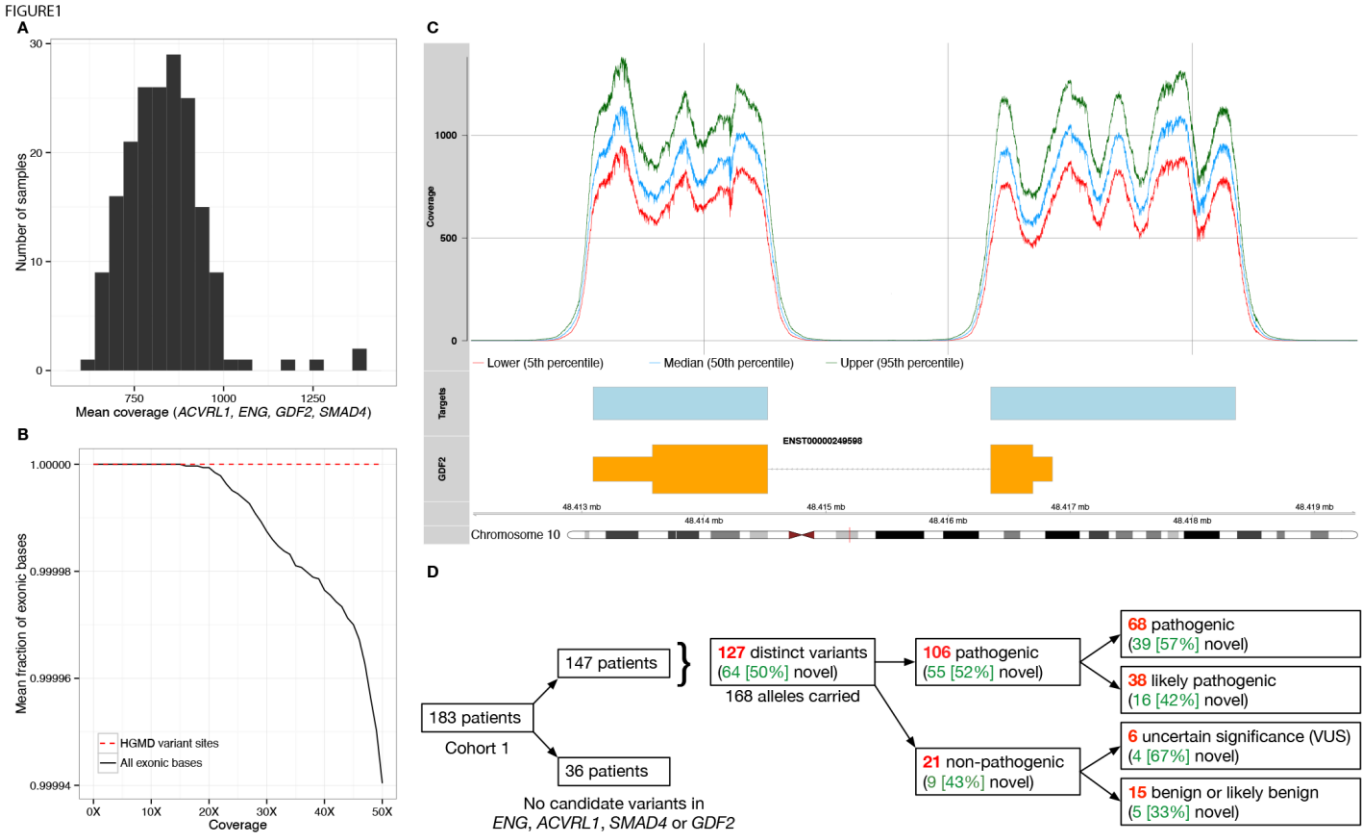
79 Corti P, Young S, Chen CY, et al. Interaction between alk1 and blood flow in the development of arteriovenous malformations. *Development.* 2011;138:1573–1582.

80 Gkatzis K, Thalgott J, Dos-Santos-Luis D, et al. Interaction Between ALK1 Signaling and Connexin40 in the Development of Arteriovenous Malformations. *Arterioscler Thromb Vasc Biol*. 2016 Apr;36(4):707-17.

81 Thalgott JH, Dos-Santos-Luis D, Hosman AE, et al. Decreased Expression of Vascular Endothelial Growth Factor Receptor 1 Contributes to the Pathogenesis of Hereditary Hemorrhagic Telangiectasia Type 2. *Circulation*. 2018 Dec 4;138(23):2698-2712.

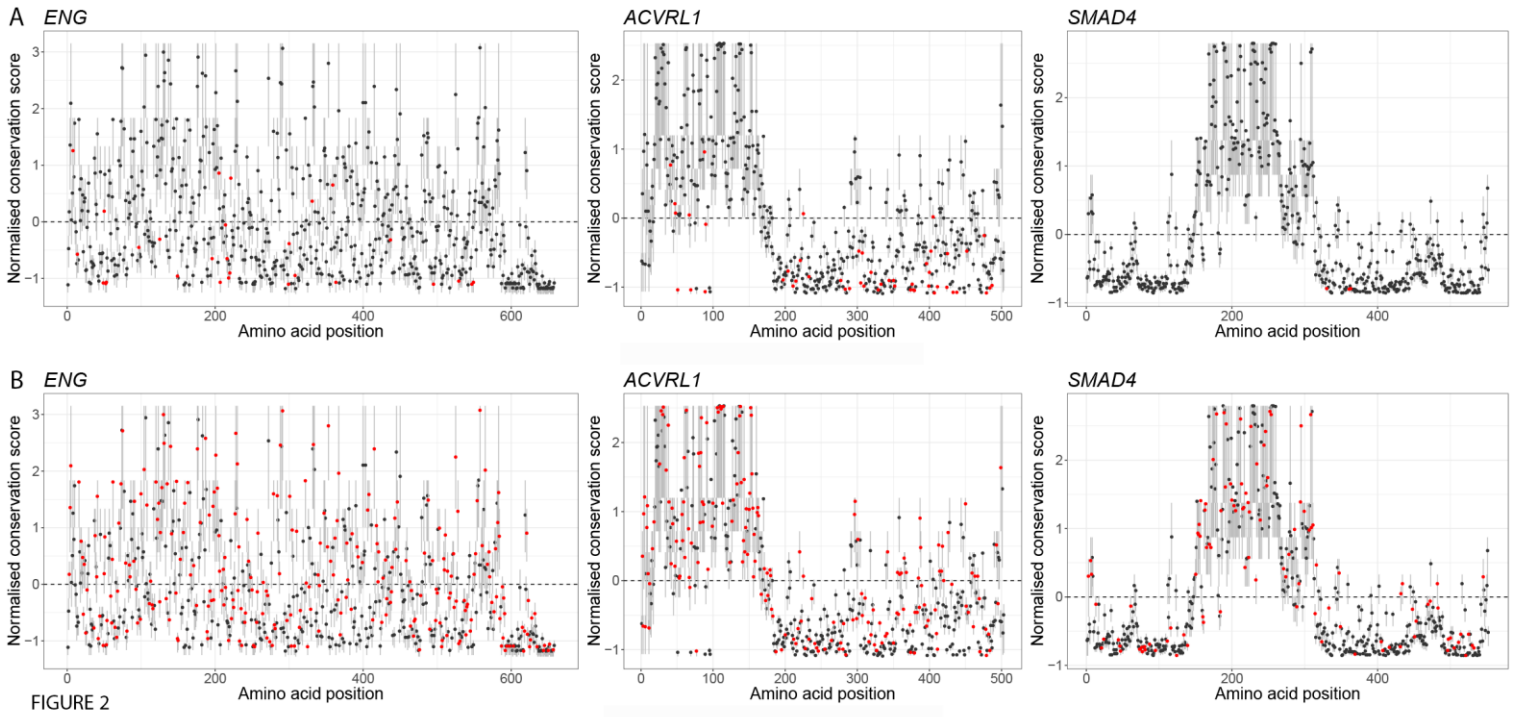
82 Crist AM, Lee AR, Patel NR, Westhoff DE, Meadows SM. Vascular deficiency of Smad4 causes arteriovenous malformations: a mouse model of Hereditary Hemorrhagic Telangiectasia. *Angiogenesis*. 2018 May;21(2):363-380.

Figure 1: **Technical evaluation and output from the HHT panel of the ThromboGenomics platform.**



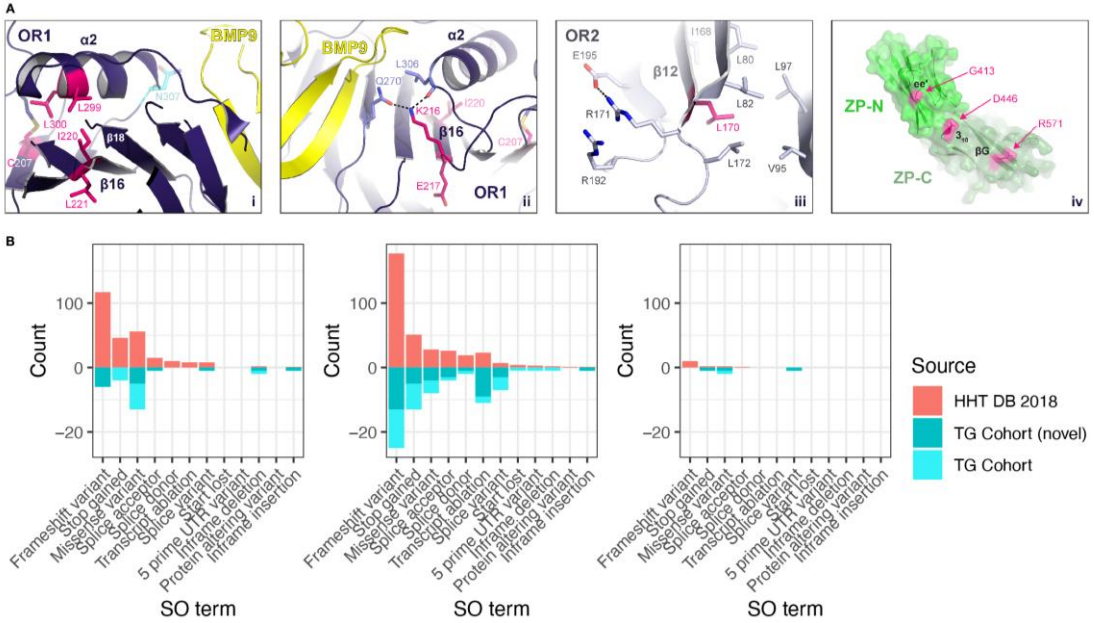
A) Histogram of mean coverage in 183 samples over the targeted regions of the 4 targeted genes (*ENG*, *ACVRL1*, *SMAD4* and *GDF2)*. **B)** The fraction of targeted exonic bases covered at the specified depth (0X–50X) or more, averaged over samples. The solid black line indicates exonic bases, and demonstrates that 99.99% of the targeted exonic bases are covered by at least 50 sequencing reads. The dashed red line indicates bases that lie within human genome mutation database (HGMD) variants and demonstrates that they are all covered by at least 50 sequencing reads. **C)** Coverage profile for the *GDF2* gene encoding BMP9 on chromosome 10, mapped against the Ensembl transcript (orange) which indicates the position and size of the two *GDF2* exons. The pale blue bars indicate the targeted region, and the three traces above, the median, 5th and 95th percentile coverage across the locus. Despite this, no pathogenic variants were identified in the cohort. Equivalent plots for *ENG ACVRL1*, and *SMAD4* are provided in *Supplemental Figure 1*. **D)** Schematic of the classification of the 127 distinct candidate variants identified by the platform.

Figure 2: Normalised amino acid conservation scores across *ENG, ALK1* and *SMAD4*.



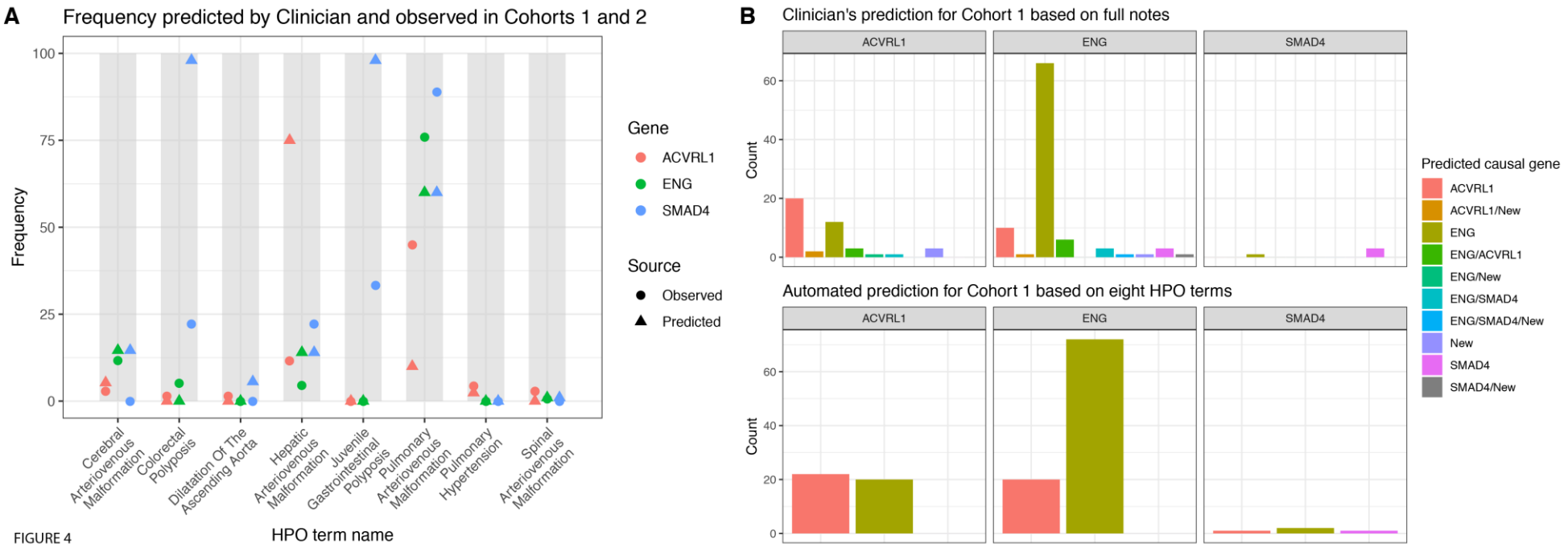
The degree of evolutionary conservation of each amino acid in the human protein sequences of ENG (NM_0011147, 658 amino acids), ACVRL1 (NM_000020.2, 503 amino acids), and SMAD4 (NM_005359.5 (552 amino acids), plotted against the respective amino acid position. The conservation, reflecting the retention of macromolecular function was plotted as normalised conservation scores and 95% confidence intervals obtained using ConSurf.[58] Lower scores indicate greater conservation. In all 6 plots, the selected amino acids are plotted in red, and all other amino acids in black. **A)** Amino acid sites of pathogenic or likely pathogenic (POLP) HHT missense substitutions from current Cohort 1 (N=18), Cohort 2 (N=15) and the 2018 HHT Mutation Database[40] (N=64) are plotted in red, and all other amino acids in black. Amino acids where pathogenic or likely pathogenic variants were located, were more conserved than amino acids with non-pathogenic variants (*ENG* mean difference -0.52 (95% CI -0.81, -0.24), p=0.00072; *ACVRL1* -0.77 (-0.94, -0.61), $p=6.6x10^{-15}$, *SMAD4* -0.80 (-0.89, -0.72), $p=3.9x10^{-61}$). Notably however, not all pathogenic or likely pathogenic variants were at conserved sites, and for endoglin, the normalised conservation scores and confidence intervals were very variable with the exception of the transmembrane domain (near amino acid 600) and C terminal cytoplasmic tail (amino acids 635–658). **B)** Amino acid sites of likely benign missense substitutions in the gnomAD database[57] plotted in red, versus all other amino acids in black. Amino acids where benign variants were sited were less conserved than other amino acids (*ENG* mean difference 0.34 (95% CI 0.19, 0.50, $p=1.6x10^{-5}$; *ACVRL1* 0.48 (0.30, 0.66), $p=3.2x10^{-7}$, *SMAD4* 0.54 (0.32, 0.77), $p=3.0x10^{-6}$).

Figure 3. Molecular characterisation of sequence variants.



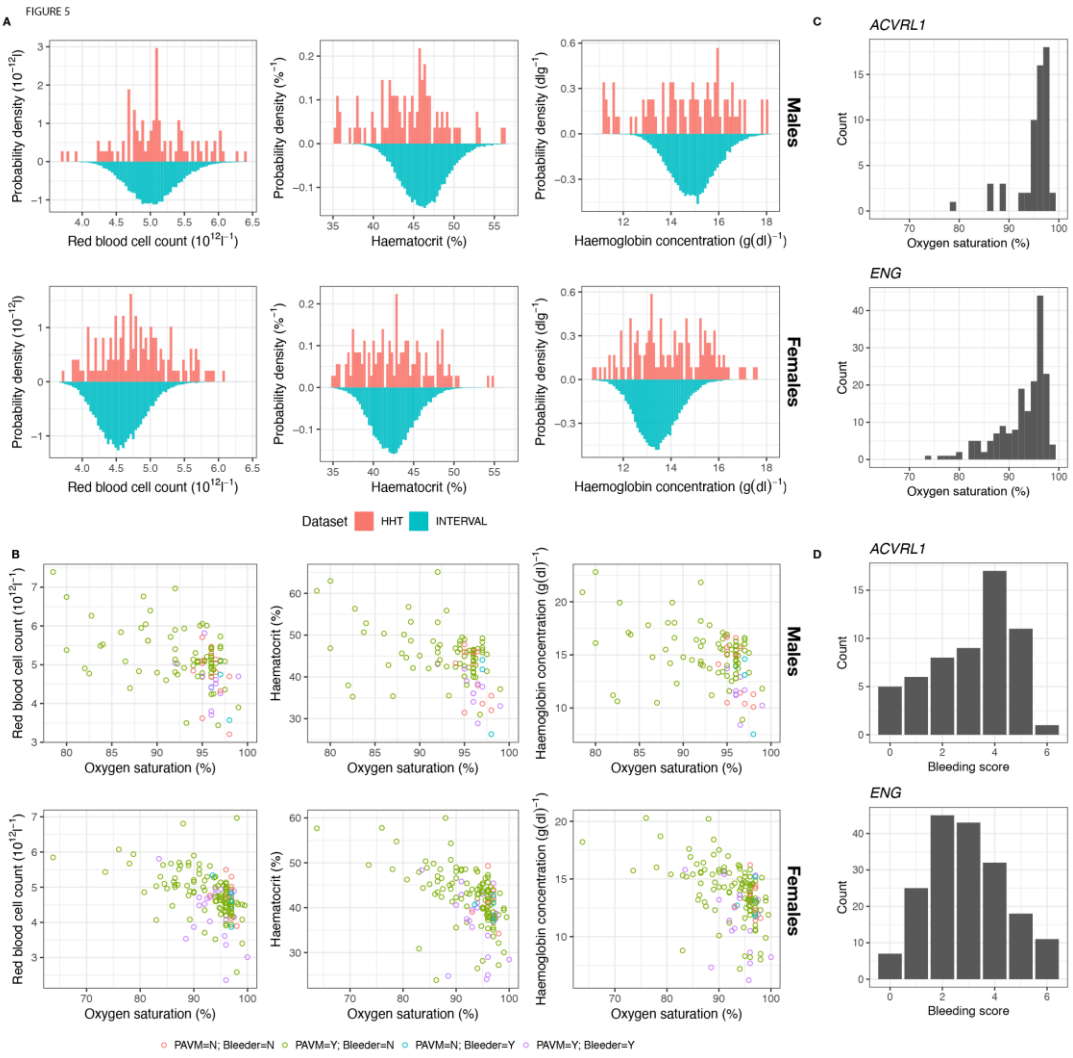
A) Mapping of *ENG* missense substitutions in Cohorts A and B onto the crystal structures of ENG and its complex with BMP9. Proteins are shown in cartoon representation (BMP9 yellow), with specific ENG amino acids depicted as stick, with carbon atoms coloured dark magenta, or for N-glycosylation site N307, cyan. **i)** The relative position of six pathogenic or likely pathogenic variants described in the present manuscript (Cys[C]207Tyr, Leu[L]300Pro, Leu[L]299Arg, Ile[I]220Asn, Leu[L]221Gln, and Asn[N]307Leu defines a hot spot for pathogenic or likely pathogenic missense variants that includes helix α2 of the ENG OR1 domain, whose C-terminal end lies close to the BMP9 binding site[37]. **ii)** A second hot spot for pathogenic or likely pathogenic variants (Lys[K]216Glu and Glu[E]217delinsGluAla) affects residues at the N-terminal end of OR1 β-strand 16, including Lys216 that connects the C-terminal end of α-helix 2 to Gln270 at the ENG/BMP9 interface via hydrogen bond interactions. This view, which depicts amino acid contacts as observed in the high-resolution structure of ENG OR [37], also shows the location of the Cys[c]207Tyr and Ile[I]220Asn mutations from a different perspective. **iii)** The duplication of Leu[L]170 affects residues in the core of ENG OR2, where Leu170 is involved in a number of hydrophobic interactions. The duplication most likely affects the folding of ENG, rather than directly impacting its function: the insertion could either disrupt the register of the N-terminal part of OR2 β-strand 12 or, more likely, cause one additional residue to be accommodated in the loop that follows the same β-strand. In the latter case, the extra Leu would take the place of Arg[R]171, disrupting a hydrogen-bond with Glu[E]195 as well as a stacking interaction with Arg[R]192. Moreover, by shifting Arg[R]171 to take the place of Leu[L]172, the duplication would replace a hydrophobic residue with a charged one at the bottom of the hydrophobic core of the OR2 β-sandwich. **iv.** Three variants that are benign in terms of HHT pathogenesis (Gly413Asp; Asp446Gly and Arg571His) affect residues that are all exposed on one face of the ENG ZP module[37]. The three variants were independently assigned as benign without reference to the tertiary structure, due to presence in the same HHT DNA as an *ENG* splice, nonsense (stop) or frameshift variant respectively. **B)** Bar plot of the number of pathogenic or likely pathogenic variants in the HHT Mutation database and in Cohort 1 broken down by sequence ontology (SO) term in each of *ENG*, *ACVRL1* and *SMAD4*. The upper bars give the number in the HHT Mutation database in 2018[40] (620 variants in total); the lower bars give the number in the current cohort, with novel variants highlighted in dark blue (106 variants in total).

Figure 4: Phenotypic prediction accuracy for samples with pathogenic variant.



A) The predicted (triangles) and observed (circles) frequencies of eight *a priori* discriminatory HPO terms in Cohorts 1 and 2. **B)** Predicted causal gene displayed for each of the observed HHT genotypes in Cohort 1. **Upper panel**: Clinician prediction using all HPO terms– see also *Supplemental Table 5*. **Lower panel:** Automated prediction through Bayesian modelling of the 8 discriminatory HPO terms (for further details see *Supplemental Methods, Supplemental Figure 3 and Supplemental Table 6*).

Figure 5: Quantitative red cell traits



A) Distributions of quantitative red cell traits in HHT and control populations: Total red blood cell count (left), hematocrit (centre) and hemoglobin (right) plotted for the HHT patients (one measurement per patient, proband and affected family members from Cohorts 1 and 2) above the respective INTERVAL population distribution from 50,000 blood donors (one result per donor).[69] Upper panel males, lower panel females. Although the median values are similar, it will be noted that HHT cases had a higher proportion of extreme red cell values (both high and low) relative to healthy controls: For red cell count, hematocrit and hemoglobin respectively, the proportion of HHT patients within the 5th-95th sex-stratified percentiles of the INTERVAL ranges were only 66%, 61% and 50% for males; 64%, 55% and 48% for females (all p values <0.0001). **B) Relationships with bleeding and hypoxemia in HHT cohort:** Patients with more severe blood losses ("bleeders") were defined by a dynamic bleeding scale ≥4, and subcategorised by the presence (purple symbols) or absence (blue symbols) of pulmonary (P)AVMs which impair gas exchange resulting in lower arterial partial pressure of oxygen (PaO_2) and hence lower oxygen saturation of haemoglobin (SaO_2). Patients with lower bleeding scale status were also categorised by the presence (green) and absence (red) of PAVMs. The graphs (upper panel males, lower panel females), plot total red blood cell count (left), hematocrit (centre) and hemoglobin (right) against same-day oxygen saturation measured by finger oximetry for 10 minutes standing using one measurement per patient (proband and affected family members). Note that in all six analyses, the patients with greater bleeding (red and purple) tended to have lower red blood cell indices ($p<0.0001$ in all cases), and there was a superimposed anticorrelation between the red cell indices and SaO_2 ($p<0.0001$ in all cases). **C) Oxygen saturation (SaO_2) in HHT patients:** Histograms of oxygen saturation (SaO_2) in *ACVRL1* and *ENG* cases. **D) Bleeding status in HHT patients:** Histograms of bleeding scores in *ACVRL1* and *ENG* cases

Mutational and phenotypic characterisation of hereditary hemorrhagic telangiectasia

Claire L Shovlin*†, Ilenia Simeoni, Kate Downes, Zoe Frazer, Karyn Megy, Maria Bernabeu Herrero, Abigail Shurr, Jennifer Brimley, Dilip Patel, Loren Kell, Jonathan Stephens, Isobel Turbin, Micheala Aldred, Christopher Penkett, Kathy Stirrups, Willem H Ouwehand, Luca Jovine and Ernest Turro*†

DATA SUPPLEMENT

SUPPLEMENTAL METHODS:

CLINICAL COHORTS OF HHT FAMILIES:

Ethical Approvals

Molecular studies on families with hereditary haemorrhagic telangiectasia and arteriovenous malformations was ethically approved by Scotland A MREC Ethics Committee as MREC 98/0/42, and subsequently 07/MRE00/19 which remains valid -the study is also registered on the National Clinical Trials Database as NCT00230685. *Case notes review on patients with hereditary haemorrhagic telangiectasia (HHT)* was ethically approved by the Hammersmith, Queen Charlotte's, Chelsea, and Acton Hospital Research Ethics Committee and the approval remains valid (PI Shovlin for all).

Cohort 1 (n=183)

Patients and family members attending a single UK institution for management of HHT were consented for research genetic studies. Particularly for earlier families (1992-2002) there was an ascertainment bias towards pulmonary arteriovenous malformation (AVM) families due to the pulmonary AVM embolisation programme that commenced in 1984.[1] DNA was extracted either during initial research studies (using phenol chloroform extraction[2-4]); or in bulk in 2015 using QIAamp DNA Blood Midi Kit. DNA was quantified using Nanodrop and QuBit: DNA concentrations ranged from 50.2-253ng/µl: for sequencing 5µg was supplied in TrisEDTA buffer.

Primary phenotyping comprised a full clinical history including personal and family symptoms/ complications from HHT, and clinical evaluation of HHT telangiectasia by an experienced HHT clinician (CLS). All patients were offered screening/investigation for pulmonary AVMs (using mean oxygen saturation (SaO_2) after standing for 7-10 minutes and thoracic CT scans); and iron deficiency anemia (using complete blood counts, serum iron, transferrin saturation index (T*f*SI) and ferritin), as described elsewhere.[5-7] Patients with symptoms suggestive of other visceral AVMs underwent relevant investigations, but formal screening was restricted to selected cases. A clinical diagnosis of HHT was made in the presence of at least three Curaçao Criteria,[8] and management conducted according to clinical need. The medical records of patients on whom DNA had been sent for sequencing were evaluated retrospectively using a templated record. Human Phenotype Ontology (HPO)[9] terms and numerical values were selected from clinical notes, patient letters, radiology systems and electronic blood records. The score for each HPO term was initially "0" if definitely excluded (only feasible if the feature develops early in life and is evident, or patient had been screened); "1" if present, and otherwise unknown ".". For 8 phenotypes, a severity weighting of '2' was incorporated to indicate particularly florid or severe disease such as: nosebleeds daily and/or needing specialised ENT treatments; pulmonary AVMs resulting in SaO_2 of 90% or less; a fatal hemorrhage; iron deficiency anemia requiring blood transfusions or intravenous iron; two or more brain abscesses, two or more ischemic strokes, or an intracranial hemorrhage at <30ys of age. A '3' was used to record a complication in pregnancy (as there are no relevant HPO terms, and by definition, hemorrhages in lungs or brain in pregnancy are more severe/life-threatening as two lives are at risk[10]). In total, 301 different HPO terms were captured across the sequenced probands and their relatives prior to the prediction. Across all 183 families, a further 156 relatives had a clinical diagnosis of HHT, and the families provided phenotypic data on a mean of 1.85 affected individuals per DNA.

Using the available family phenotypes, additional predictions were made by an experienced HHT clinician (CLS) in 2016, specifically 1) whether the family was likely to have HHT or a different inherited vasculopathy; 2) whether the condition (e.g. pulmonary AVM +/- nosebleeds) was likely to be sporadic; 3) assuming HHT was present, which gene defect was predicted (haploinsufficiency of one or more of *ACVRL1*; *ENG; SMAD4*, *GDF2*; with or without suspicion of a new gene (non *ENG/ACVRL1/SMAD4/GDF2*) due to unusual familial characteristics); 4) an indication where there was insufficient data to predict the implicated gene with confidence and 5) if a specific gene was predicted as harbouring the causal variants, a weighting indicating the subjective confidence in the prediction. The usual confidence level was 95:5, increasing to 98:2 where the prediction was supported by copious clinical details in multiple family members, reducing if less confident.

Cohort 2 (n=94)

The molecular study cohort had excluded families where the HHT pathogenic variant was already known, from the group's earlier research studies[2-4] or clinical sequencing programmes. To provide a replicate data set from the same geographical background, additional families with a prior molecular diagnosis were brought into the analysis of the clinical phenotypic spectrum, with ethical approvals (LREC 00/5792). As in the published literature, there was an *ENG* bias for pulmonary AVMs, *ACVRL1* bias for hepatic AVMs, and all juvenile polyposis cases were attributable to *SMAD4.*

Bleeding, red cell, and oxygenation indices (cohorts 1 and 2)

With ethical approvals (LREC 00/5792), for both cohorts, all available HPO, bleeding phenotypes and quantitative red cell and oxygenation indices were collected on the probands and affected family members reviewed in the service. The captured status was generally that at final clinical assessment (i.e. having had the opportunity to develop further phenotypic traits), with the exception of "HHT severity" indices such as hypoxemia prior to treatment of PAVMs, and anaemia prior to treatment. For each individual, the lowest SaO_2, both lowest and highest hemoglobin, mean corpuscular volume (MCV), and platelet volumes were recorded. For red cell indices, however, only the first recorded full dataset (bleeding scale, SaO_2 and red cell indices) for each patient were analysed, irrespective of whether this preceded natural history or therapeutic corrections.

DISTRIBUTIONAL MODELLING STAGES

Determination of discriminatory phenotypes

As described in the main text, published evidence was used to identify 8 HPO terms that were thought to discriminate between the HHT genotypes corresponding to pulmonary arteriovenous malformation (AVM), cerebral AVM, spinal AVM, hepatic AVM, pulmonary hypertension, colorectal polyposis, juvenile gastrointestinal polyposis, and dilatation of the ascending aorta.

Disease expert input:

We used independent beta distributions to model the gene-specific prior probabilities of a patient having the 8 HPO terms, incorporating recognition of local screening practices. For each discriminatory HPO term and gene, two shape hyperparameters (α and β) were assigned. *SUPPLEMENTAL FIGURE 3* shows probability density functions for 3 HPO terms -pulmonary AVMs, hepatic AVMs, and pulmonary hypertension- in 2 genes. *SUPPLEMENTAL TABLE 6* provides the full list of expert clinician-assigned hyperparameter values.

Clinician-independent prediction of causal gene for each individual

Using gene-specific prior distributions of trait frequencies for the 8 discriminatory HPO terms and the expected prevalence of *ACVRL1*, *ENG* and *SMAD4*-related HHT in the study population (27%, 67% and 6%, respectively), we inferred a posterior probability of causality of each of the 3 genes. As detailed in the methods, the gene with the highest posterior probability, was predicted to be the causal gene.

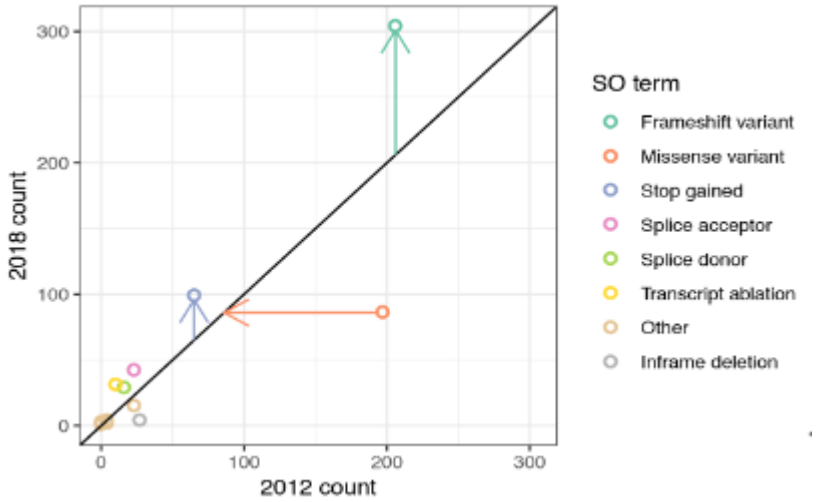
General Applicability

Out of the α and β shape parameter values and gene prevalences, 6 values were specific to the current study population, but they could be adjusted to suit local population characteristics. For example, a service that routinely screened all HHT patients for cerebral AVMs would expect a higher proportion of cases with that phenotype, in turn demanding different parameter values**.** The initial gene-specific prior distributions of trait frequencies in the study population were effectively dependent on disease expert input, but subsequent modifications can be readily employed.

REVIEW OF VARIANT CLASSIFICATION SYSTEMS

Evolution of variant classifications

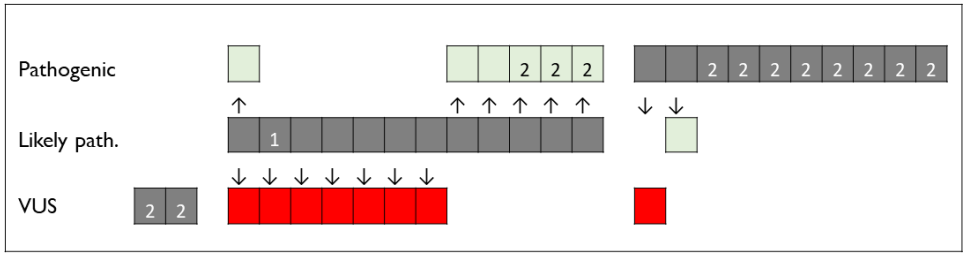
Molecular diagnostic studies have become increasingly stringent in the assignment of pathogenicity status to variants.[11,12] This is relevant to the current study since in the past, HHT-causal "mutations" in *ENG*, *ACVRL1*, and *SMAD4* were defined with less rigour. For example, in the ARUP Laboratories HHT Mutation Database,[13] the total number of variants reported in these genes increased between 2012[14] and 2018, as demonstrated by an increased number of Pathogenic frameshift and stop gain variants (*SUPPLEMENTAL FIGURE i*). Despite this increase, the total number of missense variants considered to be Pathogenic or Likely Pathogenic fell, as fewer met the more stringent requirements (*SUPPLEMENTAL FIGURE i*).



SUPPLEMENTARY FIGURE i: Evolution of the number of variants meeting Pathogenic or Likely Pathogenic status on the HHT Mutation Database[13] in the latest update (January 2018) compared to 2012 (published in [14]). SO: Sequence Ontology.

Current discrepancies in variant classifications

The current literature contains major discrepancies in pathogenicity assignments within HHT genes, depending on the applied criteria. This involves clinical diagnostic calls, in addition to recognised HHT database repositories- the ARUP Laboratory's HHT Mutation Database,[13] and ClinVar[15]. For example, of the missense variants for which our service had both clinical calls (from one of 4 different diagnostic laboratories), with the variant also present on at least one of ClinVar[15] or the HHT Mutation Database,[13] only 10/34 (29.4%) were called consistently (*SUPPLEMENTAL FIGURE ii*). Some variability (e.g. between Likely Pathogenic and Pathogenic calls) will have little consequence in clinical and research settings, but 8/32 (25%) of clinical laboratory Pathogenic or Likely Pathogenic calls were graded as a non-actionable variant in at least one of the two database repositories (*SUPPLEMENTAL FIGURE ii*).



SUPPLEMENTAL FIGURE ii. Diagram illustrating the concordance/discordance in laboratory pathogenicity assignations for 24 variants reported to our service in the last 10 years (grey boxes) with respect to the 2019 ClinVar[15] and HHT Mutation Database[13] assignments. Grey boxes indicate the Clinical Diagnostic Laboratories reports of Pathogenic (upper row), Likely Pathogenic (Likely path., middle row) or variant of uncertain significance (VUS, lower row). Consistent calls are denoted by white numbers within the grey boxes for one (1) or both (2) databases[13,15]. In addition to inconsistent Pathogenic↔Likely Pathogenic calls denoted by pale green boxes, eight Likely Pathogenic or Pathogenic clinical calls were classified as a VUS or equivalent ('pending classification') by at least one report on by ClinVar (N=3) (N=3), or by the HHT Mutation Database (N=5).

We concluded first that there was a need for a consistent approach, particularly to address the now high proportions of VUSs that are being generated for this rare disease, and secondly that there was no current gold standard reference for HHT variant calls.

We therefore evaluated available pathogenicity criteria in detail for application to the current study.

PATHOGENICITY CRITERIA CONSIDERATIONS

Overview of pathogenicity assignments

The mappings first proposed by Plon et al[11] suggested using "definitely pathogenic" if the subjective probability of pathogenicity was considered >0.99; "likely pathogenic" if between 0.95 and 0.99; "uncertain significance" if between 0.049 and 0.95, and "benign/likely benign" for the remainder.

Although some have used a subjective probability of >0.90 as sufficient to weight likely pathogenic, in our opinion this is insufficiently stringent given clinical harm could result from mis-assignment of pathogenic status and subsequent erroneous cascade testing. We therefore retained considerations in Plon et al[11] during assignments of pathogenicity.

Criteria providing evidence for Pathogenicity

The ACMG/AMP criteria for classifying pathogenic variants[12], span elements of allele frequency, consequence predictions, phenotype information and the results of cosegregation studies (*Supplemental Figure iii*). Not all criteria were appropriate to use in the current study: Of the 16 ACMG pathogenic criteria, only 9 were considered at least in part applicable– one very strong (**PVS1**); one strong (**PS3**); three moderate (**PM1**, **PM2**, **PM4**) and four supporting (**PP2**, **PP3**, **PP4** and **PP5**, yellow in table below). It was noted that **PVS1**, **PM4** and [**PS3**/**PM1**] are mutually exclusive. Study design prevented the use of **PS2**, **PS4**, **PM3**, **PM6** and **PP1**. Biological considerations led to the non-use of **PS1**, **PM5**, and more restricted application of **PP2** and **PP3**.

Further details on modifications employed for the pathogenicity criteria, and their specific consequences, are provided in the following section. The main modification was that by promoting **PM2** (absent from controls, or at very low frequency) to strong, there was no potential to allow frameshifts, stop codon and other clear nulls to be classified as variants of uncertain significance (as currently occurs *Supplemental Figure iv*), and no longer necessary to apply what we considered biologically weaker criteria of **PS1**, **PM5**, and elements of **PP2** and **PP3** in order to meet likely pathogenic status.

Pathogenic criteria	ALLELE FREQUENCY	VARIANT SPECIFIC	PATIENT PHENOYPE	FAMILY-DATA	REPUTABLE DATABASE
Very strong evidence		**PVS1-** Null : frameshift, stop gain, canonical splice, start codon loss or multiexon deletion where loss of function known mechanism			
Strong evidence	~~**PS4** Significantly increased frequency in patients compared to controls~~	~~**PS1** Same amino acid change not used because aa change may not be mol defect~~		~~**PS2** *De novo* trio study~~	
	PM2 ↑	**PS3** Data from functional studies			
Moderate evidence	**PM2** Allele absent in controls	**PM1-** Mutational hotspot/ critical domain.		~~**PM3**-Trans~~	
		PM4: Protein length change (inframe indels, loss of stop codons)		~~**PM6** Assumed de novo no parental confirmation~~	
		~~**PM5**: Different change to same aminoacid- illogical for 3ary structure as different aminoacids tolerated to different degrees~~			
		PP3 (Splicing only) ↑			
Supporting evidence		~~**PP3** Multiple lines of computational evidence re conservation/evolotion~~ Modified to splicing only	**PP4** Specific phenotype	~~**PP1** Cosegregation~~	**PP5** Reputable source reports
		~~**PP2** Missense in whole gene considerations~~ Modified to cysteine only			

Supplemental Figure iii: The 16 categories of ACMG/AMP[12] evidence for pathogenicity categorised by type of evidence, and whether used unchanged in the current study (yellow highlight, black text); modified for use in the current study (blue text/highlight); or not used (strike-through text). Note that study design was the primary reason why **PS2**, **PS4**, **PM3**, **PM6** and **PP1** were not used. Biological considerations led to non-use of **PS1**, **PM5**, and restricted application of **PP2** and **PP3**. *Supplemental Figures iv* and *Supplemental Figure v* provide further details on the implications of the modifications, separated by whether the variant was categorised as **PVS 1** (*Supplemental Figure iv*), or in the other 4 categories (**PS3**, **PM1**, **PM4** or **PP3A**, *Supplemental Figure v*).

Modifications Employed:

1) PM2 upgrade

Rationale for modification of PM2:

i) ACMG sets a precedent for modifying the strength of a criterion: In ACMG, there is a "strong" pathogenicity criterion (**PS4**) for rare alleles if there is statistical evidence of a higher prevalence in patients than controls. However, in rare disease situations, this is downgraded to moderate or even supporting evidence.

ii) If the ACMG criteria were strictly applied, a novel clear null could be assigned as pathogenic with 1 very strong + 1 moderate +1 supporting criteria **(PVS1**, **PM2** and **PP4**). However, it could be impossible to assign likely pathogenicity to a novel single exon/large inframe deletion/+4 splice site variants, as the maximum criteria could be two moderate+1 supporting criteria **([PM1 or PM4]; PM2 and PP4)**. This was considered inappropriate for a rare disease research study, and **PM2** was upgraded to "strong".

iii) PM2 is strictly used for an allele absent in control populations or at very low frequencies for recessive disorders. HHT is a condition that may be non-penetrant, and/or require expert evaluation for diagnosis. As the general population databases are likely to include occasional HHT cases, **PM2** was also applied at very low frequencies.

iv) The **PM2** upgrade enables clinical colleagues to adapt, if they require.

Consequences:

As shown in *Supplemental Figure iv*, the increased simplicity would have prevented 25 clear nulls being misclassified as a VUS or only "Likely Pathogenic" on existing databases. It also moved 31 variants in the current cohort to a Likely Pathogenic category without using **PS1**, **PM5**, or all elements of **PP2** and **PP3:**

	Current Cohort	ClinVar	HHT Mutation Database
A) PVS1 - null variants			
Likely/Benign↔VUS	0	0	0
VUS→ Pathogenic	0	4 (1.4%) *CV 1-4*	3 (0.6%) *HMDB 1-3*
Likely Pathogenic →Pathogenic	0 *a*	21 (4.3%) *CV 5-19*	3 (0.6%) *HMDB 4-6*
B) PS3, PM1, PM4 and PP3- potential null variants:			
Likely/Benign↔VUS	0	0	0
VUS→ Likely Pathogenic *b*	31 *TG 1-31*	Not calculated	Not calculated
Likely Pathogenic →Pathogenic	0	Not calculated	Not calculated

Supplemental Figure iv: Results of **PM2** upgrade on variants within the current cohort, ClinVar[15] and the HHT Mutation Database[13] accessed 03/03/2020. **A) PVS1**: multiexon structural variants, pan exon and within exon frameshifts, stop codons that would generate premature termination codons (at least 55 base pairs upstream of the final exon-exon boundary),[14] loss of the transcriptional start site, or loss of canonical AG splice acceptor or GT donor sites. **B)** Inframe indels **(PM4)**, and variants supported by functional (**PS3**, **PM1**) or computational (**PP3**) data.

Footnotes to Supplemental Figure iv :

a. 70 variants were categorised as **PVS1** in the current cohort.

b. Variants CV1-21 and HMDB 1-6 also meet **PVS1**.

c. Specific variants:

ClinVar[15] assignments:

CV1: *ACVRL1* c.1478del (p.Ser493fs); **CV2**: *ACVRL1* c.1507C>T (p.Gln503Ter); **CV3**: *ENG* c.742_743insT (p.Asp248fs); **CV4**: *SMAD4* c.1647delA (p.Gln549fs); **CV5:** *ACVRL1* c.100dup (p.Cys34fs); **CV6**: *ACVRL1* c.302del (p.Leu101fs); **CV7** *ACVRL1* c.1406_1413dup (p.Trp472fs); **CV8**: *ENG* c.-8_8del (p.Met1fs); **CV9**: *ENG* c.408del (p.Glu137fs); **CV10**: *ENG* c.583del (p.Glu195fs); **CV11**: *ENG* c.780_789dup (p.Asp264fs); **CV12**: *ENG* c.895del (p.Leu299fs); **CV13**: *ENG* c.1657del (p.Leu553fs); **CV14**: *ENG* c.259C>T (p.Gln87Ter); **CV15**: *ENG c.1479C>A (p.Cys493Ter);* **CV16**: *ENG* c.68-1G>A; **CV17**: *ENG* c.816+1G>A; **CV18**: *ENG* c.1429-2A>G; **CV19**: *SMAD4* c.955+1delG

HHT Mutation Database[13] assignments:

HMDB1: *ACVRL1* c.1436del (p.Arg479Hisfs); **HMDB2** *ACVRL1* c. 1450delinsTG (p.Arg484Trpfs); **HMDB3**. *ACVRL1* c.1453dup (p.Ile485Asnfs); **HMDB 4** *ACVRL1* c.1043_1048+1dup (p.Asp348fs)**; HMDB 5** *ENG* c. 1015_1042del (p.Ala339Argfs); **HMDB6** *ENG* c. 1311+1dup.

Current Cohort assignments (note numbers newly allocated for manuscript):

TG1: *ACVRL1* c.916_917insTCA, p.(Ala306delinsValThr; **TG2**: *ACVRL1* c.830C>A, p.(Thr277Lys); **TG3**: *ACVRL1* c.830C>G, p.(Thr277Arg); **TG4**: *ACVRL1* c.848G>T, p.(Gly283Val); **TG5**: *ACVRL1* c.905T>C, p.(Leu302Pro); **TG6**: *ACVRL1* c.1037_1039delTCG, p.(Ile346_Ala347delinsThr); **TG7**: *ACVRL1* c.1298C>A, p.(Pro433His); **TG8**: *ACVRL1* c.1427C>G, p.(Pro476Arg); **TG9**: *ACVRL1* c.1456_1458delAAG, p.(Lys486del); **TG10**: *ACVRL1* c.1459A>C, p.(Lys487Gln); **TG11**: *ACVRL1* c.269G>A, p.(Cys90Tyr); **TG12**: *ACVRL1* c.1246+5G>A; **TG13**: *ACVRL1* c.200G>A, p.(Arg67Gln); **TG14**: *ENG* c.511_512insTCC, p.(Leu170dup); **TG15**: *ENG* c.620G>A, p.(Cys207Tyr); **TG16**: *ENG* c.646A>G, p.(Lys216Glu); **TG17**: *ENG* c.763G>T, p.(Gly255Cys); **TG18**:*ENG* c.896T>G, p.(Leu299Arg); **TG19**: *ENG* c.899T>C, p.(Leu300Pro); **TG20**: *ENG* c.1484T>C, p.(Leu495Pro); **TG21**: *ENG* c.360+5G>A; **TG22**: *ENG* c.523G>T, p.(Ala175Ser); **TG23**: *ENG* c.1268A>G, p.(Asn423Ser); **TG24**: *ENG* c.1311+5G>C; **TG25**: *ENG* c.-127C>T; **TG26**: *ENG* c.991G>A, p.(Gly331Ser); **TG27**: *ENG* c.1134G>A, p.(Ala378Ala); **TG28**: *ENG* gain exon 12-13 (125 bp); **TG29**: *SMAD4* c.1058A>G, p.(Tyr353Cys); **TG30**: *SMAD4* c.1081C>T, p.(Arg361Cys; **TG31**: *SMAD4* c.1309G>A, p.(Val437Ile/splice).

The consequence of other modifications (to **PS3, PM1, PP2, PP3**) and non-use of **PS1** and **PM5**, are shown for the current cohort in *Supplemental Figure v:*

	Modified Pathogenic Criteria					Not used		Benign
	PS3 *a* research function	PM1 *b* research structure	PP2 *c* missense/ cysteines	PP3 *d* excludes conserved	PP3 *e* splicing upgraded	PS1 same αα	PM5 different Aα	BP5 *f* Second variant & others
Upgraded								
Benign/LB→VUS	0	0	0	0	0	0	0	0
Benign/LB→ Likely Pathogenic	0	0	0	0	0	0	0	0
VUS→ Likely Pathogenic	1 TG32	4 TG14-16,18	1 TG33	0	6 T31, 42-46	0	0	0
Downgraded								
Benign/LB ←VUS	0	0	0	0	0	0	0	0
VUS ←Likely Pathogenic	0	0	4 TG34-37	6 TG 38-41	0	0	0	0
Benign/LB← Likely Pathogenic	0	0	0	0	0	0	0	0

SUPPLEMENTAL FIGURE v : Potential results of other pathogenicity criteria modifications on variants within the current cohorts, depending on PM2, as discussed in detail opposite/below. αα: amino acid, LB likely benign.

a **PS3:** Using **PM2** as strong, the modification made no difference to assignments. If **PM2** was a moderate criterion, non-use of research laboratory data would affect only one variant, likely because the reputable sources had reviewed most of the same published data.

b **PM1**: Using **PM2** as strong, the modification made no difference to assignments. Variants listed referring to if **PM2** was used as a moderate criterion.

c **PP2:** If **PM2** was a moderate criterion, then the use of **PP2** only as 'supporting evidence' would leave one novel cysteine-loss missense variant as a VUS. Using **PM2** as strong as here, inclusion of **PP2** for any amino acid would make 4 VUSs (*ENG)* Likely Pathogenic.

d **PP3:** Although losing evolutionary/conservation elements would not have resulted in any changes in the current cohort with **PM2** left as moderate, it did move 6 otherwise "Likely Pathogenic" to VUS, using **PM2** as strong.

e **PP3:** Splicing criteria applied to two variants that met **PP5** assigning as Likely Pathogenic.

f **BP5:** In the current cohort, all variants would have been assigned as Benign due to high population frequency (see data in *SUPPLEMENTAL TABLE 3*)

Variant lists by column and with **TG1-TG32** numbered as in *SUPPLEMENTARY FIGURE v*:

a **PS3:** **TG32** *ENG* c.360+4A>G;

b **PM1**:**TG14** *ENG* c.511_512insTCC, p.170dup; **TG15** *ENG* c.620G>A, p.Cys207Tyr; **TG16** *ENG* c.646A>G, p.Lys216Glu; **TG18** *ENG* c.896T>G, p.Leu299Arg;

c **PP2:** **TG33***: ACVRL1* c.269G>A, p.(Cys90Tyr); **TG34** *ENG* c.35T>C, p.(Leu12Pro); **TG35** *ENG* c.239T>C, p.(Leu80Pro)*;* **TG36** *ENG* c.497A>C, p.(Gln166Pro)*;* **TG37** *ENG* c.776T>G, p.(Val259Gly);

d **PP3: exclusions:** **TG38:** *ENG* c.35T>C, p.(Leu12Pro); **TG39** *ENG* c. 239T>C, p.(Leu80Pro); **TG40** *ENG* c.497A>C, p.(Gln166Pro); **TG41** *ENG* c.776T>G, p.(Val259Gly);

e **PP3: splicing:** **TG31**: *SMAD4* c.1309G>A; **TG42:** *ACVRL1* c .1246+5G>A; **TG43:** *ENG* c.360+5G>A; **TG44:** *ENG* c. 523G>T; **3TG45** *ENG* c.1268A>G; **TG46:** *ENG* c.1311+5G>C. (see data in *SUPPLEMENTAL TABLE 2B*)

2) PS3 Data from functional studies

Rationale for extension of PS3

For this rare disease, functional expertise lies primarily in research laboratories where appropriate experimental design can ensure that the specific variant is being evaluated, rather than a general HHT cellular phenotype. For this study, the category could not access unpublished clinical laboratory evidence, but included published research laboratory data in human cells that demonstrated variants resulted in i) endoplasmic reticulum retention,[16] failure of variant protein maturation (pulse chase experiments),[17] abnormal splicing,[18,19] or modified relevant intracellular signalling pathways.[20]

3) PM1: Data from domain predictions

Rationale for extension of PM1

The phrase "emerging" is more applicable than "well-established" for data from ref[21].

4) Modified use of PP2 (cysteine only, upgraded to moderate evidence)

Rationale for modification of PP2

i) Distinction of cysteines: The recognised role of disulphide bonding in the tertiary structures of the *ENG*, *ACVRL1*, and *SMAD4* proteins emphasise cysteine loss/gain substitutions more likely deleterious than other amino acids (upgraded to moderate).

ii) Non-use of **PP2** for other amino acids: With the exception of cysteines, the data in Figure 2 did not support the use of PP2 for any other missense substitutions.

5) PP3 not used for evolutionary/conservation, and upgraded for splicing

Rationale for modification of PP3:

i) ACMG sets the precedent for modifying the strength of a criterion (see above).

ii) While at a full gene level, scores can be informative (Figure 2), current study overlaps in SIFT, Align-GVD, and Grantham distances led us not use these elements.

iii) With the published evidence for splicing functional effects in HHT, **PP3** (splicing only) was upgraded to a criterion of moderate evidence. **PP3** was considered met for splice variants if at least 3 of Human Splicing Finder, Splice Site Finder-like (SSF), MaxEntScan, NNSplice, and GeneSplicer[22] gave a score exceeding 10% of the maximum score for the corresponding algorithm (see data in *SUPPLEMENTAL TABLE 2B*)

Consequences: The changes better reflected evidence, and meant that VUSs TG 38-41 were not assigned prematurely to Likely Pathogenic due to the upgraded **PM2** criteria.

6) Non-use of PS1

Rationale for non-use

i) Allowing the same amino acid as a previous variant but different nucleotide change does not allow for the possibility that the variant is affecting splicing or an alternate noncoding function of the gene. Our concern is illustrated by *ENG* c.1134G>A (p. Ala378Ala), and by *SMAD4* c.1309G>A (p.(Val437Ile) with nucleotide-based computational evidence that splicing is affected.

ii) Although not relevant to any variants in the current study, when using **PM2** as strong, also using **PS1** would have the potential to upgrade a variant either from Likely Pathogenic to Pathogenic (e.g. for a rare variant meeting **PS4**), or to move a variant of uncertain significance (VUS) to Likely Pathogenic status.

7) Non-use of PM5

Rationale for non-use:

i) Allowing a different amino acid at the same residue as a pathogenic substitution is not compatible with the biological knowledge that some amino acids will be better tolerated than others due greater similarity with the original amino acid. ***ii)*** Although not using **PS5** was not relevant to any variants in the current study, the recognised role of disulphide bonding in the tertiary structures of the *ENG*, *ACVRL1*, and *SMAD4* proteins support our view that a cysteine gain at a particular amino acid would be more damaging than a substitution by any other amino acid.

Criteria providing evidence for Benign Status

Again it was noted that not all criteria for classifying benign variants[12] were appropriate to use in the current study: Of the 12 ACMG benign criteria[12], only 3 were considered sufficient in the current study to move the variant towards benign status. These related to:

- Allele frequency in the general population (**BA1** >5%, **BS1** >0.02);
- Variant found in a case with an alternate molecular cause of disease (**BP5**);

  Where a DNA already harboured a pathogenic or likely pathogenic variant, we found that this was a very quick step to indicate that other variants in the same DNA *were likely* to have benign status. In the current cohort, all such variants also met stronger criteria for benign status based on allele frequency

Study design prevented the use of **BS4** and **BP2**; non-penetrant individuals with HHT prevented the use of **BS2,** molecular mechanism prevented use of **BP3,** and caution prevented our initial use of **BS3, BP4, BP6** and **BP7**. *Post hoc* review indicated that using these criteria would not have altered any calls.

COMPARISON OF CLASSIFICATIONS ON CURRENT DATA

Where current Cohort 1 (ThromboGenomics) variants were identified on either the 2020 ClinVar[15] and/or HHT Mutation Databases[13], listings were compared:

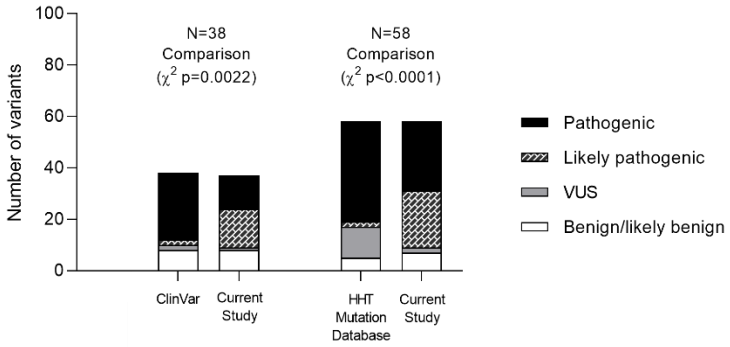
Likely Benign and Likely Pathogenic: No variant was listed in this way.

Likely Pathogenic: The HHT Mutation Database[13] and ClinVar[15] list significantly fewer Likely Pathogenic calls for the current dataset (*Supplemental Figure vi*):

- Current study Likely Pathogenic assignments reflected new insights from 3 dimensional structural modelling, overlooked published functional data, and in our view, appropriately reflect use of a subjective probability range of 0.95-0.99;
- 15 current Likely Pathogenic variants were listed as Pathogenic on [13,15] or both;
- 9 current Likely Pathogenic variants were listed as a VUS on either [13,15] or both;
- For 4 variants, the two databases[13,15] disagreed, with one assigning as a Pathogenic variant and the other as a VUS (see also *Supplemental Figure ii).*

Variants of Uncertain Significance: The HHT Mutation Database[13] and ClinVar[15] list many Likely Benign variants as VUSs (*Supplemental Figure vi*):

- Current study Likely Benign variants assignments initially reflected an alternate Pathogenic or Likely Pathogenic variant, and were then seen to be common in the latest large population databases potentially unavailable at the time of some original database calls. This supports a repository that identifies all disease gene variants found in DNAs already harbouring a (Likely) Pathogenic variant.
- 3 current Likely Benign variants were listed as a as a VUS on either[13,15] or both;
- For 1 variant, the two databases disagreed[13,15], with one assigning as a VUS and the other as Likely Benign .



Supplemental Figure vi. Listing of current cohort variants on existing databases. The left hand pair indicate the 38 variants also listed in ClinVar[15], the right hand pairing the 58 variants also listed in the HHT Mutation Database[13]. P values were calculated by χ^2 with 3 degrees of freedom. The small proportion of Likely Pathogenic calls for rare variants in HHT disease genes is also seen on the full database listings: ClinVar[15] lists 21/278 (7.6%) as Likely Pathogenic, the HHT Mutation Database[13] 3/490 (0.6%) compared to the current study 38/127 (30.0%), χ^2 p <0.0001with 2 degrees of freedom.

SUPPLEMENTARY RESULTS

SUPPLEMENTAL TABLES

SUPPLEMENTARY TABLE 1A: Pathogenic or likely pathogenic variants

GENE	DNA (*ACVRL1*: ENST00000388922)	PROTEIN	EXON	ACMG CRITERIA	CLASS	NEW*	HETEROZYGOTES
ACVRL1	c.138_139insC	p.(Arg47fs)	3	PVS1+PM2+PP4	5	1	1
ACVRL1	c.200G>A	p.(Arg67Gln)	3	PM2+PP4+PP5	4	0	2
ACVRL1	c.258delC	p.(Asn86fs)	3	PVS1+PM2+PP4	5	0	1
ACVRL1	c.269G>A	p.(Cys90Tyr)	3	PM2+PP2+PP4	4	0	1
ACVRL1	c.372_373insC	p.(Gly124fs)	4	PVS1+PM2+PP4	5	1	1
ACVRL1	c.475G>T	p.(Glu159*)	4	PVS1+PM2+PP4	5	0	2
ACVRL1	c.626-2A>G		intron 5	PVS1+PM2+PP4	5	1	1
ACVRL1	c.632G>A	p.(Gly211Asp)	6	PS3+PM2+PP4+PP5	4	0	5
ACVRL1	c.830C>A	p.(Thr277Lys)	7	PM2+PM1+PP4	4	0	1
ACVRL1	c.830C>G	p.(Thr277Arg)	7	PM2+PM1+PP4	4	1	1
ACVRL1	c.848G>T	p.(Gly283Val)	7	PM2+PM1+PP4	4	1	2
ACVRL1	c.905T>C	p.(Leu302Pro)	7	PM2+PM1+PP4	4	1	1
ACVRL1	c.916_917insTCA	p.(Ala306delinsValThr	7	PM2+ PM1+PP4	4	1	1
ACVRL1	c.1037_1039delTCG	p.(Ile346_Ala347delinsThr)	7	PM2+PM1+PP4	4	0	1
ACVRL1	c.1040delC	p.(Asp347fs)	7	PVS1+PM2+PP4	5	1	1
ACVRL1	c.1049_1055delCCACAGG	p.(Gly350Valfs)	8	PVS1+PM2+PP4	5	1	1
ACVRL1	c.1120C>T	p.(Arg374Trp)	8	PS3+PM2+PP4+PP5	4	0	4
ACVRL1	c.1171G>T	p.(Glu391*)	8	PVS1+PM2+PP4	5	0	1
ACVRL1	c.1231C>T	p.(Arg411Trp)	8	PS3+PM2+PP4+PP5	4	0	5
ACVRL1	c.1232G>A	p.(Arg411Gln)	8	PS3+PM2+PP4+PP5	4	0	1
ACVRL1	c.1246+5G>A		intron 8	PM2+PP3+PP4	4	1	1
ACVRL1	c.1298C>A	p.(Pro433His)	9	PM2+PM1+PP4	4	0	1
ACVRL1	c.1427C>G	p.(Pro476Arg)	10	PM2+PM1+PP4	4	1	1
ACVRL1	c.1435C>T	p.(Arg479*)	10	PVS1+PM2+PP4	5	0	3
ACVRL1	c.1448delT	p.(Leu483Argfs)	10	PVS1+PM2+PP4	5	1	1
ACVRL1	c.1456_1458delAAG	p.(Lys486del)	10	PM2+PM1+PP4	4	1	1
ACVRL1	c.1459A>C	p.(Lys487Gln)	10	PM2+PM1+PP4	4	1	1
ACVRL1	c.1468C>T	p.(Gln490*)	10	PVS1+PM2+PP4	5	0	1
GENE	**DNA (*ENG*:ENST00000373203)**	**PROTEIN**	**EXON**	**ACMG CRITERIA**	**CLASS**	**NEW***	**HETEROZYGOTES**
ENG	c.-127C>T		5'	PM2+PP4+PP5	4	0	2
ENG	c.1A>G	p.(Met1Val)	1	PVS1+PM2+PP4	5	0	2
ENG	c.39delG	p.(Leu14Trpfs)	1	PVS1+PM2+PP4	5	1	1
ENG	c.63_64insC	p.(Thr22Hisfs)	1	PVS1+PM2+PP4	5	1	1
ENG	c.146delT	p.(Val49Ala)	2	PVS1+PM2+PP4	5	1	1 ...../

…../ GENE	**DNA (*ENG:* ENST00000373203)**	**PROTEIN**	**EXON**	**ACMG CRITERIA**	**CLASS**	**NEW***	**HETEROZYGOTES**
ENG	c.155G>A	p.(Gly52Asp)	2	PS3+PM2+PM1+PP4+PP5	4	0	1
ENG	c.159C>A	p.(Cys53*)	2	PVS1+PM2+PP4	5	0	1
ENG	c.164delC	p.(Ala55Valfs*26)	2	PVS1+PM2+PP4	5	0	1
ENG	c.166C>T	p.(Gln56*)	2	PVS1+PM2+PP4	5	0	1
ENG	c.208G>T	p.(Glu70*)	2	PVS1+PM2+PP4	5	1	1
ENG	c.277C>T	p.(Arg93*)	3	PVS1+PM2+PP4	5	0	6
ENG	c.298_299delAG	p.(Ser100Cysfs)	3	PVS1+PM2+PP4	5	0	1
ENG	c.328C>T	p.(Gln110*)	3	PVS1+PM2+PP4	5	0	1
ENG	c.360+1G>A		intron 3	PVS1+PM2+PP4	5	0	2
ENG	c.360+4A>G		intron 3	PS3+PM2+PP4	4	0	2
ENG	c.360+5G>A		intron 3	PM2+PP3+PP4	4	0	1
ENG	c.370del	p.(Leu124Trpfs)	4	PVS1+PM2+PP4	5	0	3
ENG	c.496_497insC	p.(Gln166Profs)	4	PVS1+PM2+PP4	5	0	1
ENG	c.496delC	p.(Gln166Argfs)	4	PVS1+PM2+PP4	5	0	1
ENG	c.511C>T	p.(Arg171*)	4	PVS1+PM2+PP4	5	0	1
ENG	c.511_512insTCC	p.(Leu170dup)	4	PM1+PM2+PP4	4	1	1
ENG	c.519_544delCCAAGGTCAGTTTCCCCAGCAACCT	p.(Gln174Hisfs)	4	PVS1+PM2+PP4	5	1	1
ENG	c.523G>T	p.(Ala175Ser)	4	PM2+PP3+PP4	4	1	1
ENG	c.620G>A	p.(Cys207Tyr)	5	PM2+PM1+PP2+ PP4	4	0	2
ENG	c.646A>G	p.(Lys216Glu)	5	PM2+PM1+PP4	4	0	1
ENG	c.667_668insG	p.(Val223_Leu224fs)	5	PVS1+PM2+PP4	5	1	1
ENG	c.689+1G>A		intron 5	PVS1+PM2+PP4	5	1	1
ENG	c.690_706delGTATGGCTCTCGCCCCG	p.(Pro231Glufs)	6	PVS1+PM2+PP4	5	1	1
ENG	c.690-2A>G		intron 5	PVS1+PM2+PP4	5	1	1
ENG	c.736delG	p.(Asp246Ilefs)	6	PVS1+PM2+PP4	5	0	1
ENG	c.763G>T	p.(Gly255Cys)	6	PM2+PM1+PP2+PP4	4	1	2
ENG	c.767_768insAGGGTCC	p.(Pro257Glyfs)	6	PVS1+PM2+PP4	5	1	1
ENG	c.772delT	p.(Tyr258Thrfs)	6	PVS1+PM2+PP4	5	0	1
ENG	c.774C>A	p.(Tyr258*)	6	PVS1+PM2+PP4	5	0	1
ENG	c.778_779insGT	p.(Ser260Cysfs)	6	PVS1+PM2+PP4	5	1	1
ENG	c.817-1G>A		intron 6	PVS1+PM2+PP4	5	1	2
ENG	c.896T>G	p.(Leu299Arg)	7	PM2+PM1+PP4	4	1	1
ENG	c.899T>C	p.(Leu300Pro)	7	PM2+PM1+PP4+PP5	4	0	1
ENG	c.899_904_delinsCGGG	p.(Leu300fs)	7	PVS1+PM2+PP4	5	1	1
ENG	c.923C>A	p.(Ala308Asp)	7	PS3+PM2+PP4+PP5	4	0	1 …./

…./GENE	DNA (*ENG:*ENST00000373203)	PROTEIN	EXON	ACMG CRITERIA	CLASS	NEW*	HETEROZYGOTES
ENG	c.967_968delGT	p.(Val323Leufs)	7	PVS1+PM2+PP4	5	0	2
ENG	c.991G>A	p.(Gly331Ser)	7	PM2+PP3+PP4+PP5	4	0	1
ENG	c.992-2A>G		intron 7	PVS1+PM2+PP4	5	0	1
ENG	c.1080_1083delGACA	p.(Thr361Serfs)	8	PVS1+PM2+PP4	5	0	3
ENG	c.1134G>A	p.(Ala378Ala)	8	PM2++PP3+ PP4+PP5	4	0	1
ENG	c.1166_1168delTCT	p.(Phe389del)	9	PVS1+PM2+PP4	4	0	1
ENG	c.1183G>T	p.(Glu395*)	9	PVS1+PM2+PP4	5	1	2
ENG	c.1195delA	p.(Arg399Glyfs)	9	PVS1+PM2+PP4	5	0	1
ENG	c.1268A>G	p.(Asn423Ser)	9	PM2+PP3+PP4	4	0	1
ENG	c.1273-2A>T		intron 9	PVS1+PM2+PP4	5	1	1
ENG	c.1306C>T	p.(Gln436*)	10	PVS1+PM2+PP4	5	0	1
ENG	c.1311+5G>C		intron 10	PM2+PP3+PP4	4	1	1
ENG	c.1334delT	p.(Met445Argfs)	11	PVS1+PM2+PP4	5	0	1
ENG	c.1365C>A	p.(Tyr455*)	11	PVS1+PM2+PP4	5	1	1
ENG	c.1384C>T	p.(Gln462*)	11	PVS1+PM2+PP4	5	1	1
ENG	c.1411C>T	p.(Gln471*)	11	PVS1+PM2+PP4	5	1	1
ENG	c.1456_1457insGTCCCCATCCGTCTCCGAGT	p.(Phe486Cysfs)	12	PVS1+PM2+PP4	5	1	1
ENG	c.1472_1475delACAG	p.(Asp491Alafs)	12	PVS1+PM2+PP4	5	1	1
ENG	c.1484T>C	p.(Leu495Pro)	12	PM2+PM1+PP4	4	1	1
ENG	c.1490_1491insT	p.(Leu497fs)	12	PVS1+PM2+PP4	5	1	1
ENG	c.1513G>T	p.(Glu505*)	12	PVS1+PM2+PP4	5	0	1
ENG	c.1582_1583delCC	p.(Pro528Alafs)	12	PVS1+PM2+PP4	5	0	1
ENG	c.1626delC	p.(Thr544Profs)	12	PVS1+PM2+PP4	5	1	1
GENE	**DNA: human genome reference GRCh37**	**INDEL**	**EXON**	**ACMG CRITERIA**	**CLASS**	**NEW***	**HETEROZYGOTES**
ENG	LOSS:9:130605269-130618535 (13,266 bp)	del exon 1 & 2	1-2	PVS1+PM2+PP4	5	1	1
ENG	LOSS:9:130605269-130609751 (4,482 bp)	del exon 2 (152 bp, fs)	2	PVS1+PM2+PP4	5	0	1
ENG	LOSS:9:130586957-130592332 (5,375 bp)	del exon 3-7	3-7	PVS1+PM2+PP4	5	1	1
ENG	LOSS:9:130586252-130592332 (16,080 bp)	del exon 3-8	3-8	PVS1+PM2+PP4	5	0	1
ENG	LOSS:9:130586252-130593094 (6,842 bp)	del exon 3-8 #2	3-8	PVS1+PM2+PP4	5	1	1
ENG	LOSS:9:130581520-130593094 (11,574 bp)	del exon 3-10	3-10	PVS1+PM2+PP4	5	1	1
ENG	LOSS:9:130577259-130593094 (15,835 bp)	del exon 3-15	3-15	PVS1+PM2+PP4	5	1	1
ENG	LOSS:9:<130,587,973->130,588,140 (500 bp)	del exon 5 (166 bp, fs)	5	PVS1+PM4+PP4	5	1	1
ENG	LOSS:9:130586957-130588561 (1,604 bp)	del exon 5-7	5-7	PVS1+PM2+PP4	5	1	1
ENG	LOSS:9:130577259-130582329 (5,070bp)	del exon 9_15	9-15	PVS1+PM2+PP4	5	1	1
ENG	GAIN:9:130579377-130580669 (1,292 bp)	gain exon 12-13 (125 bp)	12-13	PM2+ PM4+PP4	5	1	2

GENE	DNA (*SMAD4* ENST00000398417)	PROTEIN	EXON	ACMG CRITERIA	CLASS	NEW*	HETEROZYGOTES
SMAD4	c.303G>A	p.(Trp101*)	4	PVS1+PM2+PP4	5	1	1
SMAD4	c.1058A>G	p.(Tyr353Cys)	10	PM2+PM1+PP2+PP4	4	1	1
SMAD4	c.1081C>T	p.(Arg361Cys)	10	PM2+PM1+PP2+PP4	5	0	1
SMAD4	c.1309G>A	p.(Val437Ile/splice)	12	PM2+PM1+PP3+PP4	4	1	1

SUPPLEMENTARY TABLE 1B: Benign or likely benign variants*

ACVRL1	ENST00000388922 c.88C>T	p.(Pro30Ser)	3	BS1 +BP2	2	0	1
ACVRL1	ENST00000388922 c.948G>C	p.(Glu316Asp)	7	BS1 +BP2	2	1	1
ACVRL1	ENST00000388922 c.1153A>C	p.(Ile385Leu)	8	BS1 +BP2	2	1	1
ENG	ENST0000037320 c.7C>T	p.(Arg3Cys)	1	BS1 +BP2	2	0	1
ENG	ENST0000037320 c.388C>T	p.(Pro130Ser)	4	BS1 +BP2	2	0	1
ENG	ENST0000037320 c.572G>A	p.(Gly191Asp)	5	BS1 +BP2+BP6	2	0	6
ENG	ENST0000037320 c.1019C>T	p.(Pro340Leu)	8	BS1 +BP2	2	0	1
ENG	ENST0000037320 c.1238G>A	p.(Gly413Asp)	9	BS1 +BP2	2	1	1
ENG	ENST0000037320 c.1337A>G	p.(Asp446Gly)	11	BS1 +BP2	2	1	1
ENG	ENST0000037320 c.1712G>A	p.(Arg571His)	13	BS1 +BP2+BP6	2	0	1
ENG	ENST0000037320 c.1844C>T	p.(Ser615Leu)	14	BS1 +BP2+BP6	1	0	2
GDF2	ENST00000249598 c.326T>C	p.(Val109Ala)	1	BS1 +BP2	2	0	1
GDF2	ENST00000249598 c.776A>G	p.(Asn259Ser)	2	BS1 +BP2	2	0	1
GDF2	ENST00000249598 c.997C>T	p.(Arg333Trp)	2	BS1 +BP2	2	0	1
SMAD4	ENST00000398417 c.1573A>G	p.(Ile525Val)	12	BS1 +BP2	2	1	1

SUPPLEMENTARY TABLE 1C: Rare Variants of Uncertain Significance (VUS)

ENG	ENST00000373203 c.35T>C	p.(Leu12Pro)	1	PM2+PP4~~+PP3~~	3	0	1
ENG	ENST0000037320 c.219+22C>T		3	PM2+PP4	3	0	1
ENG	ENST0000037320 c.239T>C	p.(Leu80Pro)	3	PM2+PP4~~+PP3~~	3	1	1
ENG	ENST0000037320 c.497A>C	p.(Gln166Pro)	4	PM2+PP4~~+PP3~~	3	1	1
ENG	ENST0000037320 c.776T>G	p.(Val259Gly)	6	PM2+PP4~~+PP3~~	3	1	1
GDF2	ENST00000249598 c.348T>C	p.(Asp116Asp)	2	PM2+PP4	3	1	1

SUPPLEMENTAL TABLE 1: List includes variant categories for ACMG/AMP Standards of Pathogenicity (defined in[12]): Very strong evidence (**PVS1**: null variant defined by variant type); strong evidence (**PS3**: well established functional studies, **PM2** [upgraded, see section above]: Absent from controls in Exome Sequencing Project, 1000 Genomes and ExAc); moderate evidence (**PM1**: located in a critical and well established functional domain); supporting evidence (**PP2** missense variant where low rate of benign variants: [loss of cysteine only, when upgraded see section above]; **PP3**: multiple lines of computational evidence[splicing only, when upgraded see section above]; **PP5**: reputable source reports as pathogenic). For Benign: Standalone (**BA1**: allele frequency >5% (removed by algorithm); strong evidence (**BS1**: allele frequency greater than expected); Supporting evidence (**BP2**: alternate higher impact allele [used instead of **BP5**: alternate molecular cause], **BP6**: reputable source report. Italicised: present but not needed for pathogenicity assignments. * All benign variants also met **PP4** based on patient origin, but this was not considered sufficient to override the other, stronger evidence and move the variant back to a VUS. Further details on individual assignments are provided in *SUPPLEMENTAL TABLE 2 (A-C).*

Supplemental Table 2A: MDT algorithm summary

Class 1	Class 2	←Term(s)		Class 3: Variant of uncertain significance (VUS)	Term(s)→		Class 4	Class 5					
Benign	Likely Benign	allele frequency	variant specific	all met PP4- specific phenotype	allele frequency	variant specifics	Likely pathogenic	Pathogenic					
				127					*ACVRL1*	*ENG*	*SMAD4*	*GDF2*	*Totals*
				↓									
				Transcript ablation (multi or pan/fs exon CNV)	PM2	PVS1	----------	**11**	0	11	0	0	11
				Frameshift variant	PM2	PVS1	----------	**31**	6	25	0	0	42
				Stop gained (premature termination codon)	PM2	PVS1	----------	**18**	4	13	1	0	60
				Start lost	PM2	PVS1	----------	**1**	0	1	0	0	61
				Splice acceptor (-1,-2) or donor (+1,+2) loss	PM2	PVS1	----------	**7**	1	6	0	0	68
				↓									
				59									
				↓									
				Functional data of loss of function	PM2	PS3	**7**		4	3	0	0	75
				↓									
				52									
				Predictive data critical structures									
				Further ALK1 kinase domain	PM2	PM1	**11**		11	0	0	0	86
				SMAD4 dwarfin domain	PM2	PM1	**3**		0	0	3	0	89
				ENG crystal domain	PM2	PM1	**7**		0	7	0	0	96
				Missense Cysteine loss	PM2	PP2	**1**		1	0	0	0	97
				Computational other splice site variant	PM2	PP3	**5**		1	4	0	0	102
				HHT Mutation Database pathogenic	PM2	PP5	**4**		0	4	0	0	106
				↓									
				21									
				↓									
	0	BA1		**Population allele frequency > 5%**					0	0	0	0	106
	15	BS1	BP5	**Alternate pathogenic variant in DNA**					3	8	1	3	121
	0	BS1	BP6	**HHT Mutation Database benign**					0	0	0	0	121
				↓									
				6									127
Benign or Likely Benign				**Variant of uncertain significance (VUS)**			**Pathogenic or Likely Pathogenic**						
15				6			106						

SUPPLEMENTAL TABLE 2A: Order in which criteria were applied to 127 separate rare variants entered the MDT algorithm at the top. These were moved from the VUS status towards pathogenic (blue) or benign (orange) status based on the ACMG/AMP criteria with minor modifications. For further details of how criteria were used in the assignments, see Supplementary Methods and *SUPPLEMENTAL FIGURES iii-iv.* The distribution in *ACVRL1*, *ENG*, *SMAD4* and *GDF2* is illustrated on the right of the Class assignment columns, and the final column indicates the running total of variants assigned at each step Many variants met more than the stated criteria (as outlined in *SUPPLEMENTAL TABLE 1*).

Supplemental Table 2B: Splice site variant details

Gene	Nucleotide change	Consequence	Site	Final Class	SpliceSite Finder-like	Max Ent Scan	NN SPLICE	Gene Splicer	Human Splicing Finder	HHT Mutation Database.[13]	Variant Number
Splice acceptor											
ACVRL1	c.626-2A>G	Splice site loss	intron 5	5	-100%	-100%	-100%	-100%	-100%	-	TG46
ENG	c.690-2A>G	Splice site loss	intron 5	5	-100%	-100%	-100%	-100%	-100%	Pathogenic	TG47
ENG	c.817-1G>A	Splice site loss	intron 6	5	-100%	-100%	-100%	-100%	-100%	-	TG48
ENG	c.992-2A>G	Splice site loss	intron 7	5	-100%	-100%	-100%	-100%	-100%	Pathogenic	TG49
ENG	c.1273-2A>T	Splice site loss	intron 9	5	-100%	-100%	-100%	-100%	-100%	-	TG50
Splice donor											
ENG	c.360+1G>A*	Splice site loss	intron	5[A]	-100%	-100%	-100%	-100%	-100%	Pathogenic	TG51
ENG	c.689+1G>A	Splice site loss	intron 5	5	-100%	-100%		-100%	-100%	-	TG52
Other splice variants											
ENG	**c.360+4A>G***	**Splice loss: exon 3 skip** [5]	**intron**	**5[B]**	**-100%**	**-77.0%**	**-100%**	**-100%**	**-9.70%**	**Pending class.**	**TG32**
ENG	c.360+5G>A*	Splice site loss	intron 3	4[C]	-100%	-100%	-100%	-100%	-14.2%	Pending class.	TG21
ENG	c.1311+5G>C*	Splice site loss	intron 10	4[D]	-100%	-84.2%	-100%	-50%	-13.2%	-	TG24
ACVRL1	c.1246+5G>A*	Splice site loss	intron 8	4[E]	-100%	-78%	-100%	-31%	-100%	-	TG52
ENG	**c.991G>A***	**Splice site loss, p.Gly331Ser**	**exon 7**	**4[F]**	**-100%**	**-59.0%**	**-47.4%**	**-45.2%**	**-11.7%**	**Pathogenic**	**TG53**
ENG	**c.1134G>A***	**Splice site loss, p.Ala378Ala**	**exon 8**	**4[G]**	**-14.6%**	**-52.1%**	**-50%**	**-35.7%**	**-88.4%**	**Pathogenic**	**TG54**
ENG	c.523G>T	Splice site loss, p.Ala175Ser	exon 4	4[H]	-14.0%	-78.2%	-100%	-79.2%	-11.5%	-	TG55
ENG	c.1268A>G	Splice site gain, p.Asn423Ser	exon 9	4[I]	.	+100% (gain)		+100% (gain)	+100% (gain)	Uncertain	TG23
SMAD4	c.1309G>A	Splice site loss, p.Val437Ile	exon 12	4[J]	-4.5%	-23.7%	-10%	-12.3%	-3.6%	-	TG31

Supplemental Table 2B: Bold indicates variants referred to specifically in text, * variant illustrated in *Supplemental Figure 2*. Note variant numbers newly allocated for manuscript.
A: All 5 algorithms predicted 100% loss for all 7 invariant splice donor and acceptor site variants.
B: Four of 5 algorithms (SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer, and HumanSplicingFinder) correctly predicted splice site loss with ENG c.360+4A>G which we have previously shown to result in an exon 3 skip in Epstein-Barr Virus (EBV)-transformed lymphoblastoid cells isolated from a different patient.[18] Although an alternate splice site was potentially generated, this would result in an out of frame transcript and likely nonsense mediated decay.[14,18]
C-E: Three +5 splice donor variants had similarly high predictions to canonical splice acceptor and splice donor sites (complete loss by ≥2 algorithms plus ≥2 predictions of > 50% loss).
F-J: For exon sequences adjacent to splice junctions, two of the identified variants (*ENG* c.1134G>A (**F**) and *ENG* c.991G>A (**G**)) are reported as Pathogenic on the HHT Mutation Database.[13] Three further novel variants were assigned as Likely Pathogenic based on similar predictions. Note *SMAD4* c.1309G>A (p.Val437Ile) is also listed in *Supplemental Table 2C* below.

Supplemental Table 2C Characteristics of pathogenic missense substitutions and inframe indels

GENE	EXON	DNA	Protein	Critical tertiary structure	Class	Functional	Images current ms	Class HHT Mut. DB [3]	General pop. Allele Frequency	Grantham distance	Align GVGD	SIFT
ENG	2	c.155G>A	p.Gly52Asp	In OR2 β-strand 3. Introduces a negative charge into the protein's hydrophobic core; causes clashes with Met190, Val49, Gln188.	4	Saito et al[21] Pece et al[17]		Pathogenic	-	94	C65	0
ENG	4	c.511_512 insTCC	p.170dup	In the core of OR2. The duplication could either disrupt the register of the N-terminal part of OR2 β-strand 12 or, more likely, cause one additional residue to be accommodated in the loop that follows, taking the place of Arg171, disrupting a hydrogen-bond with Glu195, a stacking interaction with Arg192 and replacing a hydrophobic residue with a charged one at the bottom of the hydrophobic core of the OR2 β-sandwich	4	Saito et al[21]	Fig 2Aiii	-	-	-	-	-
ENG	5	c.620G>A	p.Cys207Tyr	Destroys conserved OR1 disulfide bond with Cys30.	4	Saito et al[21]	Fig 2Ai	Pending classification	-	194	C65	0
ENG	5	c.646A>G	p.Lys216Glu	Just before β-strand 16 of OR1. Changes local charge, so may disrupt interaction with Gln270 at the interface with BMP9 as well as cause electrostatic repulsion of residues such as Asp248.	4	Saito et al[21]	Fig 2Aii	Pending classification	-	56	CO	0.21
ENG	6	c.763G>T	p.Gly255Cys	At beginning of the OR1 loop following β-strand 18. Substitution would clash with nearby residues, and likely interfere with correct ENG disulfide bond formation.	4			-	-	159	C0	0
ENG	7	c.896T>G	p.Leu299Arg	In α2 of OR1. Would change the local charge and cause clashes with nearby residues, destabilizing the packing of the helix against the OR1 β-sheet that includes strands 16 and 18.	4		Fig 2Ai	-	-	102	C0	0 ..../

GENE	EXON	DNA	Protein	Critical tertiary structure	Class	Functional	Images current	HHT mutation database2018	Allele Frequency	Grantham distance	Align GVGD	SIFT
ENG	7	c.899T>C	p.Leu300Pro	In α2 of OR1. Would interfere with the helical fold and diminish packing of the Leu against conserved disulfide C30-C207. May also indirectly affect ligand binding as the C-terminal end of the helix is near the interface with BMP9.	4		Fig 2Ai	Pathogenic	-	98	C0	0.01
ENG	7	c.923C>A	p.Ala308Asp	In the loop between OR1 α2 and β-strand 23. Would disrupt the hydrophobic packing of α2 against β-23 as well as β-16 and β-18.	4	Ali et al [16]		Suspected Pathogenic	-	126	C0	0.05
ENG	12	c.1484T>C	p.Leu495Pro	In the hydrophobic core of the ZP-C domain. Likely distorts β-strand C of ENG ZP-C.	4			-	-	98	C65	0.01
ACVRL1	3	c.200G>A	p.Arg67Gln	-	4			Pathogenic	-	43	C35	0
ACVRL1	3	c.269G>A	p.Cys90Tyr	-	4			Pathogenic	-	194	C65	0
ACVRL1	6	c.632G>A	p.Gly211Asp	Kinase domain	4	Alaa et al[19]		Pathogenic	-	94	C65	0
ACVRL1	7	c.830C>A	p.Thr277Lys	Kinase domain	4			Pending classification	-	78	C65	0
ACVRL1	7	c.830C>G	p.Thr277Arg	Kinase domain	4			-	-	71	C65	0
ACVRL1	7	c.848G>T	p.Gly283Val	Kinase domain	4			-	-	109	C65	0
ACVRL1	7	c.905T>C	p.Leu302Pro	Kinase domain	4			-	-	98	C45	0
ACVRL1	8	c.1120C>T	p.Arg374Trp	Kinase domain	4	Ricard et al[20]		Pathogenic	-	101	C65	0
ACVRL1	8	c.1153A>C	p.Ile385Leu	Kinase domain	4			-	-	5	C0	0
ACVRL1	8	c.1231C>T	p.Arg411Trp	Kinase domain	4	Alaa et al. [19]		Pathogenic	-	101	C65	0
ACVRL1	8	c.1232G>A	p.Arg411Gln	Kinase domain	4	Ricard et al[20]		Pathogenic	ExAc AFR:0.0065%	43	C35	0
ACVRL1	9	c.1298C>A	p.Pro433His	Kinase domain	4			Pending classification	-	77	C65	0
ACVRL1	10	c.1427C>G	p.Pro476Arg	Kinase domain	4			-	-	103	C65	0
ACVRL1	10	c.1459A>C	p.Lys487Gln	Kinase domain	4			-	-	53	C45	0
SMAD4	10	c.1058A>G	p.Tyr353Cys	Cysteine introduced to dwarfin domain	4			-	-	194	C15	0.11
SMAD4	10	c.1081C>T	p.Arg361Cys	Cysteine to dwarfin domain	4			Pathogenic		180	C65	0
SMAD4	12	c.1309G>A	p.Val437Ile	Dwarfin domain	4			-	-	29	C25	0

Supplemental Table 3: Patients with multiple variants in *ENG, ACVRL1, SMAD4* or *GDF2*

	Pathogenic or likely pathogenic variant				Second variant			
DNA	**Gene variant**	**Exon**	**Class**	**Class in HHT Mutation Database.[13]**	**Second non pathogenic variant**	**Highest general pop. prevalence**	**Tertiary structure**	**Class in HHT Mutation Database.[13]**
#1	*ENG* c.667_668insG (p.Val223fs)	*5*	5	-	*ACVRL1* c.88C>T (p.Pro30Ser)	ExAC AMR:0.12%	-	-
#2	*ENG* c.1166_1168delTCT (p.Phe389del)	9	5	-	*ACVRL1* c.948G>C (p.Glu316Asp)	ESP EurAm:0.02%	Kinase domain	-
#3	*ENG* c.370del (p.Leu124Trpfs*39)	4	5	-	*ACVRL1* c.1153A>C (p.Ile385Leu)	ExAC All 0.00041%	Kinase domain	-
					& ENG c.388C>T (p.Pro130Ser)	ExAC SAS:0.27%	-	-
#4	*ENG* c.511C>T (p.Arg171*)	5	5	*Pathogenic*	*ENG* c.7C>T (p.Arg3Cys)	ExAC AMR:0.095%	-	-
#5	*ACVRL1* c.1171G>T (p.Glu391*)	8	5	*Pathogenic*	*ENG* c.572G>A (p.Gly191Asp)	ExAC NFE:2.49%	-	*Benign*
#6	*SMAD4* c.1058A>G (p.Tyr353Cys)	10	4	*Pend class.*	*ENG* c.572G>A (p.Gly191Asp)			
#7	*ACVRL1* c.1120C>T (p.Arg374Trp)	8	5	*Pathogenic*	*ENG* c.572G>A (p.Gly191Asp)			
#8	*ENG* c.360+4A>G (splice)	Intron 3	4	*Pathogenic*	*ENG* c.572G>A (p.Gly191Asp)			
#9	*ACVRL1* c.905T>C (p.Leu302Pro)	7	4	-	*ENG* c.572G>A (p.Gly191Asp)			
#10	*ENG* c.1456_1457ins20nt;p.Phe486fs*12	12	5	-	*ENG* c.1019C>T (p.Pro340Leu)	ExAC SAS:0.0065%	-	
#11	*ENG* c.277C>T (p.Arg93*)	3	5	*Pathogenic*	*ENG* c.1238G>A (p.Gly413Asp) [a]	ExAC NFE:0.0015%	-	-
#12	*ENG* c.991G>A (p.Gly331Ser; splice.)	7	4	-	*ENG* c.1337A>G (p.Asp446Gly) [a]	ExAC NFE:0.0015%	-	-
#13	*ENG* del exon 3-8	3-8	5	-	*ENG* c.1712G>A (p.Arg571His) [a]	ExAC EAS:0.15%	-	*Susp. benign*
#14	*ENG* c.690-2A>G (splice)	Intron 5	5	*Pathogenic*	*ENG* c.1844C>T (p.Ser615Leu)	ExAC FIN:0.16%	-	*Benign*
#15	*ACVRL1* c.632G>A (p.Gly225Asp)	6	5	-	*ENG* c.1844C>T (p.Ser615Leu)			
#16	*ENG* c.496delC (p.Gln166Argfs*56)	4	5	*Pathogenic*	*GDF2* c.326T>C (p.Val109Ala)	ExAC EAS:0.13%	-	-
#17	*ENG* c.776T>G (p.Val259Gly)	6	3	-	*GDF* c.348T>C (p.Asp116Asp)	-	? Splice[a]	-
#18	*ENG* c.511_512insTCC (p.Leu170dup)	4	4	-	*GDF2* c.776A>G (p.Asn259Ser)	ExAC NFE:0.054%	-	-
#19	*ENG* c.360+5G>A (splice)	Intron 3	4	*Pathogenic*	*GDF2* c.997C>T (p.Arg333Trp)	ExAC AFR:0.38%	-	[b]
#20	*ACVRL1* c.830C>G (p.Thr277Arg)	7	4	-	*SMAD4* c.1573A>G (p. Ile525Val)	ExAC All 0.056%	Dwarfin domain	-

SUPPLEMENTAL TABLE 3: Note DNA numbers newly allocated for manuscript. pop., population, ExAC- annotations provided in Lek et al.[23] Susp., suspected. [a] See main manuscript *FIGURE 4Aiv* [b] Reported as pathogenic in the HHT Mutation database.[13]

Supplemental Table 4: *GDF2* variant considerations

Supplemental Table 4A: Details of the 4 *GDF2* variants and concurrent *ENG* variants

	GDF2 **variant**					**Second variant**			
Sample#	***GDF2*** **Nucleotide change**	**BMP9 Protein change**	**dbSNP**	**ExAc highest**	**Clin Var**	**Gene**	**Nucleotide change**	**Protein change**	**Class**
#19	c.997C>T	Arg333Trp	rs35129734	AFR:0.38%	RCV000074346.2[a]	*ENG*	c.360+5G>A	Splice [b]	Pathogenic
#16	c.326T>C	Val109Ala	rs782118500	EAS:0.13%	*NI*	*ENG*	c.496delC	Gln166Argfs*56	Likely Pathogenic
#18	c.776A>G	Asn259Ser	rs140271276	OTH:0.11%	*NI*	*ENG*	c.511_512insTCC	Leu170dup [c]	Likely pathogenic
#17	c.348T>C	None: Asp116Asp [b]	-	-	-	*ENG*	c.776T>G	Val259Gly	VUS

Note DNA numbers newly allocated for manuscript.
[a] Listed as Pathogenic HHT type 5 on ClinVar[15,24]
[b] For splicing predictions see also *Supplemental Tables 2A and 4B*
[c] For crystal structure see *Figure 4Aiii*, *Supplemental Table 2C*. *NI, not in*

Supplemental Table 4B: *ENG* c.360+5G>A splicing predictions

Alamut Visual Report:	**SSF [0-100]**	**MaxENT[0-12]**	**NNSplice[0-1]**	**GeneSplicer[0-15]**	**HSF[0-100]**
Wildtype (G)	**78.2**	7.77	0.98	4.74	85.34
Variant (A)	0	0	0	0	0
Change	>99.99%	>99.99%	>99.99%	>99.99%	>99.99%

Supplemental Table 4C: Phenotypic data

There was no suggestion of a possible phenotypic difference between cases with a *GDF2* variant in addition to an *ENG* pathogenic/ likely pathogenic variant *versus* an *ENG* pathogenic/ likely pathogenic in isolation (see below). There was also no discernible difference in phenotypes for DNAs with a single or two HHT variants.

	Number of Probands	Bleeding score			Pulmonary AVM (%)		Cerebral AVM (%)		Spinal AVM (%)	
		Median	Mean	SD	Mean (%)	SD	Mean	SD	Mean	SD
Single *ENG* POLP	89	3	3.2	1.3	84.8	36.1	11.2	31.8	0	0
ENG POLP + *GDF2* variant	4	3	3.3	0.7	66.7	58	0	0	0	0

Supplemental Table 5: Clinician positive and negative predictive values

There was a trend for ENG stop codon variants to be called more accurately than other variants but no difference in accuracy when limited to predictions with the highest (>85%) degree of confidence

A) *ACVRL1*, *ENG* and *SMAD4* pathogenic variants

	N	PPV	NPV	Sensitivity	Specificity
a) Any prediction					
SMAD4	4	80.0	99.2	80.0	96.9
ENG	85	85.0	75.4	85.0	73.0
ACVRL1	42	68.9	82.4	68.9	89.9
b) If sole prediction					
SMAD4	4	80.0	99.1	80.0	98.2
ENG[a]	75	86.2	76.9	86.2	74.1
ACVRL1	36	69.2	83.2	69.2	89.8
c) If ≥90% confident					
SMAD4	4	80.0	99.2	80.0	97.6
ENG	81	85.3	75.4	85.3	74.1
ACVRL1	39	69.6	83.3	69.6	89.5

a: See *Supplemental Table 5B* for *ENG* breakdown for molecular subtypes

B) *ENG* subtype-specific pathogenic variants

ENG	**N**	**PPV**	**NPV**	**Sensitivity**	**Specificity**
Frameshift	27	84.4	87.5	84.4	61.4
Stop gained	16	88.9	80	88.9	88.9
Missense	11	84.6	91.3	84.6	67.7

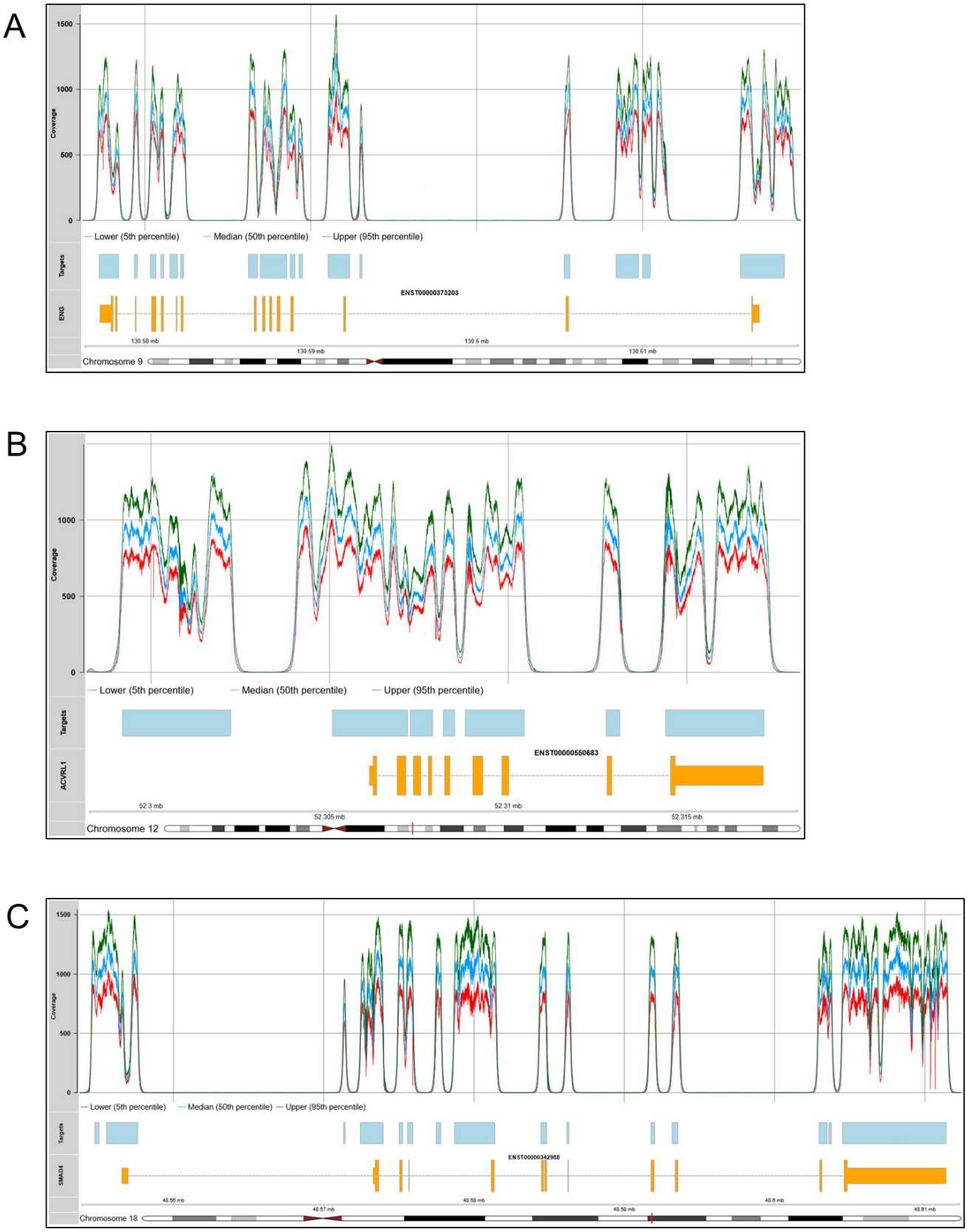
Supplemental Table 6: Application of the beta distribution to ThromboGenomics HHT cohort

HPO term name	Gene	Source	Expected Frequency	Observed Frequency	alpha	beta	Certainty (%)
Pulmonary AVM	*ENG*	Predicted	60	78.50	12	5	50-90
Cerebral AVM	*ENG*	Predicted	14.6	10.75	6	40	5-20
Spinal AVM	*ENG*	Predicted	1	0	6	300	0.5-5
Hepatic AVM	*ENG*	Predicted	14	6.45	3	20	3-30
Pulmonary hypertension	*ENG*	Predicted	0.00005	0	2	200	0-2
Colorectal polyposis	*ENG*	Predicted	0.01	6.45	3	150	0.01-10
Juvenile gastrointestinal polyposis	*ENG*	Predicted	0.01	0	1.5	1000	0-0.5
Dilatation of the ascending aorta	*ENG*	Predicted	0.01	0	1.5	100	0-0.5
Pulmonary AVM	***ACVRL1***	**Predicted**	**10**	**50**	**6**	**30**	**5-30**
Cerebral AVM	***ACVRL1***	**Predicted**	**5.3**	**2.38**	**8**	**300**	**1-5**
Spinal AVM	***ACVRL1***	**Predicted**	**0.01**	**4.76**	**3**	**250**	**0.1-3**
Hepatic AVM	***ACVRL1***	**Predicted**	**75**	**9.52**	**5**	**12**	**6-60**
Pulmonary hypertension	***ACVRL1***	**Predicted**	**2.4**	**7.14**	**4**	**80**	**1-10**
Colorectal polyposis	***ACVRL1***	**Predicted**	**0.01**	**2.38**	**3**	**150**	**0.01-10**
Juvenile gastrointestinal polyposis	***ACVRL1***	**Predicted**	**0.01**	**0**	**1.5**	**1000**	**0-0.5**
Dilatation of the ascending aorta	***ACVRL1***	**Predicted**	**0.01**	**2.38**	**1.5**	**100**	**0-0.5**
Pulmonary AVM	*SMAD4*	Predicted	60	75	12	5	50-90
Cerebral AVM	*SMAD4*	Predicted	14.6	0	6	40	5-20
Spinal AVM	*SMAD4*	Predicted	1	0	6	300	0.5-5
Hepatic AVM	*SMAD4*	Predicted	14	25	5	20	3-30
Pulmonary hypertension	*SMAD4*	Predicted	0.00005	0	2	200	0-2
Colorectal polyposis	*SMAD4*	Predicted	98	25	6	3	25-90
Juvenile gastrointestinal polyposis	*SMAD4*	Predicted	98	25	7	3	25-90
Dilatation of the ascending aorta	*SMAD4*	Predicted	5.6	0	3	20	1-30

SUPPLEMENTAL TABLE 6: Expected frequencies denote clinician-predicted frequencies prior to modelling, based on earlier studies.[25-36] Observed frequencies refer to the prevalence in the cohort. Alpha and beta terms were selected as described in *SUPPLEMENTAL FIGURE 3*. Certainty (%), is the range in which the phenotype was considered certain to occur in the current population. Note that in the current cohort, due to historical referral bias and universally recommended pulmonary AVM screening, the pulmonary AVM prevalence was higher than observed in other cohorts. Conversely, due to local screening practices, hepatic AVMs were under-ascertained.

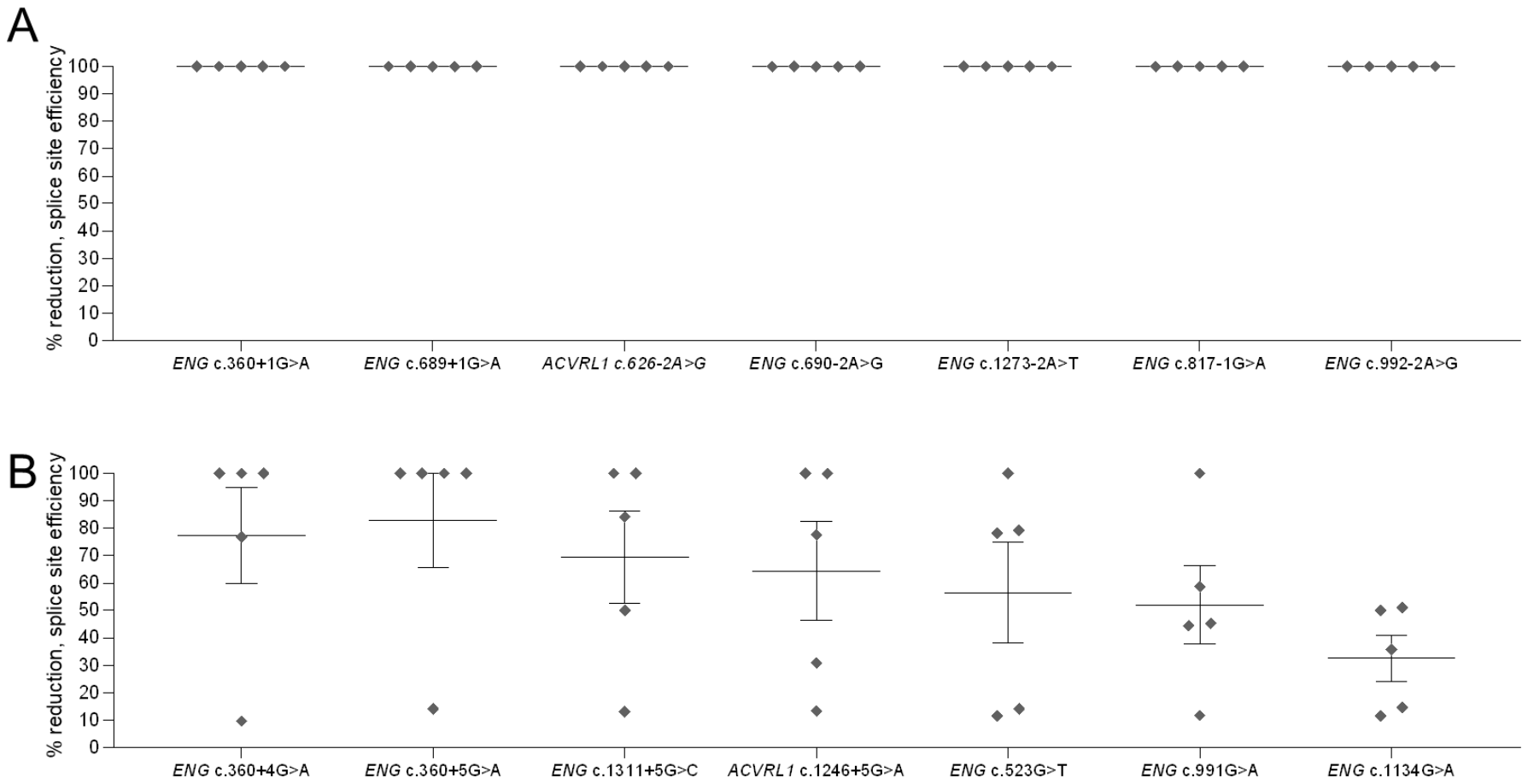
SUPPLEMENTAL FIGURES

Supplemental Figure 1: Coverage profile for the *ENG, ACVRL1* and *SMAD4* genes.



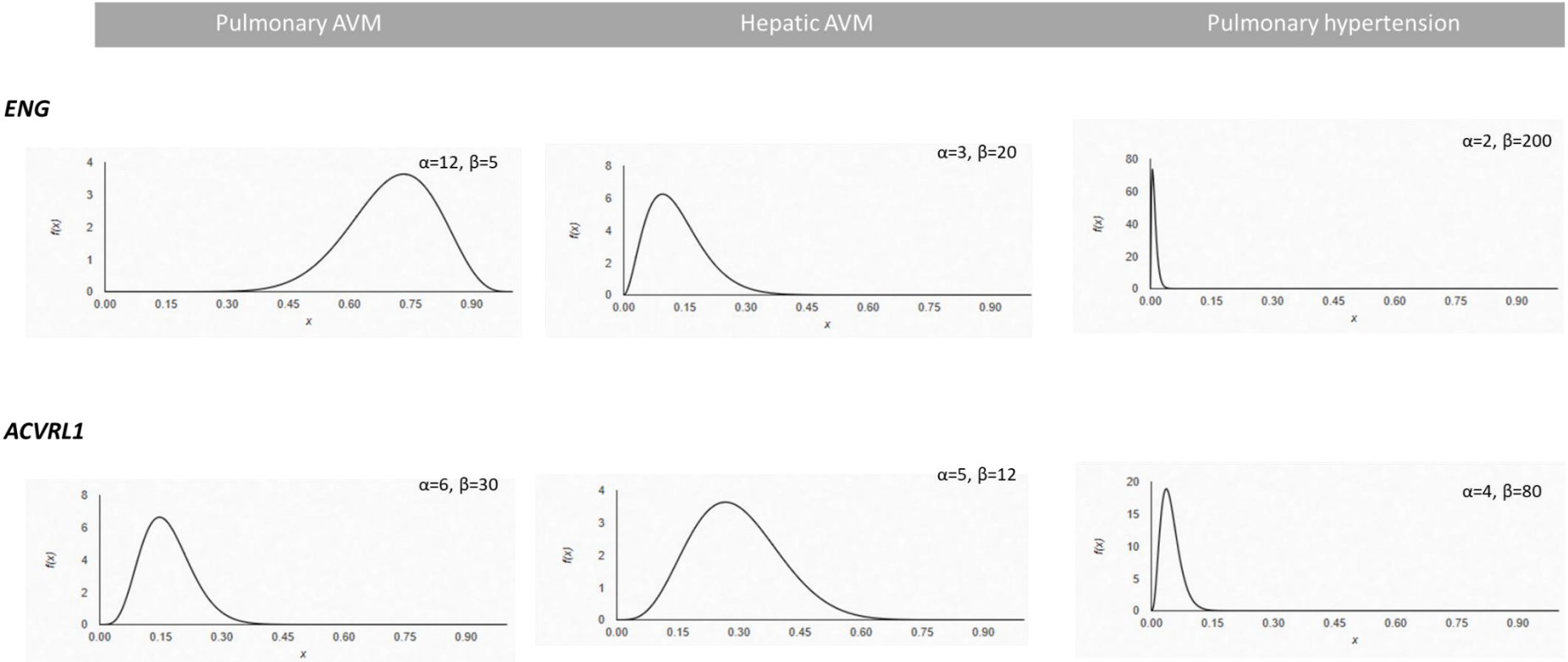
SUPPLEMENTAL FIGURE 1: Plots show the chromosome position, transcript, targeted regions and coverage profile for **A)** *ENG*, **B)** *ACVRL1* and **C)** *SMAD4*, respectively. *GDF2* is illustrated in the main manuscript, *FIGURE 1*.

Supplemental Figure 2: Selected splice site variant illustrations



SUPPLEMENTAL FIGURE 2: All predictions were made using Alamut Visual v2.9 Interactive BioSoftware[37] from which the splice site predicted changes (as a %) are illustrated for SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer, and HumanSplicingFinder. **A)** Consensus splice site variants**. B)** Splice region variants. Note that for *ENG* c.360+4A>G, a previous publication demonstrated that in the mRNA of a different HHT patient with the same variant the exon 3 is skipped,[18] and that *ENG* c.991G>A and *ENG* c.1134G>A variants are defined as Pathogenic[13] on the HHT Mutation Database (see also *SUPPLEMENTAL TABLE 2B*).

Supplemental Figure 3: Example prior distributions for HHT phenotypes in the cohort



SUPPLEMENTAL FIGURE 3: Examples of model distribution selections for three phenotypes known to differ in prevalence between the two listed genotypes. The α and β shape parameters are listed in *SUPPLEMENTAL TABLE 6* which were based on earlier studies as previously described.[25-36]

SUPPLEMENTAL REFERENCES

1. Shovlin CL, Buscarini E, Hughes JMB, Allison DJ, Jackson JE. Long-term outcomes of patients with pulmonary arteriovenous malformations considered for lung transplantation, compared with similarly hypoxaemic cohorts. *BMJ Open Respir Res*.2017 Oct 13;4(1):e000198.

2. Cole SG, Begbie ME, Wallace GM, Shovlin CL. A new locus for hereditary haemorrhagic telangiectasia (HHT3) maps to chromosome 5. *J Med Genet.* 2005 Jul;42(7):577-82.

3. Govani FS, Shovlin CL. Fine mapping of the hereditary haemorrhagic telangiectasia (HHT)3 locus on chromosome 5 excludes VE-Cadherin-2, Sprouty4 and other interval genes. *J Angiogenes Res.* 2010 Aug 11;2:15.

4. Shovlin CL, Prigoda N, Govani FS. A family with a novel chromosomal rearrangement: a resource to identify a further gene for pulmonary arteriovenous malformations and hereditary haemorrhagic telangiectasia. *Thorax* 2007 (62) A115

5. Livesey JA, Manning RA, Meek JH, *et al.* Low serum iron levels are associated with elevated plasma levels of coagulation factor VIII and pulmonary emboli/deep venous thromboses in replicate cohorts of patients with hereditary haemorrhagic telangiectasia. *Thorax.* 2012 Apr;67(4):328-33.

6. Shovlin CL, Chamali B, Santhirapala V, *et al.* Ischaemic strokes in patients with pulmonary arteriovenous malformations and hereditary hemorrhagic telangiectasia: associations with iron deficiency and platelets. *PLoS One.* 2014 Feb 19;9(2):e88812.

7. Boother EJ, Brownlow S, Tighe HC, Bamford KB, Jackson JE, Shovlin CL. Cerebral abscess associated with odontogenic bacteremias, hypoxemia, and iron loading in immunocompetent patients with right-to-left shunting through pulmonary arteriovenous malformations. *Clin Infect Dis.* 2017 Apr 19. doi: 10.1093/cid/cix373

8. Shovlin CL, Guttmacher AE, Buscarini E, *et al.* Diagnostic criteria for hereditary hemorrhagic telangiectasia (Rendu-Osler-Weber syndrome). *Am J Med Genet.* 2000 6;91(1):66-7

9. Köhler S, Doelken SC, Mungall CJ, *et al*. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D966-74

10. Shovlin CL, Sodhi V, McCarthy A, Lasjaunias P, Jackson JE, Sheppard MN. Estimates of maternal risks of pregnancy for women with hereditary haemorrhagic telangiectasia (Osler-Weber-Rendu syndrome): suggested approach for obstetric services. *BJOG.* 2008 Aug;115(9):1108-15.

11. Plon SE, Eccles DM, Easton D, *et al.* Unclassified Genetic Variants Working Group. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat*. 2008 Nov;29(11):1282-91.

12. Richards S, Aziz N, Bale S, *et al*. 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 May;17(5):405-24.

13. University of ARUP Department of Pathology and ARUP Laboratories. HHT Mutation Database (http://www.arup.utah.edu/database/HHT/) and SMAD4 (http://www.arup.utah.edu/database/SMAD4/SMAD4_display.php), accessed 20th September 2019)

14. 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 Apr;4(4):184-96.

15. National Center for Biotechnology Information. ClinVar available at https://www.ncbi.nlm.nih.gov/clinvar/

16. 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

17 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 Nov;8(12):2171-81

18. 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 Jul;61(1):68-79.

19. Alaa El Din F, Patri S, Thoreau V, *et al.* Functional and splicing defect analysis of 23 ACVRL1 mutations in a cohort of patients affected by Hereditary Hemorrhagic Telangiectasia. *PLoS One.* 2015 Jul 15;10(7):e0132111.

20. Ricard N, Bidart M, Mallet C, *et al.* Functional analysis of the BMP9 response of ALK1 mutants from HHT2 patients: a diagnostic tool for novel ACVRL1 mutations. *Blood*. 2010 Sep 2;116(9):1604-12.

21. Saito T, Bokhove M, Croci R, *et al.* Structural Basis of the Human Endoglin-BMP9 Interaction: Insights into BMP Signaling and HHT1. *Cell Rep.* 2017 May 30;19(9):1917-1928.

22. Fortier N. Revisiting the Five Splice Site Algorithms used in Clinical Genetics. http://blog.goldenhelix.com/nfortier/splice-site-algorithms/ (accessed 20th September 2019)

23. Lek M, Karczewski KJ, Minikel EV, *et al.* Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016 Aug 18;536(7616):285-91.

24. Wooderchak-Donahue WL, McDonald J, O'Fallon B, *et al.* BMP9 mutations cause a vascular-anomaly syndrome with phenotypic overlap with hereditary hemorrhagic telangiectasia. *Am J Hum Genet.* 2013 Sep 5;93(3):530-7.

25. Kjeldsen AD, Møller TR, Brusgaard K, *et al.* Clinical symptoms according to genotype amongst patients with hereditary haemorrhagic telangiectasia. *J Int Med*. 2005;258:349-355

26. Letteboer TG, Mager JJ, Snijder RJ, *et al.* Genotype-phenotype relationship in hereditary haemorrhagic telangiectasia. J *Med Genet.* 2006 Apr;43(4):371-7

27. Bayrak-Toydemir P, McDonald J, Markewitz B, *et al.* Genotype-phenotype correlation in hereditary hemorrhagic telangiectasia. *Am J Med Genet A.* 2006;140:463-470

28. Bossler AD, Richards J, George C, Godmilow L, Ganguly A. Novel mutations in ENG and ACVRL1 identified in a series of 200 individuals undergoing clinical genetic testing for hereditary hemorrhagic telangiectasia (HHT): correlation of genotype with phenotype. *Hum Mutat.* 2006;27:667-675

29. Sabbà C, Pasculli G, Lenato GM, *et al.* Hereditary hemorrhagic telangiectasia: clinical features in ENG and ALK1 mutation carriers. *J Thromb Haemost.* 2007;5:1149-1157.

30. Letteboer TG, Mager HJ, Snijder RJ, *et al.* Genotype-phenotype relationship for localization and age distribution of telangiectases in hereditary hemorrhagic telangiectasia. *Am J Med Genet A*. 2008 Nov 1;146A(21):2733-9.

31. Revuz S, Decullier E, Ginon I, *et al.* Pulmonary hypertension subtypes associated with hereditary haemorrhagic telangiectasia: Haemodynamic profiles and survival probability. *PLoS One.* 2017 Oct 5;12(10):e0184227.

32. 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.

33. Gallione C, Repetto GM, Legius E, *et al.* A combined syndrome of juvenile polyposis and hereditary haemorrhagic telangiectasia is associated with mutations in MADH4 (SMAD4). *Lancet.* 2004;363:852-859.

34. Heald B, Rigelsky C, Moran R, *et al.* Prevalence of thoracic aortopathy in patients with juvenile Polyposis Syndrome-Hereditary Hemorrhagic Telangiectasia due to SMAD4. *Am J Med Genet A.* 2015 Aug;167A(8):1758-62.

35. Vorselaars VMM, Diederik A, Prabhudesai V, *et al.* SMAD4 gene mutation increases the risk of aortic dilation in patients with hereditary haemorrhagic telangiectasia. *Int J Cardiol.* 2017

36. Jelsig AM, Tørring PM, Kjeldsen AD, *et al.* JP-HHT phenotype in Danish patients with SMAD4 mutations. *Clin Genet.* 2016 Jul;90(1):55-62. doi: 10.1111/cge.

37. Alamut Visual Alamut® Visual 2.7 (2018 Interactive Biosoftware)