blood

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

Mutational and phenotypic characterisation of hereditary hemorrhagic telangiectasia

Tracking no: BLD-2019-004560R1

Claire Shovlin (Imperial College London, United Kingdom) Ilenia Simeoni (University of Cambridge, United Kingdom) Kate Downes (University of Cambridge, United Kingdom) Zoe Frazer (Imperial College London, United Kingdom) Karyn Megy (University of Cambridge, United Kingdom) María Bernabéu-Herrero (Imperial College London, United Kingdom) Abigail Shurr (Imperial College London, United Kingdom) Jennifer Brimley (Imperial College Healthcare NHS Trust, United Kingdom) Dilipkumar patel (imperial college, United Kingdom) Loren Kell (University College London, United Kingdom) Jonathan Stephens (University of Cambridge, United Kingdom) Isobel Turbin (Imperial College London, United Kingdom) Micheala Aldred (Indiana University School of Medicine, United States) Christopher Penkett (University of Cambridge, United Kingdom) Willem Ouwehand (University of Cambridge, United Kingdom) Luca Jovine (Karolinska Institutet, Sweden) Ernest Turro (University of Cambridge, United Kingdom)

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, 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.

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: 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, generated Supplementary Data, developed the text and Figures with ET and LJ, and revised and approved the final manuscript. IS ThromboGenomics platform design, managed ThromboGenomics, processed samples, reviewed and approved the final manuscripts KD Managed ThromboGenomics, reviewed and approved the final manuscripts KM ThromboGenomics platform design, gene and transcript curation, variant submission to public repository, and reviewed and approved the final manuscript ZF Extracted DNAs, contributed to data discussions, and reviewed and approved the final manuscript MBH Participated in MDT, contributed to data discussions, and approved the final manuscript JB Recruited study participants, contributed to data discussions, and reviewed and approved the final manuscript DP Extracted DNAs, contributed to data discussions, and reviewed and approved the final manuscript LK Assisted with study phenotyping, and reviewed and approved the final manuscript JS Processed Thrombogenomics samples, performed deletion validations and approved the final manuscript. IT Contributed to data discussions, and reviewed and approved the final manuscript MA Participated in MDT, contributed to data discussions, and reviewed and approved the final manuscript CP Contributed to data discussions, and reviewed and approved the final manuscript WHO Contributed to data discussions, and reviewed and approved the final manuscript LJ Performed structural analysis, generated Figure 3A and associated text, contributed to data discussions, and reviewed and approved the final manuscript. ET Preformed data analyses, developed the statistical models, contributed to data discussions, generated Figures, co-authored the text, and approved the final manuscript.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: The variants are being submitted to ClinVar and will also be available on the HHT Mutation Database post publication

Clinical trial registration information (if any):

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 3944 words

Prior Abstract/Poster presentation: NONE

Prior Presubmission website listing: NONE

KEY POINTS

- We integrate systematic variant identification by deep high-throughput DNA sequencing with ACMG Guidelines for pathogenicity assignments.
- 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, 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 carry any 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, generated Supplementary Data, developed the text and Figures with ET and LJ, and revised and approved the final manuscript.	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 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 LEGENDS

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, 5^{th} and 95^{th} 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.

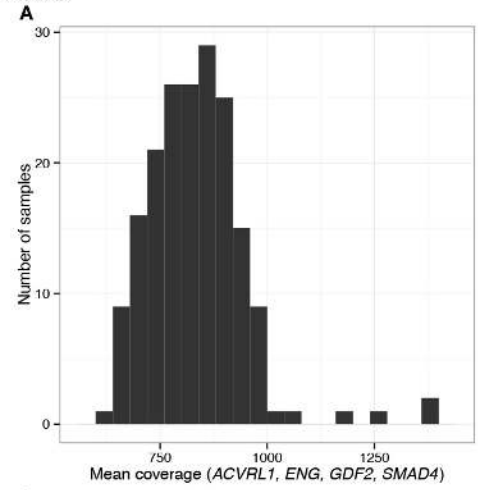
A) The predicted (triangles) and observed (circles) frequencies of eight *a priori* discriminatory HPO terms in Cohorts 1 and 2. **B)** Predicted causal gene displayed for each of the observed HHT genotypes in Cohort 1. **Upper panel**: Clinician prediction using all HPO terms– see also *Supplemental 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*).

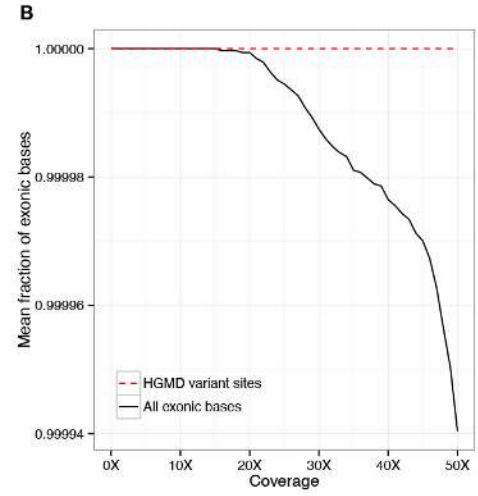
Figure 5: Quantitative red cell traits

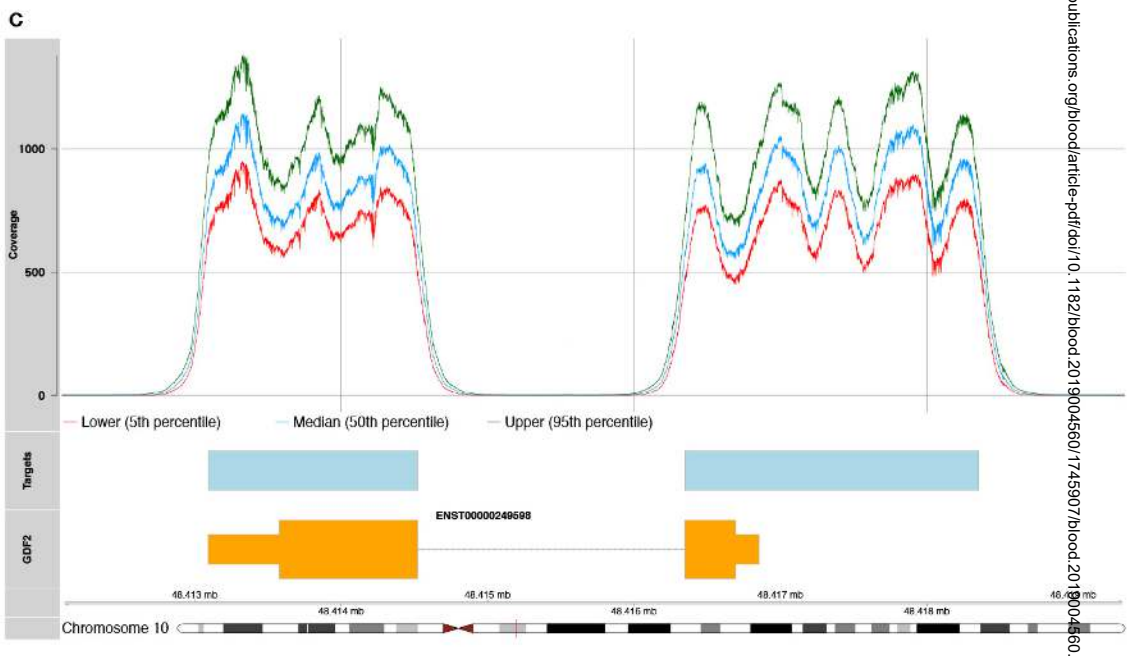
A) Distributions of quantitative red cell traits in HHT and control populations: Total red blood cell count (left), hematocrit (centre) and haemoglobin (right) plotted for the HHT patients (one measurement per patient, proband and affected family members from Cohorts 1 and 2) above the respective INTERVAL 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 5^{th}-95^{th} 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),

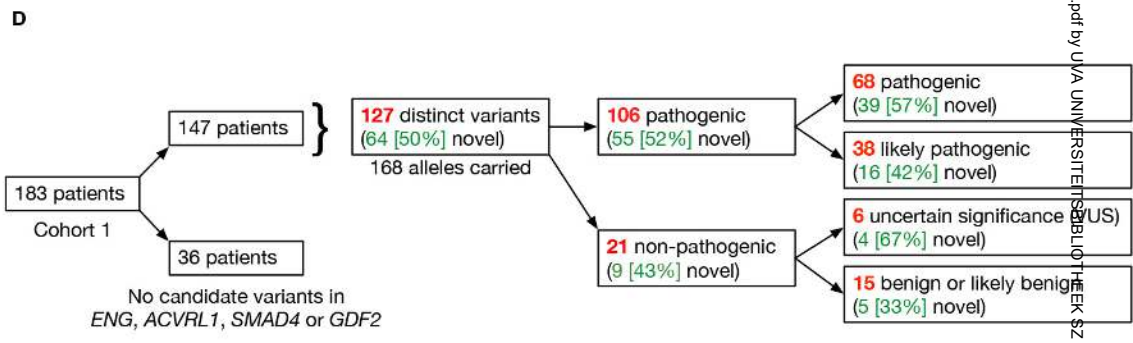
hematocrit (centre) and hemoglobin (right) against same-day oxygen saturation measured by finger oximetry for 10 minutes standing using one measurement per patient (proband and affected family members). Note that in all six analyses, the patients with greater bleeding (red and purple) tended to have 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

FIGURE1









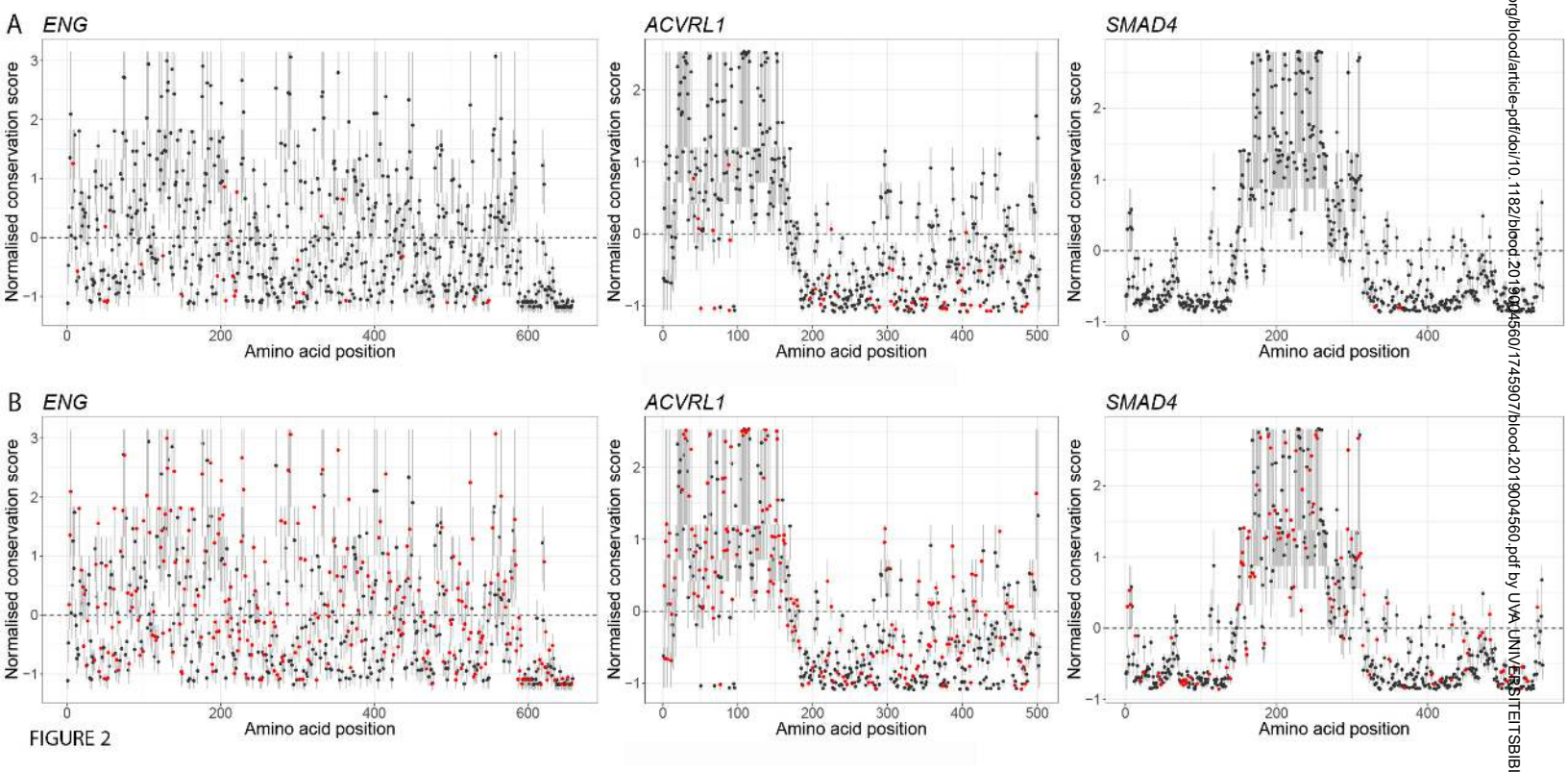
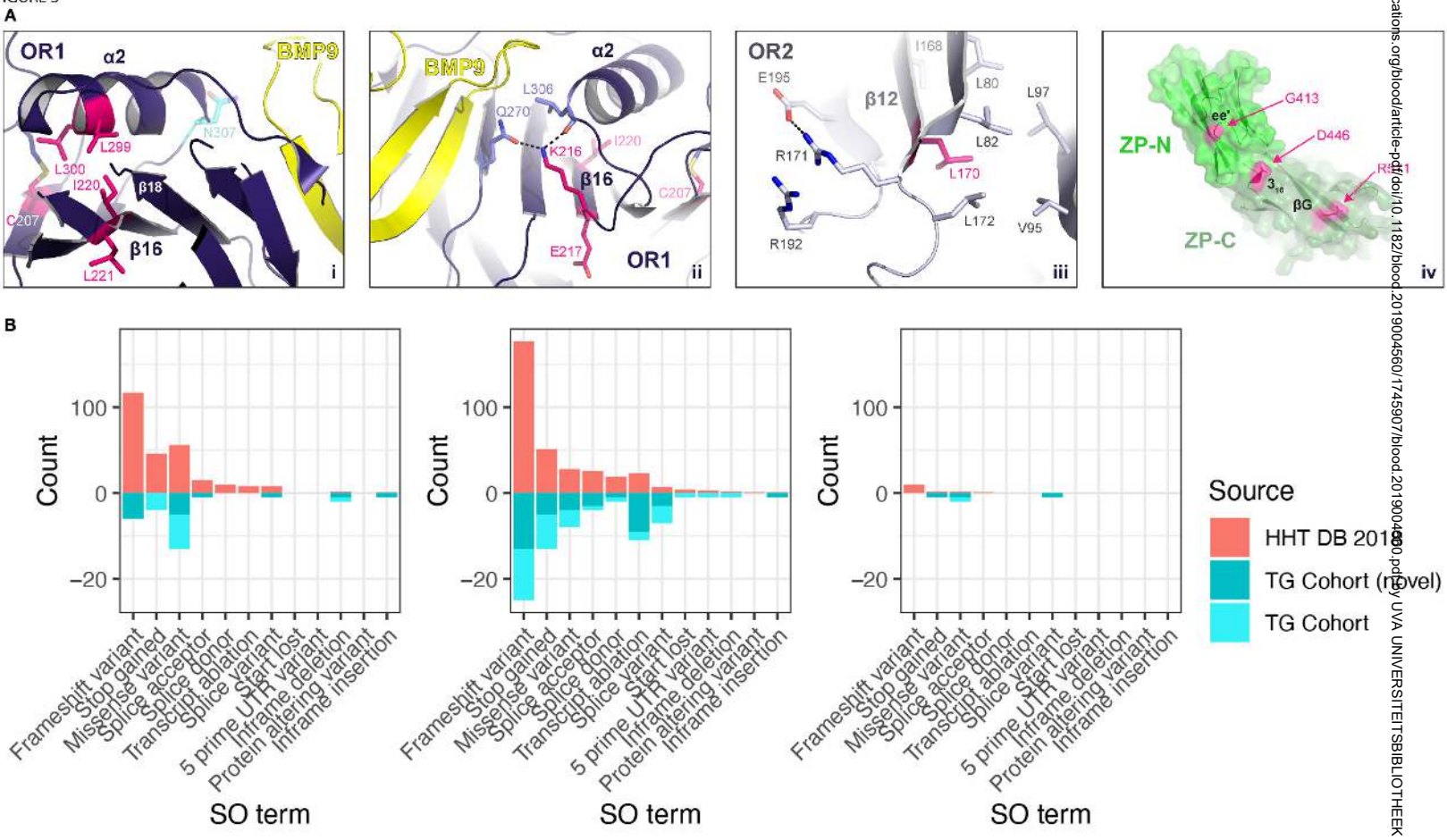


FIGURE 2

FIGURE 3



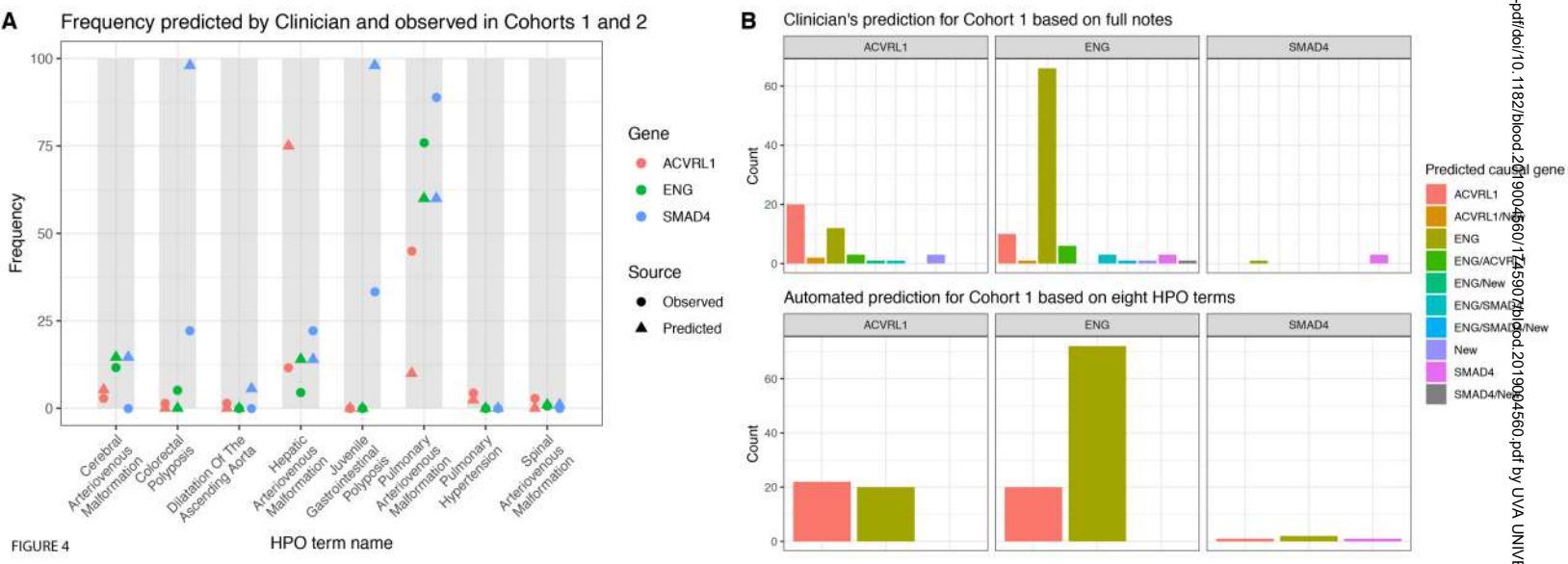


FIGURE 4

FIGURE 5