ORIGINAL ARTICLE

Genome sequencing reveals a deep intronic splicing *ACVRL1* mutation hotspot in Hereditary Haemorrhagic Telangiectasia

Whitney L. Wooderchak-Donahue,[1,2] Jamie McDonald,[2,3] Andrew Farrell,[4] Gulsen Akay,[1,5] Matt Velinder,[4] Peter Johnson,[1] Chad VanSant-Webb,[1] Rebecca Margraf,[1] Eric Briggs,[1] Kevin J Whitehead,[3,6] Jennifer Thomson,[7] Angela E Lin,[8] Reed E Pyeritz,[9] Gabor Marth,[4] Pinar Bayrak-Toydemir[1,2]

► Additional material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2018-105561).

For numbered affiliations see end of article.

Correspondence to
Dr Pinar Bayrak-Toydemir, Molecular Genetics Department, ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT 84108, USA; pinar.bayrak-toydemir@aruplab.com

WL.W-D and JMD contributed equally.

Received 26 June 2018
Revised 14 August 2018
Accepted 23 August 2018
Published Online First 22 September 2018

ABSTRACT

Introduction Hereditary haemorrhagic telangiectasia (HHT) is a genetically heterogeneous disorder caused by mutations in the genes *ENG*, *ACVRL1*, and *SMAD4.* Yet the genetic cause remains unknown for some families even after exhaustive exome analysis. We hypothesised that non-coding regions of the known HHT genes may harbour variants that disrupt splicing in these cases.

Methods DNA from 35 individuals with clinical findings of HHT and 2 healthy controls from 13 families underwent whole genome sequencing. Additionally, 87 unrelated cases suspected to have HHT were evaluated using a custom designed next-generation sequencing panel to capture the coding and non-coding regions of *ENG*, *ACVRL1* and *SMAD4*. Individuals from both groups had tested negative previously for a mutation in the coding region of known HHT genes. Samples were sequenced on a HiSeq2500 instrument and data were analysed to identify novel and rare variants.

Results Eight cases had a novel non-coding *ACVRL1* variant that disrupted splicing. One family had an *ACVRL1*intron 9:chromosome 3 translocation, the first reported case of a translocation causing HHT. The other seven cases had a variant located within a ~300 bp CT-rich 'hotspot' region of *ACVRL1*intron 9 that disrupted splicing.

Conclusions Despite the difficulty of interpreting deep intronic variants, our study highlights the importance of non-coding regions in the disease mechanism of HHT, particularly the CT-rich hotspot region of *ACVRL1*intron 9. The addition of this region to HHT molecular diagnostic testing algorithms will improve clinical sensitivity.

To cite:
Wooderchak-Donahue WL, McDonald J, Farrell A, *et al*. *J Med Genet* 2018;**55**:824–830.

INTRODUCTION

Hereditary haemorrhagic telangiectasia (HHT) is an autosomal dominant vascular dysplasia that occurs in 1 in 5000 individuals.[1] HHT is characterised by (1) recurrent epistaxis; (2) telangiectases of the nasal mucosa, fingertips, lips and oral cavity, (3) arteriovenous malformations (AVMs) in the lungs, liver, gastrointestinal tract and brain and (4) a family history of HHT. Patients who present with three or more of these criteria are considered clinically diagnosed with HHT according to 'Curaçao criteria'.[2]

HHT is a genetically heterogeneous disorder caused by mutations in one of multiple genes in the transforming growth factor-beta signalling pathway. Endoglin (*ENG*), activin A receptor type II-like 1 (*ACVRL1/ALK1*) and *SMAD4* mutations cause HHT1 (OMIM 187300), HHT2 (OMIM 600376) and the combined juvenile polyposis/HHT syndrome (OMIM 175050), respectively.[3–5] Mutations in these genes lead to an underproduction of their respective proteins and results in excessive abnormal angiogenesis.[6]

ENG and *ACVRL1* mutations account for roughly equal percentages of the disorder,[7] and are found in ~85% of cases submitted for molecular genetic testing for HHT.[8 9] Mutations in *SMAD4* cause only 1%–2% of cases.[10] Numerous private mutations have been described across all coding regions of *ENG* and *ACVRL1*,[11] including large exonic deletions or duplications that occur in ~6%–10% of cases.[8 9 12] More recently, non-coding mutations in the 5' untranslated region (UTR) of *ENG* were shown to cause HHT in ~1%–2% of cases,[13 14] indicating the need to include this non-coding region in routine HHT molecular testing algorithms.[15]

Molecular diagnostic testing is important in the medical management of individuals with HHT. If a pathogenic *ENG*, *ACVRL1* or *SMAD4* mutation is identified in an affected proband, diagnostic testing is then available for at-risk family members. HHT is difficult to diagnose on physical examination and medical history alone in the first two decades of life, so identification of a familial mutation allows for early screening and detection of internal AVMs that can be treated prior to the development of serious complications or death.

Efforts to identify new HHT causing genes and regions have proven difficult. Linkage analysis identified two additional HHT loci at chromosome 5q31[16] and chromosome 7p14[17] over a decade ago, but the genes remain unknown. In 2013, our group reported mutations in the bone morphogenetic 9 (*BMP9)/GDF2* gene in three unrelated individuals suspected to have HHT who had previously tested negative for *ENG*, *ACVRL1* and *SMAD4*.[18] Although this has been designated as HHT, type 5 (OMIM 615506), these cases represented <1% of samples submitted to our laboratory for suspected

HHT, and the clinical findings in these individuals differed from those found in 'classical HHT' as defined by the Curaçao criteria. *RASA1* mutations which cause capillary malformation-AVM (CM-AVM) syndrome (OMIM 608354) have been reported in several individuals clinically suspected to have HHT who had epistaxis and dermal lesions described as telangiectases.[19 20] A second form of CM-AVM (CM-AVM2) caused by germline loss-of-function mutations in *EPHB4* was recently reported.[21] CM-AVM2 overlaps with *RASA1*-related CM-AVM as well as HHT in that clinical features include multifocal capillary malformations, AVMs, telangiectases and epistaxis (reported in several cases).[21 22] Because of the phenotypic overlap of these disorders with HHT, next-generation sequencing (NGS) molecular diagnostic panels for HHT and other vascular malformation syndromes (see online supplementary table 1) have become more widely used. Despite these advancements, the molecular diagnosis is still unknown for some HHT cases.

We hypothesised that deep intronic non-coding variants in *ENG* and *ACVRL1* may disrupt splicing and cause HHT in individuals without an identifiable *ACVRL1, ENG* or *SMAD4* mutation by current molecular diagnostic testing protocols. Here, we used genome sequencing to evaluate 13 multigeneration families with HHT and a custom NGS panel designed to interrogate the coding and non-coding regions of *ENG*, *ACVRL1* and *SMAD4* in 87 additional unrelated individuals suspected to have HHT with no detectable mutation. Although deep intronic variants can be difficult to interpret and characterise, such variants identified in these cases may alter splicing and reveal a hidden disease mechanism of HHT.

MATERIALS AND METHODS

Subjects

Individuals and families included in the study were clinically confirmed or suspected to have HHT, and had tested negative at ARUP Laboratories for a mutation in a known HHT gene by sequencing of coding regions and intron/exon borders, and duplication/deletion analysis. The majority of individuals from 13 families who underwent genome sequencing were examined at the University of Utah HHT Center of Excellence (Salt Lake City, Utah, USA). The other 10 families were considered suspicious for a 'telangiectasia syndrome', but not necessarily HHT. None of the probands or family members of these ten families met three or more diagnostic criteria for HHT. These families had also previously undergone exome sequencing with negative results. Eighty-seven additional individuals with clinical suspicion for HHT were identified from samples submitted to our laboratory for HHT testing. Patient history forms are required with submission of samples and ask ordering clinicians to specify the type, location and number of vascular malformations, as well as other related clinical findings such as epistaxis. Based on information provided on this form, 25/87 had three diagnostic criteria for HHT, 47/87 had two criteria and 15/87 had one criteria. In some cases, ARUP genetic counsellors contacted the ordering clinician to clarify reported findings or obtain more information.

Genome sequencing and data analysis using RUFUS software

Genomic DNA was extracted from peripheral blood for 37 individuals using a Gentra Puregene Blood Kit (Qiagen, Valencia, California, USA). Whole genomes from 35 individuals confirmed with or suspicious for HHT, and 2 unaffected individuals from 13 families were sequenced on a HiSeq2500 instrument (Illumina, San Diego, California, USA) using 2×151 paired-end sequencing to ~60× average read depth. Variants were identified using RUFUS, a k-mer-based method capable of detecting variants independent of alignment to a reference sequence (https://github.com/jandrewrfarrell/RUFUS). Variants were prioritised through population allele frequency, predicted functional consequence and segregation within and between families.

Custom NGS panel and data analysis

Genomic DNA was extracted from peripheral blood using a Gentra Puregene Blood Kit (Qiagen) for 87 individuals on whom testing in our clinical lab had been ordered which included the HHT genes, with negative results. Custom 120 nucleotide RNA baits were designed to specifically target the coding and non-coding regions (including introns, 5'UTRs and 3'UTRs) for *ACVRL1*, *ENG* and *SMAD4*. Coding regions for *GDF2* and *RASA1* were also included. RNA baits were tiled at 5x spacing and were in replicates of 10 to increase hybridisation efficiency of the targeted region (~0.1 Mb). Genomic DNA (3 µg) was sheared using a Covaris S220 ultrasonicator instrument (Covaris, Woburn, Massachusetts, USA) to 180 bp fragments. Illumina adapters were added using the Bravo automated instrument and SureSelect XT kit reagents (Agilent Technologies, Santa Clara, California, USA). Adapter ligated DNA underwent hybridisation with the biotinylated RNA baits for 24 hours at 65°C. Hybridised DNA targets of interest were captured using streptavidin-coated magnetic beads. DNA targets of interest were eluted and barcode/indexed after a series of washes to remove the non-targeted, unbound genome. DNA quality and quantity were assessed using the TapeStation (Agilent Technologies). Samples were pooled (1:1) and sequenced on a HiSeq2500 instrument (Illumina) using 2×100 paired-end sequencing to ~500× average read depth.

Sequences were aligned to the human genome reference (hg19) sequence using the Burrows-Wheeler Alignment tool (0.5.9) with default parameters.[23] PCR duplicates were removed using the Samtools package,[24] and base quality score recalibration, local realignment and variant calling were performed using the Genome Analysis Toolkit (GaTK V.1.3).[25] Alamut software was used to predict the effect of rare and novel variants on splicing. Potential pathogenic variants were confirmed using Sanger sequencing and further characterised using cDNA sequencing.

Sanger sequencing

Custom primers were designed to confirm non-coding variants of interest identified. Primer sequences are available on request. Amplicon fragments were bidirectionally sequenced with universal M13 primers using the Big Dye Terminator V.3.1 cycle sequencing kit and an ABI 3730 DNA Analyzer (Life Technologies, Carlsbad, California, USA). Sequences were compared with the *ACVRL1* and *ENG* reference sequences (NM_000020.2 and NM_000118.3, respectively) using Mutation Surveyor (SoftGenetics, State College, Pennsylvania, USA).

cDNA sequencing

RNA was extracted from peripheral blood and converted into cDNA for select cases using reverse transcriptase with random primers. Exonic primers were used to specifically amplify and Sanger sequence the region of *ACVRL1* that had the potential splicing defect. Primer sequences are available on request.

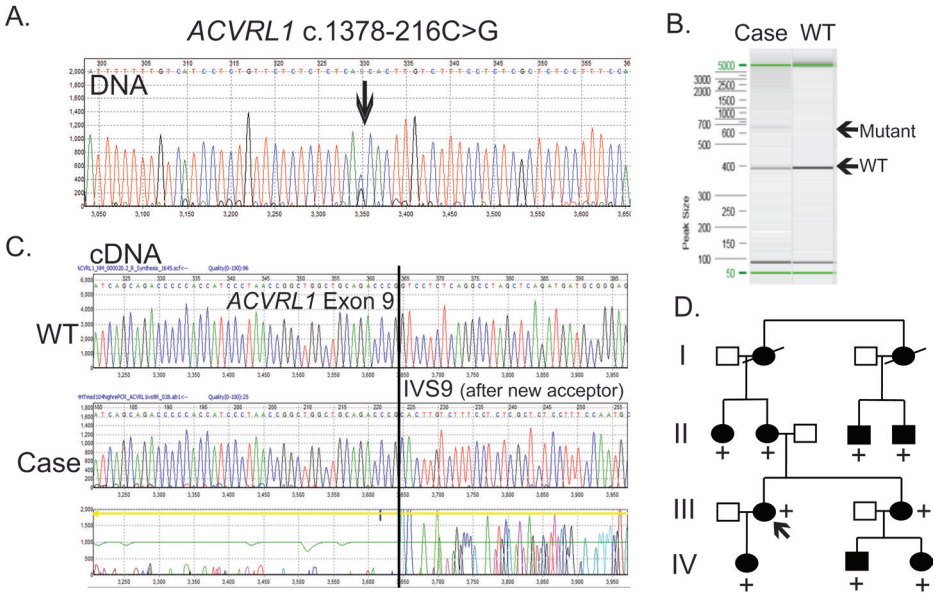

Figure 1 Novel *ACVRL1* intron 9 variant (c.1378-216C>G) identified by genome sequencing disrupts splicing causing hereditary haemorrhagic telangiectasia. (A) Sanger sequencing confirmed the novel *ACVRL1* intron 9 (intervening sequencing (IVS)9) variant. (B) A gel depicts amplified cDNA with the patient having two bands compared with one in the wild type (WT). The top band corresponds to the aberrantly spliced product with the partial retention of intron 9 (confirmed by cDNA sequencing in (C)). (D) The splice variant tracked through the family (+).

RESULTS

Genome sequencing reveals non-coding variants in *ACVRL1* that disrupt splicing

Thirty-five symptomatic individuals and 2 asymptomatic family members from 13 families with HHT or suspected to have HHT, who had previously tested negative by exome sequencing, underwent genome sequencing. All families had also previously tested negative for a mutation in *ENG*, *ACVRL1* and *SMAD4* by Sanger sequencing and deletion/duplication analysis by multiplex ligation-dependent probe amplification (MLPA). Genomes were sequenced to ~60× average coverage and novel and rare variants were analysed.

A novel deep intronic heterozygous variant (*ACVRL1* c.1378-216C>G) was identified and confirmed by Sanger sequencing in one four-generation family (family 1) with HHT (figure 1A and table 1, case 1). Eighteen members of this family had been evaluated at the University of Utah HHT Clinic. Nine were affected with HHT on clinical grounds, while six were suspected affected and three considered unknown. In addition to epistaxis and characteristic telangiectases, six of the affected had pulmonary AVM, five requiring treatment by transcatheter embolisation. Previous linkage analysis had shown that the disease in the family linked to chromosome 12q, consistent with linkage to *ACVRL1* on chromosome 12q13 (data not shown). This deep intronic *ACVRL1* variant (*ACVRL1* c.1378-216C>G) was predicted by multiple splice site prediction programmes to create a new 'AG' splice acceptor site and alter splicing (table 1). To confirm the predicted splicing defect in this family, RNA was obtained from an affected family member and converted to cDNA. Exon-specific primers showed the presence of two mRNA species, one correlating in size to the wild-type band and another higher mutant band (figure 1B). Sanger sequencing of the higher band confirmed the predicted partial retention of *ACVRL1* intron 9 in the aberrantly spliced product (figure 1C). This variant tracked with the disease in the family, as all nine affected family members had the variant (figure 1D). Of note, a rare benign variant (rs111710113, 0.004 minor allele frequency) at the same location but a different nucleotide change (*ACVRL1* c.1378-216C>T) was not predicted to affect splicing as it does not create a new 'AG' acceptor site.

Novel *ACVRL1* translocation identified using genome sequencing

Another family (family 2) that underwent genome sequencing was found to carry a novel heterozygous *ACVRL1* intron 9:chromosome 3 translocation (figure 2 and table 1, case 2). The proband aged 27 years has epistaxis, telangiectases in characteristic locations and a pulmonary AVM. Her mother has epistaxis and telangiectases. PCR primers flanking the t(12,3)(q13,p21) translocation were used to confirm the translocation in the affected family members by Sanger sequencing their genomic DNA across the breakpoints (chr12:52313774_chr3:48252144) (figure 2A). The chromosome 3 breakpoint was located in an intergenic region; however, the chromosome 12 breakpoint was located in *ACVRL1* intron 9 (c.1348-769). The translocation is expected to interrupt the *ACVRL1* gene, because it would lead to a loss of the last coding exon (exon 10). The proband and her affected mother carried the variant, whereas her unaffected grandmother did not carry the variant (figure 2B). These translocation-specific primers did not give amplification in the unaffected family member or wild-type control (see online supplementary figure 1). This is the first report of a translocation causing HHT. Although Sanger sequencing confirmed the breakpoints of the translocation for chromosome 12, no additional blood sample was available for karyotyping to confirm at the chromosomal level.

Based on genome sequencing results in these original 13 families, two families with clinically confirmed HHT had a deep intronic *ACVRL1* variant involving intron 9 that caused HHT. The remaining 11 families (10 of whom had been considered to be atypical in some way for HHT) did not have an identifiable causal variant in the non-coding regions of *ENG*, *ACVRL1* or *SMAD4*, and genes examined as part of the HHT clinical

Table 1 Clinical and molecular findings of eight unrelated probands with a novel *ACVRL1* intron 9 variant

Case #	Gene	IVS #	Nucleotide change	Protein effect	Predicted splicing scores*	Variant classification	Age, sex	Fam Hx	E	T	AVM	Other
1	*ACVRL1*	9	c.1378-216C>G	Splicing† Partial retention of IVS9	89.2, 84.8	Pathogenic	63 F	Yes	x		PAVM, HAVM	Tracked in family
2	*ACVRL1*	9	t(12,3)(q13,p21) translocation	Splicing†	NA	Pathogenic	27 F	Yes	x	x	PAVM	Tracked in family
3	*ACVRL1*	9	c.1378-274C>G	Splicing† Partial retention of IVS9	87.9, 79.5	Pathogenic	41 M	Unknown	x	x	PAVM	Multiple T in characteristic locations
4	*ACVRL1*	9	c.(1378-156C>A; 1378-155T>G]	Splicing† Partial retention of IVS9	88.5, 85.1	Pathogenic	64 M	Yes	x	x	Colon AVM	
5	*ACVRL1*	9	c.1378-131C>G	Splicing† Partial retention of IVS9	80.3, 73.8	Pathogenic	47 F	Yes	x	x	PAVM, HAVM	E monthly in childhood; T on lips
6	*ACVRL1*	9	c.1378-78T>G	Partial retention of IVS9 predicted	91.7, 80.0	Pathogenic	70 F	Yes		x	PAVM, HAVM	GI and skin T
7	*ACVRL1*	9	c.1378-69C>A	Splicing† Partial retention of IVS9	91.5, 84.5	Pathogenic	49 M	Yes	x		Diffuse PAVM	Mother has E and a stroke
8	*ACVRL1*	9	c.1378-69C>A	Splicing† Partial retention of IVS9	91.5, 84.5	Pathogenic	80 F	Yes	x	x	PAVM	T on lips and hands; Fam Hx of E, T, CAVM and PAVM

*Human Splicing Finder and SpliceSite Finder scores compared with native splice site (89.5, 78.8).
†Molecular characterisation using cDNA sequencing was performed to confirm the splicing defect.
IVS, intervening sequencing; ESE, exon-splicing enhancer; Fam Hx, family history; E, epistaxis; T, telangiectasia; AVM, arteriovenous malformation; PAVM, pulmonary AVM; HAVM, hepatic AVM; CAVM, cerebral AVM, GI, gastrointestinal.

differential (*GDF2* and *RASA1*) were also negative. Genome data are still being analysed for these families for gene discovery purposes.

ACVRL1 intron 9 mutation hotspot discovered using NGS panel sequencing

Given the genome sequencing results in the initial 13 families, 87 additional unrelated individuals whose samples had been submitted to our clinical laboratory for testing which included the HHT genes, with a negative result, were evaluated using an 'HHT genome' panel. This small custom NGS panel was designed to specifically target the coding and non-coding regions including the 5'UTR and 3'UTR of *ENG*, *ACVRL1* and *SMAD4*, as well as the coding regions of *BMP9*/*GDF2* and *RASA1*.

Six additional cases (cases 3–8) were identified as having a novel heterozygous *ACVRL1* intron 9 variant predicted by multiple splice prediction programmes to affect splicing (table 1). Case 3 had the deepest intronic *ACVRL1* variant (c.1378-274C>G) (figure 3A). Two unrelated cases, 7 and 8, had the same *ACVRL1* variant (c.1378-69C>A) and similar phenotypic findings. Clinical and molecular findings for these cases are summarised in table 1. All of these cases had three or more Curaçao clinical diagnostic criteria[2]; and all had a solid organ AVM, with pulmonary AVMs the most common in 5 of 6 cases (83%) followed by hepatic AVMs in 2 of 6 cases (33%).

All novel *ACVRL1* intron 9 variants identified using the custom capture panel were predicted to create a new 'AG' splice acceptor and ultimately result in the partial retention of the intron 9 sequence directly following each new acceptor site. Predicted splicing defects for each case were confirmed by amplifying the cDNA and sequencing the aberrantly spliced species (see online supplementary figure 2) except for case 6, in which a subsequent RNA patient sample could not be obtained.

For case 6 with the *ACVRL1* c.1378-78T>G variant, the Human Splicing Finder and SpliceSite Finder prediction scores (91.7 and 80, respectively) were stronger compared with the native intron 9 splice acceptor (89.5, 78.8). The remaining 81 cases did not have an identifiable causal variant in the non-coding regions of *ENG*, *ACVRL1* or *SMAD4*, and genes examined as part of the HHT clinical differential (*GDF2* and *RASA1*) were also negative.

DISCUSSION

These results reveal that non-coding region variants play a larger role in HHT than previously thought. Overall, 8% of these cases (8/100) with some suspicion of HHT, with previous negative molecular testing, had a novel non-coding *ACVRL1* intron 9 variant that was proven to disrupt splicing. However, it should be emphasised that this cohort consists of cases that were considered suspicious for HHT by a clinician; but only 28/100 were reported to have three or more cirteria for HHT, 47/100 with two and 15/100 with only one clinical criteria. Thus, 8% clearly underestimates the detection rate of pathogenic intron 9 variants in individuals who meet clinical diagnostic criteria for HHT, but have previously tested 'negative' in a clinical genetics laboratory. The detection rate for a non-coding region variant of *ACVRL1* intron 9 was 29% (8/28) for individuals/families reported to have three or more Curaçao criteria.

It is noteworthy that all eight cases in which a non-coding region variant was identified in this study met clinical diagnostic criteria for HHT. Seven of the eight cases/families were diagnosed at an HHT Center of Excellence in North America, with the eighth case diagnosed by a medical geneticist with a particular interest in HHT. This despite the fact that overall the cases sent from an HHT Center or Genetics Clinic were in the minority. It is our experience that the accuracy and rigorousness with which the Curaçao diagnostic criteria are applied is very dependent on

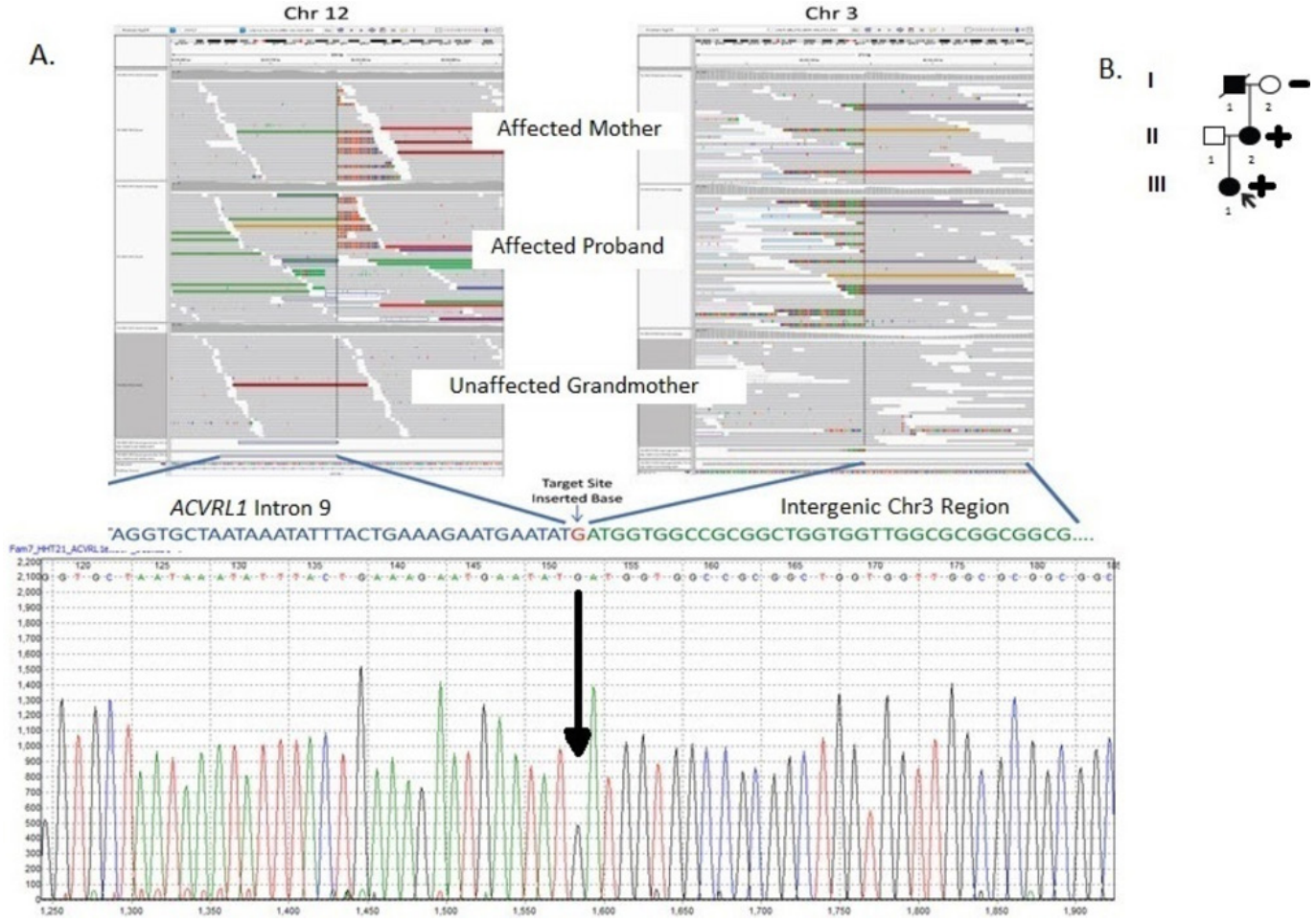

Figure 2 Novel *ACVRL1* intron 9:chromosome 3 translocation identified by genome sequencing. (A) Next-generation sequencing data depict the novel t(12,3)(q13,p21) translocation identified in the affected family members but not the unaffected family member. Sanger sequencing was used to confirm the translocation (lower panel). (B) The translocation tracked through the family (+).

a clinician's experience with vascular malformation syndromes; in particular, knowing what constitutes cutaneous lesions and nosebleeds that are typical for HHT.

Based on our results, we estimate that pathogenic variants in *ACVRL1* intron 9 alone account for about one-third (~29%) of cases with clinically confirmed HHT who do not have a genetic diagnosis using testing currently available in most molecular diagnostic laboratories. Furthermore, we estimate that *ACVRL1* intron 9 variants account for approximately 1% of all HHT cases. This is based on our experience that ~97% of cases with clinically confirmed HHT have a causative variant in the coding region or exon/intron boundaries of *ACVRL1*, *ENG* or *SMAD4* (McDonald, unpublished results). To increase clinical sensitivity, molecular diagnostic testing algorithms should be expanded to include this ~300 bp region. Any novel variants identified that create a new 'AG' splice acceptor site should be investigated further using confirmatory RNA studies to confirm the splicing defect.

One family had an *ACVRL1* intron 9:chromosome 3 translocation, the first translocation identified to cause HHT. The discovery of this translocation represents a new disease mechanism in HHT. In this family, *ACVRL1* transcription is predicted to be disrupted since the translocation is located before the last exon of the gene. While this novel translocation was identified using genome sequencing, this event could have been detected using karyotyping. It is likely that the prevalence of translocations in HHT is very rare.

Our findings suggest that deep intronic splice variants in *ACVRL1* intron 9 indicates a previously unrecognised mutational hotspot. Intriguingly, all variants identified in our cohort except for the translocation were located in a CT-rich region of *ACVRL1* intron 9 (figure 3B), and all result in the generation of a new 'AG' splice acceptor. It is well known that the canonical acceptor splice site region contains high CT-rich sequences; however, this CT-rich region usually does not extend to deep intronic regions.[26] All identified splice site variants in our clinical cases are located within 275 bases from exon 10 within intron 9. A sequence structure analysis up to 317 bases from exon 10 revealed that the CT ratio of this region is 85%, which is similar to the canonical acceptor splice site.[26] The high CT ratio in *ACVRL1* intron 9 decreases after 317 bp.

It is likely that any variant residing in the ~300 bp CT-rich region that creates a new 'AG' sequence in *ACVRL1* intron 9 will activate a new cryptic splice site. To test this hypothesis, we simulated this effect by altering every possible variant in this region to create a new 'AG' site using Alamut software in which multiple splice prediction scores were obtained and compared with the native 'AG' intron 9 acceptor splice site. In nearly all (42) 'AG' variant simulations, the simulated variants were predicted to create a strong cryptic acceptor splice site (figure 3C). When

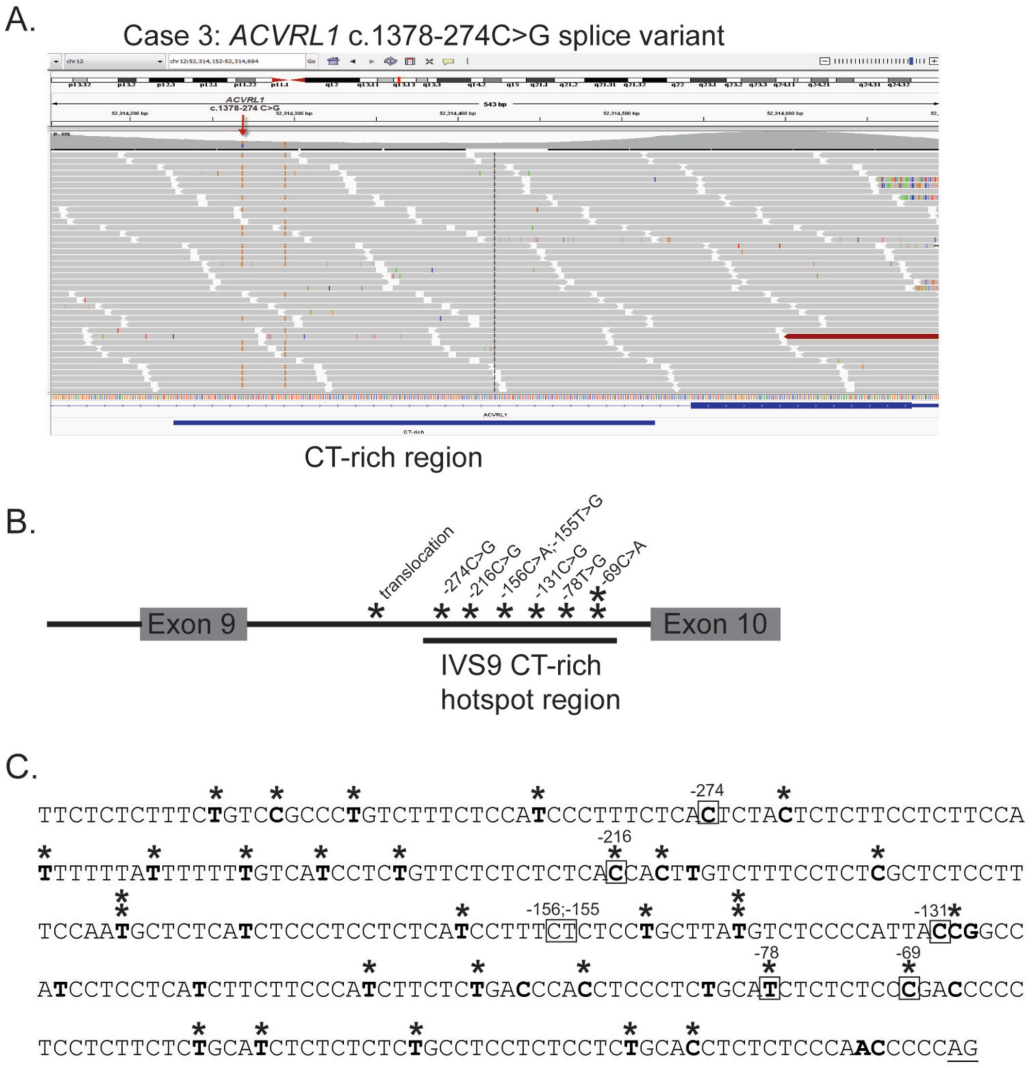

Figure 3 *ACVRL1* intron 9 mutation hotspot. (A) *ACVRL1* intron 9 variant (c.1378-274C>G) identified by panel sequencing depicts the CT-rich region below. (B) Diagram of *ACVRL1* intron 9 variants (*) in which most create a new 'AG' splice acceptor in the CT-rich region causing aberrant splicing and hereditary haemorrhagic telangiectasia. (C) Sequence of the *ACVRL1* intron 9 hotspot region depicting the bases in bold in which variants at that location were simulated by mutation to create a new 'AG' acceptor site and evaluated using Alamut. Starred (*) bases had a splice site prediction score from either the Human Splicing Finder or the SpliceSite Finder that was higher than the score from the native splice acceptor, AG. Boxed bases represent the location of a variant that affected splicing in a clinical case. Two base locations are indicated with two stars because that one location could create two 'AG' sites depending on the in silico variant. IVS, intervening sequencing.

we compared the prediction scores of these simulated 'AG' variants with the scores from the native acceptor site, we found that 30 of 42 simulated AG sites had a higher prediction score than the native site (figure 3C, see starred bases). This effect was not observed when the same exercise was applied to other introns in *ACVRL1* such as in intron 6 where only 13 'AG' simulated variants were predicted to create a new cryptic splice site.

The polypyrimidine tract is a well-known motif for recruiting spliceosome proteins to the pre-mRNA.[26] However, it has not been well studied if the polypyrimidine tract or simply the presence of a CT-rich sequence may increase the chance of a new AG acceptor site as a disease-causing mechanism. It requires more computational and some functional studies to answer this question.

In this study, we have shown the importance of non-coding region changes on the HHT disease mechanism as well as detailed clinical phenotyping to reveal these regions or variants. Genome and custom panel sequencing revealed a non-coding region pathogenic variant in 8 of 100 individuals/families suspected to have HHT, with no causative variant detected to date. We previously reported that non-coding mutations in the 5'UTR region of *ENG* cause HHT in ~1%–2% of cases.[13] Although non-coding variants can be difficult to interpret and characterise, our studies highlight the importance of including these regions in molecular diagnostic testing for HHT.

CONCLUSION

Non-coding variants in *ENG* and *ACVRL1* play a greater role in causing HHT than previously thought, suggesting molecular testing expansion to include non-coding regions, particularly the *ACVRL1* intron 9 CT-rich mutation hotspot. Variants in these regions may explain many of the families with definite HHT according to Curaçao criteria who have not had a causative genetic variant detected to date in a clinical laboratory. Using RUFUS analysis of genome data, a new causal mechanism of

HHT was identified in one family who had a novel *ACVRL1* intron 9:chromosome 3 translocation. Despite the difficulty of interpreting deep intronic variants, our study highlights the importance of non-coding regions in the disease mechanism of HHT. In particular, the addition of the ~300 bp CT-rich region of *ACVRL1* intron 9 to HHT molecular diagnostic testing algorithms will improve clinical sensitivity.

Author affiliations
[1]ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT, Salt Lake City, USA
[2]Department of Pathology, University of Utah, Salt Lake City, Utah, USA
[3]HHT Center, Department of Radiology, University of Utah, Salt Lake City, Utah, USA
[4]USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, Utah, USA
[5]Department of Pediatric Genetics, Zeynep Kamil Maternity and Children's Training and Research Hospital, Istanbul, Turkey
[6]Division of Cardiovascular Medicine, Department of Medicine, University of Utah, Salt Lake City, Utah, USA
[7]Yorkshire Regional Genetics Service, Chapel Allerton Hospital, Leeds, UK
[8]Medical Genetics, Department of Pediatrics, Mass General Hospital for Children, Boston, Massachusetts, USA
[9]Department of Internal Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correction notice Thsi article has been corrected since it was published online first. Details related to Dr Farrell's funding have ben added to the 'Funding' section.

Acknowledgements The authors would like to thank the patients and their family members for participating in this research. The authors would also like to thank the members of the ARUP Molecular Genetics and Genomics Clinical Laboratories for assisting in the sequence analysis of these patients.

Contributors WWD planned, conducted all experiments, analysed the data and drafted the manuscript. JM and PBT planned experiments, obtained all samples, interpreted data and drafted the manuscript. AF, MV, RM and GM performed the bioinformatics analysis, assisted in the interpretation of results and reviewed the manuscript. GA, PJ, CVW and EB assisted with experiments, data analysis and reviewed the manuscript. KW, JT, AEL and REP saw the patients, obtained samples and reviewed the manuscript.

Funding This study was funded by Cure HHT, the Chan Soon-Shiong Family Foundation (Heritage 1K Utah Genome Project) and the ARUP Institute for Clinical and Experimental Pathology. Dr Gulsen Akay was supported by the Scientific and Technological Research Council of Turkey (TUBİTAK), with 2219 Postdoctoral Research Fellowship. Dr Andrew Farrell was funded from a TL1 fellowship (TL1 TR001066).

Competing interests None declared.

Patient consent Not required.

Ethics approval This study was approved by the University of Utah Institutional Review Board (IRB #7275, 35637 and 20480).

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Kjeldsen AD, Vase P, Green A. Hereditary haemorrhagic telangiectasia: a population-based study of prevalence and mortality in Danish patients. *J Intern Med* 1999;245:31–9.
2 Shovlin CL, Guttmacher AE, Buscarini E, Faughnan ME, Hyland RH, Westermann CJ, Kjeldsen AD, Plauchu H. Diagnostic criteria for hereditary hemorrhagic telangiectasia (Rendu-Osler-Weber syndrome). *Am J Med Genet* 2000;91:66–7.
3 Shovlin CL, Hughes JM, Tuddenham EG, Temperley I, Perembelon YF, Scott J, Seidman CE, Seidman JG. A gene for hereditary haemorrhagic telangiectasia maps to chromosome 9q3. *Nat Genet* 1994;6:205–9.
4 Johnson DW, Berg JN, Baldwin MA, Gallione CJ, Marondel I, Yoon SJ, Stenzel TT, Speer M, Pericak-Vance MA, Diamond A, Guttmacher AE, Jackson CE, Attisano L, Kucherlapati R, Porteous ME, Marchuk DA. Mutations in the activin receptor-like kinase 1 gene in hereditary haemorrhagic telangiectasia type 2. *Nat Genet* 1996;13:189–95.
5 Gallione CJ, Repetto GM, Legius E, Rustgi AK, Schelley SL, Tejpar S, Mitchell G, Drouin E, Westermann CJ, Marchuk DA. A combined syndrome of juvenile polyposis and hereditary haemorrhagic telangiectasia associated with mutations in MADH4 (SMAD4). *Lancet* 2004;363:852–9.
6 Abdalla SA, Letarte M. Hereditary haemorrhagic telangiectasia: current views on genetics and mechanisms of disease. *J Med Genet* 2006;43:97–110.
7 Bayrak-Toydemir P, Mao R, Lewin S, McDonald J. Hereditary hemorrhagic telangiectasia: an overview of diagnosis and management in the molecular era for clinicians. *Genet Med* 2004;6:175–91.
8 Richards-Yutz J, Grant K, Chao EC, Walther SE, Ganguly A. Update on molecular diagnosis of hereditary hemorrhagic telangiectasia. *Hum Genet* 2010;128:61–77.
9 McDonald J, Damjanovich K, Millson A, Wooderchak W, Chibuk JM, Stevenson DA, Gedge F, Bayrak-Toydemir P. Molecular diagnosis in hereditary hemorrhagic telangiectasia: findings in a series tested simultaneously by sequencing and deletion/duplication analysis. *Clin Genet* 2011;79:335–44.
10 Gallione CJ, Richards JA, Letteboer TG, Rushlow D, Prigoda NL, Leedom TP, Ganguly A, Castells A, Ploos van Amstel JK, Westermann CJ, Pyeritz RE, Marchuk DA. SMAD4 mutations found in unselected HHT patients. *J Med Genet* 2006;43:793–7.
11 ARUP Laboratories. ENG mutation database. http://arup.utah.edu/database/ENG/ENG_welcome.php (accessed 1 Apr 2018).
12 Fontalba A, Fernández-Luna JL, Zarrabeitia R, Recio-Poveda L, Albiñana V, Ojeda-Fernández ML, Bernabéu C, Alcaraz LA, Botella LM. Copy number variations in endoglin locus: mapping of large deletions in Spanish families with hereditary hemorrhagic telangiectasia type 1. *BMC Med Genet* 2013;14:121.
13 Damjanovich K, Langa C, Blanco FJ, McDonald J, Botella LM, Bernabeu C, Wooderchak-Donahue W, Stevenson DA, Bayrak-Toydemir P. 5'UTR mutations of ENG cause hereditary hemorrhagic telangiectasia. *Orphanet J Rare Dis* 2011;6:85.
14 Albiñana V, Zafra MP, Colau J, Zarrabeitia R, Recio-Poveda L, Olavarrieta L, Pérez-Pérez J, Botella LM. Mutation affecting the proximal promoter of Endoglin as the origin of hereditary hemorrhagic telangiectasia type 1. *BMC Med Genet* 2017;18:20.
15 McDonald J, Wooderchak-Donahue W, VanSant Webb C, Whitehead K, Stevenson DA, Bayrak-Toydemir P. Hereditary hemorrhagic telangiectasia: genetics and molecular diagnostics in a new era. *Front Genet* 2015;6:1.
16 Cole SG, Begbie ME, Wallace GM, Shovlin CL. A new locus for hereditary haemorrhagic telangiectasia (HHT3) maps to chromosome 5. *J Med Genet* 2005;42:577–82.
17 Bayrak-Toydemir P, McDonald J, Akarsu N, Toydemir RM, Calderon F, Tuncali T, Tang W, Miller F, Mao R. A fourth locus for hereditary hemorrhagic telangiectasia maps to chromosome 7. *Am J Med Genet A* 2006;140:2155–62.
18 Wooderchak-Donahue WL, McDonald J, O'Fallon B, Upton PD, Li W, Roman BL, Young S, Plant P, Fülöp GT, Langa C, Morrell NW, Botella LM, Bernabeu C, Stevenson DA, Runo JR, Bayrak-Toydemir P. BMP9 mutations cause a vascular-anomaly syndrome with phenotypic overlap with hereditary hemorrhagic telangiectasia. *The American Journal of Human Genetics* 2013;93:530–7.
19 Hernandez F, Huether R, Carter L, Johnston T, Thompson J, Gossage JR, Chao E, Elliott AM. Mutations in RASA1 and GDF2 identified in patients with clinical features of hereditary hemorrhagic telangiectasia. *Hum Genome Var* 2015;2:15040.
20 Wooderchak-Donahue WL, Johnson P, McDonald J, Blei F, Berenstein A, Sorscher M, Mayer J, Scheuerle AE, Lewis T, Grimmer JF, Richter GT, Steeves MA, Lin AE, Stevenson DA, Bayrak-Toydemir P. Expanding the clinical and molecular findings in *RASA1* capillary malformation-arteriovenous malformation. *Eur J Hum Genet* 2018;29.
21 Amyere M, Revencu N, Helaers R, Pairet E, Baselga E, Cordisco M, Chung W, Dubois J, Lacour JP, Martorell L, Mazereeuw-Hautier J, Pyeritz RE, Amor DJ, Bisdorff A, Blei F, Bombei H, Dompmartin A, Brooks D, Dupont J, González-Enseñat MA, Frieden I, Gérard M, Kvarnung M, Hanson-Kahn AK, Hudgins L, Léauté-Labrèze C, McCuaig C, Metry D, Parent P, Paul C, Petit F, Phan A, Quere I, Salhi A, Turner A, Vabres P, Vicente A, Wargon O, Watanabe S, Weibel L, Wilson A, Willing M, Mulliken JB, Boon LM, Vikkula M. Germline loss-of-function mutations in EPHB4 cause a second form of capillary malformation-arteriovenous malformation (CM-AVM2) deregulating RAS-MAPK signaling. *Circulation* 2017;136:1037–48.
22 Yu J, Streicher JL, Medne L, Krantz ID, Yan AC. EPHB4 mutation implicated in capillary malformation-arteriovenous malformation syndrome: a case report. *Pediatr Dermatol* 2017;34:e227–e230.
23 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
24 Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 2008;18:1851–8.
25 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
26 Lim LP, Burge CB. A computational analysis of sequence features involved in recognition of short introns. *Proc Natl Acad Sci U S A* 2001;98:11193–8.