

## ARTICLE OPEN



# uAUG creating variants in the 5'UTR of ENG causing Hereditary Hemorrhagic Telangiectasia

Omar Soukarieh<sup>1,2</sup>✉, Emmanuelle Tillet<sup>3</sup>, Carole Proust<sup>1</sup>, Charlène Dupont<sup>1</sup>, Béatrice Jaspard-Vinassa<sup>2</sup>, Florent Soubrier<sup>4,5</sup>, Aurélie Goyenville<sup>6</sup>, Mélanie Eyries<sup>4,5</sup> and David-Alexandre Trégouët<sup>1</sup>

Hereditary Hemorrhagic Telangiectasia (HHT) is a rare, autosomal dominant, vascular disorder. About 80% of cases are caused by pathogenic variants in *ACVRL1* (also known as *ALK1*) and *ENG*, with the remaining cases being unexplained. We identified two variants, c.-79C>T and c.-68G>A, in the 5'UTR of *ENG* in two unrelated patients. They create upstream AUGs at the origin of upstream overlapping open reading frames (uoORFs) ending at the same stop codon. To assess the pathogenicity of these variants, we performed functional assays based on the expression of wild-type and mutant constructs in human cells and evaluated their effect on *ALK1* activity in a BMP-response element assay. This assay is mandatory for molecular diagnosis and has been so far only applied to coding *ENG* variants. These variants were associated with a decrease of protein levels in HeLa and HUVEC cells and a decreased ability to activate *ALK1*. We applied the same experiments on three additional uoORF-creating variants (c.-142A>T, c.-127C>T and c.-10C>T) located in the 5'UTR of *ENG* and previously reported in HHT patients. We found that all the analyzed variants alter protein levels and function. Additional experiments relying on an artificial deletion in our mutated constructs show that identified uAUGs could initiate the translation indicating that the associated effect is translation-dependent. Overall, we have identified two 5'UTR *ENG* variations in HHT patients and shed new light on the role of upstream ORFs on *ENG* regulation. Our findings contribute to the amelioration of molecular diagnosis in HHT.

npj Genomic Medicine (2023)8:32; <https://doi.org/10.1038/s41525-023-00378-5>

## INTRODUCTION

Hereditary Hemorrhagic Telangiectasia (HHT), also known as Osler–Weber–Rendu syndrome, is a rare vascular disorder characterized by autosomal dominant transmission. It causes severe complications including bleeding, recurrent epistaxis, iron deficiency, anemia, mucocutaneous telangiectasias, and arteriovenous malformations (AVM) in different organs (e.g., lungs, liver, brain, digestive tract). In some cases, AVM can be asymptomatic and often goes undetected in HHT patients, thus increasing morbidity and mortality. In addition, there is a considerable inter- and intra-family variation in symptoms and clinical severity, even in cases resulting from an identical pathogenic variant<sup>1–3</sup>. HHT requires a multidisciplinary approach as it is a multiorgan disease for which management and treatment depend on individual case evaluation. International guidelines have been elaborated for the diagnosis and management of HHT<sup>4,5</sup>. According to these guidelines, anti-angiogenic and anti-fibrinolytic drugs are now recommended to treat bleeding and epistaxis in HHT patients<sup>5–8</sup> but treatment improvements are still awaited.

HHT affects about 1 in 5000–8000 people worldwide<sup>9–11</sup> and is caused by pathogenic variants in genes coding for proteins of the bone morphogenetic proteins (BMPs) signaling pathway belonging to the transforming growth factor  $\beta$  (TGF- $\beta$ ) superfamily. For molecular diagnosis, three main genes, *ENG*, *ACVRL1* and *SMAD4* are routinely screened. More recently, HHT-causing variants have been identified in *GDF2* (encoding BMP9) that is now also screened in molecular diagnosis of HHT<sup>10,12</sup>. It has been estimated that 80% of HHT cases are caused by pathogenic variants in *ENG* and *ACVRL1* genes encoding respectively endoglin and *ALK1*

proteins<sup>13,14</sup>. Endoglin and *ALK1* are part of a crucial signal pathway in endothelial cells. Briefly, *ALK1* is a serine-threonine kinase activating the phosphorylation of Smad1 and Smad5 in response to BMP9 or BMP10 ligands. Phosphorylated Smad1 and Smad5 will then bind to Smad4 and translocate into the nucleus to initiate specific gene transcription regulation. Endoglin has no kinase activity by itself but is able to enhance *ALK1* signaling. A cellular-based assay has been developed to characterize coding missense variants in *ACVRL1* and *ENG*<sup>15,16</sup> but has never been used for non-coding variants. The developed assay evaluates the potential effect of coding *ACVRL1* or *ENG* variants on the *ALK1* response to BMP9 by using a transfected reporter gene (luciferase) driven by a BMP-response element (BRE). This assay has been used to discriminate pathogenic from non-pathogenic variants and is now considered as a mandatory step in the molecular diagnosis of HHT. At least 15% of HHT cases remain without molecular explanations in exonic/flanking intronic regions of the three main causative genes<sup>17</sup> leading to diagnosis uncertainty for patients and their families. The remaining cases could be explained either by pathogenic variants in hypothetical undiscovered genes<sup>18–20</sup>, or non-coding variants in known HHT genes. Indeed, the interpretation of non-coding variants, including promoter, 5'UTR, 3'UTR, and deep intronic variants, is a real challenge in molecular diagnosis and their characterization can contribute to resolve unexplained cases of HHT.

In this study, we identified two variants in the 5'UTR of *ENG* in two heterozygous and unrelated HHT patients. These two variants, c.-79C>T and c.-68G>A, were predicted to create upstream AUG (uAUGs) at the origin of upstream overlapping open reading

<sup>1</sup>Univ. Bordeaux, INSERM, Bordeaux Population Health Research Center, UMR 1219, F-33000 Bordeaux, France. <sup>2</sup>Univ. Bordeaux, INSERM, BMC, U1034, F-33600 Pessac, France. <sup>3</sup>INSERM UMR U1292, Laboratoire BIOSANTE, Université Grenoble Alpes, Inserm, CEA, Grenoble, France. <sup>4</sup>Département de Génétique, Hôpital Pitié-Salpêtrière, Sorbonne Université, Assistance Publique-Hôpitaux de Paris, Paris, France. <sup>5</sup>Sorbonne Université, INSERM, UMR\_S 1166, Institute of Cardiometabolism And Nutrition (ICAN), Paris, France. <sup>6</sup>Université Paris-Saclay, UVSQ, Inserm, END-ICAP, Versailles, France. ✉email: omar.soukarieh@inserm.fr

**Table 1.** Bioinformatics predictions and clinical data for *ENG* 5'UTR variants analyzed in this study.

Position <sup>a</sup>	dbSNP ID	MAF	Created ORF (size)	Database	Classification	Associated phenotype(s)	Segregation data	References
c.-142A>T	NA	NA	uoORF (270 nts)	NA	NA	E, T, CAVM, and/or PAVM	Yes	<sup>21</sup>
c.-127C>T	rs1060501408	None	uoORF (255 nts)	HGMD HHT database ClinVar	Pathogenic/likely pathogenic/pathogenic	T, PAVM, CAVM, and/or spinal AVM	Yes	<sup>2,21,22</sup>
<b>c.-79C&gt;T</b>	rs1564466502	None	uoORF (207 nts)	ClinVar	Uncertain significance	E, T, and PAVM → stroke	No	This study
<b>c.-68G&gt;A</b>	NA	NA	uoORF (195 nts)	NA	NA	T and PAVM	No	This study
c.-10C>T	rs756994701	0.0000066	uoORF (138 nts)	HGMD HHT database ClinVar	Likely pathogenic/pathogenic	E and PAVM	Yes	<sup>23</sup>
c.-9G>A	rs368423516	0.00041	eCDS (+9 nts)	HGMD HHT database ClinVar	Conflicting interpretations/pathogenic mild	Infrequent E and T (heterozygous carriers) Daily E and T (homozygous carriers)	Yes	<sup>2</sup>

MAF minor allele frequency reported in GnomAD, NA non-applicable, None no MAF described in GnomAD, uoORF upstream overlapping open reading frame, eCDS elongated CoDing Sequence, nts nucleotides, HHT Hereditary Hemorrhagic Telangiectasia, E epistaxis, T telangiectasia, AVM arteriovenous malformation, CAVM cerebral AVM, PAVM pulmonary AVM.

<sup>a</sup>Variant position in the L-ENG NM\_001114753.3 transcript (c.1 corresponds to the A of the main AUG).

Bolded variants correspond to newly identified variants in HHT patients and analyzed here for the first time.

frames (uoORFs) ending at the same stop codon located at position c.125 within the coding sequence (CDS). Using different functional in vitro assays, including original ones, we demonstrated that these variants alter endoglin levels and function, thus arguing in favor of a pathogenic effect of these variants.

Interestingly, additional very rare non-coding variants creating uAUGs in the 5'UTR of *ENG* have been previously identified in HHT patients and reported in the literature<sup>2,21–23</sup> and in databases (<https://www.ncbi.nlm.nih.gov/clinvar/?gr=1>; <https://www.hgmd.cf.ac.uk/ac/all.php>; [https://arup.utah.edu/database/ENG/ENG\\_display.php](https://arup.utah.edu/database/ENG/ENG_display.php); Table 1). Three of these variants (c.-142A>T; c.-127C>T and c.-10C>T) are predicted to generate uoORFs also ending at the same stop codon at position c.125. Two of these variants (c.-142A>T and c.-127C>T) have been analyzed in vitro in published studies and have been associated with a decrease of endoglin levels<sup>2,21,22</sup>. While c.-127C>T and c.-10C>T have been reported as pathogenic in public databases, the c.-142A>T is absent from databases but suggested to be pathogenic in Ruiz-Llorente et al.<sup>21</sup>. An additional uAUG-creating variant (c.-9G>A) has been reported in the 5'UTR of *ENG*<sup>2</sup>. As the created uAUG is in frame with the CDS, it is predicted to generate an elongated CDS. This variant has been associated with a very mild decrease of the endoglin levels in vitro<sup>2</sup> and its clinical interpretation is considered as “conflicting” or “pathogenic mild” in HGMD and HHT databases, respectively. By deploying our functional assays on these four previously reported uAUG-creating *ENG* variants, our study provides additional arguments supporting a pathogenic role of these variants and sheds new light on the non-canonical regulation of *ENG*.

## RESULTS

### uoORF-creating *ENG* variants are associated with a decrease of protein levels in vitro

We identified two very rare candidate variants located in the 5'UTR of *ENG* in two unrelated individuals with unresolved molecular diagnosis for HHT (Table 1) among 274 sequenced individuals. The

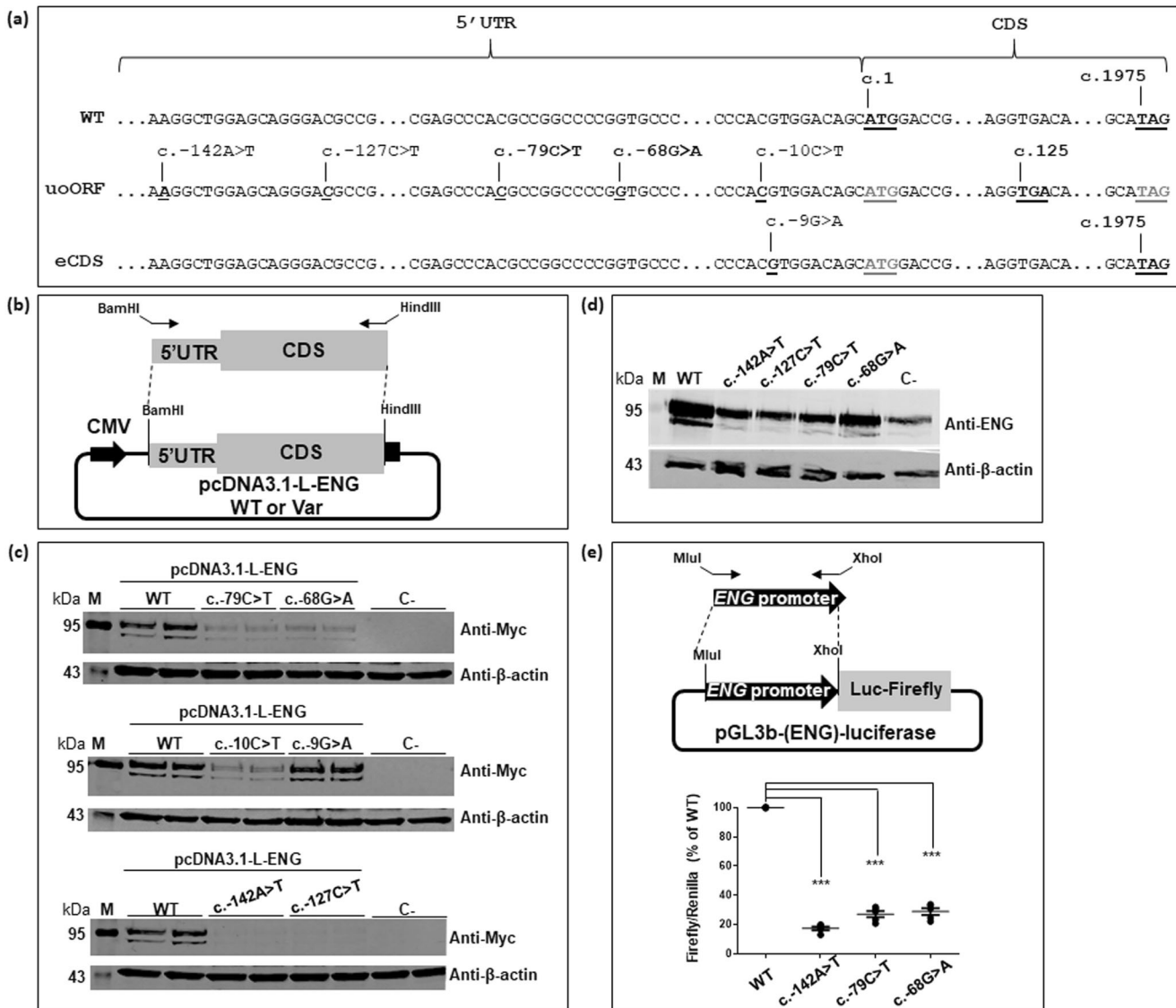
carrier of *ENG* c.-79C>T was diagnosed with HHT as she presented with epistaxis, mucosal and cutaneous telangiectasias and a pulmonary arteriovenous malformation that led to stroke. Her 3 monozygotic triplet sons were examined and had epistaxis, and one of them had 2 cutaneous telangiectasias suggestive of HHT. The carrier of the c.-68G>A variant presented with several telangiectasias, a pulmonary arteriovenous malformation and a familial history of epistaxis since her paternal grand-father, her father, her brother and her daughter were reported to have recurrent epistaxis.

The Variant Allele Frequency (VAF) of the identified c.-79C>T variant was 0.52 with 138 reads harboring the alternate T allele and 129 harboring the wild-type C allele. Corresponding VAF for the identified c.-68G>A variant was ~0.50 with 229 and 227 reads harboring the G and A alleles, respectively. No other pathogenic coding/splice candidate variants with minor allele frequencies <1% was identified in HHT genes for these patients (Supplementary Table 1).

Both variants, c.-79C>T and c.-68G>A, are predicted to create uAUGs that are out-of-frame with the CDS and predicted to create uoORFs ending at the same stop codon located within the CDS (Fig. 1a). The presence of these variants has been confirmed by Sanger sequencing.

These variants represent relevant candidates to explain HHT in these individuals, but their pathogenicity still needed to be demonstrated. Apart from these two variants, other 5'UTR variants have been reported in *ENG*. Among them, four are at the origin of new uAUGs, which generate either uoORFs ending at the same stop codon in the CDS (c.-142A>T, c.-127C>T and c.-10C>T) or an elongated CDS (c.-9G>A) since the predicted uAUG is in frame with the CDS (Fig. 1a and Table 1)<sup>2,21–23</sup>. To contribute to the classification of the identified variants, we started by evaluating their potential effect on in vitro endoglin levels, in parallel with the published variants indicated above (Fig. 1a–d).

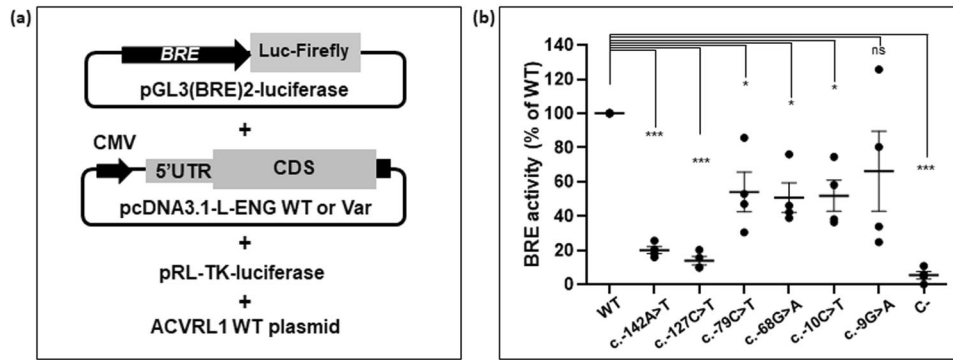
First, we observed a decrease of the protein steady-state levels associated with both c.-79C>T and c.-68G>A variants in transfected HeLa cells, reduced to 17% and 42% relative to the



**Fig. 1** uAUG-creating variants are associated with a decrease of ENG levels in vitro. **a** cDNA of the long (NM\_001114753.3) isoform of ENG transcripts. Positions of the identified 5'UTR uAUG-creating variants from this project and published studies<sup>2,21–23</sup> as well as position of the associated uStop codon at position c.125, and of main start (c.1) and stop (c.1975) codons are indicated. 5'UTR 5' untranslated region, CDS CoDing Sequence, WT wild type, uoORF upstream overlapping open reading frame, eCDS elongated CDS. **b** Schematic presentation of the pcDNA3.1-L-ENG constructs prepared and used for the evaluation of ENG steady-state levels in HeLa cells. Arrows indicate specific primers targeting the cDNA of ENG with extra sequences containing restriction sites to allow specific cloning in the expression vector. Position of the Myc-His tag is represented by the black square on the plasmid. CMV cytomegalovirus promoter, WT wild type, Var variant, 5'UTR 5' untranslated region, L long. **c, d** Western blot results on total proteins extracted from transfected HeLa cells with 1 µg of pcDNA3.1-L-ENG constructs or from transduced HUVEC cells with 20 MOI of lentiviruses containing ENG, respectively. Two bands of different molecular weights are observed for endoglin likely corresponding to more glycosylated (upper band) and less/non glycosylated (lower band) ENG monomers<sup>16</sup>. Anti-Myc and anti-ENG correspond to the used antibodies for the target protein on HeLa and HUVECs, respectively, and anti-β-actin corresponds to the antibody used against the reference protein. kDa kilodalton, M protein ladder, WT wild type, C- negative control corresponding to pcDNA3.1- empty vector. Shown results are representative of 5 independent experiments. Uncropped blots are shown in Supplementary Fig. 2. All blots were processed in parallel and derive from the same experiments. **e** Decrease of luciferase activity observed with uAUG-creating variants in ENG. Schematic presentation of the pGL3b-(ENG)-luciferase prepared and used in this assay is shown in the upper panel. Arrows indicate specific primers targeting the promoter of ENG with extra sequences containing restriction sites to allow specific cloning in the expression vector. Luc luciferase. In the lower panel, shown results with standard error of the mean correspond to Firefly/Renilla ratios normalized to the wild-type (WT) in 5 independent experiments. \*\*\**p* value < 10<sup>-3</sup> (two-factor ANOVA followed by Tukey's multiple comparison test of variants versus WT).

WT, respectively (Fig. 1c and Supplementary Figs. 1a and 2a). Interestingly, the published variant c.-10C>T was also associated with a decrease of endoglin levels in our assay (ENG levels ~32%; Fig. 1c and Supplementary Figs. 1a and 2b). Moreover, we observed a drastic effect (<10% of WT levels)

associated with c.-142A>T and c.-127C>T variants and a very moderate effect (~80% of WT levels) observed with the c.-9G>A. RT-qPCR results on RNAs extracted from transfected cells showed similar levels of ENG between WT and variants (Supplementary Figs. 1b and 2b, c).



**Fig. 2** ALK1 response to BMP9 stimulation is affected by 5'UTR *ENG* variants. **a** Schematic presentation of the co-transfected constructs in this assay. BRE BMP-response element, CMV cytomegalovirus promoter, WT wild type, Var variant, 5'UTR 5' untranslated region, L Long. **b** Decrease of BRE activity observed with all the analyzed variants in co-transfected NIH3T3 cells stimulated with BMP9 (5 pg/ml). Shown results with standard error of the mean correspond to the quantification of Firefly/Renilla and normalized to the wild-type (WT) ( $n = 4$ ). \*\*\* $p$  value  $< 10^{-3}$ ; \* $p$  value  $< 5 \times 10^{-2}$ , ns non-significant (two-factor ANOVA followed by Tukey's multiple comparison test of variants versus WT).

Similar results were obtained after overexpression of *ENG* variants in human umbilical vein endothelial cells (HUVECs). One can assume that protein levels obtained with the empty vector (C-on Fig. 1d) reflect the endogenous levels of endoglin and are not disturbed by the transduction. Thus, western blot analysis on total proteins extracted from transduced cells revealed drastic effects for c.-142A>T and c.-127C>T variants with similar levels to those of the endogenous endoglin (Fig. 1d and Supplementary Figs. 1c, d and 2d), followed by c.-79C>T with slightly higher levels (around 40% of WT levels), and then c.-68G>A associated with less pronounced effect (around 60% of WT levels).

Finally, we assessed the effect of *ENG* variants in the context of the promoter in a luciferase assay. We used this assay in order to study the potential effect of *ENG* variants on luciferase activity by altering the promoter activity and/or the translational mechanism in HeLa cells<sup>24,25</sup>. We observed a decrease of the luciferase activity with the tested variants (c.-142A>T; c.-79C>T and c.-68G>A; Fig. 1e). Importantly, the levels of luciferase appeared correlated to the protein levels detected for the same variants in the previous assays. At least for these variants, one could assume that the obtained effect on luciferase activity is related to the predicted uORFs, also observed in the context of the pGL3b-(*ENG*)-luciferase construct. Again, these results show the alteration of *ENG* levels by the analyzed variants, in concordance with the two previous assays.

#### Decreased levels of *ENG* related to 5'UTR variants alter the ability of *ENG* to activate BMP9-stimulated ALK1 receptor

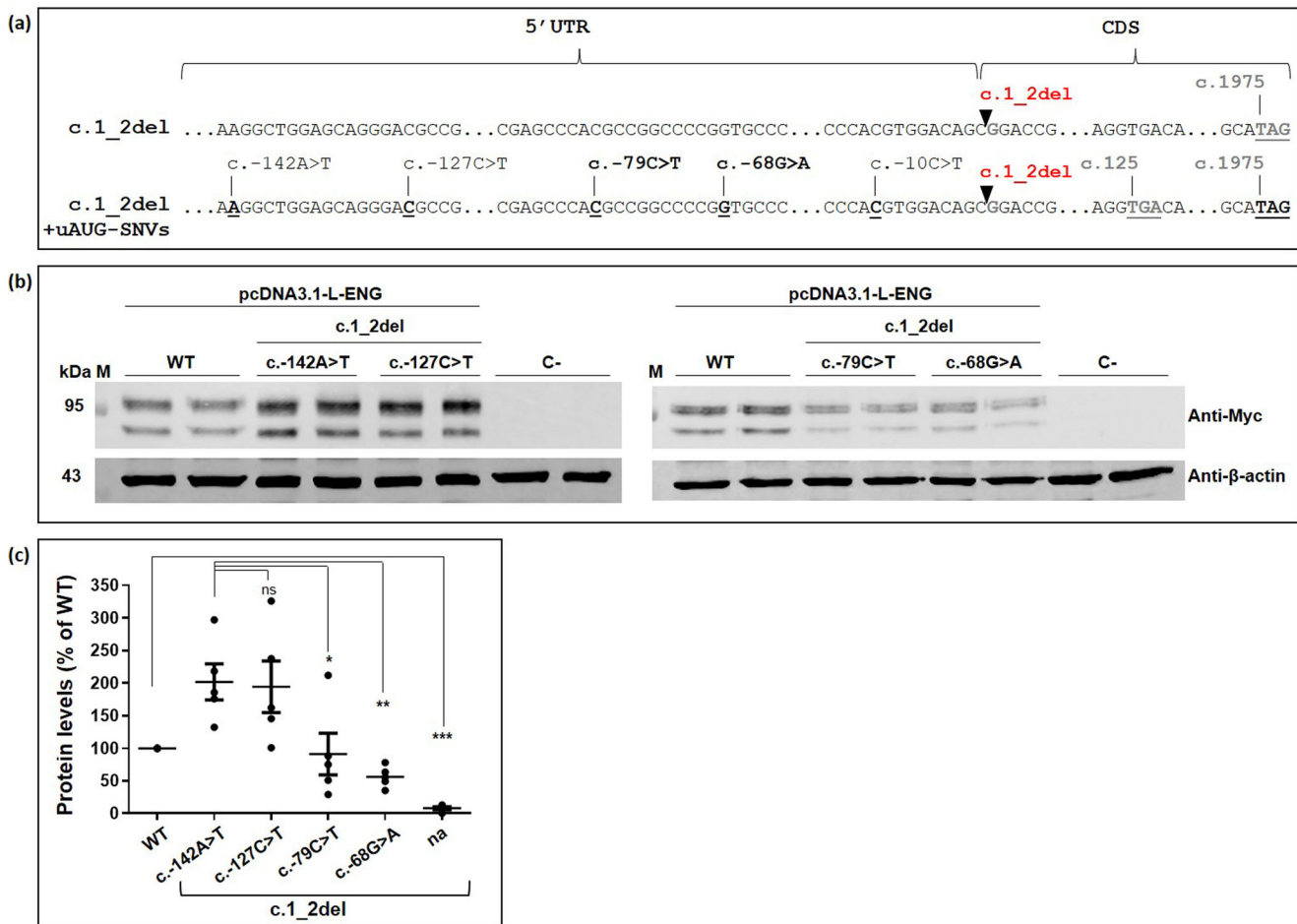
The ability of *ENG* variants (Table 1) to enhance ALK1-mediated BMP9 response was assessed using a BRE assay in NIH-3T3 cells. First, we demonstrated that pcDNA3.1-L-*ENG*-WT construct used in this study had similar stimulation efficiency than the WT *ENG* clone used as a reference in previous studies (Supplementary Fig. 3a). We then tested six different 5'UTR variants and showed that the 5'UTR variants had a reduced ability to stimulate ALK1 (Fig. 2). More precisely, c.-142A>T and c.-127C>T variants were associated with the lowest BRE activity detected ( $\leq 20\%$  of WT levels), followed by variants c.-79C>T, c.-68G>A and c.-10C>T which induced a moderate activity ( $\sim 50\%$  of WT levels) and finally variant c.-9G>A which was associated with a modest decrease of activity ( $\sim 70\%$  of WT levels) (Fig. 2). Of note, BRE activity obtained with this last variant was highly variable between independent experiments, leading to conflicting interpretations of the alteration of activity by this variant. Altogether, the decrease of BRE activity obtained with the *ENG* 5'UTR variants correlated with the *ENG* levels measured in vitro. Moreover, BRE stimulation by *ENG* variants was hampered only when low levels of expression

plasmid were transfected (below or equal to 1 ng/well; Supplementary Fig. 3b), highlighting that the lack of activity is directly linked to low *ENG* levels in the cells. Thus, these results demonstrate that uORF-creating variants could be considered as hypomorphic variants.

#### The created uAUGs in the 5'UTR of *ENG* are able to initiate the translation

We next investigated whether the uAUGs created by *ENG* variants in the 5'UTR could be used to initiate the translation of a new protein. In order to answer this question, we started by suppressing the main ATG of *ENG* by deleting the dinucleotide *ENG* c.1\_2del (Fig. 3). This deletion is predicted to completely abolish the translation of the CDS. In addition, in the presence of any of the *ENG* uAUG-creating variants, this deletion should transform the uORFs into an elongated CDS as the created uAUGs become in frame with the main stop codon. This leads to the identification of potentially translated proteins from uAUGs in vitro. pcDNA3.1-L-*ENG* carrying the deletion with or without 5'UTR variants were transfected in HeLa cells and *ENG* levels were assessed in western blot and RT-qPCR as described above.

Interestingly, we detected high protein levels associated to c.-142A>T and c.-127C>T variants in the presence of the deletion ( $>190\%$  in comparison to the WT construct) and less important levels with c.-79C>T and c.-68G>A ( $\sim 90\%$  and  $50\%$ , respectively) (Fig. 3b, c and Supplementary Fig. 4). Similar results were also obtained for c.-10C>T (Supplementary Figs. 5a, b and 6). The observed proteins appear to be of similar size to the WT product. Considering that the predicted size difference between the predicted elongated proteins initiated at the different uAUG and the WT protein is only 0.4 to 4.8 kDa, they would be undistinguishable in our Western blots. Importantly, these levels are inversely proportional to the levels of endoglin detected in our in vitro assays in association with the studied variants. Indeed, the highest protein levels obtained in this test were associated with variants that had the most drastic effect on *ENG* levels in vitro (Figs. 1 and 3). These results would suggest that the obtained effects with *ENG* variants are related to a potential competition between created uAUGs and the main AUG at the translation level. A very low amount of proteins ( $<10\%$  of the WT levels) is also detected with the pcDNA3.1-L-*ENG*-c.1\_2del construct that could result from the use of a non-canonical translation initiation site in the 5'UTR of *ENG* (Fig. 3c and Supplementary Fig. 5a). Apart for c.-68G>A, relative *ENG* transcript levels were similar in presence of the deletion with or without uAUG-creating variants as detected by RT-qPCR (Supplementary Fig. 5c). Combined effect of these variants (at least for the c.-68G>A) on the transcription and/or RNA stability cannot be excluded.



**Fig. 3 Created uAUGs in the 5'UTR of ENG seem to be able to initiate the translation.** **a** cDNA of the long (NM\_001114753.3) isoform of ENG transcripts. Positions of the identified 5'UTR uAUG-creating variants from this project and published studies<sup>2,21–23</sup> as well as position of the associated uStop codon (c.125), the introduced deletion c.1\_2del, and of main start (c.1) and stop (c.1975) codons are indicated. 5'UTR 5' untranslated region, CDS coding sequence, WT wild type, uoORF overlapping upstream open reading frame. **b** Western blot results on total proteins extracted from transfected HeLa cells with 1 µg of pcDNA3.1-L-ENG constructs. Two bands of different molecular weights are observed for endoglin likely corresponding to more glycosylated (upper band) and less/non glycosylated (lower band) ENG monomers<sup>16</sup>. Anti-Myc and anti-βactin correspond to the used antibodies for the target and the reference proteins, respectively. kDa kilodalton, M protein ladder, WT wild type, C- negative control corresponding to pcDNA3.1- empty vector. Shown results are representative of 5 independent experiments. Uncropped blots are shown in Supplementary Fig. 4. All blots were processed in parallel and derive from the same experiments. **c** Quantification of protein steady-state levels obtained in (b) and probably resulting from translation initiation at predicted uAUGs. For quantification, the average of each duplicate has been calculated from the quantified values and ENG levels for each sample have been normalized to the corresponding β-actin levels then to the WT (%). na, no 5'UTR variant introduced. Graphs with standard error of the mean are shown and are representative of 5 independent experiments. \*\*\**p* value < 10<sup>-3</sup>; \*\**p* value < 10<sup>-2</sup>; \**p* value < 5 × 10<sup>-2</sup>, ns non-significant (two-factor ANOVA followed by Tukey's multiple comparison test of variants versus WT).

## DISCUSSION

The starting point of this project was the identification of two variants in the 5'UTR of ENG (c.-79C>T and c.-68G>A) in two independent patients diagnosed with HHT but with unresolved molecular diagnosis. Here, we evaluated the effect of these variants on endoglin quantity and function using *in vitro* assays. Data accumulated in this study suggest that these two variants are responsible for a decrease in endoglin levels (17–42% of WT levels in HeLa cells). These results were reproduced in HUVECs, thus showing the relevance of the *in vitro* assay in HeLa cells. Moreover, we found that these variants are associated with a decrease of BRE activity *in vitro* (50% of the WT). Mallet et al. have previously considered that missense ENG variants associated with partial response to BMP9 (40–60%) were pathogenic. Indeed, they showed that variants they studied with partial response are associated with confirmed diagnosis of HHT and three of them showed familial co-segregation<sup>16</sup>. Based on this study, one can

consider that the two hypomorphic variants we identified are likely responsible of the HHT phenotypes in these patients. In addition, we here evaluated for the first time the functional effects of ENG c.-10C>T variant. We show that this variant is associated with ~32% of protein steady-state levels in HeLa cells and with ~50% of BRE response. This variant is classified as likely pathogenic in HGMD based on family history and patient's phenotype<sup>23</sup>. Consequently, data accumulated in this work suggest classifying c.-79C>T and c.-68G>A variants also as likely pathogenic. Furthermore, our results obtained for ENG c.-142A>T that parallel those we observed for ENG c.-127C>T, classified as pathogenic in HHT database, and combined with published data<sup>21</sup> add strong support for this variant being pathogenic. The observed effects are concordant with those obtained by other groups<sup>2,21,22</sup>. However slight difference in the quantified endoglin levels could be observed and may be due to differences in the used constructs and cells<sup>2,21,22</sup>.

While the two uoORFs we identified and three out of the four published uoORFs start at different uAUGs located within the 5'UTR, they all end at the same stop codon located at position c.125. The only exception holds for c.-9G>A variant, that creates a uAUG in frame with the CDS, and generates an elongated CDS, probably at the origin of a longer form of the ENG protein carrying three additional amino acids. In this study, we assessed the potential functional effect of all of these variants. Curiously, a weak decrease (~20%) of ENG levels and BRE activity have been associated with c.-9G>A variant compared to a drastic reduction observed for c.-142A>T and c.-127C>T and moderate reduction for c.-79C>T, c.-68G>A, and c.-10C>T. Interestingly, uoORF-creating variants (Fig. 1a and Table 1) are all associated with confirmed HHT diagnosis associated with severe symptoms. Concerning the three published uoORF-creating variants, molecular findings are consistent with clinical and familial data, suggesting that uoORF-creating variants in *ENG* can cause a severe form of HHT. This hypothesis still needs to be confirmed by analyzing all possible uoORF-creating variants in *ENG*, even those associated with non-uAUG codons. Of note, all uoORF-creating variants studied in this report and identified in severe forms of HHT were associated with ENG protein levels  $\leq 40\%$  in vitro (HeLa cells) and with  $\leq 50\%$  of BRE activity. Additional analysis would be interesting in order to define thresholds of pathogenicity of *ENG* variants in these assays. Furthermore, additional data will be mandatory to clarify the classification of c.-9G>A eCDS-creating variant.

Cellular-based BRE activity assay has been used to characterize *ENG* missense variants identified in HHT patients. Here, we adapted this assay to study 5'UTR variants in *ENG* and we observed that this kind of variants could alter the response of ALK1 to BMP9. However, while missense variants seem to alter the expression of ENG at the membrane and/or have negative effect on wild-type ENG, 5'UTR variants more likely cause a decrease of ENG levels in cells, which then results in an impaired response to BMP9. Thus, one could suggest that uAUG-creating variants associated with a decrease of ENG levels in vitro will probably alter the BRE activity and be likely pathogenic.

In total, three complementary assays have been used in this project to evaluate the functional effects of *ENG* 5'UTR variants and they all provided concordant results, suggesting that they could be reliably used for the functional characterization of uoORF-creating variants in *ENG*.

Again, we demonstrated that uAUG-creating variants in *ENG* could alter protein levels and function. However, this could happen via different mechanisms<sup>26,27</sup>. Indeed, upstream ORFs are part of the most known translational regulatory elements. They could, for example, enter in competition with the main coding sequence and affect the translation of this latter. To assess the potential translation of the created uoORFs, we started by deleting the main ATG, in the presence of uAUG-creating variants. Our results suggest that the created uAUGs could initiate the translation. More precisely, by deleting two nucleotides, we transformed uoORFs into elongated CDS. That led us to identify proteins that are probably initiated at the uAUGs. We combined these data with bioinformatics predictions to estimate the translation confidence of the created uAUGs in the 5'UTR of *ENG* by using the PreTIS tool based on the combination of 44 features calculated from mRNA sequence (<https://service.bioinformatik.uni-saarland.de/pretis/>)<sup>28</sup>. We applied PreTIS predictions on the six variants and obtained 0.67 (Low) to 0.96 (High) scores of translation confidence for the created uAUGs (Supplementary Table 2). At least for the variants evaluated in this study, translation could be initiated at uAUGs carrying PreTIS scores  $\geq 0.67$  in *ENG*. However, these scores do not seem to be able to predict the strength of the translation initiation and still need to be evaluated. Moreover, we analyzed the Kozak sequence surrounding the created uAUGs and found that those resulting from the c.-142A>T and c.-127C>T variants are surrounded with

stronger kozak sequences comparing to the 3 other uoORF-creating variants (c.-79C>T, c.-68G>A and c.10C>T) (Supplementary Table 2). These observations are consistent with our results obtained in vitro showing a higher amount of protein associated with the c.-142A>T and c.-127C>T variants (Fig. 3). The precise identification of the detected proteins will require further investigations using supplemental methods. In addition, it would be interesting to assess the functional potential of these proteins in the BRE activity assay. Indeed, if these proteins show some restoration of ENG functions, this may lead to new therapeutic approaches common to all uoORF-creating variants ending at the same uSTOP. Lastly, Kim and collaborators showed a decrease of RNA levels in carriers of the c.-127C>T variant<sup>22</sup> and c.-142A>T has been predicted to create a binding site for the transcription regulatory factor HOXA3<sup>21</sup>, suggesting that potential effect of the identified variants on the transcription cannot be excluded.

Finally, *ENG* is one of the rare examples of genes that are rich in uAUG-creating variants in the 5'UTR. Our study, along with others, demonstrated that 5'UTR variants predicted to create uoORFs should not be neglected in molecular diagnosis of genetic diseases. While we only studied uAUG-creating variants in the 5'UTR of *ENG* here, we are aware that 5'UTR variants creating non-canonical translation initiation codons or disrupting existing upstream ORFs, should also be given more attention.

## METHODS

### Clinical data of HHT patients and variant identification

Clinical diagnosis of HHT is determined based on the Curaçao criteria established by the HHT international committee<sup>29</sup>. These criteria include the presence of epistaxis (spontaneous and recurrent nose bleeds), telangiectasias (multiple, at characteristic sites such as lips, oral cavity, fingers and nose), visceral vascular lesions (gastrointestinal telangiectasias and/or arterio-venous malformations), and family history (a first-degree relative with HHT). The diagnosis of HHT is definite if three of the criteria are present, possible or suspected if two are present and unlikely if fewer than two are present.

As part of a molecular diagnosis routine conducted at the genetics department of the Pitié-Salpêtrière Hospital (Paris, France), 274 individuals with rare vascular diseases among which 53 are with suspected HHT have been screened for candidate pathogenic variants using a custom next-generation sequencing (NGS) targeted gene panel including HHT genes (*ACVRL1*, *ENG*, *SMAD4*, *GDF2*, *RASA1* and *EPHB4*) and additional genes related to other hereditary vascular diseases (Supplementary Table 3). Sequencing was performed on genomic DNA extracted from whole blood. VCF files from sequenced individuals were scrutinized using the MORFEE bioinformatics tool<sup>30</sup> in order to detect and annotate non-coding SNVs creating uAUGs (uAUG-SNVs) in the 5'UTR of the sequenced genes.

### Ethics declaration

The patients provided written informed consent for their DNA material to be used for genetic analysis in the context of molecular diagnosis in accordance with the French bioethics' laws (Commission Nationale de l'Informatique et des Libertés no 1774128). The aforementioned committee approved the study.

### Nomenclature

DNA sequence variant nomenclature follows current recommendations of the HGVS<sup>31</sup>.

### Plasmid constructs and expression in human cells

In order to evaluate the potential effect of *ENG* variants on the protein steady state levels, we performed the in vitro functional

assay described by Labrousche et al.<sup>32</sup>. To do so, we started by the amplification of the long isoform of ENG (L-ENG; NM\_001114753.3) cDNA from HeLa cells by using specific primers covering the entire 5'UTR and the CDS lacking its stop codon (ENG c.-303\_c.1974), and cloned it in the pcDNA3.1/myc-His(-) plasmid (Invitrogen) in frame with a Myc-His tag to obtain the wild-type (WT) clone (Supplementary Table 4). The PCR reaction was performed using Phusion High Fidelity (HF) DNA polymerase (ThermoFisher) and the cloning was carried out after double digestion of the inserts and plasmid with BamHI and HindIII. Mutated clones carrying ENG variants identified in this study or described in the literature (Fig. 1a and Table 1)<sup>2,21–23</sup> were prepared by directed mutagenesis on the WT generated clone, pcDNA3.1-L-ENG-WT (Fig. 1b), using specific back to back primers (Supplementary Table 4) and the Phusion HF DNA polymerase, followed by DpnI digestion of the template, phosphorylation of the generated PCR product, ligation and transformation of competent cells to obtain unique clones. In addition, we generated a supplemental construct in which we deleted the start codon of ENG (pcDNA3.1-L-ENG-c.1\_2del, Supplementary Table 4) and introduced separately ENG variants (c.-142A>T, c.-127C>T, c.-79C>T, c.-68G>A and c.-10C>T) in the latter construct. In presence of ENG c.1\_2del, the created uAUGs become in frame with the Myc tag, allowing the detection of potential proteins translated from the uAUGs. All the recombinant plasmids were verified by Sanger sequencing (Supplementary Table 4) (Genewiz).

In order to evaluate the effect of the identified variants on protein levels, HeLa cells (ATCC as original source) were transfected with the prepared pcDNA3.1-L-ENG constructs. HeLa were prepared in 6-well plates 24 h before the transfection with  $4.5 \times 10^5$  cells/well in RPMI medium (Gibco-Invitrogen) supplemented with 10% fetal calf serum (Gibco-Invitrogen). Transfections were performed in duplicate with 1  $\mu$ g of each plasmid using JetPRIME® reagent (Polyplus Transfection) according to the manufacturer's recommendations. Empty pcDNA3.1/myc-His(-) plasmid was used as negative control. The day of the transfection, cell confluence was at 60–80%. Cells were harvested and lysed 48 h after transfection to extract total RNA and protein. For this purpose, cells have been scraped and collected in 500  $\mu$ l of PBS. Then, they were split in 2 aliquots: 100  $\mu$ l aliquot for RNA extraction and 400  $\mu$ l aliquot for protein extraction, as indicated below.

As endoglin is mainly known for its endothelial function in HHT, we also assessed the ENG steady-state levels in HUVECs (Lonza). For this purpose, we first used the generated pcDNA3.1-L-ENG constructs to subclone the WT and mutant cDNA of ENG in pRRLsin-MND-MCS-WPRE plasmid, upstream of a MND promoter. Unlike for pcDNA3.1-L-ENG constructs, pRRLsin-MND-MCS-WPRE-L-ENG ones do not contain any tag and the CDS of ENG ends at its own stop codon (ENG c.-303\_c.1977). Then, lentiviruses were produced at the platform of vectorology (Vect'UB) of the University of Bordeaux (<https://www.tbmcore.u-bordeaux.fr/vectorologie/>) for the WT and ENG c.-142A>T, c.-127C>T, c.-79C>T and c.-68G>A variants. Twenty-four hours before transduction, HUVECs were prepared in 6-well plates with  $2.8 \times 10^5$  cells/well in EGM-2 medium (Endothelial Cell Growth Medium-2, Lonza). Cells were transduced in duplicate with 20 MOI (Multiplicity of Infection) of the generated lentiviruses by using 3.2 mg of protamine sulfate. Transduced cells were harvested with trypsin 72 h post-transduction and each well was transferred to a P100 plate. For each construct, one plate has been used for whole protein extraction, 48 h after the transfer to P100 plates and the duplicate plate was used to freeze transduced cells.

### Protein preparation and western blot analysis

Whole protein extractions were performed with RIPA supplemented with protease inhibitors for a pellet collected from one well of HeLa cells or one P100 plate of HUVECs. Concentrations were

measured by using the BCA protein assay kit (Pierce™) following the manufacturer's instructions. Proteins were loaded on 10% SDS-PAGE gels in parallel with a protein prestained ladder (Euromedex) and transferred onto PVDF membranes (Bio-Rad) by using the trans-blot turbo transfer system (Bio-Rad). Membranes were incubated with monoclonal anti-(c-Myc Tag) antibody (Merck Millipore, #05–419) (HeLa extracts), anti-ENG (Abcam, #ab169545) (HUVECs extracts) to probe endoglin, and anti- $\beta$ -actin (Cell Signaling, #4970) (HeLa and HUVECs extracts) as a loading control. These antibodies were used at 1/3000, 1/2000 and 1/3000 dilutions, respectively. Fluorescent goat anti-mouse IgG Alexa Fluor 700 (ThermoFisher, #A-21036, 1/5000) was used against anti-(c-Myc Tag) and goat anti-Rabbit IgG (H + L) Alexa Fluor 750 (Invitrogen, #A-21039, 1/5000) was used against anti-ENG and anti- $\beta$ -actin. Odyssey Infrared Imaging System (Li-Cor Biosciences) in 700 and 750 channels was used to scan, reveal, and quantify the blots. For quantification, the average of each duplicate was computed from the quantified values and ENG levels for each sample were normalized to the corresponding  $\beta$ -actin levels then to the WT or the negative control levels (%). The two bands obtained for the Endoglin, corresponding to the more glycosylated (upper band) and less/non glycosylated (lower band) ENG monomers<sup>16</sup>, were taken together for the quantification. All blots were processed in parallel and derive from the same experiments.

### RNA isolation and RT-qPCR analysis

In order to evaluate ENG transcript levels in transfected HeLa cells, total RNA was isolated from the collected pellets by using the RNeasy mini kit (Qiagen) following the manufacturer's instructions. Extracted RNA was quantified and equal quantities were used for reverse transcription reaction, performed with the M-MLV reverse transcriptase (Promega). Then 20 ng of cDNA of each sample was used for the qPCR reaction (duplicate/sample) with ENG- or  $\alpha$ -tubulin-specific primers (Supplementary Table 4) and the GoTaq qPCR Master mix (Promega), in the presence of CXR reference dye, in a final volume of 10  $\mu$ l and 40 cycles of amplification on QuantStudio3 Real-Time PCR System (ThermoFisher). QuantStudio design and analysis software was used to analyze the results and transcript levels were normalized to the reference  $\alpha$ -tubulin gene.  $2^{-\Delta\Delta CT}$  method was used to calculate the relative amounts of ENG to  $\alpha$ -Tubulin in different samples.  $\Delta\Delta CT$  were calculated by taking into account the mean of qPCR duplicates followed by the mean of transfection duplicates for each sample. Reaction efficiency (90–110%) and melting curves were evaluated for each couple of primers.

### Luciferase assay

A complementary in vitro assay was deployed to evaluate the effect of the identified variants on the promoter activity. For this purpose, the ENG promoter, containing the basal promoter and the region carrying major transcriptional regulatory elements, as defined in Rius et al.<sup>24</sup>, was amplified by using specific primers (Supplementary Table 4). This promoter sequence corresponds to the 805 nucleotides located upstream of the main ATG. The amplified promoter has been cloned in pGL3-basic vector containing the CDS of Firefly luciferase to obtain the WT clone (Fig. 1e). Then, ENG variants c.-142C>T, c.-79C>T and c.-68G>A were introduced in parallel by directed mutagenesis, as described above for pcDNA3.1-L-ENG vectors. All the recombinant plasmids were verified by Sanger sequencing by using specific primers (Supplementary Table 4) (Genewiz). WT or mutant clones were co-transfected with a plasmid containing Renilla luciferase in triplicate in 96-well plate of HeLa cells. Forty-eight hours after the transfection, luciferase activity was measured by using the dual-glo luciferase assay system (Promega) directly in transfected wells by detecting luminescence with both Firefly and Renilla luciferases. Mean of the triplicates of Firefly/Renilla ratios of each sample was normalized to the WT.

### Functional effect of ENG variants on BRE activity in vitro

The BRE assay described in Mallet et al.<sup>16</sup> has been modified to assess the functionality of the 5'UTR ENG variants. Briefly, NIH-3T3 were seeded in 96-wells white plates (15,000 cells/well) in DMEM containing 1% fetal calf serum and transfected the following day by a mixture of plasmid (i) BRE luciferase reporter plasmid (75 ng) (ii) pRL-TK luc encoding Renilla luciferase (20 ng), pcDNA3-ALK1 (0.15 ng) and pcDNA3.1-L-ENG WT or 5'UTR-mutated constructs (0–10 ng/well). Four hours after transfection, cells were stimulated overnight with 5 pg/ml of BMP9 in serum-free medium (R&D Systems) and luciferase activity was measured with the twinlite Firefly and Renilla Luciferase Reporter Gene Assay System (PerkinElmer). Means of triplicate were calculated for each sample and Firefly/Renilla ratio of stimulated wells was normalized to that obtained in non-stimulated wells.

### Statistical data analysis

Differential protein and RNA levels, luciferase activity, and BRE activity were assessed using two-factor analysis of variance followed by Tukey's multiple comparison test. A threshold of  $p < 0.05$  was used to declare statistical significance.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

ENG constructs generated during this study are available upon request by email to the corresponding author (omar.soukarieh@inserm.fr).

### CODE AVAILABILITY

The initial version of MORFEE tool is available at <https://github.com/daissi/MORFEE>.

Received: 27 February 2023; Accepted: 3 October 2023;

Published online: 17 October 2023

### REFERENCES

- Kritharis, A., Al-Samkari, H. & Kuter, D. J. Hereditary hemorrhagic telangiectasia: diagnosis and management from the hematologist's perspective. *Haematologica* **103**, 1433–1443 (2018).
- Damjanovich, K. et al. 5'UTR mutations of ENG cause hereditary hemorrhagic telangiectasia. *Orphanet J. Rare Dis.* **6**, 85 (2011).
- Faughnan, M. E. et al. International guidelines for the diagnosis and management of hereditary haemorrhagic telangiectasia. *J. Med. Genet.* **48**, 73–87 (2011).
- Shovlin, C. L. et al. The European Rare Disease Network for HHT Frameworks for management of hereditary haemorrhagic telangiectasia in general and speciality care. *Eur. J. Med. Genet.* **65**, 104370 (2022).
- Faughnan, M. E. et al. Second International Guidelines for the Diagnosis and Management of Hereditary Hemorrhagic Telangiectasia. *Ann. Intern. Med.* **173**, 989–1001 (2020).
- Al-Samkari, H. Hereditary hemorrhagic telangiectasia: systemic therapies, guidelines, and an evolving standard of care. *Blood* **137**, 888–895 (2021).
- Al-Samkari, H. et al. An international, multicenter study of intravenous bevacizumab for bleeding in hereditary hemorrhagic telangiectasia: the InHIBIT-Bleed study. *Haematologica* **106**, 2161–2169 (2021).
- Dupuis-Girod, S. et al. European reference network for rare vascular diseases (VASCERN): when and how to use intravenous bevacizumab in Hereditary Haemorrhagic Telangiectasia (HHT)? *Eur. J. Med. Genet.* **65**, 104575 (2022).
- Dakeishi, M. et al. Genetic epidemiology of hereditary hemorrhagic telangiectasia in a local community in the northern part of Japan. *Hum. Mutat.* **19**, 140–148 (2002).
- Shovlin, C. L. et al. Mutational and phenotypic characterization of hereditary hemorrhagic telangiectasia. *Blood* **136**, 1907–1918 (2020).
- Kitayama, K. et al. Mutational and clinical spectrum of Japanese patients with hereditary hemorrhagic telangiectasia. *BMC Med. Genomics* **14**, 1–9 (2021).
- Balachandar, S. et al. Identification and validation of a novel pathogenic variant in GDF2 (BMP9) responsible for hereditary hemorrhagic telangiectasia and pulmonary arteriovenous malformations. *Am. J. Med. Genet. A* **188**, 959–964 (2022).
- Olivieri, C. et al. Analysis of ENG and ACVRL1 genes in 137 HHT Italian families identifies 76 different mutations (24 novel). Comparison with other European studies. *J. Hum. Genet.* **52**, 820–829 (2007).
- Komiyama, M., Ishiguro, T., Yamada, O., Morisaki, H. & Morisaki, T. Hereditary hemorrhagic telangiectasia in Japanese patients. *J. Hum. Genet.* **59**, 37–41 (2014).
- Ricard, N. et al. Functional analysis of the BMP9 response of ALK1 mutants from HHT2 patients: a diagnostic tool for novel ACVRL1 mutations. *Blood* **116**, 1604–1612 (2010).
- Mallet, C. et al. Functional analysis of endoglin mutations from hereditary hemorrhagic telangiectasia type 1 patients reveals different mechanisms for endoglin loss of function. *Hum. Mol. Genet.* **24**, 1142–1154 (2015).
- Prigoda, N. L. et al. Hereditary haemorrhagic telangiectasia: mutation detection, test sensitivity and novel mutations. *J. Med. Genet.* **43**, 722–728 (2006).
- Bayrak-Toydemir, P. et al. A fourth locus for hereditary hemorrhagic telangiectasia maps to chromosome 7. *Am. J. Med. Genet. A* **140**, 2155–2162 (2006).
- Cole, S. G., Begbie, M. E., Wallace, G. M. F. & Shovlin, C. L. A new locus for hereditary haemorrhagic telangiectasia (HHT3) maps to chromosome 5. *J. Med. Genet.* **42**, 577–582 (2005).
- Wallace, G. M. & Shovlin, C. L. A hereditary haemorrhagic telangiectasia family with pulmonary involvement is unlinked to the known HHT genes, endoglin and ALK-1. *Thorax* **55**, 685–690 (2000).
- Ruiz-Llorente, L. et al. Characterization of a family mutation in the 5' untranslated region of the endoglin gene causative of hereditary hemorrhagic telangiectasia. *J. Hum. Genet.* **64**, 333–339 (2019).
- Kim, M.-J. et al. Clinical and genetic analyses of three Korean families with hereditary hemorrhagic telangiectasia. *BMC Med. Genet.* **12**, 130 (2011).
- Bossler, A. D., 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.* **27**, 667–675 (2006).
- Rius, C. et al. Cloning of the promoter region of human endoglin, the target gene for hereditary hemorrhagic telangiectasia type 1. *Blood* **92**, 4677–4690 (1998).
- Graulich, W., Nettelbeck, D. M., Fischer, D., Kissel, T. & Müller, R. Cell type specificity of the human endoglin promoter. *Gene* **227**, 55–62 (1999).
- Silva, J., Fernandes, R. & Romão, L. Translational regulation by upstream open reading frames and human diseases. *Adv. Exp. Med. Biol.* **1157**, 99–116 (2019).
- Hinnebusch, A. G., Ivanov, I. P. & Sonenberg, N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* **352**, 1413–1416 (2016).
- Reuter, K., Biehl, A., Koch, L. & Helms, V. PreTIS: a tool to predict non-canonical 5' UTR translational initiation sites in human and mouse. *PLoS Comput. Biol.* **12**, e1005170 (2016).
- Shovlin, C. L. et al. Diagnostic criteria for hereditary hemorrhagic telangiectasia (Rendu-Osler-Weber syndrome). *Am. J. Med. Genet.* **91**, 66–67 (2000).
- Aissi, D. et al. MORFEE: a new tool for detecting and annotating single nucleotide variants creating premature ATG codons from VCF files. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.03.29.012054> (2020).
- den Dunnen, J. T. et al. HGVS recommendations for the description of sequence variants: 2016 update. *Hum. Mutat.* **37**, 564–569 (2016).
- Labrousse-Colomer, S. et al. A novel rare c.-39C>T mutation in the PROS1 5'UTR causing PS deficiency by creating a new upstream translation initiation codon. *Clin. Sci.* **134**, 1181–1190 (2020).

### ACKNOWLEDGEMENTS

This project was supported by the GENMED Laboratory of Excellence on Medical Genomics [ANR-10-LABX-0013], a research program managed by the National Research Agency (ANR) as part of the French Investment for the Future, the EPIDEMIO-M-VT Senior Chair from the University of Bordeaux initiative of excellence IdEX, the INSERM GOLD Cross-Cutting program, and the Lefoulon-Delalande Foundation.

### AUTHOR CONTRIBUTIONS

O.S. and D.-A.T. conceived the project. O.S., C.P., and E.T. designed the experiments. O.S., E.T., C.P., and C.D. performed the experiments. O.S., E.T., and C.P. analyzed the data. B.J.-V. and A.G. provided technical support and suggestions on the project and the experiments. F.S. and M.E. were in charge of clinical management of HHT patients and provided vcf files. O.S. and D.-A.T. drafted the paper that was further shared to co-authors who read/corrected and approved the final manuscript.



## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41525-023-00378-5>.

**Correspondence** and requests for materials should be addressed to Omar Soukarieh.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023