

Non-canonical regulation of Endoglin by rare and common variants: new molecular and clinical perspectives for Hereditary Hemorrhagic Telangiectasia and beyond

Omar Soukarieh^{1,2#}, Gaëlle Munsch^{1#}, Clémence Deiber¹, Caroline Meguerditchian¹, Carole Proust¹, Ilana Caro¹, Maud Tusseau³, Alexandre Guilhem^{4,5}, Shirine Mohamed⁶, INVENT consortium, Béatrice Jaspard-Vinassa², Aurélie Goyenvalle⁷, Stéphanie Debette^{1,8}, Sophie Dupuis-Girod^{3,9}, David-Alexandre Trégouët^{1*}

¹ Univ. Bordeaux, INSERM, Bordeaux Population Health Research Center, UMR 1219, F-33000 Bordeaux, France

² Univ. Bordeaux, INSERM, Biology of Cardiovascular Diseases, U1034, F-33600 Pessac, France

³ Hospices Civils de Lyon, French National HHT Reference Center and Genetics department, Hôpital Femme-Mère-Enfant, 69677, Bron, France

⁴ Service de Médecine Interne et Immunologie Clinique, Centre de compétence maladie de Rendu-Osler, Centre Hospitalo-Universitaire Dijon Bourgogne, Dijon, France

⁵ Université de Bourgogne, INSERM, EFS BFC, UMR1098, RIGHT Interactions Greffon-Hôte-Tumeur/Ingénierie Cellulaire et Génique, Dijon, France

⁶ Département de Médecine interne et Immunologie Clinique, CHRU BRABOIS, Vandoeuvre-lès-Nancy, France

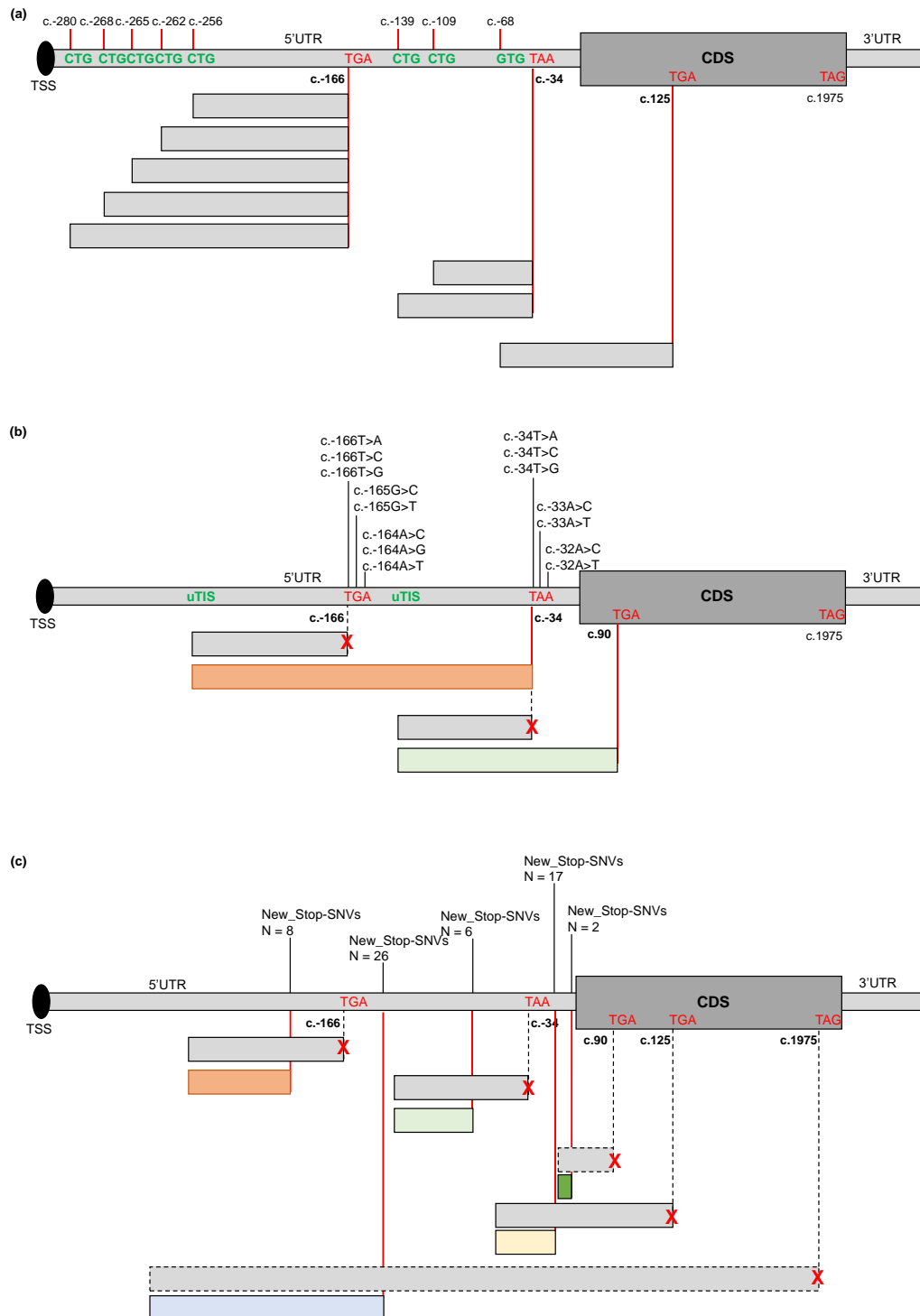
⁷ Université Paris-Saclay, UVSQ, Inserm, END-ICAP, Versailles, France

⁸ Department of Neurology, Institute for Neurodegenerative Diseases, Bordeaux University Hospital, France

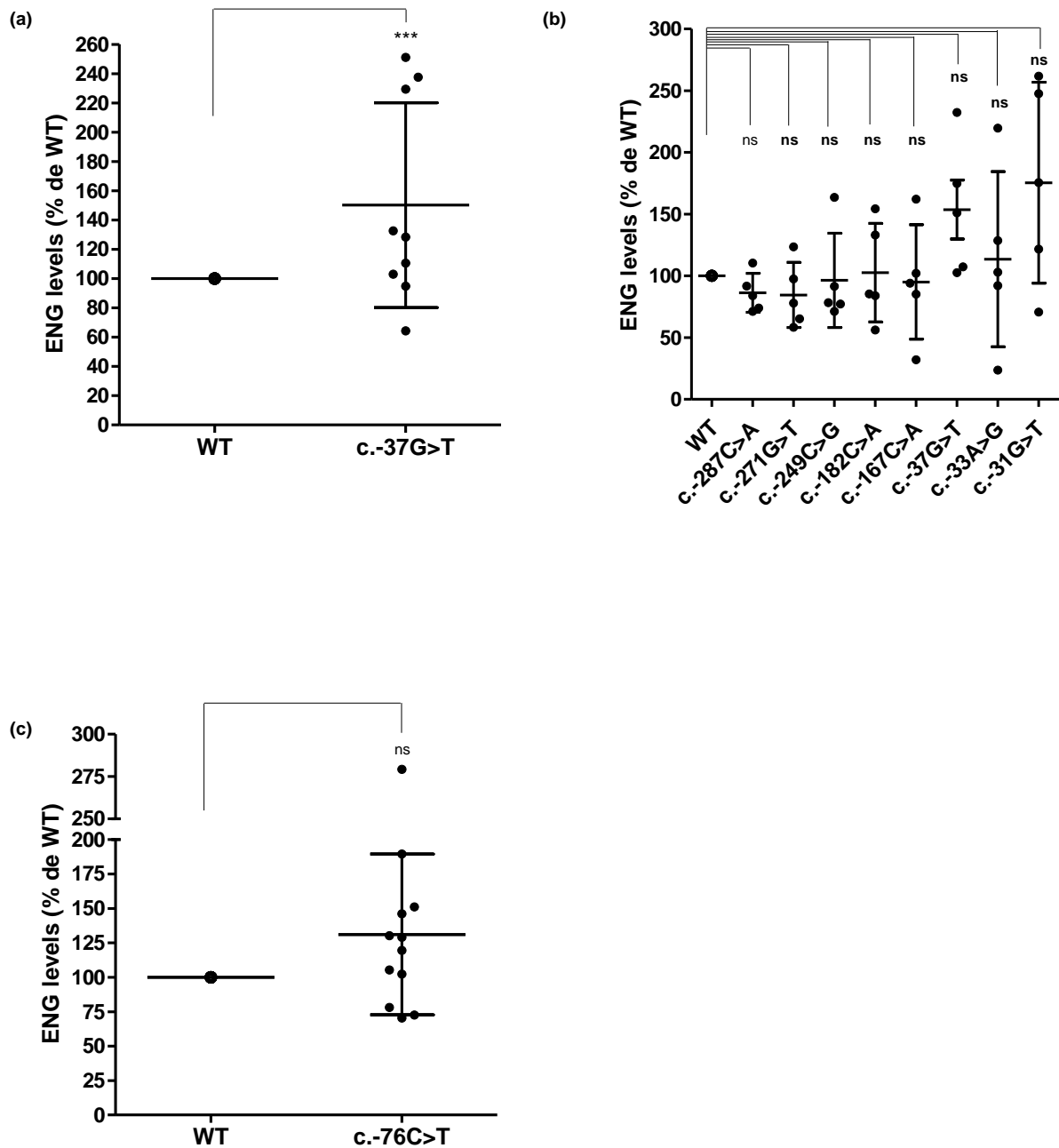
⁹ Univ. Grenoble Alpes, Inserm, CEA, Laboratory Biology of Cancer and Infection, F-38000 Grenoble, France.

These authors have contributed equally to this work

* Joint corresponding authors: omar.soukarieh@inserm.fr; david-alexandre.tregouet@u-bordeaux.fr



Supplementary Figure 1. Illustration of naturally existing upstream Open Reading Frames (upORFs) (a) uStop-deleting variants (b) and new-Stop-creating variants in the 5'UTR of ENG. The type of upstream Translation Initiation Sites (uTIS) and position of stop codons associated with the existing or the new upORFs are indicated. The red cross represents the deleted uStop replaced by another one downstream (b) or the shortening of existing upORFs by the creation of upstream stop codons. The number of variants creating stop codons in the same frame is indicated. TSS, translation start site; CDS, CoDing Sequence; SNVs, single nucleotide variants.



Supplementary Figure 2. Quantification of ENG levels in HeLa cells from Western blot (a) AND rt-qPCR (c) analyses. ENG steady-state level in HeLa cells is significantly increased with the c.-37G>T variant in comparison to the wild-type (WT) construct. For quantification, the average of the duplicate has been calculated from the quantified values and ENG levels have been normalized to the corresponding β -actin levels then to the WT (%). The two bands obtained for the Endoglin, corresponding to the more glycosylated (upper band) and less/non glycosylated (lower band) ENG monomers, were taken together for the quantification. The graph with standard error of the mean is representative of 9 independent experiments. (b and c) RNA levels of Endoglin do not variate between wild-type and variants. Normalized $2^{-\Delta\Delta CT}$ to the WT are shown. Graphs with standard error of the mean are representative of 5 (b) and 12 (c) independent experiments. ***, p-value < 10^{-3} , ns, non-significant (two-factor ANOVA followed by Tukey's multiple comparison test of variants versus WT).

Supplementary Table 3. Primers used in this study for cloning, directed mutagenesis (DM), qPCR, and sequencing. Restriction sites are underlined. Variations are in lowercase.

Technique	Variant	Primer name	Primer sequence (5'-3')
Cloning	*	ENG-Iso1-WT-BamHI-F	tagtggatccATACCACAGCCTTCATCTGCGC
Cloning	**	ENG_SacII-R	CGTGCCGCGGTCCATGCTG
Cloning	***	ENG-BIpl-R	TTCAGCTCAGCAGCAGAGGTG
DM	c.-287C>A	ENG-c.-287CA-BamHI-F	tagtggatccATACCACAGCCTTCATaTGCGCC
DM	c.-271G>T	ENG-c.-271GT-F	TGGGGCCATgACTGCTG
DM	c.-271G>T	ENG-c.-271GT-R	CAGCAGTCaTGGCCCCA
DM	c.-249C>G	ENG-c.-249CG-F	TCACTGCCATgCATTGGAGCC
DM	c.-249C>G	ENG-c.-249CG-R	GGCTCCAATGcATGGCAGTGA
DM	c.-182C>A	ENG-c.-182CA-F	CCCGCGTCCaTGTGTCCAC
DM	c.-182C>A	ENG-c.-182CA-R	GTGGACACATGGACGCGGG
DM	c.-167C>A	ENG-c.-167CA-F	TCCAATTCTCaTGACCCCTCG
DM	c.-167C>A	ENG-c.-167CA-R	CGAGGGGTCAgAGAAAGTGA
DM	c.-76C>T	ENG-c.-76CT-F	AGCCACGctGGCCCCGGT
DM	c.-76C>T	ENG-c.-76CT-R	ACCGGGCCaGCGTGGGCT
DM	c.-37G>T	ENG-c.-37GT-F	TGGACACATGATAAGGCCAGCG
DM	c.-37G>T	ENG-c.-37GT-R	CGCTGGGCCTTATCaTGTGTCCA
DM	c.-33A>G	ENG-c.-33AG-F	TGGACACAGGATgAGGCCAGCG
DM	c.-33A>G	ENG-c.-33AG-R	CGCTGGGCCTcATCCTGTGTCCA
DM	c.-31G>T	ENG-c.-31GT-F	TGGACACAGGATAaTGCAGCG
DM	c.-31G>T	ENG-c.-31GT-R	CGCTGGGCaTTATCCTGTGTCCA
qPCR	All	ENG-qPCR-Ex9-F	ACAAGTTTGTCTTGCGCAGT
qPCR	All	ENG-qPCR-Ex11-R	CTGTCCATGTTGAGGCAGTG
qPCR	All	hu_TUBA F	GATGCTGCAATAACTATGCCGAG
qPCR	All	hu_TUBA R	GAAAACCAAGAAGCCCTGAAGACGG
Sequencing	All	CMV-F	CGCAAATGGGCGGTAGGCGTG
Sequencing	***	ENG-seq1-F	CATGTCCTTCTCTGGAGTTC

*, all variants except c.-287C>A; **, c.-287C>A, c.-271G>T, c.-249C>G, c.-182C>A, c.-167C>A; ***, c.-76C>T, c.-37G>T, c.-33A>G and c.-31G>T.

Supplementary Table 8. Clinical data for *ENG* 5'UTR variants analyzed in this study.

Position¹	Created uTIS	Patients (sex)	Associated symptoms	Familial data
c.-76C>T	uCUG	Patient 1 (F)	2 DVT, stroke, E, T (lips and skin)	na
		Patient 2 (F)	2 PAVM, E (once a month), ≤ 2 T (face and hands)	Asymptomatic father (age: 72) (carrier of the variant)
c.-33A>G	uAUG	Patient 3 (F)	PAVM, recurrent E, numerous T (fingers and lips)	na

¹, variant position in the L-ENG NM_001114753.3 transcript (c.1 corresponds to the A of the main AUG); uTIS, upstream Translation Initiation site; F, female; na, non-applicable; DVT, deep venous thrombosis. E, epistaxis; T, telangiectasia; PAVM, pulmonary arteriovenous malformation.

Supplementary Table 10. Existing upstream Translation Initiation Sites (uTIS) in the 5'UTR of ENG, their Kozak sequences and predictive scores obtained with TIS-predictor (KSS scores) and PreTIS. uTIS position in the L-ENG; NM_001114753.3 transcript (c.1 corresponds to the A of the main AUG) is indicated. Nucleotides at positions -3 and +4 relative to the predicted uAUGs are in bold when corresponding to the most conserved nucleotide. The uTIS are underlined. Bolded scores are those higher than the predefined thresholds.

uTIS position	Kozak sequence	KSS score	PreTIS score
c.-280	G <u>CCCTGG</u>	0.8	NA
c.-268	G <u>GACTGC</u>	0.59	NA
c.-265	CT <u>GCTGC</u>	0.49	NA
c.-262	CT <u>GCTGT</u>	0.49	NA
c.-256	T <u>CACTGC</u>	0.48	NA
c.-139	A <u>GGCTGG</u>	0.78	0.68
c.-109	CG <u>CCTGC</u>	0.6	0.44
c.-68	CC <u>GGTGC</u>	0.61	0.64