

## Testa et al. – Supplementary Data

### Supplementary Methods

#### Variant databases

The inspected databases were the European Association of Hemophilia and Allied Disorders (EAHAD) FVIII Gene Variant Database (<https://f8-db.eahad.org/index.php>),<sup>37</sup> the CDC Hemophilia Project (CHAMP, <https://www.cdc.gov/ncbddd/hemophilia/champs.html>),<sup>38</sup> and the Human Genome Mutation Database (HGMD, <http://www.hgmd.cf.ac.uk/ac/all.php>).<sup>11</sup> Additional inspections for nucleotide missense changes in the *F8* gene were performed in the genome aggregation database (gnomAD, <https://gnomad.broadinstitute.org/>).

#### Cloning strategy

All oligonucleotides used to generate the constructs used in this study are listed in Supplementary Table S1.

The human *F8* (*hF8*) cDNA was cloned (*Sall*-*NheI* restriction) from the pBS-*hF8* construct (Supplementary Figure S1A) to the pCMV6XL4-MCS plasmid with the designed multi-cloning site (MCS) (Supplementary Figure S1B), generated by self-annealing and PCR amplification, and inserted (*NotI* restriction) as a *NotI*-MCS-*NotI* segment into an empty pCMV6XL4 vector to obtain the pCMV6XL4-*hF8* expression plasmid (Supplementary Figure S1C). For modification purposes, by taking advantage of the two *hF8* natural *KpnI* and *EcoRV* restriction sites, the *hF8* sequence was cleaved in three segments (F8A, *Sall*-*KpnI*; F8B, *KpnI*-*EcoRV*; F8C, *EcoRV*-*NheI*) that were subcloned into the pCMV6XL4-MCS plasmid (Supplementary Figure S1B) to create the pCMV6XL4-F8A, pCMV6XL4-F8B, and pCMV6XL4-F8C plasmids (Supplementary Figure S1D). The F8C segment, containing the natural *hF8* termination codon (NTC, position 2352), was then modified by two mutagenesis steps, which included removal of the NTC (NTCsup), followed by cloning (Supplementary

Figure S1E). The *F8* NTC (TGA) and the downstream *NheI* and *NotI* restriction sites were removed and the *AsiI* restriction site inserted to obtain the pCMV6XL4-C1 construct (Supplementary Figure S1E, panel i). The *MluI* restriction site, designed to introduce the *Gaussia* cDNA, as well as the surrounding *PmlI* and *NheI* sites, were inserted downstream of the F8C1 segment, also by adding an additional nucleotide downstream of the *AsiI* restriction site (GCGAT↓CGC) to maintain the reading frame, to create the pCMV6XL4-C2 construct (Supplementary Figure S1E, panel ii). The modified F8C2 segment was then cloned (*EcoRV-AsiI*) into the pCMV6XL4-hF8 vector to replace the original F8C segment and generate the pCMV6XL4-F8D plasmid containing the whole hF8 sequence and the downstream restriction sites (*AsiI*, *PmlI* and *MluI*) designed to allow production of the chimeric constructs (Supplementary Figure S1E, panel iii). The *Gaussia* sequence was then cloned (*MluI* sites) downstream of hF8 to obtain the pCMV6XL4-hF8-GL chimeric construct encoding the factor VIII (FVIII)-*Gaussia* (GL) (FVIII-GL) fusion protein (Supplementary Figure S1F). For comparison analysis, a second version of the FVIII-GL construct was produced by cloning the *Gaussia* sequence including the signal peptide (SP) downstream of hF8 (Supplementary Figure S1G). Further, a glycine-serine (GS) linker, known to improve secretion efficiency of chimeric proteins containing coagulation factors by separating fusion partners,<sup>43,44</sup> was added between hF8 and *Gaussia* sequences by self-annealing oligonucleotides with flanking *PmlI* restriction sites to obtain the pCMV6XL4-hF8-linker-GL expression plasmid (Supplementary Figure S1H). Finally, the pCMV6XL4-hF8-Stop-GL plasmid (Supplementary Figure S1I), designed as negative control, was produced by re-introducing the hF8 NTC between hF8 and *Gaussia*.

Nonsense/missense nucleotide changes were introduced in the pCMVXL4-F8A, pCMVXL4-F8B or pCMVXL4-F8C2 plasmids by site-directed mutagenesis, and then sub-cloned into the pCMV6XL4-hF8-GL construct, through *SalI-KpnI* (F8A), *KpnI-EcoRV* (F8B) or *EcoRV-*

AsiSI (F8C2) restriction sites, by replacing the corresponding wild-type segment. All constructs were validated by sequencing.

### **Optimization of the luciferase-based expression system**

The Gaussia-based assay provides a highly sensitive system for detecting intracellular and secreted luciferase activity by oxidation of the coelenterazine substrate to the coelenteramide product. The light output of the FVIII-GL fusion protein in the assay is related to the level of readthrough over the PTC inserted in the F8-coding region (Supplementary Figure S2).

To optimize the luciferase-based system we first compared expression levels of the FVIII-GL fusion protein containing the sequence encoding mature Gaussia luciferase, with or without the signal peptide (SP), which drives targeting to endoplasmic reticulum and the subsequent intracellular cleavage. As expected, the FVIII-GL chimera with the SP produced higher luciferase activity and a remarkable dispersion of replicate values as compared to the FVIII-GL protein containing only mature Gaussia (Supplementary Figure S3A), which was selected as fusion partner for this study. We then evaluated expression levels at different time points of FVIII-GL, which showed time-dependent secreted luciferase activity (Supplementary Figure S3B). Subsequently, based on the knowledge of the effects of linker sequences on fusion protein features,<sup>43,44</sup> we compared the expression at 48 hours of FVIII-GL, in which FVIII and Gaussia are directly fused, with that of FVIII-linker-GL, containing an interposed glycine-serine linker. The secretion efficiency of the two chimeric proteins was similar (Supplementary Figure S3C), which prompted us to select the direct FVIII-GL fusion protein as scaffold for the insertion of the designed nonsense/missense changes. To this purpose, and based on our hypothesis, we also evaluated expression levels at different time points of the FVIII-GL scaffold bearing two paradigmatic *F8* PTCs differing in features and localization. These studies (Supplementary Figure S3D), where fusion proteins showed a

time-dependent expression, allowed us to select the time point at 48 hours to magnify signal detection for those *F8* PTCs associated with low or very low secreted protein levels, as in the case of Arg15\*. Aware of potential background confounding effects for detection of very low levels, as further controls we also produced FVIII-GL constructs containing the *F8* natural termination codon, two or three stop codons between FVIII and Gaussia. We observed an inverse relationship between the levels of detected readthrough and the number of stop codons (Supplementary Figure S3E), which validated our expression system and indicated a very low background useful to detect very low levels.

## Supplementary Table S1.

Oligonucleotides used to design the FVIII-GL constructs and introduce nonsense and missense changes in the *hF8* sequence.

Oligonucleotide	Use	Sequence (5'-3')
MCS-F	multi-cloning site generated by self-annealing and cloning ( <i>NotI</i> ) in pCMV6XL4	tatatGCGGCCGC GTCGACGAATTCGGATTCGGT ACCAAGCTTAGATCTGATATC
MCS-R	multi-cloning site generated by self-annealing and cloning ( <i>NotI</i> ) in pCMV6XL4	atataGCGGCCGC AAGCTAGCCTCGAGTCTAGAG ATATCAGATCTAAGCTTGGTAC
MutNTCdel-insAsiSI	Deletion of <i>hF8</i> TGA and of downstream <i>NheI</i> and <i>NotI</i> sites, and insertion of <i>AsiSI</i> site	GCACAGGACCTCTACGCGATCGCTCATAGCTG TTTCCTG
MutF8C2MluPmlNhe	Insertion of <i>MluI</i> site (Gaussia cloning) and flanking <i>PmlI</i> and <i>NheI</i> restriction sites	GACCTCTACGCGATCGCTCACGTGACGCGTGC TAGCTCATAGCTGTTTCCTG
Gaussia-Mlu-F	Amplification of cDNA sequence encoding mature Gaussia for cloning ( <i>MluI</i> ) downstream of <i>hF8</i>	aaACGCGTAAGCCCACCGAGAACAACGAAG
GaussiaSP-Mlu-F	Amplification of cDNA sequence including Gaussia signal peptide (SP) for cloning ( <i>MluI</i> ) downstream of <i>hF8</i>	aaACGCGTATGGGAGTCAAAGTTCTGTTTGCCC
Gaussia-Mlu-R	Amplification of cDNA Gaussia terminal sequence for cloning ( <i>MluI</i> ) downstream of <i>hF8</i>	aaACGCGTTAGTCACCACCGGCCCC
GS-Linker-F	Insertion ( <i>PmlI</i> ) of the glycine-serine (GSSGSGSGSSGSSG) linker between <i>hF8</i> and Gaussia	aaaGTGgggtcttctgggggtctgggggtcttctgggtcttctggg gggCAC
GS-Linker-R	Insertion ( <i>PmlI</i> ) of the glycine-serine (GSSGSGSGSSGSSG) linker between <i>hF8</i> and Gaussia	aaaGTGccccccagaagaccagaagaccccccagaccccc agaagaccCAC
MutInsF8NTC2352	Insertion of <i>hF8</i> natural stop codon between <i>hF8</i> and Gaussia	CAGGACCTCTACTGAGCGATCGCTCAC
MutR15X	R15X nonsense variant	CTGTGCCTTTTGTGATTCTGCTTTAGTG
MutW157X	W157X nonsense variant	CATACATATGTCTGACAGGTCCTGAAAGAG
MutS229X	S229X nonsense variant	GAAAAGTTGGCACTGAGAAACAAAGAAC
MutW247X	W247X nonsense variant	CATCTGCTCGGGCCTGACCTAAAATGCACAC
MutR355X	R355X nonsense variant	GGAACCCCAACTATGAATGAAAATAATG
MutR446X	R446X nonsense variant	GTACAAAAAAGTCTGATTTATGGCATAAC
MutS507X	S507X nonsense variant	CGTCCTTTGTATTGAAGGAGATTACC

MutR602X	R602X nonsense variant	GTATTTGATGAGAACTGAAGCTGGTACCAAG
MutQ611X	Q611X nonsense variant	CTCACAGAGAATATATAACGCTTTCTCCC
MutQ645X	Q645X nonsense variant	GTTTTTGATAGTTTGTAGTTGTCAGTTTGTTTG
MutQ772X	Q772X nonsense variant	CCTAGCACTAGGTAAAAGCAATTTAATG
MutW791X	W791X nonsense variant	GACTGACCCTTGATTGCACACAGAAC
MutR814X	R814X nonsense variant	GTTGATGCTCTTGTGACAGAGTCCTACTC
MutQ815X	Q815X nonsense variant	GATGCTCTTGCGATAGAGTCCTACTCCAC
MutS872X	S872X nonsense variant	GTATTTACCCCTGAGTGAGGCCTCCAATTAAG
MutL878X	L878X nonsense variant	CTCCAATTAAGATGAAATGAGAAACTG
MutW979Xtga	W979X(tga) nonsense variant	CAAGAAAGTTCATGAGGAAAAAATGTATC
MutW979Xtag	W979X(tag) nonsense variant	CTCCAATTAAGATGAAATGAGAAACTG
MutW1048X	W1048X nonsense variant	GTCCATCAGTCTAGCAAATATATTAG
MutK1059X	K1059X nonsense variant	GTGACACTGAGTTTTAAAAAGTGACACC
MutQ1098X	Q1098X nonsense variant	GAAATGGTCCAATAGAAAAAGAGGGC
MutQ1109X	Q1109X nonsense variant	CCACCAGATGCATAAAATCCAGATATG
MutW1127X	W1127X nonsense variant	GAATCAGCAAGGTGAATACAAAGGACTC
MutE1165X	E1165X nonsense variant	GAATTTCTTGTCTTAGAAAAACAAAG
MutQ1232X	Q1232X nonsense variant	GTAGTTTTGCCTTAGATACATACAGTG
MutK1289X	K1289X nonsense variant	CAGCTCATTTCTCATAAAAAGGGGAGGAAG
MutQ1336X	Q1336X nonsense variant	GAGCTTTGAAATAATTCAGACTCCC
MutQ1386X	Q1386X nonsense variant	CCATTACTTAGTCTCCCTTATC
MutE1597X	E1597X nonsense variant	CCAAAAGAATAGTGGAATCC
MutE1634X	E1634X nonsense variant	TAGCAGCAATAAATTAGGGACAAAATAAGCC
MutW1645X	W1645X nonsense variant	GAAGTCACCTGAGCAAAGCAAGGTAG
MutQ1705X	Q1705X nonsense variant	GATGAGGATGAAAATTAGAGCCCCGCAGC
MutR1715X	R1715X nonsense variant	CTTCAAAGAAAACATGACACTATTTTATTG
MutW1726X(tga)	W1726X(tga) nonsense variant	CAGTGGAGAGGCTCTGAGATTATGGGATGAG
MutW1726X(tag)	W1726X(tag) nonsense variant	CAGTGGAGAGGCTCTAGGATTATGGGATGAG
MutW1836X	W1836X nonsense variant	CAAACCTTACTTTTGAAAAGTGCAACATC
MutR1960X	R1960X nonsense variant	GATCAAAGGATTTGATGGTATCTGCTC
MutR1985X	R1985X nonsense variant	CATGTGTTCACTGTATGAAAAAAGAGGAG
MutG2076X	G2076X nonsense variant	GACTTCATTATTCCTGATCAATCAATGCCTG
MutW2081X	W2081X nonsense variant	CAATCAATGCCTGAAGCACCAAGGAGC
MutR2135X	R2135X nonsense variant	GTGGCAGACTTATTGAGGAAATTCCAC

MutR2166X	R2166X nonsense variant	CTCCAATTATTGCTTGATACATCCGTTTGC
MutL2197X	L2197X nonsense variant	GTTGCAGCATGCCATAGGGAATGGAGAG
MutR2228X	R2228X nonsense variant	CTCCTTCAAAAGCTTGACTTCACCTCCAAG
MutE2247X	E2247X nonsense variant	GTGAATAATCCAAAATAGTGGCTGCAAGTG
MutQ2303X	Q2303X nonsense variant	CAAAGTAAAGGTTTTTTAGGGAAATCAAGAC
MutR2326X	R2326X nonsense variant	GACTCGCTACCTTTGAATTCACCCCC
MutR446W	R446W missense variant	AAGTCTGGTTTATGGCATACT
MutR814W	R814W missense variant	TCTTGTGGCAGAGTCCTA
MutK1289Y	K1289Y missense variant	ATTTCTCATATAAAGGGGAGG
MutK1289Q	K1289Q missense variant	TTTCTCACAAAAGGGGAGGA
MutW1726Y	W1726Y missense variant	AGGCTCTATGATTATGGGAT
MutR1985W	R1985W missense variant	ACTGTATGGAAAAAGAGGAG
MutR2135W	R2135W missense variant	GACTTATTGGGGAAATTCCA

For oligonucleotides used in PCR, forward (F) and reverse (R) sequences are indicated.

For oligonucleotides used in site-directed mutagenesis (indicated as Mut), in which forward and reverse primers are designed to be complementary, only the forward sequence is indicated.

## Supplementary Table S2.

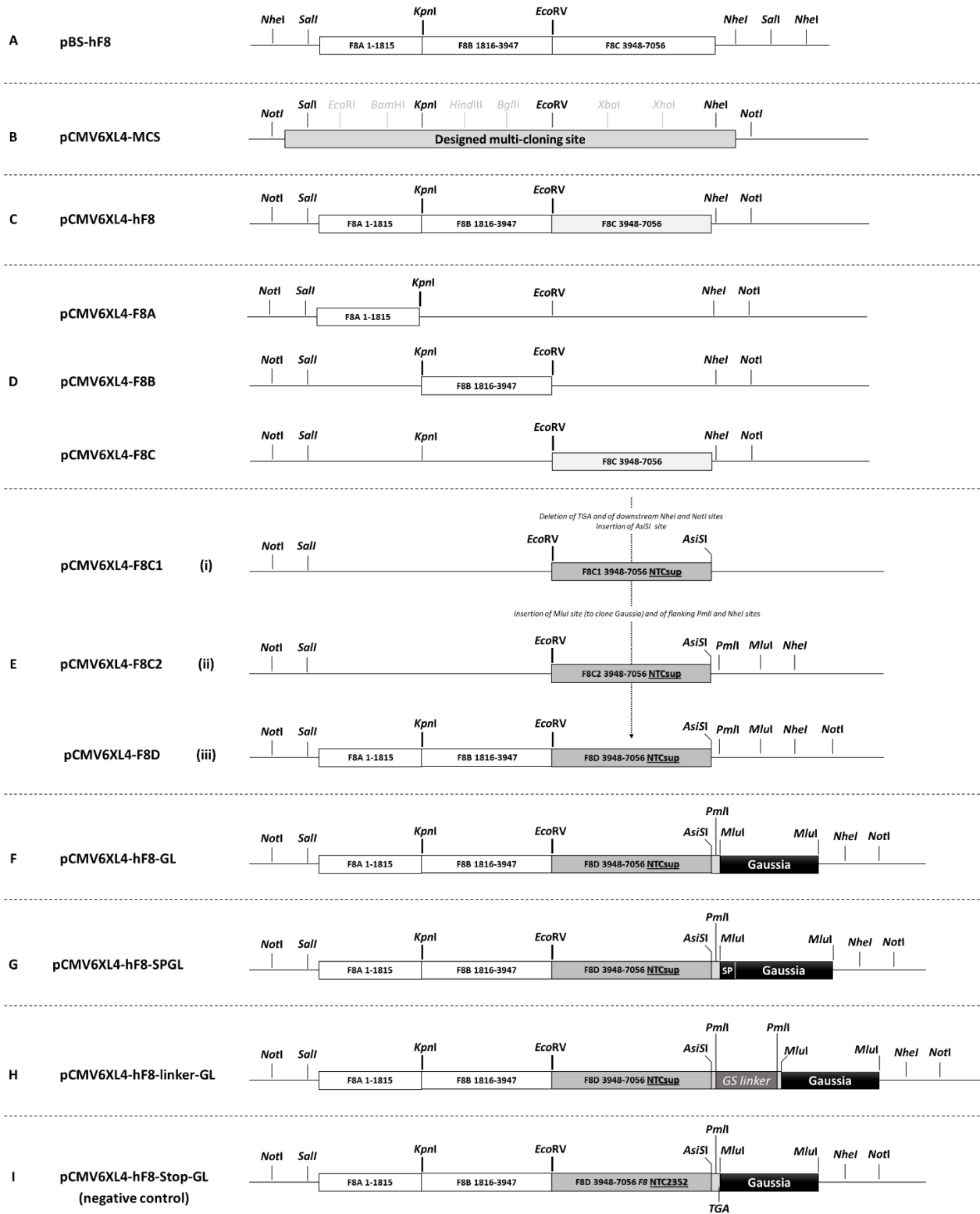
### Comparison of secreted levels of HA plasma and recombinantly expressed FVIII variants.

<i>Patients' PTCs</i>				
<b>Protein variant (levels)</b>	p.Arg446* (0.2±0.04%)	p.Trp1726* (0.7±0.2%)	p.Arg1985* (0.4±0.1%)	p.Arg2135* (0.3±0.1%)
p.Arg814* (1.5±0.2%)	<0.0001	0.0019	0.0001	<0.0001
p.Lys1289* (2.1±0.1%)	<0.0001	<0.0001	<0.0001	<0.0001
<i>Recombinant PTCs</i>				
<b>Protein variant (levels)</b>	Arg446* (1.2±0.3%)	Trp1726* (0.4±0.1%)	Arg1985* (0.6±0.2%)	Arg2135* (0.4±0.1%)
Arg814* (1.9±0.3%)	0.0148	0.0002	0.0008	0.0002
Lys1289* (2.1±0.2%)	0.0009	<0.0001	<0.0001	<0.0001
<i>Missense variants</i>				
<b>Protein variant (levels)</b>	Arg446Trp (35.1±6.1%)	Trp1726Tyr (4.8±1.1%)	Arg1985Trp (11.2±4.9%)	Arg2135Trp (15.5±3.2%)
Arg814Trp (57.1±8.7%)	0.006	<0.0001	<0.0001	<0.0001
Lys1289Tyr (63.1±6.35)	0.0007	<0.0001	<0.0001	<0.0001
Lys1289Gln (99.8±16.9%)	0.0004	<0.0001	<0.0001	<0.0001

The Table compares antigen levels of i) nonsense variants measured in HA patients' plasma (patients' PTCs), ii) nonsense variants expressed through our luciferase-based system (Recombinant PTCs), and iii) missense variants predicted to arise from readthrough of patients' PTCs. The secreted levels (indicated in parenthesis) of variants in each column are compared with secreted levels of variants in each row, with the relative p value reported for each comparison.

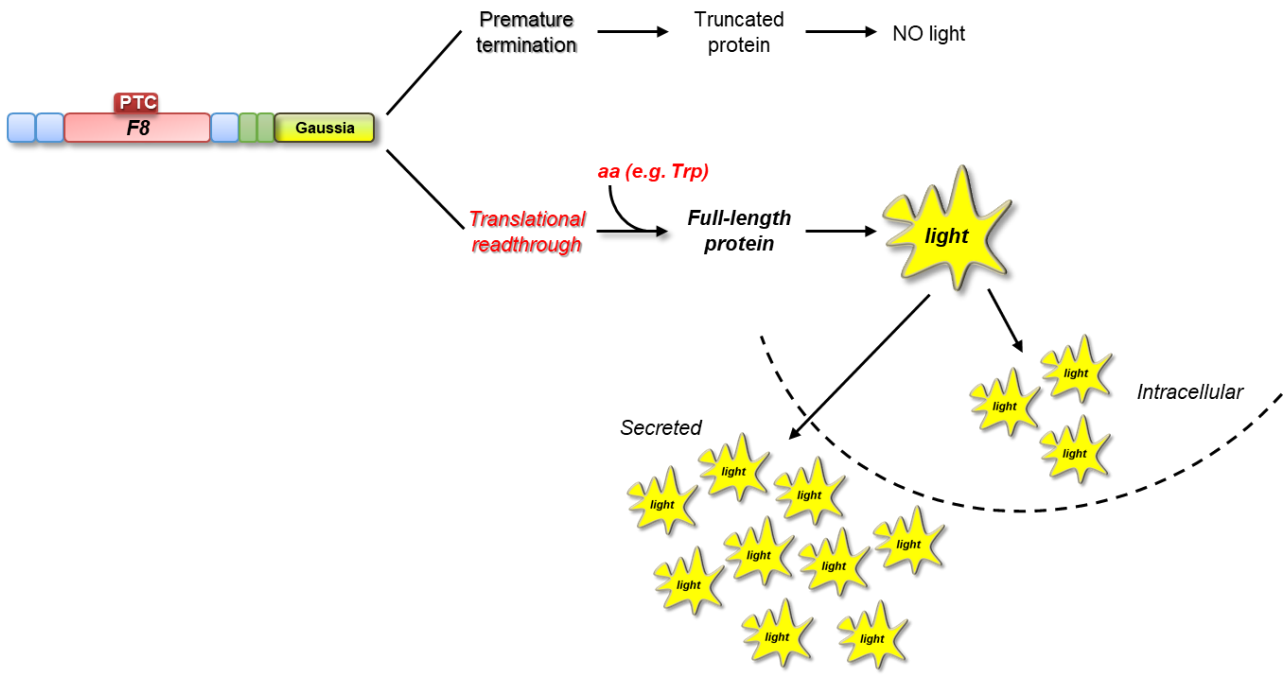
Protein levels are expressed as percentage of the reference (pooled normal plasma, PNP, for patients' plasma) or wild-type FVIII-GL for expression studies (n=4 replicates) and are indicated as mean ± standard deviation.





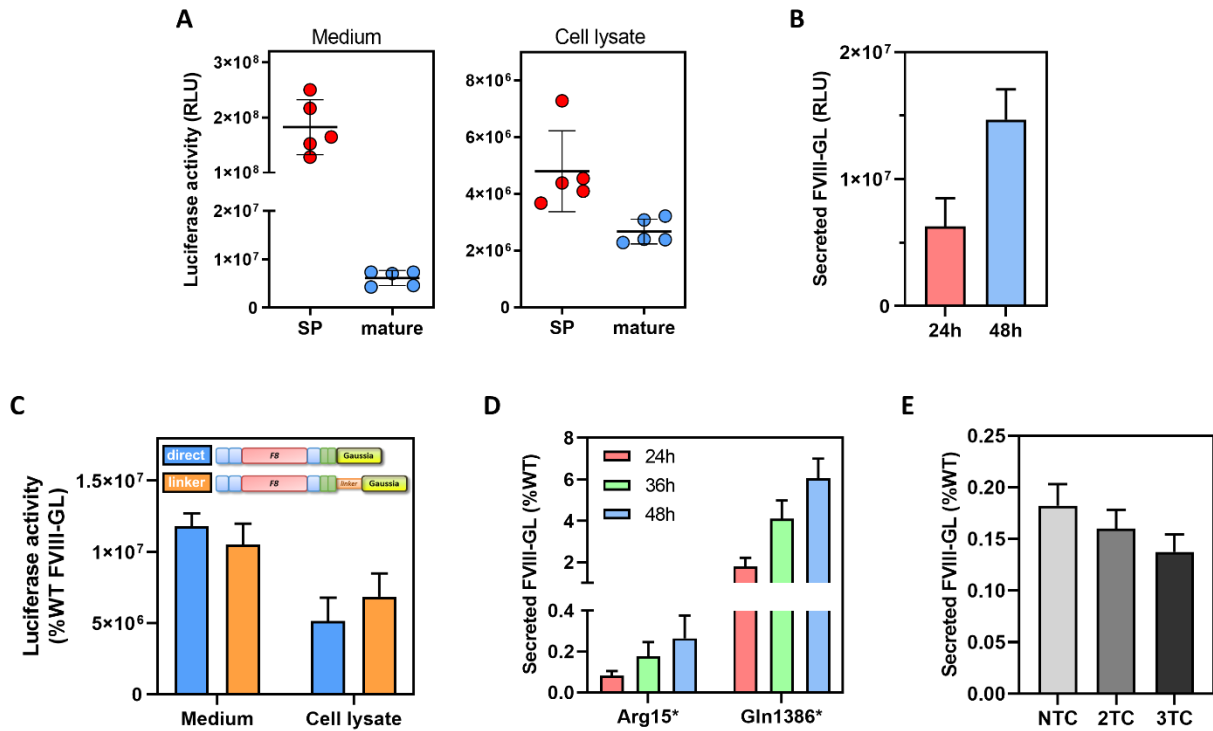
## Supplementary Figure S1. Cloning strategy to generate the FVIII-GL chimera.

Cloning/mutagenesis steps to generate the chimeric FVIII-GL constructs. Full description is provided in Supplemental Methods.



**Supplementary Figure S2. The FVIII-GL fusion protein and the luciferase-based assay.**

Schematic representation of the optimized fusion protein joining FVIII and the Gaussia (GL) luciferase (FVIII-GL). In the scheme is depicted the expected protein output resulting from premature termination of translation (upper part) or translational readthrough leading to insertion of a subset of amino acids resulting in the production of full-length proteins (middle part), in turn remaining in the intracellular compartment or being secreted (lower part).



**Supplementary Figure S3. Optimization of the luciferase-based expression system.**

A) Comparison of expression levels (24 hours) of the FVIII-GL fusion protein containing the Gaussia luciferase sequence with (SP) or without (mature) the signal peptide.

B) Expression studies on the FVIII-GL fusion protein at different time points.

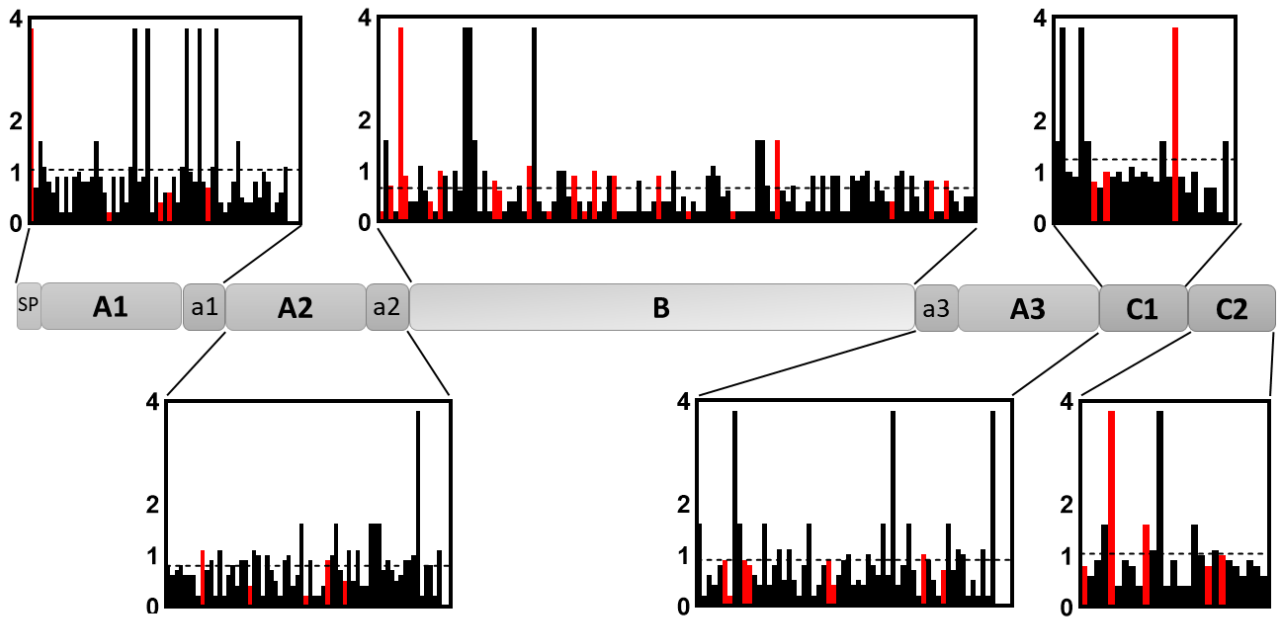
C) Comparison of FVIII-GL (direct fusion) with FVIII-linker-GL (interposed GS linker) fusion proteins at the selected 48-hour time point.

Top, schematic representation of fusion proteins directly fuses or through the GS linker.

D) Comparison of secreted levels of the selected FVIII-GL scaffold bearing the Arg15\* and Gln1386\* *F8* PTCs.

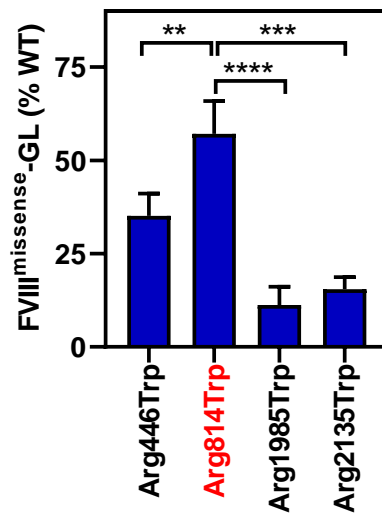
E) Evaluation of residual expression levels from the FVIII-GL construct bearing *F8* natural termination codon (NTC), two (2TC) or three (3TC) termination codons between FVIII and Gaussia.

Results (mean  $\pm$  standard deviation) are expressed as Relative Luminescence Units (RLU) or percentage of WT FVIII-GL (panels B-E, n=4).



**Supplementary Figure S4. Predicted readthrough score of *F8* PTCs.**

Schematic representation of FVIII domain organization and predicted readthrough score (according to Manuvakhova et al.<sup>22</sup>) of *F8* PTCs indicated for each domain. The mean readthrough score for each domain is indicated as dotted line. Red bars represent PTC variants expressed in the FVIII-GL chimera.



**Supplementary Figure S5. Missense variants bearing the non-conservative arginine-to-tryptophan substitution arising from readthrough of TGA PTCs.**

Comparison of luciferase activity levels of missense variants bearing the arginine-to-tryptophan substitution as predicted for readthrough at TGA PTCs (Arg446\*, Arg814\*, Arg1985\* and Arg2135\*). The B-domain Arg814\* variant is highlighted in red.

Results (mean  $\pm$  standard deviation; n=4) are expressed as percentage of WT FVIII-GL. \*\*, p<0.01; \*\*\*, p<0.001; \*\*\*\*, p<0.0001.