Supplemental Data

Methods

Copy number variation (CNV) identification and breakpoint mapping

Index cases were screened for presence of CNV using multiplex ligation-dependent probe amplification (MLPA) and data analyzed using the Manchester National Genetics Reference Laboratory methodology as previously described.[1] Data for ~120 single nucleotide variants, genotyped using the MassARRAY™ matrix-assisted laser desorption ionization–time of flight system (Sequenom Inc., San Diego, CA, USA) as described,[2] was used to reduce maximal CNV size for the exon 32-34 and 33-34 deletions (Supplementary Tables 1 and 2). Long-range PCR (primer sequences and reaction conditions available on request) utilizing either One*Taq*® Hot Start DNA polymerase (New England Biolabs (UK) Ltd., Hitchin) or the Expand Long Template PCR System (Roche Diagnostics Ltd., Burgess Hill, UK) followed by direct DNA sequence analysis was used to determine CNV breakpoints.

Generation of von Willebrand factor (VWF) p.Pro2722Ala expression plasmid

A QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies LDA UK Ltd., Stockport) was used to generate the *VWF* variant c.8164C>G (p.Pro2722Ala) in the pCI-neo-VWF wild-type (WT) expression plasmid. Successful mutagenesis was confirmed via DNA sequence analysis.

***In vitro* expression of recombinant p.Pro2722Ala VWF**

HEK293T cells were maintained in Dulbecco's modified Eagle medium containing GlutaMAX™ supplemented with 10% v/v fetal bovine serum (Thermo Fisher Scientific, Paisley, UK). Cell transfections were conducted in 9.6 cm^2 wells using Lipofectamine™ LTX reagent (Thermo Fisher). To summarize, 6 µg WT or p.Pro2722Ala VWF expression plasmid (or 3 µg WT and p.Pro2722Ala for 50:50 co-transfections) was transfected and left to incubate at 37°C / 5% CO_2 for 24 h before replacement of the culture media. Following a further 48 h incubation, culture media was collected and the cells lysed in 1 mL 1x passive lysis buffer (Promega UK Ltd., Southampton). Measurement of VWF:Ag in collected media and lysates was performed using a matched-pair antibody human VWF ELISA (Enzyme Research Laboratories Ltd., Swansea, UK) according to the manufacturer's specifications.

PCR analysis of large deletion familial segregation

For the deletions of exon 3, exons 32-34 and exons 33-34, multiplex PCR assays (primer sequences and reaction conditions available on request) utilizing One*Taq*® Hot Start DNA polymerase (New England Biolabs) were designed to genotype each deletion and corresponding WT allele. For the deletion of exons 4-5, PCR utilizing the deletion-specific primers previously described[3] was performed using Phusion High-Fidelity DNA polymerase (Thermo Fisher).

***In silico* analysis of large deletion breakpoints**

RepeatMasker analysis (undertaken via the University of California Santa Cruz human genome browser; http://genome-euro.ucsc.edu/index.html, accessed May 2018) was used to determine the presence of any repetitive elements flanking deletion breakpoint junctions. QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/analyze.php, accessed May 2018) was used to identify G-quadruplex forming sequences that can predispose DNA to form secondary structures causing initial replication forks to collapse or stall.[4] The DNA Pattern Find feature of the Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/dna_pattern.html, accessed May 2018) was used to identify motifs known to be associated with DNA recombination.[5]

Table S1. Single nucleotide variant genotype data used to reduce the maximal size of the exon 32-34 deletion

Variant	Genotype in P9F3 family member						
	I1 (UFM)	I2 (IC)	I3 (AFM)	II1 (AFM)	II2 (UFM)	III3 (UFM)	III4 (UFM)
rs542993 (c.5842+352G>A)	G/G	ND	G/G	G/G	G/G	G/G	G/G
rs216814 (c.5842+2523C>A)	A/A	C/A	C/A	A/A	C/A	A/A	C/A
rs216812 (c.5842+3572C>T)	T/T	ND	C/C	C/T	C/T	C/T	C/C
rs216811 (c.5842+5509C>T)	T/T	ND	C/C	C/T	C/T	C/T	C/C
rs216809 (c.5842+6500G>T)	T/T	ND	G/G	G/T	G/T	G/T	G/G
rs216801 (c.5843-3125C>T)	T/T	ND	C/C	C/T	C/T	C/T	C/C
rs216905 (c.5843-867G>A)	G/G	ND	G/G	G/G	G/G	G/A	G/G
rs216904 (c.5843-630A>G)	G/G	ND	A/A	A/G	A/G	A/G	A/A
rs216903 (c.5843-111A>G)	G/G	ND	A/A	A/G	A/G	A/G	A/A

AFM, affected family member; IC, index case; ND, not determined; UFM, unaffected family member.

Table S2. Single nucleotide variant genotype data used to reduce the maximal size of the exon 33-34 deletion

Variant	Genotype in P6F1 family member						
	II1 (IC)	**II2 (UFM)**	**II3 (UFM)**	**II4 (UFM)**	**II5 (UFM)**	**III1 (UFM)**	**III2 (AFM)**
rs216812 (c.5842+3572C>T)	C/C	C/T	C/T	C/T	C/T	C/T	C/T
rs216811 (c.5842+5509C>T)	C/C	C/T	C/T	C/T	C/T	C/T	C/T
rs216809 (c.5842+6500G>T)	G/T	G/T	G/T	G/T	G/T	G/T	T/T
rs216801 (c.5843-3125C>T)	C/T	T/T	C/T	C/T	C/T	C/T	T/T
rs216905 (c.5843-867G>A)	G/G	G/A	G/A	G/A	G/A	G/A	G/A
rs216904 (c.5843-630A>G)	A/G	G/G	A/G	A/G	A/G	A/G	G/G
rs216903 (c.5843-111A>G)	G/G	G/G	A/G	G/G	G/G	G/G	G/G

AFM, affected family member; IC, index case; UFM, unaffected family member.

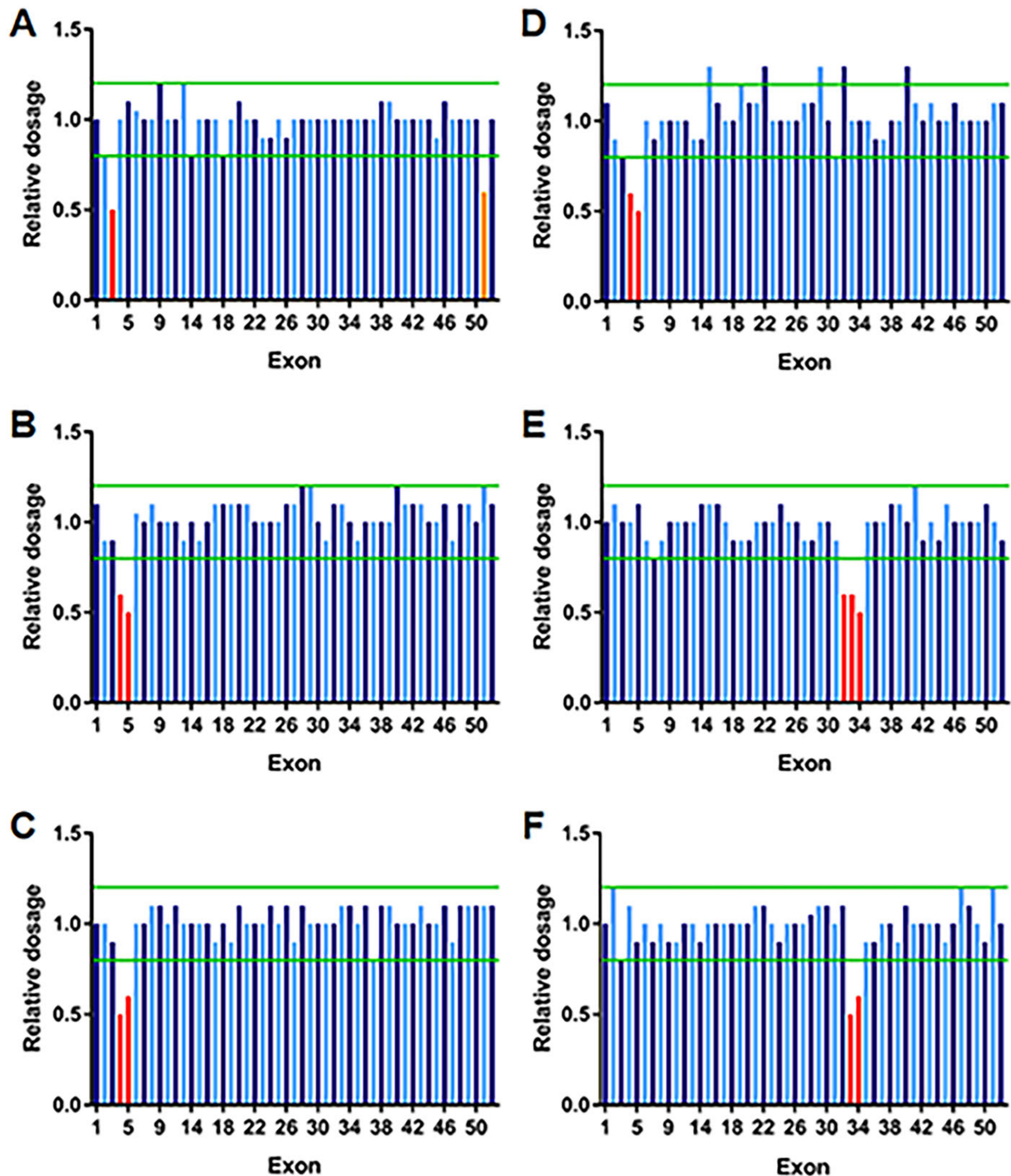

Figure S1. Identification of large heterozygous deletions in IC following dosage analysis by MLPA. (A) ex3del in IC P9F11II1 (reduced exon 51 signal due to presence of variant c.8164C>G at probe binding site). (B) ex4-5del in IC P1F5III1. (C) ex4-5del in IC P6F10II1. (D) ex4-5del in IC P12F9III1. (E) ex32-34del in IC P9F3I2. (F) ex33-34del in IC P6F1II1. Relative dosage calculated based on 23 genomic reference probes and in comparison with healthy control individuals (normal dosage range 0.8-1.2). There is no probe for exon 12.

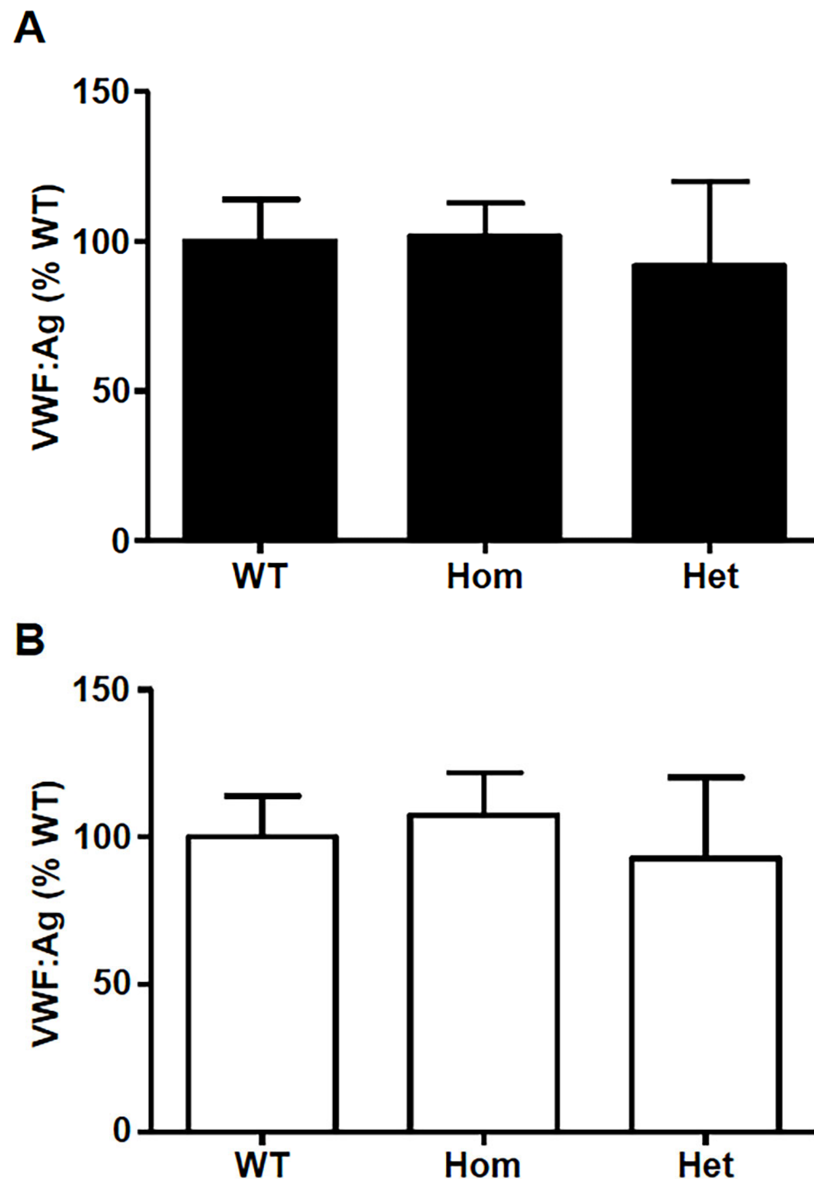

Figure S2. ***In vitro*** **expression of recombinant p.Pro2722Ala VWF in supernatants (A) and lysates (B).** No significant differences in VWF:Ag levels were observed (one-way analysis of variance). Mean values for n=3 triplicate measurements are shown (bars indicate standard deviation). Het, heterozygous expression; Hom, homozygous expression; WT, wild-type expression.

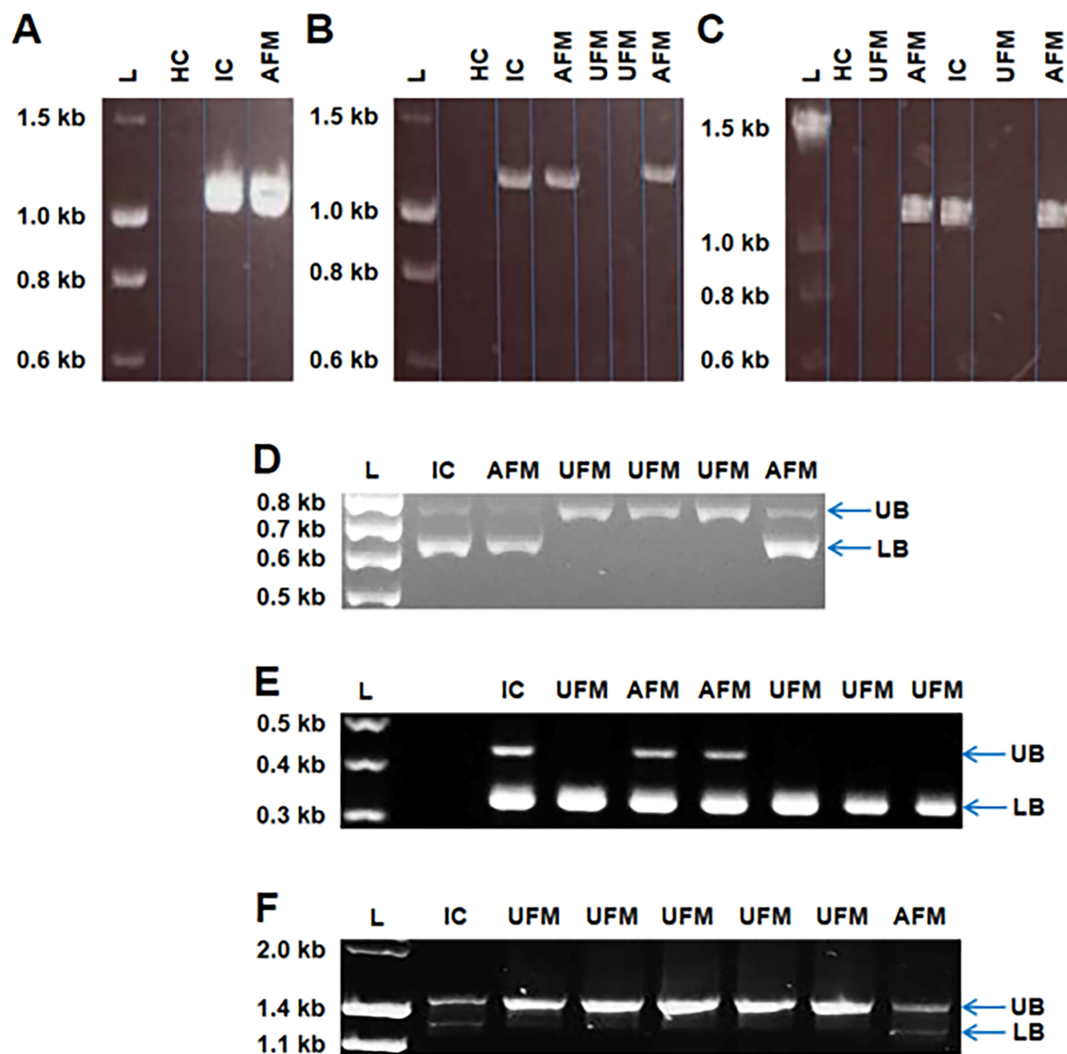

Figure S3. Familial segregation analysis of identified large deletions. (A-C) Exon 4-5 deletion in pedigrees P1F5, P6F10 and P12F9 respectively (deletion-specific 1084 bp band shown). (D) Exon 3 deletion in pedigree P9F11 (upper band (UB), 725 bp wild-type; lower band (LB), 613 bp deletion-specific). (E) Exon 32-34 deletion in pedigree P9F3 (UB, 427 bp deletion-specific; LB, 329 bp wild-type). (F) Exon 33-34 deletion in pedigree P6F1 (UB, 1497 bp wild-type; LB, 1210 bp deletion-specific). AFM, affected family member; HC, healthy control individual; IC, index case; L, size marker (A-C, HyperLadder 1 kb; D-E, HyperLadder 100 bp; F, Lambda Hind III / phiX Hae III marker); UFM, unaffected family member.

A

```
Int 2  ATGTTGGGCAGGCTGGTCTTGAATTCCTGACCTCAGGTGATCTGCCCACC
   IC  ATGTTGGGCAGGCTGGTCTTGAATTCCTGACCTCAGGTGATCTGCCCACC

Int 3  ATATTGGCCAGGCTGGTCTCGAActcctGACCTCAGGTGATCCACCCGCC

Int 2  TCGGCCTCCCAAAGTGCTGGGATTATAGGCGTGAGCCACTGTGCCCGGCC
   IC  TCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCATGCCTGGTC
Int 3  TCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCATGCCTGGTC

Int 2  AGCTTCTTTGTTTTCTTCTGTATGCCCCAAATTTTTAATAATGGACATGA

   IC  TGAAGATGTTTACAGAGTGTGTGAGGAGCCCTAGGGAAGGAGCTTTGA
Int 3  TGAAGATGTTTACAGAGTGTGTGaggagCCCTAGGGAaggagCTTTGA
```

B

```
AluSx1  ATGTTGGGCAGGCTGGTCTTGAATTCCTGACCTCAGGTGATCTGCCCACC
AluSx1  ATATTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCGCC

AluSx1  TCGGCCTCCCAAAGTGCTGGGATTATAGGCGTGAGCCACTGTGCCCGGCC
AluSx1  TCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCATGCCTGGTC

AluSx1  AGCTTCTTTGTTTTCTTCTGTATGCCCCAAATTTTTAATAATGGACATGA
AluSx1  TGAAGATGTTTACAGAGTGTGTGAGGAGCCCTAGGGAAGGAGCTTTGA
```

C

```
Int 31  ATCCTATATCTTTATAACTAATTATAATCTTTAATTTctccaCATTTTCCC
    IC  ATCCTATATCTTTATAACTAATTATAATCTTTAATTTCTCCACATTTTCCC

Int 34  TCGATGTCATTTCAAAGAGAGCTTCCTAATACTTCACTGTGTACCATAGGT

Int 31  TTTGTGTCCaggagTGATCATACCT*CTATGCCCTTTCTTTCTAAAGTAGG
    IC  TTTGTGTCCAGGAGTGATCATACCT*GCGTTTAGGCTGTTTCCAGTGTTTT
Int 34  TACCCAACTGTTCTTCTATGGATGT*GCGTTTAGGCTGTTTCCAGTGTTTT

Int 31  ATTCGTGGTCTCTTACTAAAATGTCATGTCTGAGGCCAGAGTGATTCATTG

    IC  GTAATGACAAGCAATGCTGCAATGAGTAAACCTGAACATACGAACGAATTT
Int 34  GTAATGACAAGCAATGCTGCAATGAGTAAACCTGAACATACGAACGAATTT
```

```
D  Int 32 ACGCTTTTTTTGTGGCGGGGGCCTGGGTTTGTATATTTTCCCGTTACTAGA
          |||||||||||||||||||||||||||||||||||||||||||||||||||
       IC ACGCTTTTTTTGTGGCGGGGGCCTGGGTTTGTATATTTTCCCGTTACTAGA

   Int 34 TTTTTTTTTTTTTGCTATTGAGTTGTATGAGTTCCTAATATATTTTGGATA

   Int 32 TGTAAGTCAAAACCTGCATAAAGCT*ACTGTCCTTCGGGGGAATAAGTCAA
          |||||||||||||||||||||||||*
       IC TGTAAGTCAAAACCTGCATAAAGCT*TGCAAATATTTTCTTCCATTCTGTA
                                   *|||||||||||||||||||||||||
   Int 34 TTAACCCTTTATCAAATGTACGGTT*TGCAAATATTTTCTTCCATTCTGTA

   Int 32 TGCAAGTTTGCCCTTAAAGGGCAATAACTCTATGCAAGTTTTGACTTATAG

       IC GGCTGCCTTTTTATTCTGTTGATTGTTTCTTTGGCTGCACAGAAGCTTTTT
          |||||||||||||||||||||||||||||||||||||||||||||||||||
   Int 34 GGCTGCCTTTTTATTCTGTTGATTGTTTCTTTGGCTGCACAGAAGCTTTTT
```

Figure S4. *In silico* analysis of large deletion breakpoints. (A) Exon 3 deletion breakpoint. (B) Sequence flanking exon 3 deletion breakpoint located within *Alu*Sx1 short interspersed nuclear elements demonstrating 65% homology around the breakpoint junction. Deletion likely to have resulted from *Alu*-mediated homologous recombination.[4] (C) Exon 32-34 deletion breakpoint. Lack of homology and random inserted sequence (*, GGACACA) around the breakpoint junction suggests deletion resulted from non-homologous end joining.[4] (D) Exon 33-34 deletion breakpoint. Lack of homology and random inserted sequence (*, GCAGCATAAGCATAAAGC) around the breakpoint junction suggests deletion resulted from non-homologous end joining.[4] Location of G-quadruplex forming sequences (shaded yellow), deletion hotspot consensus sequences (solid underline), DNA polymerase arrest sites (lowercase) and DNA polymerase a/b hotspots (red italics) within 75bp of the deletion breakpoints are shown.

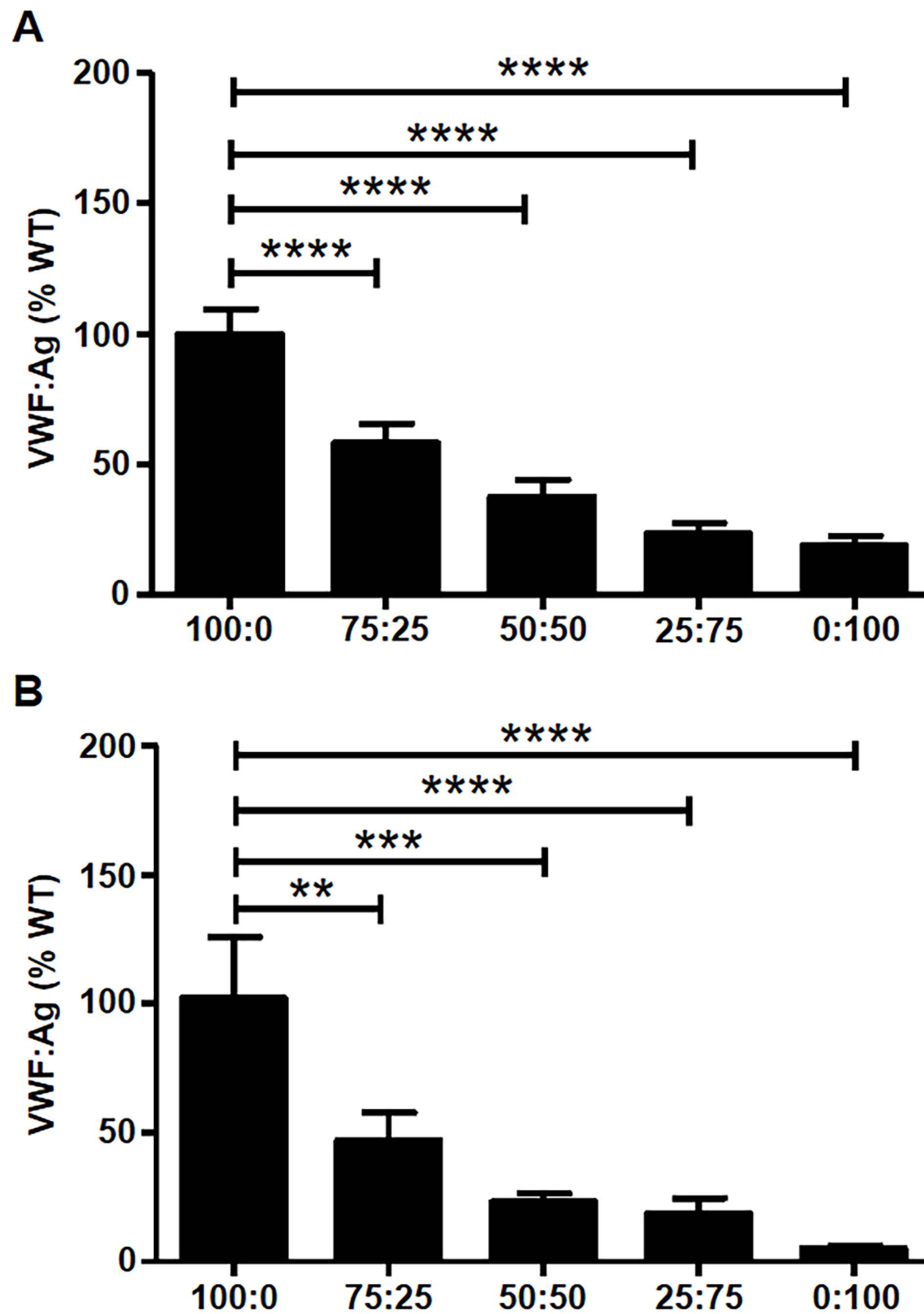

Figure S5. *In vitro* expression of rVWF3del (A) and rVWF32-34del (B) at various WT:mutant titrations. Mean values for n=3 triplicate measurements are shown (bars indicate standard deviation). **, $p<0.01$; ***, $p<0.001$; ****, $p<0.0001$ (unpaired Student *t* test).

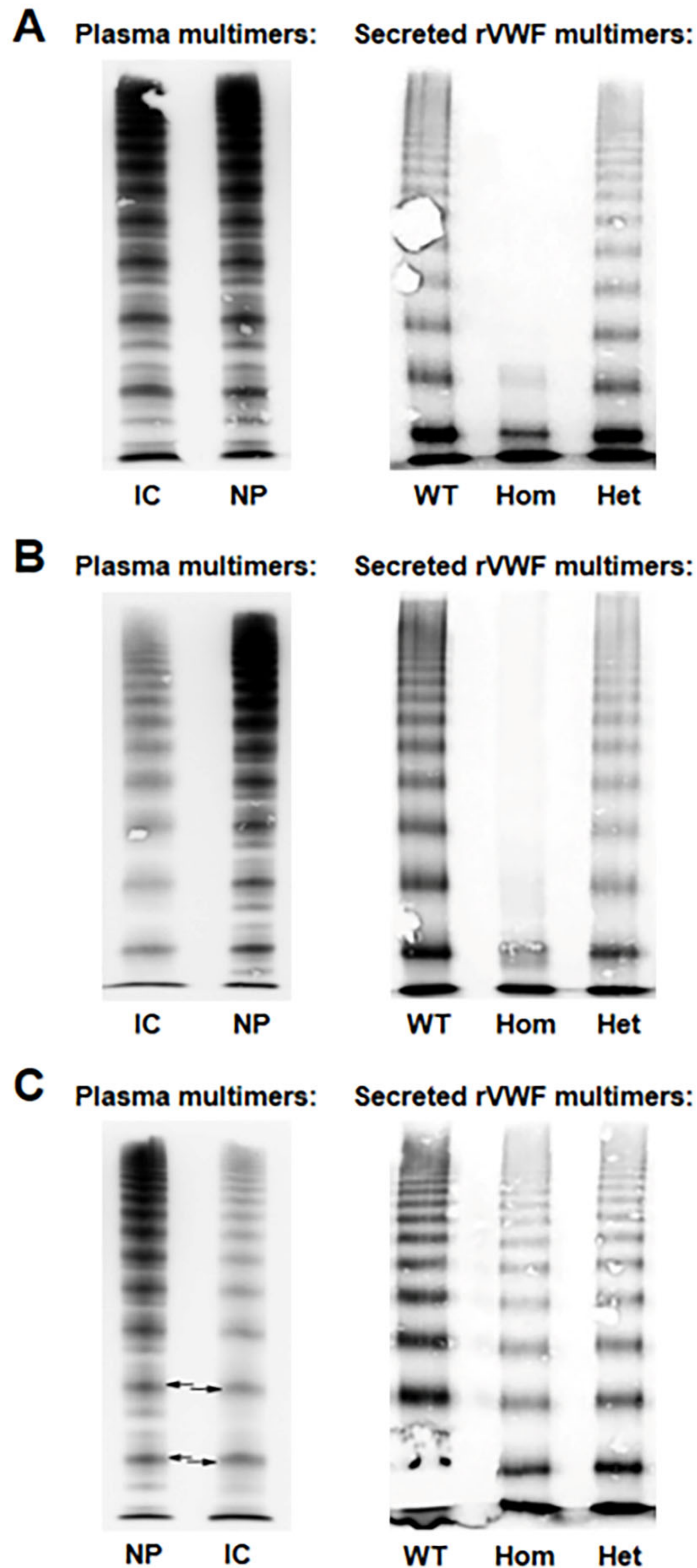

Figure S6. Patient plasma-derived and recombinant expression (rVWF)-derived multimer profiles. (A) ex3del. (B) ex32-34del. (C) ex33-34del. Multimer analysis was performed via electrophoresis on 1.6% (w/v) SDS-agarose gels. Het, heterozygous; Hom, homozygous; IC, index case; NP, normal plasma; WT, wild-type.

References

1. Hampshire DJ, Abuzenadah AM, Cartwright A, et al. Identification and characterisation of mutations associated with von Willebrand disease in a Turkish patient cohort. *Thromb Haemost*. 2013;110(2):264-274.
2. Johansson AM, Halldén C, Säll T, Lethagen S. Variation in the *VWF* gene in Swedish patients with type 1 von Willebrand Disease. *Ann Hum Genet*. 2011;75(4):447-455.
3. Sutherland MS, Cumming AM, Bowman M, et al. A novel deletion mutation is recurrent in von Willebrand disease types 1 and 3. *Blood*. 2009;114(5):1091-1098.
4. Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet*. 2016;17(4):224-238.
5. Vissers LELM, Bhatt SS, Janssen IM, et al. Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Hum Mol Genet*. 2009;18(19):3579-3593.