
Article:
Cartwright, A., Peake, I.R., Goodeve, A.C. et al. (1 more author) (2016) In silico analysis highlights the copy number variation mechanism responsible for the historically reported VWF exon 42 deletion. Haemophilia. ISSN 1351-8216

https://doi.org/10.1111/hae.13059

In silico analysis highlights the copy number variation mechanism responsible for the historically reported *VWF* exon 42 deletion

A. Cartwright, I. R. Peake, A. C. Goodeve and D. J. Hampshire
Haemostasis Research Group, Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK

Correspondence: Dr. Daniel J. Hampshire, Haemostasis Research Group, Department of Infection, Immunity and Cardiovascular Disease, Faculty of Medicine, Dentistry and Health, University of Sheffield, Beech Hill Road, Sheffield S10 2RX, UK.
Tel.: +44 (0)114 226 1433; fax: +44 (0)114 271 1863; e-mail: d.hampshire@shef.ac.uk

Main text: 1062 words; 1190 including figure legends
Figures: 2
Tables: 0
References: 10

Changes in DNA copy number (i.e. large deletions or duplications) within the human genome result from structural aberrations of chromosomal DNA, whereby previously independent regions of DNA sequence are joined together, resulting in loss or gain of DNA when compared to the reference genome [1]. Exonic deletions of the von Willebrand factor gene (*VWF*) contribute to all three types of von Willebrand disease (VWD), for example deletion of exons 4-5 in type 1 [2], exons 26-34 in type 2A [3] and exons 17-18 in type 3 [4]. Where investigated, *VWF* deletion breakpoints show a clear transition from 5' to 3' flanking DNA sequence and the mechanism resulting in the structural aberration is either *Alu*-mediated homologous recombination [2] or non-homologous (microhomology-mediated) end joining [4].

In 1990, Peake and colleagues [5] reported an out-of-frame homozygous deletion of exon 42 (ex42del) in a type 3 VWD patient, proposed to introduce a premature stop codon into the *VWF* sequence within exon 43. Interestingly, unlike other *VWF* deletion breakpoints described, the breakpoint for this copy number variant (CNV) also had a novel 182bp insertion at the breakpoint junction derived from an unknown location within the human genome [5].

In order to determine the mechanism by which this unusual ex42del CNV had arisen, the novel 182bp insertion previously reported was analysed using the Basic Local Alignment Search Tool (BLAST; http://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed March 2016), a sequence similarity search program, to identify the genomic source of the inserted DNA sequence. This analysis revealed that the 182bp sequence matched only intron 38 of *VWF* (c.6799-2026_6799-1845; National Center for Biotechnology Information (NCBI) reference sequence NM_000552.3), with a 96% homology (Fig. 1). The reason for not finding 100% homology between the reported 182bp sequence and the reference *VWF* intron 38 sequence could be explained by known single nucleotide variants (SNV) for 4/7 of the observed mismatches (Fig. 1). For the remaining 3 mismatches observed, it was possible that they represented rare unreported SNV or, for the AGA>GAG mismatch, an error when the reported sequence was originally analysed. Alternatively, an increased rate of *de novo* SNV and insertion/deletion (indel) variants had been reported around CNV breakpoints [1, 6], particularly when replication-based mechanisms were involved in their generation.

RepeatMasker analysis (undertaken via the University of California Santa Cruz (UCSC) human genome browser; http://genome-euro.ucsc.edu/index.html, accessed March 2016) was also performed to determine the presence of any repetitive elements flanking the ex42del breakpoint junction and within the inserted *VWF* intron 38 sequence. The intron 42 breakpoint was found to reside within an L3b (CR1 family) long interspersed nuclear element (LINE), however intron 41 contained no repetitive elements excluding homologous recombination as the mechanism giving rise to the ex42del. The complexity of the CNV also precluded a non-homologous end joining mechanism. The first 96bp of the intron 38 inserted sequence was shown to derive from a 407bp L1ME1 (L1 family) LINE with the remainder comprising non-repetitive sequence. However, given the truncated size of the L1ME1 element and the lack of a poly(A) tract in the inserted sequence, L1 retrotransposition was unlikely.

With the identity of the 182bp insertion confirmed, the ex42del breakpoint junction was also re-evaluated which indicated that DNA sequence flanking the reported breakpoint junction (10bp of 5' flanking *VWF* intron 41 sequence and 2bp of 3' flanking *VWF* intron 42 sequence) was also present either side of the reported 182bp inserted sequence from *VWF* intron 38 (Fig. 2A). The extension of the inserted sequence to 194bp removed the T>A and A>T mismatches previously reported by Peake and colleagues as occurring between the flanking *VWF* intron 41 sequence and the sequence observed in the index case [5]. In addition, the extension also highlighted regions of microhomology around the breakpoint junction (Fig. 2A). However, the complexity of the CNV precluded a microhomology-mediated end joining mechanism.

Given all these observations, a replication-based mechanism appeared to be the likely cause of the ex42del. Break-induced replication requires homologous recombination, so the mechanism is therefore likely to be microhomology-mediated break-induced replication (MMBIR). During cell division, genomic DNA undergoes replication to generate two genetically identical 'daughter' cells. The point at which the replication machinery disassociates the double-stranded 'parental' DNA strand to expose single-stranded DNA to be used as templates for the synthesis of the two

new DNA strands is referred to as the replication fork. MMBIR results from either collapse or stalling of the replication fork during DNA synthesis [1, 7]. In the instance associated with the ex42del, to overcome the replication fork error the replication machinery would have restarted DNA synthesis via a different replication fork located in *VWF* intron 38 before switching back to the original replication fork in intron 42; microhomology of a few nucleotides at the breakpoints allowing this switching to occur [1, 7].

G-quadruplexes can predispose DNA to form secondary structures causing initial replication forks to collapse or stall [1]. Analysis of the DNA sequence flanking the ex42del breakpoint junction utilising QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/analyze.php, accessed June 2016) highlighted two G-quadruplex forming sequences flanking the intron 41 breakpoint and one flanking the intron 42 breakpoint (Fig. 2B). The presence of a simple repetitive sequence close to the breakpoint junction, i.e. the intron 40 tandem repeat region in *VWF* (VWFdb STR Registry: http://www.vwf.group.shef.ac.uk/str.html, accessed March 2016), might have also predisposed the DNA to form secondary structures [1, 8]. In addition, utilising the DNA Pattern Find feature of the Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/dna_pattern.html, accessed March 2016) motifs known to be associated with DNA recombination [9], facilitating the replacement of the collapsed or stalled replication fork with a replacement fork to continue DNA synthesis, were shown to be located within 75bp of the breakpoint (Fig. 2B).

According to Human Genome Variation Society nomenclature guidelines (http://www.hgvs.org/mutnomen/, accessed March 2016), the ex42del structural aberration can be described as either c.7081+78_7287+1044delinsATCCATGATGCTGTCTGTTTTGATAGTTTTGACCTTCTCATTGCTAGGTAGTATTCCACGGTGTGTGTGTATCACAATTTATTTTTCTCATTCAGATTTTCACGAATGAGTCTTATTTCCTCAACCTGACTGTCAGCCATTCGAGGGCTAGGACGGTGTGTTCGAGCCTGCCCATGATGGGCACTGTGTT (according to NCBI reference sequence NM_000552.3) or more simply as g.146601_148929delinsNG_009072.1:g.141972_142161 (according to NCBI genomic reference sequence NG_009072.1).

To our knowledge, this is the first reported case of a replication-based mechanism, specifically MMBIR, associated with a CNV within *VWF*. Considering the increased understanding of the human genome and advances in genetic analysis, further examination of historically reported *VWF* CNV would not only extend our understanding of the mechanisms responsible but might also highlight whether more complex structural aberrations have a more significant impact on patient phenotype due to the unbalanced nature of the rearrangement, for example resulting in the activation of a pseudoexon [10].

Acknowledgements

This study was supported by the UK Medical Research Council (AC) and the National Institutes of Health program project grant HL081588 Zimmerman Program for the Molecular and Clinical Biology of VWD (IRP, ACG and DJH).

Author contributions

IRP initiated and coordinated the original study; all authors participated in the study design; AC, IRP and DJH analysed and interpreted results; AC and DJH wrote the manuscript; all authors revised and approved the final version of the manuscript.

Disclosures

The authors stated that they had no interests which might be perceived as posing a conflict or bias.

References

1. Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet* 2016; **17:** 224-38.
2. Sutherland MS, Cumming AM, Bowman M, Bolton-Maggs PHB, Bowen DJ, Collins PW*, et al.* A novel deletion mutation is recurrent in von Willebrand disease types 1 and 3. *Blood* 2009; **114:** 1091-8.
3. Bernardi F, Patracchini P, Gemmati D, Pinotti M, Schwienbacher C, Ballerini G*, et al.* In-frame deletion of von Willebrand factor A domains in a dominant type of von Willebrand disease. *Hum Mol Genet* 1993; **2:** 545-8.

4. Hampshire DJ, Abuzenadah AM, Cartwright A, Al-Shammari NS, Coyle RE, Eckert M*, et al.* Identification and characterisation of mutations associated with von Willebrand disease in a Turkish patient cohort. *Thromb Haemost* 2013; **110:** 264-74.
5. Peake IR, Liddell MB, Moodie P, Standen G, Mancuso DJ, Tuley EA*, et al.* Severe type III von Willebrand's disease caused by deletion of exon 42 of the von Willebrand factor gene: family studies that identify carriers of the condition and a compound heterozygous individual. *Blood* 1990; **75:** 654-61.
6. Carvalho CMB, Pehlivan D, Ramocki MB, Fang P, Alleva B, Franco LM*, et al.* Replicative mechanisms for CNV formation are error prone. *Nat Genet* 2013; **45:** 1319-26.
7. Hastings PJ, Ira G, Lupski JR. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* 2009; **5:** e1000327.
8. Lee JA, Carvalho CMB, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 2007; **131:** 1235-47.
9. Vissers LELM, Bhatt SS, Janssen IM, Xia Z, Lalani SR, Pfundt R*, et al.* Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Hum Mol Genet* 2009; **18:** 3579-93.
10. Khelifi MM, Ishmukhametova A, Khau Van Kien P, Thorel D, Méchin D, Perelman S*, et al.* Pure intronic rearrangements leading to aberrant pseudoexon inclusion in dystrophinopathy: a new class of mutations? *Hum Mutat* 2011; **32:** 467-75.

Figure legends

Fig. 1. Alignment of the previously published 182bp insertion [5] (light grey) and *VWF* intron 38 sequence (dark grey). Mismatches (dotted borders) that are known SNV: i) rs216882 (c.6799-1985G>A); ii) rs216881 (c.6799-1881T>C); iii) rs216880 (c.6799-1860C>T); iv) rs526881 (c.6799-1845C>T).

Fig. 2. Re-evaluation of the ex42del CNV breakpoint. A) The inserted intron 38 sequence observed in the index case (underlined) and flanking intron 41 and intron

42 sequence at the breakpoint junction (sequence identified to also originate from intron 38 shaded) with regions of microhomology highlighted (dotted borders). B) Location of the inserted sequence (*) in addition to G-quadruplex forming sequences (shaded), deletion hotspot consensus sequences (solid underline), DNA polymerase arrest sites (lowercase) and DNA polymerase a/b hotspots (dotted underline) [9] within 75bp of the breakpoint junction. #Full sequence GGAGAAGCGGGAACAAGTCTAGGAGG.