Critical Reviews in Oncology/Hematology 104 (2016) 1–8

Contents lists available at ScienceDirect

Critical Reviews in Oncology/Hematology

journal homepage: www.elsevier.com/locate/critrevonc

Principles of genetic variations and molecular diseases: applications in hemophilia A

N. Lannoy[a,b,*], C. Hermans[a]

[a] Hemostasis and Thrombosis Unit, Hemophilia Clinic, Division of Hematology, Cliniques Universitaires Saint-Luc, Brussels, Belgium
[b] Institut de Recherche Expérimentale et Clinique (IREC), Université Catholique de Louvain, Brussels, Belgium

Contents

ARTICLE INFO

Article history:
Received 26 May 2015
Received in revised form 7 March 2016
Accepted 14 April 2016

Keywords:
Hemophilia A
Gene targeting
F8 gene mutations
Mutation analysis
CNV
Inversion
Molecular Biology methods

ABSTRACT

DNA structure alterations are the ultimate source of genetic variations. Without them, evolution would be impossible. While they are essential for DNA diversity, defects in DNA synthesis can lead to numerous genetic diseases. Due to increasingly innovative technologies, our knowledge of the human genome and genetic diseases has grown considerably over the last few years, allowing us to detect another class of variants affecting the chromosomal structure. DNA sequence can be altered in multiple ways: DNA sequence changes by substitution, deletion, or duplication of some nucleotides; chromosomal structure alterations by deletion, duplication, translocation, and inversion, ranging in size from kilobases to mega bases; changes in the cell's genome size. If the alteration is located within a gene and sufficiently deleterious, it can cause genetic disorders.

Due to the *F8* gene's high rate of new small mutations and its location at the tip of X chromosome, containing high repetitive sequences, a wide variety of genetic variants has been described as the cause of hemophilia A (HA). In addition to the *F8* intron 22 repeat inversion, HA can also result from point mutations, other inversions, complex rearrangements, such as duplications or deletions, and transposon insertions causing phenotypes of variable severity characterized by complete or partial deficiency of circulating FVIII.

This review aims to present the origins, mechanisms, and consequences of *F8* alterations. A sound understanding of the multiple genetic mechanisms responsible for HA is essential to determine the appropriate strategy for molecular diagnosis and detected each type of genetic variant.

* Corresponding author at: Hemostasis and Thrombosis Unit, Hemophilia Clinic, Division of Hematology, Cliniques Universitaires Saint-Luc, Brussels, Belgium
E-mail address: nlannoy@dmsnet.be (N. Lannoy).

http://dx.doi.org/10.1016/j.critrevonc.2016.04.005
1040-8428/

1. Introduction

In the last 25 years of the twentieth century, our knowledge about human genetic variations was primarily limited to restriction, identifying single nucleotide and microsatellite/minisatellite variants by means of traditional polymer chain reaction- (PCR) based DNA sequencing. In recent years, the rapid development and expanded use of microarray technologies, such as an array comparative genomic hybridization (aCGH) and next-generation sequencing (NGS), have led to the discovery of submicroscopic structural variations that are not identified by classical sequencing and not visible using traditional light microscopes. The size of these rearrangements, termed copy-number variations (CNVs), is estimated to range from kilobases (kb) to megabases (mb). These rearrangements can involve deletions, duplications or insertions of DNA sections, and account for a significant amount of the individual variability within species. It is estimated that over 13% of the human genome is affected by numerous CNVs. These appear to be the main source of genetic diversity, competing with the single nucleotide variants (SNVs) (Stankiewicz and Lupski, 2010). On account of this, clearly the sequence of a gene can be altered in a number of ways: small-scale mutations, such as those affecting a gene in one or a small number of nucleotides, including point mutations (missense, nonsense, splicing, and small deletions/duplications); large-scale mutations that alter the chromosomal structure, causing large duplications/deletions, translocations, inversions, and insertions. Thus, if the variant is sufficiently deleterious to affect the gene structure and the protein synthesis associated with it, this causes genetic disease (Cooper et al., 2007; Lupski and Stankiewicz, 2005; Lee and Lupski, 2006).

This study sought to review the different variants reported in the *F8* gene, which encodes the coagulation Factor VIII, and their underlying mechanisms. Hemophilia A (HA) is an X-linked congenital bleeding disorder, caused by a lack or dysfunction of coagulation Factor VIII, and is classified as severe (< 1%), moderate (1–5%), or mild (5–40%), according to the FVIII plasma activity. The hemophilia An incidence is estimated at approximately 1 in 5,000 live male births, and no ethnic or geographic predisposition has yet been identified. Approximately 60% of hemophilia A patient have a family history of hemophilia A, the remainder consisting of sporadic cases caused by *de novo* mutations. In 90% of the latter, the mutation occurred in the parents or grandparents (Becker et al., 1996). The genetic alterations responsible for hemophilia can be classified into three categories. The first consists of alterations that change the sequence of the gene's components, or **gene variants**, including the promoter, exons, and introns. The large size of the *F8* gene results in a high rate of new small mutations (2.5×10^{-5}–4.2×10^{-5}, *versus* the median mutation rate estimated at approximately 1×10^{-6}/gene/cell division). The *F8* location at the tip of X chromosome (Xq28), which contains high repetitive sequences in close proximity to each other, renders this gene region prone to rearrangement of the second category, classified as **chromosome variants**, thus accounting for the wide variety of large genetic alterations observed in HA patients (Bauters et al., 2008; Vandewalle et al., 2009; Vogel and Motulsky, 1997). The third source of DNA damage corresponds to the insertion of mobile elements, termed "transposons,ẅhich has also been described among HA patients.

2. Genomic organization of *F8* at locus Xq28

First cloned in 1984, the Factor VIII gene (*F8*) was mapped to the distal end of the long arm of the X chromosome (Xq28) on the minus strand. This gene spans 186 kb (hg19: chrX:154064064-154250998 UCSC genome browser, see http://genome.ucsc.edu/) of genomic DNA. The gene is divided into 26 exons and transcribed into an mRNA of 9029 bp, with a coding sequence of 7053 nucleotides, encoding a mature protein of 2332 amino acid residues that are arranged within six domains, organized as follows: A1-A2-B-A3-C1-C2. Compared to all other exons, and with the exception of exon 26 that mostly codes for the 3′ untranslated region, exon 14 is very long, measuring 3106 bp, and encodes most of the B domain (Fig. 1). The *F8* gene contains the particularly large intron 22, measuring 32.8 kb, which exhibits several particularities. These include the presence of a bidirectional promoter who initiates the transcription of expressed genes (*F8A* and *F8B*) and whose function is not yet well understood. The intronless *F8A* (OMIM 305423) spans 2 kb and is transcribed in the opposite direction to *F8*. It encodes a 40-kDa Huntingtin-associated protein (Peters and Ross, 2001), thought to be involved in the aberrant nuclear localization of the huntingtin protein observed in Huntington's disease. *F8 B* (OMIM 305424) is transcribed in the same direction using a private exon within intron 22 that is spliced to exons 23 through 26, with the *F8* reading frame creating a final overlapping transcript spanning 2.5 kb. The function of the *F8 B* transcript and its potential translated product remain unknown (Graw et al., 2005). This arrangement is further complicated by the association between these three sequences (*F8A*, the first exon of *F8B*, and their common promoter) within a 9.5 kb fragment labelled *int22h-1* that is duplicated at two positions towards the Xq-telomere (*int22h-2* and *int22h-3*), situated more telomerically at approximately 488 and 566 kb (Naylor et al., 1995). *Int22h-2* and *int22h-3* demonstrate 99.93% overall similarity, while the homology between *Int22h-1* and *int22h-3* is of 99.24% and that of the other *int22h-2* repeats 99.18%. Interestingly, the statistics and distribution of the sequence differences between the homologous copies strongly support theories that the copy in the *F8* gene was introduced by a duplication more than 25 million years ago (Bagnall et al., 2005). Similar to the int22 h sequence, a different repeat has been identified within the *F8* intron 1, labelled *int1h-1* and located approximately 15.26 kb downstream of exon 1. An inverted homologous copy of this was found approximately 125 kb upstream of *F8* (labelled *int1h-2*). Bagnall et al. postulated that *int1 h*, like *int22 h*, duplicated more than 25 million years ago (Bagnall et al., 2005).

FVIII requires a scaffold of protein chaperones within the endoplasmic reticulum (ER) to help it create its complex tertiary structure. These intracellular interactions also serve as a quality control mechanism in order to retain misfolded FVIII protein until it can achieve its properly folded structure or to target this protein to the ER degradation pathways. Null mutations (non-sense, deletion/duplication out-frame mutations), which introduce a premature termination codon in the mRNA, are thus either degraded by nonsense-mediated decay (NMD) or translated into a truncated protein (Mendell and Dietz, 2001; Baker and Parker, 2004). Predictably, missense mutations significantly disrupt the tertiary structure of FVIII molecules, which are primarily retained and degraded, resulting in severe hemophilia A. In mild/moderate HA; missense mutations affect the functional properties of FVIII, resulting in reduced secretion and a less severe clinical phenotype. These missense mutations are clustered in regions known to interact with Factors IXa and X, the Von Willebrand factor (VWF), and the phospholipid surface, alternatively can affect the stability of FVIII after activation by thrombin (d'Oiron et al., 2008).

Genetic analysis of hemophilia disease requires several technical approaches to cover the entire field of HA defects, thus necessitating numerous expensive and time-consuming procedures. As a brief overview, 25% of all HA cases are caused by an inversion resulting from homologous recombination involving intron 1 or 22 and related sequences outside the F8 gene. Several methods have been developed that are now frequently used to detect these inversions (targeted PCR, Southern blot, LR-PCR, and IS-PCR). In order to detect approximatively 75% of the other muta-