



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/103690/>

Article:

Edison, E., Konkle, B.A. and Goodeve, A.C. (2016) Genetic analysis of bleeding disorders. *Haemophilia*, 22 (S5). pp. 79-83. ISSN: 1351-8216

<https://doi.org/10.1111/hae.13024>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Genetic Analysis of Bleeding Disorders

Eunice Edison,¹ Barbara A. Konkle,² Anne C. Goodeve^{3,4}

¹Department of Haematology, Christian Medical College, Vellore, India; ²Bloodworks Northwest and University of Washington, Seattle, WA, USA; ³Sheffield Diagnostic Genetics Service, Sheffield Children's NHS Foundation Trust and ⁴Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, UK

Running head: Genetic Laboratory for Bleeding Disorders

Corresponding author:

Prof Anne Goodeve, BSc, PhD

Haemostasis Research Group

Department of Infection, Immunity and Cardiovascular Disease

Faculty of Medicine, Dentistry and Health

Beech Hill Road

SHEFFIELD, S10 2RX, UK

Word count: 3188, including abstract- references

References: 14

Tables: 3

Email: a.goodeve@shef.ac.uk

Tel: 0044 114 271 2679 / 7005

Fax: 0044 114 271 1863

Secretary: Margaret Beckett

Email: m.beckett@sheffield.ac.uk

Keywords: bleeding disorder, confirmation-sensitive gel electrophoresis, dosage analysis, genetic analysis, next generation sequencing, pathogenicity prediction, F8 inversion.

Abstract

Molecular genetic analysis of inherited bleeding disorders has been practiced for over 30 years. Technological changes have enabled advances, from analyses using extragenic linked markers to next-generation DNA sequencing and microarray analysis. Two approaches for genetic analysis are described, each suiting their environment. The Christian Medical Centre in Vellore, India uses confirmation sensitive gel electrophoresis mutation screening of multiplexed PCR products to identify candidate mutations, followed by Sanger sequencing confirmation of variants identified. Specific analyses for *F8* intron 1 and 22 inversions are also undertaken.

The MyLifeOurFuture US project between the American Thrombosis and Hemostasis Network, the National Hemophilia Foundation, Bloodworks Northwest and Biogen uses molecular inversion probes (MIP) to capture target exons, splice sites plus 5' and 3' sequences and to detect *F8* intron 1 and 22 inversions. This allows screening for all *F8* and *F9* variants in one sequencing run of multiple samples (196 or 392). Sequence variants identified are subsequently confirmed by a diagnostic laboratory.

After having identified variants in genes of interest through these processes, a systematic process for determining their likely pathogenicity should be applied. Several scientific societies have prepared guidelines. Systematic analysis of the available evidence facilitates reproducible scoring of likely pathogenicity. Documentation of frequency in population databases of variant prevalence and in locus-specific mutation databases can provide initial information on likely pathogenicity. Whereas null mutations are often pathogenic, missense and splice-site variants often require *in silico* analyses to

predict likely pathogenicity and using a standard suite of tools can help standardise their documentation.

Introduction

Genetic analysis is a well-established process in the analysis of inherited bleeding disorders. New technology continually enables the process to become more automated in laboratories that can access these new techniques. A number of variations on next generation DNA sequencing (NGS) are being used for bleeding disorder genetic analysis. Simpler screening techniques also retain utility in many laboratories and although their mutation detection rate may be slightly lower, these remain an important contribution to genetic analysis, facilitating identification of mutations in many laboratory situations. The approaches to analysis of diverse bleeding disorders in Vellore, India, and of haemophilia A and B simultaneously in Seattle, US are presented.

Once candidate pathogenic mutations have been identified in index cases, the possible pathogenicity of the variant(s) can be assessed and where relevant, reported back to the requesting clinician with an assessment of likely contribution to the patient's disease. An approach to undertaking pathogenicity assessment and its documentation is described.

Proficient genetic laboratory for diagnosis of bleeding disorders at Christian Medical College, Vellore,

India-Eunice Edison

Hereditary bleeding disorders are common in southern India due to high consanguinity rates. The Christian Medical College (CMC), Vellore provides advanced hemostasis diagnostic and treatment facilities to a population of ~250 million, where over 1000 new patients with hereditary bleeding disorders are diagnosed annually. This includes all coagulation factor deficiencies and platelet function disorders. Patients require accurate diagnosis and their families need carrier detection and prenatal

diagnosis. The Centre has therefore gradually developed advanced hematological and molecular genetic diagnostic facilities over the last 20 years.

Evaluating molecular genetics of hemophilia and other hemostatic disorders was initiated in 1995.

Analysis was initially confined to intron 22 inversion genotyping and linkage-based analysis of mutations in hemophilia A and B [1]. In 2000, with support from Dr Carol Kasper, ex-Vice President of the World Federation of Hemophilia, enhancement of the technologies available was initiated and Dr Shaji who undertook training in haemophilia genetic analysis at the international haemophilia training centre (IHTC) in Sheffield, UK.

Over the next 5-6 years, technology for mutation detection for a wide range of bleeding disorders was established. This included the hemophilias, where the inverse PCR for intron 22 [2] and the 2-tube test for the intron 1 inversions [3] are used. Gene dosage analysis of exons 3, 6, 8, 14, 20 and 24 of *F8* and exons 6 and 8 of *F9* uses albumin as the control gene. Establishment of multiplex ligation-dependent probe amplification (MLPA) for both *F8* and *F9* to analyze large deletions/duplications is underway.

Analysis includes many rare coagulation disorders and also common platelet-function disorders such as Glanzmann thrombasthenia. The laboratory uses the well-established conformation sensitive gel electrophoresis (CSGE) technique, a simple mutation screening method [4]. Through modifying existing protocols by multiplexing PCRs, a comprehensive and cost-effective mutation screen for haemophilia and other rare bleeding disorders has been established [5-7]. Subsequently, Bernard-Soulier and Wiskott-Aldrich syndromes were included in the laboratories repertoire [8]. Molecular diagnosis of von Willebrand disease (VWD) has recently been introduced using a similar CSGE-sequencing approach [9]. The percentage of type 3 VWD patients (62%) is higher than those of types 1 (21.5%) and 2 (15.5%) in our study population in contrast to reports from western populations where type 3 VWD is less frequent than types 2 and 1 [9]. The approach used has the major advantage that it is cost-effective and readily feasible for most patients. Molecular genetic studies of the coagulation factors and disorders of

hemostasis have contributed to better understanding of the biology of these disorders, facilitating accurate detection of carriers and genetic counselling in many families (Table 1).

The establishment of the molecular genetic facilities has not only aided accurate diagnosis, but furthered understanding the complexity of genetic changes, genotypic-phenotypic heterogeneity and also devising therapeutic options in patients with bleeding disorders. This is well documented in scientific contributions to national (n=12) and international conferences (n=13).

The laboratory provides training in molecular diagnosis of bleeding disorders. We have instructed fourteen individuals (South Africa-1, South Korea-3, Malaysia-2, Sri Lanka-2, Bangladesh-2 and India-4) in molecular diagnosis of haemophilia; four of these were trained through IHTC fellowships.

With the increase in the number of laboratories offering molecular diagnosis, the need for a proficiency testing program was felt. An external quality assurance scheme for haemophilia (A and B) and genetic markers of thrombosis was initiated in 2006 when three laboratories were included. This recently increased to six laboratories; surveys are sent twice yearly. Thirty four samples (haemophilia A and B) have been dispatched in fifteen surveys. Response rate is between 70 and 80% while accuracy ranges between 70-80%.

The techniques have been combined to provide comprehensive genetic analysis of bleeding disorders and facilities and techniques are continually expanded to enable CMC Vellore to be the most proficient hemostasis genetic laboratory in the country.

The MyLifeOurFuture Haemophilia Genetic Analysis Project-Barbara Konkle

The MyLifeOurFuture (MLOF) initiative (www.mylifeourfuture.org) was begun in the US in 2012 through partnership between the American Thrombosis and Hemostasis Network (ATHN), the National Hemophilia Foundation (NHF), Bloodworks Northwest (BWNW, formerly the Puget Sound Blood Center), and Biogen, with a goal of providing genotyping to US haemophilia A and B patients and to establish a

data repository and samples to facilitate research in haemophilia and its complications. At that time, surveys of consumers by NHF and of providers by ATHN found that approximately 20% of patients had *F8* or *F9* DNA analysis performed, confirming the need for genetic analysis to support clinical care. In addition, the program was built on the commitment by the partners to expand research in haemophilia and its treatment.

Each of the four partners brings distinct expertise and together they govern the project. ATHN, working with 135+ affiliated Haemophilia Treatment Centers (HTC), provides HTC provider education, secure data collection infrastructure known as ATHN Clinical Manager, and point of access for future research proposals. Participating HTC providers contracted through ATHN enroll patients, obtain samples for genotyping, and provide clinical results to patients. NHF, a US national patient advocacy organization with 52 chapters, educates the bleeding disorders community about the initiative through its publications, annual meeting and local chapter events. This keeps the community informed about project status and supports recruitment. BWNW facilitates receipt of samples from sites, serves as the central genotyping laboratory and houses the research sample repository. Biogen, a biotechnology company with products for treatment of haemophilia, provides scientific collaboration and financial support.

Patients and for minors, their parents (patients/parents), are informed of the project through NHF and their area HTC, which establishes a contract with ATHN. Patients/parents are offered enrollment in the institutional review board approved research repository. Blood for DNA extraction is sent to BWNW for genetic analysis and, for those consenting to research, additional samples are stored in the repository. Haemophilia severity is determined by local laboratory testing. Initial *F8* and *F9* DNA analysis is performed utilizing a next generation sequencing (NGS) approach employing molecular inversion probes (MIP) for DNA capture [10, 11], targeting all exons, splice sites, and sequences 5' and 3' to the coding region. MIP also detect *F8* intron 1 and 22 inversions using an approach similar to the inverse-shifting

PCR methodology described by Rossetti et al [12]. This allows screening for all *F8* and *F9* variants in one sequencing run of multiple samples (196 or 392). Variants identified by NGS are confirmed in the BWNW CLIA-certified laboratory by another method specific to the variant. In males where no likely deleterious DNA variant is identified, and in females with moderate/severe disease and only one identified deleterious variant, MLPA is used. A clinical report is returned to the HTC, and results are transmitted electronically through ATHN to the HTC's Clinical Manager database. Variant results are classified per American College of Medical Genetics and Genomics (ACMG) guidelines for pathogenicity [13].

As of February 2016, 4651 patients have enrolled in MLOF, and 3694 of those participants/parents consented to the sample and data research repository. Ninety four HTCs were engaged in the project, with 81 sites actively enrolling patients. Of the 4649 patients, 3636 have haemophilia A [1924 severe (S) Male (M):1913, Female (F):11], 659 moderate (Mo) (M:650, F:9), 1053 mild (Mi)(M:907, F:146)], 906 have haemophilia B [205 S (M:203, F:2), 336 Mo (M:335, F:1), 365 Mi (M:319, F:46)], and 107 are unclassified as part of a carrier pilot. Once optimized, our NGS approach has detected all variants, including large deletions, except for partial gene duplications of which it detected 6 of 9 found in 16 patients, all with haemophilia A. In the first 3000 patient samples analyzed, 228 previously unreported variants were found. A potentially causative variant was found in all patients except 27 with haemophilia A (S:7, Mo:2, Mi:18) and 2 with haemophilia B (S:0, Mo:1, Mi:1). In sequencing both genes simultaneously we found a potentially causative variant in the other gene in 5 patients, 3 of whom had been previously reported as associated with haemophilia. Factor levels in those patients and prevalence in a population database (ExAC) raise questions regarding their pathogenicity.

During this project, the ACMG published new standards and guidelines for interpreting DNA sequence variants, which uses specific evidence to assign pathogenicity [13]. For variants predicted to result in null alleles, the interpretation of the variant is fairly straightforward. For other variants, interpretations

of pathogenicity or likely pathogenicity (>90% certainty) require additional evidence, and given the past practice to genetically test only one affected family member, required segregation data can be lacking. Within MLOF, additional data from the HTC's will be obtained to maximize knowledge regarding the variants, which will inform deter pathogenicity assessment and deepen knowledge concerning hemophilia-causing variants.

In the future, with enrollment of 5000 subjects in the research repository, investigators worldwide will be able to apply to access the de-identified genetic data, repository samples, and phenotypic data through ATHN. Coded data and biologic samples will be linked to coded clinical data from the ATHNdataset, gathered separately from this study. An independent, multidisciplinary research review committee managed by ATHN will evaluate proposals for scientific integrity and feasibility of research proposal and govern release of the data and sample repository. In concert, ~2200 samples are undergoing whole genome sequencing through the US National Heart, Lung and Blood Institute Transomics in Precision Medicine (TOPMed) program and results will be deposited in the NIH Database of Genotypes and Phenotypes (dbGap) for use for scientific investigation.

In summary, MLOF is providing genetic information for patients and their families to help inform reproductive planning and clinical care. In addition, through the research repository and the ATHN clinical database, an invaluable resource for research in hemophilia and associated disorders is being built.

Pathogenic or Not? Classification of genetic variants in haemostasis-Anne Goodeve

Once a genetic variant(s) have been identified in a bleeding disorder of interest, an assessment must be made regarding its likely pathogenicity prior to reporting the variant to the requesting clinician. This can be straightforward where the variant has been reported a number previously in other patients with the same disorder and similar severity. In these cases, documenting the genotype-genotype relationship for

the variant can contribute substantially to evidence of pathogenicity. However, making a judgement on possible pathogenicity is more challenging when the variant has not been previously reported.

Organizations including the UK Association for Clinical Genetic Science (ACGS) and the ACMG have produced guidelines on sequence variant classification [13, 14].

Standardized methods to collect and document information available on the sequence variant can be very useful in determining its likely pathogenicity. The ACGS recommend a classification on a scale of 1-5, where a classification of 1 indicates that the variant is common in the normal population and therefore not pathogenic, whereas 5 indicates that the variant has been seen previously in other patients with the disorder and that there is evidence of its pathogenicity through other analyses such as *in vitro* expression [14]. Categories between these extremes are 2; unlikely pathogenic, 3; variant of unknown significance and 4; likely pathogenic (Table 2). Both the ACGS and ACMG guidelines [13, 14] recommend combining different evidence types to reach a conclusion regarding pathogenicity.

Two resources that can be most useful at the outset of these investigations are locus-specific mutation databases (LSDB) for the gene(s) of interest and sequence variant databases where many individuals have been analyzed for their sequence variation. Resources are listed in Table 3. Classification can be most challenging when a variant has not been previously reported. *In silico* predictions should be employed to help predict possible pathogenicity in these instances. Many different algorithms using different methodologies have been designed for both missense mutations and possible splice site mutations (PSSM) and these resources are listed in Table 3 and in the guidelines [13, 14].

Amino acid conservation examines the extent of conservation of the residue across several species, typically using sequences from several mammals, but also including chicken, frog and fish to encompass around 500 million years of evolution. Strongly conserved residues are much more likely to be structurally/functionally important than those that display significant variation. PSSM may be analyzed using a range of different splice site prediction tools. In both of these predictions, analyses using

different algorithm types should be used, and generally at least five different tools are recommended to help reach a consensus. For these analyses to be most useful for future patients, a standardized documentation format can be used to record the findings from each part of the data analysis that can be updated when further information is identified.

For a small proportion of mutations, functional analysis may be available through publications that can include *in-vitro* mutagenesis and missense mutation expression analysis. This can be helpful in pathogenicity prediction, as can analysis of mRNA expression analysis, seeking any aberrant transcript(s) produced as a result of a mutation affecting splicing. In some instances, where predicted pathogenicity is inconclusive (category 3, variant of unknown significance), the report can suggest analysis of additional family member(s) to determine possible co-segregation of the variant with the disorder. Similar analysis of relatives can also be undertaken where the phase of inheritance is uncertain when two or more potentially pathogenic variants are present.

Utilization of the laboratory analytical techniques along with application of external quality assessment and systematic analysis of variant pathogenicity can lead to robust systems for bleeding disorder genetic analysis.

Acknowledgements

MyLifeOurFuture is funded by Biogen and Dr. Konkle is supported by Biogen for her effort in this project. ACG has received honoraria from Octapharma and Baxalta. CSL-Behring fund upkeep and development of the VWF mutation database hosted in Sheffield.

References

1. Jayandharan G, Shaji RV, George B, Chandy M, Srivastava A. Informativeness of linkage analysis for genetic diagnosis of haemophilia A in India. *Haemophilia* 2004; **10**: 553-9.
2. Rossetti LC, Radic CP, Larripa IB, De Brasi CD. Genotyping the hemophilia inversion hotspot by use of inverse PCR. *Clin Chem* 2005; **51**: 1154-8.
3. Bagnall RD, Waseem N, Green PM, Giannelli F. Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A. *Blood* 2002; **99**: 168-74.
4. Williams IJ, Abuzenadah A, Winship PR, Preston FE, Dolan G, Wright J, *et al.* Precise carrier diagnosis in families with haemophilia A: use of conformation sensitive gel electrophoresis for mutation screening and polymorphism analysis. *Thromb Haemost* 1998; **79**: 723-6.
5. Jayandharan G, Viswabandya A, Baidya S, Nair SC, Shaji RV, George B, *et al.* Six novel mutations including triple heterozygosity for Phe31Ser, 514delT and 516T-->G factor X gene mutations are responsible for congenital factor X deficiency in patients of Nepali and Indian origin. *J Thromb Haemost* 2005; **3**: 1482-7.
6. Jayandharan GR, Viswabandya A, Baidya S, Nair SC, George B, Mathews V, *et al.* Mutations in coagulation factor XIII A gene in eight unrelated Indians. Five novel mutations identified by a novel PCR-CSGE approach. *Thromb Haemost* 2006; **95**: 551-6.
7. Jayandharan G, Spreafico M, Viswabandya A, Chandy M, Srivastava A, Peyvandi F. Mutations in the MCFD2 gene are predominant among patients with hereditary combined FV and FVIII deficiency (F5F8D) in India. *Haemophilia* 2007; **13**: 413-9.
8. David S, Jayandharan GR, Abraham A, Jacob RR, Devi GS, Patkar N, *et al.* Molecular basis of Wiskott-Aldrich syndrome in patients from India. *Eur J Haematol* 2012; **89**: 356-60.
9. Ghosh K, Shetty S. Epidemiology, diagnosis, and management of von Willebrand disease in India. *Semin Thromb Hemost* 2011; **37**: 595-601.

10. Turner EH, Lee C, Ng SB, Nickerson DA, Shendure J. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nature Meth* 2009; **6**: 315-6.
11. Boyle EA, O'Roak BJ, Martin BK, Kumar A, Shendure J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics* 2014; **30**: 2670-2.
12. Rossetti LC, Radic CP, Larripa IB, De Brasi CD. Developing a new generation of tests for genotyping hemophilia-causative rearrangements involving int22h and int1h hotspots in the factor VIII gene. *J Thromb Haemost* 2008; **6**: 830-6.
13. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; **17**: 405-24.
14. Wallis Y, Payne S, McAnulty C, Bodmer D, Sistermans E, Robertson K, *et al*. Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics. Association for Clinical Genetic Science and the Dutch Society of Clinical Genetic Laboratory Specialists. 2013. Available from http://www.acgs.uk.com/media/774853/evaluation_and_reporting_of_sequence_variants_bpgs_june_2013_-_finalpdf.pdf

Table 1. Summary of genetic diagnosis carried out in the CMC Vellore laboratory on patients with disorders of hemostasis from 1996-2016.

Disorder	Patients Analyzed (n)	Prenatal Diagnosis (n)
Haemophilia A	667	232
Haemophilia B	143	26
FVII Deficiency	5	4
FX Deficiency	2	-
FXIII Deficiency	6	2
Glanzmann Thrombasthenia	18	6
Bernard-Soulier Syndrome	4	-
Wiskott-Aldrich Syndrome	65	8

Table 2. Five ACGS variant pathogenicity classes

Class	Description *	Suggested report wording
1	Clearly not pathogenic	Not pathogenic (not reported)
2	Unlikely to be pathogenic	Unlikely pathogenic (not reported)
3	Variant of unknown significance (VUS)	Uncertain pathogenicity
4	Likely to be pathogenic	Likely pathogenic
5	Clearly pathogenic	Predicted to be pathogenic

Wallis, et al [14]

Table 3. Classes and examples of evidence used to help determine predicted pathogenicity.

Evidence type	Examples	Web address
Population frequency	Exome Variant Server; EVS	http://evs.gs.washington.edu/EVS/
	Single nucleotide polymorphism database; dbSNP	http://www.ncbi.nlm.nih.gov/snp/
	Exome aggregation consortium; ExAC	http://exac.broadinstitute.org/
Presence on LSDB	1000 genomes project	http://browser.1000genomes.org
	Factor FVIII variant database	http://www.factorviii-db.org/
	CDC Hemophilia A Mutation Project (CHAMP) and Hemophilia B Mutation Project (CHBMP)	http://www.cdc.gov/ncbddd/hemophilia/champs.html
	Hemobase HA, HB and VWD (Spanish registry)	http://www.hemobase.com/EN/index.htm
	Factor IX variant database	http://www.factorix.org/
	von Willebrand factor variant database	https://grenada.lumc.nl/LOVD2/VWF/home.php?select_db=VWF
Amino acid prediction	MutPred	http://mutpred.mutdb.org/
	Sorting intolerant from tolerant (SIFT)	http://sift.jcvi.org/
	PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/
	SNAP	https://www.rostlab.org/services/snap/
	MutationAssessor	http://mutationassessor.org/
Splice site prediction	Human Splicing Finder	http://www.umd.be/HSF3/
	NNSplice	http://omictools.com/nnssplice-tool
	GeneSplicer	http://www.cbc.umd.edu/software/GeneSplicer/gene_spl.shtml
	MaxEntScan	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html