UNIVERSITÀ DEGLI STUDI DI MILANO

PhD Program in Translational Medicine

XXXVII Cycle

Department of Biomedical Sciences for Health

COMPUTATIONAL APPROACHES IN STUDYING INHIBITOR DEVELOPMENT IN HAEMOPHILIA A PATIENTS

PhD candidate:
Himani Chand

Matricola N.:
R13656

Supervisor:
Prof. Flora Peyvandi

Co-supervisor:
Dr. Roberta Palla

PhD Program Director:
Prof. Mario Clerici

Academic Year 2024 – 2025

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Computational methods used in translational medicine

Translational medicine is an interdisciplinary field in biomedical research that aims to connect basic science with practical clinical applications. The main objective of translational medicine is to translate laboratory findings into diagnostic methods, tools, treatments, and preventive measures that enhance patient care. Disease progression is influenced by the interplay of genetic, epigenetic, environmental, and lifestyle factors, which require a comprehensive understanding of disease mechanisms.

In recent decades, the advancement of biomedical data has transformed the understanding of disease processes and treatment strategies. Computational approaches have become essential in translational medicine, helping in the transition of fundamental scientific discoveries into clinical use. These approaches serve as a crucial link between laboratory research and patient treatment, enabling scientists to investigate complex biological systems, predict disease progression, and identify potential therapeutic targets. The integration of computational methods into medical research has revolutionised the conventional approaches to disease investigation. Advancements in high-throughput technologies like next-generation sequencing (NGS) and other omic technologies such as transcriptomics, proteomics, etc, have led to a generation of vast amounts of data that requires sophisticated computational and bioinformatic tools for effective analysis, offering unparalleled opportunities to examine diseases at molecular, cellular, and systemic levels.

This shift towards data-driven research in the field of biomedical sciences has given rise to the use of various computational and bioinformatic techniques that can process and analyse biological information to understand the complexities of the progression of disease and help in accelerating the development of effective therapies.[1] This chapter consists of an overview of the primary computational and bioinformatics methods employed in translational medicine and their applications in investigating disease mechanisms.

Major computational and bioinformatic approaches are used to study disease mechanisms

Machine Learning (ML) and Artificial Intelligence (AI):

Machine learning algorithms are often used in translational research. Supervised learning algorithms are used for disease classification and to create prediction models. Unsupervised learning techniques like clustering are used to get insights into the patterns in a dataset. Meanwhile, deep learning algorithms are also being used in medical image studies.[2]

Network Biology:

Diseases often involve the interaction of multiple molecular pathways and cellular processes. In network biology, various methodologies are used to understand interactions between biological systems as interconnected networks, such as protein-protein interactions, gene regulatory networks, and metabolic pathways. Network analysis helps to identify key molecules, pathways, and interactions involved in an underlying condition, as well as in identifying drug targets in a disease network.[3]

Omic studies and multi-omics data integration:

Advancements in next-generation sequencing technologies have helped to provide a comprehensive image of the genetic and molecular basis of diseases. Omics include genomics, transcriptomics, proteomics, metabolomics, single-cell transcriptomics, single-cell multi-omics, spatial transcriptomics, and other methods. These technologies provide valuable insights into the biological pathways and processes that distinguish disease states from healthy controls. Integration of data from the multi-omics methods helps in establishing a more detailed cause-and-effect relationship.[4] For instance, a study published by Iperi et al. combined data from genomics, transcriptomics, proteomics, and metabolomics to comprehensively analyse B lymphocyte metabolism in systemic lupus erythematosus (SLE) patients. Multi-omics integration revealed metabolites such as lactate and glutamine correlated with disease activity and severity.[5]

Different methods of these key computational approaches are often used together, as per convenience, to better understand disease mechanisms and treatment response.

Applications in translational medicine

Biomarker Discovery:

Potential biomarkers are identified from various datasets using multi-omics, while ML algorithms are used to analyse gene expression data to identify signatures associated with a disease state. These methods could be used to discover biomarkers for the diagnosis and prevention of several rare as well as common disorders.

Drug Discovery and Development:

High-throughput methods are often used to do virtual screening of compound libraries and for the prediction of drug-target interactions. These methods reduce the time and cost associated with drug development.

Precision Medicine:

Predictive models based on patient-specific genetic variant data are used for the prediction of treatment response and identification of disease subtypes to recommend tailored treatments.

Applications of computational approaches used in the thesis

The studies performed in my thesis used computational and bioinformatic approaches to study inhibitor development in haemophilia A neonates at an early stage of life. The approaches contribute to translational medicine by improving disease prediction, biomarker discovery and personalised treatment strategies. The following are the approaches used in this thesis and their key applications:

- **Prediction of inhibitor development in HA**
  1. Statistical modelling and ML algorithm on Mimotope-variation Assay- The chapter used Mimotope-variation analysis for statistical modelling and machine-learning algorithms to develop prediction models for identifying patients at risk of inhibitor development before FVIII exposure, suggesting early intervention strategies which can help in creating a personalised treatment plan by tailoring FVIII therapy based on the immune risk profile of the patient.
  2. Epitope Mapping for therapeutic target identification- Mimotope-variation assay using random peptide phage display library has been used to find epitopes on the FVIII protein associated with inhibitor development. It identified epitopes on FVIII domains that produce a stronger antibody response. This observation supports the design of less immunogenic FVIII variants and tolerogenic therapies that reduce the risk of inhibitor development in HA patients.
- **Investigating molecular mimicry**

  Sequence similarity- This chapter used sequence similarity analysis using the pathogen epitope database (IEDB) to test whether pathogen-derived epitopes mimic FVIII, which could contribute to inhibitor formation. No evidence of such mimicry in the study refined our understanding of immune-modulation.
- **DNA methylation for biomarker discovery**

  Next-generation sequencing using Methyl-seq- This chapter dealt with DNA methyl-seq for the identification of differentially methylated CpG sites in patients developing inhibitors, suggesting potential for epigenetic biomarkers to stratify patients based on inhibitor risk.

The integration of bioinformatics, NGS and machine learning in this thesis illustrates how computational approaches used in translational medicine could help in improving disease understanding, risk prediction, as well as treatment optimisation.

Haemophilia A and Inhibitor development

Haemophilia A (HA) is a hereditary bleeding disorder caused by a deficiency or dysfunction of Factor VIII (FVIII), which is a critical protein involved in the blood coagulation cascade. HA is an X-linked recessive disorder which affects nearly 1:5,000 male births.[6] HA is characterised by spontaneous bleeding, prolonged bleeding after an injury and an increased risk of soft tissue and joint bleeding. HA is categorised into three levels of severity based on the remaining FVIII activity in plasma: severe (less than 1% FVIII activity), moderate (1-5%), and mild (6-40%). The risk of formation of alloantibodies (inhibitors) against FVIII in patients with severe HA is higher, and about one-third of the patients develop inhibitors during the FVIII infusion treatment. This section offers a broad overview of Haemophilia A, the role of FVIII and the inhibitor development against FVIII.

Haemophilia A: Genetic, Molecular and Clinical Background

Haemophilia A is caused by mutations in the *F8* gene coding for FVIII protein, located on the X chromosome (Xq28).[7] This gene is approximately 186kb, containing 26 exons. The gene encodes a protein of 2,351 amino acids, which undergoes extensive post-translational modifications. The mature FVIII protein consists of domains arranged as A1-A2-B-A3-C1-C2, B domain is removed for coagulation activity. Mutations in the *F8* gene have been identified, including point mutations, insertions, deletions, and inversions. The most common mutation, accounting for approximately 45% of severe Haemophilia A cases, is an intron 22 inversion. These genetic alterations lead to reduced or absent FVIII activity, impairing the blood coagulation process.

Clinically, Haemophilia A manifests as spontaneous bleeding into joints (hemarthrosis), muscles, and soft tissues, as well as prolonged bleeding after trauma or surgery. Repeated bleeding episodes can lead to chronic joint damage, arthropathy, and disability. The diagnosis of Haemophilia A is based on clinical history, family history, and laboratory tests, including FVIII activity assays and genetic testing. Early diagnosis and prophylactic treatment with FVIII replacement therapy have significantly improved the quality of life for patients with Haemophilia A.

Factor VIII: Structure and Function

FVIII is synthesised as a single-chain protein that undergoes processing:

1. Signal peptide removal
2. Heavy chain formation (A1-A2-B domains)
3. Light chain formation (A3-C1-C2 domains)
4. Metal ion binding (primarily copper and calcium)

FVIII is a key cofactor in the intrinsic pathway of the blood coagulation cascade. It circulates in the blood as an inactive precursor bound to von Willebrand factor (VWF), which stabilises and protects FVIII from degradation. Upon vascular injury, FVIII is activated by thrombin or factor Xa, leading to its dissociation from VWF.

Treatment therapies

Haemophilia A treatment strategies include

- **FVIII replacement therapy**- infusion of recombinant or plasma-derived FVIII
- **Extended half-life FVIII products**- genetically modified FVIII with a longer half-life, therefore reducing the infusion frequency for prophylaxis.
- **Non-factor therapies** - includes FVIII independent approaches for prophylaxis.
  1. FVIII mimetic antibody- Emicizumab is a bispecific antibody therapy using two monoclonal antibodies that mimics FVIII function by bridging factor IXa and factor X. Mim8, which is under trials, is a bispecific humanised IgG4 antibody that bridges FX and FIXa.
  2. Anticoagulant quenchers- Anti-TFPI treatment uses a humanised monoclonal antibody against tissue factor pathway inhibitor (TFPI). TFPI, which is a natural anticoagulant, inhibits FXa and the tissue factor-FVIIa complex crucial for clot formation.

  Another RNAi therapy under clinical trials is Fitusiran, which targets the transcription of antithrombin by decreasing hepatic synthesis and plasma levels to rebalance coagulation.
- **Gene therapy**- uses viral vectors to deliver a functional FVIII gene to enable long-term endogenous FVIII production (AAV-based therapy). Roctavian is an adeno-associated virus 5 (AAV5) vector-based gene therapy designed to carry a functional copy of FVIII targeting liver cells for long-term expression of FVIII.

Extended half-life FVIII products reduce infusion frequency, improving adherence. Non-factor therapies like Emicizumab offer convenient subcutaneous prophylaxis and are effective for both inhibitor-positive and inhibitor-negative patients. Gene therapy, such as AAV-based approaches, has the potential to provide long-term endogenous FVIII production, reducing or eliminating the need for regular infusions, though challenges like durability and accessibility remain. Despite advancements, FVIII replacement therapy remains the standard for managing acute bleeds, surgical procedures, immune tolerance induction (ITI) in inhibitor-positive patients, and inhibitor management, ensuring its continued relevance in haemophilia A treatment.

Factor VIII Replacement Therapy

The management and prevention of bleeding in HA patients mainly involve FVIII replacement therapy, which includes the administration of exogenous FVIII to restore normal coagulation [8]. FVIII products are derived from plasma or produced using recombinant DNA technology.

Prophylactic FVIII replacement therapy, administered regularly to prevent bleeding episodes, has become the standard of care for patients with severe Haemophilia A. This approach has been shown to reduce the frequency of bleeding episodes, prevent joint damage, and improve overall quality of life. However, the development of inhibitors, which are neutralising antibodies against FVIII, remains a significant challenge in the management of Haemophilia A.

Inhibitor Development in Haemophilia A

Inhibitors are a major complication of FVIII replacement therapy, occurring in approximately one-third of patients with severe Haemophilia A. Inhibitors are polyclonal IgG antibodies that bind to FVIII, neutralising its coagulant activity and rendering replacement therapy ineffective. The development of inhibitors is influenced by both genetic and environmental factors. Genetic risk factors include the type of *F8* mutation, family history of inhibitors, and polymorphisms in immune regulatory genes. Environmental factors, such as the intensity of FVIII exposure and the presence of inflammation or infection, also play a role. One of the most important risk factors remains the type of FVIII used for the treatment.

The immune response to FVIII involves the recognition of FVIII as a foreign antigen by antigen-presenting cells (APCs), which present FVIII peptides to T cells. This triggers the activation of B cells and the production of anti-FVIII antibodies. The epitopes targeted by inhibitors are primarily located in the A2, A3, and C2 domains of FVIII, which are critical for its interaction with FIXa, phospholipids, and VWF.

The presence of inhibitors complicates the management of Haemophilia A, as patients no longer respond to standard FVIII replacement therapy. For patients with inhibitors experiencing bleeding episodes, alternative treatments known as bypassing agents are employed. These include recombinant activated factor VII (rFVIIa) and activated prothrombin complex concentrate (aPCC), which stimulate thrombin production without relying on FVIII. Immune tolerance induction (ITI), a treatment strategy aimed at eradicating inhibitors. This method involves administering high FVIII doses regularly over an extended period, ranging from months to years. Emicizumab facilitates significantly the prophylaxis in children with fewer infusions and allowing for subcutaneous administration. Therefore, during the emicizumab therapy, particularly in children, the incidence of spontaneous bleeding is minimised, leading to reduced exposure to FVIII. However, during

emicizumab therapy, prophylaxis primarily relies on this approach, with FVIII reserved only for managing acute bleeding episodes.

Although ITI proves successful in roughly 70% of cases, it is expensive and demands substantial dedication from both patients and their caregivers.

CHAPTER 2

AIMS

This PhD thesis aimed to gain insights into the inhibitor development against FVIII in Haemophilia A patients as well as illustrate the use of bioinformatics and computational approaches in the field of translational medicine and biomedical sciences. The thesis seeks:

1. To use a novel immunoglobulin G epitope mapping method to explore the factor VIII (FVIII)–specific epitope profile in the SIPPET cohort population and to develop an epitope mapping–based inhibitor prediction model.
2. To assess if FVIII-specific B-cell epitopes share sequence similarity with pathogen-derived linear B-cell epitopes.
3. To understand whether CpG sites that are differentially methylated in peripheral blood mononuclear cells (PBMCs) of HA patients could have a role in inhibitor development, to better understand the biological pathways that lead to inhibitor development.

CHAPTER 3

PREDICTING INHIBITOR DEVELOPMENT USING A RANDOM PEPTIDE PHAGE-DISPLAY LIBRARY APPROACH IN THE SIPPET COHORT

Shermarke Hassan[1,2*], Guido Baselli[3*], Luca Mollica[4], Riccardo L. Rossi[5], **Himani Chand**[1], A. El-Beshlawy[6], M. Elalfy[7], V. Ramanan[8], P. Eshghi[9], M. Karimi[10], Roberta Palla[1], Frits R. Rosendaal[2], Flora Peyvandi[1,11]

[1]Università degli Studi di Milano, Department of Pathophysiology and Transplantation, Milan, Italy
[2]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands
[3]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Translational Medicine, Department of Transfusion Medicine and Hematology, Milan, Italy
[4]Department of Medical Biotechnologies and Translational Medicine, L.I.T.A/University of Milan, Milan, Italy.
[5]Bioinformatics, Istituto Nazionale Genetica Molecolare "Romeo ed Enrica Invernizzi", 20122, Milan, Italy
[6]Pediatric Hematology Department, Cairo University Pediatric Hospital, Cairo, Egypt
[7]Faculty of Medicine, Ain Shams University - Department Pediatrics, Cairo, Egypt
[8]Jehangir Clinical Development Centre, Department of Hematology, Jehangir Hospital Premises, Pune, India
[9]Congenital Pediatric Hematologic Disorders Research Centre, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[10]Hematology Research Centre, Shiraz University of Medical Sciences, Shiraz, Iran
[11]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Angelo Bianchi Bonomi Hemophilia and Thrombosis Center, Milan, Italy

Blood Adv **(2024) 8 (11): 2880–2889.**
https://doi.org/10.1182/bloodadvances.2023011388

Abstract:

Inhibitor development is the most severe complication of haemophilia A (HA) care and is associated with increased morbidity and mortality. This study aimed to use a novel immunoglobulin G epitope mapping method to explore the factor VIII (FVIII)–specific epitope profile in the SIPPET cohort population and to develop an epitope mapping–based inhibitor prediction model. The population consisted of 122 previously untreated patients with severe HA who were followed up for 50 days of exposure to FVIII or 3 years, whichever occurred first. Sampling was performed before FVIII treatment and at the end of the follow-up. The outcome was inhibitor development. The FVIII epitope repertoire was assessed by means of a novel random peptide phage-display assay. A least absolute shrinkage and selection operator (LASSO) regression model and a random forest model were fitted on posttreatment sample data and validated in pretreatment sample data. The predictive performance of these models was assessed by the C-statistic and a calibration plot. We identified 27 775 peptides putatively directed against FVIII, which were used as input for the statistical models. The C-statistic of the LASSO and random forest models were good at 0.78 (95% confidence interval [CI], 0.69-0.86) and 0.80 (95% CI, 0.72-0.89). Model calibration of both models was moderately good. Two statistical models, developed on data from a novel random peptide phage display assay, were used to predict inhibitor development before exposure to exogenous FVIII. These models can be used to set up diagnostic tests that predict the risk of inhibitor development before starting treatment with FVIII.

Introduction:

Recent advances in the treatment of patients with haemophilia A (HA) have greatly improved clinical outcomes and quality of life. Nevertheless, one of the greatest treatment complications in severe HA is still the development of anti–factor VIII (anti-FVIII) alloantibodies that neutralise FVIII (also called inhibitors). At least one-third of patients treated with FVIII replacement therapy develop an inhibitor during the first 20 to 30 days of exposure to FVIII (also called exposure days or EDs),[8] making treatment with FVIII ineffective. This in turn leads to increased morbidity and mortality among these patients.[8]

This complication is the result of a multicausal immune response involving both patient- and treatment-related factors.[8] The type of FVIII product is one of the most important risk factors for inhibitor development, with the SIPPET randomised clinical trial showing that patients treated with recombinant FVIII (rFVIII) have an almost twofold higher risk of developing an inhibitor than those treated with plasma-derived FVIII (pdFVIII) products.[9] The pathophysiological mechanisms behind this increased immunogenicity remain unknown. Some plausible biological explanations have been postulated, such as the different posttranslational modifications caused by the use of different cell lines during the manufacturing process of rFVIII products and the protective role played by von Willebrand factor (VWF) in pdFVIII products.[10]

Mature FVIII consists of 6 major domains (A1, A2, B, A3, C1, and C2) and 3 acidic linking regions (a1, a2, and a3); A1-a1-A2-a2-B-a3-A3-C1-C2. The VWF-FVIII complex forms through a high-affinity interaction between the FVIII light chain and the VWF D′D3 domains.[11] FVIII is activated by limited proteolysis through thrombin cleavage of 3 peptide bonds at Arg391 (a1-A2 junction), Arg759 (a2-B junction), and Arg1708 (a3-A3 junction).[12] After thrombin cleavage, activated FVIII (without the B domain) is released from VWF and binds to phosphatidylserine on the extracellular surface of activated platelets.[13,14]

The anti-FVIII humoral immune response is highly polyclonal and consists primarily of immunoglobulin G (IgG) antibodies recognizing variable multiple epitopes among patients and even in the same patient over time.[15] Several studies have examined the immunogenicity of FVIII and the mechanisms underlying inhibitor development during treatment with FVIII.[10,16,17] The role of FVIII B-cell epitopes in inhibitor development has been previously investigated using different techniques. Specific regions in the A2 (region encompassing Arg484-Ile508[18]), A3 (Gln1778-Asp1840[19]), C1 (Lys2065-Trp2212[20] and residues 2063-2071[21]), and C2 (residues Glu2181-Val2243[22] as well as other residues[23-25]) FVIII domains were shown to be target domains for FVIII alloantibodies interaction using several methods including low resolution immunoprecipitation, western blotting,

and antibody neutralization assays,[15,26] as well as high resolution methods such as hydrogen-deuterium exchange mass spectrometry,[25] crystallographic studies,[27] surface plasmon resonance–based methods,[24] and phage display.[28-31]

In recent years, quantitative immunoproteomics has developed rapidly, offering high-throughput analyses at relatively low cost. The aim of this study was to use a novel high-throughput epitope mapping technique based on a random peptide phage-display method to explore the overall FVIII epitope profile and to develop an epitope–based inhibitor prediction model.

Methods:

Patient population

Study samples were obtained from patients enrolled in the SIPPET trial, which was designed to investigate the immunogenicity of different FVIII products in patients with severe HA who were previously untreated with any FVIII concentrates (PUPs) or minimally treated with blood components.[9] Samples from 122 patients were used for this study. These patients were treated with 8 different FVIII products (4 pdFVIII products and 4 rFVIII products). Inhibitor development was measured using the Bethesda assay with Nijmegen modification.[32] Thirty-nine of 122 individuals developed an inhibitor.

One sample of citrated plasma was collected at baseline before exposure to FVIII (pretreatment) and on sample at the end of the study (posttreatment). As previously described,[9] in inhibitor-positive patients, the end of the study was the time of inhibitor development. In inhibitor-negative patients, the study ended when the patient reached 50 EDs or after 3 years of follow-up, whichever occurred first.

Approval for this study was obtained from the medical ethics committee at each study center, and informed consent was obtained from all parents/guardians of patients.

Mimotipe Variation Analysis

Assay set-up

The total IgG epitope repertoire was assessed using mimotope-variation analysis (MVA), a phage display–based method (Protobios, Tallinn, Estonia) as described previously.[33] A combinatorial library of randomized linear 12-mer peptides fused to the pIII minor coat protein of M13 phages (Ph.D.-12, New England Biolabs, Ipswich, MA) was used according to the manufacturer's protocol. Two μL of plasma was incubated with 5 μL of phage library (~5×10^{10} phage particles; overnight at 4°C). The human IgG–captured phages were pulled down by protein G-coated magnetic beads (NEB, S1506S). Phage DNA was extracted, enriched and samples were barcoded by polymerase chain reaction amplification. Pooled samples were analyzed by Illumina sequencing (50-bp single end read, Genohub). The resulting DNA sequences were in silico translated to 12 amino acid (aa) long peptide sequences. To correct for differences in sequencing depth among the samples, the total count of each unique peptide per sample was normalized in its counts per 3 million. The resulting output consisted of a database of 12-mer peptides with varying degrees of apparent affinity for IgG antibodies. In the context of the assay, apparent affinity was defined as the frequency with which a 12-mer peptide was detected (ie, the peptide count). These peptides are often referred to in the literature as "mimotopes," due to the fact that they may mimic the structure of an epitope.

Two versions of the assay were performed, the standard MVA assay (described above) and a competition assay. In the MVA competition assay, the same FVIII products that were used to treat the patient (Alphanate [Grifols], Fanhdi [Grifols], Emoclot [Kedrion Biopharma], Factane [LFB], Advate [Baxalta], Kogenate FS [Bayer AG], ReFacto AF [Pfizer], or Recombinate [Baxalta]) were also used to precondition study samples before competition analyses. In detail, respective FVIII products (final concentration, 3 uM) were incubated with 2 μL of plasma for 2 hours at room temperature before proceeding with the MVA assay as described above.

Removal of TUPs

One issue in conducting phage display experiments is the presence of so-called target-unrelated peptides (TUPs). These are false-positive results caused by selection-related TUPs, which are peptides binding to materials and reagents used in the assay (for example, plastic surfaces and albumin), or propagation-related TUPs, caused by faster propagation of some phage clones, resulting in a higher peptide count for some peptides. To minimize the effect of these TUPs, we removed all peptides that were predicted to be TUPs using the SAROTUP software tool.[34] Using this tool, known TUPs were filtered out exploiting the TUPscan and the mimosearch algorithms. Peptides with a high likelihood ($P > .8$) to bind to polystyrene, as assessed by the PSBinder algorithm, were also filtered out.

Quality control using intra-assay and inter-assay replicates

To improve the assay signal to noise ratio, we focused on the most abundant peptides in the data set. To establish an optimal abundance threshold, we relied on technical replicates of a control sample within the same MVA plate (intra-assay) and across the different assay plates (inter-assay). Each technical replicate was compared with its intra-assay and inter-assay littermates using each possible filtering threshold. As similarity metric, we accounted for the proportion of peptides in the data set found in only 1 replicate at each possible threshold.

Identification of FVIII mimotopes

To find peptides identified using the MVA competition assay that bind selectively to FVIII-specific antibodies (FVIII mimotopes), the posttreatment sample was analyzed twice, once using the standard MVA assay and once using the MVA competition assay. FVIII mimotopes were defined as peptides that were present in the posttreatment sample in which the standard MVA assay was performed but not in the posttreatment sample in which the MVA competition assay was performed. Consequently, the abundance of each peptide in the standard MVA assay vs the MVA competition assay was

compared using the Fisher exact test. Adjustment for multiple testing was done using the Bonferroni method. An adjusted *P* value < .05 was considered statistically significant. Only peptides significantly underrepresented in the MVA competition assay samples when compared with the standard MVA assay samples were considered to be FVIII mimotopes and used for further analyses.

Clustering workflow

Each FVIII epitope can be conceptualized as being represented by multiple peptide sequences, each containing the antibody-binding motif. Therefore, the Hammock algorithm was used to cluster peptides based on sequence similarity before further analyses.[35] A complete linkage clustering algorithm was used for the initial clustering step. Cluster iterative merging was based on 3 iterations, maximum alignment length was set at 150% of that of the input peptide, and 5% of the initial clusters were used as seeds for cluster merging. Applying the algorithm resulted in clusters of highly similar peptides. For each cluster, a consensus motif was generated based on the multiple sequence alignment of the sequences. Each consensus motif can be interpreted as representing an epitope motif. Highly conserved residues (>60%) were denoted with an uppercase symbol, whereas moderately conserved residues (30%-60%) were denoted with a lower case symbol. Columns in the multiple sequence alignment in which no single residue had a prevalence of >30% were denoted with "x." The total peptide count of each cluster was calculated as the sum of the count of each peptide included in a cluster. Clusters with an epitope motif that contained <4 conserved residues were filtered out from the data set.

Alignment of epitope motifs to FVIII

Epitope motifs from the remaining clusters were then aligned to the linear amino acid sequence of FVIII (UniProt database ID P00451). Local pairwise alignment using the Smith-Waterman algorithm was used (R package; "Biostrings" version 2.40.2). The degree to which a given residue on FVIII was surface accessible was calculated using the GETAREA algorithm[36] using as input the crystal structure of a B-domain–deleted FVIII molecule[37] (PDB ID 3CDZ). Based on the literature,[38] a relative solvent accessibility of ≥20% was used as a cutoff for defining whether a residue was buried.

Statistical analysis

Descriptive statistics

For the descriptive analyses, data were summarized using the mean and standard deviation, or median and interquartile range, or as proportions.

Prediction modeling

To find biomarkers that were able to predict inhibitor development before the start of FVIII therapy, 2 statistical prediction models were fitted to the data. Both models were trained on data generated from the posttreatment samples and were validated on data generated from the pretreatment samples. Firstly, a logistic regression model using L1 regularization (R package; "glmnet" version 4.1.7), also called least absolute shrinkage and selection operator (LASSO) logistic regression, was evaluated using all clusters as the input. Leave-one-out cross-validation was used to select the optimal value for the regularization parameter. All clusters were used as the input for the model, the variables were centered and scaled before model fitting. Secondly. A random forest model (R package; "randomForest" version 4.7-1.1) was evaluated using all clusters as the input. Values for the number of trees in the model and the number of variables at each split were selected by fitting models with different values of these parameters and then selecting the parameter value that minimized the out-of-bag error rate.

Evaluation of predictive performance

Predictive performance of these models was evaluated in 2 ways. Firstly, we assessed the degree to which a model could discriminate between patients with and without inhibitors, using a receiver operating characteristic (ROC) curve. Secondly, we evaluated the degree to which the predicted cumulative incidence of inhibitor development matched the observed cumulative incidence using a calibration plot.

Selection of important clusters

To identify the clusters that were most important for inhibitor development, we first ranked the importance of each variable in the random forest model for model prediction by permutation feature importance. We then selected all the clusters with a feature importance score in the 90th percentile. From this set, we then selected all the clusters that were also present (ie, with a nonzero model coefficient) in the final LASSO regression model. We then generated descriptive statistics for this final set of clusters.

Approval for this study was obtained from the medical ethics committee at each study center, and informed consent was obtained from all parents/guardians of patients.

Results:

Mimotope Variation Analysis

Of the 122 previously untreated patients with haemophilia selected from the SIPPET study cohort for this analysis, 39 patients developed an inhibitor during follow-up (Table 1). Inhibitor-positive patients were slightly more likely to have a null mutation in the *F8* gene and slightly more likely to use a recombinant FVIII product than inhibitor-negative patients, although these changes were not significant (Table 1).

Table 1. Patient characteristics.

	Inhibitor-negative (N=83)	**Inhibitor-positive (N=39)**
Age at first treatment (months)		
Mean (SD)	20.7 (17.6)	17.6 (13.0)
Family history of inhibitor development		
No	67 (80.7%)	31 (79.5%)
Yes	12 (14.5%)	4 (10.3%)
Unknown	4 (4.8%)	4 (10.3%)
***F8* gene mutation (null vs. non-null)**		
Non-null mutation	15 (18.1%)	1 (2.6%)
Null mutation	67 (80.7%)	35 (89.7%)
Unknown	1 (1.2%)	3 (7.7%)
***F8* gene mutation (detailed)**		
Frameshift mutation	10 (12.0%)	7 (17.9%)
Intron 1 inversion	4 (4.8%)	0 (0%)
Intron 22 inversion	32 (38.6%)	20 (51.3%)
Large deletion	5 (6.0%)	2 (5.1%)
Nonsense mutation	16 (19.3%)	6 (15.4%)
Missense mutation	8 (9.6%)	0 (0%)
Splice site mutation	4 (4.8%)	1 (2.6%)
Non-null mutation, type unknown	3 (3.6%)	0 (0%)
Unknown	1 (1.2%)	3 (7.7%)
FVIII product (type)		
Recombinant FVIII product	39 (47.0%)	22 (56.4%)
Plasma-derived FVIII product	44 (53.0%)	17 (43.6%)

SD, standard deviation.

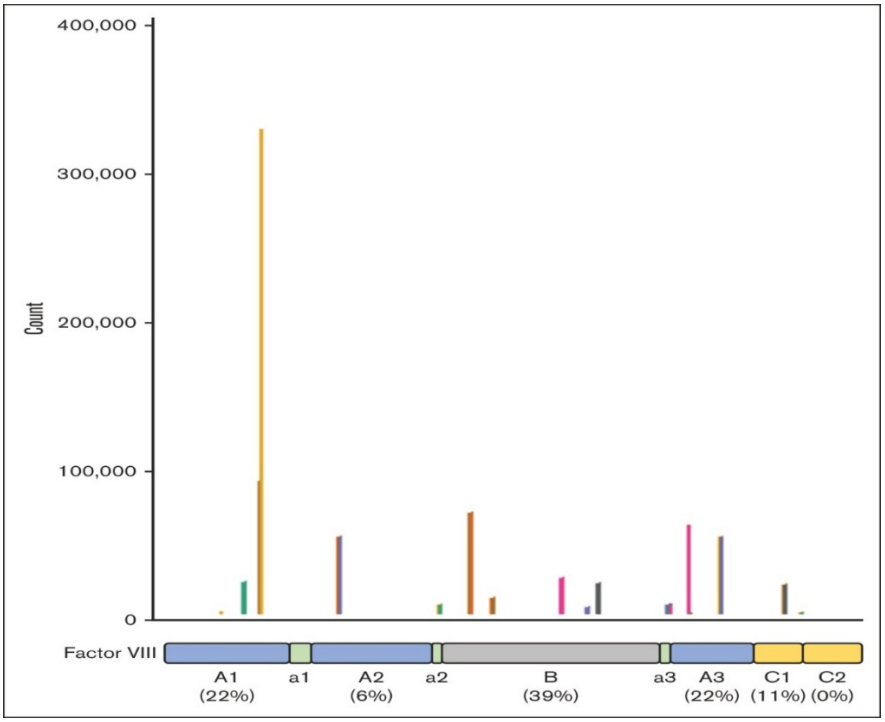
The mean number of unique peptides generated from each patient's posttreatment samples was 356,365. After removing potential TUPs, the mean number of unique peptides generated for each

patient decreased to 313,340. As shown in Supplemental Figure 1, both intra-assay and inter-assay reproducibility started dropping when considering peptides with abundance <250 reads. Therefore, only peptides found with abundance >250 in at least 1 sample were considered for further analyses. This yielded 27,775 unique peptides that were identified as being FVIII mimotopes, with a median number of 266 (range, 4-1101) peptides per patient. These 27,775 peptides were then clustered (as described in "Methods"), which resulted in 223 clusters.

Location of clusters on FVIII

Using pairwise local alignment, 18 of 223 clusters were mapped with acceptable alignment to the linear sequence of FVIII. Most of these clusters were mapped against the B domain (39%; Figure 1). Of the 10 clusters that were mapped to parts of the linear sequence for which information on surface accessibility was available, 9 were aligned to positions on FVIII that were partially surface accessible (Table 2). All B-domain–aligned peptide clusters had nonzero mean peptide counts in patients using a B-domain–deleted product. Furthermore, of the 7 peptide clusters that were aligned to the B domain, 4 peptide clusters had a higher mean peptide count among patients receiving treatment with full-length FVIII than among patients receiving B-domain–deleted FVIII. (Table 3).

Figure 1. Alignment of epitope motifs on the linear sequence of FVIII.



The plot shows the alignment of 18 epitope motifs on the linear sequence of FVIII. The x-axis represents the linear sequence of FVIII, from position 20 to 2351. The y-axis shows a count for each position of the FVIII sequence. The count for each position was defined as the weighted sum of each cluster whose epitope motif

was mapped to that position, using the peptide count of each cluster as weights. For example, if 2 clusters with a peptide count of 20 and 10, respectively, were mapped to a given position, then the total score for that position would be cluster 1 ∗ 20 + cluster 2 ∗ 10 = 30. The number of epitope motifs mapped to each domain, as a proportion of all 18 aligned epitope motifs, is shown at the bottom of the figure.

Table 2. Linear alignment of mimotope clusters on FVIII

Mimotope core motif sequence	FVIII sequence	Alignment	Residue number, start	Residue number, end	Domain	No. of surface-accessible residues*
nxRRPfflnsg	LNSG	LNSG	187	190	A1	0
glggLi	LPGLI	L?GLI	261	265	A1	3
dPxqtll	QTLL	QTLL	316	319	A1	2
GLGqLL	LGQFL	LGQ?L	322	326	A1	4
nqkms	NQIMS	NQ?MS	583	587	A2	5
pdtppSxp	PPSMP	PPS?P	925	929	B	NA†
txxKtxIxTxt	TNRKTHI	T??KT?I	1028	1034	B	NA†
Ppdixspp	PPDAQNP	PPD???P	1105	1111	B	NA†
KVFRxp	KQFRLP	K?FR?P	1335	1340	B	NA†
VFRlpxtxt	FRLP	FRLP	1337	1340	B	NA†
TxltRtls	LTRVL	LTR?L	1423	1427	B	NA†
qNLsl	NLSL	NLSL	1461	1464	B	NA†
ydkadnerarlg	YDEDENQSPR	YD???N???R	1699	1708	other	NA†
dRxeLNmxxxl	RGELN	R?ELN	1768	1772	A3	3
lNEvLv	LNEHL	LNE?L	1771	1775	A3	3
hTnln	HTNTLN	HTN-LN	1878	1883	A3	4
kxDiLaxl	KVDLLA	K?D?LA	2091	2096	C1	3
KxDssGP	DSSG	DSSG	2150	2153	C1	3

SD, standard deviation.

∗The number of surface-accessible residues for each alignment was calculated using the GETAREA algorithm, as described in "Methods."

†No information on surface-accessibility of B domain and some adjacent residues.

Table 3. Mean peptide count (SD) of peptide clusters that were aligned to the B domain, stratified by use of B-domain–deleted or full-length FVIII.

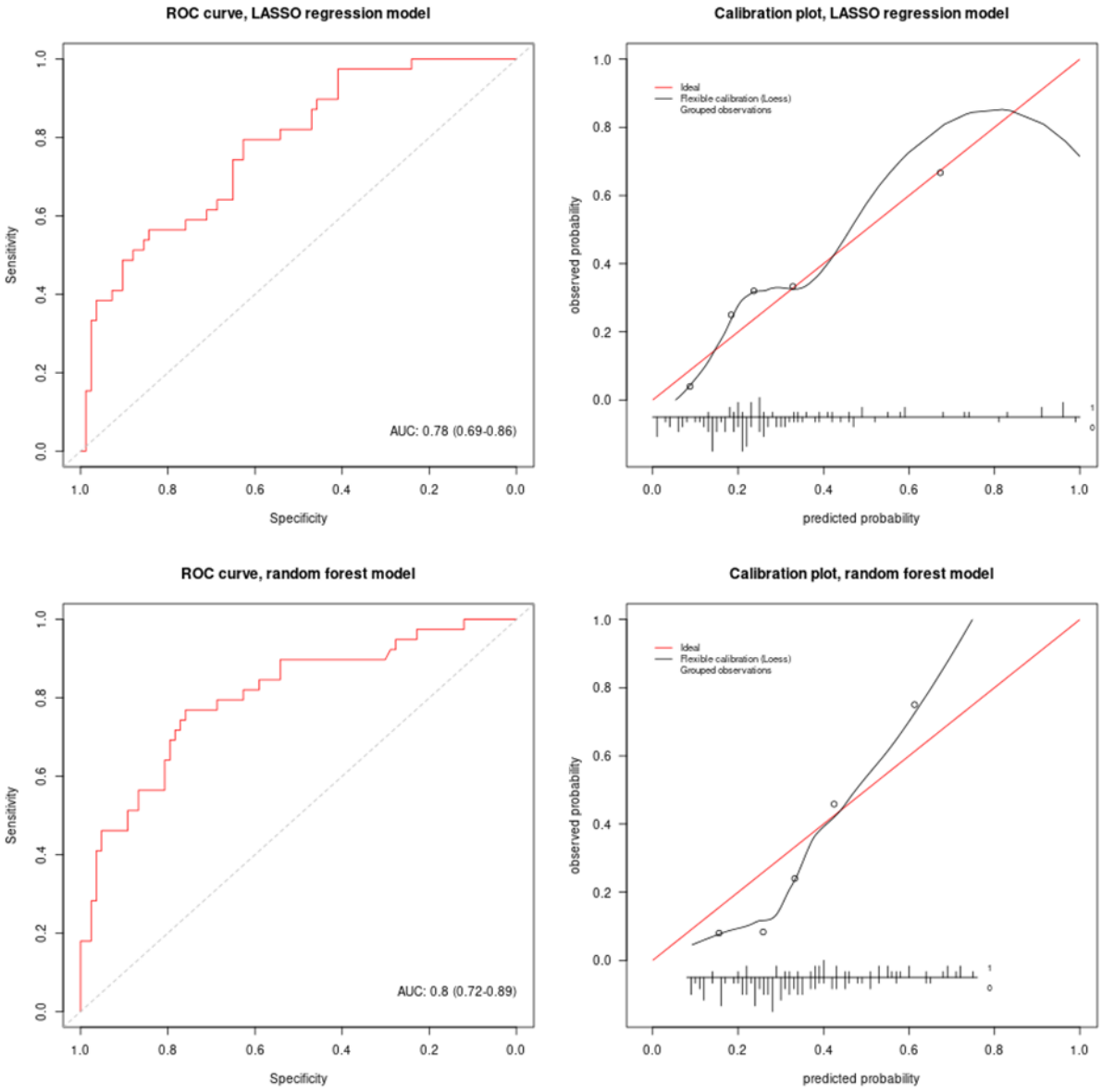
	B-domain–deleted FVIII (n = 7)	Full-length FVIII (n = 115)
qNLsl	175 (56.0)	349 (543)
KVFRxp	405 (627)	383 (948)
Ppdixspp	63.6 (89.0)	59.2 (288)
TxltRtls	119 (248)	1 160 (10 800)
txxKtxIxTxt	51.1 (63.3)	304 (1790)
VFRlpxtxt	157 (391)	395 (1870)
pdtppSxp	27.7 (27.4)	291 (716)

SD, standard deviation.

Predictive value of clusters

Next, we constructed 2 statistical prediction models to assess the degree to which the presence of these clusters in patients' samples were able to predict inhibitor development. First, a LASSO logistic regression model was fitted to all 223 clusters in the posttreatment patient samples. The fitted model was then used to predict inhibitor development using the pretreatment patient samples. The C-statistic was 0.78 (95% confidence interval, 0.69-0.86; Figure 2A). Model calibration was good because the cumulative incidence of inhibitor development predicted by the model was roughly in line with the observed incidence across the entire risk range (Figure 2C). Next, a random forest model was fitted to all 223 clusters in the posttreatment patient samples. The fitted model was then used to predict inhibitor development using the pretreatment patient samples. The C-statistic was 0.80 (95% confidence interval, 0.72-0.89; Figure 2B). Model calibration was moderate, due to the model somewhat overpredicting the observed cumulative incidence of inhibitor development in the lower-risk range (Figure 2D).

Figure 2. Evaluation of the degree to which the logistic regression model and random forest classifier model can predict inhibitor development.



Model discrimination (ie, the degree to which a model assigns a higher risk to an inhibitor-positive patient vs an inhibitor-negative patient) was assessed by plotting ROC curves and by calculating the area under the curve (AUC). The AUC varies between 0.5 (no discrimination) to 1 (perfect discrimination). (A-B) The ROC curves of the LASSO logistic regression model and the random forest model are shown in (A) and (B), respectively. Model calibration (ie, the degree to which the predicted cumulative incidence of inhibitor development matched the observed cumulative incidence) was assessed using a calibration plot. For each quintile of predicted cumulative incidence, we plotted the mean predicted cumulative incidence of inhibitor development in a group against the observed cumulative incidence of inhibitor development in that group. In addition, we plotted a LOESS (locally estimated scatterplot smoothing) line in the same figure to assess model calibration across the full risk range. Ideally, all points should lie exactly on the diagonal line (which represents perfect

agreement between predicted and observed values). (C-D) The calibration plots of the LASSO logistic regression model and the random forest model are shown in (C) and (D), respectively.

There were 12 clusters that had a feature importance score in the 90th percentile in the random forest model and were also part of the final LASSO logistic regression model (Table 4). Of these 12 clusters, only 2 mapped with good alignment to the linear sequence of FVIII. These 2 clusters were mapped to the A2 and A3 domain (Table 4). Ten of 12 clusters had a higher peptide count in inhibitor-positive patients than in inhibitor-negative patients (Table 4). No clear differences were seen in the mean peptide counts of the clusters when measured in patients treated with plasma-derived FVIII vs patients treated with rFVIII. (Table 5).

Table 4. Peptide clusters were used as predictors in both the logistic regression model and the random forest classifier model.

ore motif	Mean peptide count in INH^- group	Mean peptide count in INH^+ group	Fold change	FVIII domain	Number of unique peptides in cluster (%)	Peptide count of cluster (%)
kxPxstw	133	1826	13.70	-	24 (0.09%)	55 764 (0.16%)
hntMels	38	119	3.10	-	11 (0.04%)	20 460 (0.06%)
hTnln	151	454	3.01	A3	48 (0.17%)	66 660 (0.19%)
nqkms	140	365	2.60	A2	31 (0.11%)	66 660 (0.19%)
dxxYxlxm	438	1114	2.54	-	34 (0.12%)	63 368 (0.18%)
Yvntxxxt	193	475	2.46	-	11 (0.04%)	25 499 (0.07%)
LtqM	159	302	1.90	-	31 (0.11%)	61 253 (0.17%)
pQyxnxxk	454	738	1.63	-	51 (0.18%)	52 743 (0.15%)
sxnKP	325	483	1.49	-	53 (0.19%)	52 537 (0.15%)
WDVpPxxxxt	301	445	1.48	-	21 (0.08%)	33 105 (0.09%)
KxxHyxk	459	459	1.00	-	15 (0.05%)	39 173 (0.11%)
qTAkfh	41	40	0.96	-	41 (0.15%)	48 928 (0.14%)

The total number of unique peptides was 27,775. Total peptide count was 35,452,858. The "-" indicates a lack of good alignment on the linear sequence of FVIII.

Table 5. Mean count of peptide clusters, stratified by inhibitor status, *F8* gene mutation, and FVIII product type.

	Inhibitor status				*F8* Gene mutation*		FVIII product type	
	Inhibitor-negative (n = 83)	**Inhibitor-positive, total (n = 39)**	**Inhibitor-positive, low-titer (n = 15)**	**Inhibitor-positive, high-titer (n = 24)**	**Non-null mutation (n = 16)**	**Null mutation (n = 102)**	**rFVIII product (n = 61)**	**pdFVIII product (n = 61)**
Mean peptide count (SD)								
kxPxstw	133 (119)	1830 (9810)	4360 (15800)	240 (337)	145 (119)	779 (6070)	1190 (7850)	164 (196)
hntMels	38.5 (44.8)	119 (266)	256 (398)	33.4 (28.7)	88.5 (264)	61.6 (139)	91.8 (212)	36.6 (61.5)
hTnln	151 (372)	454 (1540)	850 (2370)	206 (564)	295 (590)	247 (985)	350 (1260)	145 (345)
nqkms	140 (199)	365 (1240)	168 (139)	489 (1570)	159 (215)	223 (784)	267 (996)	157 (223)
dxxYxlxm	438 (1590)	1110 (3850)	451 (840)	1530 (4850)	332 (440)	717 (2770)	889 (3140)	420 (1750)
Yvntxxxt	193 (290)	475 (891)	448 (523)	492 (1070)	185 (167)	299 (615)	286 (532)	280 (608)
LtqM	159 (483)	302 (1000)	159 (316)	391 (1250)	61.3 (106)	234 (752)	187 (797)	222 (573)
pQyxnxxk	454 (1170)	738 (1890)	109 (157)	1130 (2340)	176 (189)	614 (1560)	683 (1540)	407 (1320)
sxnKP	325 (674)	483 (858)	717 (995)	337 (746)	208 (293)	413 (794)	459 (888)	293 (543)
WDVpPxxxxt	301 (464)	445 (1130)	277 (277)	549 (1430)	331 (468)	358 (791)	330 (520)	364 (917)
KxxHyxk	459 (3120)	459 (1880)	99.5 (155)	684 (2380)	65.6 (164)	532 (3030)	617 (3630)	301 (1510)
qTAkfh	41.5 (103)	39.7 (82.9)	39.3 (97.9)	40.0 (74.3)	88.9 (164)	34.3 (81.9)	44.8 (104)	37.0 (88.6)

SD, standard deviation.

∗Four patients were excluded from this analysis as their *F8* gene mutation was unknown.

Discussion:

We assessed the FVIII-specific epitope profile of 122 previously untreated patients with HA, using a novel random peptide phage-display assay. Our results show that the apparent FVIII-specific antibody response is highly polyclonal, with many different epitope motifs. Among the 18 epitope motifs that were mapped to the linear sequence of FVIII, most of the mimotopes aligned with A1, A3, and B domain sequences. Using information on the presence of these epitope motifs in patient samples, 2 statistical prediction models (developed on posttreatment samples and validated in pretreatment samples) were found to be predictive for inhibitor development.

Seven of the 18 epitope motifs (39%) with good alignment to FVIII were mapped to the B domain. It is important to note that the alignments were not confirmed in vitro (eg, by antibody-binding assays using FVIII proteins with mutations at the relevant residues). All B-domain–aligned peptide clusters had nonzero mean peptide counts in patients using a B-domain–deleted product. This provides evidence against the hypothesis that the motifs of these peptide clusters truly represent targets of antibodies that are highly specific to a linear epitope on the B domain. That being said, these results are based on very small numbers because there were only 7 patients using a B-domain–deleted FVIII product (all ReFacto).

Previous studies have suggested that antibodies against the B domain might be predominantly of the non-neutralizing type,[39-41] because the B domain is not essential for the role of FVIII in blood clotting and is cleaved off after FVIII is activated. We could not verify this in our data set because almost all epitope motifs that were most important for inhibitor prediction (Table 4) did not map well to the linear sequence of FVIII.

In the past, several studies have tried to develop models to predict inhibitor development. Most of these models were based on either clinical parameters (eg, the age of FVIII treatment initiation, the type of FVIII product used, or the intensity of the first treatment moment with FVIII) or genetic parameters (eg, family history of inhibitor development, the type of *F8* gene mutation, HLA type, or gene polymorphisms in immunoregulatory genes such as *IL10* or *CTLA4*).[42] Previously, a study in the SIPPET cohort also assessed the predictive value of the presence of non-neutralizing antibodies detected before treatment as part of a larger clinical prediction model for inhibitor development.[43] None of the aforementioned prediction models were able to accurately predict inhibitor development. In addition, some prediction models were only implementable after starting treatment with FVIII, due to the inclusion of treatment-related predictors (eg, information on treatment intensity can only be obtained after a couple of days of exposure to FVIII). This limits the applicability of these models as one would ideally want to have an idea about the risk of inhibitor development before FVIII treatment

is initiated so that certain types of treatment modalities (eg, exposure to exogenous FVIII) can be avoided. The models in this publication only use information on the pretreatment epitope repertoire of the patient and can therefore be used before FVIII treatment initiation.

The presence of peptides that specifically bind to anti-FVIII antibodies in samples taken before treatment with FVIII might seem unexpected at first glance. However, several studies have reported the presence of non-neutralizing anti-FVIII antibodies in healthy controls.[44] In addition, a previous study using pretreatment samples of the current cohort reported that ~10% of patients had measurable anti-FVIII antibodies.[45] This suggests that natural autoreactivity against endogenous FVIII is relatively common in patients as well as healthy controls. Another hypothesis could be that the detected antibodies were not initially directed against FVIII but were the result of previous exposure to a pathogen (eg, a bacteria or virus) that contained a similar epitope. This cross-reactivity of the antibody response has been previously reported in several autoimmune disorders.[46] Our results indicate that the presence of anti-FVIII antibodies before treatment with FVIII might be a risk factor for inhibitor development in a subset of patients.

Limitations:

This approach has some limitations. Firstly, it has been shown that only a handful of contact residues within an epitope make a significant contribution to antibody binding.[47] In this study, we tried to identify these residues by clustering highly similar FVIII mimotopes and generating a consensus motif. Using alanine walk mutational analysis, the study by Kahle et al[31] showed that there was reasonable agreement between a given consensus motif and the crucial binding residues of an epitope. Therefore, the consensus motifs derived from the multiple sequence alignment of each cluster of peptide sequences can, in theory, be considered to be potential epitope motifs. However, the accuracy of this approach is unknown, and further verification is needed to identify the exact residues involved in binding to an antibody.

Secondly, the final epitope motifs were mapped to FVIII by aligning the motifs to the linear sequence of FVIII. However, it has been reported that the majority of B-cell epitopes are conformational[48,49] (although the exact proportion of B-cell epitopes purported to be conformational is unknown). In this case, clustering based on sequence similarity might yield the correct conformational epitope motif, but the linear alignment procedure will produce faulty alignment. An alternative approach would involve mapping the epitope motifs to the 3-dimensional structure of FVIII, using an in-silico approach. However, a recent study that assessed a set of B-cell epitope prediction algorithms against a benchmark data set reported that all algorithms performed relatively poorly at mapping a potential

epitope to the right location on an antigen.[50] That being said, knowing the correct location of these putative B-cell epitopes is not needed if the goal is only to predict inhibitor development.

We removed all peptides that were predicted to be target unrelated (based on software exploiting publicly available repositories[34]) from the final peptide database. However, the residual impact of TUPs that were not removed from the database on the results is difficult to quantify. In addition, some peptides can bind to both elements of assay as well as an IgG antibody (ie, they can be classified as both target-unrelated and target-related peptides). By removing these peptides, we might have inadvertently also removed some important peptides from the initial database.

From the output of the assay, only peptides with a count >250 were selected; this resulted in a much smaller data set. The cutoff was based on the intra-assay and inter-assay replicability (supplemental Figure 1). It is possible that many peptides that were the target of a FVIII-specific antibody were removed in this step.

Lastly, our analysis of the immune response did not include nonpeptidic epitopes (such as the glycans present on the surface of FVIII). One difference between rFVIII and pdFVIII is in their respective glycosylation patterns.[51] Unfortunately, our approach does not allow for the assessment of the impact of differing glycosylation patterns on immunogenicity.

Conclusion:

Two statistical models, developed on data from a novel random peptide phage display assay, were used to predict inhibitor development before exposure to exogenous FVIII. These models can be used to set up diagnostic tests that predict the risk of inhibitor development before starting treatment with FVIII.

CHAPTER 4

MOLECULAR MIMICRY AND INHIBITOR DEVELOPMENT IN PATIENTS WITH SEVERE HAEMOPHILIA A

Shermarke Hassan[1,2], **Himani Chand[1]**, Guido Baselli[3], Luca Mollica[4], Riccardo L. Rossi[5], A. El-Beshlawy[6], M. Elalfy[7], V. Ramanan[8], P. Eshghi[9], M. Karimi[10], Roberta Palla[1], Flora Peyvandi[1,11]

[1]Università degli Studi di Milano, Department of Pathophysiology and Transplantation, Milan, Italy
[2]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands
[3]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Translational Medicine, Department of Transfusion Medicine and Hematology, Milan, Italy
[4]Department of Medical Biotechnologies and Translational Medicine, L.I.T.A/University of Milan, Milan, Italy.
[5]Bioinformatics, Istituto Nazionale Genetica Molecolare "Romeo ed Enrica Invernizzi", 20122, Milan, Italy
[6]Pediatric Hematology Department, Cairo University Pediatric Hospital, Cairo, Egypt
[7]Faculty of Medicine, Ain Shams University – Department Pediatrics, Cairo, Egypt
[8]Jehangir Clinical Development Centre, Department of Hematology, Jehangir Hospital Premises, Pune, India
[9]Congenital Pediatric Hematologic Disorders Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[10]Hematology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[11]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Angelo Bianchi Bonomi Hemophilia and Thrombosis Center, Milan, Italy

In preparation for submission

Abstract:

Previous exposure to a pathogen may induce inhibitor development during later exposure to FVIII due to sequence similarity between the organism and factor VIII (FVIII). Aim of this study is to assess if FVIII-specific B-cell epitopes share sequence similarity with pathogen-derived linear B-cell epitopes.

The population consisted of 39 inhibitor-positive and 83 inhibitor-negative patients with severe haemophilia A that were followed-up for 50 days of exposure to FVIII. A random peptide phage-display library assay was used to identify a set of 12-mer linear peptide sequences with high affinity for FVIII-specific antibodies (i.e. FVIII mimotopes). Using a custom method, these FVIII mimotopes were screened against pathogen-derived linear B-cell epitopes reported in the Immune Epitope Database (IEDB). The presence of a common sequence motif among FVIII mimotopes was assessed using the MEME algorithm and their predictive value was assessed using the C-statistic. No B-cell epitope was significantly overrepresented (in terms of alignment) among FVIII mimotopes. The most common pathogens from which the top 10 linear B-cell epitopes were derived were human gammaherpesvirus 4 (3x), human herpesvirus 5, human betaherpesvirus 6B, Plasmodium vivax, Dengue virus type 2, hepatitis B virus, human T-cell leukaemia virus type I and trypanosoma cruzi. No strong epitope motif was found among mimotopes aligned to the top 10 B-cell epitopes and their presence in pre-treatment samples was not strongly predictive of inhibitor development (C-statistic: 0.58, 95%CI: 0.47-0.70).

There was no evidence of molecular mimicry between a set of FVIII mimotopes and pathogen-derived linear B-cell epitopes reported in the IEDB.

Introduction:

Recent advances in the treatment of patients with haemophilia A (HA) have greatly improved clinical outcomes and quality of life. Nevertheless, one of the greatest treatment complications in severe HA is still the development of anti-factor VIII (FVIII) alloantibodies that neutralize FVIII (also called inhibitors). At least one third of patients treated with FVIII replacement therapy develop an inhibitor during the first 20-30 days of exposure to FVIII (EDs)[8], making treatment with FVIII ineffective. This in turn leads to increased morbidity and mortality among these patients.[8]

This complication is the result of a multi-causal immune response involving both patient- and treatment-related factors. [8] The type of FVIII product is one of the most important risk factors for inhibitor development, with the SIPPET randomised clinical trial showing that patients treated with recombinant FVIII (rFVIII) have an almost twofold higher risk of developing an inhibitor than those treated with plasma-derived FVIII (pdFVIII) products.[9] The pathophysiological mechanism behind this increased immunogenicity remains unknown. Some plausible biological explanations have been postulated, such as the different post-translational modifications caused by the use of different cell lines during the manufacturing process of rFVIII products and the protective role played by proteins such as Von Willebrand factor (VWF) in pdFVIII products. [10]

FVIII seems to be more immunogenic than other biologicals[52] , even compared to other clotting factor products such as factor IX. A previous study showed that a proportion of patients with severe haemophilia A already have circulating anti-FVIII antibodies before exposure to exogenous FVIII.[45] Furthermore, a proportion of healthy patients also seem to develop FVIII-specific antibodies.[44] This high level of immunogenicity cannot be explained by already known genetic- and treatment-related factors.[7]

It has been reported that exposure to certain infectious agents is an important risk factor for the development of certain autoimmune disorders. (e.g. Guillain-Barre syndrome, multiple sclerosis, diabetes type 1 and several rheumatic diseases).[46] One important mechanism by which these infectious agents may elicit an autoimmune response is due to sequence similarity between antigens present in an organism and self-antigens.

Up to now, no studies have assessed the phenomenon of molecular mimicry in patients with haemophilia. Therefore, the aim of this study was to use a novel high-throughput epitope mapping technique based on a random peptide phage-display method in order to assess if previous exposure to infectious agents increases the risk of inhibitor development in patients with haemophilia.

Methods:

Patient population

Study samples were obtained from patients enrolled in the SIPPET trial, which was designed to investigate the immunogenicity of different FVIII products in patients with severe haemophilia A who were previously untreated with any FVIII concentrates (PUPs) or minimally treated with blood components.[9] Samples from 122 patients were used for this study. These patients were treated with 8 different FVIII products (4 pdFVIII products and 4 rFVIII products). Inhibitor development was measured using the Bethesda assay with Nijmegen modification.[32] Thirty-nine out of 122 individuals developed an inhibitor.

One sample of citrated plasma was collected at baseline (T0) and two samples at the end of the study (EOS). As previously described[9], in inhibitor-positive patients the end of the study was the time of inhibitor development. In inhibitor-negative patients the study ended when the patient reached 50 EDs or after three years of follow-up (whichever came first).

Approval for this study was obtained from the medical ethics committee at each study centre and informed consent was obtained from all parents/guardians of patients.

Mimotope-variation analysis

Assay set-up

The total IgG epitope repertoire was assessed using mimotope-variation analysis (MVA), a next generation phage display method. (Protobios, Tallinn). [33] MVA was conducted as previously described. Briefly, 2 µl of plasma was incubated with 5 µl of phage library (~5×10^{10} phage particles, derivative of Ph.D.-12, NEB, UK) overnight at +4 °C. The human immunoglobulin G (IgG)-captured phages were pulled down by protein G-coated magnetic beads (NEB, S1506S). Phage DNA was extracted, enriched and samples were barcoded by PCR amplification. Pooled samples were analysed by Illumina sequencing (50-bp single end read, Genohub, USA). The resulting DNA sequences were *in silico* translated to 12 amino acid (aa) long peptide sequences. To correct for differences in sequencing depth among the samples, the total count of each unique peptide sequence per sample was normalized in its counts per three million.

The resulting output consisted of a database of 12-mer peptide sequences with varying degrees of affinity for IgG antibodies. These peptide sequences are often referred to in the literature as "mimotopes", since they may mimic the true epitope of an antibody.

Two versions of the assay were performed, the standard MVA assay (described above) and a competition assay. In the MVA competition assay, different factor VIII products (Alphanate (Grifols),

Fanhdi (Grifols), Emoclot (Kedrion Biopharma), Factane (LFB), Advate (Baxalta), Kogenate FS (Bayer AG), ReFacto AF (Pfizer), Recombinate (Baxalta)) were used to precondition study samples before competition analyses. In detail, respective FVIII products (final concentration: 3 uM) were incubated with 2 µl of plasma for 2 hours at room temperature before proceeding with the MVA assay as described above.

Filtering steps

Removing peptide sequences with low counts

Firstly, peptide sequences with a count of <20 were filtered out before any further analyses.

Removal of target unrelated peptides (TUPs)

Next, target-unrelated peptides (TUPs) were removed from the dataset. These are false-positive results caused by peptide sequences binding to materials and reagents used in the assay (for example, plastic surfaces, albumin), or by faster propagation of some phage clones, resulting in a higher peptide count for some peptide sequences. To minimize the effect of these TUPs, we removed all peptide sequences that were predicted to be TUPs using the SAROTUP software.[34] Briefly, known TUPs were filtered out exploiting the TUPscan and the mimosearch algorithms. Peptides with a high likelihood (P>0.8) to bind to polystyrene, as assessed by the PSBinder algorithm, were also filtered out.

Removal of peptide sequences present in negative controls

Samples from three controls were also analysed. Controls were patients with haemophilia that were not ELISA-positive for FVIII-specific antibodies during the first 50 days of exposure to FVIII. Peptide sequences found in samples of these patients were removed from the entire dataset.

Identification of peptide sequences with high affinity for FVIII specific antibodies (FVIII mimotope sequences)

We then selected the peptide sequences with a positive count in the end-of-study sample in which the standard MVA assay was performed and a count of zero in the end-of-study sample in which the MVA competition assay was performed (which was depleted of FVIII-specific antibodies) for further analyses. For each patient, we then selected the 5 remaining peptide sequences with the highest count in the end-of-study sample in which the standard MVA assay was performed. These sequences will be referred to as FVIII mimotope sequences.

Identification of FVIII mimotope sequences with good alignment to FVIII

These FVIII mimotope sequences were then aligned (ungapped alignment) against the linear sequence of a B-domain deleted FVIII molecule [37] (PDB ID: 3CDZ) using standard local pairwise alignment (using the Smith-Waterman algorithm) to identify FVIII mimotope sequences with high similarity to the linear sequence of FVIII. The degree to which a given residue on FVIII was surface accessible was calculated using the GETAREA algorithm [36] using as input the crystal structure of the aforementioned B-domain deleted FVIII molecule [37] (PDB ID: 3CDZ). Based on the literature[22], a relative solvent accessibility of ≥ 20% was used a cut-off for defining whether a residue was buried or not. FVIII mimotope sequences that were aligned to positions on FVIII that were completely buried (i.e. surface inaccessible) were removed from the analysis, leaving only FVIII mimotope sequences mapped to positions on FVIII with at least one surface accessible residue.

Construction of B-cell epitope database

From the Immune Epitope Database, which is a repository for data generated from experimental studies on B- and T-cell epitopes [53], we extracted all linear B-cell epitopes derived from antigens on bacteria, viruses, helminths, protozoan parasites or fungi. (extracted March 3rd, 2021) Discontinuous B-cell epitopes were not included.

Screening method

We then aimed to identify epitope motifs that were present in both the FVIII mimotope sequences and the linear B-cell epitopes present in the IEDB database. One approach could be to first cluster the FVIII mimotope sequences based on sequence similarity[24]. Each cluster would then consist of a multiple sequence alignment of highly similar FVIII mimotope sequences. A consensus sequence (which can be interpreted as the epitope motif) could then be generated from each cluster, and the IEDB linear B-cell epitope database could then be searched for these motifs. This approach hinges upon first identifying clusters that yield a clear consensus sequence, which is not always the case.
Therefore, we tried to identify epitope motifs that were present in both the FVIII mimotope sequences and the linear B-cell epitopes present in the IEDB database using an alternative approach. (also described in Figure 1)
First, each linear B-cell epitope in the IEDB database was fragmented into every possible 12-mer fragment using a sliding window approach. (Figure 1A) Next, each FVIII mimotope sequence was aligned against each IEDB B-cell epitope fragment and the number of matching residues was calculated. The same was also done a for a set of randomly generated 12-mer peptide sequences.

(Figure 1B) For each IEDB B-cell epitope fragment, we then calculated the total number of FVIII mimotope sequences with at least 4 matching residues to that IEDB B-cell epitope fragment and the total number of randomly generated sequences with at least 4 matching residues to that IEDB B-cell epitope fragment. (Figure 1C) Fisher's exact test was used to determine if the proportion of peptide sequences with at least 4 matches to an IEDB B-cell epitope fragment was significantly higher among the FVIII mimotope sequences compared to the randomly generated peptide sequences. Bonferroni's correction was applied to correct for the multiple testing problem. (Figure 1D) We then selected the top 10 IEDB B-cell epitope fragments for further analysis.

Figure 1. Mapping strategy used to identify linear B-cell epitopes derived from the IEDB sharing sequence similarity with FVIII mimotope sequences derived from a phage display experiment.

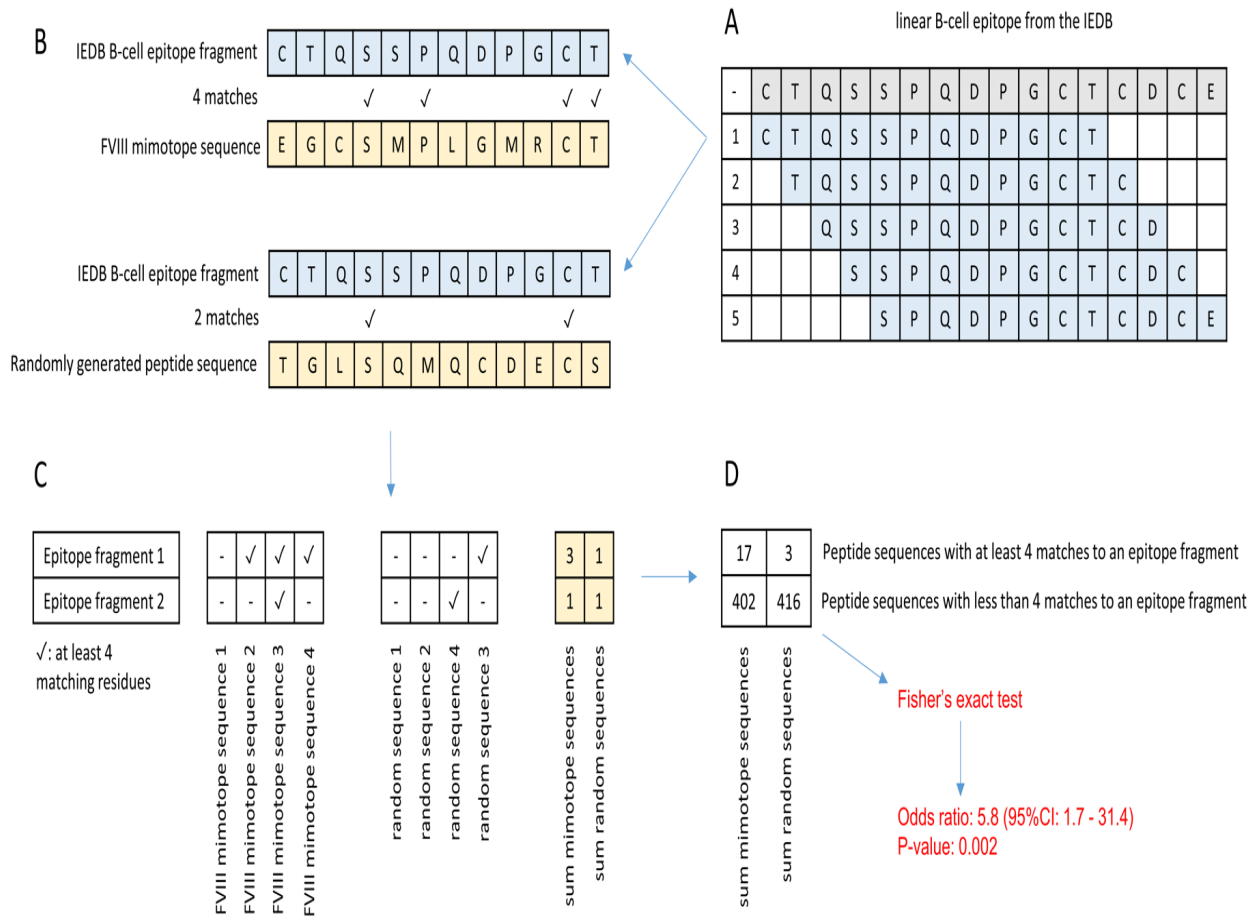


Figure legend: Figure 1 shows the mapping strategy used to identify linear B-cell epitopes derived from antigens on bacteria, viruses, helminths, protozoan parasites or fungi found in the Immune Epitope Database (IEDB) sharing significant sequence similarity with a set of 12-mer linear peptide sequences derived from a phage display experiment that are known to be bound by FVIII-specific antibodies (i.e. FVIII mimotope sequences). First, each linear B-cell epitope in the IEDB database was fragmented into every possible 12-mer

fragment using a sliding window approach. (A) Next, each FVIII mimotope sequence was aligned against each IEDB epitope fragment and the number of matching residues was calculated. The same was also done a for a set of randomly generated 12-mer peptide sequences. (B) For each IEDB epitope fragment, we then calculated the total number of FVIII mimotopes sequences/randomly generated sequences with at least 4 matching residues to that IEDB epitope fragment. (C) Fisher's exact test was used to determine if the proportion of peptide sequences with at least 4 matches to an IEDB epitope fragment was significantly higher among the FVIII mimotope sequences compared to the randomly generated peptide sequences. Bonferroni's correction was applied to correct for the multiple testing problem. (D)

Generation of the epitope motif

Next, for each of these 10 IEDB B-cell epitope fragments, a peptide motif was generated from the set of peptide sequences that had least 4 matches to that epitope fragment using the MEME algorithm. [54]

Evaluating the predictive value of selected FVIII mimotope sequences

Next, we assessed if each set of FVIII mimotope sequences mapped to a given B-cell epitope fragment could predict inhibitor development. This was done in two ways.

Firstly, we fitted a logistic regression model to model the relationship between the pre-treatment sample count of the FVIII mimotope sequences mapped to the top 10 B-cell epitopes and the risk of inhibitor development. Separate models were fitted for each FVIII mimotope sequence and the predictive value of the models was generated by calculating the C-statistic of each model.

Secondly, to identify the FVIII mimotope sequences that were most important for predicting inhibitor development, we fitted a model to the data consisting of all FVIII mimotope sequences combined. Variable selection was done by fitting a logistic regression model with L2 regularization (also called "ridge regression") to the data. The final value of the regularization parameter was chosen using k-fold cross-validation. The model was first fitted to the post-treatment samples and then validated on the pre-treatment samples. The predictive value of the model was generated by calculating the C-statistic.

Visualization of alignment location of mimotope sequences

Lastly, the location of all mimotope sequences that mapped to any of the top 10 B-cell epitope fragments that also had good alignment against FVIII (defined as having at least 4 matching residues against the linear sequence) was shown visually using VMD.

Results:

Patient characteristics

For this analysis, 39 patients who were inhibitor-positive were selected. The mean age of this group at the time of initiation of treatment with FVIII was 17.6 months (SD: 13.0). Most patients (89.7%) had a non-null mutation in the *F8* gene. Among patients with a non-null mutation, roughly half of them had an intron-22 inversion. Furthermore, roughly half of patients (56.4%) were treated with recombinant FVIII while the other group (43.6%) was treated with plasma-derived FVIII. (Table 1)

Table 1. Patient characteristics.

	Inhibitor-negative (N=83)	**Inhibitor-positive (N=39)**
Age at first treatment (months)		
Mean (SD)	20.7 (17.6)	17.6 (13.0)
Family history of inhibitor development		
No	67 (80.7%)	31 (79.5%)
Yes	12 (14.5%)	4 (10.3%)
Unknown	4 (4.8%)	4 (10.3%)
***F8* gene mutation (null vs. non-null)**		
Non-null mutation	15 (18.1%)	1 (2.6%)
Null mutation	67 (80.7%)	35 (89.7%)
Unknown	1 (1.2%)	3 (7.7%)
***F8* gene mutation (detailed)**		
Frameshift mutation	10 (12.0%)	7 (17.9%)
Intron 1 inversion	4 (4.8%)	0 (0%)
Intron 22 inversion	32 (38.6%)	20 (51.3%)
Large deletion	5 (6.0%)	2 (5.1%)
Nonsense mutation	16 (19.3%)	6 (15.4%)
Missense mutation	8 (9.6%)	0 (0%)
Splice site mutation	4 (4.8%)	1 (2.6%)
Non-null mutation, type unknown	3 (3.6%)	0 (0%)
Unknown	1 (1.2%)	3 (7.7%)
FVIII product (type)		
Recombinant FVIII product	39 (47.0%)	22 (56.4%)
Plasma-derived FVIII product	44 (53.0%)	17 (43.6%)

After applying the phage display method to samples from these 39 inhibitor-positive patients, the total number of unique peptide sequences generated was 33,068,599. The median number of unique peptide sequences per patient sample was 215,263 (range: 88,009-1,116,682). After filtering out

peptide sequences with a peptide count lower than 20, peptide sequences that were also found in control samples, and peptide sequences with a high likelihood of being target-unrelated, 3,020,449 unique peptide sequences (9.1% of the initial dataset) were left. Next, we selected the 5 peptide sequences that were most likely to be targets of FVIII-specific antibodies for each patient (i.e. FVIII mimotope sequences) as described in the methods section which resulted in 195 FVIII mimotope sequences being left (0.00059% of the initial dataset). These 195 FVIII mimotope sequences were aligned against the linear sequence of FVIII. Of the 195 FVIII mimotope sequences, 157 were not completely surface-inaccessible. (0.00047% of the initial dataset) These FVIII mimotope sequences were used for further analyses.

Selection epitopes of interest using screening method

The remaining 157 FVIII mimotope sequences were then screened against the database of epitopes extracted from the Immune Epitope Database (IEDB) website. The database contained 35,185 B-cell epitopes, of which 34,866 were unique. The screening method is described in Figure 1. Among the top 10 B-cell epitopes (based on their unadjusted P-value), the most common organisms from which the B-cell epitopes were derived were human gammaherpesvirus 4 (3x), human herpesvirus 5, human betaherpesvirus 6B, Plasmodium vivax, Dengue virus type 2, hepatitis B virus, human T-cell leukaemia virus type I and trypanosoma cruzi. Since no B-cell epitope was statistically significantly overrepresented among FVIII mimotope sequences compared to the randomly generated peptide sequences after correcting for multiple testing, we selected the top 10 B-cell epitopes with the smallest unadjusted P-values for further analyses.

Generating an epitope motif for sets of FVIII mimotope sequences

For each set of FVIII mimotope sequences that was mapped to one of the 10 B-cell epitopes selected previously, we generated an epitope motif. (Figure 2) We found different epitope motifs for each set of FVIII mimotope sequences. In addition, none of the epitope motifs showed a very strong signal.

Evaluating the predictive value of selected FVIII mimotope sequences

We then assessed the predictive value of the FVIII mimotope sequences that were mapped to the top 10 B-cell epitopes using two different modelling approaches. As a control group for these analyses, we used 83 inhibitor-negative patients. (see Table 1 for patient characteristics of the control group)
Firstly, we evaluated the predictive performance of each individual FVIII mimotope sequence (measured in the pre-treatment samples) for inhibitor development using logistic regression. There

were 45 FVIII mimotope sequences that were mapped to one of the 10 B-cell epitopes. Of these, 8 were not present in any of the pre-treatment samples and were therefore excluded from further analyses. The predictive value of the remaining 37 FVIII mimotope sequences, measured using the C-statistic, varied
from 0.50 to 0.54.

Next, we evaluated if a logistic regression model containing of all the FVIII mimotope sequences mapped to the 10 B-cell epitopes would have good predictive power. Ridge regularization was used to minimize overfitting. The C-statistic of this model (with 37 predictors) was 0.73 (95%CI: 0.62-0.83) in the post-treatment samples. (Figure 3) The fitted model was then validated in the pre-treatment samples. The C-statistic in the post-treatment samples was 0.58 (95%CI: 0.47-0.70). (Figure 3) Of these 37 mimotope sequences, 13 had good alignment against the linear sequence of FVIII (which was defined as at least 4 matching residues in the alignment). The alignment location of these mimotope sequences on FVIII are shown in Figure 4.

Table 2. Top ten organisms containing B-cell epitopes deposited in the IEDB database with sequence similarity to FVIII mimotope sequences derived from the phage-display experiment.

Organism	Antigen	Epitope	12-mer Fragment	score of FVIII-associated sequences *	score of randomly generated sequences *	P-value**	Adjusted P-value**
Human gamma herpesvirus 4	EBNA-3C	STPERPGPS KQPSEP	ERPGPSK QPSEP	8	0	3.57e-03	1
Human gamma herpesvirus 4	EBNA-3C	EEAQSTPER PGPSKQ	EAQSTPE RPGPS	7	0	7.30e-03	1
Plasmodium vivax Sal-1	hypothetical protein PVX_081792	QMEGFQKQ LDRLSDSLS KIQKALGE YL	QKQLDRL SDSLS	7	0	7.30e-03	1
Human herpesvirus 5 strain AD169	HCMVUL50	AGGAAAGP RPPPPPM	GAAAGPR PPPPP	6	0	1.49e-02	1
Dengue virus 2 Thailand/NGS-C/1944	Genome polyprotein	ALCEALTL ATGPISTLW	EALTLAT GPIST	6	0	1.49e-02	1
Hepatitis B virus	Major surface antigen precursor	ASTNRQSG RQPTPISPPL RDSHPQ	GRQPTPIS PPLR	6	0	1.49e-02	1
Human T-cell leukemia virus type I	envelope glycoprotein	DISQLTQAI VKNHKNLL K	QAIVKNH KNLLK	6	0	1.49e-02	1
Human beta herpesvirus 6B	DNA polymerase catalytic subunit	DKKQLALK TTCNSVYG VT	QLALKTT CNSVY	6	0	1.49e-02	1
Human gamma herpesvirus 4	Epstein-Barr nuclear antigen 4	EPWQPLSW PHETRVI	EPWQPLS WPHET	6	0	1.49e-02	1
Trypanosoma cruzi strain CL Brener	surface antigen 2 (CA-2)	FGQAAAGG KPSPFGQ	QAAAGG KPSPFG	6	0	1.49e-02	1

*The alignment procedure is described in Figure 1. Hypothesis testing was done using Fisher's exact test. Adjustment for multiple testing was done using the Bonferroni correction method.

Figure 2. Epitope motifs identified in each set of FVIII mimotope sequences mapped to a linear B-cell epitope.

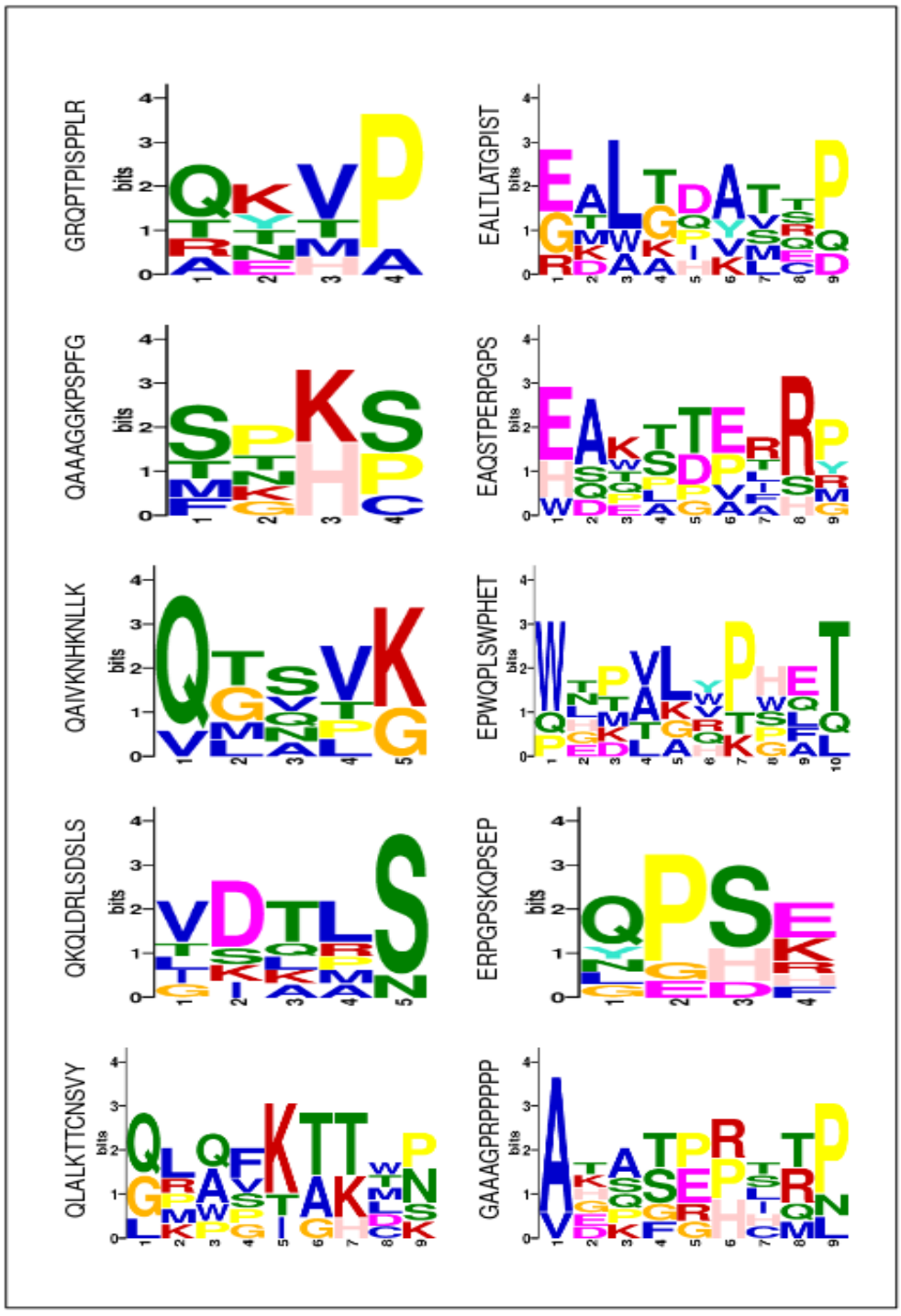


Figure legend: An epitope motif was generated for each set of peptide sequences that had at least for 4 matching residues for one of the 10 B-cell epitope fragments shown in Table 2. The fragment on which each subfigure was based is shown to the left of each figure. Motif search was performed using MEME version 5.4.1.

Figure 3. ROC plot of regression model containing 37 FVIII mimotope sequences as predictors.

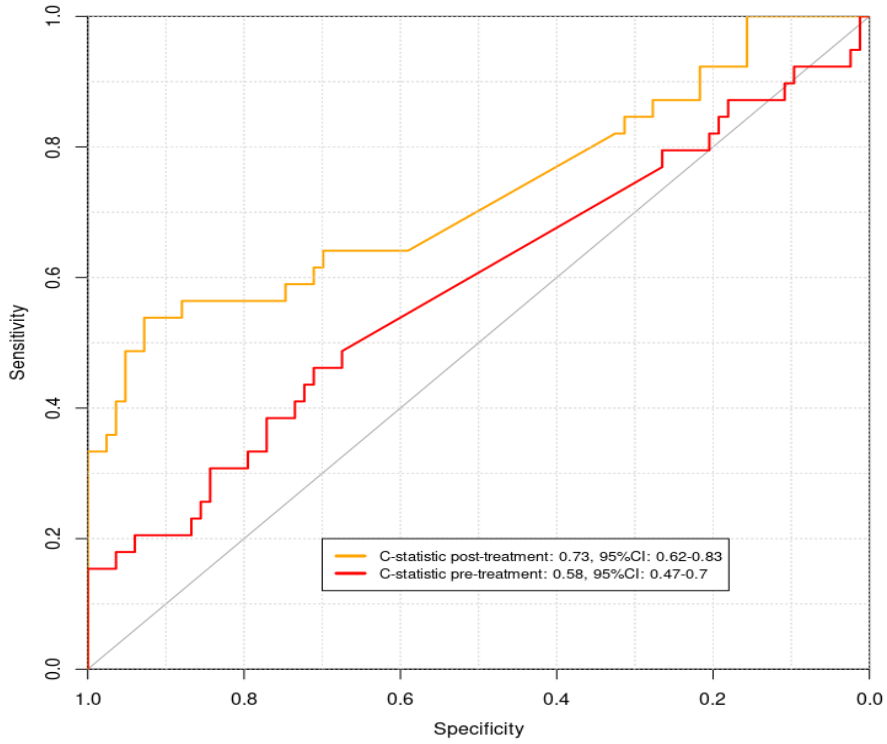
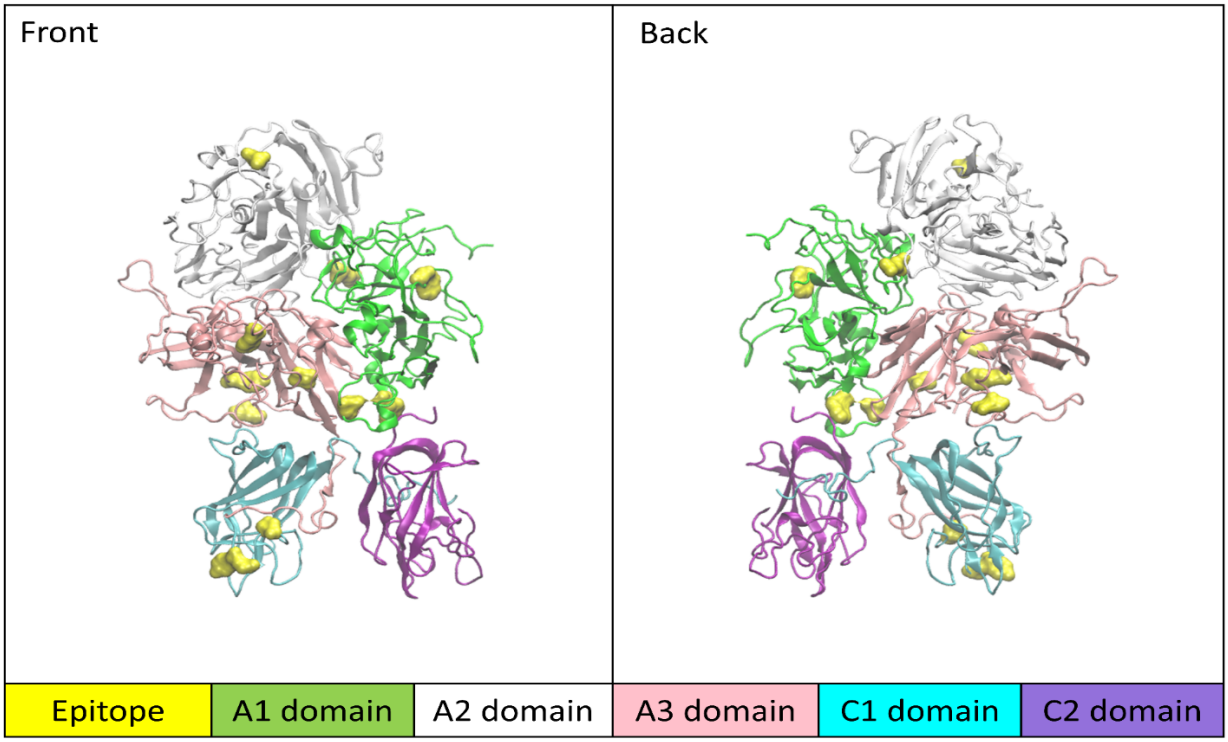


Figure 4. Location of mimotope sequences on a B-domain deleted FVIII (PDB ID: 3CDZ).



Discussion:

This is, to our knowledge, the first study to directly assess the role of molecular mimicry in inhibitor development in patients with congenital haemophilia A.

A novel next generation sequencing phage display method was used to assess the total IgG repertoire of 39 patients with haemophilia who developed neutralizing antibodies against FVIII in the first 50 days of treatment with FVIII. From this assay, we identified a set of FVIII mimotope sequences. We then applied a custom screening method to assess if there was any sequence similarity between the set of FVIII mimotope sequences and a dataset of linear B-cell epitopes downloaded from the Immune Epitope Database. No B-cell epitope was statistically significantly overrepresented among the FVIII mimotope sequences derived from the experiment when compared to the randomly generated peptide sequences.

Further analyses on the FVIII mimotope sequences matched to the top 10 most significantly overrepresented B-cell epitopes showed that there was no strong epitope motif present. Furthermore, the presence of these FVIII mimotope sequences in pre-FVIII exposure samples (both individually and as a group) seemed not to be strongly predictive of inhibitor development after exposure to FVIII.

Results from other studies

There are very few studies that have investigated the link between infection with a pathogen and inhibitor development in patients with congenital haemophilia. One recent study found no significant difference in the proportion of patients with an inhibitor in HCV-positive patients (N = 58), compared to HCV-negative patients (N = 96). [55] However, it was unclear whether HCV infection preceded inhibitor development. As of yet, no other studies have directly assessed the relationship between a pathogen and inhibitor development.

Several case reports have linked cases of acquired haemophilia with infections such as Mycoplasma pneumonia [56], Streptococcus pneumoniae [57] ,human herpesvirus 4 (Epstein-Barr virus) [58] ,Helicobacter cinaedi [59] , Hepatitis C [60,61], Hepatitis B [62] , SARS-CoV-2 [63] , HIV [64] , Cytomegalovirus [65] ,human alphaherpesvirus 3 (Herpes Zoster) [65] , Human Herpesvirus 8 [66] , Influenza A [67] and Escherichia coli [67] and Parvovirus B19 [68] . Due to the rarity of acquired haemophilia, and the lack of controls in these studies (they were all case reports), identifying the causative agent is very difficult and many of the pathogens listed above are more than likely to be completely unrelated to the disease. Interestingly, some of the pathogens reported in those previous case-reports seem to be part of the herpesviridae family. In our study, 5 of the top 10 B-cell epitopes were derived from viruses in the Herpesviridae family; namely human gammaherpesvirus 4 (found 3x), human herpesvirus 5 and human betaherpesvirus 6B. Several viruses in this family have been reported to be associated with auto-immune disease, especially human herpes virus 4 (Epstein-Barr virus). [69]

Limitations:

The FVIII mimotope sequences we identified were linear and were screened against a set of linear B-cell epitopes. However, the majority of B-cell epitopes seem to be conformational and not linear. [70,71] Therefore, if there is any molecular mimicry between certain antigens on pathogens and FVIII and if this is due to similarity of conformational parts of the antigen and FVIII, then our approach will not be able to identify these epitopes.

Furthermore, this approach only focused on sequence similarity with linear B-cell epitopes deposited in the IEDB. The advantage of this approach is that the sequences deposited in the IEDB are already known to be immunogenic. The downside is that approach precludes evaluation of organisms not present in the IEDB.

Lastly, this approach is focused on amino acid epitopes. If there is some similarity between carbohydrate epitopes on foreign antigens derived from pathogens and FVIII, then this approach wouldn't be able to find these epitopes.

Conclusion:

There was no evidence of molecular mimicry between a set of FVIII mimotopes obtained from a novel random peptide phage-display library assay and linear B-cell epitopes from a set of pathogens obtained from the IEDB.

CHAPTER 5

EFFECT OF DNA METHYLATION ON INHIBITOR DEVELOPMENT IN PEOPLE WITH HAEMOPHILIA A TREATED WITH FVIII CONCENTRATES

Himani Chand[1], Shermarke Hassan[1], Andrea Cairo[2], Roberta Palla[1], Flora Peyvandi[1,2]

[1]Università degli Studi di Milano, Department of Pathophysiology and Transplantation, Milan, Italy

[2]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Angelo Bianchi Bonomi Hemophilia and Thrombosis Center, Milan, Italy

Research and Practice in Thrombosis and Haemostasis, Volume 9, Issue 3, 102858 DOI: 10.1016/j.rpth.2025.102858

Abstract:

Haemophilia A (HA) is a hereditary X-linked haemorrhagic disorder. Following the first treatment with exogenous FVIII, one-third of patients with severe HA develop anti-FVIII antibodies (inhibitors), which render treatment ineffective. Recent findings underlined the critical role of DNA methylation in several autoimmune diseases by altering gene expression profiles. This study was designed to evaluate potential differences in DNA methylation profiles of HA patients who develop inhibitors against FVIII and those who do not, with the aim of identifying immune-regulatory genes that may contribute to the risk of inhibitor formation. In this study, we aimed to understand whether CpG sites are differentially methylated in peripheral blood mononuclear cells (PBMCs) of HA patients could have a role in inhibitor development, to better understand the biological pathways that lead to inhibitor development. A case-control study was performed using 45 inhibitor-positive HA patients and 67 inhibitor-negative HA patients from the SIPPET study cohort. Enrichment bisulfite sequencing was performed on DNA samples from PBMCs of HA patients and differentially methylated CpG sites (DMCs) were identified with bioinformatic approach.Overall, information on 621,121 CpG sites was obtained. 2,772 sites were significantly differentially methylated (unadjusted P-value < 0.05). Association of CpG sites to a few genes involved in active immune response (*JAK1, CD1C, PIGR, TOLLIP, BLNK, CD44, IL23R, IFNLR1, SOCS2, TLR1*, etc.) was seen in inhibitor-positive patients but it did not indicate pathways associated with inhibitor development.

DMCs were identified in PBMC samples from HA patients with inhibitors. However, our data could not confirm the role of these CpG sites affecting immune regulatory pathways.

Introduction:

Haemophilia A (HA) is a recessive X-linked haemorrhagic bleeding disorder caused by mutations in the *F8* gene. HA patients suffer from spontaneous bleeding manifestations within muscles, joints, and sometimes vital organs caused by the absence or defect in coagulation factor VIII (FVIII). To manage and prevent bleeding episodes in haemophilia A patients, exogenous FVIII concentrates are commonly administered as a standard replacement therapy.[8] Although this treatment has significantly improved the quality of life for patients, following the treatment with exogenous FVIII, a major challenge arises in the development of neutralising antibodies, known as inhibitors, against the administered FVIII. Despite the significant advancements in haemophilia management, roughly one-third of patients with severe HA develop anti-FVIII antibodies (inhibitors) during the first 50 exposure days to FVIII concentrates that render treatment ineffective.[72]

The development of inhibitors in haemophilia A patients remains a complex and poorly understood phenomenon. It is a multifaceted process with intricate involvement between genetic factors and external environmental factors. Several genetic risk factors (for instance the type of *F8* gene mutation, family history, HLA type, and variants in immune-related genes e.g. *IL-10*), as well as treatment-related risk factors (such as the source of FVIII,[9] and the intensity of FVIII treatment at first exposure), have been discussed [73] as triggers of the immune response against FVIII in some HA patients. Lacroix-Desmazes et al highlights mechanisms driving inhibitor development, including innate and adaptive immune responses, and strategies to induce immune tolerance, such as immune tolerance induction (ITI) therapy and innovative approaches like engineered FVIII products or tolerogenic nanoparticles.[17] However, the main causal pathways and additional causal factors leading to inhibitor development remain unknown.

The cellular immune response against FVIII is complex involving both T-cell and B-cell response. Studies also exhibit the vital role of activated T cells in the memory B-cell response to FVIII and the requirement of direct T-cell contact to re-stimulate these cells.[74] Recent studies have suggested that genetic factors, environmental influences, and immune responses play pivotal roles in shaping this phenomenon.[10,75] However, one emerging epigenetic mechanism that has gained attention in various diseases is DNA methylation.

DNA methylation modification is characterised by the addition of a methyl group to cytosine residues by DNA methyltransferases (DNMTs) enzymes, which play a critical role in gene expression regulation. Changes in DNA methylation patterns are linked with immune

dysfunction and autoimmune disorders exhibiting improper gene expression. It has been demonstrated in several studies that immune components like CD4+ T cells and co-stimulatory molecules undergo epigenetic modifications and play a crucial role leading to dysfunction and regulation issues in cellular immune system in systemic autoimmune rheumatic diseases.[76] In patients with systemic lupus erythematosus (SLE) studies have described hypomethylation of neutrophils and granulocytes, especially at the gene locus of the interferons MX1 and IF144.[77] In rheumatoid arthritis, early methylation studies revealed that patients' T-cells, similar to SLE, are characterised by global hypomethylation.[78,79]

Studies involved in the association of the role of DNA methylation in inhibitor development are limited. In haemophilia research, only one non-peer-reviewed study, published on a preprint server in 2019, has assessed the effect of changes in methylation on inhibitor development.[80] This study found that a differentially methylated CpG site (DMC) in the *IL4* gene was associated with inhibitor development. However, the study only assessed the methylation status of thirteen CpG sites in several regulatory regions that were previously shown to be associated with inhibitor development.

Our current study seeks to identify DNA regions in peripheral blood leukocytes that are differentially methylated in patients with and without an inhibitor, using a genome-wide approach, and understand how DNA methylation may influence inhibitor development in haemophilia A patients undergoing FVIII replacement therapy. By utilising next-generation sequencing methods we intend to investigate the functional implications of differentially methylated genes, allowing us to gain a deeper understanding of the underlying biological pathways contributing to inhibitor development and possibly to identify biomarkers that could predict the occurrence of inhibitor development.

Methods:

Sample Population

DNA samples of 111 patients were obtained from the Survey of Inhibitors in Plasma-Product Exposed Toddlers (SIPPET) randomised trial.[9] SIPPET study was performed on previously untreated patients (PUPs) or minimally treated patients suffering from severe HA (FVIII< 0.01 IU mL−1). Minimally treated patients were the patients that received blood components as a form of treatment for <5 times, were not treated with investigational drugs and tested negative for FVIII inhibitors.[9] Administration of either plasma-derived FVIII or recombinant FVIII as replacement therapy was randomly assigned to each patient. Follow-up on these patients was done for 50 exposure days or 3 years for patients that did not develop inhibitors. For inhibitor-positive patients end of study was considered when inhibitor development was confirmed. Among 111 patients, 45 were inhibitor-positive (cases) and 66 were inhibitor-negative (controls). For this study DNA samples were obtained from end of study samples of citrated plasma stored at -80°C.

Sample preparation

Genomic DNA was extracted from peripheral blood leukocytes (PBMCs). 2µg of high-quality genomic DNA was used for methylation assay.

Methylation assay

Target enrichment bisulfite-sequencing was performed using Agilent SureSelectXT Human Methyl-Seq assay. Genomic DNA was sheared into smaller fragments using the Covaris E-series system. The recommended methyl-Seq protocol and specific enzymes were used to repair the ends, adenylate the 3' ends, and ligate methylated adaptors. Genomic DNA library hybrids were prepared. The hybrids were captured on biotinylated RNA baits and purified using streptavidin beads (New England BioLabs), this step was performed to capture target-enriched genomic DNA library which consisted only of CpG sites and genomic regions susceptible to methylation in the whole genome. The target sequences were bisulfite converted using the EZ DNA Methylation-Gold kit (Zyma Research) as described in the methyl-Seq protocol. During the bisulfite conversion step only the unmethylated Cytosine bases get converted to Uracil, but the methylated Cytosine bases remain unchanged. Target sequences were then indexed. The sequencing of libraries was then performed on Illumina NextSeq2000.

This assay captures about 3.7 million CpGs which consists of about 84Mb of the human genome.

Bioinformatic pipeline

Pre-processing of bisulfite sequencing data

BCL files obtained from the sequencer were converted to FASTQ format using bcl2fastq conversion software. Quality check was performed with FastQC.[81] Low-quality reads (phred <20) were removed, and adapters were trimmed from the fastq files using Trim-Galore! [82] to increase the quality of sequences. Adapter trimmed fastq files were aligned to bisulfite converted human reference genome (Hg38) and methylation call files were generated using Bismark.[83] Quality check on alignment files was performed with qualimap [84] to get information about the coverage of reads for each sample.

Data cleaning

Methylation call files (.cov format) were used for data analysis using the MethylKit [85] R package. A quality check was done on all the methylation call files for any low coverage. Filtering of bases with coverage less than 10X was done to avoid unreliable statistical results. Similarly, bases with very high coverage were also removed in case of overamplification and/or PCR bias. All the files were normalized taking the median of coverage as the scaling factor.

Differentially methylated CpG sites detection

After normalization, CpG sites that were present across all the samples were merged into a methylKit data frame. Differential methylation was calculated using the beta-binomial model, this model considers the variability of the sample as well as epigenetic variability and corrected for multiple testing using the Benjamini-Hochberg method. Then, all the CpG sites with very low methylation differences among them, were filtered out. A new data frame was created where CpG sites were selected with p-values < 0.05 as well as methylation difference of at least 5%. The Reactome database was used to perform over-representation analysis on the differentially methylated sites as well as Enrichr web-based tool [86] was used to get GO molecular functions.

Results:

Patient population

The patient population of our study comprises individuals diagnosed with haemophilia A, and they exhibit various mutations of the *F8* gene, as shown in Table 1. Moreover, the study encompasses patient samples from different ethnic backgrounds, with a majority hailing from Egypt (39.64%%) and India (36.03%) (Table 1).

Table 1. Characteristics of F8 Gene Mutations and Ethnicity of HA Patients (n=111).

	Inhibitor-Positive (n=44)	**Inhibitor-Negative (n=67)**	**Percentage of Total Study Population (n=111)**
***F8* Gene Mutation (n=111)**			
Int 22 Inv	25	28	47.75%
Missense	2	6	7.21%
Splicing	2	4	5.40%
Frameshift	6	9	13.52%
Large deletion	3	4	6.30%
Polymorphism	-	2	1.80%
Nonsense	5	10	14.42%
Int 1 Inv	1	3	3.60%
Ethnicity (n=111)			
Argentina	-	1	0.90%
Austria	1	1	1.80%
Brazil	2	-	1.80%
Chile	2	1	2.70%
Egypt	23	21	39.64%
India	12	28	36.03%
Iran	1	5	5.40%
Italy	-	3	2.70%
Mexico	-	1	0.90%
Spain	3	1	3.60%
Turkey	-	1	0.90%
USA	-	4	3.60%

General data quality

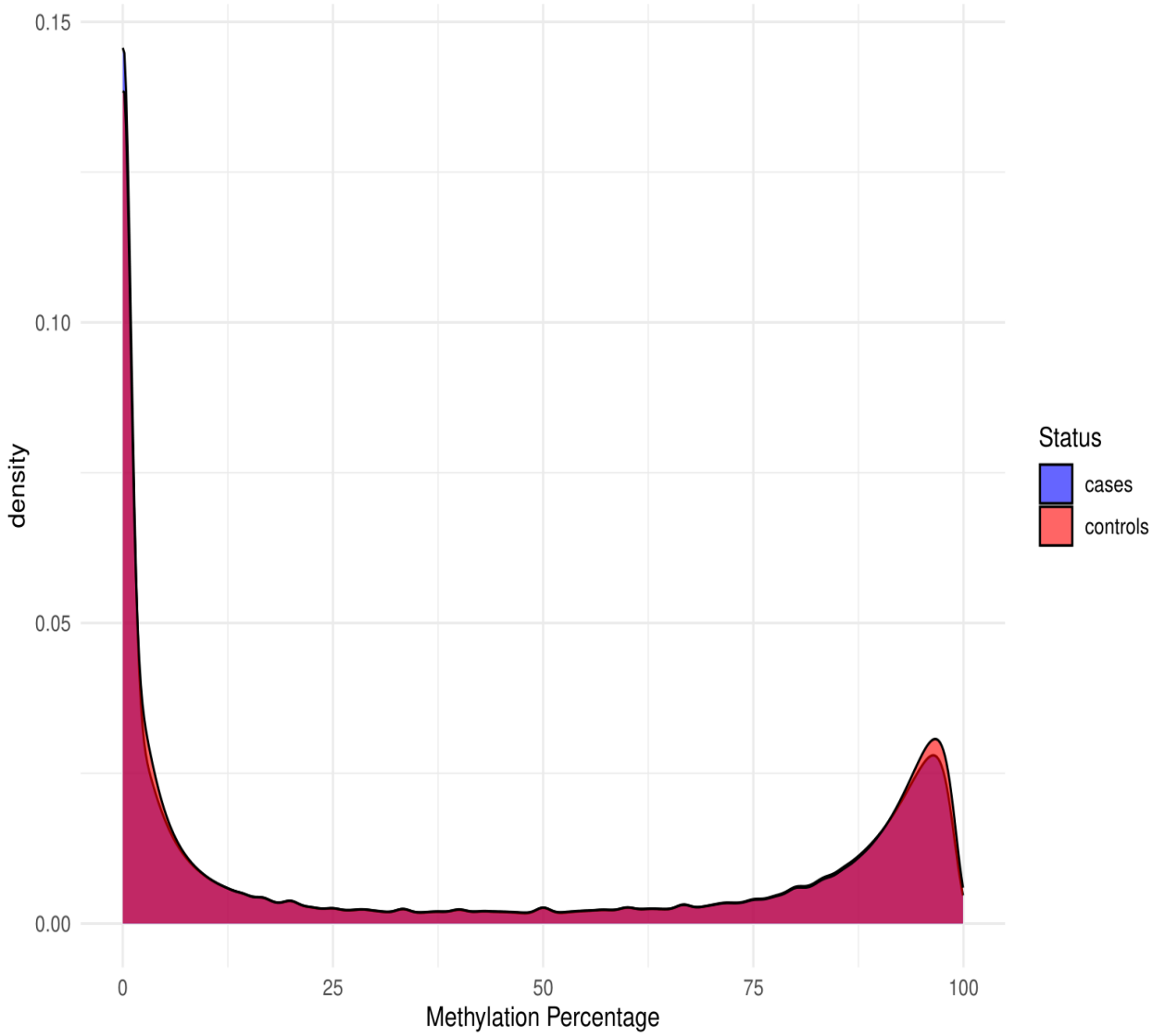
Quality control of the data revealed read fragment sizes from 70-151 bp long. The average Bismark alignment rate of the reads to the converted genome was ˜80%. Mean coverage to the targeted sites was found to be 50X. Reads with at least 10X coverage were found to be between 51%-77% with a mean of 2 million CpG sites in the data.

Exploratory analysis result

Principal component analysis (PCA) performed on the whole data revealed no distinction between cases and controls suggesting little variation between the two groups. The density distribution graph of CpG sites for both cases and control showed very little difference between the two groups (Figure 1). The graph shows how the CpG sites are distributed across its range (0%- 100%, where 0% depicts unmethylated CpG sites while 100% depicts methylated CpG sites) and gives insight into the distribution of methylation among the population. From this graph we can interpret the differential methylation among inhibitor-positive and inhibitor-negative patients on a wider scale. Density represents the relative likelihood of methylation values of cpg sites at different levels of methylation. This graph gives two peaks, one peak at 0% denoting unmethylated cpg sites and the other peak at 100% giving information of methylated cpg sites. However, it indicated that the inhibitor-positive group (cases) was less methylated than the inhibitor-negative group (control). Overall information on 621,121 CpG sites was obtained. From the statistical beta-binomial testing method, 2,772 differentially methylated CpG sites (DMCs) were selected based on p-value < 0.05 and at least a 5% methylation difference.

From the obtained DMCs, 1,680 sites were hypermethylated and 1,092 CpG sites were hypomethylated in inhibitor-positive samples. However, these sites were found to be statistically insignificant on a genome-wide level (q-value > 0.05). Of the total DMCs, 30% were found in the introns, 13% in the intergenic region, 48% in promoters, and 9% in the exon region. Most of these DMCs were outside the CpG island regions (Figure 2). The plot indicates that intergenic regions exhibit differential methylation, with hypomethylation predominantly observed in CpG shelves and hypermethylation in CpG islands. This suggest that methylation changes are most notable in non-CpG island regions, potentially influencing gene regulation and contributing to the differences between the two sample groups.

Figure 1. Distribution plot of percentage methylation of CpG sites.



Distribution plot of percentage methylation of CpG sites showing the control group(inhibitor-negative) having slightly higher methylation levels than the case(inhibitor-positive) group

The x-axis represents percentage methylation of CpG sites, ranging from 0% (completely unmethylated) to 100% (fully methylated). The y-axis represents probability density, indicating the relative frequency of CpG sites at different methylation levels.

Figure 2. Annotation of CpGs according to their locations in the CpG islands context.

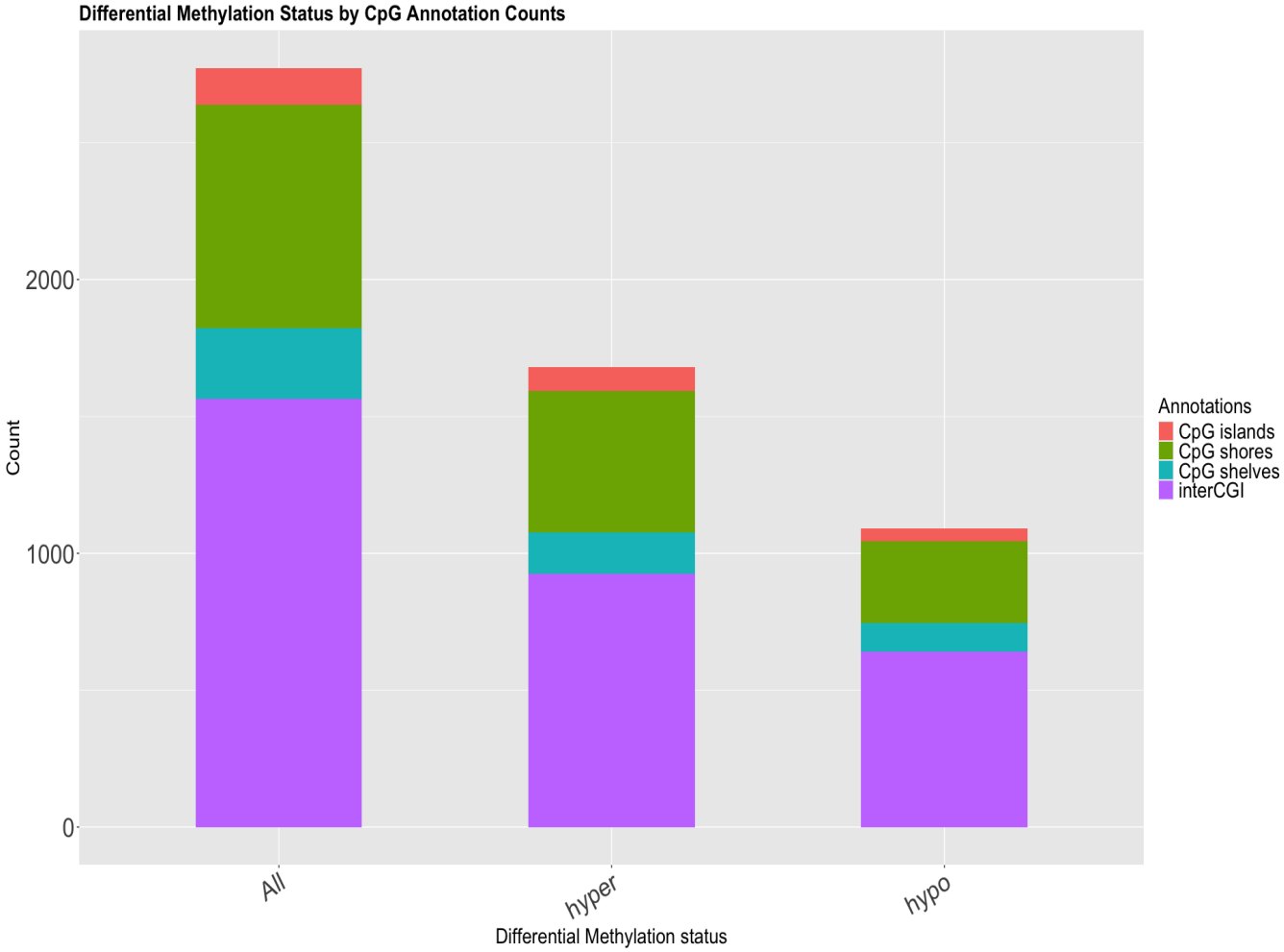


Figure legend: The bar plot depicts the distribution of differentially methylated CpG sites (DMCs) across various genomic regions (CpG Islands, Shores, Shelves, and InterCGI) based on their methylation status. The y-axis represents the count of CpG sites annotated to each genomic region. The x-axis categorizes the data into three groups:

- **All**: Represents the total number of differentially methylated CpG sites irrespective of their methylation status.
- **Hyper**: Refers to CpG sites that exhibit increased methylation in the inhibitor-positive group compared to the inhibitor-negative group. **Hyper-methylation**: CpG sites with significantly higher methylation percentages in the inhibitor-positive group compared to the inhibitor-negative, based on a predefined statistical threshold (p-value < 0.05, methylation difference > 5)
- **Hypo**: Refers to CpG sites that exhibit decreased methylation in the inhibitor-positive group compared to the inhibitor-negative group. **Hypo-methylation**: CpG sites with significantly lower methylation percentages in the inhibitor-positive group compared to the inhibitor-negative group (p-value <0.05, methylation difference < -5.

CpG shores are defined as 2Kb upstream/downstream from the ends of the CpG islands, CpG shelves are defined as another 2Kb upstream/downstream of the farthest upstream/downstream limits of the CpG shores, remaining genomic regions make up the inter-CGI annotation.

Over-representation analysis

An association with a gene was found for 2,762 out of 2,772 DMCs. Association with a gene is determined if DMC is present in the regulatory region or gene body region. We then used the biomaRt [87] package to annotate these DMCs to the gene it is associated with. Reactome pathway analysis identified 144 genes involved in the immune pathway and 59 genes involved in haemostasis. Keeping aside the p-value, hypermethylation of 105 genes and hypomethylation of 52 genes involved in the immune system were found, while 46 hypermethylated and 22 hypomethylated genes related to haemostasis were observed. Genes involved in the immune system were selected from the Reactome database for both hypermethylated and hypomethylated sites and the Enrichr web-based tool was used to see enriched molecular functions of the related genes (Table 2, Table 3). From this analysis we identified gene ontology (GO) of molecular functions performed by genes which match the genes annotated to a specific molecular function term in the GO database. It also highlights molecular functions that are enriched in our gene set. GO Molecular Function revealed many hypomethylated genes to be involved in Cytokine Receptor Binding (GO:0005126) (p-value 0.000442) [*SOCS2, IL23R, TLR9, JAK1*], Toll-like Receptor Binding (GO:0035325) (p-value 0.000669) [*TLR1, TOLLIP*] and Cytokine Receptor Activity (GO:0004896) (p-value 0.002504) [*IL23R, IFNLR1, CD44*] suggesting a possible role of these genes in inhibitor positive patients.

Table 2. Table of top 10 significant GO Molecular Function of hypermethylated genes in Inhibitor-positive samples as observed by Enrichr web-based tool.

Term	p-value	q-value	Overlap_genes
Ubiquitin-Protein Transferase Activity (GO:0004842)	3.064843e-09	6.865249e-07	UBE2H, LRSAM1, TRIM41, UBE3D, CUL3, MIB2, TNFAIP3, LMO7, FBXL14,FBXL22, TRIM6, TRAF3, TRIM5, RNF182, UBE2O, RNF130, HERC6
Ubiquitin-Like Protein Transferase Activity (GO:0019787)	1.013918e-08	1.135588e-06	UBE2F, UBE2H, LRSAM1, CUL3, MIB2, TNFAIP3, LMO7, FBXL14, TRAF3, TRIM5, RNF182, UBE2O, RNF130
Ubiquitin Protein Ligase Activity (GO:0061630)	1.469754e-04	1.071939e-04	LRSAM1, TRIM6, TRIM41, UBE3D, TRAF3, TRIM5, CUL3, MIB2, UBE2O, RNF130, FBXL22, HERC6
Ubiquitin-Like Protein Ligase Activity (O:0061659)	1.914176e-06	1.071939e-04	LRSAM1, TRIM6, TRIM41, UBE3D, TRAF3, TRIM5, CUL3, MIB2, UBE2O, RNF130, FBXL22, HERC6
Phosphotyrosine Residue Binding (GO:0001784)	6.456476E-06	2.892501E-04	SYK, SLA, PIK3R2, GRB2, PIK3R1
Cytokine Receptor Activity (GO:0004896)	1.309272e-05	4.243176e-04	IFNAR2, IL1RL2, IL18RAP, CCR6, IL17RC, CRLF1
Protein Phosphorylated Amino Acid Binding (GO:0045309)	1.325992e-05	4.243176e-04	SYK, SLA, PIK3R2, GRB2, PIK3R1
Kinase Binding (GO:0019900)	3.121649e-04	8.740617e-03	IFNAR2, TRIM6, CDKN1A, STK11IP, TNIP2, TRIM5, CASP1, GRB2, PRKACA, AP2A2, PTPN2
Neurotrophin TRK Receptor Binding (GO:0005167)	4.528891e-04	1.014472e-02	GRB2, PIK3R1
Neurotrophin TRKA Receptor Binding (GO:0005168)	4.528891e-04	1.014472e-02	GRB2, PIK3R1

Table 3. Table of top 10 significant GO Molecular Function of hypomethylated genes in Inhibitor-positive samples as observed by Enrichr web-based tool.

Term	p-value	q-value	Overlap_genes
Growth Hormone Receptor Binding (GO:0005131)	0.000343	0.022875	SOCS2, JAK1
Cytokine Receptor Binding (GO:0005126)	0.000442	0.022875	SOCS2, IL23R, TLR9, JAK1
Regulatory RNA Binding (GO:0061980)	0.000446	0.022875	TRIM71, HNRNPA2B1, TLR9
Toll-like Receptor Binding (GO:0035325)	0.000669	0.025771	TLR1, TOLLIP
1-Phosphatidylinositol-3-Kinase Regulator Activity (GO:0043274)	0.001100	0.032500	SOCS2, SOCS3
Phospholipase Binding (GO:0043274)	0.001266	0.032500	BLNK, FYN
Phosphatidylinositol-3-Kinase Regulator Activity (GO:0035014)	0.001633	0.035274	SOCS2, SOCS3
Hormone Receptor Binding (GO:0051427)	0.001832	0.035274	SOCS2, JAK1
Cytokine Receptor Activity (GO:0004896)	0.002504	0.042839	IL23R, IFNLR1, CD44
Ubiquitin Conjugating Enzyme Binding (GO:0031624)	0.004425	0.058792	RNF144B, TOLLIP

The table 2 and Table 3 summarizes the results of the enrichment analysis.

Term: The specific GO Molecular Function term describes the activity performed by gene products associated with the input gene list. Each term represents a molecular function enriched among the gene list provided.

p-value: The statistical significance of the enrichment for the corresponding GO term. A lower p-value indicates a stronger association between the input gene set and the GO term.

q-value: The false discovery rate (FDR)-adjusted p-value, which accounts for multiple testing correction.

Overlap_genes: The list of genes from the input dataset that overlap with the genes annotated to the specific GO Molecular Function term in the reference database. These are the genes contributing to the observed enrichment and are potentially involved in the processes described by the term.

Subpopulation analysis

Since our patient population consisted of different ethnic backgrounds, a sub-study using the same methodology was also performed to observe any methylation difference among the patient samples based on their ethnicities as environmental factors play a major role in epigenetic modification of an individual. Clustering of patient samples from Egypt and India based on CpG sites was performed using Ward's method. The analysis did not reveal any pattern in the cluster suggesting that the geographic origin of an individual does not play a major role in DNA methylation leading to inhibitor development.

No new enriched pathway was observed when performing a similar analysis on samples with Int22Inv mutation of the *F8* gene (n=53). This was performed to observe whether any enrichment of a particular pathway could be involved in relation to a specific mutation.

Although the analysis did not provide any methylation pattern among inhibitor-positive and inhibitor-negative groups based on their ethnic background or *F8* mutation, we acknowledge that the cohort size is limited. This small sample size may reduce the statistical power to comprehensively evaluate the contribution of certain genetic or epigenetic factors, such as population-specific combinations of SNPs.

Discussion:

The process of inhibitor development in haemophilia A remains an intricate puzzle, involving the complex interaction of genetics and environmental factors. This study presents a comprehensive genome-wide analysis of the potential role of DNA methylation in influencing inhibitor development among haemophilia A patients undergoing FVIII replacement therapy. Additionally, our focus on exploring the functional implications of differentially methylated genes enabled us to delve deeper into the underlying biological pathways that might underpin inhibitor development. The study comprises patients with different ethnicities as well as different *F8* gene mutations. Our results demonstrate the complexity of the epigenetic landscape concerning inhibitor development. Whilst the influence of genetic factors and external triggers on inhibitor development has been recognized, the involvement of epigenetic mechanisms such as DNA methylation has been only studied in recent years. The role of DNA methylation in the regulation of gene expression adds new complexity to the already multifaceted process of inhibitor development. The study demonstrates a slight difference in the methylation status between the two groups. The inhibitor-positive group was moderately less methylated as compared to the inhibitor-negative group, indicating that patients who develop inhibitors have higher gene expression of certain genes. This study also showed that several CpG sites associated with genes related to immune regulation and haemostasis were either hypomethylated or hypermethylated in patients who developed inhibitors as compared to those who did not. While most of the hypomethylated genes are a part of the signalling cascade a few genes that are part of the immune system are also included for example *JAK1, CD1C, PIGR, TOLLIP, BLNK, CD44, IL23R, IFNLR1, SOCS2, TLR1*, etc. These genes are an indicator of the involvement of DNA methylation in a stronger immune response in patients who develop inhibitors against FVIII. Hypermethylation of genes involved in Ubiquitin-related pathways (eg. *UBE2H, TRIM6, TNFAIP3, CUL3, TRAF3,* etc.) (Table 2) indicates increased inflammatory responses, dysregulated immune system activation and impaired antigen presentation. Genes such as *TRIM6*, *TNFAIP3*, and *TRAF3* are also involved in NF-κB immune signaling pathway indicating immune dysregulation since NF-κB is one of the major regulators of inflammation and immunity.

Our findings revealed specific differentially methylated CpG sites in relation to genes that could be associated with the formation of inhibitors like genes associated with Cytokine Receptor Binding, Toll-like Receptor Binding, and Cytokine Receptor Activity, although these associations were not statistically significant at the genomic level. This suggests that although

DNA methylation may play a role in increasing the gene expression of genes that heighten the immune response, it is likely part of a larger network of factors contributing to inhibitor development. However, no clear mechanism or pathway has been determined.

Importantly, our study highlights the diversity of the patient population, which includes people of different ethnic origins and presents with different *F8* gene mutations. This diversity could be extremely important because genetic background plays a role in the modulation of the manifestation and severity of haemophilia A. However, the observed lack of distinct clustering based on geographic origin indicates that ethnicity might not be a predominant factor in influencing inhibitor development in our study population or it could be due to our small sample size to provide statistical power to address the possibility. While not predominant, such factors could potentially influence inhibitor susceptibility and warrant further investigation in larger, more diverse cohorts. Future studies with expanded sample sizes could help validate these findings and explore the role of genetic variability in greater detail.

In a study published by Liu W et al. in 2022, comparing gene expression and methylation levels between haemophilia A patients with and without inhibitors, it was found that differentially expressed genes were significantly enriched in immune response regulation, particularly involving pathways such as Toll-like receptor and tumour necrosis factor signalling. [88] Additionally, methylation analysis revealed lower DNA methylation status in certain gene regions in inhibitor-positive patients, potentially implicating a role for epigenetic modifications in inhibitor development.[88] However, in our study, involving 111 patients, despite the lack of significant q-values in the analysis, enrichment analysis revealed several hypomethylated genes to be involved in crucial functions such as cytokine receptor binding and Toll-like receptor binding, suggesting a potential role of these genes in inhibitor-positive patients.

The data in Table 2 and Table 3 provide important insights into the molecular functions linked to hypermethylated and hypomethylated genes in inhibitor-positive samples, shedding light on potential immune system implications. Table 2 points to significant pathways like ubiquitin-protein transferase activity, ubiquitin-like protein transferase activity, and ubiquitin ligase activity, which are central to processes such as protein degradation, immune signalling, and maintaining cellular balance. Changes in these pathways could disrupt antigen processing and presentation, potentially impairing immune surveillance. Table 3 highlights pathways enriched in hypomethylated genes, including cytokine receptor binding, Toll-like receptor binding, and PI3K regulation—all critical for immune signalling, inflammatory responses, and communication between immune cells. Genes like *SOCS2, JAK1,* and *TLR9* within these pathways may reflect disruptions in immune regulation specific to inhibitor-positive samples.

These findings provide a starting point for future studies to explore how such pathway alterations might contribute to immune system dysregulation, potentially leading to weakened immune responses or increased susceptibility to autoimmune issues. Investigating these pathways further, especially their roles in antigen presentation and immune activation, could help identify new therapeutic targets. Overall, the results underscore the importance of DNA methylation as a regulator of immune-related pathways in disease, offering valuable leads for clinical research.

Limitations:

Despite its contributions, this research is not without limitations. Our sample size, though modest, takes into consideration a lot of CpG sites that might have no role in inhibitor development, contributing to higher q-values increasing insignificance. A larger sample size might impact the generalizability of findings. Moreover, while our data provides a preview of DNA methylation patterns, a longitudinal study design of how methylation of DNA changes over time and in response to treatment could offer a better understanding as described by Verlinden et al in their study.[89] Additionally, since the primary FVIII immune response is T-cell dependent, focusing on peripheral blood leukocytes might not fully capture methylation patterns that could play a role in inhibitor development. DNA methylation study targeting T lymphocytes and B lymphocytes might provide a finer differential methylation pattern of immunoregulatory genes. Furthermore, while the study explores DNA methylation as a potential mechanism, it is important to recognize that other epigenetic modifications, such as histone modifications, could also contribute to the observed outcomes. Acknowledging the limitations of our study, further research and a more comprehensive methodology is required to elucidate the intricate connections between genetics, epigenetics, and inhibitor development, potentially revealing novel avenues for intervention and personalised therapeutic approaches.

Conclusion:

Concluding the study, although the observed associations were not genome-wide significant, differentially methylated sites associated with the immune system and haemostasis especially the hypomethylation of genes involved in cytokine receptor activities and toll-like receptor activity suggest the complexity of the underlying mechanisms and highlight the need for larger and more comprehensive investigations. It also indicates that patients with DNA methylation of genes involved in active immune response might be associated with the development of inhibitors against FVIII.

CHAPTER 6

THESIS CONCLUSION

This thesis has explored various aspects of inhibitor development in haemophilia A (HA) using a combination of multiple approaches, including epitope mapping, molecular mimicry analysis and DNA methylation studies. Each chapter has contributed unique insights into the complex mechanisms underlying the immune response to factor VIII (FVIII). By exploring these mechanisms, the study deepens our understanding of inhibitor formation, and it highlights potential predictive and therapeutic strategies.

In Chapter 3, we employed a novel random peptide phage-display assay called mimotope variation assay (MVA) to assess the FVIII-specific epitope profile of previously untreated HA patients. Our findings demonstrated that the FVIII-specific antibody response is highly polyclonal, with numerous distinct epitope motifs aligning primarily with the A1, A3, and B domains of FVIII. Importantly, our statistical models using pre-treatment epitope repertoire data were able to predict inhibitor development, suggesting a potential diagnostic tool to assess inhibitor risk before FVIII treatment initiation. However, limitations such as the inability to confirm B-domain-specific antibody binding in vitro and the challenge of distinguishing conformational from linear epitopes suggest that further studies are required to refine our understanding of FVIII-targeting antibody responses.

Chapter 4 investigated the role of molecular mimicry in inhibitor development, utilizing a next-generation sequencing phage-display approach to examine potential cross-reactivity between FVIII mimotope sequences and known B-cell epitopes from pathogens. Despite analysing a large dataset, no strong evidence of molecular mimicry was found, suggesting that pathogen-derived epitopes may not be a major driver of inhibitor formation in congenital HA patients. While prior case reports have suggested associations between acquired haemophilia and certain infections, the lack of significant sequence similarity in our study underscores the need for alternative explanations, such as broader immune dysregulation mechanisms, in understanding inhibitor development.

In Chapter 5, we examined the role of DNA methylation in modulating the immune response leading to inhibitor formation. Genome-wide analysis revealed differential methylation patterns in inhibitor-positive versus inhibitor-negative patients, with hypomethylated genes primarily involved in immune regulation and inflammatory signalling pathways, such as JAK1, CD44, and IL23R. These findings suggest that epigenetic modifications may contribute to a

heightened immune response in inhibitor-positive patients, although the exact mechanisms remain unclear. Although the limitations posed by a modest sample size and the inability to establish causal relationships, this study underscores the significance of epigenetic factors in FVIII immunogenicity and calls for further investigations incorporating larger cohorts and additional epigenetic markers.

In this thesis, computational tools were essential in studying large-scale data analysis, predictive modelling, and epitope mapping. The use of machine learning algorithms improved our ability to identify potential risk factors for inhibitor formation, improving the diagnostic accuracy. The integration of high-throughput sequencing and bioinformatics allowed for a more comprehensive analysis of immune responses, which provided valuable insights that would have been challenging to obtain through conventional laboratory methods alone.

From a clinical perspective, the results obtained from the studies of this thesis have significant translational implications for HA treatment. The predictive models developed for inhibitor risk assessment could be integrated into clinical practice to guide personalized treatment strategies. Additionally, the identification of immune and epigenetic markers provides potential therapeutic targets for modulating inhibitor responses, leading improved management of HA patients. This research provides valuable datasets and analytical frameworks that can serve as a valuable source for studying FVIII immunogenicity and immune tolerance induction. The methodologies employed, from epitope mapping to methylation analysis, offer a framework for investigating immune responses in other haematological and autoimmune conditions.

Taken together, the findings presented in this thesis highlight the complexity of inhibitor development in HA and underscore the multifactorial nature of this process. While predictive models based on epitope profiles show promise, the absence of molecular mimicry and the role of epigenetic modifications suggest that a combination of genetic, immunological, and environmental factors drives inhibitor formation. The integration of advanced molecular profiling techniques provides new avenues for personalized risk assessment and therapeutic interventions. While this study advances our understanding of inhibitor development in HA, several limitations remain. The relatively small sample size in methylation studies necessitates validation in larger cohorts. Additionally, the absence of direct functional validation of identified biomarkers limits the immediate clinical applicability of some findings. The non-availability of cells in the SIPPET study prevents further analysis of immune system using single cell RNA seq and transcriptional analysis. Future research should focus on expanding patient cohorts, integrating multi-omics approaches, and developing targeted

immunomodulatory therapies to mitigate inhibitor formation. By addressing these challenges, future studies can further refine predictive models and therapeutic strategies for HA patients.

SUMMARY

This PhD thesis explores inhibitor development in haemophilia A using several computational and bioinformatic approaches on the SIPPET group. We used mimotope variation analysis from a random peptide phage-display library and observed stronger antibody response against A1, A3 and B domains of FVIII. The epitope motifs mapped against FVIII were then used to predict inhibitor development. Two statistical models developed on data from a novel random peptide phage display assay, were used to predict inhibitor development before exposure to exogenous FVIII. The research also investigated if FVIII-specific B-cell epitopes share sequence similarity with pathogen-derived linear B-cell epitopes to study the role of molecular mimicry in inhibitor development. From the study we found no evidence of molecular mimicry between a set of FVIII mimotopes obtained from a novel random peptide phage-display library assay and linear B-cell epitopes from a set of pathogens obtained from the IEDB. Furthermore, the thesis analysed methyl-seq data to understand the effect of DNA methylation on inhibitor development in haemophilia A patients treated with FVIII concentrates. The study identified association of CpG sites to few genes involved in active immune response (*JAK1, CD1C, PIGR, TOLLIP, BLNK, CD44, IL23R, IFNLR1, SOCS2, TLR1*, etc.) in inhibitor-positive patients. This approach provides deeper insights into immune regulation, tolerance, and antibody development, which are central to inhibitor formation in HA. Overall, this research contributes to a deeper understanding of immune system dynamics in haemophilia A and underscores the importance of integrating computational and experimental methods for studying complex diseases.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to everyone who has supported and contributed to the completion of this PhD thesis.

First and foremost, I am grateful to my supervisor, Prof. Flora Peyvandi, for her guidance and insightful feedback. Her expertise and support have been invaluable to the success of this work.

I would also like to extend my heartfelt thanks to my co-supervisor, Dr. Roberta Palla, for her continuous support and mentorship. Her expertise has been critical in helping me navigate complex challenges throughout this project. I greatly appreciate her patience and encouragement which has been instrumental in the completion of this thesis.

I am thankful to my colleague Dr. Shermarke Hassan for collaborating with me and his valuable contributions to this research. His input, whether through data analysis, discussions, or feedback on manuscript drafts, has significantly enhanced the quality of the work presented in this thesis. It has been a pleasure working with such dedicated and insightful individual.

I am equally thankful to the members of MSCA-ITN “EDUC8” consortium, for their constructive criticism and helpful suggestions that have significantly improved the quality of my research.

I am thankful to my colleagues in the lab, especially Andrea Cairo, for creating a supportive and collaborative environment. Their assistance, both technical and moral, has been essential to the completion of this thesis.

Special thanks to all the collaborators involved in the SIPPET study, whose data were fundamental to this research. I would like to acknowledge the funding bodies and organisations that supported my research.

I dedicate this work to the loving memory of my mother, whose love shaped who I am, I carry you with me in everything I do. To my father and brother, thank you for being my anchor through the hardest moments, your strength and quiet support have meant more than words can say. To my best friend, Deepak, without whom I would have been lost during this journey, thank you for being my constant source of happiness. To my friends, Aparna, Srishti and Ishita thank you for your encouragement during the challenges of this journey.

Thank you all for your contributions to this work.

PEER REVIEW INFORMATION

I would like to thank Prof. Michael Makris for his thoughtful evaluation and contribution to the review of this thesis.

BIBLIOGRAPHY

1. Mcguire, M.F., Sriram Iyengar, M., Mercer, D.W.: Computational Approaches for Translational Clinical Research in Disease Progression.
2. Shivahare, B.D., Singh, J., Ravi, V., Chandan, R.R., Alahmadi, T.J., Singh, P., Diwakar, M.: Delving into Machine Learning's Influence on Disease Diagnosis and Prediction. Open Public Health J. 17, (2024).
3. Siminea, N., Czeizler, E., Popescu, V.B., Petre, I., Păun, A.: Connecting the dots: Computational network analysis for disease insight and drug repurposing, (2024)
4. Chen, C., Wang, J., Pan, D., Wang, X., Xu, Y., Yan, J., Wang, L., Yang, X., Yang, M., Liu, G.P.: Applications of multi-omics analysis in human diseases, (2023)
5. Iperi, C., Fernández-Ochoa, Á., Pers, J.O., Barturen, G., Alarcón-Riquelme, M., Quirantes-Piné, R., Borrás-Linares, I., Segura-Carretero, A., Cornec, D., Bordon, A., Jamin, C.: Integration of multi-omics analysis reveals metabolic alterations of B lymphocytes in systemic lupus erythematosus. Clinical Immunology. 264, (2024).
6. Castro, H.E., Briceño, M.F., Casas, C.P., Rueda, J.D.: The History and Evolution of the Clinical Effectiveness of Haemophilia Type A Treatment: A Systematic Review. Indian Journal of Hematology and Blood Transfusion. 30, 1–11 (2014).
7. Gouw, S.C., Van Den Berg, H.M., Le Cessie, S., Van Der Bom, J.G.: Treatment characteristics and the risk of inhibitor development: A multicenter cohort study among previously untreated patients with severe hemophilia A. Journal of Thrombosis and Haemostasis. 5, 1383–1390 (2007).
8. Peyvandi, F., Garagiola, I., Young, G.: The past and future of haemophilia: diagnosis, treatments, and its complications. The Lancet. 388, 187–197 (2016).
9. Peyvandi F , Mannucci PM , Garagiola I , et al. A randomized trial of factor VIII and neutralizing antibodies in hemophilia A . N Engl J Med . 2016 ; 374 (21): 2054 - 2064 .
10. Cormier M , Batty P , Tarrant J , Lillicrap D . Advances in knowledge of inhibitor formation in severe haemophilia A . Br J Haematol . 2020 ; 189 (1): 39 - 53 .
11. Yee A , Oleskie AN , Dosey AM , et al. Visualization of an N-terminal fragment of von Willebrand factor in complex with factor VIII . Blood . 2015 ; 126 (8): 939 - 942 .
12. Pezeshkpoor B , Schreck U , Biswas A , et al. An in silico and in vitro approach to elucidate the impact of residues flanking the cleavage scissile bonds of FVIII . PLoS One . 2017 ; 12 (7): e0180456 .

13. Lollar P , Parker CG . Subunit structure of thrombin-activated porcine factor VIII . Biochemistry . 1989 ; 28 (2): 666 - 674 .
14. Van Dieijen G , Tans G , Rosing J , Hemker HC . The role of phospholipid and factor VIII(a) in the activation of bovine factor X . J Biol Chem . 1981 ; 256 (7): 3433 - 3442 .
15. Lollar P . Pathogenic antibodies to coagulation factors. Part one: factor VIII and factor IX . J. Thromb. Haemost . 2004 ; 2 (7): 1082 - 1095 .
16. Lai J , Hough C , Tarrant J , Lillicrap D . Biological considerations of plasma-derived and recombinant factor VIII immunogenicity . Blood . 2017 ; 129 (24): 3147 - 3154 .
17. Lacroix-Desmazes S , Voorberg J , Lillicrap D , Scott DW , Pratt KP . Tolerating factor VIII: recent progress . Front Immunol . 2020 ; 10 : 2991 .
18. Healey JF , Lubin IM , Nakai H , et al. Residues 484-508 contain a major determinant of the inhibitory epitope in the A2 domain of human factor VIII . J Biol Chem . 1995 ; 270 (24): 14505 - 14509 .
19. Barrow RT , Healey JF , Gailani D , Scandella D , Lollar P . Reduction of the antigenicity of factor VIII toward complex inhibitory antibody plasmas using multiply-substituted hybrid human/porcine factor VIII molecules . Blood . 2000 ; 95 (2): 564 - 568 .
20. Gish JS , Jarvis L , Childers KC , et al. Structure of blood coagulation factor VIII in complex with an anti–C1 domain pathogenic antibody inhibitor . Blood . 2021 ; 137 (21): 2981 - 2986 .
21. Batsuli G , Deng W , Healey JF , et al. High-affinity, noninhibitory pathogenic C1 domain antibodies are present in patients with hemophilia A and inhibitors . Blood . 2016 ; 128 (16): 2055 - 2067 .
22. Healey JF , Barrow RT , Tamim HM , et al. Residues Glu2181-Val2243 contain a major determinant of the inhibitory epitope in the C2 domain of human factor VIII . Blood . 1998 ; 92 (10): 3701 - 3709 .
23. Nguyen P-CT , Lewis KB , Ettinger RA , et al. High-resolution mapping of epitopes on the C2 domain of factor VIII by analysis of point mutants using surface plasmon resonance . Blood . 2014 ; 123 (17): 2732 - 2739 .
24. Lin JC , Ettinger RA , Schuman JT , et al. Six amino acid residues in a 1200 Å2 interface mediate binding of factor VIII to an IgG4κ inhibitory antibody . PLoS One . 2015 ; 10 (1): e0116577 .
25. Sevy AM , Healey JF , Deng W , et al. Epitope mapping of inhibitory antibodies targeting the C2 domain of coagulation factor VIII by hydrogen–deuterium exchange mass spectrometry . J. Thromb. Haemost . 2013 ; 11 (12): 2128 - 2136 .

26. Lavigne-Lissalde G , Rothschild C , Pouplard C , et al. Characteristics, mechanisms of action, and epitope mapping of anti-factor VIII antibodies . Clin Rev Allergy Immunol . 2009 ; 37 (2): 67 - 79 .
27. Walter JD , Werther RA , Brison CM , et al. Structure of the factor VIII C2 domain in a ternary complex with 2 inhibitor antibodies reveals classical and nonclassical epitopes . Blood . 2013 ; 122 (26): 4270 - 4278 .
28. Villard S , Lacroix-Desmazes S , Kieber-Emmons T , et al. Peptide decoys selected by phage display block in vitro and in vivo activity of a human anti-FVIII inhibitor . Blood . 2003 ; 102 (3): 949 - 952 .
29. Villard S , Piquer D , Raut S , et al. Low molecular weight peptides restore the procoagulant activity of factor VIII in the presence of the potent inhibitor antibody ESH8 . J Biol Chem . 2002 ; 277 (30): 27232 - 27239 .
30. Mühle C , Schulz-Drost S , Khrenov A V , et al. Epitope mapping of polyclonal clotting factor VIII-inhibitory antibodies using phage display . Thromb. Haemost . 2004 ; 91 (3): 619 - 625 .
31. Kahle J , Orlowski A , Stichel D , et al. Epitope mapping via selection of anti-FVIII antibody-specific phagepresented peptide ligands that mimic the antibody binding sites . Thromb. Haemost . 2015 ; 113 (2): 396 - 405 .
32. Duncan E , Collecutt M , Street A . Nijmegen-Bethesda assay to measure factor VIII inhibitors . Methods Mol Biol . 2013 ; 992 : 321 - 333 .
33. Sadam H , Pihlak A , Jaago M , et al. Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients . EBioMedicine . 2021 ; 64 : 103211 .
34. He B , Chen H , Li N , Huang J . Sarotup: a suite of tools for finding potential target-unrelated peptides from phage display data . Int J Biol Sci . 2019 ; 15 (7): 1452 - 1459 .
35. Krejci A , Hupp TR , Lexa M , Vojtesek B , Muller P . Hammock: a hidden Markov model-based peptide clustering algorithm to identify protein-interaction consensus motifs in large datasets . Bioinformatics . 2016 ; 32 (1): 9 - 16 .
36. Fraczkiewicz R , Braun W . Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules . J Comput Chem . 1998 ; 19 (3): 319 - 333 .
37. Ngo JCK , Huang M , Roth DA , Furie BC , Furie B . Crystal structure of human factor VIII: implications for the formation of the factor IXa-factor VIIIa complex . Structure . 2008 ; 16 (4): 597 - 606 .

38. Savojardo C , Manfredi M , Martelli PL , Casadio R . Solvent accessibility of residues undergoing pathogenic variations in humans: from protein structures to protein sequences . Front Mol Biosci . 2021 ; 7 : 626363 .
39. Whelan SFJ , Hofbauer CJ , Horling FM , et al. Distinct characteristics of antibody responses against factor VIII in healthy individuals and in different cohorts of hemophili A patients . Blood . 2013 ; 121 (6): 1039 - 1048 .
40. Lebreton A , Lapalud P , Chambost H , et al. Prevalence and epitope specificity of non-neutralising antibodies in a large cohort of haemophilia A patients without inhibitors . Thromb. Haemost . 2011 ; 105 (6): 954 - 961 .
41. Vincent AM , Lillicrap D , Boulanger A , et al. Non-neutralizing anti-FVIII antibodies: different binding specificity to different recombinant FVIII concentrates . Haemophilia . 2009 ; 15 (1): 374 - 376 .
42. Bachelet D , Albert T , Mbogning C , et al. Risk stratification integrating genetic data for factor VIII inhibitor development in patients with severe hemophilia A . PLoS One . 2019 ; 14 (6): e0218258 .
43. Hassan S , Palla R , Valsecchi C , et al. Performance of a clinical risk prediction model for inhibitor formation in severe haemophilia A . Haemophilia . 2021 ; 27 (4): e441 - e449 .
44. Abdi A , Bordbar MR , Hassan S , et al. Prevalence and incidence of non-neutralizing antibodies in congenital hemophilia a—a systematic review and meta-analysis . Front Immunol . 2020 ; 11 : 563 .
45. Cannavò A , Valsecchi C , Garagiola I , et al. Nonneutralizing antibodies against factor VIII and risk of inhibitor development in severe hemophilia A . Blood . 2017 ; 129 (10): 1245 - 1250 .
46. Rojas M , Restrepo-Jiménez P , Monsalve DM , et al. Molecular mimicry and autoimmunity . J Autoimmun . 2018 ; 95 : 100 - 123 .
47. Novotny J . Protein antigenicity: a thermodynamic approach . Mol Immunol . 1991 ; 28 (3): 201 - 207 .
48. Barlow DJ , Edwards MS , Thornton JM . Continuous and discontinuous protein antigenic determinants . Nature . 1986 ; 322 (6081): 747 – 748 .
49. Haste Andersen P , Nielsen M , Lund O . Prediction of residues in discontinuous B-cell epitopes using protein 3D structures . Protein Sci . 2006 ; 15 (11): 2558 - 2567 .
50. Sun P , Chen W , Huang Y , et al. Epitope prediction based on random peptide library screening: benchmark dataset and prediction tools evaluation . Molecules . 2011 ; 16 (6): 4971 - 4993 .

51. Peyvandi F , Miri S , Garagiola I . Immune responses to plasma-derived versus recombinant FVIII products . Front Immunol . 2021 ; 11 : 591878 .
52. Hassan S, Fijnvandraat K, Van Der Bom JG, Gouw SC. Preventing or Eradicating Factor VIII Antibody Formation in Patients with Hemophilia A: What Can We Learn from Other Disorders? Semin. Thromb. Hemost. 2018;44(6):531–543.
53. Vita R, Mahajan S, Overton JA, et al. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 2019;47(D1):D339–D343.
54. L. Bailey T, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Second Int. Conf. Intell. Syst. Mol. Biol. 1994;28–36.
55. Bolina-Santos E, Chaves DG, da Silva-Malta MCF, Carmo RA, Barbosa-Stancioli EF, Lobato Martins M. HCV infection in hemophilia A patients is associated with altered cytokines and chemokines profile and might modulate the levels of FVIII inhibitor. J Med Virol. 2022;94(2):683-691. doi:10.1002/jmv.27432
56. Min SK, Kilgore PE, Ju SK, et al. Transient acquired hemophilia associated with Mycoplasma Pneumoniae pneumonia. J. Korean Med. Sci. 2008;23(1):138–141.
57. Cheng Q, Yu G, Ye Y. Acquired hemophilia A in a patient associated with community-acquired pneumonia. Blood Coagul. Fibrinolysis. 2013;24(7):774–776.
58. Yamamoto M, Shindo M, Sumi C, et al. Acquired hemophilia A associated with Epstein-Barr-virus-associated T/natural killer-cell lymphoproliferative disease: A case report. Medicine (Baltimore). 2021;100(16):e25518.
59. Matsuoka A, Sasaki Y, Kubodera A, et al. Acquired Hemophilia A Presenting with Infectious Aortic Aneurysms Due to an Underlying Helicobacter cinaedi Infection. Intern. Med. 2021;60(24):3947–3952.
60. Martínez Pascual C, Antón Ródenas G, Ortiz Sánchez ML, Pons Miñano JA. [Acquired hemophilia A in a patient with chronic hepatitis C virus infection receiving treatment with peginterferon α-2a]. Med. Clin. (Barc). 2013;141(5):228–229.
61. Mauge L, Pavie J, Batisse D, Darnige L. Acquired hemophilia A resolution in a hepatitis C virus/human immunodeficiency virus-coinfected patient with cure of hepatitis C by direct-acting antiviral agents. Hepatology. 2018;67(2):794–796.
62. Thévenot T, Lambert M, Canva-Delcambre V, Sugishita K, Goudemand J. [Acquired hemophilia and viral hepatitis B infection]. Gastroenterol. Clin. Biol. 2006;30(5):800–802.
63. Cittone MG, Battegay R, Condoluci A, et al. The statistical risk of diagnosing coincidental acquired hemophilia A following anti-SARS-CoV-2 vaccination. J. Thromb. Haemost. 2021;19(9):2360–2362.

64. Rattanathammethee T, Norasetthada L, Tantiworawit A, et al. Acquired hemophilia A in the HIV-infected patient: a case report and literature review. Blood Coagul. Fibrinolysis. 2015;26(2):207–210.
65. Baudo F, Caimi T, de cataldo F. Diagnosis and treatment of acquired haemophilia. Haemophilia. 2010;16(102):102–106.
66. Marietta M, Pozzi S, Luppi M, et al. Acquired haemophilia in HIV negative, HHV-8 positive multicentric Castleman's disease: a case report. Eur. J. Haematol. 2003;70(3):181–182.
67. F P-M, E P, O C, et al. Acquired Haemophilia A in Association with Influenza A and Urinary Tract Infection. Eur. J. case reports Intern. Med. 2020;7(9):.
68. Mo L, Bao GC. Acquired factor VIII deficiency: two case reports and a review of literature. Exp. Hematol. Oncol. 2017;6(1):.
69. Houen G, Trier NH. Epstein-Barr Virus and Systemic Autoimmune Diseases. Front. Immunol. 2021;11:.
70. Barlow DJ, Edwards MS, Thornton JM. Continuous and discontinuous protein antigenic determinants. Nature. 1986;322(6081):747–748.
71. Haste Andersen P, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci. 2006;15(11):2558–2567.
72. Coppola A, Santoro C, Tagliaferri A, Franchini M, DI Minno G. Understanding inhibitor development in haemophilia A: towards clinical prediction and prevention strategies. Haemophilia. 2010 Jan;16 Suppl 1:13-9. Doi: 10.1111/j.1365-2516.2009.02175.x. PMID: 20059564.
73. Garagiola I, Palla R, Peyvandi F. Risk factors for inhibitor development in severe hemophilia a. Thromb Res. 2018 Aug;168:20-27. Doi: 10.1016/j.thromres.2018.05.027. Epub 2018 May 25. PMID: 29879570.
74. André S, Meslier Y, Dimitrov JD, Repessé Y, Kaveri SV, Lacroix-Desmazes S, Dasgupta S. A cellular viewpoint of anti-FVIII immune response in hemophilia A. Clin Rev Allergy Immunol. 2009 Oct;37(2):105-13. Doi: 10.1007/s12016-009-8117-2. PMID: 19165635.
75. Ghosh K, Shetty S. Immune response to FVIII in hemophilia A: an overview of risk factors. Clin Rev Allergy Immunol. 2009 Oct;37(2):58-66. Doi: 10.1007/s12016-009-8118-1. PMID: 19148784.
76. Mazzone, R., Zwergel, C., Artico, M. et al. The emerging role of epigenetics in human autoimmune disorders. Clin Epigenet 11, 34 (2019). Doi: 10.1186/s13148-019-0632-2
77. Coit P, Yalavarthi S, Ognenovski M, Zhao W, Hasni S, Wren JD, Kaplan MJ, Sawalha AH. Epigenome profiling reveals significant DNA demethylation of interferon signature genes in

lupus neutrophils. J Autoimmun. 2015 Apr;58:59-66. Doi: 10.1016/j.jaut.2015.01.004. Epub 2015 Jan 28. PMID: 25638528; PMCID: PMC4363276.

78. Richardson B, Scheinbart L, Strahler J, Gross L, Hanash S, Johnson M. Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis. Arthritis Rheum. 1990 Nov;33(11):1665-73. Doi: 10.1002/art.1780331109. PMID: 2242063.
79. Corvetta A, Della Bitta R, Luchetti MM, Pomponio G. 5-Methylcytosine content of DNA in blood, synovial mononuclear cells and synovial tissue from patients affected by autoimmune rheumatic diseases. J Chromatogr. 1991 May 31;566(2):481-91. Doi: 10.1016/0378-4347(91)80265-e. PMID: 1939459.
80. de Souza TB, de Souza TL, Santos Ferreira C dos, da Silva CFO, Rossetti LC, et al. A Differentially Methylated CpG Site in the IL4 Gene Associated with Anti-FVIII Inhibitor Antibody Development in Hemophilia A. bioRxiv, bioRxiv: 2019. P. 550566.
81. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
82. Krueger F, Trimgalore (2021), GitHub repository: https://github.com/FelixKrueger/TrimGalore
83. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011 Jun 1;27(11):1571-2. Doi: 10.1093/bioinformatics/btr167. Epub 2011 Apr 14. PMID: 21493656; PMCID: PMC3102221.
84. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016 Jan 15;32(2):292-4. Doi: 10.1093/bioinformatics/btv566. Epub 2015 Oct 1. PMID: 26428292; PMCID: PMC4708105.
85. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 2012 Oct 3;13(10):R87. Doi: 10.1186/gb-2012-13-10-r87. PMID: 23034086; PMCID: PMC3491415.
86. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016 Jul 8;44(W1):W90-7. Doi: 10.1093/nar/gkw377. Epub 2016 May 3. PMID: 27141961; PMCID: PMC4987924.
87. Durinck, S., Spellman, P., Birney, E. et al. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc 4, 1184–1191 (2009).
88. Liu W, Lyu C, Wang W, Xue F, Chen L, Li H, Chi Y, Ma Y, Wu R, Fang Y, Zhang L, Yang R. Risk factors for inhibitors in hemophilia A based on RNA-seq and DNA methylation. Res

Pract Thromb Haemost. 2022 Sep 5;6(6):e12794. Doi: 10.1002/rth2.12794. PMID: 36090157; PMCID: PMC9445143.

89. Verlinden I, Güiza F, Derese I, Wouters PJ, Joosten K, Verbruggen SC, Van den Berghe G, Vanhorebeek I. Time course of altered DNA methylation evoked by critical illness and by early administration of parenteral nutrition in the paediatric ICU. Clin Epigenetics. 2020 Oct 20;12(1):155. Doi: 10.1186/s13148-020-00947-w. PMID: 33081814; PMCID: PMC7576729.

APPENDIX

SIPPET study group

Amal El-Beshlawy[1] z Mohsen Elalfy,[2] Vijay Ramanan,[3] Peyman Eshghi,[4] Suresh Hanagavadi,[5] Ramabadran Varadarajan,[6] Mehran Karimi,[7] Mamta V. Manglani,[8] Cecil Ross,[9] Guy Young,[10] Tulika Seth,[11] Shashikant Apte,[12] Dinesh M. Nayak,[13] Elena Santagostino,[14] Maria Elisa Mancuso,[14] Adriana C. Sandoval Gonzalez,[15] Johnny N. Mahlangu,[16] Santiago Bonanad Boix,[17] Monica Cerqueira,[18] Nadia P. Ewing,[19] Christoph Male,[20] Tarek Owaidah,[21] Veronica Soto Arellano,[22] Nathan L. Kobrinsky,[23] Suvankar Majumdar,[24] Rosario Perez Garrido,[25] Anupam Sachdeva,[26] Mindy Simpson,[27] Mathew Thomas,[28] Ezio Zanon,[29] Bulent Antmen,[30] Kaan Kavakli,[31] Marilyn J. Manco- Johnson,[32] Monica Martinez,[33] Esperanza Marzouka,[34] Maria Gabriella Mazzucconi,[35] Daniela Neme,[36] Angeles Palomo Bravo,[37] Rogelio Paredes Aguilera,[38] Alessandra Prezotti,[39] Klaus Schmitt,[40] Brian M. Wicklund,[41] Bulent Zulfikar[42].

[1]Pediatric Hematology Department, Cairo University Pediatric Hospital, Cairo, Egypt;

[2]Faculty of Medicine, Ain Shams Center, University - Department Pediatrics, Cairo, Egypt;

[3]Jehangir Clinical Development Centre, Department of Hematology, Jehangir Hospital Premises, Pune, India;

[4]Pediatric Congenital Hematologic Disorders Research Center, Shahid Beheshti University, Tehran, Iran;

[5]J J M Medical College, Davangere, India;

[6]Centre for Blood Disorders, Chennai, India;

[7]Hematology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran;

[8]Lokmanya Tilak Municipal Medical College & General Hospital, Sion, Mumbai, India;

[9]St. John's Medical College Hospital, Bangalore, India;

[10]Children's Hospital Los Angeles (CHLA), Los Angeles, USA;

[11]All India Institute of Medical Sciences, Department of Hematology, New Delhi, India;

[12]Sahyadri Speciality Hospital, Pune, India;

[13]Melaka Manipal Medical College (Manipal campus), Manipal University, Karnataka, India;

[14]Fondazione IRCCS Ca' Granda Ospedale Maggiore, A. B. Bonomi Hemophilia and Thrombosis Center, Milan, Italy;

[15]Hospital de Especialidades UMAE, Istituto Mexicano del Seguro Social (IMSS), Monterrey, Mexico;

[16]Faculty of Health Sciences, School of Pathology, University of the Witwatersrand, NHLS and Charlotte Maxeke Johannesburg Academic Hospital, Johannesburg, South Africa;

[17]Hospital Universitario La Fe, Unidad Coagulopatias Congenitas, Valencia, Spain;

[18]Centro de Pesquisa Clinica HEMORIO - Instituto Estadual de Hematologia Arthur de Siqueira Cavalcanti, Rio de Janeiro, Brazil;

[19]City of Hope National Medical Center, Duarte, California, USA;
[20]Medizinische Universität Wien, Department of Paediatrics, Wien, Austria;
[21]King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia;
[22]Centro de Hemofilicos del Hospital de Niños Dr. Roberto del Rio, Santiago, Chile;
[23]Sanford Roger Maris Cancer Center, Fargo, North Dakota, USA;
[24]University of Mississipi Medical Center, Division of Pediatric Hematology/Oncology, Jackson, Mississippi, USA;
[25]Hospital Universitario Virgen del Rocio, Unidad de Hemofilia, Sevilla, Spain
[26]Pediatric Hematology Oncology and Bone Marrow Transplantation Institute For Child Health Sir Ganga Ram Hospital, New Delhi, India;
[27]Rush Hemophilia & Thrombophilia Center, Rush University Medical Center, Chicago, Illinois, USA;
[28]Kerala lnstitute of Medical Science, Kerala, India;
[29]Clinica Medica II, Azienda Ospedaliera di Padova, Centro Emofilia, Padova, Italy;
[30]Cukurova Universitesi, Tip Fakultesi Pediatrik Hematoloji B.D., Adana, Turkey;
[31]Ege Universitesi Tip Fakultesi Cocuk Sagligive Hastalikari Anabilim Dali, Pediatrik Hematoloji Bilim, Izmir, Turkey;
[32]Hemophilia & Thrombosis Center, University of Colorado Denver, Aurora, Colorado, USA;
[33]Hospital de Ninos Sor Maria Ludovica La Plata, Servicio de Hematologia, Buenos Aires, Argentina;
[34]Hospital de Ninos Dr. Luis Calvo Mackenna, Centro Hemofilico, Santiago, Chile;
[35]UO Diagnostica Speciale e Terapia delle Malattie dell'Emostasi e della Trombosi, Università Sapienza, Policlinico Umberto I, Rome, Italy;
[36]Fundación de la Hemofilia, Buenos Aires, Argentina;
[37]Hospital Regional Universitario Carlos Haya, Malaga, Spain;
[38]Instituto National de Pediatria, Ciudad de México, Mexico;
[39]Centro de Hematologia e Hemoterapia do Espírito Santo - Hemoes, Vitoria, Brazil;
[40]Department of Pediatric and Adolescent Medicine, Kepler University Clinic, Linz, Austria;
[41]Children's Mercy Hospital, Kansas City, Missouri, USA;
[42]Istanbul Universitesi Cerrahpasa Tip Fakultesi, Pediatrik Hematoloji B.D., Istanbul, Turkey