UNIVERSIDAD DE MURCIA
ESCUELA INTERNACIONAL DE DOCTORADO

New genetic defects and mechanisms involved in antithrombin deficiency

Nuevas alteraciones genéticas y mecanismos implicados en la deficiencia de antitrombina.

Belén de la Morena Barrio

2021

May, 2021

UNIVERSIDAD DE MURCIA
FACULTAD DE MEDICINA

New genetic defects and mechanisms involved in antithrombin deficiency

Nuevas alteraciones genéticas y mecanismos implicados en la deficiencia de antitrombina.

Belén de la Morena Barrio

Director: Javier Corral de la Calle

Tutor: Vicente Vicente García

May, 2021

A mi familia

Agradecimientos

Igual que ha sido difícil abarcar y resumir todo el trabajo de esta investigación en una tesis, igual, o más difícil resulta agradecer como se merece a todos los que la han hecho posible. Esta tesis es resultado no sólo de 4 años de doctorado, sino de más de 20 años de trabajo y colaboración de muchas personas. A cada uno de ellos quiero dar las gracias, porque su colaboración ha sido fundamental.

Gracias al Dr. Vicente Vicente, nuestro jefe de grupo, por darme la oportunidad de comenzar mi andadura en la investigación científica en su grupo, por su apoyo en todo momento, su cercanía y su gran ejemplo de trabajo.

Gracias de manera especial a mi director de tesis, Javier Corral, por su dedicación incondicional, su trabajo, su motivación, su gran creatividad y su ejemplo de servicio y dedicación al grupo. Gracias a su empuje he llegado a hacer cosas que nunca habría imaginado. Gracias por su implicación e interés no solo en lo exclusivamente científico sino en lo personal y profesional. He aprendido mucho de él y lo seguiré haciendo.

Gracias también en especial a mi hermana Uge, que ha sido siempre un gran apoyo y ejemplo desde el inicio de mi carrera científica y siempre; enseñándome y allanándome el camino que ella ya había pisado. Gracias a sus consejos y aportaciones que me han servido de orientación y aprendizaje. Aunque nos confundan por la voz o por lo mucho que nos parecemos, siempre irá ella por delante.

Gracias a Padi, con él, hemos emprendido grandes desafíos en el mundo de la genómica, con el desarrollo de nanoporo. Al compás de "Los Miserables" hemos ido aprendiendo y dominando el MinION, desde sus inicios, con viajes a la UMU o el IMIB, hasta su uso ahora, que hacemos casi con los ojos cerrados; siempre con buen humor, atención y paciencia. Gracias a Toñi, como no, pieza clave en esta tesis y en todo el CRH. Gracias por su trabajo incansable, su orden, su alegría, su cariño y atención a todos.

Gracias a todo el grupo antitrombina, esta tesis es fruto y consecuencia del trabajo de todos. Como el trabajo en equipo para salir de un scaperoom. Me falta agradecer a Carlos por sus brillantes aportaciones, por su trabajo y constancia; a Rosa, también trabajadora y encantadora, recién incorporada al equipo con mucho futuro y mucho pasado en el grupo, gran fichaje y Pedro, nuestro recién incorporado y ¡primer bioinformático del grupo!

Gracias al grupo de Informática de la Universidad de Murcia, a Fran, Jose, Javier y Jesualdo que nos han apoyado en el análisis de la secuenciación por nanoporos. Gracias también al grupo de Informática y Sistemas del IMIB,

Ángel, Fernando, Carmen y María.

Gracias a las bibliotequers, hemos compartido mucho estos años y han sido un gran apoyo siempre: Esther, Sensi, Julia, Vero, ahora Ana... les deseo lo mejor, en breve les tocará a ellas pasar por aquí... y las que ya han pasado de la biblio a otra etapa profesional y con las que hemos aprendido y compartido tanto.

Gracias a los grupos que forman parte del equipo de Hematología y Oncología, por todos estos años. A los seniors: Constant, Rocio, Sonia, Pepe, Marisa, Irene, MªCarmen, Andrés, Paqui, Raul, Francisco; los predocs y los técnicos: Laura, Ernesto, David, Salva, Tzu, Pedro, Natalia, Nuria... De verdad, gracias a cada uno en particular porque he aprendido y disfrutado mucho con vosotros.

Gracias a los médicos del Servicio de Hematología y Oncología, porque gracias a ellos podemos hacer investigación dirigida al paciente, por aportar su visión clínica a nuestros estudios.

Gracias al grupo liderado por Francisco Vidal del Banc de Sang i Teixits de Barcelona, por todo lo que aprendí y lo bien que me acogieron en mi estancia en su laboratorio. Gracias al grupo del profesor Willem Ouwehand de la Universidad de Cambridge, por todo lo que me enseñaron, que ha sido tan importante para mi formación y para el desarrollo de mi tesis doctoral.

Para acabar, los más importantes, mi familia: mi referente fundamental. Ya me encargué de ponerlo en los titulares de los periódicos locales y por si queda duda, lo reitero en mis agradecimientos. Sobre todo, gracias a mis padres, porque siguiendo su ejemplo y gracias a su apoyo, cada uno de sus hijos Pilar, Isabel, Uge, Gonzalo y yo estamos donde estamos, dando lo mejor de nosotros mismos como ellos nos han enseñado.

Resumen en castellano

El sistema hemostático es un delicado balance entre elementos procoagulantes y anticoagulantes que es crucial para organismos con un sistema circulatorio, pero que también es muy sensible a pequeñas distorsiones con graves consecuencias patológicas. La enfermedad tromboembólica, la principal causa de morbimortalidad de nuestra sociedad, se desarrolla como consecuencia del desequilibrio de este sistema hemostático favor de una respuesta coagulante. La gravedad e incidencia de esta enfermedad ha incentivado la búsqueda de alteraciones asociadas con variaciones en el funcionamiento del sistema hemostático. Gran parte de estos estudios se han centrado en los propios elementos de dicho sistema, y muchos se han dirigido a elementos genéticos, dada la alta heredabilidad de la propia enfermedad tromboembólica.

La antitrombina es un potente anticoagulante natural, miembro de la superfamilia de las serpinas, que inhibe la mayoría de las proteasas procoagulantes, principalmente el FXa o la trombina, pero también el FVIIa, FIXa, FXIa y FXIIa. Como miembro de la superfamilia de las serpinas, el mecanismo inhibitorio de la antitrombina es extraordinariamente efectivo. Todas estas características justifican el papel crucial que tiene esta serpina anticoagulante en la hemostasia y explican que incluso reducciones moderadas de antitrombina aumenten significativamente el riesgo de trombosis venosa. Además, se ha demostrado que la ausencia total de este anticoagulante esta asociada a letalidad embrionaria en un modelo de ratón Knock Out (KO). La deficiencia de antitrombina fue la primera trombofilia congénita identificada en 1965 por O. Egeberg en una familia noruega, y es hoy por hoy la trombofilia más grave, la que tiene mayor riesgo trombótico asociado. Estos datos justifican el profundo estudio realizado sobre este anticoagulante y su deficiencia durante más de 55 años. Conocemos en detalle aspectos bioquímicos como su estructura cristalográfica, mecanismo de activación e inhibición, y modificaciones post-traduccionales. También conocemos a fondo aspectos moleculares de la antitrombina, como las mas de 352 mutaciones identificadas y descritas en la base de datos Human Gene Mutation Database (HGMD) en pacientes con deficiencia de antitrombina. Pero todavía existen aspectos que se desconocen y retos por desentrañar de este crucial anticoagulante. Quizás los más importantes son conocer la base de la heterogeneidad clínica descrita entre los casos con deficiencia de antitrombina e identificar la base molecular del 25% de casos en los que todavía no se ha identificado alteración en el gen codificante, *SERPINC1*. Por otra parte, existe un grupo de alteraciones moleculares que están muy poco caracterizadas en deficiencia de antitrombina, y en todas las enfermedades genéticas; las variantes estructurales. Se tratan de alteraciones de más de 50 pares de bases, menos abundantes que alteraciones genéticas puntuales, pero con mayor impacto patológico debido a que afectan a un mayor numero de bases. Estas variantes pueden ser: deleciones, duplicaciones, inversiones, inserciones, traslocaciones o variantes más complejas. Estas alteraciones no se identifican con facilidad y no se han caracterizado profundamente debido a las limitaciones de los métodos moleculares disponibles hasta hace unos años, la mayoría de ellos basados en el estudio de carga genética (MLPA, FISH, array de CGH entre otros), o de secuenciación masiva de pequeños fragmentos. Sin embargo, gracias al desarrollo tecnológico de la última década, con la secuenciación masiva, que incluyen grandes amplificaciones y

nuevos sistemas de secuenciación de tercera generación, deberían ser analizadas en detalle.

El objetivo general de esta tesis ha sido aportar nueva información sobre la antitrombina y su deficiencia, intentando correlacionar datos genéticos y clínicos en una gran cohorte de pacientes con deficiencia de antitrombina. El estudio aborda dos objetivos específicos:

1) Describir el impacto clínico de la deficiencia de antitrombina.
En este punto, evaluamos dos escenarios diferentes: por un lado, el impacto de este trastorno en trombosis pediátrica para intentar identificar y caracterizar las alteraciones con mayor riesgo trombótico y las características, factores de riesgo asociados y por otro lado, estudiamos una nueva variante genética de difícil diagnóstico y moderada clínica trombótica que sin embargo presenta una gran prevalencia en una población para conocer mecanismos de expansión de alteraciones protrombóticas moderadas.
2) Estudiar nuevos defectos moleculares implicados en deficiencia de antitrombina.
Este objetivo se focaliza principalmente en caracterizar en profundidad las variantes estructurales implicadas en la deficiencia de antitrombina e identificar nuevos defectos moleculares responsables de este trastorno. Para abordar estos objetivos, hemos estudiado una de las mayores cohortes de pacientes con este raro trastorno reclutados durante más de 22 años en dos centros de referencia para la deficiencia de antitrombina de España y Bélgica. En total hemos estudiado 676 casos índice y más de 1200 sujetos con deficiencia de antitrombina tras los estudios familiares. En la cohorte Belga, identificamos un subgrupo de 12 pacientes de origen africano portadores de la misma alteración molecular p.Thr147Ala (polimorfismo rs2227606 en *SERPINC1*) y analizamos 100 sujetos sanos de origen africano. Finalmente, nuestro estudio también incluyó un grupo de 9 pacientes con deficiencia de antitrombina causada por variantes estructurales detectadas por MLPA procedentes de cohortes de Finlandia y Francia. En todos los pacientes, la deficiencia de antitrombina se realizó en diferentes hospitales, pero se validó y caracterizó en los centros de referencia mediante estudios funcionales, bioquímicos y moleculares. En todos los casos además se recopilaron datos demográficos y clínicos. Los estudios genéticos para el diagnóstico molecular de la deficiencia de antitrombina incluyen secuenciación con el método de Sanger de los 7 exones de *SERPINC1* y regiones flanqueantes, estudio de grandes alteraciones estructurales mediante MLPA, amplificación mediante grandes PCRs (LR-PCR) del gen *SERPINC1* seguida de secuenciación masiva empleando una plataforma de Illumina, y secuenciación masiva por NGS en la plataforma PGM de Ion Torrent. Para el estudio específico de variantes estructurales y búsqueda de nuevos defectos genéticos en deficiencia de antitrombina en casos seleccionados, se ha realizado array de CGH y secuenciación por nanoporos empleando dos plataformas MinION y PromethION. Para el estudio en profundidad de la variante p.Thr147Ala se empleó un modelo de expresión recombinante de antitrombina en células eucariotas y una novedosa técnica para el análisis de haplotipos basada en secuenciación por nanoporos de amplicones grandes.

Los resultados de este estudio muestran que la deficiencia de antitrombina incrementa notablemente el riesgo de trombosis pediátrica, ya que los portadores de este desorden tienen 300 veces mayor riesgo de sufrir trombosis antes de los 18 años que el descrito en la población general. En estos pacientes se observan dos periodos de mayor incidencia de trombosis pediátrica: el neonatal, con factores de riesgo y localización de la trombosis diferente a los descritos en pacientes adultos con trombosis, destacando la trombosis en senos venosos; y la adolescencia, con factores de riesgo (uso de anticonceptivos, embarazo, inmovilización...) y localizaciones (trombosis venosa profunda o tromboembolismo pulmonar) comunes a los descritos en pacientes adultos con trombosis. Además, describimos una alta morbi-mortalidad en trombosis pediátrica en pacientes con deficiencia de antitrombina, incluyendo eventos fatales, y alta recurrencia. Las mutaciones que provocan una deficiencia sin variante en circulación (deficiencia tipo

I) y las más graves causantes de deficiencia con una antitrombina con menor o nula actividad anticoagulante (deficiencia tipo II) como la variante Budapest 3 en homocigosis, son las que mayor riesgo tienen de sufrir trombosis en edad pediátrica. Por el contrario, hemos caracterizado una deficiencia leve de antitrombina de difícil diagnóstico empleando los sistemas funcionales rutinarios disponibles en hospitales causada por la mutación p.Thr147Ala. El estudio profundo de esta alteración en el modelo recombinante confirmó que provoca una pequeña alteración estructural en la molécula que reduce la actividad funcional. El estudio del haplotipo de la variante con una novedosa técnica de secuenciación de grandes fragmentos de PCR en los portadores ha permitido definir 13 marcadores ligados a ella y presentes en el alelo mutado de todos los portadores de la p.Thr147Ala, demostrando su efecto fundador en población africana. Por otra parte, hemos estudiado de la mayor cohorte mundial de variantes estructurales que causan deficiencia de antitrombina (N= 39). Y lo hemos hecho empleando diferentes métodos moleculares, que incluyen MLPA, LR-PCR y secuenciación por la plataforma Illumina, array de CGH y la secuenciación del genoma completo con tecnología de nanoporos empleando la plataforma MinION y PromethION. Este estudio ha mostrado que las variantes estructurales son heterogéneas en tamaño: afectando desde 193 pb hasta 8 Mb y en tipo, incluyendo duplicaciones, deleciones inserciones y variantes estructurales complejas, esta última alteración, rara en enfermedades genéticas, detectada por primera vez en este trastorno gracias a la secuenciación por nanoporos. Hemos desarrollado una PCR especifica para la detección de una variante relativamente frecuente en deficiencia de antitrombina: la duplicación en tándem del exón 6 de *SERPINC1*. Hemos descrito un mapa detallado de variantes estructurales afectando el gen *SERPINC1* definiendo zonas calientes intra e intergénicas. El estudio mediante secuenciación por nanoporos de las variantes estructurales ha permitido definir los puntos de cohorte de las variantes a nivel nucleótido. Esto ha permitido demostrar la presencia de elementos repetitivos, así como microhomologías y/o inserciones o duplicaciones en los puntos de corte de las variantes estructurales. La presencia de estos elementos sugiere la existencia de un mecanismo común basado en replicación, (BIR/MMBIR/FoSTeS) que involucra elementos repetitivos, para la formación de variantes estructurales causantes de este desorden. Se han identificado variantes estructurales recurrentes comunes en varios pacientes demostrando su efecto fundador mediante estudios adicionales. Por último, la secuenciación por nanoporos ha permitido detectar la base molecular de tres pacientes sin mutación detectada con los métodos moleculares empleados de forma rutinaria en el diagnóstico molecular de la deficiencia de antitrombina. Detectamos por primera vez la inserción de un elemento (SINE- VNTR- Alu) SVA en el intron 6 de *SERPINC1*. Este descubrimiento aumenta el número de patologías causadas por la inserción de retrotransposones, lo que, junto a la dificultad diagnóstica para su identificación, sugiere que puede ser un mecanismo patogénico subestimado en este desorden y posiblemente cualquier desorden de base genética. Además, la resolución nucleotídica obtenida por la secuenciación por nanoporos ha permitido el estudio completo mediante ensamblaje de novo del SVA insertado, demostrando que se trata de un nuevo elemento, hemos podido realizar también estudios filogenéticos del mismo. Además, hemos demostrado la existencia de un efecto fundador en esta variante en los tres pacientes portadores, mediante la identificación de short tándem repeats (STR) comunes en todos ellos.

Nuestro estudio proporciona información nueva y original sobre la deficiencia de antitrombina, la trombofilia más grave, generando las siguientes conclusiones:

1) El estudio de la mayor cohorte de casos con deficiencia de antitrombina aporta evidencia sobre el alto riesgo de trombosis pediátrica asociada a esta trombofilia. Por tanto, apoya el cribado de ciertos tipos de deficiencia de antitrombina en niños de familias afectadas, sobre todo en el caso de las deficiencias tipo I, para beneficiarse de estrategias preventivas y evitar factores de riesgo como el parto asistido en el caso de los neonatos o el uso de anticonceptivos en la adolescencia.

2) Las mutaciones fundadoras expandidas en diferentes poblaciones, como la descrita en nuestro estudio en población africana, causan deficiencias leves de difícil diagnóstico funcional.

3) Nuestro estudio aporta nueva y detallada información sobre las variantes estructurales implicadas en deficiencia de antitrombina. Diseccionamos las variantes estructurales implicadas en deficiencia de antitrombina, mostrando su gran heterogeneidad en tipo y tamaño. Identificamos regiones calientes intra e intergénicas para la formación de variantes estructurales en *SERPINC1*. Identificamos y caracterizamos un nuevo mecanismo que causa deficiencia de antitrombina: la inserción de un nuevo retrotransposón SVA en el intrón 6 de *SERPINC1*, ampliando así el espectro de enfermedades causadas por este tipo de variantes. Mostramos además un posible mecanismo común basado en replicación (BIR/MMBIR/FoSTeS) implicado en la formación de variantes estructurales que afectan al gen *SERPINC1* provocando deficiencia de antitrombina. Nuestro estudio reveló que la combinación de diferentes métodos de diagnóstico molecular es útil para la correcta caracterización de estos grandes defectos genéticos, que nos permite sugerir un nuevo algoritmo diagnóstico para la deficiencia de antitrombina y nos hace sugerir que la secuenciación de nanoporos es el método más adecuado para su completa caracterización. Además, la secuenciación por nanoporos ha permitido también el estudio de haplotipos de una manera mas sencilla y barata para la definición del efecto fundador de una variante.

Abstract

INTRODUCTION

Antithrombin is a natural anticoagulant that plays a crucial role in hemostasis. The wide range of target procoagulant proteases and the potent inhibitory mechanism of this serpin explain that even moderate reductions of antithrombin significantly increase the risk of venous thrombosis. Thus, antithrombin deficiency is the first and the strongest thrombophilia that has been widely studied during more than 50 years. However, there are still some challenges concerning this anticoagulant, such as the factors involved in the clinical heterogeneity, to do a better characterization of structural variants involved in antithrombin deficiency, and to describe the molecular base of up to 25% of cases with no genetic defect. In this thesis we have tried to correlate genetic and clinical data in a large cohort of patients with antithrombin deficiency.

OBJECTIVE

The objective of this thesis is to supply new and original information of this rare disorder. Specifically, we aimed to describe the clinical impact of antithrombin deficiency. Moreover, we also want to deeply characterize structural variants involved in antithrombin deficiency, and to identify new molecular defects involved in this disorder.

METHODS

We have studied one of the largest cohorts of patients with this rare disorder, recruited during more than 20 years. For all patients, a complete set of functional, biochemical and molecular studies have been done and clinical data has been collected. Specific studies included recombinant expression and a novel approach for haplotype analysis. New molecular technologies, including third generation sequencing and long range PCR, have been used for characterization of molecular defects in *SERPINC1*.

RESULTS

Antithrombin deficiency causes very high risk of pediatric thrombosis (300-fold compared with the general population). In contrast, we have characterized a mild antithrombin deficiency of difficult functional diagnosis caused by p.Thr147Ala, a mutation with founder effect in African population.

The study by different methods, including whole genome sequencing with nanopore technology, of the largest cohort of structural variants causing antithrombin deficiency showed that they are heterogeneous in size and type, and identified for the first time a complex structural variant and an insertion of a new SVA element. We described a detailed map of structural variants with intra and intergenic *hotspots* and we found a common mechanism for formation of structural variants causing antithrombin deficiency that involved repetitive elements.

CONCLUSIONS

Our study supplies new and original information on the strongest thrombophilia.

The high risk of pediatric thrombosis supports the screening of certain types of antithrombin deficiency in children of affected families to benefit from preventive strategies. In contrast, founder mutations expanded in different populations like the first one described by our study in African population will probably cause a mild deficiency of difficult functional diagnosis.

We identified and characterized a new mechanism causing antithrombin deficiency: the insertion of a retrotransposon in an intron. Moreover, we dissected the structural variants involved in antithrombin deficiency. Our study also reveals that the combination of different methods is useful for the characterization of these gross gene defects, but nanopore sequencing is the most suitable method to fully characterize them.

Abstract castellano

INTRODUCCIÓN

La antitrombina es un anticoagulante natural que juega un papel crucial en la hemostasia. Las diferentes proteasas procoagulantes diana y el potente mecanismo inhibidor de esta serpina explican que incluso reducciones moderadas de antitrombina aumenten significativamente el riesgo de trombosis venosa. Así, la deficiencia de antitrombina es la primera y más grave trombofilia, estudiada ampliamente durante más de 50 años. Sin embargo, aún existen retos en torno a este anticoagulante, como conocer los factores implicados en la heterogeneidad clínica, caracterizar las variantes estructurales implicadas y describir la base molecular del 25% de casos sin defecto genético conocido. En esta tesis hemos intentado correlacionar datos genéticos y clínicos en una gran cohorte de pacientes con deficiencia de antitrombina.

OBJETIVO

El objetivo de esta tesis es aportar información nueva y original sobre este trastorno. Específicamente, pretendíamos describir el impacto clínico de la deficiencia de antitrombina. Además, quisimos caracterizar en profundidad las variantes estructurales implicadas en la deficiencia de antitrombina, e identificar nuevos defectos moleculares responsables de este trastorno.

MÉTODOS

Hemos estudiado una de las mayores cohortes de pacientes con esta enfermedad rara reclutada durante más de 20 años. En todos los pacientes, se ha realizado estudios funcionales, bioquímicos y moleculares y se han recopilado datos clínicos. Para estudios específicos se empleó expresión recombinante y una novedosa técnica para el análisis de haplotipos. Además, se han utilizado diferentes tecnologías para la caracterización de los defectos moleculares en *SERPINC1*, incluyendo la secuenciación de tercera generación y la PCR larga.

RESULTADOS

La deficiencia de antitrombina incrementa notablemente el riesgo de trombosis pediátrica (300 veces mayor que el de la población general). Por el contrario, hemos caracterizado una deficiencia leve de antitrombina de difícil diagnóstico funcional causada por p.Thr147Ala, una mutación con efecto fundador en la población africana.

El estudio de la mayor cohorte de variantes estructurales que causan deficiencia de antitrombina empleando diferentes métodos, incluido la secuenciación del genoma completo con tecnología de nanoporos, mostró que son heterogéneas en tamaño y tipo e identificó por primera vez una variante estructural compleja y la inserción de un nuevo retrotransposon. Describimos un mapa detallado de variantes estructurales con zonas calientes intra e intergénicas, y encontramos un mecanismo común que involucra elementos repetitivos, para la formación de variantes estructurales causantes de este desorden.

CONCLUSIONES

Nuestro estudio proporciona información nueva y original sobre la trombofilia más grave.

El alto riesgo de trombosis pediátrica apoya el cribado de ciertos tipos de deficiencia de antitrombina en niños de familias afectadas para beneficiarse de estrategias preventivas. Las mutaciones fundadoras expandidas en difer-

entes poblaciones, como la descrita en nuestro estudio en población africana, causan deficiencias leves de difícil diagnóstico funcional.

Identificamos y caracterizamos un nuevo mecanismo que causa deficiencia de antitrombina: la inserción de un retrotransposón en un intrón. Además, diseccionamos las variantes estructurales implicadas en deficiencia de antitrombina. Nuestro estudio revela que la combinación de diferentes métodos es útil para la caracterización de estos grandes defectos genéticos, pero la secuenciación de nanoporos es el método más adecuado para su completa caracterización.

Abbreviations

AT	Antithrombin
BK	Bradikinin
C1-inh	Complement 1-inhibitor
CGH array	Comparative Genome Hybridization array
CI	Confidence intervals
CIE	Crossed immunoelectrophoresis
CNV	Vopy number variants
DVT	Deep venous thrombosis
Fg	Fibrinogen
FISH	Fluorescence in situ hybridization
GBD	Global Burden of Diseases, Injuries, and Risk Factors
GWAS	Genome wide association studies
HIT	Heparin-induced thrombocytopenia
HK	High molecular weight kininogen
HMDG	Human mutation database
LINE	Lng interspersed element
LMWH	Low molecular weight heparin
LR-PCR	Long-range PCR
MLPA	Multiplex Ligation-dependent Probe Amplification
MPN	Myeloproliferative neoplasm
NGS	Massive sequencing methods
OC	Oral contraceptives
ONT	Oxford Nanopore Technologies
ORT	Hormone replacement therapy
PAGE	Polyacrylamide gel electrophoresis
PC	Protein C
PCR	Polymerized chain reaction
PE	Pulmonary embolism
PGM	Personal Genome Machine
PK	(pre)kalikrein
PNH	Paroxysmal nocturnal hemoglobinuria
PS	Protein S
RCL	Reactive Centre Loop
serpins	SERine Protease Inhibitors
SINE	Short interspersed element
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SV	Structural variants
TAFI	Thrombin-Activatable Fibrinolysis Inhibitor.
TF	Tissue factor
TFPI	Tissue factor pathway inhibitor
UF	Unfractionated heparin
uPAR	Urokinase-type Plasminogen Activator
VDRE	Vitamin D Regulatory Elements
VT	Venous thromboembolism
WGS	Whole genome sequencing
WT	Wild-type

Contents

List of Tables

List of Figures

Introduction

1.1 Haemostatic system

The haemostatic system surged more than 400 million of years ago to give response to the lethal consequences of a vascular damage in organisms with a closed and highly pressure circulatory system. Bleeding would be fatal in minutes, without a haemostatic system. Thus, a simple but extraordinary efficient system, which is mainly constituted by two elements: platelets and the clotting cascade, responds within seconds to the signals generating a clot that controls the haemorrhage. We will focus on the clotting cascade, a serial of consecutive proteolytic reactions that transform circulating zymogens in active serine proteases that reach high levels of thrombin, the last procoagulant protease, which forms the fibrin clot and also activates platelets (Figure 1). But this system must be exquisitely regulated to initiate these procoagulant reactions only where and when it is required. Thus, a group of anticoagulants were also incorporated to the haemostatic system to generate a well-regulated balance that maintains blood flowing under normal conditions (thus dominating the anticoagulant elements) but able to quick and strongly respond to vascular damage stimulus.

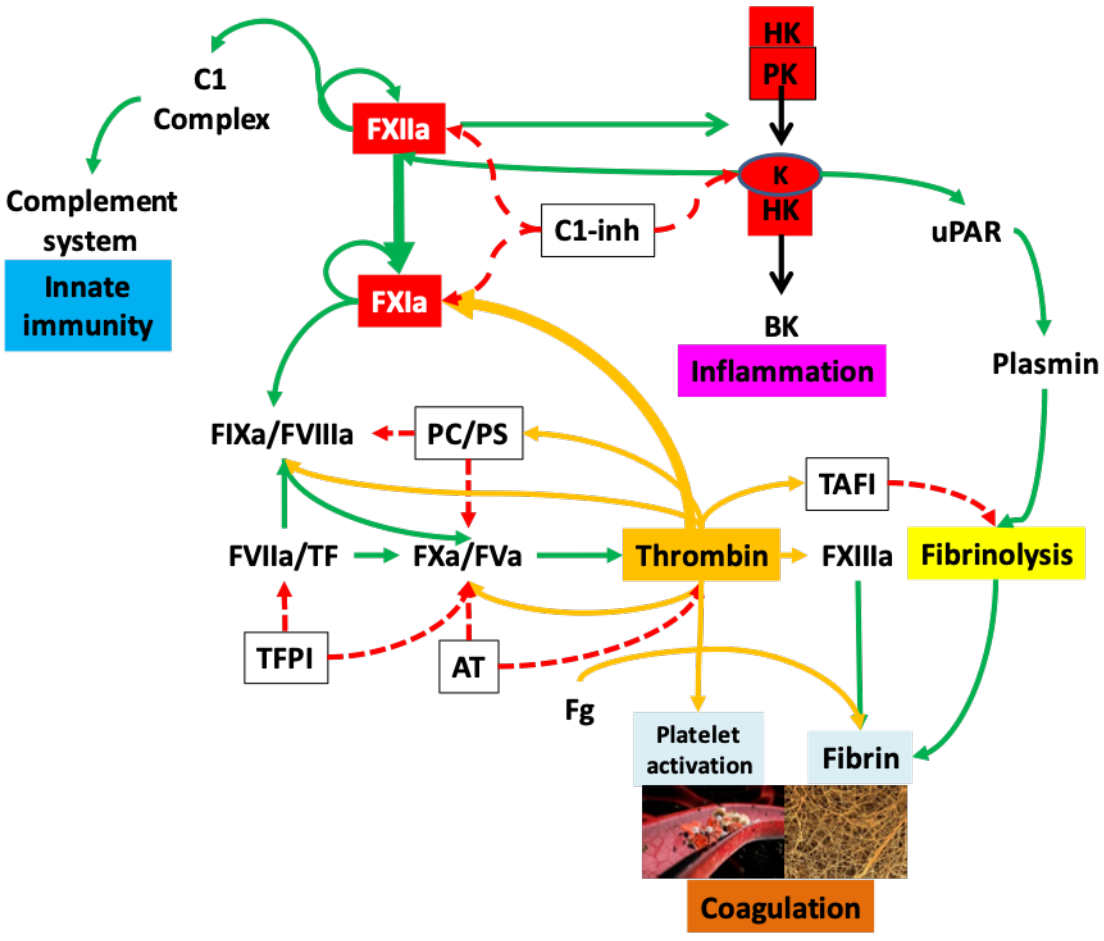


Figure 1: **Schematic representation of clotting cascade and its relationship with inflammation and the innate immunity.** Inhibitors are squared and marked with red dashed arrows. PC/PS: protein C/ protein S; Fg: fibrinogen; TFPI: Tissue factor pathway inhibitor; TF: tissue factor; PK: (pre)kalikrein; uPAR: urokinase-type Plasminogen Activator; AT: antithrombin. C1-inh: Complement 1-inhibitor; HK: High molecular weight kininogen. BK: bradikinin; TAFI: Thrombin-Activatable Fibrinolysis Inhibitor.

These features explain why even minor disturbances of the delicate haemostatic equilibrium may have pathological consequences in both sides of the haemostatic balance: bleeding, if the resulting system is not able to respond accurately, or thrombosis if the response is spontaneous without or minor signal, or uncontrolled (Figure 2)[1].

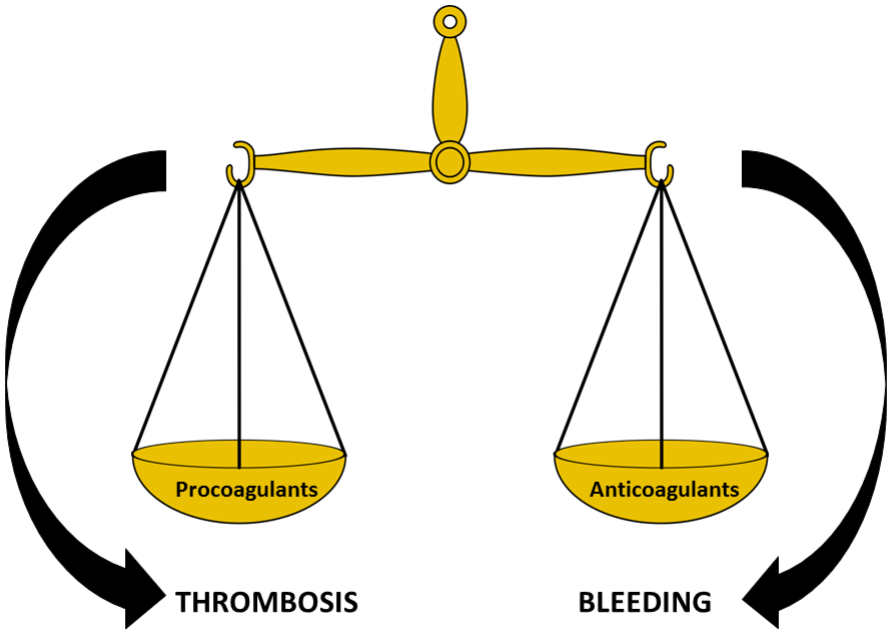


Figure 2: **Haemostatic balance of procoagulant and anticoagulant elements and the pathogenic consequences of its misbalance.**

1.2 Thromboembolism

One in four deaths worldwide registered in 2010 by the Global Burden of Diseases, Injuries, and Risk Factors (GBD) was caused by a thrombotic event in arteries or veins. The formation of a pathogenic clot, which is the cause of ischemic heart disease stroke and venous thromboembolisms, is the main leading cause of death not only in development countries but in the whole world.

Specifically, venous thromboembolism (VT), whose more frequent clinical manifestations are deep venous thrombosis (DVT) and pulmonary embolism (PE), is also a very frequent disorder: 1/1000/individuals/year. This proportion increases in older people and in subjects with specific risk factors (cancer, thrombophilia, etc). Thus, VT is a live-threatening disease that significantly impacts to the global disease burden. In Europe, 900,000 people each year can suffer from a thrombotic event. Taking into account that age is the main risk factor for thromboembolisms, and that thrombosis is a frequent complication observed in patients with tumors, we might expect that the rate of thrombosis will increase exponentially in the near years due to the aging of our population. Venous thrombosis is not only frequent in our population, but also is extremely severe. Up to 100,000 people dead each year due to blood clots. PE is a leading cause of death in pregnant women before or after their delivery. Indeed, 1 in 4 people with PE, die without warning [2].

The morbidity associated to VT is also remarkable. Despite optimum anticoagulation therapy, post-thrombotic syndrome remains one of the major late complications of DVT, with a reported prevalence from 20 to 50%. Moreover,

venous ulcers, and chronic leg swelling are common, costly sequelae of DVT [3, 4] . Also acute PE is a frequent cause of death and serious disability. The risk of PE-associated mortality and morbidity extends far beyond the acute phase of the disease. As many as 30% of the patients died during a follow-up period of up to 3 years, and up to 50% of patients continued to complain of dyspnea and/or poor physical performance 6 months to 3 years after the index event. The most feared "late sequela" of PE is chronic thromboembolic pulmonary hypertension, which can be found in up to 14% in 3 years after the acute event [5].
Finally, we can't forget that despite a correct anticoagulation both as treatment after the thrombotic event and as prophylaxis under risk situations, up to 3 in 10 cases with a first thrombotic event will have rethrombosis within 10 years, rates that can be higher in some patients [6].
In addition to the high impact in morbidity and mortality, VT has also a huge economic impact. It has been estimated that the cost/year of a newly diagnosed case is $7-10 billion in United States. Additionally, treatment can be as much as 15.000 to 29.000 € per person and often results in readmissions to their hospitals [2, 7] .
Despite these data, we must point out that the investment in research in this field is clearly insufficient. The number of resources, including scientist working in this field is low. An example of the modest political interest in this pathology is the absence of specific actions on VT generated by the European or American Scientific authorities. The consequences of the low number of resources destined to the research of this field are critical: a limited knowledge of this disorder, diagnostic methods with important limitations, and a reduced therapeutic arsenal to fight against it. Indeed, only a few numbers of genetic defects associated to VT have been discovered, the mechanisms causing this disorder are not well known yet, and the therapeutic arsenal has not been changed in the last 50 years.

1.2.1 Venous thrombosis

In this thesis, we will focus on VT, because despite it is a complex disease, the number of elements involved is lower than in other thrombotic pathologies, and the medical, economic and social impact of VT is high and exponentially growing. VT includes all ischemic disorders caused by the formation of a clot in a vein. Venous thrombosis mainly localizes in deep leg veins (DVT). PE occurs when clots break off from vein walls and travel through the heart to the pulmonary arteries. The broader term VT refers to DVT, PE, or to a combination of both of them [8] . Blood clots may also be formed in other unusual localizations, such as cerebral, retinal, upper extremity, hepatic, portal, splenic, mesenteric, and renal vein thrombosis [9] .

1.2.2 Thrombotic risk factors

VT is a multifactorial disease that might face up due to a combination of multiple risk factors.
Blood clots affects to everybody at any age, it doesn't discriminate age, races, ethnicities and it affects both men and women. However, there are certain factors and situations that may increase the risk of developing blood clots. The main risk factor for VT is the age. The annual incidence of VT in adults is 1 per 1000 adults, but this rate increases considerably after the age of 45 years [10, 11] . Thus, thrombotic events in paediatric ages are rare (0.07-0.14/10,000/year) in contrast to adults; children's thromboembolism events are associated with additional clinical factors, which in the majority of cases is central venous catheter, underlined diseases, or thrombophilia (Figure 3) [12] .

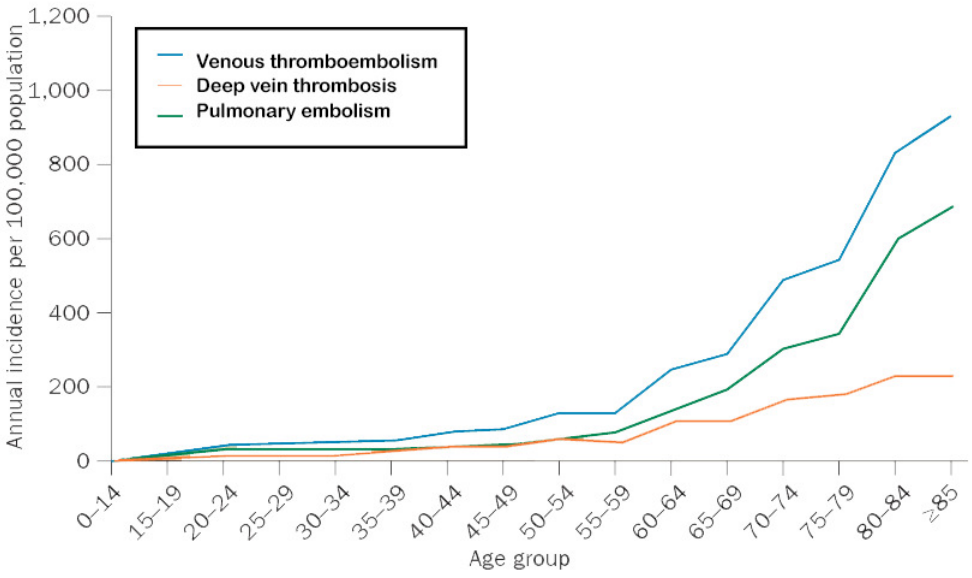


Figure 3: **Evolution of the incidence of venous thromboembolism, deep venous thrombosis and pulmonary embolism according to age.**

In adults there is a higher incidence of venous thrombosis in men than in women. The incidence male to female age-adjusted ratio is 1.2 to 1. Nevertheless, women have higher risk than men during childbearing periods. Beside age and sex, there are other factors increasing the risk of thrombosis, which generated a hypercoagulable state that favours the development of a thrombotic event.

1.3 Thrombophilia

Thrombophilia is defined as an abnormality of the coagulation or fibrinolytic system that results in a hypercoagulable state and increases the risk of an individual for a thrombotic event. It is important to remark that, as thrombosis is a complex disease, the identification of a thrombophilia in a patient, does not mean an invariable development of a thrombotic event. Some people with thrombophilia will never develop thrombosis, others remain asymptomatic until the adulthood, but others have recurrent thrombotic events during their life [13] . Thrombophilias are usually evaluated in patients with thrombosis meeting any of these criteria:
1) Development of an early thrombotic event. The cut-of age for considering a thrombophilic patient is not well defined. Some reviews consider 50 years, but others reduce the age of the first thrombotic event to 45 years.
2) Family history of thrombosis.
3) Patients with recurrent thrombosis.
4) Thrombotic events in unusual localizations: cerebral vein thrombosis, arm thrombosis, mesenteric or portal thrombosis
There are two types of thrombophilia [14] (Table 1):
a) *Acquired thrombophilia* , caused by a factor that appeared only in some moments of the life of a patient. This group of thrombophilias may be classified in two subgroups:
1) Thrombophilic states. The most important according to the incidence in the population and risk of thrombo-

Table 1: **Main congenital and acquired thrombophilias.**

Congenital thrombophilias		**Acquired thrombophilias**	
Common	Rare	Thrombophilic states	Prothrombotic agents
FV Leiden	AT deficiency	AP antibodies	Surgery
PT G202010A	PC deficiency	Cancer (MPN)	Pregnancy
ABO	PS deficiency	HPN	OC & HRT
		HIT	Immobilization

sis are antiphospholipid antibodies. Other thrompophilic states include certain cancers, mainly myeloproliferative neoplasm (MPN), rare diseases such as paroxysmal nocturnal hemoglobinuria (PNH) or treatments such as heparin-induced thrombocytopenia (HIT).
2) Prothrombotic agents. The most important agents triggering thrombosis are surgery and trauma, obesity, and immobilization, including long flights. In women, pregnancy and puerperium, oral contraceptives (OC) and hormone replacement therapy (ORT) also significantly increase the risk of venous thrombosis. Some of these risk factors are modifiable, other ones not, but understanding them is necessary in order to maximize the prevention of this disease in high-risk individuals and groups of patients for the correct prevention and treatment [14, 10] .
b) *Congenital thrombophilia*, caused by a genetic defect. Congenital thrombophilia is present during the whole life of a carrier and may be transmitted to the next generation. Table 1 shows the main congenital thrombophilias identified so far. It is important to point out that genetic variants may also be involved in modulating both the apparition and clinical impact of acquired thrombophilias.

Congenital thrombophilia
As this thesis will be focused on a specific inherited thrombophilia, a detailed introduction of congenital thrombophilias will be done.
The search of genetic predisposition to thrombosis started in 1965, when the Norwegian haematologist O Egeberg described the first haemostatic defect, antithrombin deficiency, in a family with high incidence of venous thrombosis. This study demonstrated the association between the haemostatic defects with the appearance of thrombosis in family members of different generations, supporting a dominant disorder [15] . In the 70's new studies evaluating other thrombophilic families reported new haemostatic defects affecting other natural anticoagulants, protein C and protein S, also associated to the development of thrombotic events. Some years later, when molecular methods were available to sequence the gene encoding these anticoagulants, the genetic base of these thrombophilias was confirmed. These initial molecular studies found mutations in *SERPINC1*, *PROS1* and *PROC* in patients with deficiency of these anticoagulants and thrombosis. The fact that most mutations identified were in heterozygous state and they were very rare, mainly restricted to the studied family, together with the high penetrance of these mutations make to consider congenital thrombophilia as a monogenic autosomal dominant disorder in the 80's. However, only a minor fraction of consecutive patients with venous thrombosis (3-10%) had any of the main congenital thrombophilias described. This and the high heritability of the disorder (up to 60% of the susceptibility to suffer thrombosis is explained by genetic factors [16] , strongly supported the existence of additional prothrombotic genetic risk factors.
The identification in 1994 and 1996 of two common polymorphisms with functional effects affecting elements of the haemostatic system that moderately but significantly increased the risk of venous thrombosis changed the view of the role of genetic factors not only in thrombosis, but in all complex diseases. The first polymorphism described, impairs the inactivation of FV (FV Leiden p.Arg506Gln) [17] and the second, affecting the '3-UTR of F2, the gene encoding prothrombin, associated with moderately high levels of prothrombin (PT 20210G>A) [18] . A race to

find new prothrombotic polymorphism started in the 90's. However, despite the use of massive genotyping methods such as genome wide association studies (GWAS) involving millions of polymorphisms in thousands of patients and controls. Only one additional common genetic risk factor with relatively strong prothrombotic effect was confirmed, the ABO blood group, which was already associated to VT in the early 60's!, and only few supplementary polymorphisms with very mild risk of thrombosis (OR: 1-1.2) have been found [19, 20] .
The development of massive sequencing methods (NGS), which has revolutionized the identification of genetic defects in other diseases, has only a moderate effect in the identification of new genetic risk factors involved in thrombosis. Only one recent study finds a new gene *STAB2*, as potentially involved in VT [21] .
Unfortunately, to date only few new congenital thrombophilias have been described, and their discovery followed a classical design: the analysis of thrombophilic families who had an aberrant intermediate phenotype that have guided the search for the associated genetic defect by direct sequencing to the candidate gene. Thus, one thrombophilic family with increased FIX activity had a mutation in the *F9* gene (FIX Padua) [22] , aberrant prothrombins with resistance to antithrombin caused by different mutations in *F2* have been described in few families from different countries [23, 24, 25, 26] , and two thrombophilic families with high FVIII levels caused by a partial duplication affecting *F8* have just been described [27] .
Accordingly, additional genetic defects, both in new genes as well as new genetic defect in classical genes must be identified, as we are far away to know only a minor fraction of the molecular base of VT [28] .
Finally, a correct diagnosis of congenital thrombophilia is important not only to know the molecular base of a thrombotic event. The diagnosis of congenital thrombophilias may have clinical usefulness as it may assist the clinical management of carriers.
While the routinely screening of mild prothrombotic polymorphisms in a clinical practice has been questioned because their negligible clinical usefulness [29] , the identification of severe congenital thrombophilias may guide the length of the treatment in symptomatic carriers, and certainly facilitate antithrombotic prophylaxis in asymptomatic carriers thus reducing the incidence of recurrences or first thrombotic events, respectively [30] [31] .
As antithrombin deficiency is the first and most severe thrombophilia, in this study we will focus on this thrombophilic defect.

1.4 Antithrombin

Antithrombin is a natural anticoagulant which plays an important role in the haemostatic system. The first mention to antithrombin was done by A. Schmidt in 1892, who suggested the presence of a plasma cytoglobulin that prevented blood coagulation. The first person using the term antithrombin was P. Moratowitz in 1905. But the relevance of antithrombin in haemostasis and thrombosis was launched by the association with venous thrombosis of a heterozygous deficiency observed by O. Egeberg in 1965 [15] . Since then, we have assisted to a deep characterization of this molecule from a biochemical, molecular, and clinical perspective. We can say that antithrombin is an old and well-known natural anticoagulant. However, despite more than 55 years of intense search, there are still many challenges and interesting points that remain still to explore to fully unravel this key anticoagulant.

1.4.1 Biochemical characteristics of antithrombin

Human antithrombin is a hepatic protein of 58 KDa of molecular weight, which is secreted into plasma reaching a concentration of 150 μg/mL. The half-life of antithrombin in plasma is 65 hours [32] .
Like for many other haemostatic factors, antithrombin levels in plasma show a normal distribution in the general

population, with a wide range of levels: 70-120%. Factors such as sex, body mass index, oral contraceptives or race, may affect to antithrombin levels. Furthermore, genetic factors must also play a role in determining antithrombin levels as revealed by the high heritability of this trait (h= 0.486) [33] .
Human antithrombin has 464 residues, with a signal peptide of 32 aminoacids that during the synthesis in the hepatocyte, is released by unknown proteases and directs the translation of the protein to the endoplasmic reticulum. There, the protein is folded and suffer two post-translational modifications: 3 intramolecular disulphide bonds and N-glycosylation. Antithrombin has four potential N-glycosylation sites (Figure 4).

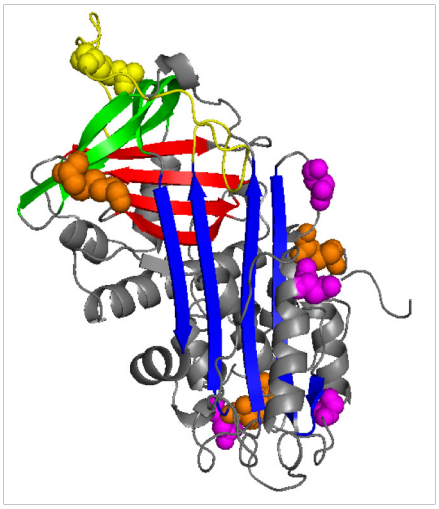


Figure 4: **Structural representation of native human antithrombin.** The reactive centre loop is colored in yellow, and the P1 residue (Arg424) is represented as yellow balls. The Asn that are glycosylated are shown as pink balls, and Cys involved in intramolecular disulphide bonds are marked as orange balls. The A-sheet is coloured in blue; the B-sheet in red and the C-sheet in green.

The presence of a serine instead of threonine at position 169 (137 in the mature protein without the signal peptide) reduces the efficiency of N-glycosylation at the consensus sequence of Asn167 leading to an inefficient glycosylation of that residue. This inefficient N-glycosylation explains the generation of two antithrombin glycoforms in plasma: alfa, with 4 N-glycans, the major glycoform of antithrombin in circulation, as it represents 90% of total plasma antithrombin; and beta, with 3 N-glycans (lacking the glycan at Asn167) representing the 10% of plasma antithrombin (Figure 5). The absence N-glycan at Asn167 increases 6-fold the affinity for heparin [34] . This explains why, despite being the minority antithrombin glycoform in plasma, β-antithrombin could be the most relevant regulator of thrombin, particularly in endothelial surface [35] , explaining at the same time, its increased clearance rate which contributes to the smaller concentration of β-antithrombin in plasma [36] .

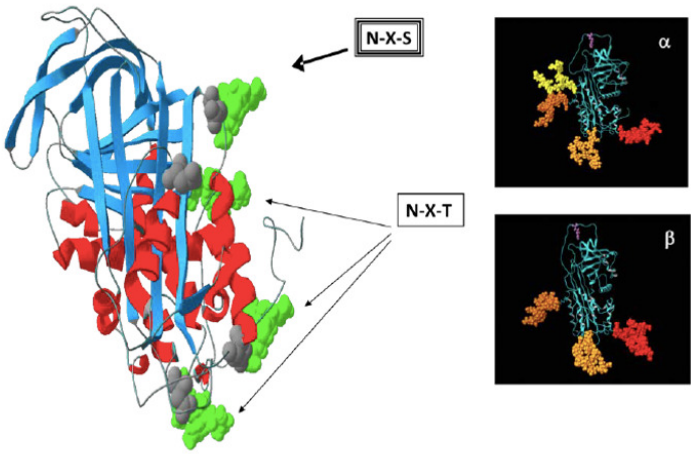


Figure 5: **Inefficient glycosylation at Asn167 explain the existence of two glycoforms of antithrombin in plasma: α with 4 N-glycans, and β with 3 N-glycans.**

1.4.2 Antithrombin: a haemostatic serpin

Antithrombin belongs to the superfamily of serpins. The superfamily of serpins (SERine Protease Inhibitors) spans more than 1,500 proteins [37] that share close molecular, structural and functional similarities to three proteins: human antithrombin, human α1-antitrypsin and chicken ovoalbumin [38] . The majority of serpins have serine protease inhibitor activity, although some of them have caspase activity, and some serpins do not have inhibitory activity [39] . Serpins, particularly inhibitory serpins, play crucial roles as regulators of many different systems, specifically those based on proteolytic cascades both, inside or outside the cell. Consequently, they play important roles in different processes such as inflammation, apoptosis, complement, and of course, in coagulation and fibrinolysis [40, 41, 42] .

Serpins share a common structure, which is required for the efficient mechanism of inhibition of these proteins. Most of the serpins, as we have seen in antithrombin, comprise 400-500 residues, a molecular weight of 37-70 KDa and an amino acid homology of 35% [43] . They are folded in a very conserved tertiary structure with 3 β-sheets (A-C) and 8 or 9 α-helixes (A-I) with a flexible and mobile exposed loop responsible for the interaction with the target protease, which is known as Reactive Centre Loop (RCL), being the P1 the residue conferring the specificity for the targeted protease [44] [45] (Figure 4).

The RCL of serpins acts as a protease-specific bait. However, unlike other families of serine protease inhibitors, which function in a simple way by extending a substrate-like RCL that passively blocks the active center of the target protease [46] , the RCL cleavage of serpins by target proteases results in inhibition by an unique and extraordinary efficient suicide mechanism that has been compared with a mouse-trap (Figure 6).

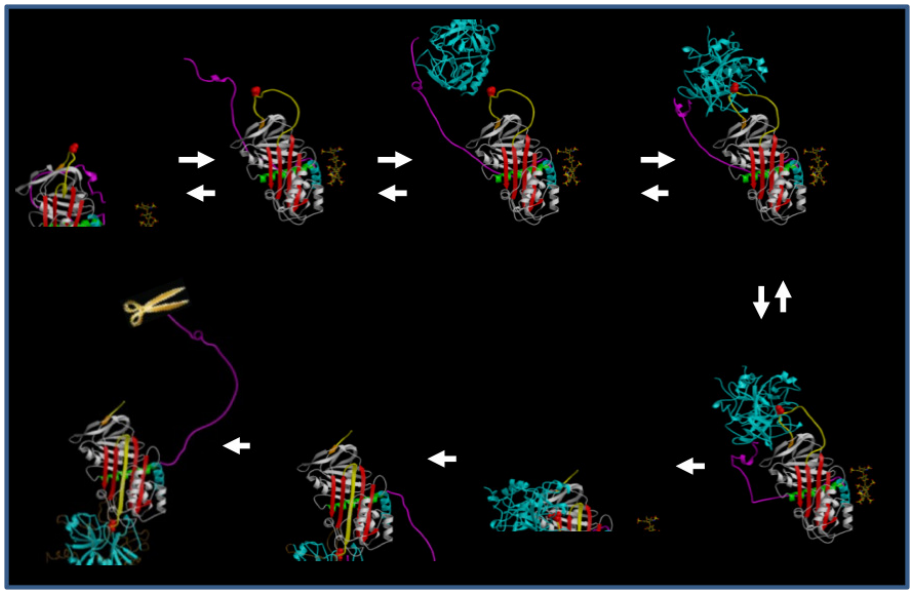


Figure 6: **Schematic representation of the inhibition of a target protease (thrombin FIIa in blue) by antithrombin after activation by heparin.** When the protease binds the serpin, it cleaves the P1-P1' bond of antithrombin (red ball). This cleavage triggers a conformational change in the serpin that introduces the cleaved RCL (yellow) into the central A-sheet. This movement translocates the protease to the other pole of the serpin, forming new covalent interactions (thrombin-antithrombin complex) that disturb the catalytic activity of the protease. When the protease binds the serpin, it cleaves the P1-P1' bond of antithrombin (red ball). This cleavage triggers a conformational change in the serpin that introduces the cleaved RCL (yellow) into the central A-sheet. This movement translocates the protease to the other pole of the serpin, forming new covalent interactions (thrombin-antithrombin complex) that disturb the catalytic activity of the protease.

Specifically, for antithrombin, the inhibition of target proteases is triggered by proteolytic cleavage of the bound between Arg-425 and Ser-426 residues (P1-P1'). Moreover, in addition to the specific RCL, antithrombin has an additional difference compared with other serpins, the RCL of native antithrombin remains partially inserted into the top of the central A β-sheet (Figure 7) [47] . This native conformation together with the saline bridge between P1 residue (Arg425) and Glu269 (237 in the mature protein) [48] justify the relative inaccessibility of the target protease to the P1 residue and so the low anticoagulant activity of the native form of antithrombin. In fact, the full anticoagulant activity of antithrombin requires the activation of this molecule triggered by heparin and other glycosaminoglycans.

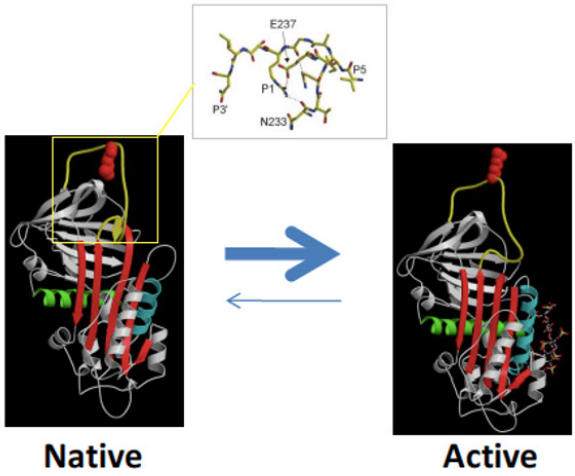


Figure 7: **Native and active forms of antithrombin.** In the native form, the RCL (yellow) is partially inserted into the central A-sheet, and the P1 residue (red balls) which is interacting with E269 (237 in the mature protein), is not exposed. Binding of heparin to antithrombin activates this anticoagulant, releasing the RCL and fully exposing the P1 residue.

The binding of heparinoid molecules to antithrombin occurs at the heparin binding site of the serpin, in which the D helix plays a key role (Figures 6 and 7) [49] . An initial weak complex with a specific pentasaccharide sequence in heparin is formed by electrostatic interactions between positively charged residues of antithrombin (mainly arginine and lysine residues) and negatively charged sulphates of heparin [48] . This interaction causes complex sequential conformational changes in the serpin. A new turn of the D helix is generated, and the extended D helix closes the central A-β, which causes the release of the partially internalized RCL. The activation of antithrombin culminates with the exposure of the P1 lateral chain adopting a more favorable conformation for binding the target serine protease (Figure 8) [49] [50] [51] .

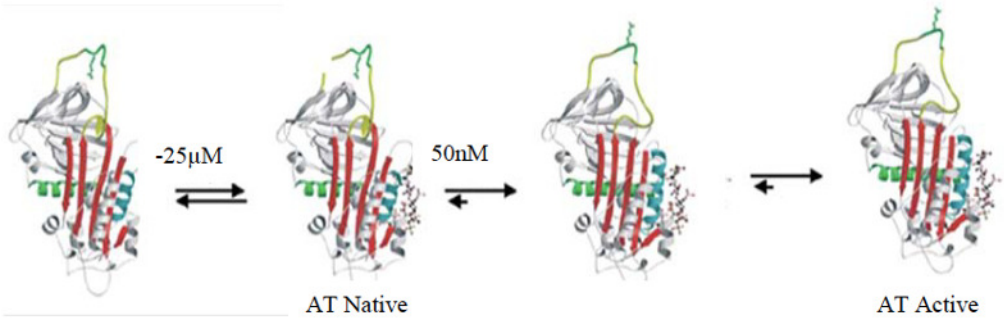


Figure 8: **Sequential conformational changes induced by heparin in antithrombin that result in its activation.**

The increased anticoagulant activity of activated antithrombin had great clinical impact and sustained the clinical use of heparins as potent anticoagulants. Three main types of heparins are available for clinical use: unfractionated

heparin, with more than 5.400 Da, does not only activate antithrombin, but also form a ternary complex with thrombin, increasing the anti-FIIa activity more than 1000-fold. Low molecular weight heparins, are smaller molecules that only activated antithrombin, thus having a main anticoagulant impact in the inhibition of FXa (300-fold compared with the activity of native antithrombin). Finally, the essential pentasacharide is the smallest heparin able to activate the anti-FXa activity of antithrombin in a similar way (Figure 9) [52] .

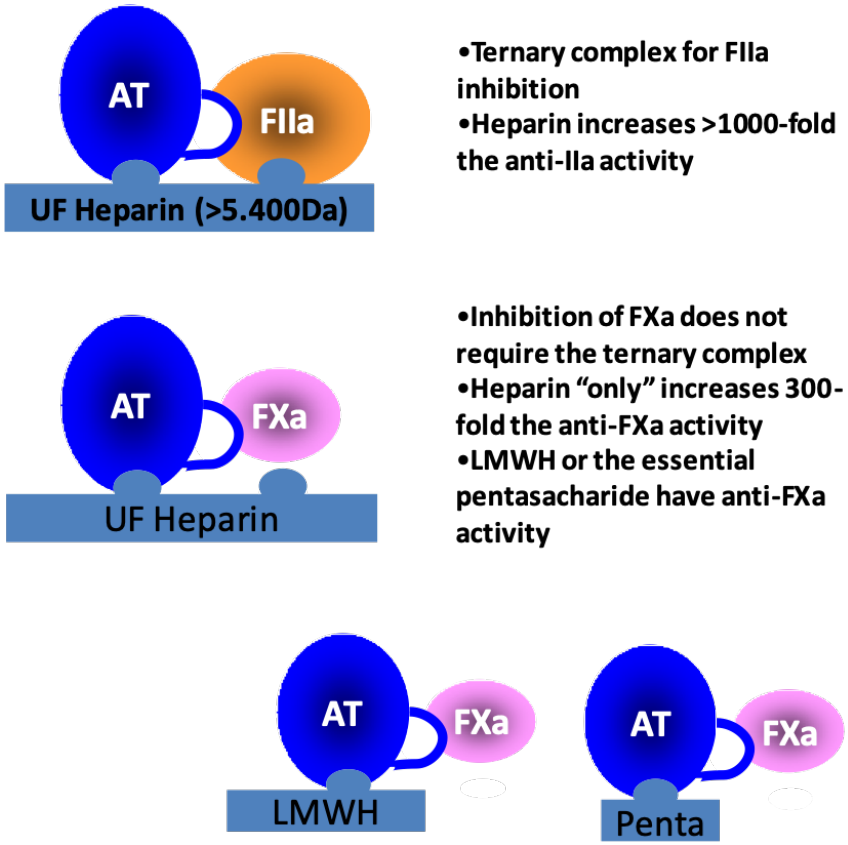


Figure 9: **Mechanism of action of heparins with clinical use.** Unfractionated heparin (UF), low molecular weight heparin (LMWH) and the essential pentasacharide (Penta), AT: antithrombin.

Antithrombin needs an extraordinary structural flexibility in order to properly fulfil its inhibitory mechanism. However, this flexibility also makes antithrombin susceptible to suffer conformational changes that inactivates its anticoagulant activity. The native conformation of antithrombin, with the RCL partially inserted makes antithrombin much more sensitive than other serpins to even to minor structural changes that could potentially result in a miss-folded conformation or inactive forms. Thus, in the normal native state, antithrombin contains a 5-strand β -sheet (β -sheet A) and a surface-exposed RCL (bridging the C-terminus of strand 5A to the N-terminus of strand 1C). However, this conformation is not the most stable. An astounding increase in thermodynamic stability (best estimate 32 kcal mol-1) [53] can be achieved through the incorporation of the RCL into β-sheet A, triggered through extension of strand 1C, to form the so-called latent conformation, a relaxed conformation which is not obviously inhibitory (Figure 10). The identification of up to 5% of plasma antithrombin in a latent conformation has suggested that this transformation could be a consequence of the senescence of the native molecule [54, 55] .

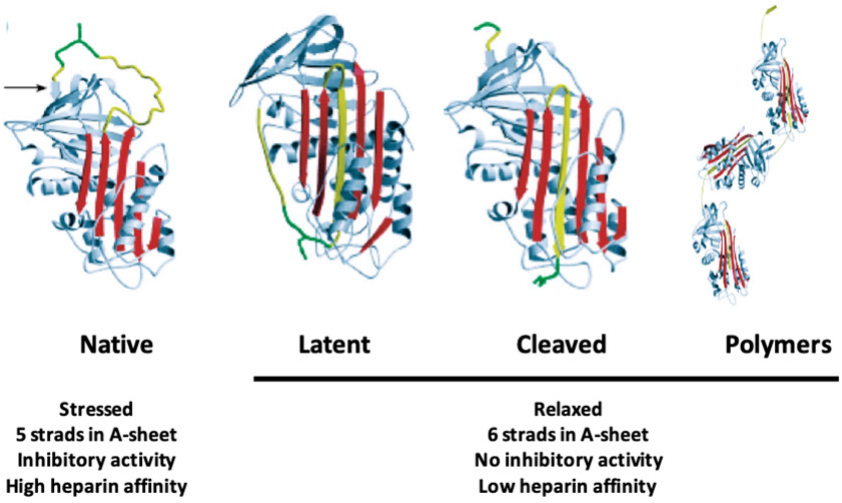


Figure 10: **Conformations of antithrombin.** The structural and functional characteristics of these conformations are also shown.

Moreover, the cleavage of the RCL by non-targeted proteases, such as elastase, induces the insertion of the cleaved RCL into the central A-sheet, generating a new relaxed conformation (cleaved antithrombin), that also has no anticoagulant activity. Both, latent and cleaved antithrombin had low heparin affinity (Figure 10).
Finally, as other serpins, the RCL of antithrombin may also be inserted into the central A sheet but from other molecule, leading to the formation of polymers, which usually are formed intracellularly during abnormal folding, and are not secreted (Figure 10) [56] . Mutations of certain residues involved in the folding to the native conformation of antithrombin may facilitate or induce the transition to the latent or polymer conformation, with obvious pathogenic consequences [57] .

1.4.3 Antithrombin functions

In addition to its key role in coagulation, antithrombin has also been described to play other biological functions. The relaxed conformations of this molecule (cleaved or latent) have strong anti-angiogenic activity [58] . Besides, a potent anti-inflammatory activity has also been tightly associated with its heparin-binding domain [14] . An antiviral activity of antithrombin has been suggested [59], and recently, a key anti-bacterial activity of antithrombin, particularly the beta glycoform, has been demonstrated [60] . Finally, our group has found evidences suggesting a potential anti-apoptotic and anti-tumoral role of antithrombin [61, 62] .

1.4.4 Genetics

SERPINC1, the gene encoding antithrombin (GenBank X68793.1, OMIM # 107300) is located at chromosome 1q23-25.1, coordinates: chr1:173,903,519-173,917,378 (hg38) in minus strand. *SERPINC1* comprises 7 exons, encompassing 13.5 Kb of genomic DNA (Figure 11).

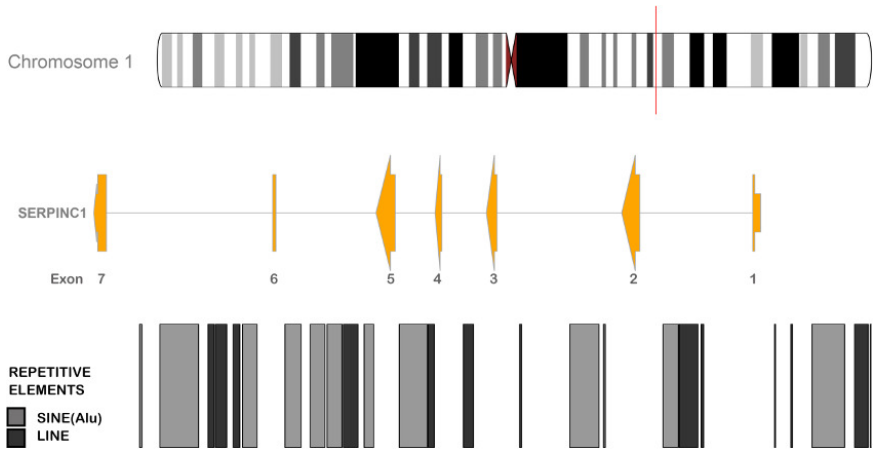


Figure 11: **Localization and organization of the *SERPINC1* gene in chromosome 1.** Repetive elements are also shown.

The promoter region of this gene is largely unknown. The promoter lacks TATA element and GC-rich regions [63, 64, 65] . Only few studies have screened for potential regulatory elements in *SERPINC1*, most of them using deletional analysis and reporter or DNase I footprint assays. These studies have identified few regions potentially involved in the transcriptional regulation of this gene. The analysis of cases with moderate antithrombin deficiency allowed our group to identify regulatory mutations in introns of *SERPINC1* [66, 67] .
The genetic diversity of *SERPINC1* is also remarkable by the low genetic variability described in general populations. So far, only 267 polymorphisms have been identified in *SERPINC1*, most of them located in introns, and therefore with functional consequences unknown for most of them. Our group has described that a single nucleotide polymorphism (SNP), rs2227589, located in intron 1 explains up to 7.2% of the interindividual variation of antithrombin [68] . The search for new genetic factors associated with the interindividual variability of antithrombin, must continue by evaluating other *SERPINC1* polymorphisms, and other genes that may influence indirectly to plasma levels or the anticoagulant function of this key haemostatic molecule. Interestingly, a high number of SINE (short interspersed element) and LINE (long interspersed element) repetitive elements are present in and surrounding this gene (Figure 11). These repetitive elements SINE and LINE are retrotrasposable DNA sequences or mobile DNA which accounts almost half of the human DNA sequence and are spread throughout the genome. These elements are colloquially known as “jumping genes” due to their ability to jump and move over the genome. Retrotransposable elements are copied into RNA, and then copied back into DNA, which is then integrated in the genome. Mobile DNA has been instrumental in shaping the structure, function, and evolution of the human genome [69] .

1.5 Antithrombin deficiency

As previously described, the potent and wide inhibitory mechanism of antithrombin explains the key haemostatic role of antithrombin, and therefore, why even moderate reductions of antithrombin activity increase the risk of venous thrombosis [70] . Moreover, congenital antithrombin deficiency, which may be considered when the anticoagulant activity of one patient is below the lowest value found in the general population, usually $< 80\%$ of a reference plasma generated pooling plasma of more than 100 healthy subjects, is the strongest inherited throm-

bophilia, that increases between 10 to 50 fold the risk of thrombosis at any age (Figure 12).

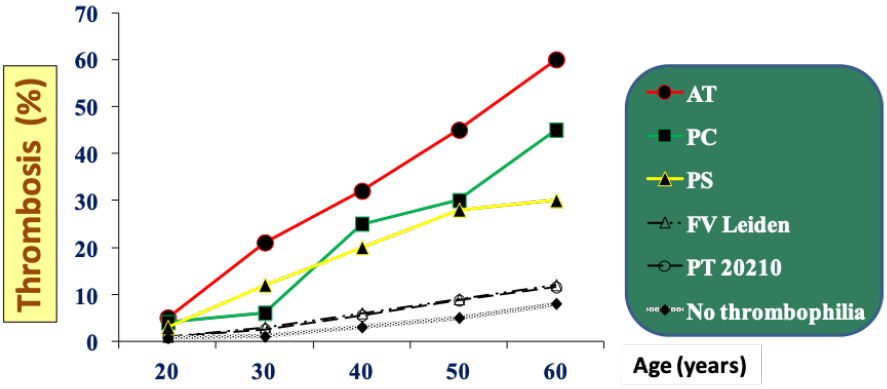


Figure 12: **Risk of thrombosis according to age in different congenital thrombophilias.**

The strong risk of thrombosis associated to congenital antithrombin deficiency also explains by this was the first inherited thrombophilia described more than 55 years ago (Figure 13).

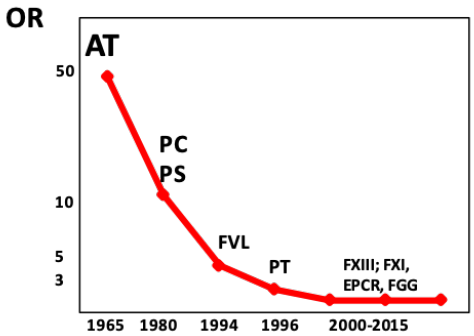


Figure 13: **Risk of thrombosis according to age in different congenital thrombophilias.**

Finally, another indicator of the relevance of antithrombin, is the lethal consequences of the complete absence of this anticoagulant. Knock out (KO) animal models associate with embryonic lethality [71] . No patients with very low levels of antithrombin due to congenital defects have been described and only few homozygous for mild deficiencies have been described [72] . Our group has recently described the first severe compound antithrombin deficiency in a dead human embryo [73] . However, as we will see in this study, the clinical impact of antithrombin deficiency is quite wide, from very mild, to severe clinical phenotype revealed by early (pediatric), recurrent and even lethal thrombosis.

As for the general population, and all congenital thrombophilias, age also increase the risk of thrombosis in patients with antithrombin deficiency, but this is the congenital thrombophilia most sensitive to age, and up to 60% of patients with antithrombin deficiency older than 60 years have developed a thrombotic event (Figure 12) [30].

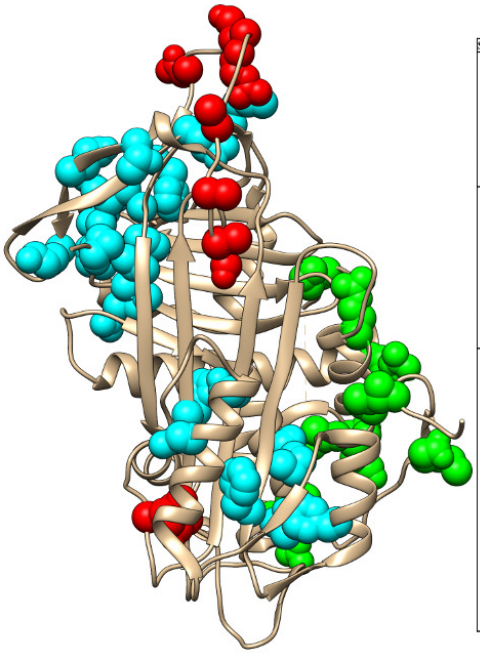
1.5.1 Types of antithrombin deficiency

Antithrombin deficiency may have two different causes: acquired or congenital. External factors affecting the synthesis, secretion, or consumption of antithrombin may cause an acquired antithrombin deficiency. Among these factors, we remark some drugs, such L-asparaginase [74], and some pathogenic conditions such as severe trauma, the thrombotic event that cause consumption of this anticoagulant, or liver diseases such as cirrhosis, that significantly impairs its synthesis [75]. Acquired antithrombin deficiency is a transient disorder, and normal levels may be reached once the factor disappears. Congenital antithrombin deficiency is, in most of cases, maintained during the whole life of the patient since it is caused by genetic defects. Most of genetic defects causing antithrombin deficiency affect the coding gene *SERPINC1*. However, it is important to point out that other gene defects may also cause antithrombin deficiency, as indicated latter.
Classically, two types of congenital antithrombin deficiency are defined according to the levels of the mutant antithrombin detected in plasma (Table 2).

Table 2: **Characteristics of different types of antithrombin deficiency** Characteristics of different types of antithrombin deficiency

Type	**Subtype**	**Plasma antithrombin**				**Gene variants**
		Anticoagulant activity		Antigen	Heparin affinity (CIE)	
		Heparin cofactor activity	Progressive activity			
I		40-60%	40-60%	40-60%	No	Nonsense, frameshift, splicing, gross gene defects, missense, small INDEL
II	RS	40-60	40-60	80-100	No	Missense (Mainly in the RCL)
II	HBS	40-60	80-100	80-100	Yes	Missense (Mainly in the heparin binding domain)
II	PE	40-60	40-60	80-100	Yes	Missense (Mainly in the C-sheet)

• **Type I deficiency** (quantitative deficiency). Antithrombin in plasma showed reduced levels (around 50% of the levels detected in reference plasmas) and similar reductions in anticoagulant activity (around 50%). Type I antithrombin deficiency is easily explained by nonsense, frameshift, or splicing mutations, as well as large deletions of the *SERPINC1* gene. Interestingly, up to 26 missense mutations associate with type I deficiency. Type I deficiency is usually severe and very rare in the general population.
• **Type II deficiency** (qualitative deficiency). This type of deficiency is considered when the variant protein is synthesized and secreted to the plasma with a normal or slightly reduced rate, but they can be detected at least by antigenic methods as the molecule is functionally defective. Type II are usually the consequence of missense single amino acid changes. There is a wide clinical heterogeneity among carriers of a type II deficiency, from very mild to very severe defects [75] . Within type II deficiency, three subtypes can be distinguished according to the functional defect, which is usually also associated with the location of the pathogenic mutation and the functional defect (Figure 14):
a) IIa or reactive site (RS), when the variant antithrombin has impaired reactivity with the target protease or the inhibition process is not effective, and the stoichiometry of the inhibition is higher than 1. Most mutations causing this deficiency affect residues of the RCL, but not all, as recently described by our group [76] .
b) IIb or heparin binding site (HBS), if the mutation, usually located at the heparin binding domain, disturbs the interaction of heparin to antithrombin or the activation induced by this binding. Type II HBS deficiency had impaired anticoagulant activity in assays with heparin (heparin cofactor activity) but nearly normal activity in assays without heparin (progressive activity).
c) IIc or pleiotropic (PE), when the mutation affects both the reactivity and heparin affinity. In a recent manuscript with my authorship, our group showed evidences supporting that mutations causing type II PE deficiency induce or cause the transition to the latent conformation, the ideal pleiotropic form, as it has no reactivity and has low heparin affinity [77] .



Subtype	Residue	Mutation	Position	Comments
RS	Ser412	Arg	P14 RCL	
	Ala414	Thr/Deletion	P12 RCL	
	Ala416	Ser/Pro	P10 RCL	Ala416Ser detected in homozygosis
	Ala419	Val	P7 RCL	
	Gly424	Asp	P2 RCL	
	Arg425	His/Pro/Cys/Deletion	P1 RCL	
	Asn430	Asp	P2' RCL	
	Ser426	Leu	P6' RCL	
	Ser397	Leu	Loop s5A	
	Ile239	Thr	Loop s3A	
HBS	Ile39	Ans	N-term	
	Arg45	Trp	N-term	
	Arg46	Cys	N-term	
	Arg56	Cys	N-term	
	Pro73	Leu	N-term	
	Arg79	Cys/Ser/His	Helix A	
	Leu131	Phe/Val	Helix C	Leu131Phe detected in homozygosis
	Lys146	Glu	Helix D	
	Ser148	Pro	Helix D	
	Ile151	Phe	Helix D	
	Ala156	Asp	Helix D	
	Arg161	His/Gln/Leu	Helix D	
PE	Ala120	Val	Helix B	
	Met121	Thr/Ile	Helix B	
	Ala126	Gly	Helix B	
	Phe179	Cys	s2A	
	Tyr190	Cys	Helix E	
	Asn219	Asp/Lys	Helix F	
	Thr250	Ile	s3A	
	Glu269	Lys	s4C	
	Lys273	Glu	s4C	
	Met283	Ile/Val	s3C	
	Met347	Lys	s2C	
	His401	Tyr	s5A	
	Phe434	Ser/Lys/Leu	s1C	
	Ala436	Thr/Asp/Gly	s4B	
	Asn437	Lys/Ser	s4B	
	Arg438	Gly/Met	s4B	
	Pro439	Thr/Leu/Ala/Arg	s4B	
	Phe440	Ser	s4B	
	Arg457	Thr	s5B	
	Val458	Gly	s5B	
	Pro461	Leu/Ser	s5B	Pro461Leu detected in homozygosis

Figure 14: **Localization in the three dimensional structure of antithrombin of mutations described in *SERPINC1* causing all type II deficiencies.** Variants associated to type II RS are shown in red, those associated to type II HBS are shown in green and those associated to type II PE are shown in cyan.

Most cases with antithrombin deficiency are caused by mutations in *SERPINC1*. In the free access version of the Human mutation database (HMDG) (last access January 2021) there are 352 different mutations described in *SERPINC1* and identified in patients with antithrombin deficiency. Most of them (90,6%) are point mutations or small INDELs (affecting less than 50 bp). Only 34 gross genetic variants (9,7%) have been reported in *SERPINC1* as causal of antithrombin deficiency 15). As antithrombin is a dominant negative disorder, most cases have heterozygous mutations. Homozygous mutations have only been detected for mild type II variants such as the type II HBS Budapest 3 (p.Phe131Leu), or Toyama (p.Arg79Cys) variants. These homozygous variants have very high risk of thrombosis and thrombotic events are usually very early and recurrent [78] . Our group has also identified homozygous carriers of antithrombin Cambridge II (p.Ala416Ser) a mild type II RS deficiency [79] and Dublin (p.Val30Glu) a defect that only shows its pathogenic effect under stress conditions [80] . Compound heterozygous are very rare and usually associate with very severe thrombosis [81] and we have recently described, in a paper in which I am coauthor, the first compound heterozygosis identified in a natural abortion [73] the first evidence of the lethality of very severe deficient states in humans.

Most mutations have pathogenic consequences by a loss-of-function mechanism affecting both, the levels or function of the variant. However, there are mutations that also cause a dominant negative effect. The best example is antithrombin London, a type II RS deficiency caused by the deletion of Arg425. The loss of function caused by this mutation is obvious as this residue is P1, crucial for the inhibitory function of antithrombin. Accordingly, this variant that is secreted normally to the plasma, but it has no anticoagulant activity at all. Interestingly, this deletion also significantly increases the heparin affinity of the variant. Thus, under low concentrations of heparin, the cofactor is bound to the inactive variant, which avoids the efficient activation of the wild-type molecule in heterozygous carriers [82] . We have also identified variants that get a dominant negative effect by a different conformational

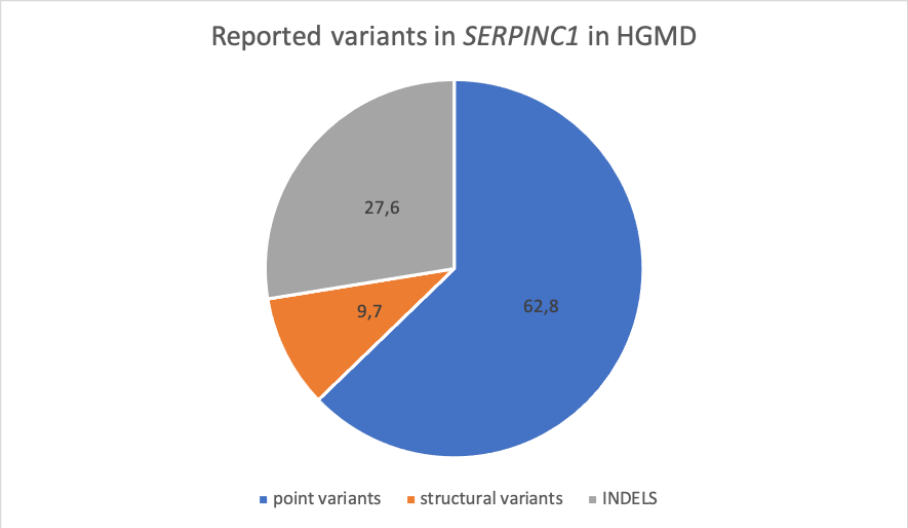


Figure 15: **Proportion of different types of *SERPINC1* mutations in patients with antithrombin deficiency.**

mechanism, as the mutation facilitates polymer formations that may also incorporate wild type monomers [83, 84] . Finally, we suspect that certain mutations in addition to the loss of function may get a gain of function with potential pathogenic consequences. The best example is the mutation affecting the initiation codon (Met1). The proposed missense change, for example we identified the p.Met1Ile mutation, never takes place, as the ribosome does not start the translation at the mutated codon, but continues the scanning of the RNA to find downstream alternative initiation codons. The result, as demonstrated by our group in a recombinant model, is a smaller protein lacking the signal peptide, that does not enter in the endoplasmic reticulum and thus it is not glycosylated, and is not secreted. The functions of this aberrant cytoplasmic molecule are unknown, but could gain a new prothrombotic function as it has been identified in the probably most severe family with antithrombin deficiency we have ever studied [85] . However, despite the high number of variants reported in *SERPINC1*, up to 20% of patients with antithrombin deficiency are still without molecular base. This could be due to the presence of other variants not detected by current diagnostic methods. Regulatory elements of *SERPINC1* or SVs affecting introns are good candidates as these regions are not routinely tested by sequencing or Multiplex Ligation-dependent Probe Amplification (MLPA). But other genes could also be involved in antithrombin deficiency. In fact, our group has reported that disorders of N-glycosylation explain up to 5% of cases with antithrombin deficiency [86] since a correct N-glycosylation is crucial for an efficient folding and intracellular traffic. Our preliminary results show that hypoglycosylated molecules are less efficiently secreted.

1.5.2 Diagnosis

Diagnosis of antithrombin deficiency relies on the classical functional, biochemical and molecular analysis, as for all severe thrombophilic states [87] . First, it requires demonstration of low plasma antithrombin activity; below normal range: $<80\%$ by using a functional assay, usually anti-FXa in the presence of heparin [88] . A correct diagnosis of antithrombin deficiency requires a positive functional finding in more than one sample, which never must be collected during or close to the thrombotic event. After that, characterization of the type of deficiency usually requires from additional biochemical analysis (progressive activity, antigen values and heparin affinity). Finally, genetic analysis for identification of a pathogenic variation in *SERPINC1*, which is only done by few reference centers, is not only confirmatory but may supply additional information that might facilitate the clinical management of the patient [88] .

Molecular analysis of cases with antithrombin deficiency usually starts by Sanger sequencing the 7 exons and flanking regions of *SERPINC1*, as this approach, which is fast and cheap, get positive results in 70% of cases. Only MLPA is done in cases with negative results. Figure 16 shows the diagnostic algorithm recommended for an accurate diagnosis of antithrombin deficiency.

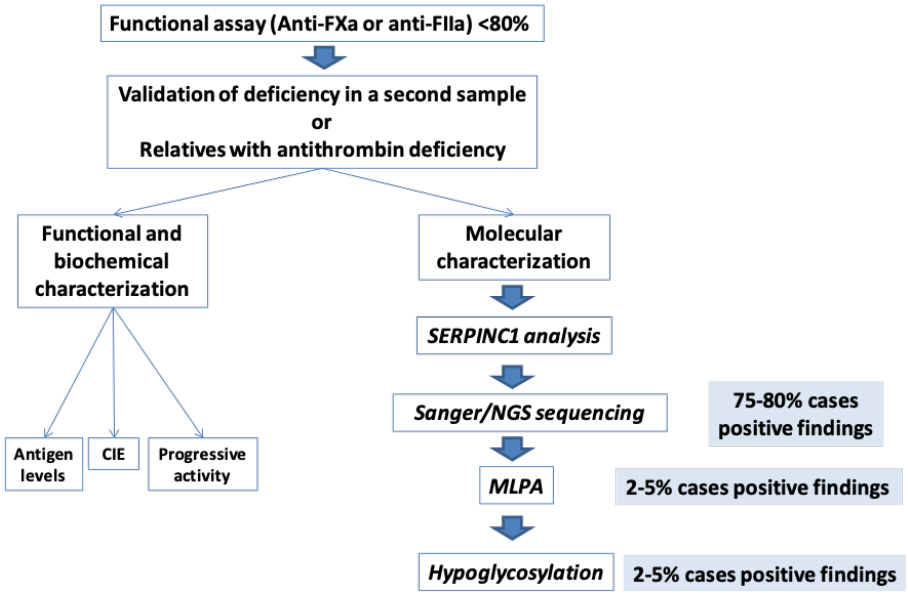


Figure 16: **Current algorithm for the diagnosis and characterization of antithrombin deficiency.**

Testing of antithrombin deficiency has clinical usefulness. First, a positive finding allows the use of a specific treatment: a substitutive therapy with concentrates of plasma or recombinant antithrombin [1] . Moreover, diagnosis of antithrombin deficiency has been demonstrated to be useful in: 1) symptomatic patients to prolong the anticoagulant treatment aiming to avoid recurrence and 2) asymptomatic family members of a proband with known congenital antithrombin deficiency, for thromboprophylaxis under risk situations aiming to avoid the first event. However, some authors still show some concerns about the diagnosis of any thrombophilia, including antithrombin deficiency, particularly who must be screened and when the screening must be done. These questions are especially controversial for the diagnosis of antithrombin deficiency in children, and the prenatal diagnosis of this disorder [29] .

1.6 Sequencing technologies

Watson and Crick, in 1953 began the revolution of DNA understanding by solving the three-dimensional structure of DNA working from crystallographic data produced by Rosalind Franklin and Maurice Wilkins [89] , which allows the knowledge of DNA replication and encoding protein in nucleic acid. However, it wasn't until 1968 when Wu and collaborators resolved the first DNA sequence: 12 bases of bacteriophage lambda [90] . This was the first step of a race to unravel the human genome. A race that was slow during the first decade. In 1973, Gilbert and Maxam were able to report 24 bases of the lactose repressor binding site. This study took two years, one base per month! [91] . The development of two methods that decode hundreds of bases in an afternoon transformed this field in around 1976 and accelerated the rhythm of advances of DNA sequencing technologies. These two methods were the chain terminator procedure developed by Sanger and Coulson and the chemical cleavage procedure devel-

oped by Maxam and Gilbert giving rise to what now we know as "First Generation Sequencing". This brought to different important milestones in sequencing methods that are shown in figure 17 which, as we can observe, in the last years the technology has been growing faster [92] .

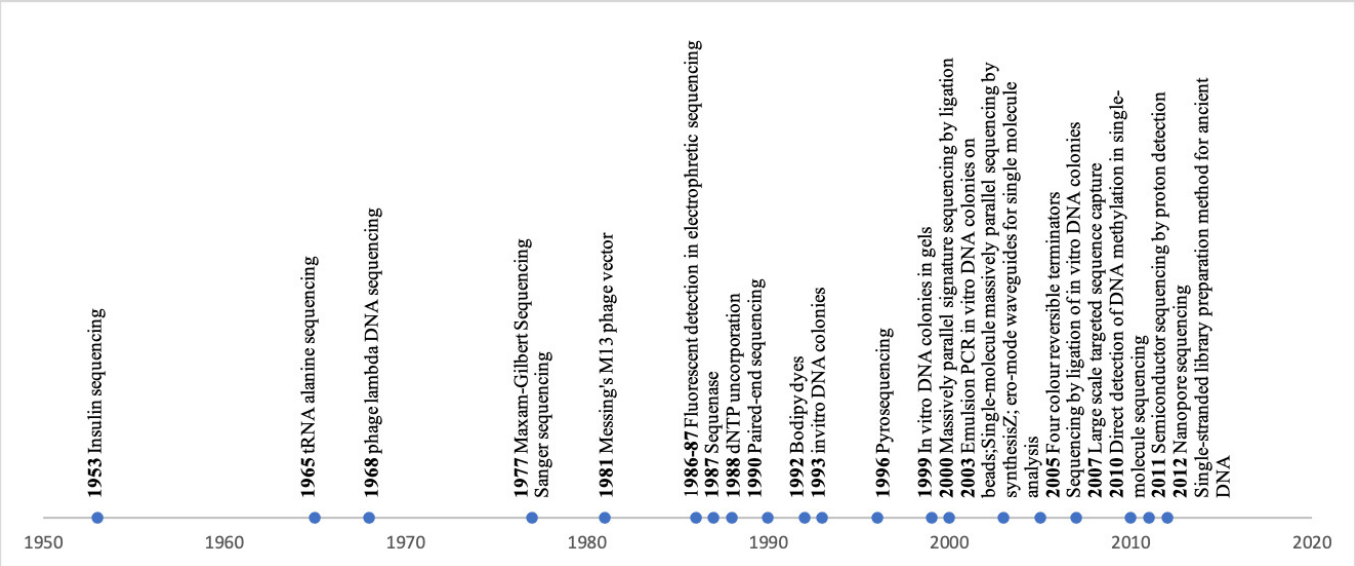


Figure 17: **Main hits of molecular studies that have allowed a fulgurate progression of molecular medicine.**

An important hit for sequencing technologies was the Human Genome Project, which was published in 2001. This huge multicentric project involved 20 research institutes from 6 different countries and required thirteen years to have a first detailed draft of the human genome [93] . The completion of the Human Genome Project revealed the need of greater and advanced technologies to answer complex biological questions that arose; however, limited throughput and the high costs of sequencing remained major barriers.

Some of these advances were reached soon: the development of DNA polymerases that generate longer amplicon sizes (generated by Invitrogen SequalPrep, AccuPrime, TaKaRa PrimeSTAR GXL, LA Taq Hot Start, KAPA Long Range HotStart and QIAGEN LongRange polymerized chain reaction (PCR) I Polymerase, among others). These enzymes could amplify up to 15 Kb or longer genomic DNA. Long-range PCR (LR-PCR) arrived as a flexible, fast, efficient and cost-effective choice for sequencing candidate genomic regions in a small number of samples [94] .

In the mid-2000s, the release of the first truly high-throughput sequencing, the so called "Second or Next Generation Sequencing" platform heralded a 50,000 fold drop in the cost of human genome. Over the past decade, NGS technologies have continued to evolve — increasing capacity by a factor of 100–1,000 — [95] and have incorporated revolutionary innovations to tackle the complexities of genomes. Finally, some years ago, in 2012 arrived the next step: "Third generation sequencing" with the capacity to sequence real-time and single-molecule without breaking the DNA nor amplifying it, as it was done with previous techniques. The first companies with this approach were PacBio, a company that developed a technique based in a polymerase-mediated synthesis in real time, and Oxford Nanopore Technologies (ONT), a company that generated a technology based on the change of ion potential generated by the pass of nucleic acids through nanopores.

1.6.1 Use of sequencing technologies

The sequencing technology of election will depend on the objective and the expected results for each study, in order to adjust the cost-effectiveness of them. Here we are going to focus on the technology used to this project: the diagnosis of antithrombin deficiency.

Despite the huge improvement on sequencing technologies, Sanger sequencing is still the gold standard method of election for the molecular analysis of cases with antithrombin deficiency. This is because the gene is small, and only 7 small PCRs cover all coding exons and flanking regions with a small cost (around 60€). Moreover, this method can detect the genetic defect responsible for the deficiency in up to 70% of cases. Regarding more complex diseases such as hereditary cancer when a panel of genes are needed to study, or when a high depth is needed, NGS is the method of choice. The technical sensitivity and specificity of the assay are paramount as clinicians use results to make important clinical management and treatment decisions [96] . Sanger Sequencing is also the method used to validate results from NGS sequencing. However, there is nowadays a significant debate within the diagnostic community regarding the necessity of Sanger validation, some experts say that Sanger sequencing is useful to confirm and define INDELS detected by NGS, but confirmatory analysis with this technique is unnecessary [97, 98] , others still see the necessity of Sanger validation due to the false positives obtained by NGS techniques [99, 96] .

Nowadays, whole exome sequencing or even whole genome sequencing is increasingly used as a primary tool for diagnosing Mendelian disorders, and probably these methods will be used routinely in a near future to get the genetic ID of each patient in order to get a much more accurate diagnosis. Indeed, even for monogenic diseases as antithrombin deficiency, other genetic factors may modulate the consequences of a mutation in the main coding gene.

1.6.2 Technologies for structural variants detection

Special mention is needed to talk about structural variants (SV) and their detection. Particularly since SV are also be involved in antithrombin deficiency. SV are genetic defects that affect >50 bp, showing a change in copy number or genomic location. SVs include copy number variants (CNV); deletions and duplications, insertions, inversions, translocations, mobile elements, repetitive sequence expansions, and complex combinations (Figure 18).

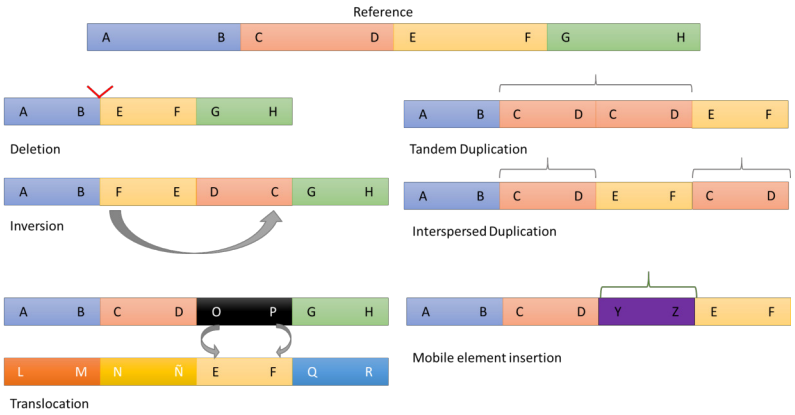


Figure 18: **Types of structural variants.**

SVs contribute ∽3.4 times more nucleotides to human genetic variation than the far more numerous single nucleotide variats (SNV) [100] . Moreover, a SV has more likely possibility to be pathogenic than a SNV [101] .

Indeed, SVs have been associated with many diseases, both germ line and somatic.
The accurate detection of SVs is a challenging task for short read sequencing. Different methods not providers of sequence have been developed all over the years, such as cytogenetics, developed in 1960, detecting Philadelphia chromosome for the first time. The use of advanced fluorescence in situ hybridization (FISH) developed in the 1980s, allows the further identification of chromosomal alterations that are unresolved by the karyotyping methods [102, 103] .
In 2001, different approach appeared: genome-wide array-based comparative genomic hybridization (CGH array) analysis of genomic DNA [104] . Its implementation to clinics together with SNP array, enable Genome-wide detection of copy number changes in a high resolution, and therefore microarray has been recognized as the first-tier test for patients with intellectual disability or multiple congenital anomalies. These two techniques are based on comparative hybridization of different probes around the genome comparing controls and patient's samples. They have a practical resolution of 50-2000 Kb allowing the detection of CNV [105] . Finally, gains or loss of genetic material targeted to specific locus can be assessed by MLPA. This method utilizes probes, each specific for a different DNA sequence (mainly exons of a specific gene of interest), to evaluate the relative copy number of each DNA sequence compared to controls [106] .
However, despite improvements in genomic technologies that we have reviewed, characterization of SVs remains challenging and the full spectrum of SVs is not achieved by routine methods used in the clinic such as FISH, karyotype, CGH array or other targeted sequencing approaches. Third generation sequencing, recently developed seems to be the most suitable and complete method to fully characterize SV, at a nucleotide resolution. Long reads, which can span repetitive or other problematic regions, allow the identification and characterization of SVs [107, 108, 109].

1.6.3 Third Generation Sequencing

As we have seen, the recent arrival of Third Generation Sequencing has been revolutionary for genome sequencing. In this thesis we have applied nanopore sequencing (ONT, Nanopore Sequencing Technologies) to analyze some particular cases with antithrombin deficiency.
Nanopore sequencing is based on the change of the current flow triggered by the DNA molecule when it passes through the nanopore. This pattern or magnitude is observed and characterized and informatically translated to "bases" or basecalled (Figure 19).

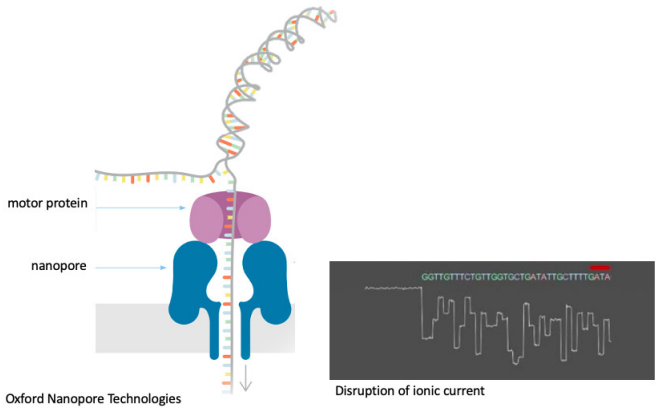


Figure 19: **Representation of the disruption of ionic current caused by the pass of the DNA strand through the nanopore.**

The principal advantages over the existing technologies are: that is a real-time and single-molecule sequencing, it generates very long-reads (the length is only dependent on the fragment of the DNA sample), it doesn't require DNA manipulation and it can be done in handle devices. The current MinION device only weighs 90 g, measuring only 3.3 × 10.5 × 2.3 cm in size, fitting in the palm of one's hand (Figure 20). The cost of the device is also much lower compared to other massively parallel sequencers. The initial cost being only around $1,000, including the device and initial set of reagents. Further nanopore sequencing devices have been developed, being the last one the PromethION, where a single flow cell can yield 50-100 Gb (typical yield of MinION system is 5-10 Gb) and 24 flow cells can be run in parallel. The cost of PromethION is $195,455 with the initial reagents [110, 111] (Figure 20) .

Figure 20: **Different devices for nanopore sequencing developed by Oxford Nanopore Technologies indicating estimated costs and uses.**

Derived from the very long reads (up to 2.3 Mb) of DNA without manipulation, several applications can be achieved for nanopore sequencing:

1. This method enable *de novo* genome assembly without preparation of complex mate-pair libraries typically required with short read sequencing.
2. Long reads are also useful to determine sequences of genomic region containing long repetitive sequences, which is difficult with short read sequencing.
3. Long reads also allow the study of SVs within the genome [109] . The nucleotide-level resolution achieved also facilitates the design of primers for PCR validation.
4. The detection of the ionic current changes shifted by the nucleotides passing through the nanopore is not limited to the canonical four nucleobases of adenine, guanine, cytosine, and thymine.
5. Taking advantage of the nature of unamplified direct sequencing, nanopore sequencing can directly observe base modifications such as methylation [112, 113] , and even direct sequencing of RNA molecules containing uracil bases [107] .

Consequently, many applications have been developed for this technology such as: targeted sequencing (with PCR

products or CRISPR-Cas9), direct RNA sequencing, cDNA sequencing, metagenomics... Besides, further studies can be achieved with informatic analysis of these sequences: epigenetics, phasing variants even for DNA, or RNA reads.

The largest limitation of the nanopore sequencer is the comparatively lower read accuracy when compared to short read sequencers, especially for INDELS. Because insertions and deletions are included in the errors, nanopore reads *per se* are not optimal for SNV detection. However, it has been demonstrated that, with high coverage reads SNV genotyping with nanopore sequenced reads [114, 115] .

Objectives

2.1 Hypothesis

Although antithrombin deficiency is the most severe thrombophilia and it has been widely studied during the last 55 years, there is a significant clinical heterogeneity that must be explored. Moreover, the real incidence and characteristics of structural variants causing this disorder is largely unknown. Finally, up to 25% of patients with antithrombin deficiency still don't have molecular diagnosis. Novel molecular methodology may complement the information obtained by current technology used in the routine diagnosis particularly to identify and characterize structural variants. We think that the analysis of a large cohort of patients with antithrombin deficiency may identify elements involved in the severity of the disease. Moreover, the selection of patients with particular features could help to identify new mechanisms and characterize structural variants involved in this disorder.

2.2 Objectives

The main objective of this project is to gain new information concerning the most severe thrombophilia, congenital antithrombin deficiency, by the analysis of a large cohort of patients with this disorder and using novel molecular methods. We aim to assess the different factors determining the clinical impact of antithrombin deficiency and to explore new mechanisms and variants involved in antithrombin deficiency.

2.2.1 Specific objectives

For those aims, we have these specific objectives:

1. Clinical impact of antithrombin deficiency.
    - To evaluate the severe impact and features in patients with antithrombin deficiency and pediatric thrombosis.
    - To study a variant of low severity and wide spreading, p.Thr137Ala, responsible for antithrombin deficiency.
2. New molecular defects involved in antithrombin deficiency.
    - To analyse of structural variants in antithrombin deficiency.
    - To implement new technologies to discover new variants involved in antithrombin deficiency.

Methods

3.1 Patients

In this study, the following cohorts of patients and controls have been included:

1. A cohort of 398 index cases with antithrombin deficiency consecutively recruited during a period of 26 years (from 1994 to 2020) in a reference center for this disease in Spain. Additional studies were done in 488 relatives with antithrombin deficiency.
2. A similar cohort of 278 index cases with antithrombin deficiency consecutively recruited by a reference center from Belgium during 30 years (from 1990 to 2020). Family studies were done in 156 relatives, carriers of antithrombin deficiency. In this cohort, a specific study was conducted in 12 patients carrying a specific *SERPINC1* defect, the rs2227606 polymorphism. For this study, a control population consisting of 100 healthy subjects of African ancestry was randomly collected in Belgium.
3. Additional 9 patients with antithrombin deficiency caused by SVs, detected by MLPA, from Finland and French origin were included in this study.

For these patients, the disease history was evaluated to find possible risk factors for thrombosis, such as oral contraceptives, pregnancy, complicated delivery, bad pregnancy outcome, obesity, immobilization, infection, surgery and trauma. Information about antithrombotic therapy and family history of venous and arterial thromboembolism was also collected. Besides anti-FXa activity of antithrombin, laboratory data from additional thrombophilic parameters (protein C activity, free protein S antigen, resistance to activated protein C, Factor V Leiden and prothrombin G20210A mutation) were collected when available. Thrombotic events were objectively diagnosed and confirmed by experienced radiologists through established imaging procedures such as Doppler-ultrasonography, computed or magnetic resonance tomography for venous thrombosis and spiral computed pulmonary angiography or lung perfusion scintigraphy for pulmonary embolism.
All patients gave their informed consent to enter in the study which was performed in accordance with the Declaration of Helsinki and approved by the Ethics Committees of the Hospital Universitario Reina Sofia (8/2013) and the "Universitair Ziekenhuis Brussel".

3.2 Blood sampling and DNA purification

Blood samples were collected long after the acute event and in absence of any anticoagulant treatment. Blood were obtained by venopuncture collection into 1:10 volume of trisodium citrate. Platelet poor plasma fractions were obtained (within 5 min after blood collection) by centrifugation at 4ºC for 20 min at 2200xg and stored at -70ºC. Genomic DNA was purified by the salting out procedure and stored at -20ºC.

3.3 Antithrombin levels

Antithrombin anti-FXa activity was determined in citrated plasma by chromogenic methods following the manufacturer's instructions (HemosIL Antithrombin, Werfen, Barcelona, Spain and Innovance Antithrombin, Siemens, Marburg, Germany) [74] and Biophen AT anti-FIIa (Hyphen Biomed, Neuville-sur-Oise, France). Innovance assay contains human FXa while HemosIL Liquid AT contains bovine FXa. Biophen AT anti-FIIa uses bovine FIIa [116]. In specific studies (Chapter 4.1.2: "Antithrombin p.Thr147Ala: The First Founder Mutation in People of African origin Responsible for Inherited Antithrombin Deficiency"), antithrombin activity was measured with varying incubation times or after heat incubation of plasma, using the anti-FXa-based assay from HemosIL Antithrombin

(Werfen). For all chromogenic assays, after adding the chromogenic substrate, the absorbance was measured at 405 nm with a plate reader (Synergy HTX,Biotek; Multiskan FC instrument; Thermo Fisher Scientific, Waltham, United States). Reactions were performed in triplicate.
Antigen levels were measured by rocket immunoelectrophoresis and/or ELISA. For both parameters, activity and antigen levels, values were expressed as a percentage of the result observed in a pool of citrated plasma from 100 healthy subjects. Heparin affinity of plasma antithrombin variants was evaluated by crossed immunoelectrophoresis (CIE) in the presence of heparin, as described elsewhere [54] , but we used two different ionic strength conditions: 150mM (physiological) and 0.8M NaCl.
Polyacrylamide gel electrophoresis (PAGE) was performed using both native (with and without 5M urea) and denaturing conditions, as previously reported [55] . Proteins were transferred to PVDF membranes and antithrombin was detected by Western Blot using a polyclonal anti-human antithrombin antibody (A9522, Sigma-Adrich). The conditions for Western blot analysis have been described elsewhere [117] .

3.4 Genetic analysis of SERPINC1

3.4.1 Sanger sequencing

PCR covering the 7 exons and flanking regions of *SERPINC1* was done in all patients using LongAmp® Taq DNA Polymerase (New England Biolabs, US). Primers used are described in table 3 . PCR products were purified and sequenced with ABI Prism Big Dye Terminator v3.1 Cycle sequencing kit and resolved on a 3130xl Genetic Analyzer (Applied Biosystems, Spain). Comparison with the reference sequence (GenBank accession number NG_012462.1) was performed with SeqScape v2.5 software (Applied Biosystems, Spain).

Table 3: ***SERPINC1* primers covering the 7 exons and flanking regions, and the primers designed to specifically amplify the duplication of exon 6 in tandem.**

Method	Exon	Primers 5'-3'	Localization	PCR size
PCR-sequencing	1	**CTCTGGAACCTCTGCGAGA**	NG-012462.1:5017-5035	197
		GAAAGCTCACCCCTCTTAC	NG-012462.1:5193-5213	
	2	**TGCAGCCTAGCTTAACTTGGCA**	NG-012462.1:7422-7443	500
		GGTTGAGGAATCATTGGACTTG	NG-012462.1:7890-7911	
	3	**TGTGCTCACCACCCATGTTA**	NG-012462.1:10297-10316	320
		ATGCTGTTTCTCCACCTCCT	NG-012462.1:10637-10656	
	4	**AAGCCAATTGAATAGCACAGG**	NG-012462.1:11438-11458	210
		AAGGGGGTAAGCTGAAGAG	NG-012462.1:11669-11687	
	5	**TGTGTTCTTACTTTGTGATTCTCT**	NG-012462.1:12408-12431	402
		AAGGGAGGAAACTCCTTCCTAG	NG-012462.1:12834-12855	
	6(6F/6B)	**TTCTCCCATCTCACAAAGAC**	NG-012462.1:14825-14844	153
		CATGCATCTCCTTTCTGTACC	NG-012462.1:14957-14977	
	New 6 (6F/6B2)	**TTCTCCCATCTCACAAAGAC**	NG-012462.1:14825-14844	232
		CCACAGGCCTGCTATAATACAG	NG-012462.1:15035-15056	
	7	**CTGTGGATGATTTACCTGCC**	NG-012462.1:18244-18260	351
		GCCCCAATAGCATGTTTCCCC	NG-012462.1:18612-18632	

Specific amplifications A specific set of primers were designed for the detection of potentially all tandem duplication of exon 6 (Table 3).

Moreover, for validation of different SVs and retrotransposon insertion, specific primers covering the breakpoint were designed (Table 4).

Table 4: **Specific set of primers used to validate structural variants detected by nanopore in Chapter 4.2.4: "Dissection of structural variants involved in antithrombin deficiency".** IDs are named in Chapter 4.2.4 * Specific sequence present only in inserted SVA sequence of patients P37-39. *** Size of the WT too large for possible amplification.

ID validation	Orientation	Primers 5'->3'	Genomic localization	Size MUT (bp)	Size WT (bp)
P37-P39	F	CCAGGGATCAAACTCAGAGGAAAGTGG	1:173905956-173905983	420	***
	R	GGCACCATTGAGCACTGAGTG	*		
P37-P39	F	CTCAGGGTTAGATGGATTAAGG	*	2,500	***
	R	GGTTACTGGGTAAAGAACAACTTGAAGG	1:173903702-173903730		
P6	F	CCCACACAACTCTCTCTTGCTCAGCC	1:173926009-173926035	287	46,505
	R	GAGGGCTCCGCAGTCTAATGTTTCAG	1:173879530-173879556		
P3	F	CCAATCAAATACCTGACAGTTTCTGACC	1:174816582-174816611	493	***
	R	ATGAGGCTCACATATACAGACCTCACC	1:173847777-173847804		
P24	F	AGTTTTGTTAGGCGAATCACAGCAGGG	1:173908135-173908163	535	195,262
	R	TCCATCCAGAAGGAACATCTTTCTTTCC	1:174103370-174103397		
P20	F	CGCAGGAACCATAAACCAGCTGGAGG	1:173908135-173908163	838	8,124
	R	TCCATCCAGAAGGAACATCTTTCTTTCC	1:173916233-173916259		
P2	F	CTCATTTAAGTTTTACTGTCCAGTTTAGG	1:174849313-174849342	6200	***
	R	ATGACATGCTTCATCCAACAGAAAGACC	1:172981159-172981187		
P25	F	TTCCTAGTGCTCATCAGCGGTTTTAGG	1:173935758-173935786	2,157	21,265
	R	AGGTGGCTGGGCAGAAGACCTTTGG	1:173914521-173914546		
P7	F	CCTGCTTCTGAAAGTGCTGTAGTTACAG	1:173950319-173950347	575	99,837
	R	GTTGATGCTGAACAATGCCTTACTCTCC	1:173850510-173850538		

In silico analysis of splicing *In silico* prediction of potential effects of a gene variation on the splicing was evaluated with the Human Splicing Finder (HSF, version 3.1) software[118] . *In silico* prediction of protein translated was performed with Translate tool in ExPASY Bioinformatic Resource Portal (Swiss Institute of Bioinformatics).

3.4.2 NGS PGM

Next-generation sequencing of the whole *SERPINC1* gene (13,810 bp) was done using Ion Torrent technology small amplicons covering up to 95% of the whole *SERPINC1* gene were sequenced in a Personal Genome Machine (PGM) equipment essentially as described previously [119] .

3.4.3 LR-PCR and NGS ILLUMINA

The whole *SERPINC1* gene was amplified by four overlapping LR-PCR, using primers and conditions indicated in Table 5 and represented in Figure 21 . PCR amplicons were visualized in agarose gels and were equimolarly mixed in a single tube in a total amount of 250 ng.

Subsequently, the libraries were fragmented and the adapter and barcodes were ligated using the NGSgo protocol (GenDX, Utrecht, Netherlands) following the manufacturer's recommendations. Resulting libraries were combined and sequenced on a MiSeq platform (Illumina, San Diego, CA). After sequencing, barcoded sequences were de-multiplexed and analyzed individually. The paired sequence files (fastq format) were used as input for analysis with the CLC Genomic Workbench v.11 software (Qiagen, Aarhus, Denmark) and analyzed using InDels and Structural Variants tool.

Table 5: **Oligonucleotides, conditions and length of for LR-PCR productsamplification of *SERPINC1*.** The primers used to verify the deletion of intron 1 are also shown. AT: Annealing Temperature. E. T: Extension Time. Sec: Seconds. Min: minutes. Bp: Base pairs.

	Primer Forward	Primer reverse	A.T. (ºC)	E.T.	Cycles	PCR length (bp)	SERPINC1
Standard amplification of SERPINC1	CTCTGGAACCTCTGCGAGA	GAAAGCTCACCCCTCTTAC	54	30 sec	30	197	Exon 1
	TGCAGCCTAGCTTAACTTGGCA	GGTTGAGGAATCATTGGACTTG	54			500	Exon 2
	TGTGCTCACCACCCATGTTA	ATGCTGTTTCTCCACCTCCT	54			320	Exon 3
	AAGCCAATTGAATAGCACAGG	AAGGGGGTAAGCTGAAGAG	54			210	Exon 4
	TGTGTTCTTACTTTGTGATTCTCT	AAGGGAGGAAACTCCTTCCTAG	54			402	Exon 5
	TTCTCCCATCTCACAAAGAC	CCACAGGCCTGCTATAATACAG	59			232	Exon 6
	CTGTGGATGATTTACCTGCC	GCCCCAATAGCATGTTTCCCC	54			351	Exon 7
Long-range PCR	GAAACATGGAAATAACAGGGAGGGAGG	TGACTTCCTGCACAGCATCTTTACACC	60	18 min +20 sec	30	5269	Promoter-Exon 1
	CAAGGAGTCCTTGATCACACAGCAGG	AACACCCTTTGGTATATATTTGAAGTTGG				6658	Exon 1-exon 3
	CTGTCCCAGGTACTGTGCTTGAAGG	CATGCCTTAACACTGGAAACAGGCC				4795	Exon 3-exon 6
	GATATGTCTTGTGTCAATAACTATCCTCC	GGTTACTGGGTAAAGAACAACTTGAAGG				6308	Exon 5-3'UTR
Verification of intron 1 deletion	CATTTCAAGTGCTCTCCCTCC	CAAGTCCAATGATTCCTCAACC	60	18 sec	30	630	intron 1-exon 2

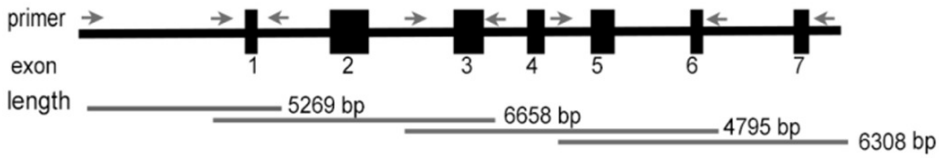


Figure 21: **Representation of 4 LR-PCR amplicons covering *SERPINC1*.** Primers localizations are represented as arrows. Sizes are also shown.

3.4.4 Multiplex Ligation-dependent Probe Amplification

Multiplex Ligation-dependent Probe Amplification technique, a multiplex PCR method able to identify gross gene defects affecting *SERPINC1*, was done using the SALSA® MLPA® Kit P227 SerpinC1 (MRC-Holland) following the manufacturer's procedure in a 3130xl Genetic Analyzer (Applied Biosystems, Spain). The analysis of the data obtained was done using Coffalyser.net software.

3.4.5 CGH array

Study of Copy Number Variations in some selected cases with SVs was done with a high-density CytoScan® HD Array (Thermo Fisher Scientific, In), which includes 2.67 million markers for CNV analysis, 750,000 SNP probes and 1.9 million non-polymorphic probes for comprehensive whole-genome coverage. Following hybridization, the laser scanner (GeneChip® 3000 Scanner) was used for scanning the arrays, and the images were extracted and analyzed using Affymetrix Gene Chip Command Console software (version 4.0) and Chromosome analysis software (ChAS v.1.2/na33.2) (Fisher Scientific, Inc.). Results were interpreted with the aid of the UCSC genome browser (http://genome.ucsc.edu/; Human Feb. 2009 GRCh37/hg19 assembly).

3.4.6 Nanopore sequencing

The characterization of SVs in chapters 4.2.3 "Long-read sequencing resolves structural variants in *SERPINC1* causing antithrombin deficiency" and 4.2.4: "Dissection of structural variants involved in Antithrombin deficiency"

was done in selected cases by long-read whole genome sequencing (WGS) using the MinION (n= 5) and PromethION (n= 19) platforms (Oxford Nanopore Technologies).

Whole genome sequencing

a. Library preparation and sequencing

For both devices in genomic approach, samples were prepared using the 1D ligation library prep kit (SQK-LSK109), and genomic libraries were sequenced on R9 flow cell. Read sequences were extracted from base-called FAST5 files by Guppy (versions 3.0.4 to 3.2.8) to generate FASTQ files.
The informatic analysis was different in cases done in PromethION due to the higher amount of data obtained. Moreover, some cases sequenced with PromethION had unknown molecular base and required from different bioinformatic approach.

b. SV discovery in PromethION platform sequencing

When using PromethION platform, all the samples were analyzed by a specific workflow developed in-house to discover SV, which was inspired on rules from nano-snakemake [120, 121] . This workflow, which is summarized in Figure 22 , is publicly available at https://github.com/who-blackbird/magpie. Reads were aligned against the

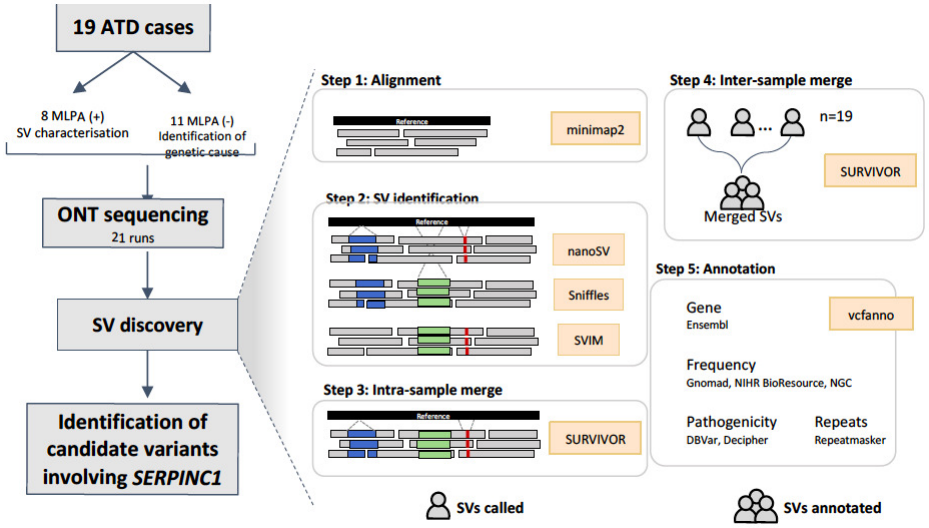


Figure 22: **Whole genome nanopore sequencing workflow using promethION platform.** An overview of the general stages of the SVs discovery workflow are shown and algorithms used are depicted in yellow boxes.

GRCh38/hg38 human reference genome using minimap2 [122] (2.17-r941) with default parameters for nanopore data ('-ax map-ont' parameter). SV discovery was done using a combination of three different algorithms:

- Sniffles v1.0.11 was executed with a supporting read evidence of 4 ('-s 4' parameter) due to coverage variability.
- NanoSV v1.2.4 was executed with default parameters. It was run on each independent chromosome in parallel to optimize computation time, with the limitation that inter-chromosomal variants were not detected.
- SVIM v1.2.0 was executed with default parameters. Resulting SV calls with a quality score of less than 10 were filtered out of the dataset and not used for analysis

SV calls were merged at two different levels: intra-sample merge, to merge calls within individuals that had been identified by all three algorithms, and inter-sample merge, to merge SV calls across individuals.
Intra-sample merge. For each of the 19 samples, SV calls from all three different algorithms were merged using SURVIVOR v.1.0.7. VCF files were concatenated into one using bcftools [123] , and SURVIVOR was run using the command 'SURVIVOR merge in.fofn 500 1 -1 -1 -1 -1 out.vcf', requiring a maximum distance of 500 bp between breakpoints. Additionally, intra-sample merge was done independently of the SV type, since different SV detection algorithms determine the type in different ways. For example, Sniffles determines canonical (DEL, DUP, INS, INV, TRA) and some complex SV types, while NanoSV calls only breakends (BND). The following options were turned-off: take the strands of SVs into account, estimate distance based on the size of SV, minimum size of SVs to be taken into account. After running SURVIVOR, an 'in-house' script was used to select the most common SV type. If there was no common type, the order of selection was NanoSV (if the SV type was not a BND) > Sniffles > SVIM type.
Inter-sample merge. Then, all the 19 samples were merged using SURVIVOR, taking the SV type into account. The command run was: 'SURVIVOR merge in.fofn 500 1 1 -1 -1 -1 out.vcf'.

c. SV in MinION platform sequencing

MinION device was used for SV validation in specific cases. It is important to remark that the analysis of the data generated by MinION was facilitated by the data obtained previously by CGH array and MLPA in the same patients, thus avoiding the limitation of the small number of reads covering the region of interest. In these cases, reads were aligned against the GRCh38/hg38 human reference genome using minimap2 [122] (2.17-r941) with default parameters for nanopore data ('-ax map-ont' parameter) and variant calling was performed with Sniffles, with a supporting read evidence of 1 to 5 due to coverage variability. Alignments were visualized and manually inspected with Integrative Genomics Viewer (IGV) [124] .

Targeted sequencing

d. Enrichment approach with MinION device

We explored two approaches to enrich sequencing to specific regions of the genome with the MinION device.

1. ***In silico enrichment.*** This method consists on an informatic enrichment of reads from a specific region of interest based on an informatic tool available in MinKNOW 1.4.3 software. This software, by "Read Until API" bioinformatic tool, analyse the reads data in real-time. Reads can be selected if they fit to a region of interest previously provided by the user. For this project, we selected the smallest size possible for this analysis (3MB), and obviously, the region selected (chr1:172299725-175426186) contained *SERPINC1* (Figure 23). Reads not fitting to the region are returned and the server can unblock the read in progress. Thus, the sequencing is optimized to reads that fit the region/s in purpose allowing them to continue passing through the sequencer. This procedure was tested in 3 cases with antithrombin deficiency caused by SV in *SERPINC1* detected by MLPA and CGH array, in order to facilitate the detection of supplementary alignments covering the SV.

2. ***One site directed amplification***. As we are preparing a patent on this method, we can't show it in detail. Briefly, our method is able to generate long-amplifications using a set of specific primers of *SERPINC1* without a second primer restricting the amplification size. These long-PCRs are sequenced using the MinION device. This design might cover any partial SV involving this gene. This method was tested in 3 cases with antithrombin deficiency caused by SVs, 2 deletions and 1 duplication.

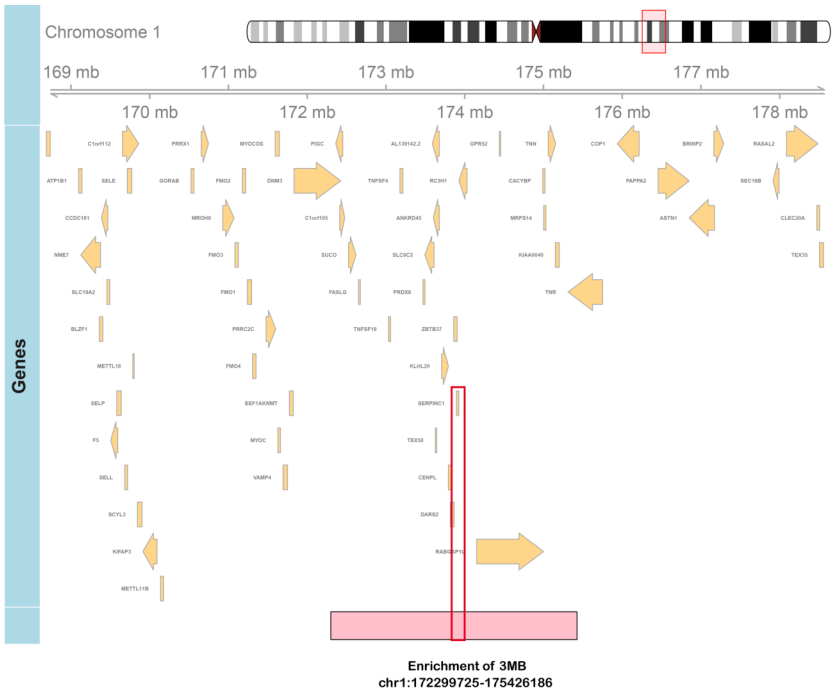


Figure 23: **Representation of the genomic region of 3Mb enriched by in silico tools.** Region of interest (chr1:172299725-175426186) is marked with a red box and *SERPINC1* localization within the region of interest is highlighted in red.

e. De novo Assembly

De novo assembly was performed to characterize the SVA insertion in one patient (P9 in Chapter 4.2.3: "Long-read sequencing resolves structural variants in *SERPINC1* causing antithrombin deficiency" and P37 in Chapter 4.2.4 "Dissection of structural variants involved in Antithrombin deficiency"). Reads within the region [GRCh38/hg38] Chr1:173,840,000-174,820,000 were extracted from the alignment of this individual and converted to a FASTQ file using Samtools [123] . *De novo* assembly was performed with wtdbg2 v2.5, using the parameters '-x ont -g 980k -X 10 -e 3' . The *de novo* contig was then aligned to the reference genome using minimap2 with default parameters for nanopore reads. The genomic sequence containing the SVA repetitive element was then extracted from the alignment and analyzed with RepeatMasker (http://www.repeatmasker.org) to characterize the type of SVA and its subelements.

f. Characterization of the breakpoints of Structural variants

For reads obtained by nanopore sequencing, and given the current relatively high error rate of long-read sequencing, specific PCR amplification on the new formed junctions followed by Sanger or NGS sequencing were done. The presence of microhomology, insertions and deletions at the breakpoints were manually characterized. The percentage of repetitive sequences was also calculated for each junction ($\pm$100 bp) by intersecting these regions with the human genomic repeat library (hg38) from RepeatMasker version open-4.0.5 (http://www.repeatmasker.org) using bedtools [125] .

g. Haplotype analysis

To investigate a possible founder effect for the p.Thr147Ala variant in Chapter 4.1.2 "Antithrombin p.Thr147Ala: The First Founder Mutation in People of African origin Responsible for Inherited Antithrombin Deficiency", a novel approach based on nanopore sequencing was used. Family studies to determine the potential haplotype associated to the mutation causing antithrombin deficiency were not possible for carriers of the p.Thr147Ala variant. So, we explored the usefulness of nanopore sequencing of LR-PCR to solve large haplotypes [126] . For this study, a combination of LR-PCR amplifications and nanopore sequencing was done.

- **Long-Range amplification** . The whole *SERPINC1* gene (14480 bp) was amplified by LR-PCR using two amplicons covering exon 3, containing the p.Thr147Ala mutation (Figure 24). Amplifications were done

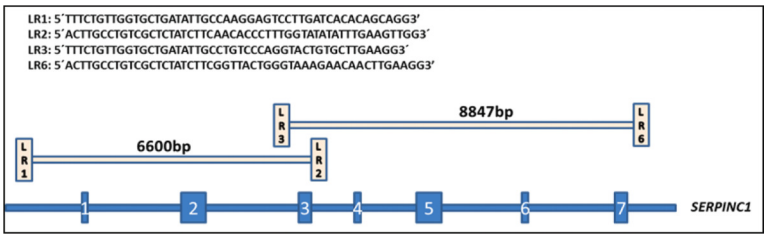


Figure 24: **Schematic representation of *SERPINC1* showing exons, introns as well as primers used for whole amplification of the gene by two amplicons.** The primers used for these LR-PCR are also shown.

with the SequalPrep Long PCR Kit (Applied BiosystemsA10498), using specific primers tailed with Oxford Nanopore universal sequence for barcoding. PCR reactions were done in a Veriti 96-well thermal cycler (Applied Biosystems) equipment following the following conditions: 94°C x 2minutes, 10 cycles of 94°C/10 seconds, 60°C/30 seconds, 68°C/1 minute/Kb; and then 20 cycles of 94°C/10 seconds, 60°C/30 seconds, 68°C/1minute/Kb plus 20 seconds per cycle; 68°C, 5minutes. PCR products were evaluated on 0.8% agarose gel electrophoresis, purified with Agencourt AMpure XP beads (Beckman Coulter A63880) and quantified with Nanodrop 2000 (Thermo Scientific).

- **Barcoding and library preparation.** Barcoding PCR was performed with 200 fmol of initial PCR product and standard conditions with PCR Barcoding expansion 1–12 kit (Oxford Nanopore EXP-PBC001) and Long Amp Taq 2x MasterMix (NEB M0287). After barcoding, PCRs were pooled taking 10 μL of each barcode and purified with Agencourt AMpure XP beads, having a final elution volume of 50 μL. Library and nanopore sequencing were performed as previously described.
- **Bioinformatic analysis.** The raw data (fastq) obtained by MinKnow software was aligned with minimap tool. Common SNVs shared by carriers of the mutation were evaluated by Illumina Variant Studio 3.0 and Integrative Genomics Viewer, IGV 2.3.72. Additionally, all reads containing the mutated allele were selected using the "grep" command in Linux to determine haplotype markers.

3.4.7 STR analysis

The analysis of a possible common origin of recurrent SVs found in Chapter 4.2.4: "Dissection of structural variants involved in Antithrombin deficiency", was done by using short tandem repeats (STR) analysis with 8 markers covering 12 Mb of chromosome 1, $\backsim$6 Mb up and downstream *SERPINC1*. A representation of STR markers evaluated in this study is shown in Figure 25 .

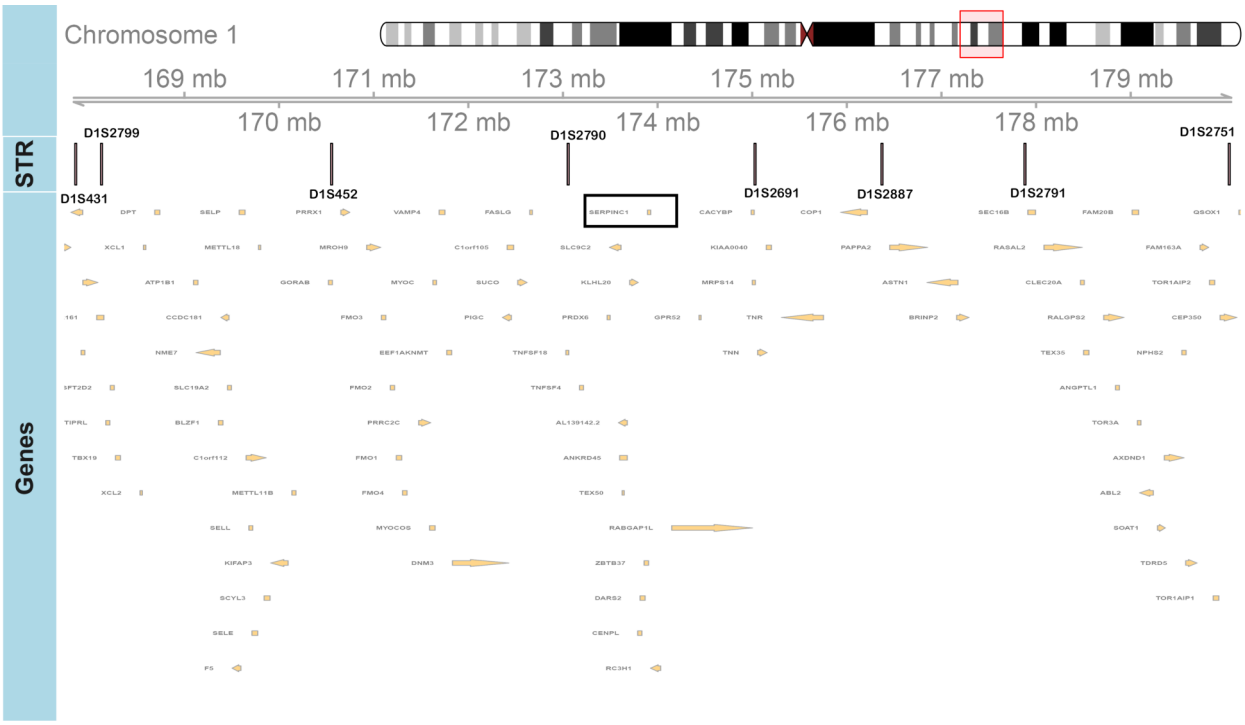


Figure 25: **Localization of STR flanking *SERPINC1* that have been evaluated in this study.** *SERPINC1* gene localization is highlighted with a squared box.

3.5 Specific analysis of p.Thr147Ala variant

3.5.1 Recombinant Expression

Wild-type (WT) and mutant p.Thr147Ala antithrombins were constructed on the β-glycoform (p.Ser169Ala) background of antithrombin to reduce glycosylation heterogeneity. Site-directed mutagenesis was performed using the QuikChange XL Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, United States) and the appropriate primers to obtain the p.Thr147Ala mutant on the pCEP4/AT-S169A plasmid, kindly provided by Dr. J. Huntington (Cambridge Institute for Medical Research, Cambridge, United Kingdom).
Both WT and mutant plasmids were transfected into HEK-293T cells using jet Prime (Polyplus Transfection, New York, United States) as previously described [127] . Stable cell lines were maintained in Dulbecco's modified Eagle's medium containing 10% fetal-bovine serum (Invitrogen, Carlsbad, United States). Cells were grown to 70 to 80 % confluence in cell factory systems before replacing the growth medium by Freestyle serum-free medium (Invitrogen). Cells were further grown at 37°C for 3 days followed by harvesting of the culture medium. The conditioned media were concentrated and dialyzed against the appropriate binding buffer (50mM Tris) with a 4MA ultrafiltration hollow-fiber cartridge (GE Healthcare, Waukesha, United States). Next, the dialyzed medium was loaded onto a HiTrap Heparin column (GE Healthcare) coupled to an ÄKTA purifier (GE Healthcare). Fractions were eluted in a 50mM Tris-HCl (pH7.4) buffer by applying a NaCl gradient from 0 to 3M. Purity of the eluted fractions was assessed by reducing SDS-PAGE and Coomassie blue staining. Finally, positive fractions were pooled, dialyzed against PBS buffer, and stored at -70°C. We measured anti-Xa activity of both purified recombinant WT and mutant p.Thr147Ala with the Innovance Antithrombin assay (Siemens) and HemosIL Antithrombin (Werfen). The activity of the variant protein was normalized to the activity of the WT protein and expressed as percentages.

3.5.2 Functional studies

We evaluated formation of thrombin-antithrombin (TAT) complexes, thermal stability and intrinsic fluorescence, and performed structural modeling.
Plasma from the carriers of the p.Thr147Ala variant and controls was incubated with 0.2 U of thrombin (Calbiochem, Merck KGaA) for 15 minutes at room temperature in the presence of 5 U of unfractionated heparin. These reactions result in the formation of TAT complexes that were detected by SDS-PAGE under reducing conditions followed by immunoblotting. These experiments were also performed on the recombinant proteins.

3.5.3 Thermal Stability

Plasma of patients and controls was incubated at 40°C for 24 hours. Anti-Xa activity was measured before and after incubation using the HemosIL Antithrombin kit (Werfen) and the Innovance Antithrombin kit (Siemens) as described previously.

3.5.4 Structural Modeling

Structural modeling was performed using UCSF Chimera software (Resource for Biocomputing, Visualization, and Informatics, San Francisco, United States) using the antithrombin-FXa-pentasaccharide complex (PDB accession number 2 gd4) as a template.

3.5.5 Statistical analysis

Statistical analyses were performed in Chapter 4.1.1. "Incidence and features of thrombosis in children with inherited antithrombin deficiency". Continuous variables were expressed as means and standard deviations and categorical data as counts and percentages. Relative risks and 95% confidence intervals (CI) were calculated using previously published formulas [128] . The significance of differences in continuous variables was tested by Mann-Withney test. Kaplan-Meier survival curves were used to illustrate the difference in thrombosis-free survival among different groups. $P<0.05$ was considered statistically significant. Statistical and graphical analysis were performed with GraphPad Prism version 7.03 (GraphPad Software, San Diego CA, USA) and SPSS, version 21 (Chicago, IL, USA).
Data presented in Chapter 4.1.2. "Antithrombin p.Thr147Ala: The First Founder Mutation in People of African origin Responsible for Inherited Antithrombin Deficiency", are presented as mean $\pm$standard deviation unless otherwise specified. Between-group comparisons were conducted using repeat measure one-way analysis of variance. A p-value of 0.05 or less was considered to indicate statistical significance. Statistical and graphical analysis was performed with GraphPad software and Analyze-it software version 2.26 (Analyze-it Software Ltd, Leeds, United Kingdom).

Results

4.1 Clinical impact of Antithrombin Deficiency

4.1.1 Incidence and features of pediatric thrombosis in patients with Antithrombin deficiency

a. Introduction

Thrombosis is an age-dependent disorder. Thus, thrombosis in children is exceedingly rare (0.07-0.14/10,000/year). Pediatric thrombosis is recognized as an important complication of severe medical conditions such as sepsis, cancer, congenital heart disease, and the use of pharmaceutical drugs such as asparaginase and estrogen-containing contraceptives. Surgery and invasive procedures, particularly placement of central venous catheters, are thrombogenic conditions involving around 50% of pediatric VTE, a number that rises to more than 90% in neonates [129]. According to data obtained from case series, case-control studies, registries and cohort studies, thrombophilia is a known risk factor for pediatric thrombosis. A meta-analysis of these studies and a recent nation-wide survey showed that children with first-onset VTE were more likely to have severe inherited thrombophilia, like deficiencies of natural anticoagulants (antithrombin, protein C and protein S), than controls [71], [130]. Accordingly, an excellent marker of the impact of a prothrombotic defect is to determine the prevalence of pediatric thrombosis among carriers of a particular thrombophilia. In order to study the incidence and clinical characteristics of pediatric thrombosis in patients with inherited antithrombin deficiency, we performed a retrospective multicentric study of one of the world-wide largest cohort of patients with this congenital thrombophilia.

b. Patients

Two reference centers for antithrombin deficiency in Spain and Belgium, recruited 441 index patients. Initial diagnosis was always confirmed by measurements of antithrombin activity (anti-FXa activity $<$80%) and genetic analysis. In 206 of them, family studies were performed and 527 first and second degree affected relatives were identified and enrolled in the study, generating a final cohort of 968 patients with antithrombin deficiency. Figure 26

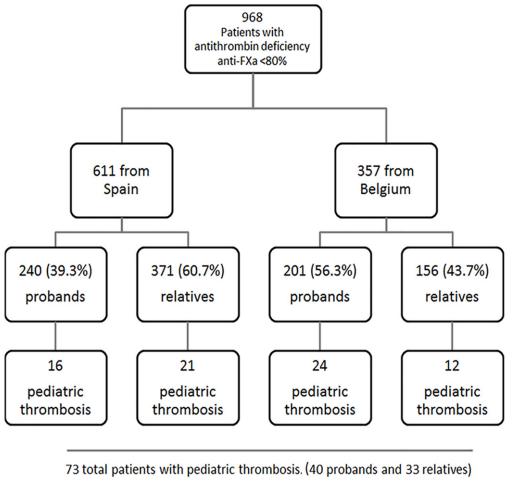


Figure 26: **Flow chart of children with congenital antithrombin deficiency who developed pediatric thrombosis, selected from the entire population.** 73 pediatric patients: 40 probands and 33 relatives.

The patient's history was evaluated in search for thrombotic events and possible provoking risk factors such as oral contraceptives, pregnancy, complicated delivery, obesity, immobilization, infection, surgery and trauma. Information about antithrombotic therapy and family history of thromboembolism was also collected. Results from additional thrombophilic parameters (protein C activity, free protein S antigen, resistance to activated protein C, Factor V Leiden and prothrombin G20210A mutation) were collected when available. Thrombotic events were objectively diagnosed by experienced radiologists through established imaging procedures such as Doppler-ultrasonography, computed or magnetic resonance tomography for venous thrombosis and spiral computed pulmonary angiography or lung perfusion scintigraphy for pulmonary embolism.
Pediatric thrombosis is defined as any objectively diagnosed thrombotic event during childhood ($\leqslant$ 18 years). Pediatric patients in this study are divided into age groups according to the proposed WHO classification. Neonates from birth to 30 days, infants from 1 month up to 2 years, children from 2 up to 12 years and adolescents from 12 to 18 years.

c. Incidence of pediatric thrombosis

Seventy three patients (37 from Spain and 36 from Belgium) out of 968 subjects with congenital antithrombin deficiency developed a first thrombotic event before the age of 19 (Table 6 and Figure 26) corresponding to a frequency of 7.5% or 4.32 cases/1000 patient years. As thrombotic events in children are unusual, further investigations are nearly always performed. As a result, 54.8% of pediatric patients included in our study were the probands of the affected families (40/73). At first thrombotic event, fifteen of the patients were neonates, one was an infant, eight were children and forty-nine were adolescents (Figure 27). Almost half of these events were provoked by additional risk factors (35/73, 47.9%) and mainly in adolescents (25/35, 71.4%)(Table 6).

Table 6: **Characteristics of the patients with antithrombin deficiency who suffered from pediatric thrombosis** Abbreviations: HBS: Heparin binding site; CSVT: Cerebral sinovenous thrombosis. Unusual thrombosis: renal veins, CSVT, deep veins of upper extremities; *Both patients carried the p.Leu131Phe in homozygosis.

Pediatric thrombosis	**SPAIN**	**BELGIUM**	**TOTAL**	**Provoked**	**Antithrombin (anti-FXa activity)**	**Unusual thrombosis**	**Deaths**
Cases	37 (6.1%)	36 (10.1%)	**73 (7.5%)**	35 (47.9%)	52.3±10.8%	17 (23.3%)	6 (8.2%)
Age at first thrombotic event (years)	11.4 ± 7	11.5 ± 7	**11.4 ± 7**	12.4 ± 7	-	2.6 ± 5.3	5.6 ± 8
Females	13 (35.1%)	20 (55.5%)	**33 (45.2%)**	21 (63.6%)	53.0±12.5%	6 (18.8%)	3 (9.1%)
Males	24 (64.9%)	16 (44.5%)	**40 (54.8%)**	14 (35%)	51.7 ±9.2%	11 (27.5%)	3 (7.5%)
Thrombosis in adolescence (12-18year)	25 (66.7%)	24 (66.7%)	**49 (67.1%)**	25 (51%)	52.3±8.5%	1 (2%)	1 (2.0%)
Thrombosis in neonates (<30 days)	6 (16.2%)	9 (25%)	**15 (20.5%)**	8 (53.3%)	46.2±13.4%	11 (73.3%)	3 (20%)
CSVT	7 (18.9%)	6 (16.7%)	**13 (17.8%)**	7 (53.8%)	51.5±9.2%	13 (100%)	1 (16.6%)
Deaths	5 (13.5%)	1 (2.8%)	**6 (8.2%)**	2 (33.3%)	41.0±19.0%	2 (33.3%)	-
Type I deficiency	32 (86.5%)	24 (66.7%)	**56 (76.7%)**	27 (48.2%)	52.2±9.7%	11 (19.6%)	3 (5.4%)
Type II deficiency	5 (13.5%)	9 (25.7%)	**14 (19.2%)**	6 (42.8%)	50.2±15.4%	6 (42.8%)	3 (21.4%)
Type II HBS deficiency	1 (2.8%)	7 (19.4%)	**8 (11.0%)**	2 (25%)	45±19.4%	1 (12.5%)	2 (25%)*

A detailed description of all 73 cases is shown in table 7.

d. Sex and age analysis of pediatric thrombosis

Analysis by sex showed a slightly higher incidence of thrombotic events in males than in females (54.8% vs 45.2%, respectively) (Table 6). This difference was even more pronounced when considering thrombosis at early age: 10 out of 15 neonates with thrombosis (66.7%) were male (Table 7). When restricting the analysis to children below 11 years, thus excluding the role of estrogen-associated thrombosis, males showed a significantly higher risk for the development of pediatric thrombosis than females (OR 3.2; 95% CI: 1.3-78; p= 0.012). These differences in

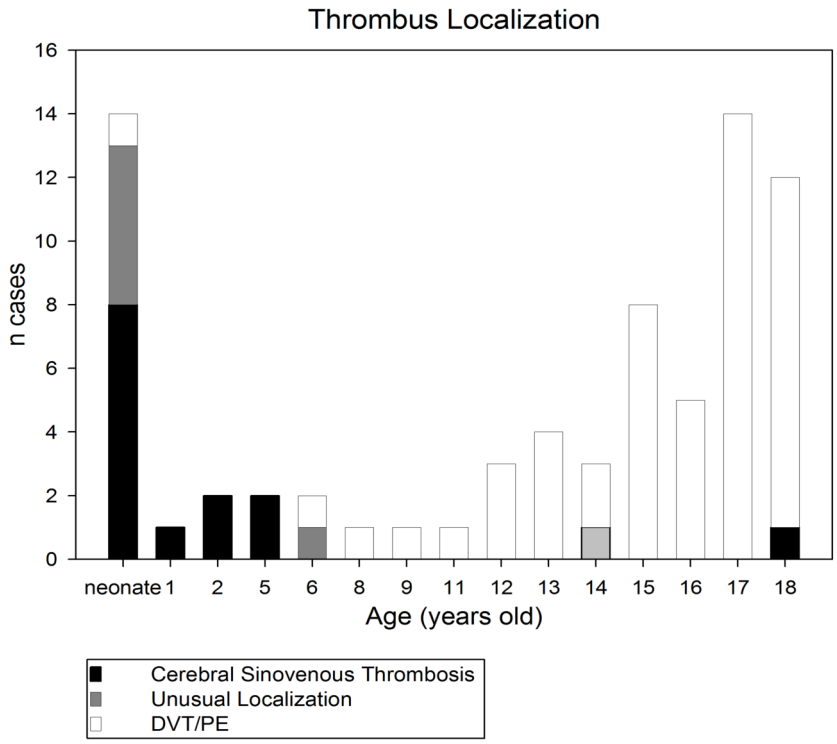


Figure 27: **Distribution of thrombotic events among children with antithrombin deficiency according to age.** The localization of the thrombosis is also represented: deep vein thrombosis of the lower limbs (DVT) and/or pulmonary embolism (PE) (white), cerebral sinovenous thrombosis (black) or unusual localizations (grey).

thrombosis-free 8 survival between males and females are also illustrated by Kaplan-Meier survival curves (Figure 28). Regarding the age of thrombotic events, the study revealed two periods with higher prevalence of thrombosis: adolescence (n= 49, 67.1%) and the neonatal period (n= 15, 20.5%) (Figure 26).

e. Thrombosis localizations

Remarkably, we observed that in neonates, thrombosis often occurred at unusual sites (11 of 15, 73.3%) (Table 6 and Figure 27) such as upper extremities, renal veins, and cerebral veins. Four patients suffered from arterial thrombosis, with associated venous thrombosis in two of them. The prevalence of cerebral sinovenous thrombosis (CSVT) was very high in our cohort (n= 13; 17.8%), especially at a young age (8 neonates and 4 children) (7 and Figure 27).
In contrast, in adolescents, clinical presentation was similar to adults: deep vein thrombosis of lower limbs and/or pulmonary embolism (n= 47) (Figure 27).

f. Risk factors associated to pediatric thrombosis

In seven neonates, the thrombotic events were idiopathic while in the other eight, possible provoking factors were identified: complicated delivery (forceps or vacuum extraction), infection/sepsis, trauma, surgery, or fetal distress. Only one thrombotic event was associated with the presence of a central venous catheter. We also point out that in cases with CSVT, three of them occurred after assisted delivery (emergency caesarian section, forceps or vacuum

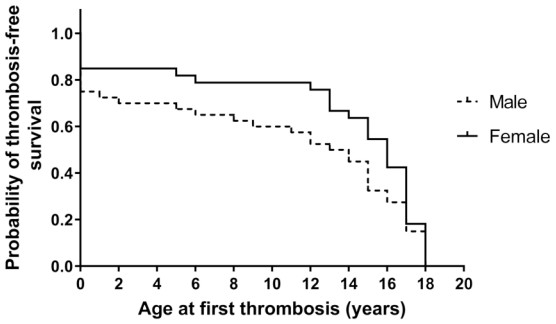


Figure 28: **Kaplan-Meier survival curves of thrombosis-free survival between male and female patients with pediatric thrombosis.**

extraction).
In the adolescence, they shared the same predisposing factors as adults: pregnancy/puerperium, oral contraceptive use, trauma, immobilization, or surgery (n= 22) (Table 7).

g. Severity of thrombosis

An interesting point to remark is the extreme severity of the events and the poor prognosis. Six children (8.2%) died as a consequence of the thrombotic event. If we only consider neonatal thrombosis, fatality rate rises to 21% (3/14) (Table 7). Interestingly, two of the neonates that deceased following the thrombotic event were unrelated homozygous carriers of the p.Leu131Phe variant, responsible for Antithrombin Budapest III, a type II heparin binding site deficiency.
Moreover, morbidity after pediatric thrombosis was strikingly high. One child needed an amputation of the arm (Figure 29), another developed serious psychomotor retardation, and one had permanent tetraplegia. Finally, VTE recurrence rate was as high as 24.6% in our cohort.

Figure 29: **Photograph of the arm of a child with thrombosis needed to be amputated.**

h. Biochemical/molecular base of pediatric thrombosis

From a molecular/biochemical point of view, the symptomatic children in our study predominantly showed type I antithrombin deficiency (76.7%). Only 14 patients presented with type II deficiency, 6 carrying genetic variations affecting the reactivity of the serpin (RS or PE) and 8 affecting the heparin-binding site (HBS) Table 7). Remarkably, in most patients (6/8) with type II HBS deficiency, the p.Leu131Phe variation was detected. Four were homozygous and the two heterozygous cases were also carriers of FV Leiden (one heterozygous and one homozygous) (Table 7).

When considering the whole cohort of subjects with AT deficiency, only 8 out of 223 (3.6%) subjects with type II HBS deficiency suffered from pediatric thrombosis and, as indicated before, most of them carried additional genetic thrombophilic risk factors or the *SERPINC1* mutation in homozygous state. In contrast, the incidence of thrombosis in children with type I deficiency was high: 56 out of 604 (9.3%). In three patients the molecular mechanism responsible for their deficiency was not found and one patient had a congenital disorder of N-glycosylation as the underlying cause of the deficiency (Table 7). Finally, we did not observe any particular recurrent mutation among our patients with pediatric thrombosis.

Three carriers from two unrelated families with p.Arg161Stop developed pediatric thrombosis, but this genetic variation was found in 24 cases from 8 unrelated families in our whole antithrombin deficient cohort. Three carriers of the p.Pro112Ser mutation from the same family presented with thrombosis during childhood, but this is a large family with 14 affected members. Two twins carrying the p.Ser223Pro mutation developed VTE at 12 and 15 years old. Finally, mutations associated with the presence of unusual disulphide-linked dimers in plasma were identified in 14 children with thrombosis, sharing the following mutations: p.Gly456Arg, p.Pro112Ser, p.Pro112Leu, c.1154-14G>A, p.Ser114Asn and p.Ser381Pro.

Case	Sex	Age of Thrombosis (years old)	Thrombosis	Risk Factors	Other Thrombophilias	Recurrence /Death /Severe Complications	Mutation	Anti-Fxa (%)	Ag AT (%)	AT Def.	Cohort
1	M	neonate	Left arm thrombosis			Arm amputation	p.Arg425His	54	93	II (RS)	1
2	M	neonate	CSVT	Forceps		Psychomotor delay	p.Met283Val	39	71	II (RS)	1
3	F	neonate	Aortic thrombosis			Deceased	p.Leu131Phe (homozygosis)	11	nd	II (HBS)	1
4	M	neonate	CSVT	Infection			p.Arg161*	52	52	I	1
5	M	neonate	Massive thrombosis	Sepsis		Deceased	p.Pro112Ser	nd	nd	I	1
6	M	neonate	DVT			Deceased	p.Leu131Phe (homozygosis)	34*	nd	II (HBS)	2
7	M	neonate	CSVT				p.Gln150Pro	55	nd	II (HBS)	2
8	F	neonate	myocardial infarction and CSVT	Clavicle facture		Recurrences	p.Gly456Arg	48	51	I	2
9	M	neonate	CSVT				p.Met284Lys	43	39	I	2
10	M	neonate	DVT and CSVT		Heterozygous Factor V Leiden		p.Phe94Ile fs X 19	50	46	I	2
11	F	neonate	DVT and CSVT	Vacuum extraction		Tetraplegic	p.Ile218Asn	68	nd	nd	2
12	F	neonate	DVT and arterial thrombosis	Cardiac surgery			p.Arg391*	45	48	I	2
13	M	neonate	CSVT and renal thrombosis				p.Arg391Profs*3	nd	nd	I	1

Case	Sex	Age of Thrombosis (years old)	Thrombosis	Risk Factors	Other Thrombophilias	Recurrence /Death /Severe Complications	Mutation	Anti-Fxa (%)	Ag AT (%)	AT Def.	Cohort
14	M	neonate	renal VT and DVT	Prematurity, fetal distress			p.Phe376Cys	43	nd	nd	2
15	M	1	CSVT and DVT		Heterozygous Factor V Leiden		p.Ile115Asn	65	50	I	1
16	M	2	CSVT			Deceased	p.Val137Ala	45	nd	I	1
17	F	2	CSVT	Cesarean section			p.Thr243Phe fs*40	46	nd	I	2
18	F	5	CSVT				p.Arg161*	59	57	I	1
19	M	5	CSVT	Infection			c.762 +1 G>A	59	nd	nd	2
20	F	6	DVT				p.Met370Thr	78	nd	I	1
21	M	6	Arterial thrombosis				p.Met370Thr	78	nd	I	1
22	M	8	DVT	Trauma		Recurrence (15 y)	p.Cys53Stop	49	nd	I	2
23	M	9	DVT			Recurrence	c.1154-14 G>A	47	44	I	2
24	M	11	DVT			Recurrences (11, 12, 17y)	PMM2-CDG	67	60	I	1
25	M	12	DVT	Immobilisation of legs			p.Ser223Pro	49	43	I	1
26	M	12	PE and DVT				p.Arg445Serfs*17	37	42	I	1
27	F	12	DVT and PE	Surgery		Recurrence	p.Trp257Arg	59	nd	I	2
28	F	13	DVT	Oral contraceptives		Recurrences	deletion exon 5	41	41	I	2
29	M	13	PE and DVT				Complete *SERPINC1* deletion	52	46	I	1

Case	Sex	Age of Thrombosis (years old)	Thrombosis	Risk Factors	Other Thrombophilias	Recurrence /Death /Severe Complications	Mutation	Anti-Fxa (%)	Ag AT (%)	AT Def.	Cohort
30	F	13	DVT				p.Arg161*	nd	nd	I	1
31	F	13	DVT				p.Leu131Phe (homozygosis)	nd	nd	II (HBS)	2
32	M	14	PE				p.Ser381Pro	56	66	I	1
33	F	14	Renal VT			Recurrence, deceased	p.Met1Ile	55	57	I	1
34	M	14	DVT	Renal agenesia		Recurrence (16y)	p.Pro318Leu	55	88	II (RS)	2
35	M	15	DVT	Long flight		Recurrences (27,42y)	intron 3 c.625-2A>G	44	47	I	1
36	M	15	DVT	May Thurner Syndrome	Heterozygous Factor V Leiden		p.Pro112Leu	40	40	I	1
37	M	15	DVT				p.Arg229*	56	nd	I	1
38	M	15	PE	Infection: varicella			p.Lys157Arg	60	60	I	1
39	F	15	DVT and PE	Oral contraceptives			p.Pro384Pro	58	nd	I	2
40	F	15	Extensive PE	Oral contraceptives			p.Thr185Profs*99	46	nd	I	2
41	F	15	DVT	Oral contraceptives			p.Pro318Leu	65	64	II (RS)	2
42	M	15	DVT				p.Ser223Pro	52	39	I	1
43	F	16	DVT and PE				c.1154-14 G>A	40	57	I	1
44	F	16	DVT	Puerperium		Recurrence	p.Pro112Ser	60	nd	I	1
45	M	16	DVT and PE				not found	66	57	I	1

Case	Sex	Age of Thrombosis (years old)	Thrombosis	Risk Factors	Other Thrombophilias	Recurrence /Death /Severe Complications	Mutation	Anti-Fxa (%)	Ag AT (%)	AT Def.	Cohort
46	F	16	DVT		Heterozygous Factor V Leiden		p.Leu131Phe	53	nd	II (HBS)	2
47	F	16	PE				Complete *SERPINC1* deletion	55	nd	I	2
48	M	16	DVT	Trauma			P.Leu131Phe (homozygosis)	nd	nd	II (HBS)	2
49	M	17	PE				p.Ser114Asn	48	47	I	1
50	F	17	DVT	Surgery		Recurrence	p.Met370Thr	64	nd	I	1
51	F	17	DVT	Puerperium		Recurrence	p.Pro112Ser	54	nd	I	1
52	F	17	DVT	Pregnancy		Recurrence	p.Pro112Ser	60	nd	I	1
53	M	17	DVT				duplication exon 6	41	53	I	1
54	M	17	DVT				p.Arg161*	61	50	I	1
55	F	17	DVT				p.Arg425del	59	100	II (RS)	1
56	F	17	DVT and PE	Oral Contraceptives and trauma			p.Met284Thr	37	38	I	2
57	F	17	DVT and PE			Recurrences (29, 36y)	p.Arg391*	59	53	I	2
58	F	17	DVT and PE	Oral contraceptives			p.Asn224Glu fs X 55	42	nd	I	2
59	M	17	PE	Trauma		Recurrence (18 y)	p.Thr185Pro fs*99	52	45	I	2
60	F	17	DVT and PE	Oral contraceptives			c.1154-14 G>A	61	46	I	2
61	M	17	DVT			Recurrence (26 y)	p.Gln70Asp fs*33	53	62	I	2

Case	Sex	Age of Thrombosis (years old)	Thrombosis	Risk Factors	Other Thrombophilias	Recurrence /Death /Severe Complications	Mutation	Anti-Fxa (%)	Ag AT (%)	AT Def.	Cohort
62	F	18	CSVT	Oral contraceptives	Heterozygous prothrombin G20210A mutation		Lys165Glufs*7	44	43	I	1
63	M	18	DVT	Infection			p.Gly456Arg	50	60	I	1
64	F	18	PE	Oral contraceptives		Deceased	p.Ser426Leu	60	nd	II (RS)	1
65	M	18	DVT and PE				p.Leu162His	54	nd	I	2
66	F	18	DVT	Oral contraceptives			c.1154-14 G>A	64	72	I	2
67	M	18	DVT	Plastering of leg		Recurrence (22y)	not found	57	48	I	2
68	M	18	DVT	Obesity(105Kg) surgery	Homozygous Factor V Leiden		p.Leu131Phe	52	78	II (HBS)	2
69	F	18	DVT	Oral contraceptives			p.Leu372Pro	44	48	I	2
70	F	18	DVT	Oral contraceptives		Recurrence (18y)	p.Arg229*	53	50	I	2
71	F	18	DVT				p.Arg79His	65	nd	II (HBS)	2
72	M	18	DVT				c.1154-2 A>G	50	54	I	2
73	M	18	PE and DVT				p.Arg445Serfs*17	31	40	I	1

Table 7: Detailed description of 73 patients with antithrombin deficiency and pediatric thrombosis. Abbreviations:CSVT (Cerebral sinovenous thrombosis); MTHFR (methylenetetrahydrofolate reductase); DVT (Deep vein thrombosis); PE (Pulmonary embolism); HBS (Heparin binding site); RS (Reactive site). Nd: not determined * patient number 6: anti-FXa measured 4 days after birth and after plasma transfusion

4.1.2 Antithrombin p.Thr147Ala: The First Founder Mutation in People of African origin Responsible for Inherited Antithrombin Deficiency

a. Introduction

There is a wide variable clinical implication in patients with congenital antithrombin deficiency. In the previous chapter, we presented severe clinical consequences of Antithrombin deficiency, here in contrast, we present results of a SNV with mild functional effect that have moderate pathogenicity but widely spreading in a specific population.

Although, antithrombin deficiency is a rare disease, the prevalence in the general population is heterogeneous depending on the type of deficiency. Type II deficiencies, caused by genetic defects not impairing the folding or secretion but the anticoagulant function, and associated to milder risk of thrombosis are more frequently found in the general population (1/500) than type I deficiencies, with no aberrant antithrombin in plasma and higher risk of thrombosis, which had lower prevalence in the general population (1/5000) [131] .

Recurrent mutations in *SERPINC1* are rare. No mutational *hotspots* have been described in *SERPINC1*, and only two relatively frequent mutations, both causing a Type II deficiency, have a founder effect, both described in European populations: one mutation from Hungarian origin, p.Leu131Phe [132] and the other one from Finish precedence: p.Pro73Leu [133].

In collaboration with a reference group for antithrombin deficiency in Belgium, we have investigated the antithrombin variant p.Thr147Ala. This variant was found in 12 patients with the same Black-African origin from a total cohort of 421 patients with congenital antithrombin deficiency in the Belgian cohort. This variant has been reported as a single nucleotide polymorphism (SNP) rs2227606 in the *SERPINC1* gene, absent in Caucasian population, but relatively frequent in African population. Moreover, *in silico* prediction tools for pathogenicity render conflicting results. Therefore, the aim of this study was to elucidate the functional effect of this mutation and to evaluate a potential founder effect. We hypothesize that this variant might be a pathogenic mutation of mild functional consequences and clinical impact, restricted to individuals of Black-African ancestry.

b. Patients

For this study we have we focused on the 12 patients carrying the rs2227606 polymorphism recruited from the Belgian cohort of patients with antithrombin deficiency. None of these carriers showed any other pathogenic mutation, large deletion or insertion in *SERPINC1*.

We collected information of these patients concerning venous and arterial thrombosis, recurrent events, age at first and recurrent events, obstetrical history, ethnicity, presence of acquired risk factors for VTE (immobilization, surgery, oral contraceptive use, pregnancy, postpartum period, obesity), acquired risk factors for arterial thrombosis (obesity, smoking, hypertension). Results of thrombophilia screening (lupus anticoagulant, anticardiolipin antibodies, anti-β2 glycoprotein I antibodies, protein C activity, free protein S antigen, resistance to activated protein C, factor V Leiden mutation,

Figure 30: **Countries of origin of the 12 patients with antithrombin deficiency carriers of the *SERPINC1* rs2227606 polymorphism**.

and prothrombin G20210A mutation) were also collected from all patients. Familial history of venous or arterial thrombosis was also saved.

The clinical features of the cohort are shown in the table 8 Eleven of the 12 subjects were female and 7 of them suffered from obstetric complications (miscarriage, preeclampsia, and failure to conceive). Five patients developed VTE and one experienced an arterial thrombotic event. Only one patient reported a familial history of thrombosis. Thrombophilia screening was normal except in patient 10 (P10), for whom reduced free protein S levels (54–57%) were measured at different occasions. Genetic analysis of the *PROS1* gene in P10 showed a homozygous p.Val510Met mutation, causative of protein S deficiency. Interestingly, all 12 patients were unrelated and of Black-African origin: 5 originate from Congo, 2 from Ruanda, 2 from Cameroun, 1 from Ivory Coast, 1 from Angola, and 1 patient did not specify the country of origin (Figure 30).

Table 8: **Clinical characteristics of the patients carrying the *SERPINC1* rs2227606 polymorphism, included in the study.** Abbreviations: CVA, cerebral vascular accident; DVT, deep vein thrombosis; F, female; M, male; OC, oral contraception; ns, not specified.

Patient	Sex	Clinical presentation			Familial history of VTE
		Age	Event	Risk factor(s)	
P1	F	29	Pulmonary embolism	OC, obesity	+
		31	Miscarriage at 12 weeks gestation (thromboprophylaxis: nadroparine 0.6mL/d)	Obesity	
P2	F	ns	Incidental finding of non-recent portal vein thrombosis	None	-
P3	F	ns	2 early miscarriages	/	-
		30	Severe preeclampsia (neurological complications, coma, epilepsy) at 5 months gestation of twin pregnancy, loss of 1 foetus	Hypertension	
P4	F	40	CVA (occlusion of left arteria cerebri media) with hemiplegia and aphasia	None	-
P5	F	ns	3 early miscarriages	Obesity	-
P6	F	ns	None (incidental finding in context of fertility work-up)	None	-
P7	F	60	Proximal DVT of right leg	Obesity	-
P8	F	38	None (incidental finding in context of fertility work-up)	None	?
P9	M	35	Thrombophlebitis of left leg	/	?
		60	Proximal DVT of left leg	Surgery (renal transplant) + corticosteroid use	
		62	Thrombosis of arteriovenous fistula	Tamoxifen treatment + corticosteroid use	
		64	Proximal DVT of left leg	Tamoxifen treatment + corticosteroid use	
P10	F	43	Pulmonary embolism	Protein S deficiency	-
P11	F	26	Early miscarriage	None	-
		38	Transient ischemic attack	Recent long-haul flight	
P12	F	24	Preterm delivery due to gestational hypertension	Obesity	-
		27	Stroke-like episode	Pregnancy, Obesity	

The SNP rs2227606 is a substitution of adenine to guanine at cDNA position 439 (c.439A>G) in exon 3 of the *SERPINC1* gene. c.439A>G causes a missense change in antithrombin: p.Thr147Ala. Minor allele frequency (MAF) data extracted from Exome Aggregation Consortium (ExAC), 1000 Genomes, and gnomAD showed a significant difference between distinct ethnic groups. The SNP is nearly absent in Caucasian and Asian populations, whereas its MAF in African population varies between 0.5 and 1.0%. (Table 9).

Table 9: **Frequency of the rs2227606 (*SERPINC1* chr1:173911984 -GRCh38.p12- T>C) single nucleotide polymorphism responsible for the antithrombin p.Thr147Ala variant.**

Study	Population	Sample size	Ref Allele	Alt Allele
1000 Genomes	Global	5008	T=0.997	C=0.003
1000 Genomes	African	1322	T=0.990	C=0.010
1000 Genomes	East Asian	1008	T=1.000	C=0.000
1000 Genomes	Europe	1006	T=1.000	C=0.000
1000 Genomes	South Asian	978	T=1.000	C=0.000
1000 Genomes	American	694	T=1.000	C=0.000
ExAC	Global	121412	T=0.99942	C=0.00058
ExAC	Europe	73354	T=1.0000	C=0.0000
ExAC	American	11578	T=0.9999	C=0.0001
ExAC	African	10406	T=0.9933	C=0.0067
ExAC	Other	908	T=1.00	C=0.00
gnomAD-Exomes	Global	251448	T=0.99959	C=0.00041
gnomAD-Exomes	European	135384	T=0.99999	C=0.00001
gnomAD-Exomes	Asian	49008	T=1.0000	C=0.0000
gnomAD-Exomes	American	34586	T=0.9998	C=0.0002
gnomAD-Exomes	African	16256	T=0.9943	C=0.0057
gnomAD-Exomes	Ashkenazi Jewish	10078	T=1.0000	C=0.0000
gnomAD-Exomes	Other	6136	T=1.000	C=0.000
gnomAD-Genomes	Global	31380	T=0.9984	C=0.0016
gnomAD-Exomes	European	18892	T=1.0000	C=0.0000
gnomAD-Exomes	African	8702	T=0.994	C=0.006
gnomAD-Exomes	East Asian	1560	T=1.000	C=0.000
gnomAD-Exomes	Other	1088	T=1.000	C=0.000
gnomAD-Exomes	American	848	T=1.00	C=0.00
gnomAD-Exomes	Ashkenazi Jewish	290	T=1.00	C=0.00
GO Exome Sequencing Project	Global	13006	T=0.9982	C=0.0018
GO Exome Sequencing Project	European American	8600	T=1.000	C=0.000
GO Exome Sequencing Project	African American	4406	T=0.995	C=0.005
TopMed	Global	125568	T=0.99814	C=0.00186

We evaluated the prevalence of the variant in 100 randomly selected subjects of African ancestry and identified one subject carrying the rs227606 variant in the heterozygous state, confirming the allele frequency of 0.5% (carrier frequency of 1%) in the African population. This subject, who had no thrombotic event, also showed a reduced antithrombin activity of 59%.

In silico prediction tools designed to evaluate pathogenicity of sequence variations render conflicting results when evaluating the p.Thr147Ala variant (Table 10) . Six of the evaluated prediction tools classify this variant as pathogenic (CADD, FATHMM, MutationTaster, PolyPhen-2, PredictSNP2, and Provean), while the remaining seven consider this variant to be benign. According to the recommendations of the American College of Medical Genetics and Genomics, this mutation should be classified as a "variant of uncertain significance" (criteria for PM1, PM2, and PP2 rules are met)[134] .The PP3/BP4 rules addressing results from *in silico* tools cannot be used since this rule requires agreement between the different computational methods.

Table 10: **Results of the different *in silico* prediction tools for the *SERPINC1* p.Thr147Ala variant.**

In Silico tool	website	Result
ALIGN GVDG	http://agvgd.hci.utah.edu/agvgd_input.php	Unlikely pathogenic
CADD	https://cadd.gs.washington.edu/score	Deleterious
FATHMM	http://fathmm.biocompute.org.uk/	Damaging
Meta-SNP	https://snps.biofold.org/meta-snp/	Neutral
Mutation Taster	http://www.mutationtaster.org/	Disease causing
PANTHER	http://www.pantherdb.org/tools/	Neutral
PhD-SNP	https://snps.biofold.org/phd-snp/phd-snp	Neutral
SIFT	https://sift.bii.a-star.edu.sg/	Tolerated
SNAP	http://www.bio-sof.com/snap	Neutral
SNPs&Go	https://snps.biofold.org/snps-and-go/snps-and-go.html	Neutral
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/	Possibly damaging
PredictSNP2	https://loschmidt.chemi.muni.cz/predictsnp2/	Deleterious
Provean	http://provean.jcvi.org/index.php	Damaging

c. Plasma Antithrombin Assays

The analysis of antithrombin antigen levels in these patients were normal or slightly reduced (91.2%; range 70±110%) while anti-Xa activity was reduced with only one specific commercial test, Innovance Antithrombin (Figure 31). Activities measured with both HemosIL Liquid AT (anti-Xa based, 98.9±10.6%) and Biophen AT anti-IIa (100.9±10.2%) were significantly higher when compared with those measured with Innovance Antithrombin (61.8±9.3%) and showed results within the normal range (80±120%). Based on the activity-to-antigen ratio, these subjects show a qualitative defect or type II phenotype.

Others and we have previously shown that incubation time of the plasma with FXa or thrombin plays a crucial

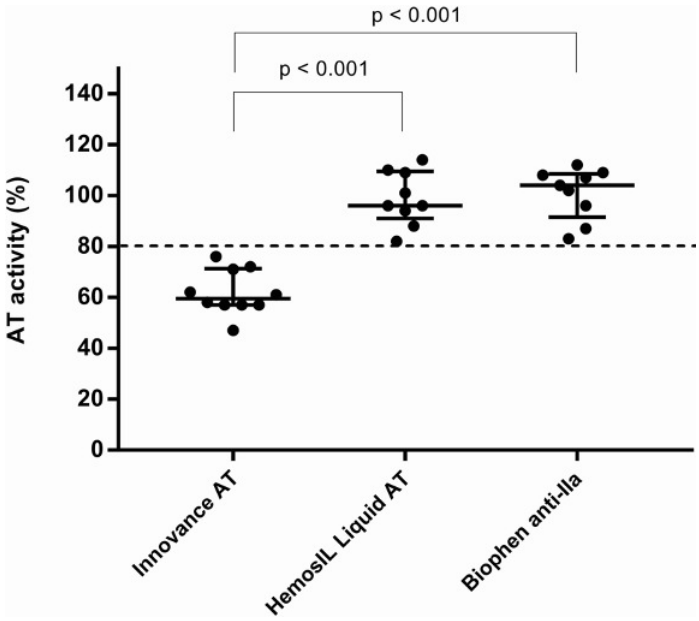


Figure 31: **Antithrombin activity measured with three different commercial assays.** Dashed line: lower limit of normal range

role in the sensitivity of the activity assays for some specific type II mutations [116, 135] . To assess if this is the case for this variant, antithrombin anti-Xa activity was measured with HemosIL Antithrombin kit using different incubation times ranging from 0 to 15 minutes. Figure 32 shows an incubation time-dependent reduction in antithrombin activity in the plasma of the patients. Every shortening in incubation time (except t1 vs. t0) resulted in a significant reduction of antithrombin activity. This effect was more pronounced when comparing the results of t0 to t15 (p= 0.0002), where we observed an average reduction of 22.6±9.2%. The antithrombin activity of normal human plasma remained constant when reducing incubation time. When using the incubation time proposed by the manufacturer, i.e., 3minutes, only three out of the nine investigated patients would show an antithrombin activity below the reference range. The heparin binding ratio of the patient plasma was clearly reduced when compared with the normal plasma, 0.78 and 1.03, respectively. This is indicative for a HBS defect. We performed CIE on

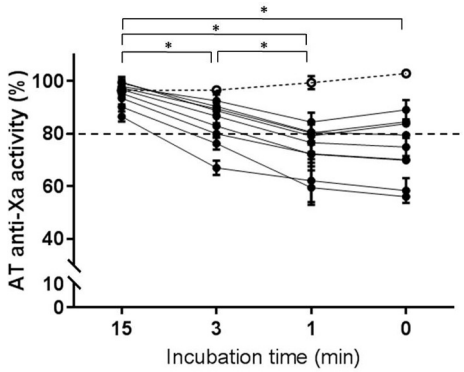


Figure 32: **Antithrombin activity measured with HemosIL Antithrombin and different incubation times**. Full lines: patient plasma; dotted line: normal pooled plasma. The symbol "*" indicates statistical significant difference. Dashed line: lower limit of normal range.

the plasma of patient 5 (Figure 33). At physiological ionic strength, the pattern is similar to that of normal plasma. When treating the plasma with a higher salt concentration to impair the binding of heparin to antithrombin, an antithrombin fraction with a lower heparin affinity appears. This fraction did however not reach the levels that were observed in a heterozygous carrier of a type II HBS deficiency such as Antithrombin Toyama (p.Arg79Cys).

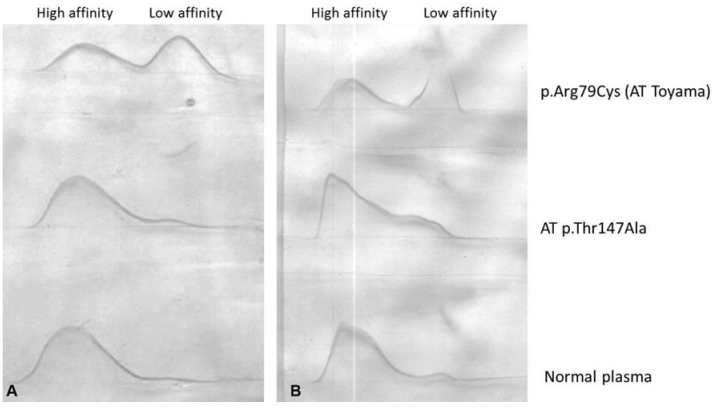


Figure 33: **Crossed immunoelectrophoresis of the plasma of a patient carrying the p.Thr147Ala variant** (P5). Plasma of a patient with type II HBS (p.Arg79Cys) and plasma from a healthy control (Normal plasma) were used as controls. (A) Physiological ionic strength (150 mM NaCl). (B) High ionic strength (0.8 M NaCl).

Plasma of four patients and four healthy controls was subjected to heat stress by incubation at 40°C for 24 hours. This induced an important reduction of anti-Xa activity in patients (29.8±12.0%) compared with the controls (13.0±6.5%) when measured with HemosIL Antithrombin. The basal antithrombin activity was already reduced in patients when determined with Innovance Antithrombin and no further reduction was observed after heat stress (Figure 34).
Heating of the plasma did not induce an increase of latent antithrombin or polymers (Figure 35).

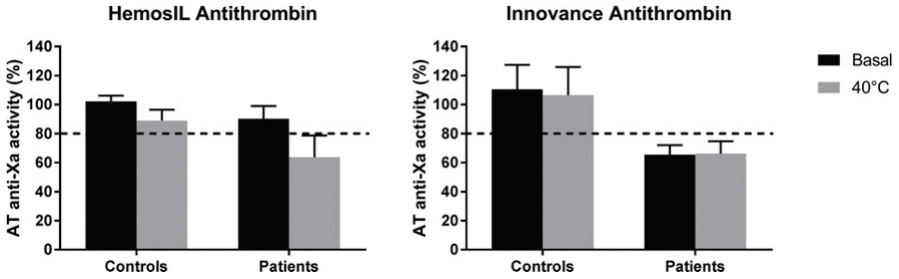


Figure 34: **Antithrombin activity in patients and controls under basal conditions and after incubation for 24 hours at 40°C.** Dashed line: lower limit of normal range.

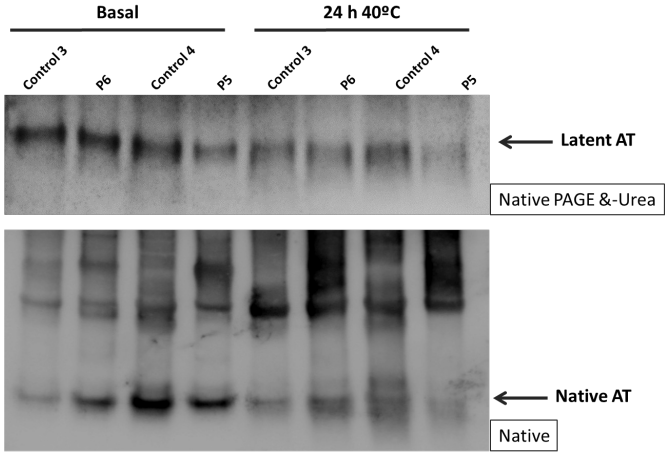


Figure 35: **Electrophoretic characterization of plasma antithrombin in patient and control plasma.** Top part shows Western Blotting of plasma samples of PAGE run under native conditions containing 6M urea. Latent antithrombin (AT) is indicated by an arrow. Bottom part of the figure shows Western Blotting of plasma samples of PAGE run under native conditions. Native antithrombin is indicated by an arrow.

d. Recombinant Expression

The p.Thr147Ala variant and WT antithrombins were expressed in a recombinant model. The anti-Xa activity was measured using two different anti-Xa assays. The purified recombinant p.Thr147Ala protein displayed an anti-Xa activity of approximately 50% of the WT recombinant antithrombin as shown in Figure 36 (51.9±2.2.% for Innovance Antithrombin and 55.0±3.3% for HemosIL Antithrombin). Interestingly, anti-Xa activity was reduced with both assays, in contrast to the data observed in plasma samples. Heparin affinity was assessed by determining the equilibrium dissociation constants of the recombinant proteins for the pentasaccharide by following changes in intrinsic fluorescence. The dissociation constants (Kd values) for the WT and mutant proteins were 1.18±0.07 and 0.84±0.06 nM, respectively, indicating no significant difference in heparin affinity.

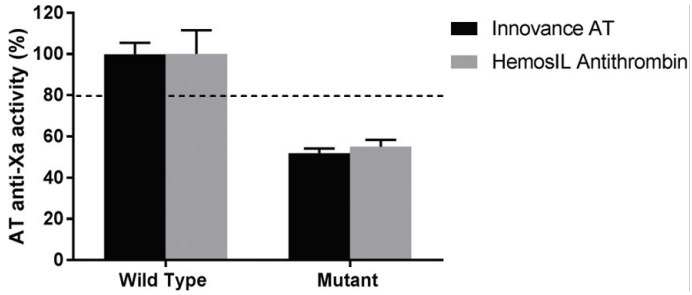


Figure 36: **Antithrombin activity of the recombinant antithrombin proteins using two different anti-Xa activity assays.** Dashed line: lower limit of normal range.

TAT Complex Formation

The formation of TAT complexes, evaluated by SDS-PAGE under reducing conditions, showed identical results in carriers and controls (Figure 37). Similar results were observed with the recombinant proteins (Data not shown).

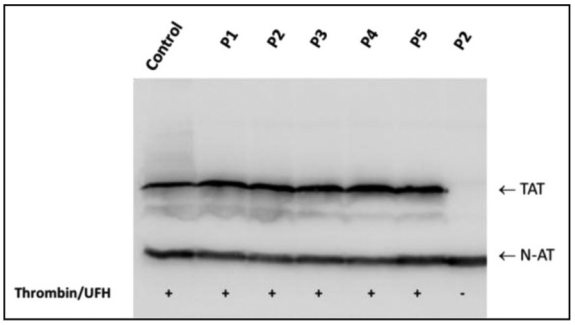


Figure 37: **8 SDS-PAGE under reducing conditions and immunoblot of plasma antithrombin of patients (P)carrying the p.Thr147Ala and a healthy control, in presence (+) or absence (-) of thrombin and unfractionated heparin (UFH).** N-AT: native antithrombin; TAT, thrombin–antithrombin complexes.

Structural Modeling

The affected residue Thr147 is a moderately conserved amino acid in the serpin superfamily, located at the bottom of helix D of the protein, in the loop connecting with helix C. This polar residue forms three hydrogen (H) bonds with amino acids Arg79, Ser144, and Glu145 (Figure38A). Mutation of this amino acid into a nonpolar alanine abolishes these three hydrogen bonds (Figure38B). The H-bond with Arg79 is of specific interest as this residue is known to directly interact with heparin. The interaction with Glu145 could also be of importance as this is a conserved residue among serpins.

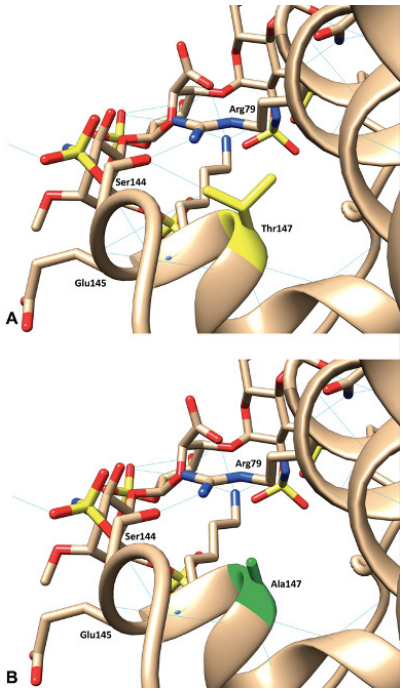


Figure 38: **Three-dimensional modeling of antithrombin**. (A) Wild-type antithrombin with residue 147 (Thr) depicted in yellow. (B) Mutant antithrombin with residue 147 (Ala) shown in green. Blue lines represent H-bonds. Heparin is shown in stick representation and antithrombin in ribbon. HBS, heparin binding site.

Haplotype Determination

Nine DNA samples of patients with the p.Thr147Ala mutation were available for haplotype analysis. LR-PCR amplicons were obtained in eight out of the nine carriers. These amplicons were sequenced for 20 minutes in the nanopore flow cell, and a mean of 367 and 460 reads, from LR1-LR2 and LR3-LR6, respectively, were obtained. Integrative Genomics Viewer and Illumina Variant Studio analyses revealed a profile of genetic variations identified in each allele, which included the same markers for the mutated allele.

This analysis revealed a common haplotype shared by all carriers of the mutation (Table 11). Reads carrying the p.Thr147Ala mutation were selected by bioinformatics tools using "grep" command in Linux and confirmed that all patients shared the haplotype described in table 11 This common haplotype contains 13 small nucleotide variants whose localization and MAF in Caucasians and Africans are shown in table 11 and figure 39.

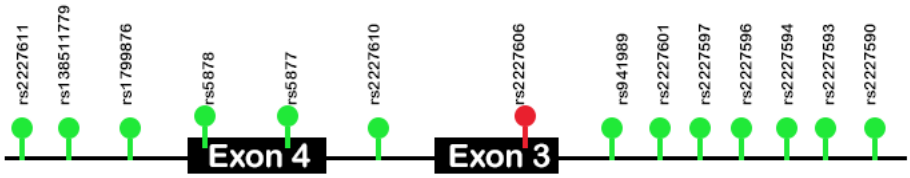


Figure 39: **Schematic representation of the common haplotype of 13 SNV present in the same allele of p.Thr147Ala mutation.** Green symbol represents the shared rs. Red symbol represents p.Thr147Ala mutation.

Table 11: **Single Nucleotide Variations linked to the p.Thr147Ala variation (italic bold) in *SERPINC1* located on the same allele.** # According to GRCh37/hg19. * Minor allele frequency according to 1000 Genomes.

dbSNP ID	Coordinate (chrom 1)#	Change	Localization	Codon	Missense change	MAF* (%) Caucasian	MAF* (%) African
rs2227611	173875467	A>G	Intron variant	–	–	0.3	27.6
rs138511779	173877033	C>T	Intron variant	–	–	0.0	26.0
rs1799876	173878471	A>G	Intron variant	–	–	33.1	83.5
rs5878	173878832	T>C	Exon 5	337	Q (synonymous)	34.7	84.4
rs5877	173878862	T>C	Exon 5	327	V (synonymous)	34.0	83.7
rs2227610	173879178	T>C	Intron variant	–	–	0.0	2.0
rs2227606	***173881122***	***T>C***	***Exon 3***	***147***	***T/A***	***0.0***	***1.0***
rs941989	173881871	C>T	Intron variant	–	–	34.2	84.4
rs2227601	173882730	T>C	Intron variant	–	–	12.3	68.5
rs2227597	173884739	C>G	Intron variant	–	–	12.1	44.8
rs2227596	173884793	T>C	Intron variant	–	–	15.9	78.7
rs2227594	173885481	C>T	Intron variant	–	–	12.1	44.8
rs2227593	173885536	G>T	Intron variant	–	–	12.1	44.9
rs2227590	173886037	G>A	Intron variant	–	–	12.1	46.9

4.2 New molecular defects involved in Antithrombin Deficiency

4.2.1 Identification of a New Mechanism of Antithrombin Deficiency Hardly Detected by Current Methods: Duplication of *SERPINC1* Exon 6

a. Introduction

Molecular analysis of antithrombin deficiency usually starts with the sequencing of the seven exons and flanking regions of *SERPINC1*. MLPA is usually restricted to cases with negative finding after the first sequencing screening. These two methods identify *SERPINC1* gene defects in up to 80% of cases with antithrombin deficiency [136] , [137]. However, a significant proportion of cases remain unexplained following this diagnostic algorithm. Other molecular defects must cause antithrombin deficiency. The analysis of cases with antithrombin deficiency but no *SERPINC1* gene defects revealed a high proportion of these cases (27%) with hypoglycosylation caused by mutations in genes involved in the generation of the N-glycan precursor [86]. We speculated that *SERPINC1* gene defects not detected by current methods might be responsible for some cases with still unknown molecular base.

b. Molecular diagnosis of antithrombin deficiency

The sequence analysis of the 7 exons and flanking regions of *SERPINC1* in the 223 unrelated cases with confirmed antithrombin deficiency recruited until 2016 in our cohort, identified mutations in 166 cases, with 95 different gene variations: 84 SNVs (53 missense, 8 non-sense, 8 deletions, 4 insertions, and 11 splicing), 10 small deletions, and 1 small insertion (Table 1). MLPA of cases without mutation revealed 6 large deletions, present at the heterozygous state: whole gene deletions in 3 patients and 3 partial deletions removing either exon 1 (n= 1), 4 (n = 1), or exons 2-5 (n= 1). Electrophoretic analysis of plasma antithrombin revealed 11 cases with increased levels of hypoglycosilated forms of antithrombin. A disorder of glycosylation was confirmed in all these cases by HPLC analysis of transferrin glycoforms. Finally, sequencing of the promoter and the first two introns of remaining 40 cases revealed 3 regulatory mutations.

Aiming to find potential gene defects located at flanking regions of exon 6, not covered by the original set of primers, we designed deeper intronic primers. This new set of primers revealed one out of the 37 cases with antithrombin deficiency but no genetic defect carrying a heterozygous insertion of 193 bp. P1 was a 42-year-old woman patient who suffered from deep venous thrombosis and had mild antithrombin deficiency (Anti-FXa activity: 75%), and family history of antithrombin deficiency. Analysis of this insertion in all available relatives showed that all carriers also had antithrombin deficiency (Figure 40). Sequencing of the PCR product revealed a duplication in tandem of 193 bp comprising exon 6: c.1154-13_ 1218+115dup (Figure 41).

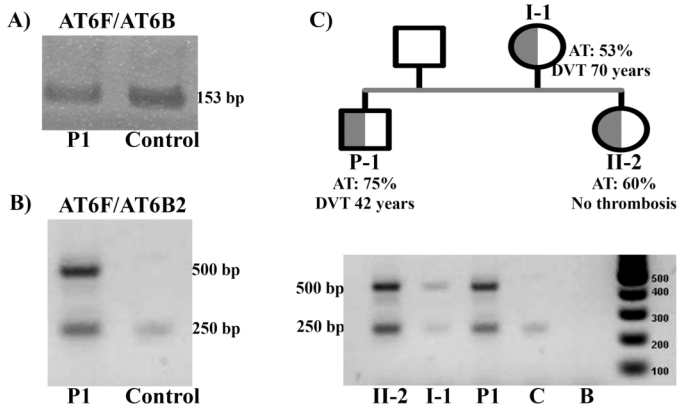


Figure 40: **Identification of a 193 bp insertion involving exon 6 in P1.** A) PCR amplification of exon 6 using primers close to exon 6. B) PCR amplification of exon 6 using primers with deeper intronic localization. C) The family pedigree of P1 showing the anti-FXa activity and the amplification of exon 6 with the second set of primers. C: Control; B: Blank; MW: Molecular weight marker. AT: Anti-FXa activity. DVT: Deep venous thrombosis.

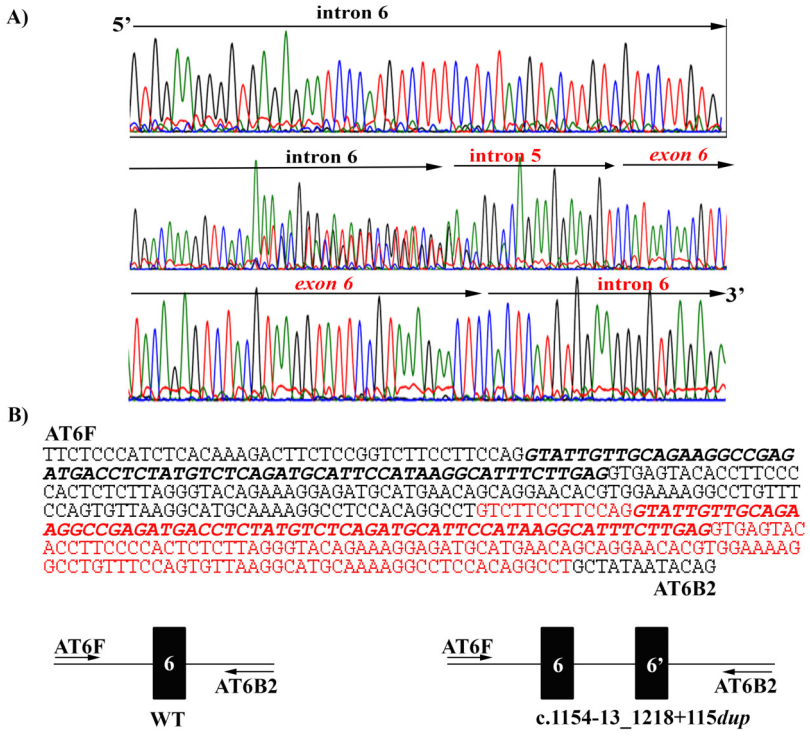


Figure 41: **Duplication in tandem of exon 6.** A) Electropherogram of the whole exon 6 PCR amplicon of P1. Exon 6 nucleotides are underlined. B) Sequence of the duplicated region and schematic representation of the duplication of exon 6 detected in P1. Duplication is marked in red, and exons are shown in bold-italic font.

This positive finding of a duplication done by PCR and sequencing and but the negative result obtained by MLPA, together with the small size of the exon 6 (65 bp), encouraged to perform fine adjustments for the MLPA assay

following the indications of the technical service of MRC-Holland. The DNA was diluted in LOW-TE, and temperature and time of injection was changed in the capillary electrophoresis (from 60 to 50ºC, and from 18s to 15s, respectively). Under these new conditions, the duplication of exon 6 in P1 was detected by MLPA (Figure 42).

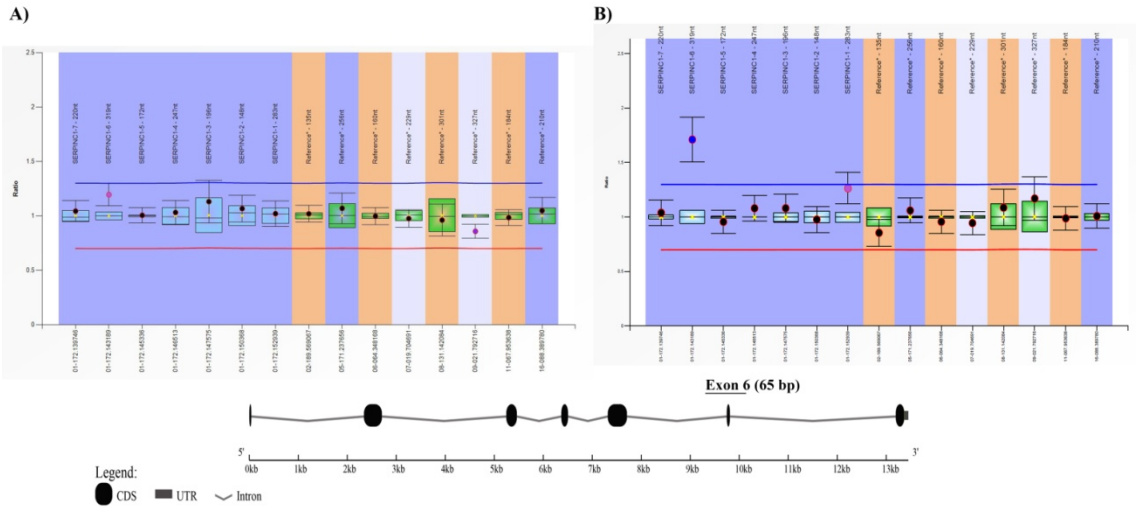


Figure 42: **MLPA results obtained in P1 using two different adjustments of capillary electrophoresis.** A) Temperature 60ºC and injection time 18s. B) Temperature 50ºC and injection time 15s. A schematic representation of the *SERPINC1* gene architecture indicating the size and localization of exons is also shown.

Interestingly, NGS of the whole *SERPINC1* using a PGM platform failed to detect this duplication despite two amplicons covered the duplication detected in P1.

c. Specific diagnosis of tandem duplication of exon 6.

As detection of duplications of exon 6 by PCR amplification might depend on the extension of the duplication and the localization of primers for PCR amplification, and MLPA might not detect the duplication of this small exon, we develop a simple and specific method that only amplified tandem duplications of exon 6. The combination of a forward primer annealing to the 3'-end of exon 6 and a reverse primer annealing to the 5' of exon 6 do not amplify the wild type allele, but could amplify tandem duplications involving exon 6. Actually, this method worked perfectly as it rendered a PCR amplicon of 193 bp in P1 (Figure 43). We evaluated potential new tandem duplications among 36 patients with still unknown molecular base, detecting a second case, P2, with a tandem duplication of bigger size (893 pb) than that of P1 (193 pb) (Figure 43).

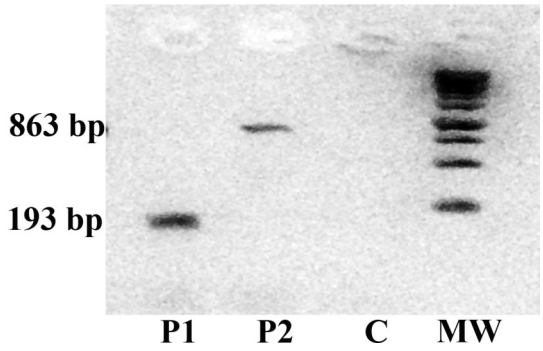


Figure 43: **Specific detection of tandem duplication of exon 6 by using a specific set of primers in P1, and detection of a new case (P2) with antithrombin deficiency carrying a different duplication of exon 6**. C: Healthy control. MW: Molecular Weight marker.

Sequencing of this PCR product confirmed the dupli-

cation (c.1154-305_ 1218+493dup) (Figure 44). P2 is a 17-year-old male patient who developed deep venous thrombosis. He had 41% of plasma anti-FXa activity and reported other relatives with antithrombin deficiency.

P2: c.1154-305_1218+493*dup* (863 bp)

TGGCTAATTTTTGTATTTTCACGATAGAGACAAGGTTTCACCATATTAGTTAGAGTGTCT
CCAACTCCTGACCTCAGGTGATCTGTACACCTTGGCCTCCCGAAGTGCTGGGATTACAGG
TGTGAGCCACTGCACCCAGCCCCCAAAGTACTTTATTATTTTTAACACATATTCATTGTG
AGAGTATGATTAGGTGAAGATTTAGGATTTCTTCTTATGTTTCAAAAAGCCCCAAAGGAT
CTCTTAATCCAAACTGAATTCCCATCTGTGGGTTGAAGCCAACTTTCTCCCATCTCACAA
AGACTTCTCCGGTCTTCCTTCCAG***GTATTGTTGCAGAAGGCCGAGATGACCTCTATGTCT***
CAGATGCATTCCATAAGGCATTTCTTGAGGTGAGTACACCTTCCCCACTCTCTTAGGGTA
CAGAAAGGAGATGCATGAACAGCAGGAACACGTGGAAAAGGCCTGTTTCCAGTGTTAAGG
CATGCAAAAGGCCTCCACAGGCCTGCTATAATACAGCCCTCTCCAAAACCTTCATGGTGT
GATTGTTCTGCCTTCCCTCCCACTACCTCTTCTGTAGCAGGTCAAGCGGGAACACAAACA
TTTAGGGAGGGTGATATAGGAAAAGAAGCCAGCAAAGGCCATCAAGAAGAAATTTACAGC
ATGAGGAGAACCAGAAGAGTATGGGGTCGCAGAAACCCAGGGAGAATTTTTTTTTTTTTT
TGAGACAGAGCTTCGTTCGCTCGTTGCCCAGGCTAGAGTGCAATGGTGCGACAGCTCACT
ACAACTTCTGCCTCCCGCGTTCAAGCGATTCTCCTGCCTCAGCCTCCTGGGTAGCTGGGA
TGACAGGCATGTGCCATCACGCCCGGCTAATTTTTGTATTTTCACGATAGAGACAAGGTT
TCACCATATTAGTTAGAGTGTCTCCAACTCCTGACCTCAGGTGATCTGTACACCTTGGCC
TCCCGAAGTGCTGGGATTACAGGTGTGAGCCACTGCACCCAGCCCCCAAAGTACTTTATT
ATTTTTAACACATATTCATTGTGAGAGTATGATTAGGTGAAGATTTAGGATTTCTTCTTA
TGTTTCAAAAAGCCCCAAAGGATCTCTTAATCCAAACTGAATTCCCATCTGTGGGTTGAA
GCCAACTTTCTCCCATCTCACAAAGACTTCTCCGGTCTTCCTTCCAG***GTATTGTTGCAGA***
AGGCCGAGATGACCTCTATGTCTCAGATGCATTCCATAAGGCATTTCTTGAGGTGAGTAC
ACCTTCCCCACTCTCTTAGGGTACAGAAAGGAGATGCATGAACAGCAGGAACACGTGGAA
AAGGCCTGTTTCCAGTGTTAAGGCATGCAAAAGGCCTCCACAGGCCTGCTATAATACAGC
CCTCTCCAAAACCTTCATGGTGTGATTGTTCTGCCTTCCCTCCCACTACCTCTTCTGTAG
CAGGTCAAGCGGGAACACAAACATTTAGGGAGGGTGATATAGGAAAAGAAGCCAGCAAAG
GCCATCAAGAAGAAATTTACAGCATGAGGAGAACCAGAAGAGTATGGGGTCGCAGAAACC
CAGGGAGAATTTTTTTTTTTTTTTGAGACAGAGCTTCGTTCGCTCGTTGCCCAGGCTAGA
GTGCAATGGTGCGACAGCTCACTACAACTTCTGCCTCCCGCGTTCAAGCGATTCTCCTGC
CTCAGCCTCCTGGGTAGCTGGGATGACAGGCATGTGCCATCACGCCCGGCTAATTTTTGT
ATTTT

Figure 44: **Sequence of the mutated allele of P2.** Duplication is marked in red. Exon 6 is indicated in bold-italic font. Repetitive Alu sequences are underlined.

4.2.2 Identification of the first large intronic deletion responsible of type I antithrombin deficiency not detected by routine molecular diagnostic methods

a. Introduction

The identification of two cases with tandem duplication of exon 6 that are hardly identified by current methods, described in the previous chapter [119] encouraged the search of new gene defects affecting *SERPINC1* by alternative methods. LR-PCR combined with NGS is a new and efficient method for detecting intragenic mutations and gross gene defects that has facilitated studies in molecular genetics [138]. In this study we aim to explore LR-PCR of *SERPINC1* as an alternative method to identify the molecular base of cases with antithrombin deficiency with no associated genetic defect according to Sanger sequencing of the 7 exons or to MLPA, and without defect of glycosylation.

b. Analysis with LR-PCR of cases with antithrombin deficiency but no genetic defect.

The analysis of *SERPINC1* by LR-PCR was done in 36 cases with antithrombin deficiency but no genetic defect only revealed one case with an abnormal amplification affecting the second amplicon covering from exon 1 to exon 3 supporting a heterozygous 2Kbp deletion (Figure 45).

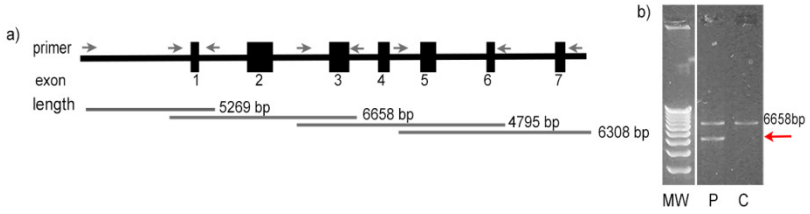


Figure 45: **Detection of a 2 Kbp deletion affecting intron 1 in *SERPINC1***. a) Schematic representation of the *SERPINC1* showing the long-range amplicons. b) Identification of a small amplification band (marked by a red arrow) in the second amplicon covering exons 1-3 of the proband (P) compared with the amplification of a control (C).

c. Clinical and analytical characteristics of the proband

The proband, a 56-year old woman suffered from deep vein thrombosis and pulmonary embolism at the age of 29 years. This thrombotic event occurred after a long car trip. The proband was smoker and was under oral contraceptives. The patient required cava vein filter. The severity of the event, the result of thrombophilic analysis (see below) and the severe family history of thrombosis (see below) recommended a lifelong treatment with acenocoumarol, which was successful, as no recurrences were recorded. Thrombophilic analysis performed in the proband, which included screening of antithrombin, protein C, protein S, deficiency, different tests of antiphospholipid antibodies, and genotyping of prothrombotic polymorphisms (FV Leiden and prothrombin G20210A), only revealed antithrombin deficiency (anti-FXa activity: 48%).
Her two sisters had neither thrombotic symptoms nor antithrombin deficiency. However, one brother with antithrombin deficiency died at the age of 28 as a consequence of pulmonary embolism (Figure 46). Her mother also died of thrombosis, although we have no further clinical or analytical data. There is also information of three maternal uncles of the proband: two had normal antithrombin levels and no thrombotic events, but one, with antithrombin deficiency, reported deep venous thrombosis during pregnancy. The proband had two siblings, the son had no antithrombin deficiency (anti-FXa: 112%), but the daughter, who is 37 years old, also had antithrombin

deficiency (anti-FXa: 50%) but no thrombotic event so far (Figure 46).

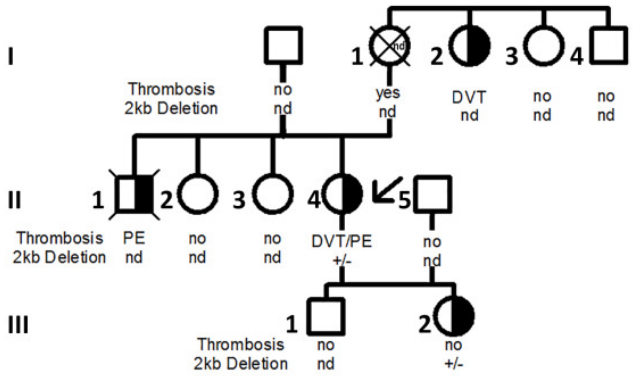


Figure 46: **Pedigree of the affected family.** The presence of thrombotic events is indicated. Relatives with antithrombin deficiency are represented with black semifilled symbols. Relatives with the 2 Kb deletion of intron 1 in heterozygosis are described as +/-. Relatives not available for molecular studies are indicated by nd (not determined). DVT: deep venous thrombosis. PE: pulmonary embolism. The proband is pointed by an arrow.

d. Characterization of antithrombin deficiency

Available samples from the proband and her daughter revealed that they had a type I antithrombin deficiency according to functional (anti-FXa: 48% and 50%, anti-FIIa: 67% and 40%, respectively) and antigenic values (51% and 52%, respectively). Moreover, analysis of plasma antithrombin by Western blot only confirmed significant reduction of levels, but without aberrant forms, or increase of the latent conformation (Figure 47)

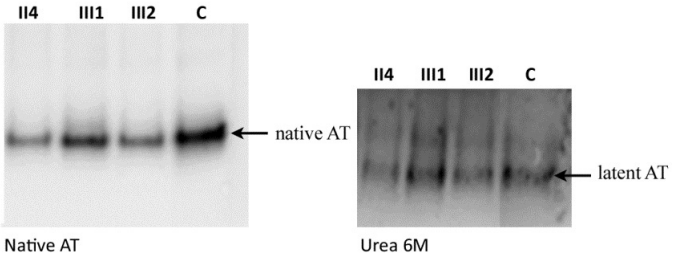


Figure 47: **Plasma antithrombin of three members of the affected family**; II-4 and III-2 with antithrombin deficiency, and III-1 with normal anti-FXa activity; and a healthy control (C) detected after electrophoretic separation on native (in the presence and absence of 6M urea) and detected by Western blot. The native and latent antithrombins are pointed by arrows.

e. Molecular characterization

As indicated before, PCR and Sanger sequencing of the 7 exons and flanking regions of *SERPINC1* in the proband revealed no significant gene defect. MLPA analysis also was normal (Figure 48). Whole gene sequencing using

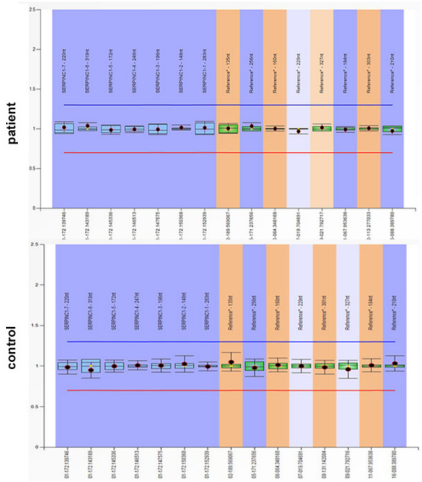


Figure 48: **Results of the MLPA analysis done in the proband and a healthy control with the SALSA *SERPINC1* kit.**

the Ion Torrent technology detected 3 intronic gene defects (c.1153+384_ 1153+386delTTA; c.624+185G>T and c.41+703_ 41+704delCT) but no one had significant pathogenic effect and none generates a cryptic splicing site according to *in silico* predictions.

LR-PCR amplification covering exon 1 to exon 3 revealed a potential 2 Kb deletion in heterozygosis (Figure 45). Analysis of the sequence of this amplicon using the CLC Genomics Workbench software shows a low number of reads covering 2,015 bp, 2,002 of them affecting the end of intron 1 and the first 13 nucleotides of exon 2 (Figure 49a). The analysis of SVs using InDels and Structural Variants tool from CLC software, showed the presence of a 1,941 bp deletion affecting only intron 1.

This intronic deletion was confirmed by specific PCR amplification using primers flanking the deletion (at intron 1 and intron 2) (Figure 49b). This specific amplification was the method used to test the presence of this deletion in the proband's relatives. As shown in Figure 49c, the 2 Kb deletion was detected in her asymptomatic daughter with antithrombin deficiency, but was absent in her son, who did not have antithrombin deficiency. Sanger sequencing of this PCR confirmed the large deletion but shows a relatively more complex rearrangement as it involved two breakpoints that caused the deletion of 1,941 bp of intron 1, the insertion of a CC dinucleotide, the presence of 34 bp of intron 1 and a new 40 bp deletion affecting the last 27 nucleotides of intron 1, including the acceptor splicing sequence and the first 13 nucleotides of exon 2(Figure 49d).

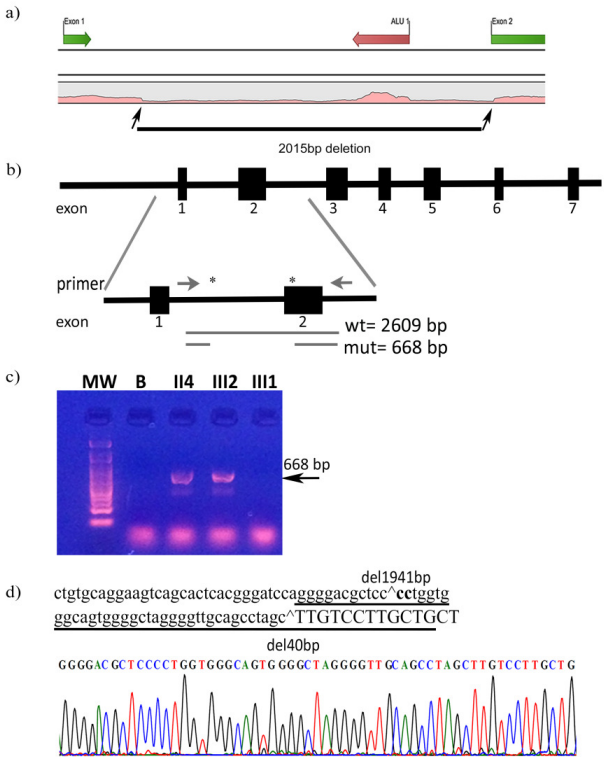


Figure 49: V**alidation and characterization of the intron 1 deletion.** a) Schematic representation of *SERPINC1* exon1-exon2 showing the reads of the long amplicon (Pink) generated by the CLC software. The deletion of 2015 bp identified is shown. The ALU sequence located in intron 1 is also shown. Bp: base pairs. b) Schematic representation of *SERPINC1* architecture showing the primers flanking the deleted region that were used to specifically amplify the mutated allele. The expected length of amplification of wild-type (wt) and mutated alleles (mut) are also shown. Asterisks show the localization of breakpoints. c) Specific PCR amplification of the deleted allele (668 bp) in the proband (II4) and her two siblings (III2 carrier and III1 not-carrier). The amplification in a tube with no DNA (B) was also used as a negative control. MW: 100 bp molecular weight marker. MW: molecular weight marker of 1 Kb. d) Sequence of the mutated allele showing the breakpoints, the deleted regions (∧symbol) and the inserted dinucleotide (marked in bold). Lower case is used for intronic sequence and exon sequence is marked in capital letter. The electropherogram of the underlined mutated sequence is also shown.

Interestingly, the CytoScan® HD Array detected no gross gene defects involving 1q25.1 (Table 12).

Table 12: **Gross gene defects detected in the proband by the CGH array.**

Type	Chromosome	Cytoband Start	Size (Kbp)	Gene Count	Genes
Loss	1	p31.1	52,945	0	
Loss	5	q21.2	96,488	0	
Loss	6	p22.1	67,788	2	HCG4B, HLA-A
Loss	7	p14.1	64,082	1	TRG-AS1
Loss	8	p11.22	139,855	2	ADAM5, ADAM3A
Loss	14	q11.2	236,639	0	
Gain	14	q32.33	279,313	6	MIR4507, MIR4538, MIR4537, MIR4539, KIAA0125, ADAM6
Gain	14	q32.33	69,797	0	
Loss	17	q21.31	85,675	1	LRRC37A4P
Gain	17	q21.31	63,795	2	KANSL1, KANSL1-AS1
Loss	20	p12.1	54,57	1	MACROD2
Gain	22	q11.22	97,368	1	MIR650
Gain	X	p11.23	136,081	5	SPACA5, SPACA5B, ZNF630-AS1, ZNF630, SSX6
Gain	X	q21.1	82,264	0	
Gain	X	q22.2	52,007	2	H2BFWT, H2BFM
Gain	X	p22.33	90,282	0	
Gain	Y	p11.2	236,416	0	
Gain	Y	p11.2	83,248	1	PCDH11Y
Gain	Y	p11.2	110,539	1	PCDH11Y

4.2.3 Long-read sequencing resolves structural variants in SERPINC1 causing antithrombin deficiency

a. Introduction

MLPA is the routine method used for SV detection in antithrombin deficiency, which are found in 5-8% of patients with this disorder [35]. Unfortunately, this method only shows information concerning the loss or gain of exons, but the exact extension of the genetic defect is not determined. Moreover, MLPA does give neither the breakpoints nor the sequence of the SVs and fails to detect intronic SVs, inversions or insertions. These genetic defects may underlie the molecular base of a proportion of the 25% of cases with antithrombin deficiency with unknown defect after analysis with current molecular methods. However, these limitations may now be addressed by third generation sequencing methods based on long reads, which can span repetitive or other problematic regions, allowing a full characterization of SVs as well as the identification of new SVs not detected by other methods [139, 108, 140, 141]. Thus, we performed long-read whole genome sequencing using nanopore technology to unravel SVs involved in antithrombin deficiency.

b. Patients and molecular methods

The study was done in 19 unrelated patients with antithrombin deficiency selected from our cohort of patients with this disorder. 8 of these patients had causal SVs identified by MLPA. In some of these patients whole genome analysis by CGH array was also done to determine the extension of the deletion. Nanopore sequencing aimed a full SV characterization including the investigation of the potential mechanisms of formation. The remaining 11 patients were selected because multiple independent genetic studies exons and flanking regions, whole *SERPINC1* sequencing by NGS (PGM and/or Miseq) and MLPA evaluating *SERPINC1* had failed to identify causal variants. (Table 13)

Table 13: **Cohort of individuals included in this study.** Results from genetic diagnostic methods correspond to findings involving *SERPINC1* gene. *SERPINC1* gene driven tests include MLPA, NGS using PGM (Ion Torrent) and Myseq (Illumina) sequencing, the last done after LR-PCR amplification. Genome wide tests are CGH array and Oxford Nanopore Technologies (ONT) sequencing. Coordinates have been confirmed by Sanger sequencing. Length refers to the extension of the SVs.CGHa: CGH array Het: Heterozygous; Ag: Antigen; bp: Base pairs.

Participant	antiFXa%	Ag (%)	Family history	Gender	*MLPA SERPINC1*	PGM Ion Torrent	CGHa	Illumina sequencing	ONT sequencing	Algorithm	Genotype	Coordinates	Length (bp)
P1	30	30	Yes	M	Deletion exon 1	-	Negative	Deletion exon 1	Deletion exon 1	nanosv; Sniffles;svim	Het	1:173916704-173935703	18999
P2	54	41	Yes	M	Deletion exon 1	-	Negative	Deletion exons 1,2	cxSV(Duplication exons 2,3-Deletion exon 1)	nanosv; sniffles	Het;Het	1:173911379-173915115; 1:173912151-173919034	3737;6884
P3	44	41	Yes	F	Complete deletion	-	Deletion 2 genes	-	Deletion 2 genes	nanosv; sniffles	Het	1:173879820-173925989	46169
P4	45	38	No	M	Complete deletion	-	Deletion 20 genes	-	Deletion 20 gene	nanosv; sniffles	Het	1:173847847-174816147	968005
P5	36	50	Yes	F	Complete deletion	-	-	-	Deletion 5 genes	nanosv	Het	1:173850996-173950174	99178
P6	61	46	Yes	M	Duplication exons 1,2 and 4; Deletion exon 6	-	Negative	Tandem duplication exons 1-5	Tandem Duplication exons 1-5	nanosv	Het	1:173908412-173919816	7045
P7	45	38	No	M	Deletion exons 1-5	-	Deletion exons 1-5 + 1 gene	-	Deletion 2 genes	nanosv; sniffles	Het	1:173908334-174103015	194389
P8	52	37	Yes	F	Deletion exons 2-5	-	-	-	Deletion exons 2-5	nanosv; sniffles	Het	1:173908218-173915405	7045
P9	56	61	Yes	F	Negative	Negative	-	Negative	Insertion SVA	nanosv	Het	1:173905922	2440
P10	50	46	Yes	F	Negative	Negative	-	Negative	Insertion SVA	visual inspection	Het	1:173905922	2440
P11	40	41	Yes	F	Negative	Negative	-	Negative	Negative				
P12	73	62	No	F	Negative	Negative	-	Negative	Negative				
P13	59	51	No	F	Negative	Negative	-	Negative	Negative				
P14	63	58	No	M	Negative	-	-	Negative	Negative				
P15	69	NA	No	F	Negative	Negative	-	Negative	Negative				
P16	56	45	Yes	F	Negative	-	-	-	Negative				
P17	68	54	No	M	Negative	Negative	-	Negative	Negative				
P18	66	67	No	M	Negative	Negative	-	Negative	Negative				
P19	67	70	No	F	Negative	-	-	-	Negative				
P20	50	70	Yes	M	Negative	-	-	-	Negative				

c. Long-read nanopore sequencing resolves structural variants involved in antithrombin deficiency

Nanopore sequencing in 21 runs produced reads with an average length of 4,499bp and median genome coverage of 16x (Figure 50 A-B). After a detailed quality control analysis (Figure51) 83,486 SVs were identified, consistent with previous reports using LR-WGS (Figure 52) [108].

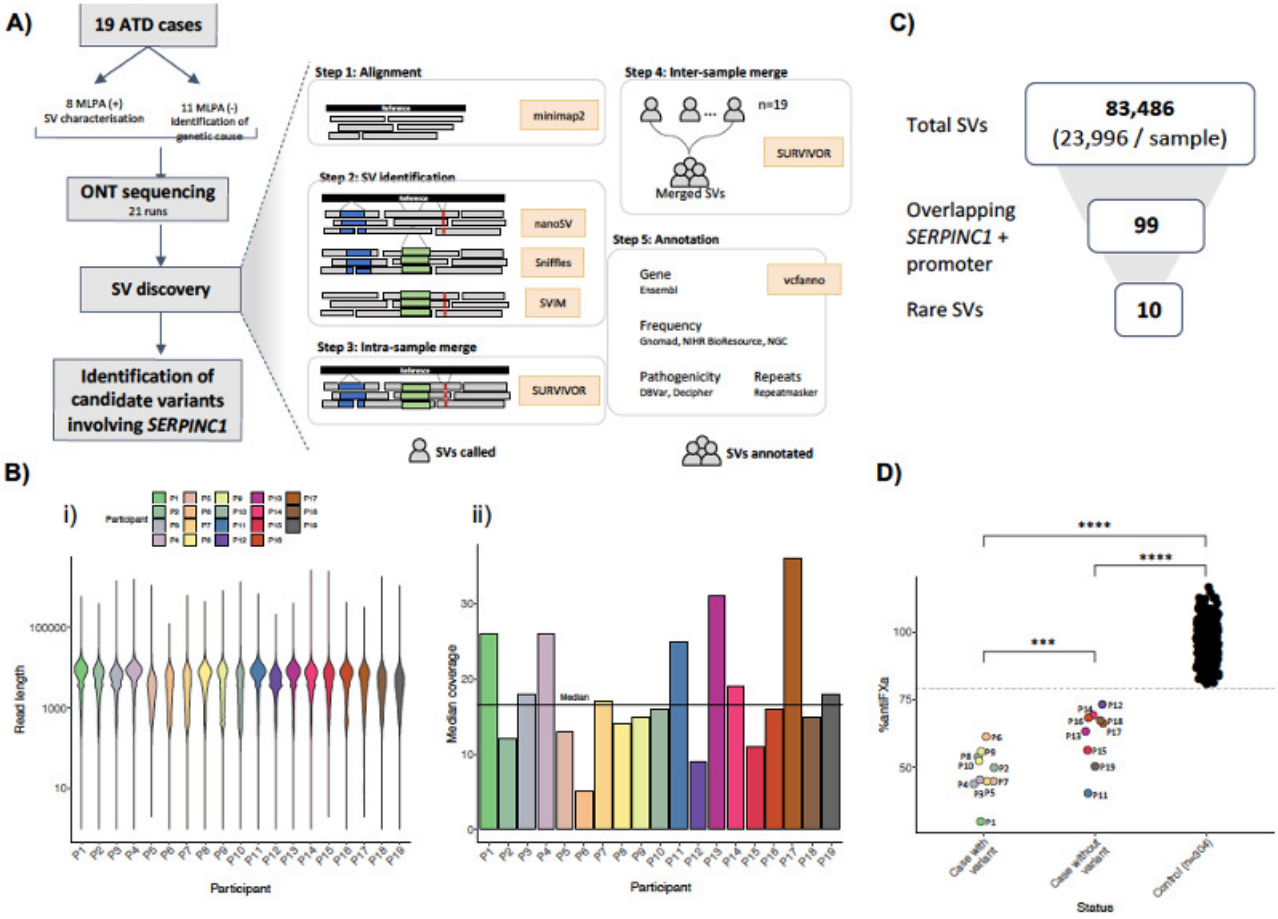


Figure 50: **Long-read nanopore sequencing workflow and results.** A) Overview of the general stages of the SVs Discovery workflow. Algorithms used are depicted in yellow boxes. B) Nanopore sequencing results. i) Sequence length template distribution. Average read length was 4,499 bp (sd $\pm$ 4,268); the maximum read length observed was 2.5Mb. ii) Genome median coverage per participant. The average across all samples was 16x (sd $\pm$ 7.7). C) Filtering approach and number of SVs obtained per step. *SERPINC1* + promoter region corresponds to [GRCh38/hg38] Chr1:173,903,500-173,931,500. D) anti-FXa percentage levels for the participants with a variant identified (P1-P10), cases without a candidate variant (P11-P19) and 300 controls from our internal database. The statistical significance is denoted by asterisks ($*$), where $***$ P$<$ 0.001, $****$ P$\leqslant$ 0.0001. p-values calculated by oneway ANOVA with Tukey's post-hoc test for repeated measures. ATD: Antithrombin Deficiency; ONT: Oxford Nanopore Technologies; SV: Structural Variant.

Focusing on rare variants (allele count $\leqq$ 10 in gnomAD v3, NIHR BioResource and NGC project)[141, 142, 143] in *SERPINC1* and flanking regions, 10 candidate heterozygous SVs were observed in 9 individuals (Figure 50C). Visual inspection of read alignments identified an additional heterozygous SV in a region of low coverage involving

SERPINC1.

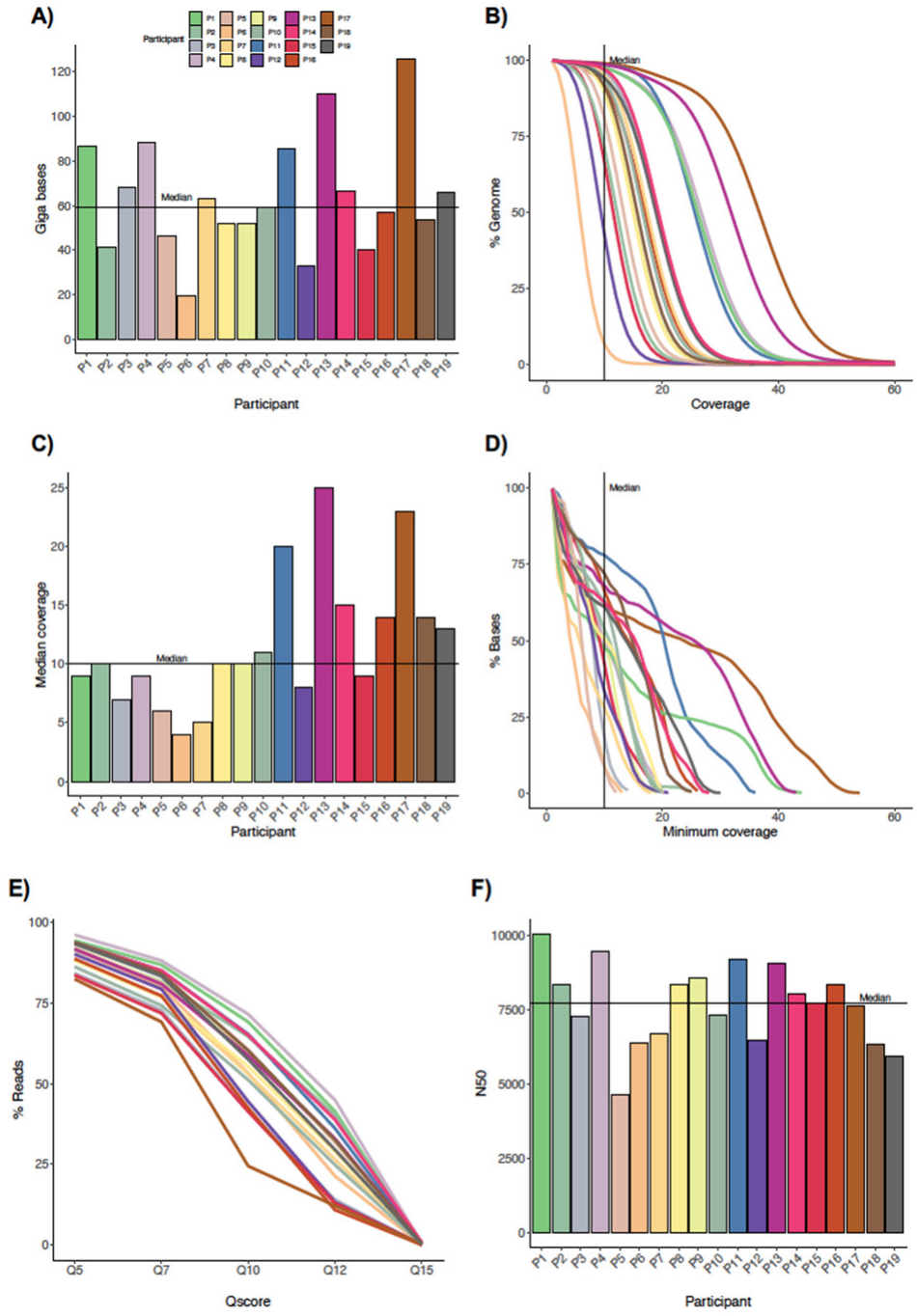


Figure 51: **Sequencing results colored by participant.** A) Giga bases sequenced B) Percentage of bases in genome sequenced at a specific minimum coverage C) Median coverage in *SERPINC1* + promoter region D) Coverage distribution in *SERPINC1* + promoter region E) Percentage of reads with a minimum Q score F) Read N50, which refers to a value where half of the data is contained within reads with alignable lengths greater than this.

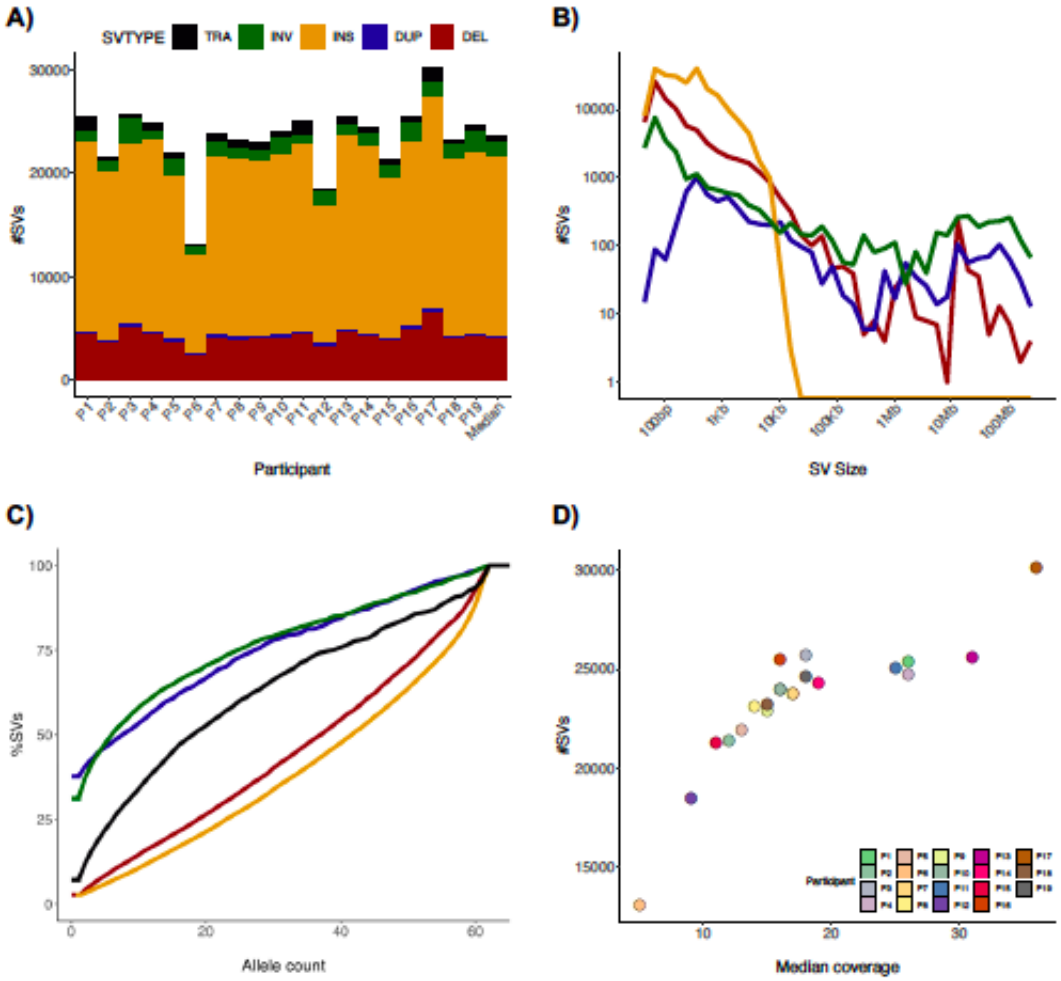


Figure 52: **Structural variants metrics.** A) Number of SVs identified by type and participant B) Number of SVs by SV size C) Fraction of SVs per allele count in our internal cohort of 62 individuals with long-read sequencing data D) Number of SVs by median coverage and participant.

d. Complete characterization of SVs involved in antithrombin deficiency

Nanopore sequencing resolved the precise configuration of SVs previously identified by MLPA in all 8 selected cases (P1-P8). SVs were identified independently of their size (from 7Kb to 968Kb, restricted to *SERPINC1* or involving neighbouring genes) and their type (six deletions, one tandem duplication and one complex SV) (Figure 53, Table 13).

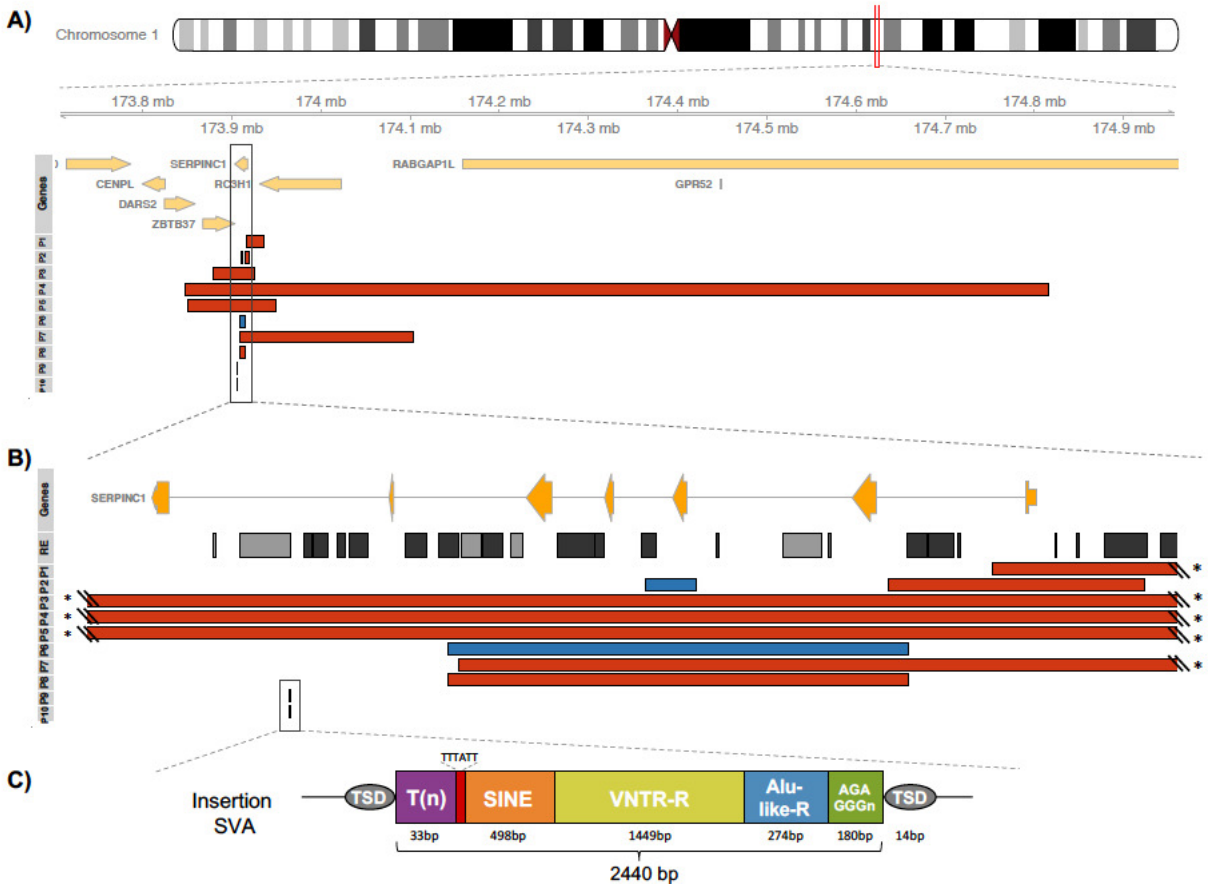


Figure 53: **Candidate SVs identified by long-read nanopore sequencing.** A) Schematic of chromosome 1 followed by protein coding genes falling in the zoomed region (1q25.1). SVs for each participant (P) are coloured in red (deletions) and blue (duplications). The insertion identified in P9 and P10 is shown with a black line. B) Schematic of *SERPINC1* gene (NM_000488) followed by repetitive elements (RE) in the region. SINEs and LINEs are coloured in light and dark grey respectively. Asterisks are present where the corresponding breakpoint falls within a RE. C) Characteristics of the antisense-oriented SVA retroelement (respect to the canonical sequence) observed in P9. Lengths of the fragments are subject to errors from nanopore sequencing. TSD:Target site duplication.

Nanopore sequencing results were fully concordant with MLPA and CGH array results in 6 cases, but, importantly, this method was able to solve the SVs responsible for antithrombin deficiency in 2 cases with previous inconsistent or ambiguous results (P2 and P6) (Table 13).

For the first case (P2) MLPA detected a deletion of exon 1, but LR-PCR followed by NGS suggested a deletion of exons 1 and 2. The discordant results were explained by Nanopore sequencing, as this method revealed a complex SV in *SERPINC1* resulting in a dispersed duplication of exons 2 and 3 and a deletion spanning exons 1 and 2, both in the same allele (Figure 54). Specific PCR amplification and Sanger sequencing validated this complex SV in the proband and a daughter also with antithrombin deficiency.

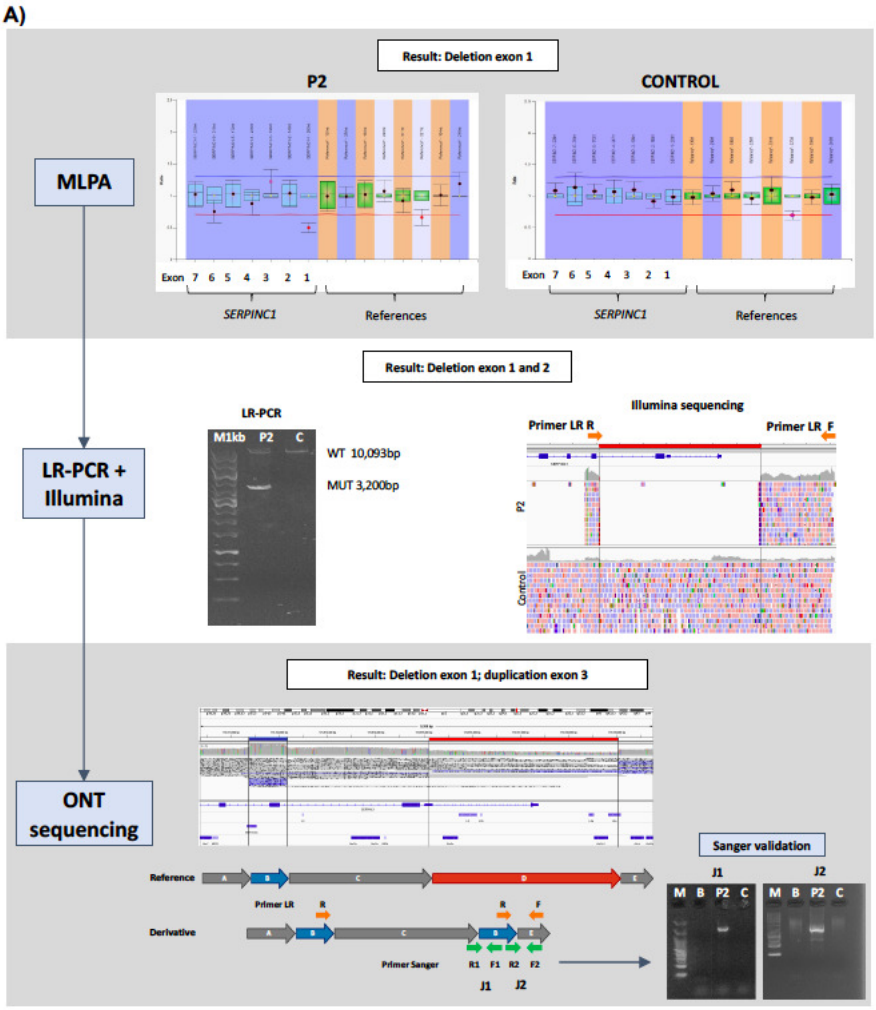


Figure 54: **SV resolution of P2.** Schematic representation of genetic diagnostic methods used to characterize the SVs. Results from MLPA, LR-PCR and nanopore are shown in white boxes. Primers used for both LR-PCR and Sanger validation experiments are shown with orange and green arrows respectively. J1 and J2 correspond to the new formed junctions. J: New junction; M: Molecular weight marker; P: patient; C: control; B: Blank.

For the second case (P6) MLPA detected a duplication of exons 1, 2 and 4 and a deletion of exon 6. Here, our sequencing approach identified a tandem duplication of exons 1 to 5, which was confirmed by l LR-PCR (Figure 55). The tandem duplication of exons 1-5 was also observed to be present in the son of P6, also with antithrombin deficiency.

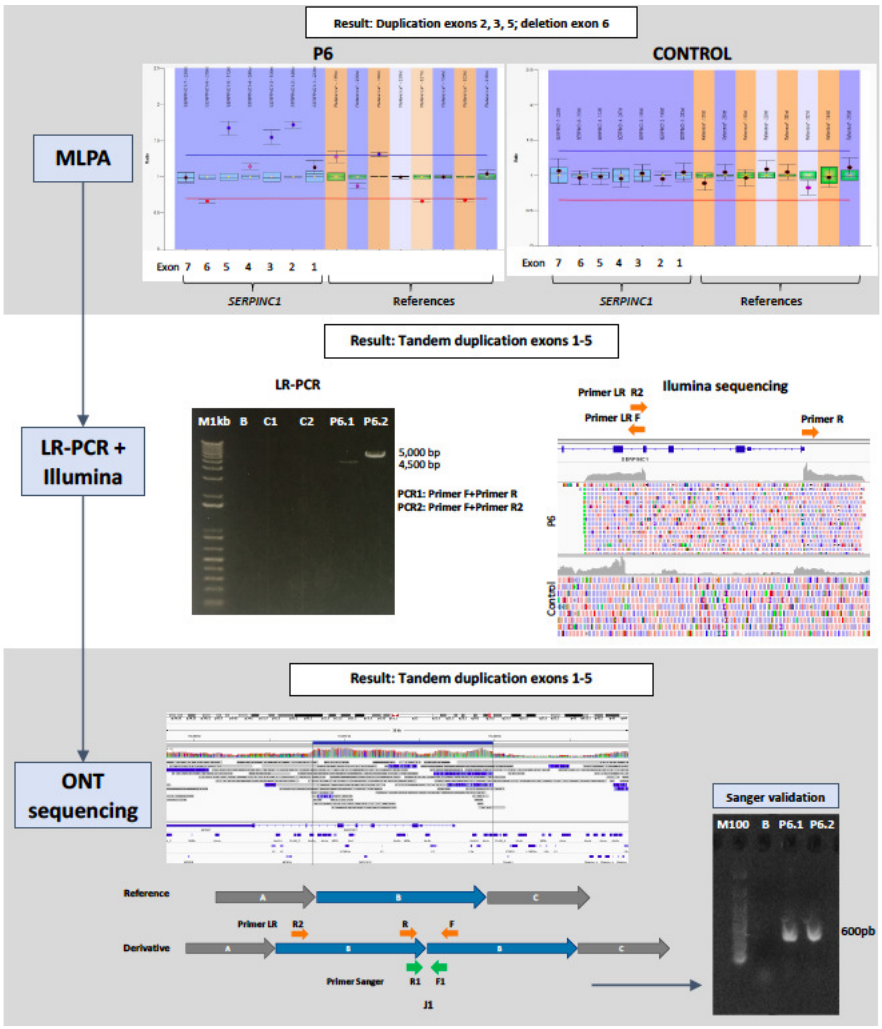


Figure 55: **SV resolution of P6.** Schematic representation of genetic diagnostic methods used to characterize the SV. Results from MLPA, LR-PCR and nanopore are shown in white boxes. Primers used for both LR-PCR and Sanger validation experiments are shown with orange and green arrows respectively. J1 correspond to the new formed junction. M100: Molecular weight marker; For the LR-PCR results, C1 and P1 correspond to PCR 1 (done with Primer F + Primer R), and C2 and P2 correspond to PCR2 (done with Primer F + Primer R2).

e. A SVA retroelement insertion is identified in two previously unresolved cases and characterized by de novo assembly

We aimed to identify new disease-causing variants in the remaining 11 participants with negative results using current molecular methods. Remarkably, two cases (P9 and P10) presented an insertion of 2,440 bp in intron 6. Blast analysis of the inserted sequence revealed a new SINE-VNTR-Alu (SVA) retroelement (Figure 53 , Table 13). De novo assembly using the data from P9 revealed an antisense-oriented SVA element flanked by a target site duplication (TSD) of 14bp (Figure 53C), consistent with a target-primed reverse transcription mechanism of insertion into the genome [144, 145]. Interestingly, the TSD in both individuals was also the same. The inserted sequence was aligned to the canonical SVA A-F sequences (Figure 56A) and it was observed to be closest to the SVA E in the phylogenetic tree (Figure 56B).

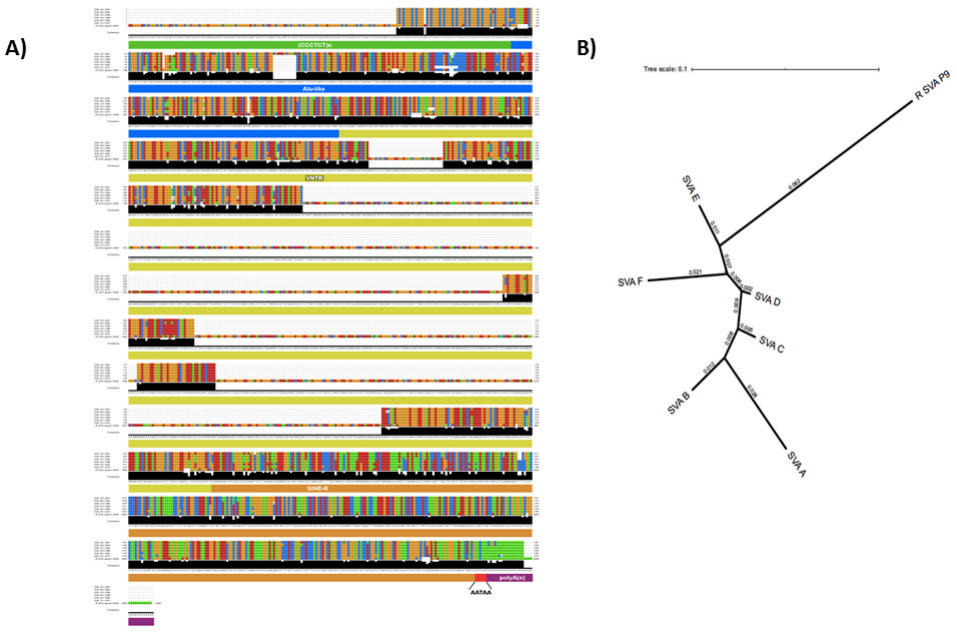


Figure 56: **SVA sequence alignments.** A) The consensus sequences of SVA A, B, C, D, E, and F were taken from RepeatMasker (www.repeatmasker.org), then aligned using MAFFT with default parameters. SVA_ query corresponds to the SVA insertion in P9. Alignments were visually inspected and colored by nucleotide using JalView. Subelements of the SVAs are indicated underneath the consensus sequence matching colors in Figure 53C. B) A phylogenetic tree was constructed with the Neighbour-Joining (NJ) algorithm using the Jukes-Cantor substitution model and visualized with iTol. The SVA insertion in P9 was observed to be closest to the SVA E in the phylogenetic tree.

Moreover, the VNTR sub-element harbored 1,449bp, which was longer than the typical ∽520bp-long VNTR in the canonical sequences. Multiple PCRs covering the retroelement were attempted to validate this insertion, but all PRCs using flanking primers failed. Only one specific PCR using an internal SVA primer, also designed with the sequence obtained from Nanopore, was able to amplify the breakpoint (Figure 57). This method was used to confirm the insertion in P9, P10 and the sequence of the retrotransposon and the breakpoint was validated by Sanger sequencing. This specific PCR amplification was also the method used to confirm the Mendelian inheritance of this SVA, as it was also present in two relatives, both with antithrombin deficiency (Figure 57).

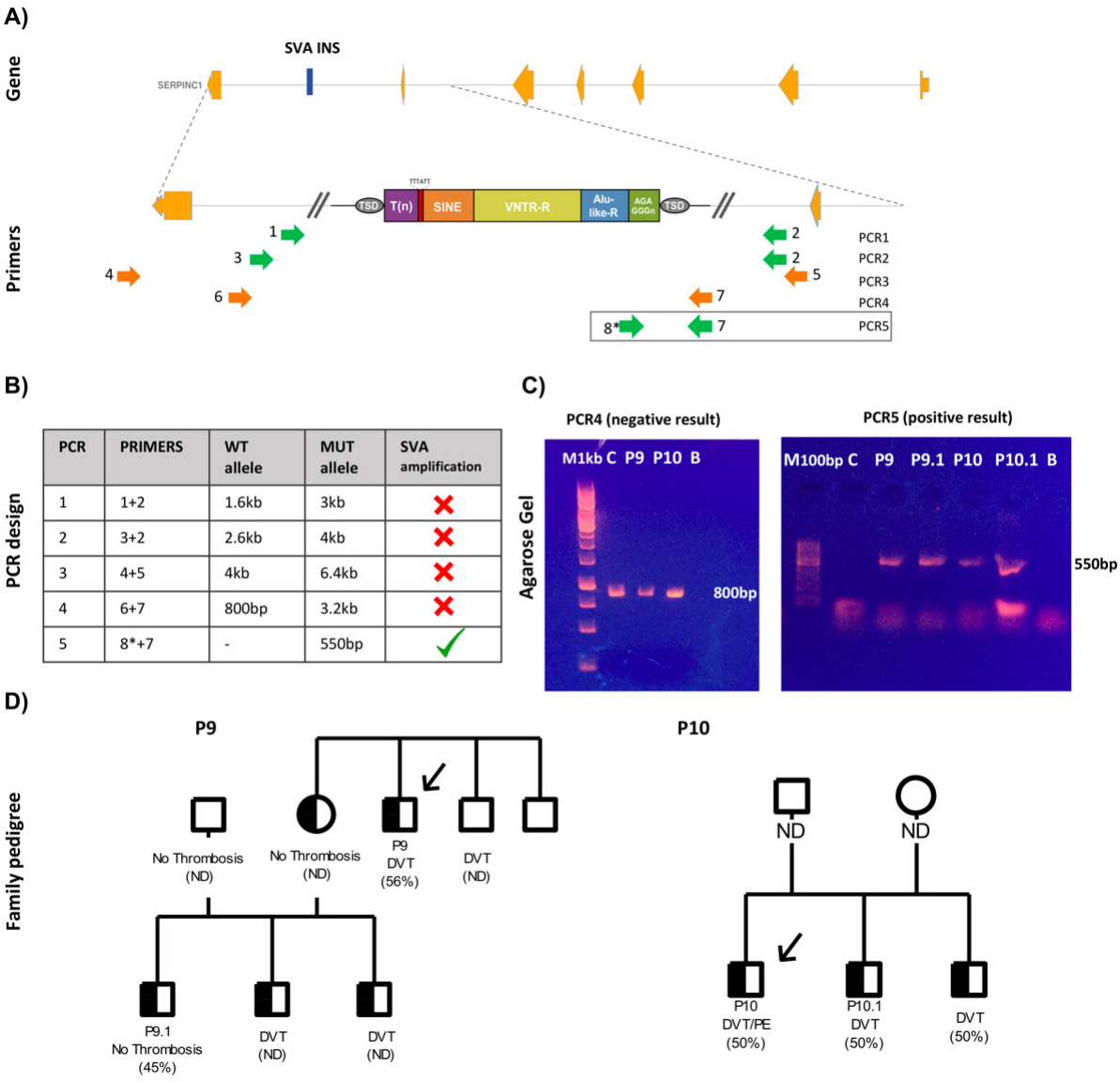


Figure 57: **PCR amplification validation of the SVA insertion in P9 and P10.** A) Schematic of *SERPINC1* gene (NM_ 000488) with zoom to intron 6 showing the SVA structure. Primers used in the long- and short-range PCRs are shown in orange and green respectively. Primer 8∗ was specifically designed within the inserted SVA sequence. Briefly, four reads of the retrotransposon present in the nanopore data for P9 and P10 were aligned to identify regions without any mismatch in order to select a 20 nucleotide sequence to be used as primer. That sequence was also checked to be present in de novo assembly alignment. B) Primer combinations for PCR amplifications and expected sizes for wild type and mutated alleles are shown in the table. PCRs 1-4 were tested under different experimental conditions, and although in all cases the wild type allele was always amplified, no amplification of the mutated allele containing the SVA was obtained in P9 or P10. C) The amplification of PCR 4 in agarose gel is shown. Only the 800 bp of the wild type allele was amplified in P9, P10 but also a healthy control. Only PCR 5, using the primer specific of the SVA rendered positive results and a specific 550bp band was obtained in P9, P10 and two relatives. D) Family pedigrees of P9 and P10, including clinical information, the diagnosis of antithrombin deficiency (semi-filled symbols) and the anti-FX activity (as % of a reference plasma). B: Blank; M: Molecular Weight Marker; DVT: Deep vein thrombosis; PE: Pulmonary embolism

f. Breakpoint analysis supports a replication-based mechanism for the majority of structural variants.

Breakpoint analysis was performed to investigate the mechanism underlying the formation of these SVs involving *SERPINC1*. Nanopore sequencing facilitated primer design to perform Sanger sequencing confirmations for all the new formed junctions, demonstrating 100% accuracy in 7/10 (70%) SVs called. Repetitive elements were detected in all the SVs, with Alu elements being the most frequent (16/24, 67%) (Table 14).

Table 14: **Repetitive elements at the SV breakpoints.** Segments column match to those in Figure S6, and refers to the relative location in the segments where the SV breakpoint is, where 5' and 3' correspond to the upstream and downstream parts respectively. Window was done for ± 150 bp from the breakpoint. Start and End in query are the relative positions to the query sequence where the repeat starts and ends respectively. Bkp: Breakpoint.

Participant	Bkp in SV	Segments	Window	Repeat coordinates	Repeat	Family	Start in query	End in query	Repeat length
P1	Deletion start	A3'-B5'	1:173916554-173916854	-					
	Deletion end	B3'-C5'	1:173935553-173935853	1:173935699-173935738	A-rich	Low_complexity	147	186	40
P2	Duplication start	A3'-B5'	1:173911229-173911529	1:173911329-173911536	MER103C	DNA/hAT-Charlie	101	300	201
	Duplication end	B3'-C5'	1:173912001-173912301	-					
	Deletion start	C3'-D5'	1:173914965-173915265	-					
	Deletion end	D3'-E5'	1:173918884-173919184	1:173918822-173919087	L2a	LINE/L2	1	204	204
P3	Deletion start	A3'-B5'	1:173879670-173879970	1:173879670-173879957	AluSx1	SINE/Alu			288
	Deletion end	B3'-C5'	1:173925839-173926139	1:173926103-173926323	AluJb	SINE/Alu	265	300	37
P4	Deletion start	A3'-B5'	1:173847697-173847997	1:173847804-173847844	HAL1	LINE/L1	108	148	41
				1:173847845-173848152	AluY	SINE/Alu	149	300	153
	Deletion end	B3'-C5'	1:174815997-174816297	1:174816148-174816445	AluSp	SINE/Alu	152	300	150
P5	Deletion start	A3'-B5'	1:173850846-173851146	1:173850956-173851264	AluSx1	SINE/Alu	111	300	191
				1:173850606-173850904	AluSx	SINE/Alu	1	59	59
	Deletion end	B3'-C5'	1:173950024-173950324	1:173949885-173950176	AluSx	SINE/Alu	1	153	153
				1:173950177-173950297	L1ME3C	LINE/L1	154	121	121
				1:173950298-173950593	AluJr	SINE/Alu	275	300	27
P6	Duplication start	A3'-B5'	1:173908262-173908562	1:173908214-173908512	AluSx1	SINE/Alu	1	251	251
				1:173908555-173908859	AluJb	SINE/Alu	294	300	8
	Duplication end	B3'-C5'	1:173919640-173919940	1:173919609-173919890	AluSz	SINE/Alu	1	251	251
				1:173919924-173920236	AluSx1	SINE/Alu	285	300	17
P7	Deletion start	A3'-B5'	1:173908184-173908484	1:173908214-173908512	AluSx1	SINE/Alu	31	300	271
	Deletion end	B3'-C5'	1:174102865-174103165	1:174102793-174102889	Tigger1	DNA/TcMar-Tigger	1	25	25
				1:174102890-174103191	AluSp	SINE/Alu	26	300	276
P8	Deletion start	A3'-B5'	1:173908068-173908368	1:173908214-173908512	AluSx1	SINE/Alu	147	300	155
	Deletion end	B3'-C5'	1:173915255-173915555	1:173915394-173915705	AluSz	SINE/Alu	140	300	162
P9	Insertion site		1:173905771-173906071	1:173905736-173905941	L2	LINE/L2	1	171	171
P10	Insertion site		1:173905771-173906071	1:173905736-173905941	L2	LINE/L2	1	171	171

Additionally, breakpoint analysis identified microhomologies (7/11, 64%) and insertions, deletions or duplications (7/11, 64%) (Figure 58)

Specific mutational signatures can yield insights into the mechanisms by which the SVs are formed. Importantly, we observed a non-random formation driven by the presence of repetitive elements in some of the SVs. For example, an Alu element in intron 1 was involved in the SVs of P6 and P8, and an Alu element in intron 5 was involved in SVs of P6, P7 and P8 (Figure 53B, Table 15 and 16). It has been suggested that repetitive elements may provide larger tracks of microhomologies, also termed "microhomology islands", that could assist strand transfer or stimulate template switching during repair by a replication-based mechanism [69]. These microhomology islands were present in the SVs of 4 cases (P4, P6-P8), highlighting the important role that repetitive elements play in the formation of non-recurrent, but non-random, SVs.

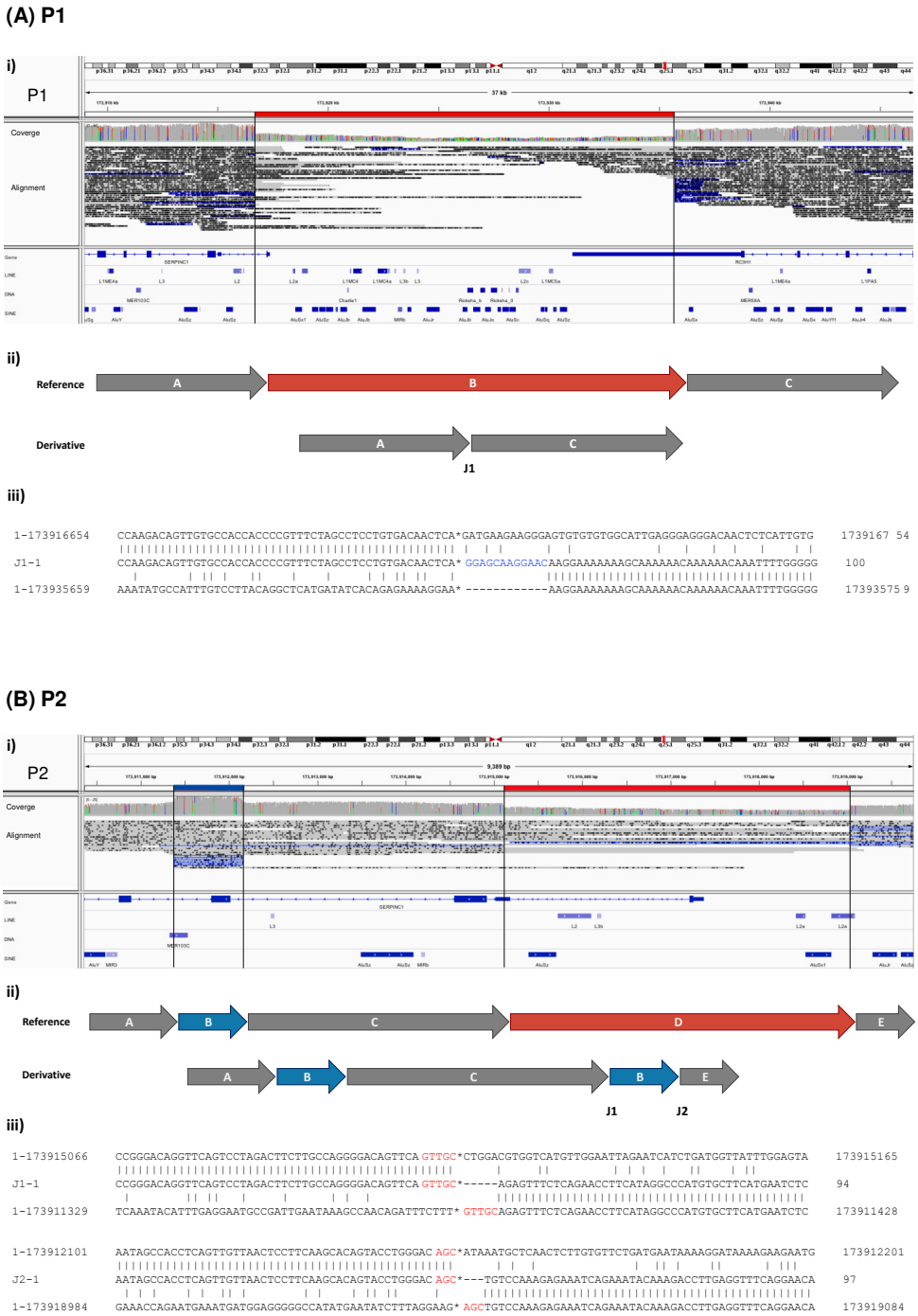
Table 15: **Structural variants breakpoint details.** New junction numbers correspond to Figure 58. Microhomology sequences (from 3 bp to 36 bp) at breakpoints and inserted sequences are shown MH: Microhomology

Parti	New junction	Microhomology at breakpoint	MH at break-point length	Deletion at break-point	Deletion at break-point length	Insertion at break-point	Insertion at break-point length	Duplication at break-point	Duplication at break-point length
P1	J1	-	-	-	-	GGAGCAAGGAAC	12	-	-
P2	J1	GTTGC	5	GTTGC	5	-	-	-	-
	J2	AGC	3	AGC	3	-	-	-	-
P3	J1	TGGN	3	-	-	GCC	3	-	-
P4	J1	TTTTTTTTTTTTTTTTTTTTTTTGAGA	26	TTT	3	-	-	-	-
P5	J1	-	-	-	-	-	-	-	-
P6	J1	GGAGGCTGAGGCAGGAGAATCACTTGAA	28	-	-	-	-	-	-
P7	J1	CAAAAATTAGCC	12	-	-	-	-	-	-
P8	J1	GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC	36	-	-	-	-	-	-
P9	J1	-	-	-	-	-	-	ACCTTGTGTTCTTT	14
P10	J1	-	-	-	-	-	-	ACCTTGTGTTCTTT	14

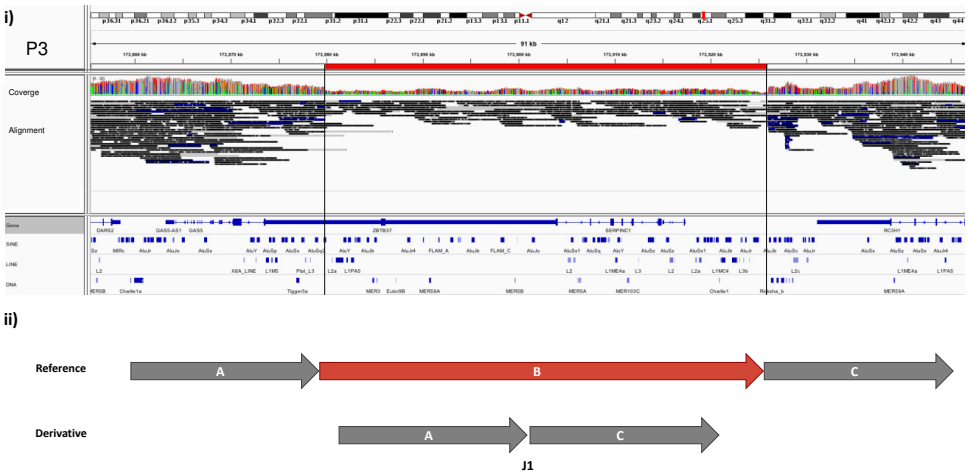
Table 16: **Detailed sequences of breakpoints of SV in cases with antithrombin deficiency shown in table 15** Expected reference sequence refers to the putative derivative sequence as shown in Figure 58. Microhomology sequences (from 3 bp to 36 bp) at breakpoints are highlighted in red, inserted sequences are in blue, and duplicated fragments are in bold. Breakpoint is marked with *.

Participant	New junction	Expected reference sequence	Observed sequence (Sanger)
P1	J1	CCAAGACAGTTGTGCCACCACCCCGTTTCTAGCCTCCTGTGACAACTCA*AAGGAAAAGGAAAAAAGCAAAAAACAAAAAACAAATTTTGGGGGTGTTAG	CCAAGACAGTTGTGCCACCACCCCGTTTCTAGCCTCCTGTGACAACTCA*{GGAGCAAGGAAC}AAGGAAAAAAGCAAAAAACAAAAAACAAATTTTGGGGGTGTTAG
P2	J1	CCCGGGACAGGTTCAGTCCTAGACTTCTTGCCAGGGGACAGTTCAGTTGC*GTTGCAGAGTTTCTCAGAACCTTCATAGGCCCATGTGCTTCATGAATCTCC	CCCGGGACAGGTTCAGTCCTAGACTTCTTGCCAGGGGACAGTTCAGTTGC*-----AGAGTTTCTCAGAACCTTCATAGGCCCATGTGCTTCATGAATCTCC
	J2	AATAGCCACCTCAGTTGTTAACTCCTTCAAGCACAGTACCTGGGACAGC*AGCTGTCCAAAGAGAACTCAGAAATACAAAGACCTTGAGGTTTCAGGAACA	AATAGCCACCTCAGTTGTTAACTCCTTCAAGCACAGTACCTGGGACAGC*---TGTCCAAAGAGAAATCAGAAATACAAAGACCTTGAGGTTTCAGGAACA
P3	J1	GTGAAACCCCATCTCTACTAAATATAAAACATTAGCCAGGCATGGTGGT*TGGACAAGGCTATTCTTTGGAGGCTGAGCAAGAGAGAGTTGTGTGGGGTCA	GTGAAACCCCATCTCTACTAAATATAAAACATTAGCCAGGCATGGTGGT{GCC}*TGGACAAGGCTATTCTTTGGAGGCTGAGCAAGAGAGAGTTGTGTGGGGTCA
P4	J1	CCTCACCTAGTTACTCTGTACCTTCTGCACAGCTGTCTTTCTGAACCTTT*TTTTTTTTTTTTTTTTTTTTTTTGAGATAGAGTTTTGCTCTTGTTGCCCAG	CCTCACCTAGTTACTCTGTACCTTCTGCACAGCTGTCTTTCTGAACC---*TTTTTTTTTTTTTTTTTTTTTTTGAGATAGAGTTTTGCTCTTGTTGCCCAG
P5	J1	CTTAAGGGTTGGCCGGGCATGGTGGCTAACGCCTGTAATCCCAGCACTTT*AAAGTACACATCTTAAAAGCAATGCTAAGTGAAAACATGAAAATGCAAGCT	CTTAAGGGTTGGCCGGGCATGGTGGCTAACGCCTGTAATCCCAGCACTTT*AAGTACACATCTTAAAAGCAATGCTAAGTGAAAACATGAAAATGCAAGCT
P6	J1	ACCTGTAATCTCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAA*TCTGGGGGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCAC	GCCTGTAATCTCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAA*TCTGGGGGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCAC
P7	J1	TTCAAGACCAGCCTGGCCAACATAGTGAAACCCCGTTTCTACTAAAAAT*ACAAAAATTAGCCAGGCATGGTGGTGAGCGCCTGTAATCCCAGCTACTCCA	TTCAAGACCAGCCTGGCCAACATAGTGAAACCCCGTTTCTACTAAAAAT*{A}CAAAAATTAGCCAGGCATGGTGGTGAGCGCCTGTAATCCCAGCTACTCCA
P8	J1	GGCAACATTTTTAGTTTTCAAAACTAGCTTAAAGACAGCCCAATAGCACA*GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCATGCAGA	GGCAACATTTTTAGTTTTCAAACTAGCTTAAAGACAGCCCAATAGCACA*GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCATGCAGA
P9	J1	AAATTATATTAATAGCATCCTTTTCTCAGATTATAACCTTGTGTTCTTT*CAAGCTATCACATTTCCTCGCTCCCGTTAAAATTCCACTTTCCTCTG	AAATTATATTAATAGCATCCTTTTCTCAGATTATAACCTTGTGTTCTTT*{INS_SVA}**ACCTTGTGTTCTTT**CAAGCTATCACTTTTCCCTGCTCCCGTTAAA-TTCCACTTTCCTCTG
P10	J1	AAATTATATTAATAGCATCCTTTTCTCAGATTATAACCTTGTGTTCTTT*CAAGCTATCACATTTCCTCGCTCCCGTTAAAATTCCACTTTCCTCTG	AAATTATATTAATAGCATCCTTTTCTCAGATTATAACCTTGTGTTCTTT*{INS_SVA}**ACCTTGTGTTCTTT**CAAGCTATCACTTTTCCCTGCTCCCGTTAAA-TTCCACTTTCCTCTG

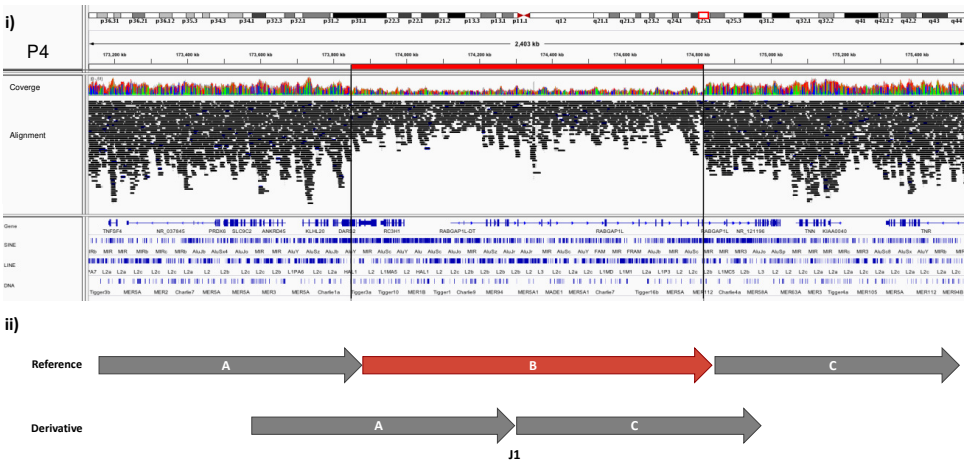
Figure 58: **Nucleotide level characterization of the candidate SVs.** Breakpoint junction sequence is aligned to the proximal and distal genomic reference sequence. Alignment is only shown for novel breakpoint junctions in the derivative chromosome. Microhomology at the breakpoint is indicated in red. Sequence in blue indicates inserted sequences at the breakpoint junction. Underline indicates repetitive elements in the reference, specified in Italic. J: Junction. (AH)P1-8, (I) P9 and P10.



(C) P3

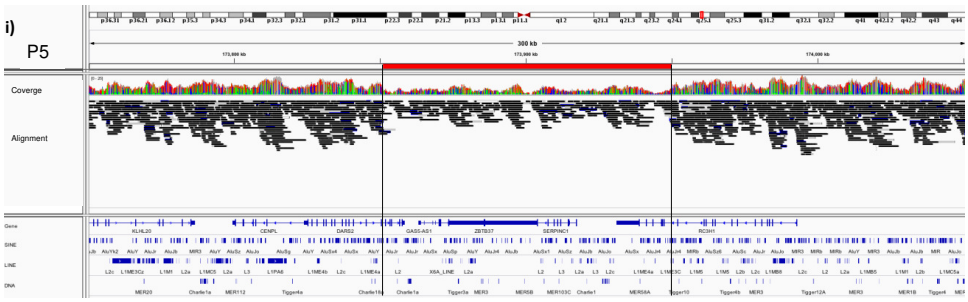

i)
P3
Coverge
Alignment
ii)
Reference
A
B
C
Derivative
A
C
J1
iii)
1-173879770 GTGAAACCCCATCTCTACTAAATATAAAACATTAGCCAGGCATGG TGGT*GGGCACCTGTAATCTCAGCTACTCGGGAGGCTGCAACAGGAGAATCACTTG 173879870
J1-1 GTGAAACCCCATCTCTACTAAATATAAAACATTAGCCAGGCATGG TGGT*GCCTGGACAAGGCTATTCTTTGGAGGCTGAGCAAGAGAGAGTTGTGTGGGG 100
1-173925939 AAATGCTTACTCATCCAGAAGAATTGACACCTTGTATGAACAGAAACAT* ---TGGACAAGGCTATTCTTTGGAGGCTGAGCAAGAGAGAGTTGTGTGGGG 173926039

(D) P4

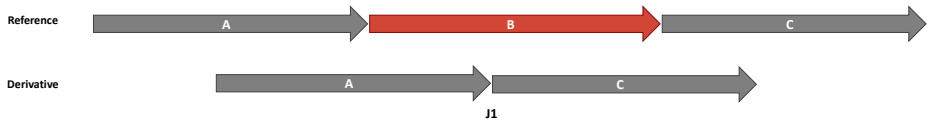

i)
P4
Coverge
Alignment
ii)
Reference
A
B
C
Derivative
A
C
J1
iii)
1-173847797 CCTCACCTAGTTACTCTGTACCTTCTGCACAGCTGTCTTTCTGAACCTTT* TTTTTTTTTTTTTTTTTTTTTTTGAGACGGGAGTCTCACTCTGTCGCCCAG 173847897
J1-1 CCTCACCTAGTTACTCTGTACCTTCTGCACAGCTGTCTTTCTGAACC ---*TTTTTTTTTTTTTTTTTTTTTTTGAGATAGAGTTTTGCTCTTGTTGCCCAG 97
1-174816097 TTGTATGAACATATGAGCATTTTGTTTTTCTTAGTAACATAATACATAGC* TTTTTTTTTTTTTTTTTTTTTTTGAGATAGAGTTTTGCTCTTGTTGCCCAG 174816197

(E) P5

i)



ii)

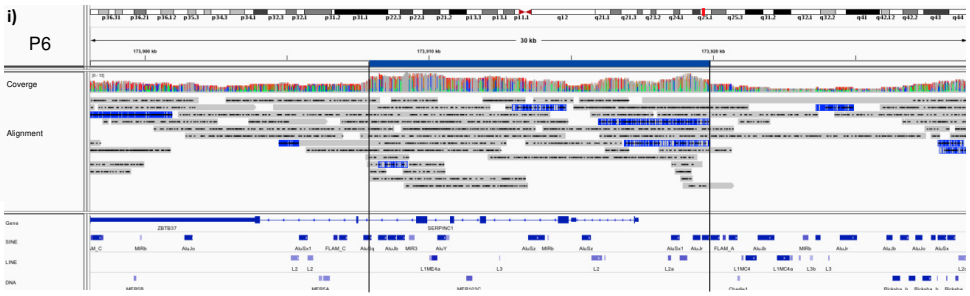


iii)

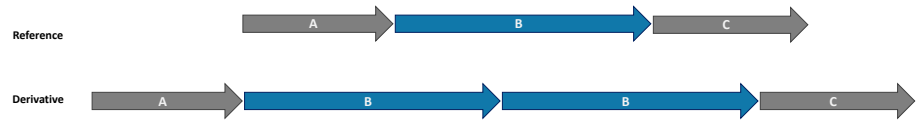
```
1-173850946   TTAAGGGTTGGCCGGGCATGGTGGCTAACGCCTGTAATCCCAGCACTTT*GGGAGGCCGAGGTGGGTGGATCGCCTGAGGTCAGGAGTTTGAGACCAGCCT   1738510 46
                                                                |  |  |     |  ||  |  |  |  ||  ||
J1-1          TTAAGGGTTGGCCGGGCATGGTGGCTAACGCCTGTAATCCCAGCACTTT* -AAGTACACATCTTAAAAGCAATGCTAAGTGAAAACATGAAAATGCAAGCT    99
                                                                |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1-173950124   TGCCACTGCACTCCAGCCTGGGTACAGAGTGGAACTCCGTCTCAAAAAA*AAAGTACACATCTTAAAAGCAATGCTAAGTGAAAACATGAAAATGCAAGCT   1739502 24
```

(F) P6

i)



ii)

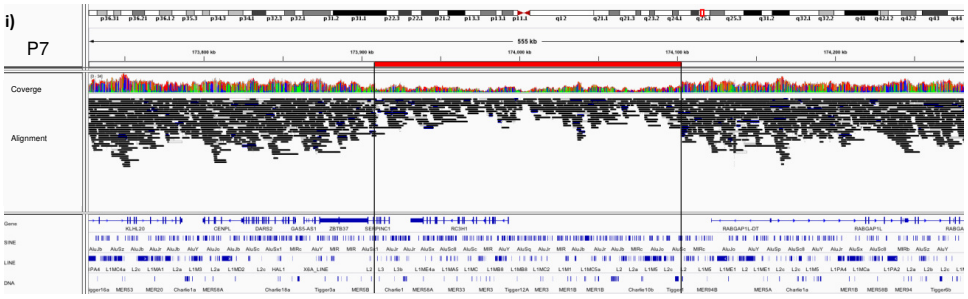


iii)

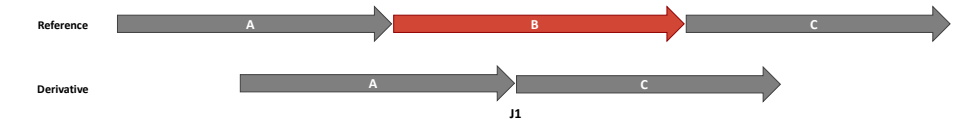
```
1-173919740   ACCTGTAATCTCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAA*CCCGGGAGGTGGAGGCTGCAGTGAGTCGAGATTACGCCACTGCACTCTAGC   173919840
                                                                |  |||  |||||||| |||||||| | |||||  | |||||||||| |||
J1-1          ACCTGTAATCTCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAA*TCTGGGGGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCACTCCAGC    100
                |||||||||  ||||  ||  |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
1-173908362   GCCTGTAATCCTAGCTCCTTGGGAGGCTGAGGCAGGAGAATCACTTGAA*TCTGGGGGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCACTCCAGC   173908462
```

(G) P7

i)



ii)

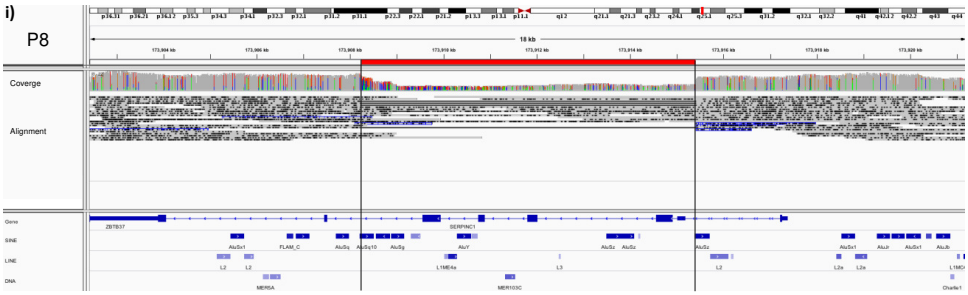


iii)

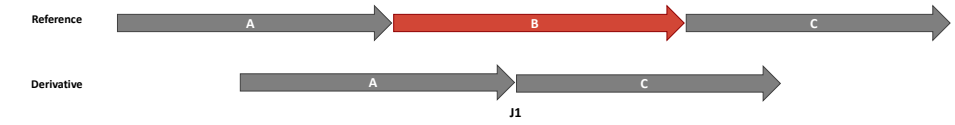
```
1-173908284    TTCAAGACCAGCCTGGCCAACATAGTGAAACCCCGTT TCTACTAAAAAT*GCAAAAATTAGCCGGGCGTGGTGGCAGGCGCCTGTAATCCTAGCTCCTTGG    173908384
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||| ||| ||||||   |||||||||||||| |||| ||
J1-1           TTCAAGACCAGCCTGGCCAACATAGTGAAACCCCGTT TCTACTAAAAAT*ACAAAAATTAGCCAGGCATGGTGGTGAGCGCCTGTAATCCCAGCTACTCCA    100
               ||||||||||||||||  ||||||| | ||||| |||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
1-174102965    GTCAAGACCAGCCTGATCAACATGGAGAAACACCGTC TCTACTAAAAAT*ACAAAAATTAGCCAGGCATGGTGGTGAGCGCCTGTAATCCCAGCTACTCCA    174103065
```

(H) P8

i)



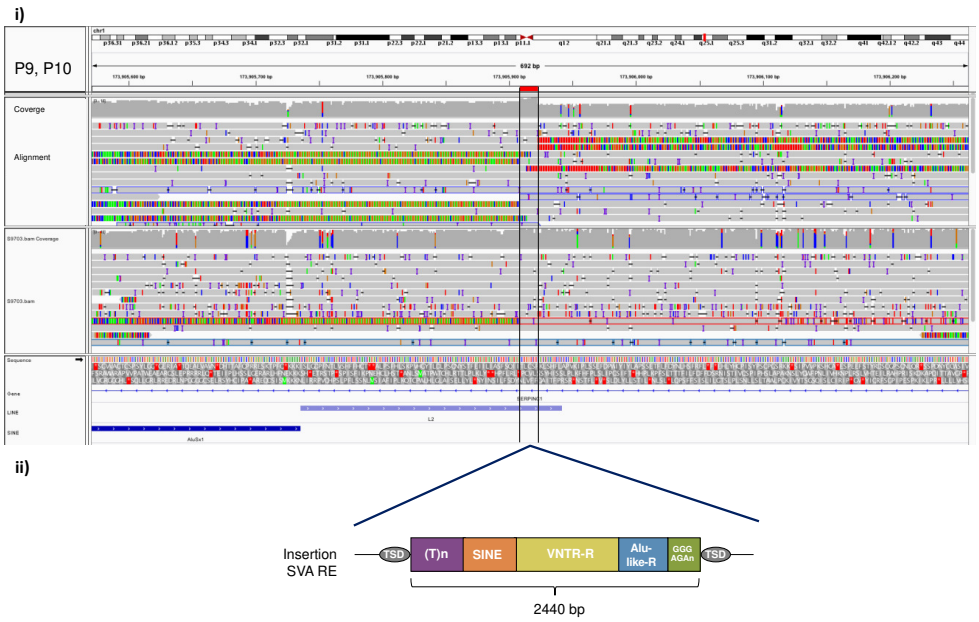
ii)



iii)

```
1-173908168    GGCAACATTTTTAGTTTTCAAAACTAGCTTAAAGACAGCCCAATAGCAC*A GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC CAAGGCAGGTAGAT    173908268
               |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||| | ||||
A3'C5'- 1      GGCAACATTTTTAGTTTTCAAA-CTAGCTTAAAGACAGCCCAATAGCAC* AGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC TGAGGCATGCAGAT    99
               | | |     |  |          |  || |    |             ||||||||||||||||||||||||||||||||||||||||||||||||||||
1-173915355    AGAAGTAAGAAGAAATGGATGGAGCAAAGGCAATATCTGGCCGGGTGCG* -GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC TGAGGCATGCAGAT    173915454
```

(I) P9 and P10


i)
P9, P10
692 bp
Coverge
Alignment
ii)
Insertion
SVA RE
TSD
(T)n
SINE
VNTR-R
Alu-
like-R
GGG
AGAn
TSD
2440 bp

iii)

```
P9 53-88:      (T)n - Simple repeat
P9 89-2366:    SVA_E  - Retrotransposon
P9 2367-2490: G-rich - Low_complexity

1-173905870   TTGAAATTATATTAATAGCATCCTTTTCTCAGATTATAACCTTGTGTTCTTT------------------------------------------------   173905926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
P9-1          TTGAAATTATATTAATAGCATCCTTTTCTCAGATTATAACCTTGTGTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTGATCATTCT   100

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-101        TGGGTGTTCTGCAGAGGGATTTGGCAGGGTCATAGGACAATAGTGGGGGAAGGTCAGCAGATAAACAAGTGAACAAAGGTCTCTGGTTTTCCTAGGCAG   200

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-201        AGGACCCTGCGGCCTTCCGCAGCGCTTGTGCCCCTGGGTACTTAAGATTAGGAGTGGTGATGACTCTCAACGAGCATGCTGCCCTCAAGCATCTGTTCAA   300

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-301        CAAAGCACATCTTGCACCGCCCTTAATCCATCTAACCCTGAGTGGACACAGCACATGTCTCAGAGAGCACAGGGTTGGGGATAAGGTCACAGATCAACAG   400

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-401        GATCCCAAGGCAGAAGAATTTTTCTTAGTACAGAACAAAATGAAAAGTCTCCCATGTCTACTTCTATCCACACAGACCCAGCAACCATCCGATTTCTCAA   500

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-501        TTTTTCCCCACCCTTCCCGCCTTTCTATTCCACAAAACCGCCATTGTCATCATGGCCCATCCTCAATGAGCCGCTGGGCACACCTCCCGACGGGCGTGGC   600

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-601        CGGGCAGAGGGGCTCCTCACTTCCCAGTAGGGCGGCCCGGCAGAAGTGCCCCTCACCTCCCAGATGGGGCGGCTGGCCGGGCGGGGGCTGACCCCTCCGC   700

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-701        CCTCCCGGACGGGCGGCTGGCCAGGCAGAGGGCTCCTCACTTCCCAGTAGGGGCGGCCGGGCAGGCGCCCTCACCTCCTGGAGGGCGGCTGGCCGGGCGG   800

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-801        GGCTGTTCCCCCACCTCCCTCCCGGACGGGGCGGCTGGCGGCAGAGGTCCTCACTCCCGAGGGCGGCCGGGCAGAGGCGCCCTCACCTCCCGGACGGGGC   900

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-901        GGCCGGCCAGGCAGGGGCTGATCCCACCCTCCCGGACGGGGCGGCTGGCCGGGCGGGGCTGCTCCCCACCTCCTCCCGGACGGGGCGGCTGGCCGGGCAG   1000

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-1001       AGGGGTCCTCACTTCCCATAGGGGCGGCCGGGAGGGCGCCTCACCTCCCGGACGGGCGGCTGGCCAGGCAGGGGGCTGATCCCCACCTCCCTCCCGGACA   1100

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-1101       GGGCGGCTGGCCGGGCGGGGGCTGACCCCCACCTCCCTCCCGGACGGGCGGCTGGCCGGGCAGGGGGCTGACTCCTCTCCCTCCCGGACGAGGCGGCTGG   1200

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-1201       CCGGGCGGGGCGTGCCCCCCACCTCCCTCCCGGATGGGGCGGCTGGCCAGGCGGGGGCTGACCCCCACCTCCCCTCCTGGGCGGGGCTGGCGGCCGGGCA   1300

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-1301       GAGGGCTCCTCACTTCCCAGTAGGGGCGGCCGGGCAGAGAGCCCTCACCTCCCGGCCGGGCGGCTGGCCGACCCCCCCTCCCTCCCGGACGGGCGGCTGG   1400

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-1401       CCGGGCAGAGGGGCTCCTCACTTCCCAGTAGGGGCGGCCGGGCAGAGGAGCCCTCACCTCCCGGACAGGGCGGCTGGCCGGGCGGGGGCTGACCCCACCT   1500

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-1501       CCTCCCAGGACGGGTGGCTGCCGGGCGGAGACGCTCCTCACTTCCCAGACGGGTGGTTGCCAGACGGAGGGGCTCTCACTTCTCAGACGGGGCGGTTGCC   1600

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-1601       AGGCAGAGGGTTTCCTCACTTCTCAGACGGAGCGGCCGGGCAGAGACACTCCTCACCTCCAGACAGGGTTGCGGCCCAGCAGAGGCGCTCCTCACATCCC   1700

1-173905926   ----------------------------------------------------------------------------------------------------   173905926
P9-1701       AGACAGGGCGGCGGGGCAGAGGTGCTCCCACATCTCAGACGATGGGCGGCCGGGCAGAGACGCTCCTCACTTCCTAGATGGGATGGCGGCGGGAAGAGGC   1800

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-1801       GCTCGCCTCCTAGATGGGATGGCGGCCGGGCAGAGAGGCCTCACTTTCCACACTGGGCAGCCAGGCAGAGGGCTCCTCATATCCCGGGACGATGGGTGGC   1900

1-173905926   ----------------------------------------------------------------------------------------------------   173905926

P9-1901       CAAGCAGAGACGCTCCTCACTTCCCAGACGGGGTGGCGGCCGGGCAGAGGCTGCAATCTCGGCTCTCCGGGAGGCCAAGGCAGGCGGCTGGGAGGTGGCT   2000

1-173905926   ----------------------------------------------------------------------------------------------------   173905926
```

```
P9-2001         GCGGAGCCGAGATCACGCCACTGCACTCCAGCCTGGGCACCATTGAGCACTGAGTGAACGGACTCCATCTGCAATCCGGCACCCCCGGGGAGGCCGAGGC     2100

1-173905926     ----------------------------------------------------------------------------------------------------     173905926

P9-2101         TGGCGGATCACTCGCGGCCAGGAGCTGGAGACCAGCCCGGCCAACACAGCGAAACCCCATCTCCACCAAAAAAACGAAAACCAGTCAGGCGTGGGCGGCG     2200

1-173905926     ----------------------------------------------------------------------------------------------------     173905926

P9-2201         CCTGCAATCGCAGGCACTCGGCAGGCTGAGGCAGGAGAATCAGGCAGGAGGTGCAGTGAGCGAGATGGCAGCAGTACCGTCCAGTTTTGGGCTCGGCATG     2300

1-173905926     ----------------------------------------------------------------------------------------------------     173905926

P9-2301         AGAGGGAGAGGGAGACGGGAGAGGGAGAGGGAGACGGAGAGGGAGAGGGAGACGGGAGAGGGAGGGGGAGCGGGGAGAGGGAGAGGGGAGAGGGGGGCG     2400

1-173905926     ----------------------------------------------------------------------------------------------------     173905926

P9-2401         GGGAGGGGGAGAGGGAGACGGGAGGCAGAGGGAGCGGGGGGAGGGTGAGGGGGGCGGGAGGGAGGAGGGAGACGGGAGAGGGAGACGGGGCCACCTTGTG     2500

1-173905926     ---TTTCAAGCTATCACATTTCCTCGCTCCCGTTAAAATTCCACTTTCCTCTGA        173905970
                   ||||||||||||||| |||||  |||||||||||| ||||||||||||||||
P9-2501         TTCTTTCAAGCTATCACTTTTCCCTGCTCCCGTTAAA-TTCCACTTTCCTCTGA        2600
```

4.2.4 Molecular dissection of structural variants involved in antithrombin deficiency

a. Introduction

The total number of gross alterations affecting *SERPINC1* that are collected in the Human Genome mutation Data Base (HGMD) is 30 (28 deletions and 2 duplications or insertions), most of them detected only by MLPA. Thus, according to the limitations of this technology, this method only identified but not fully characterized SVs causing antithrombin deficiency. In this thesis, we have analysed the worldwide largest cohort of patients with antithrombin deficiency caused by SVs (n=39), which have been detected and characterized by using different molecular methods, including whole genome nanopore sequencing.

b. Patients

Figure 59 shows the flow chart of patients selected for this study. The study was done in a large cohort of patients with antithrombin deficiency consecutively recruited mainly in two reference centers for this rare disease during 26 years (1994-2020) in Spain and 28 years in Belgium (1990-2020). For the total cohort of 398 and 278 index cases in Spain and Belgium, a complete functional and biochemical characterization was done and *SERPINC1* was sequenced by Sanger or NGS as described. Cases with negative results were further evaluated for gross gene defects by MLPA. Twenty-nine cases from these two cohorts had a positive MLPA result.

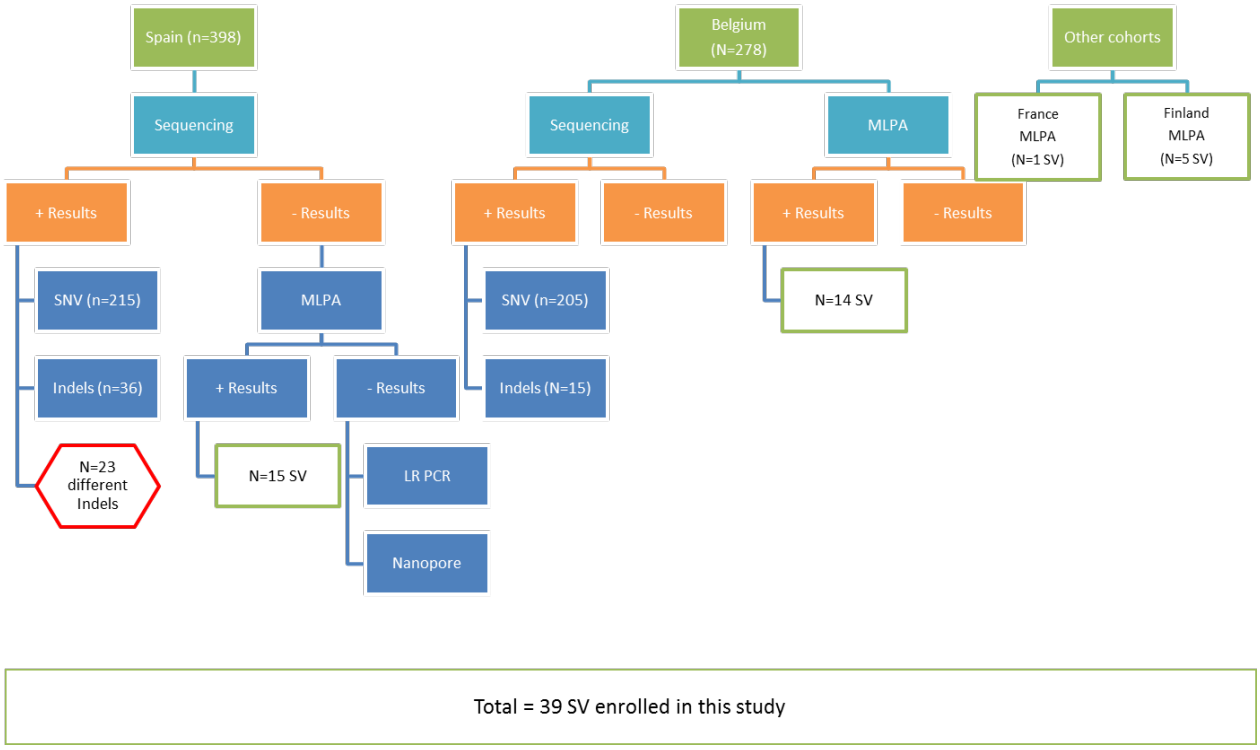


Figure 59: **Flow chart of patients selected for the analysis of structural variants causing antithrombin deficiency**

The study also included four additional cases from the Spanish cohort with SVs in *SERPINC1* not detected by

MLPA, reported in previous chapters: one case with deletion of intron 1 (Chapter 4.2.2: "Identification of the first large intronic deletion responsible of type I antithrombin deficiency not detected by routine molecular diagnostic methods"), detected by LR-PCR, and three cases with insertion of a retrotransposon, two of them identified by nanopore sequencing in the previous chapter (Chapter 4.2.3: "Long-read sequencing resolves structural variants in *SERPINC1* causing antithrombin deficiency").
The cohort of patients with antithrombin deficiency caused by SVs was completed with nine additional cases with SVs detected by MLPA recruited following the same strategy from Finish and French cohorts of patients with antithrombin deficiency.
Thus, the final number of patients with antithrombin deficiency caused by SVs was 39.
Finally, MLPA was also done in cases with antithrombin deficiency caused by small INDELs <50bp identified by sequencing methods in the Spanish cohort. Twenty three different INDEL variants were identified in 36 patients (Figure 59).

c. Clinical data of patients with SVs

The incidence of SVs and the clinical features of patients carrying these SVs was obtained using the data from the Spanish and Belgian cohorts. Nearly 5% of consecutive and unrelated cases with antithrombin deficiency recruited in the whole cohort of patients are explained by SVs, 19/398 (4.7%) in the Spanish cohort and 14/278 (5%) in the Belgian cohort, values that were similar to that described in other long cohorts (3-8%) [136, 35].
All patients carrying SVs had type I deficiency, the average anti-FXa activity was 51.9±9.0% and the mean antigen level was 49.5±10.6%. A high proportion of cases, 85.2%, presented family history of thrombosis. The severe clinical impact of SVs affecting *SERPINC1* was reflected by the early thrombotic events of carriers (median age of the first thrombotic event: 32.7 years). Indeed, 4 out of 33 carriers (12.1%) had the first thrombosis in the pediatric age (under 18 years old). Finally, it is also remarkable the incidence of recurrent thrombosis among carriers (12.1%).

d. Diversity of SVs

Our study of the whole cohort of cases with gross gene defects in *SERPINC1* (N= 39), revealed a wide diversity of SVs. Although most cases had whole or partial deletions (N= 31/39, 79.5%), four had duplications (10.3%) and three had insertions (7.7%) (Table 17).

Table 17: **Description of 39 cases with Antithrombin deficiency caused by SV.** Detailed description of results obtained by each method: MLPA, LR-PCR and nanopore sequencing. GENE left and GENE right: last gene affected in the SV in the left side and right side. CGHa: CGH array; ND: not determined; * indicates that genomic coordinates were obtained by Zoom analysis of CGH array, so they are not precise.

ID	MLPA	LR PCR data	CGHa data (del)	coord CGHa	nanopore sequencing	coord sequence bp	size SV (bp)	bp nucleotide	GENE left	GENE right	COHORT
Whole SERPINC1 deletions											
P1	Exons 1-7	ND	30 genes	1:173501975-175305010	ND		1803,032	no	SLC9C2	KIAA0040	SP
			76 genes	(Mosaic) 1:170208533-178450925	ND		8242392	no	LINC01142	RASAL2	
P2	Exons 1-7	ND	30 genes	1:172928626-174796949	yes (MinION)	1:172987296-174843232	1868323	yes	TNFSF4	RABGAP1L	SP
P3	Exons 1-7	ND	20 genes	1:173786826-174756183	yes (PromethION)	1:173848142-174816147	968005	yes	DARS2	RABGAP1L	SP
P4	Exons 1-7	ND	20 genes	1:173786826-174750889	ND	1:173848142-174816147	968005	yes	DARS2	RABGAP1L	SP
P5	Exons 1-7	ND	5 genes	1:173881395-174342043	ND		460648	no	ZBTB37	RABGAP1L	FN
P6	Exons 1-7	ND	serpinc1 and next (zoom)		yes (PromethION)	1:173879820-173925989	46169	yes	ZBTB37	SERPINC1	SP
P7	Exons 1-7	ND	ND		yes (PromethION)	1:173850996-173950174	99178	yes	DARS2	RC3H1	FR
P8	Exons 1-7	ND	ND		ND		-	no			FN
P9	Exons 1-7	ND	6 genes	1:173787033-174225421	yes (miniON)	1:173787361-174223422	436061	yes	CENPL	RABGAP1L	SP
P10	Exons 1-7	ND	ND		ND			no			BE
P11	Exons 1-7	ND	4 genes	1:173884518-174418717	ND		534199	no	ZBTB37	RABGAP1L	BE
P12	Exons 1-7	ND	4 genes	1:173884504-174418717	ND		534213	no	ZBTB37	RABGAP1L	BE
P13	Exons 1-7	ND	3 genes	1:173896668-173942868	ND		46200	no	ZBTB37	RC3H1	BE
P14	Exons 1-7	ND	3 genes	1:173888461-173971254	ND		82793	no	ZBTB37	RC3H1	BE
P15	Exons 1-7	ND	3 genes	1:173859535-173926473	ND		66938	no	DARS2	SERPINC1	BE
P16	Exons 1-7	ND	14 genes	1:173692858-176076493	yes (PromethION)	1:173686375-176083118	2396743	yes	ANKRD45	COP1	BE
			112 genes	(Mosaic)1:170194517-179599871	ND		9405354	no	LINC01681	TDRD5	
P17	Exon 1-7	no	3 genes	1:173888460-174138926	ND		250466	no	ZBTB37	RABGAP1L	FN
Partial SERPINC1 Deletions											
P18	Exons 1-2	yes	normal		ND	1:173914639-173922031	7392	yes	SERPINC1	SERPINC1	FN
P19	Exon 2	yes	ND		ND	1:173912550-173914877	2327	yes	SERPINC1	SERPINC1	BE
P20	Exon 3-4	yes	ND		ND	1:173910328-173913605	3277	yes	SERPINC1	SERPINC1	BE
P21	Exon 2-5	yes	normal		yes (PromethION)	1:173908360-173915405	7045	yes	SERPINC1	SERPINC1	SP
P22	Exon 1	yes (del ex 1-2)	normal		yes, CxSV (PromethION)	1:173911379-173915115; 1:173912151-173919034	3737; 6884	yes	SERPINC1	SERPINC1	SP
P23	Exon 2-7	no	2 genes (zoom)	1:173879820-173915405*	ND		ND	no	ZBTB37	SERPINC1	SP
P24	Exon 1-5	no	2 genes	1:173907795-174100688	yes (PromethION)	1:173908511-174102900	194389	yes	SERPINC1	RABGAP1L	FN
P25	Exon 1	no	normal		yes (PromethION)	1:173916704-173935703	18999	yes	SERPINC1	RC3H1	SP
P26	Exon 1-2	no	4 genes	1:173911377-174138926	yes (MinION)	1:173912160-174154195	242035	yes	SERPINC1	RABGAP1L	SP
P27	Exon 1	no	normal		ND		ND		SERPINC1	SERPINC1	BE
P28	Exon 7	no	normal		ND		ND		SERPINC1	SERPINC1	BE
P29	Exon 5	yes			ND	1:173909621-173910321	700	yes	SERPINC1	SERPINC1	BE
P30	Exon 4	ND	ND		ND		ND	no	SERPINC1	SERPINC1	SP
P31	Exon 6	yes	ND		yes (MinION)	1:173907040-173908412	1372	yes	SERPINC1	SERPINC1	BE
Partial SERPINC1 duplications											
P32	Exon 6	yes	normal		ND	1:173907323-173907554	198	yes	SERPINC1	SERPINC1	SP
P33	Exon 6	yes	normal		ND	1:173906975-173907838	863	yes	SERPINC1	SERPINC1	SP
P34	Exon 6	yes	ND		ND	1:173906975-173907838	863	yes	SERPINC1	SERPINC1	BE
P35	Exon 1;2;4 deletion ex6	yes (dup 1-5)	normal		yes (PromethION)	1:173908360-173915405	7,045	yes	SERPINC1	SERPINC1	SP
Normal MLPA cases											
P36	Normal	del intron 1	normal		yes (minion)	1:173914906-173916922	2016	yes	SERPINC1	SERPINC1	SP
P37	Normal	normal	ND		yes, INS SVA_E (PromethION)	1:173905922	2440	yes	SERPINC1	SERPINC1	SP
P38	Normal	normal	ND		yes, INS SVA_E (PromethION)	1:173905922	2440	yes	SERPINC1	SERPINC1	SP
P39	Normal	yes, INS SVA_E	ND		ND	1:173905922	-	yes	SERPINC1	SERPINC1	SP

Whole gene deletions

Complete deletion of *SERPINC1* was detected by MLPA in 17 cases (P1-P17) (Table 17).

CGH array, which was used to confirm the deletions detected by MLPA and to define the extension of the SVs, proved the complete deletions of the gene in 13 out of 14 available cases. All deletions involved more than 50 Kb, but showing a considerable heterogeneity of the deletion size, from 460 Kb to 1,8 Mb. Thus, the number of genes affected by deletions involving *SERPINC1* ranged from 3 to 30 genes. The deletion of *SERPINC1* defected in P6 by MLPA was not detected by CGH array because of the reduced size of the deletion (46 Kb). However, a zoom analysis of the CGH array found results supporting the deletion, which also involved part of the neighboring gene (ZBTB37). This result was confirmed by nanopore sequencing with PromethION (Figure 60 and Figure 61).

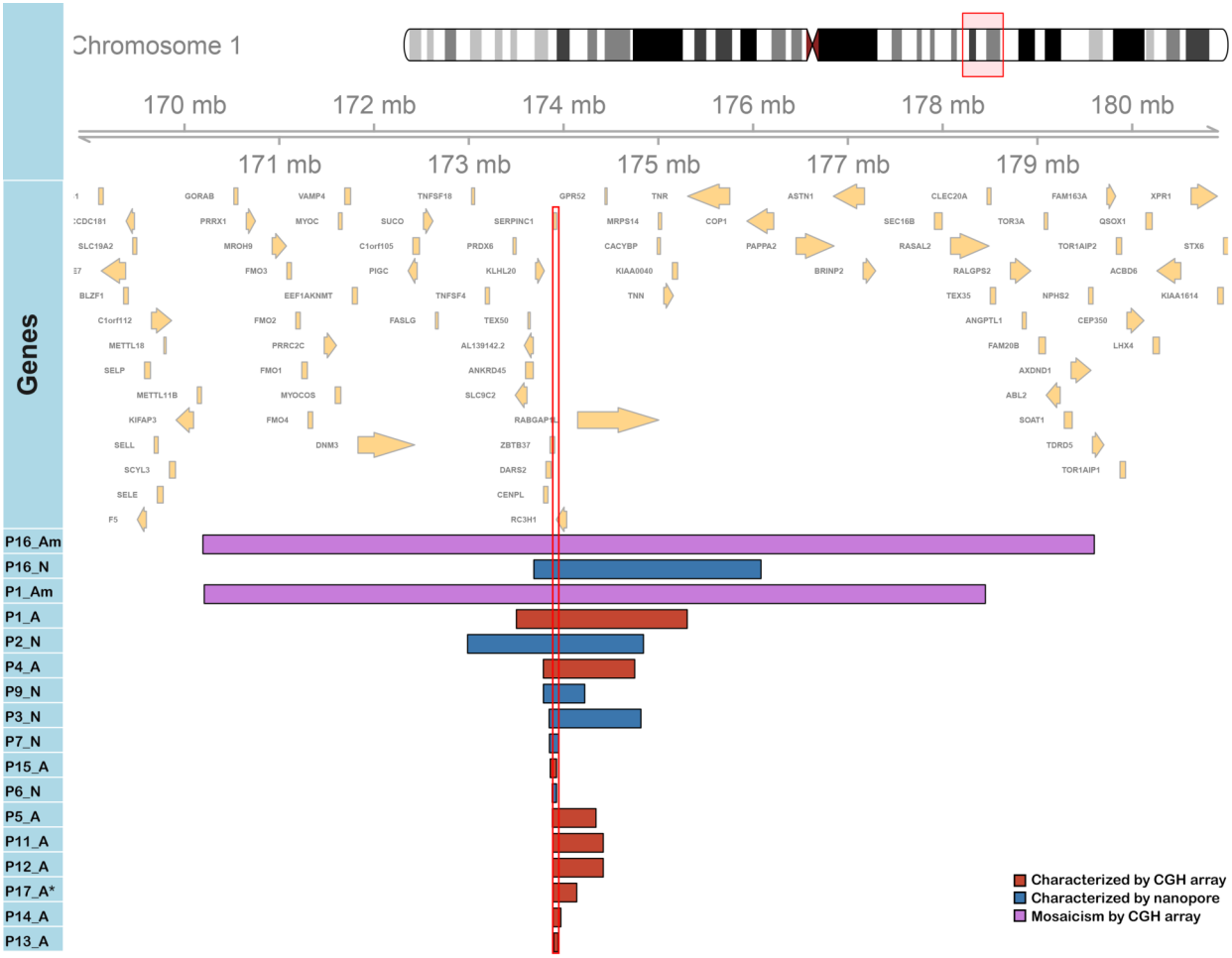


Figure 60: **Localization and extension of whole gene deletions affecting *SERPINC1* that were detected by MLPA in patients with antithrombin deficiency from our cohort. The color code and leyend of each patient show information on the method used to characterize the deletion.** For cases P1 and P16, the mosaic deletion detected by CGH array is also shown in pink. A: CGH array; N: Nanopore sequencing; m: mosaicism. $**$ indicates that genomic coordinates were obtained by Zoom analysis of CGH array, so they are not precise.

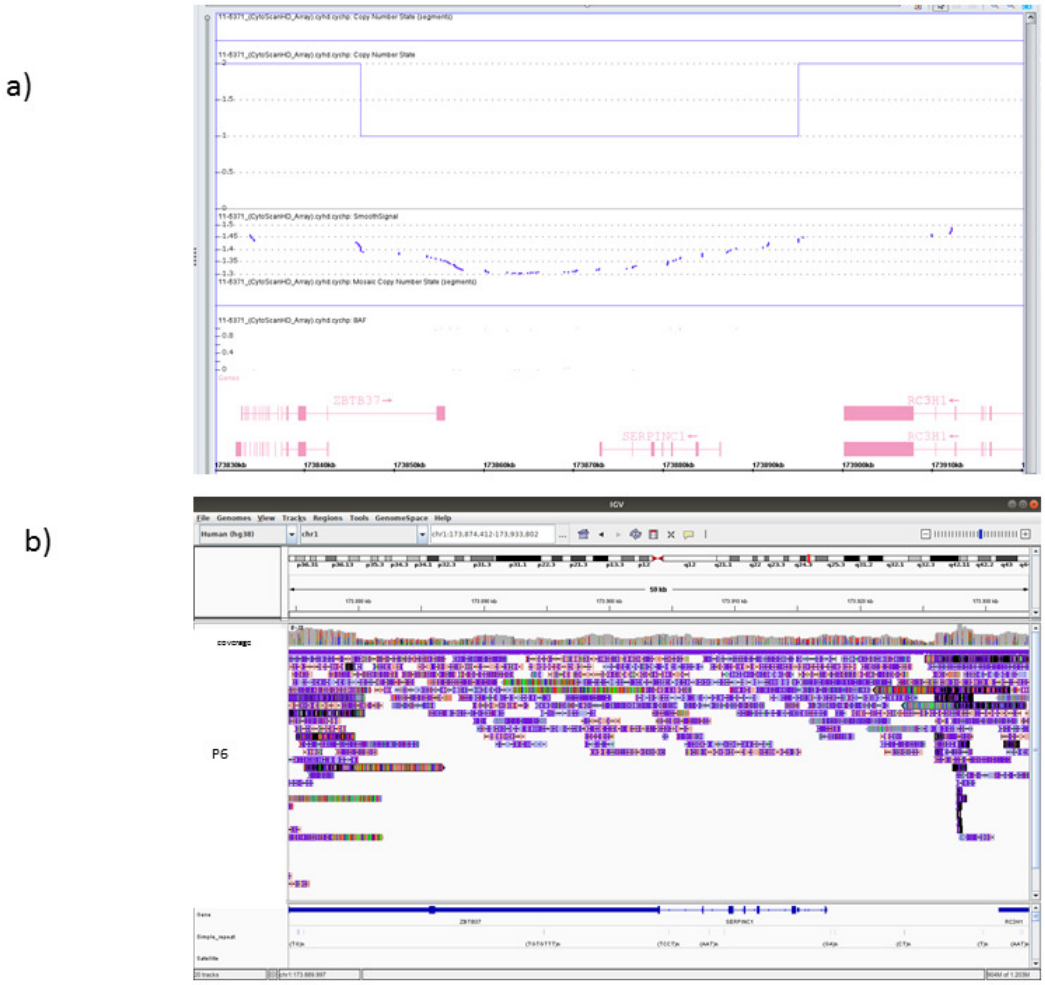


Figure 61: **Detailed description of the deletion identified in P6.** a) Zoom of the result obtained by CGH array in the region surrounding *SERPINC1* supporting the deletion. b) Nanopore alignments covering the deletion.

Further characterization of SVs involving deletion of the whole *SERPINC1* gene was done in 6 cases by nanopore sequencing: four with PromethION (P3, P6, P7 and P16) and two with MinION (P2, P9) (Figure 62). This method confirmed the extension of the deletion in all tested cases (Figure 60) and allowed the characterization of the breakpoints (See latter).

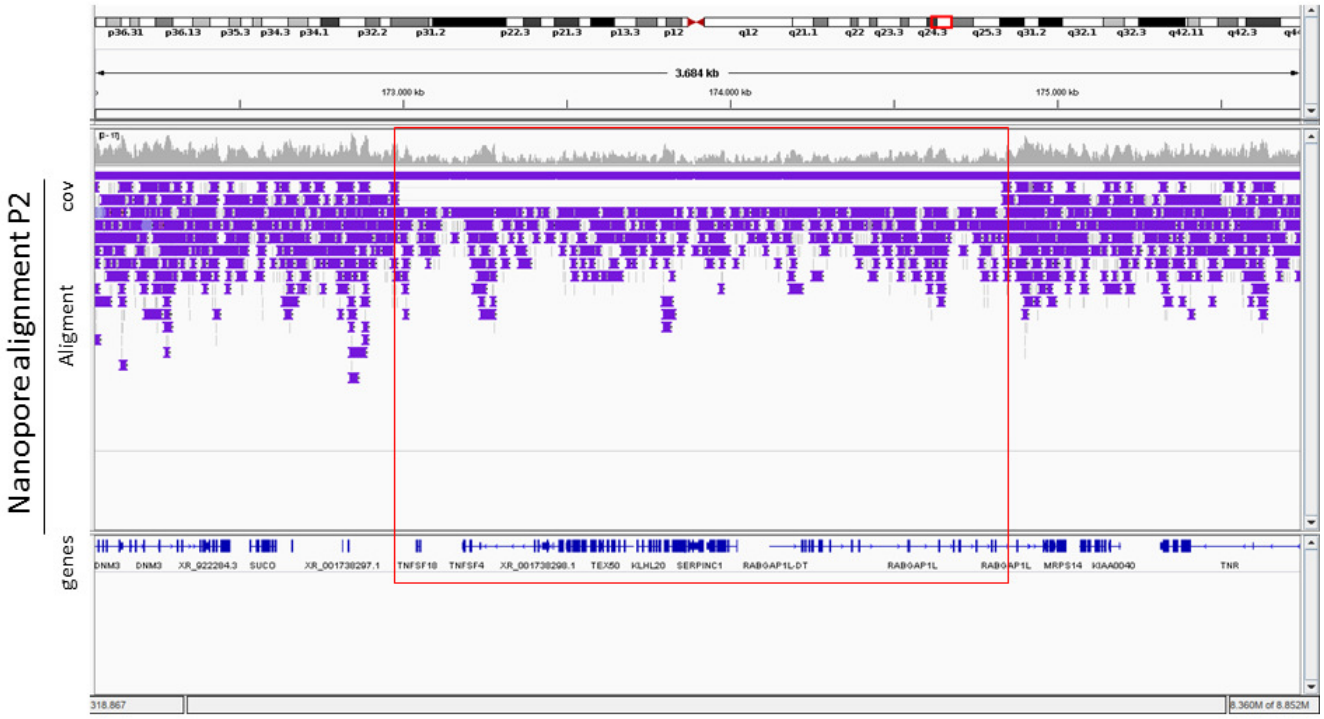


Figure 62: **Alignment of nanopore sequences generated by MinION showing the deletion covering *SERPINC1* and 19 additional genes in P2.**

Finally, we remark P1 and P16, patients who had an additional mosaic deletion detected by CGH array of 8,242,392 and 9,405,354 Mb respectively covering *SERPINC1* and 75 and 111 additional genes. In both cases the mosaic deletion covered the germline deletion (Figure 60).

Partial deletions

Fourteen cases presented partial deletions of *SERPINC1* detected by MLPA (P18-P31) (Table 17). To further characterize these SVs, LR-PCRs covering the *SERPINC1* gene were designed in available samples. However, LR-PCR was only able to cover the deletion in 7 out of 13 cases evaluated (P18-22, P29, P31). In cases P18-P22, LR-PCR results complemented the MLPA results and give us the possibility to analyze the breakpoint at nucleotide level resolution (see later). Interestingly, in case P22 a relevant discrepancy between these methods was observed. MLPA diagnosed a deletion of exon 1, which was confirmed in two affected members of the family, but LR-PCR and NGS, detected a deletion of exon 1 and 2. Nanopore sequencing showed the *SERPINC1* architecture of the mutant allele, revealing a complex SV result of a duplication of exons 2 and 3 and then a deletion of exons 1 and 2 [146].

In all cases with partial *SERPINC1* deletion in which LR-PCR amplification failed, the deletion identified by MLPA involved exons 1 or 7, suggesting that the deletion might be longer, covering primers positions designed for LR-PCR. So, further studies done in these 7 cases by CGH array and/or nanopore sequencing confirmed larger deletions involving up to 3 additional neighbouring genes (Table 17 and Figure 63).

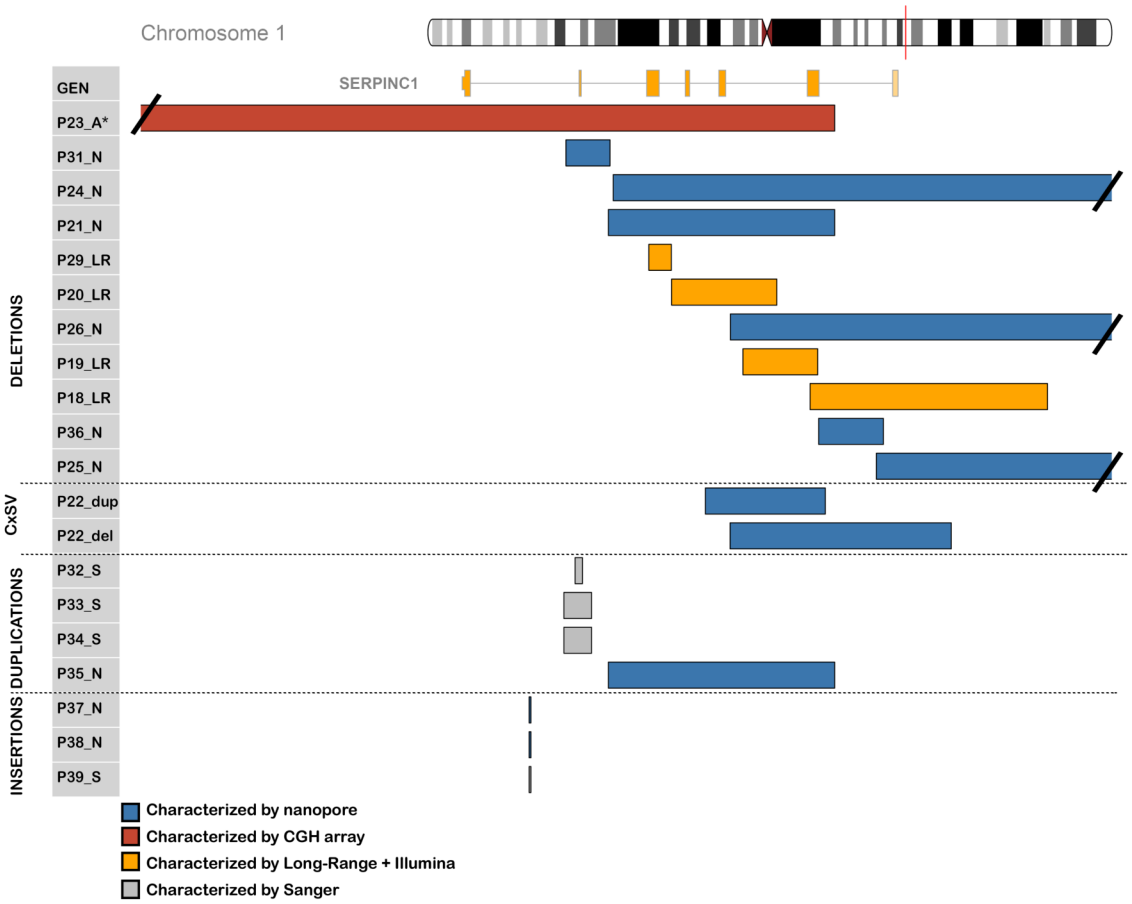


Figure 63: **Localization and extension of partial gene deletions, duplications and insertions affecting *SERPINC1* that were detected in patients with antithrombin deficiency from our cohort.** The color code and leyend show information on the method used to characterize the SV. $*$ indicates than the coordinate localization of P23 has been obtained by Zoom analysis of CGH array A: CGH array; N: Nanopore sequencing; S: Sanger sequencing, LR: LR-PCR.

In summary, partial deletions affecting *SERPINC1* were heterogeneous in size, ranging from 700 bp to 242035 bp, and may also involve neighbouring genes. Partial duplications of *SERPINC1* were identified in the remaining 4 cases with antithrombin deficiency caused by SV detected by MLPA (P32-P35) (Figure 62) . Three cases (P32-P34) carried tandem duplication of exon 6, two of them previously described in detai in previous chapters of this thesis [119]. The remaining case (P35) had a hardly interpretable partial duplication of *SERPINC1* involving exons 2, 3 and 5 and a deletion of exon 6 according to MLPA results. A range of possible LR-PCR designed to test a potential tandem duplication of exons 1-5 confirmed a 7045 bp tandem duplication of exons 1-5, which was further validated by WGS nanopore sequencing, obtaining the final architecture of the SV [146].

Structural variants in SERPINC1 not detected by MLPA

The range of SVs affecting *SERPINC1* that have been identified in patients with antithrombin deficiency included four additional cases in which either sequencing of the gene, MLPA or CGH array rendered negative results (P36-P39). The deletion of the majority of intron 1 and the first nucleotides of exon 2 was identified in P36 by LR-PCR and verified by specific PCRs in affected relatives [147] . In this study we confirmed that nanopore sequencing using the MinION platform also detects this SV, by using WGS (Figure 64).

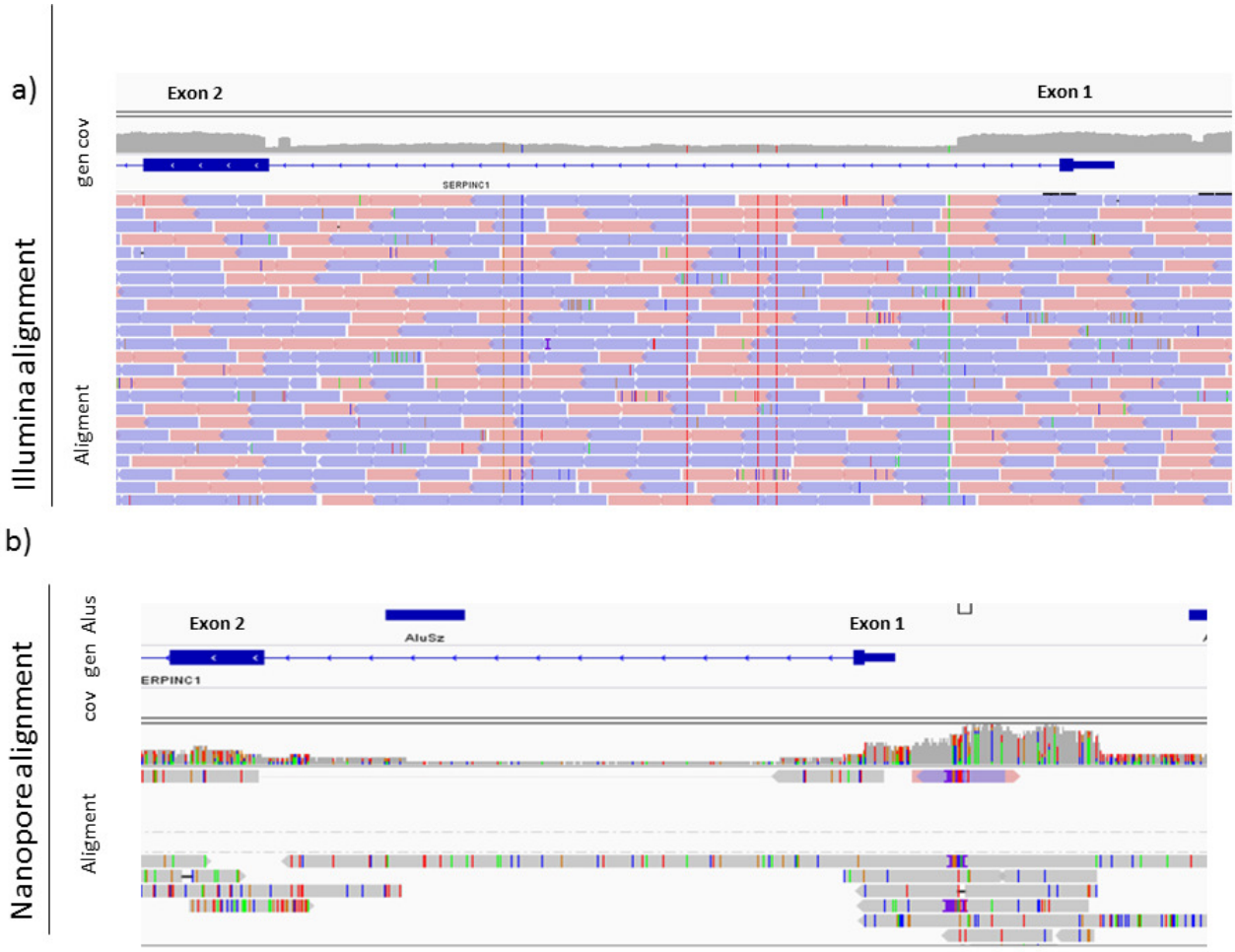


Figure 64: **Detection of the deletion of intron 1 of *SERPINC1* in P36 detected: a) by LR-PCR sequencing, and b) by nanopore sequencing in minION device.**

Moreover, the insertion of a 2.4 Kb SVA retrotransposon in intron 6 identified in two unrelated cases (P37, P38) with family history of antithrombin deficiency was detected by WGS nanopore sequencing [146] . Interestingly, the PCR designed to validate this insertion in probands and relatives [146] was used to screen new cases with antithrombin that had not known molecular defect, and interestingly, we identified a new unrelated carrier of the same insertion (P39)(Figure 65A). Moreover, using a different set of primers and PCR conditions we were able to amplify the full sequence of the insertion, characterizing both breakpoints and validating the results obtained by WGS nanopore sequencing (Figure 65B).

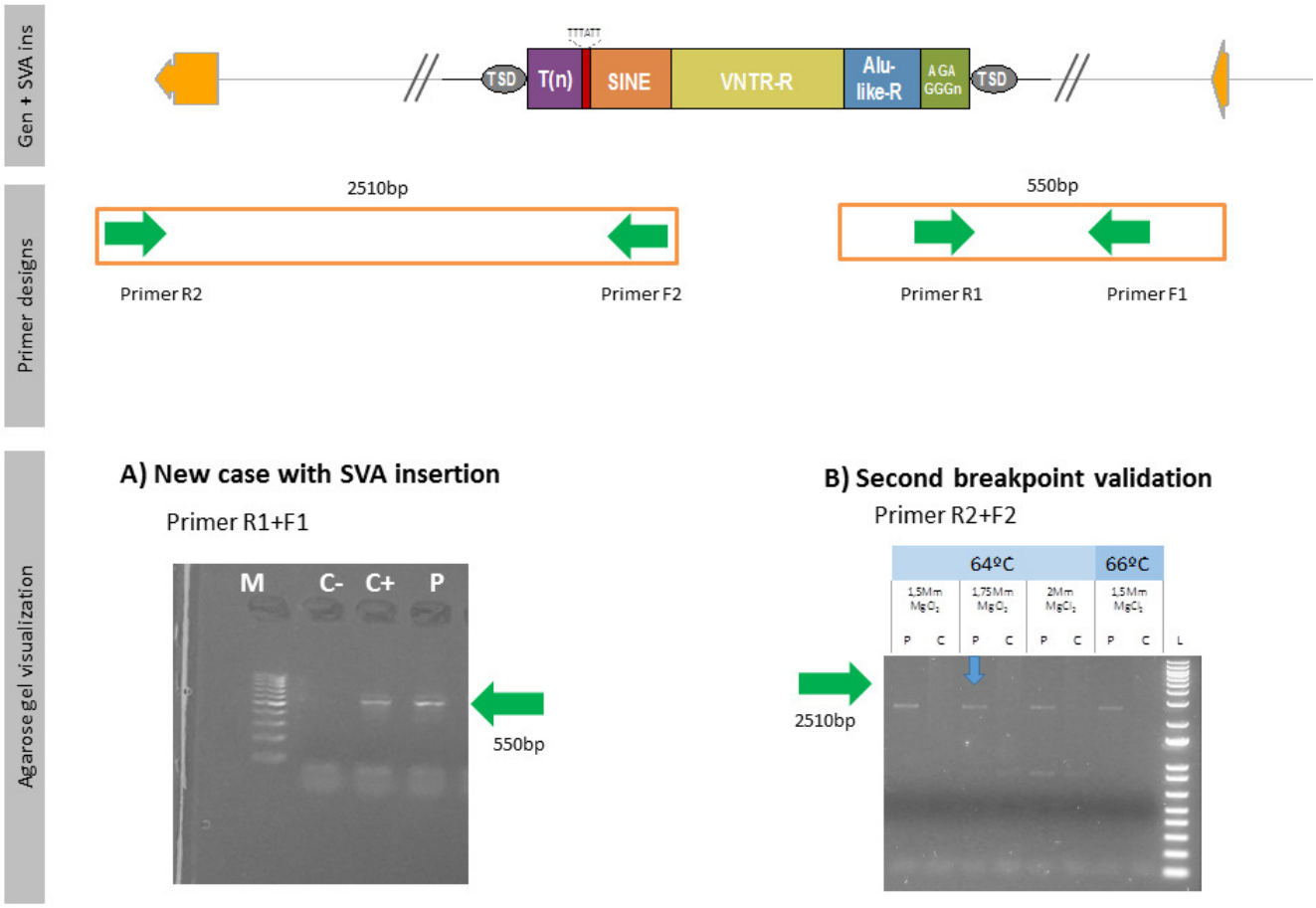


Figure 65: **Characteristics of the SVA inserted in intron 6 of *SERPINC1*.** The localization of PCR primers used to validate this insertion and screen new cases with antithrombin deficiency of unknown cause are shown. A) Identification of a new case carrying the same SV. B) PCR amplification of the 5'-end of the SVA insertion.

e. Recurrent breakpoints

In 34 cases with SVs affecting *SERPINC1*, a relative or precise breakpoint position was determined. A global view of all the breakpoints identified in this study remarked the presence of relatively common regions involved in gross gene defects affecting *SERPINC1* that resulted in antithrombin deficiency. Recurrent breakpoints included centromeric and telomeric regions flanking *SERPINC1* mainly involving 4 genes: *RABGAPL1* (9 cases);*ZBTB37* (8 cases); *RC3H1* (5 cases) and *DARS2* (4 cases) (Figure 66 and Tables 18 and 19). The exact localization of most of these breakpoints was not determined since they were identified by CGH array. Thus, these recurrent intergenic breakpoints were quite extense (up to 700 Kb) (Figure 66).

Interestingly, our study also identified much narrower (up to 1807 bp) recurrent intragenic breakpoints, and as they were studied by methods reaching nucleotide resolution, all were precisely localized. Intragenic recurrent breakpoints affected mainly 3 *SERPINC1* intronic regions and one exon: intron 5 (8 cases), intron 6 (7 cases), intron 1 (6 cases) and exon 2 (3 cases) (Figure 67 and Table 17).

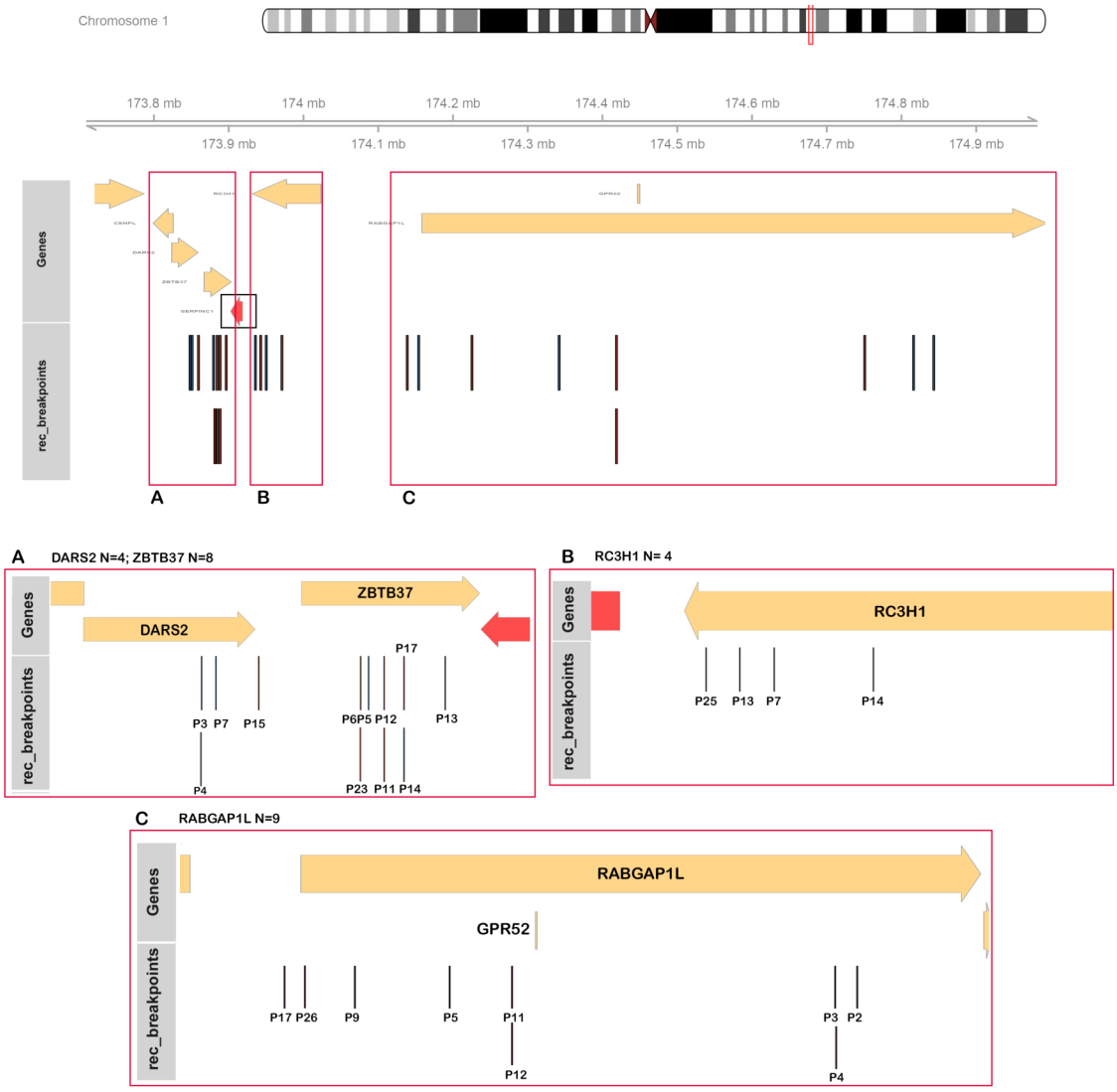


Figure 66: **Localization of recurrent breakpoints in genes flanking *SERPINC1* identified in patients with antithrombin deficiency caused by SVs.**

Figure 67: **Localization of recurrent *SERPINC1* intragenic breakpoints in patients with antithrombin deficiency caused by SVs.**

Table 18: **Analysis of the breakpoint of the SV in patients with Antithrombin deficiency.** * Indicates that the analysis was done with relative coordinates obtained by CGH array. The last column indicates the exact sequence that present microhomology or that is inserted at the breakpoint. In P5, the genomic coordinates were obtained by Zoom analysis of CGH array. ND: not determined, MH: Microhomology; INS: Insertion; DUP: duplication.

ID	technology used for bp analysis	bp nucleotide	RE at bp	START	END	MH/INS/DUP at bp	nt MH	secuence MH/INS
Whole SERPINC1 deletions								
P2	Nanopore + Sanger	YES	YES	L1PA2	L1PA2	MH	43	CTCACGCTGGGAGCTGTAGACCGGAGCTGTTCCTATTCGGCCATCTTG
P3	Nanopore + Sanger	YES	YES	AluY	AluSp	MH	26	TTTTTTTTTTTTTTTTTTTTTGAGA
P4	Sanger	YES	YES	AluY	AluSp	MH	26	TTTTTTTTTTTTTTTTTTTTTGAGA
P6	Nanopore + Sanger	YES	YES	AluSx1	-	MH	4	TGGN
P7	Nanopore + Sanger	YES	YES	AluSx	L1ME3C	No		
P9	Nanopore	YES	YES	-	Alusz	MH;INS	3;29	AAC;ACTCTGACAGGCATGCATCCATCCCAAGT
P16	Nanopore	YES	YES	-	L1PA16	MH;INS	2;1	AC; C
Partial SERPINC1 Deletions								
P18	LR-PCR+Illumina	YES	YES	AluSq2	AluJb	MH	4	GTAT
P19	LR-PCR+Illumina	YES	YES	L3	-	MH	3	CCT
P20	LR-PCR+Illumina	YES	YES	AluY	AluSz	MH	48	GGCGCGGTGGCTCNACGCCTGTAATCCCAGCACTTTGGGAGGCCAAGG
P21	Nanopore + Sanger	YES	YES	AluSx1	AluSz	MH	36	GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC
P22	Nanopore + Sanger	YES	YES	MER103C;-	-;L2a	MH	5; 3	GTTGC; AGC
P24	Nanopore + Sanger	YES	YES	AluSx1	AluSp	MH	12	CAAAAATTAGCC
P25	Nanopore + Sanger	YES	YES	-	A-rich	INS	12	
P26	nanopore	YES	YES	-	AluJr	No	No	
P29	LR-PCR+Illumina	YES	YES	L1MC1	AluY	No	No	
P31	nanopore	YES	YES	AluSp	AluSx1	MH	28	GGAGGCTGAGGCAGGAGAATCNCTTGAA
Partial SERPINC1 duplications								
P32	Sanger	YES	YES	-	-	MH	4	GGCC
P33	Sanger	YES	YES	AluSp	AluSp	MH	18	AAAATACAAAAATTAGCC
P34	Sanger	YES	YES	AluSp	AluSp	MH	18	AAAATACAAAAATTAGCC
P35	Nanopore + Sanger	YES	YES	AluSx1	AluSz	MH	28	GGAGGCTGAGGCAGGAGAATCACTTGAA
Normal MLPA								
P36	LR-PCR+Illumina	YES	No	-	-	INS	2	CC
P37	Nanopore + Sanger	YES	YES	L2	L2	DUP	14	ACCTTGTGTTCTTT
P38	Nanopore + Sanger	YES	YES	L2	L2	DUP	14	ACCTTGTGTTCTTT
P39	Sanger	YES	YES	L2	L2	DUP	14	ACCTTGTGTTCTTT

Table 19: **Detailed information about breakpoint analysis obtained with (www.repeatmasker.org) RepeatMasker for each start and end position of SV breakpoints with a window of 100bp**. Repeat element and Family name are shown. * Marks that the breakpoint has been detected by CGH array.

ID		Chr	Repeat	Family
P1*	start	chr1:173501875-173502075	MamRep1879	DNA/hAT-Tip100
	end	chr1:175304910-175305110	AluSc8	SINE/Alu
P1_mosaic*	start	chr1:170208433-170208633	THE1C	LTR/ERVL-MaLR
	end	chr1:178450825-178451025	L2d	LINE/L2
P2	start	chr1:172987196-172987396	L1PA2	LINE/L1
	end	chr1:174843132-174843332	L1PA2	LINE/L1
P3	start	chr1:173848042-173848242	AluY	SINE/Alu
	end	chr1:174816047-174816247	AluSp	SINE/Alu
P4	start	chr1:173848042-173848242	AluY	SINE/Alu
	end	chr1:174816047-174816247	AluSp	SINE/Alu
P5*	start	chr1:173881295-173881495	L1PA5	LINE/L1
	end	chr1:174341943-174342143	-	-
P6	start	chr1:173787261-173787461	AluSx1	SINE/Alu
	end	chr1:173925889-173926089	-	-
P7	start	chr1:173850896-173851096	AluSx	SINE/Alu
	end	chr1:173950074-173950274	L1ME3C	LINE/L1
P9	start	chr1:173786933-173787133	L1PA6	LINE/L1
	end	chr1:174223322-174223522	L1PA8	LINE/L1
P11*	start	chr1:173884418-173884618	AluJb	SINE/Alu
	end	chr1:174418617-174418817	AluSz	SINE/Alu
P12*	start	chr1:173884404-173884604	AluJb	SINE/Alu
	end	chr1:174418617-174418817	AluSz	SINE/Alu
P13*	start	chr1:173896568-173896768	-	-
	end	chr1:173942768-173942968	AluYf1	SINE/Alu
P14*	start	chr1:173888361-173888561	AluJr4	SINE/Alu
	end	chr1:173971154-173971354	AluSz	SINE/Alu
P15*	start	chr1:173859435-173859635	AluSx3	SINE/Alu
	end	chr1:173926373-173926573	Ricksha_b	DNA/MULE-MuDR
P16	start	chr1:173686275-173686475	HERVIP10FH-int	LTR/ERV1
	end	chr1:176083018-176083218	-	-
P16_mosaic*	start	chr1:170194417-170194617	L1M2	LINE/L1
	end	chr1:179599771-179599971	Tigger3b	DNA/TcMar-Tigger
P17*	start	chr1:173888360-173888560	AluJr4	SINE/Alu
	end	chr1:174138826-174139026	-	-
P18	start	chr1:173914539-173914739	AluSq2	SINE/Alu
	end	chr1:173921931-173922131	AluJb	SINE/Alu
P19	start	chr1:173912450-173912650	L3	LINE/CR1
	end	chr1:173914777-173914977	-	-
P20	start	chr1:173910228-173910428	AluY	SINE/Alu
	end	chr1:173913505-173913705	AluSz	SINE/Alu
P21	start	chr1:173908260-173908460	AluSx1	SINE/Alu
	end	chr1:173915305-173915505	AluSz	SINE/Alu
P22	start	chr1:173911279-173911479	MER103C	DNA/hAT-Charlie
	end	chr1:173915015-173915215	-	-
P22_2	start	chr1:173912051-173912251	-	
	end	chr1:173918934-173919134	L2a	LINE/L2
P23*	start	chr1:173879720-173879920	AluSx1	SINE/Alu
	end	chr1:173915305-173915505	AluSz	SINE/Alu
P24	start	chr1:173908411-173908611	AluSx1	SINE/Alu
	end	chr1:174102800-174103000	AluSp	SINE/Alu
P25	start	chr1:173916604-173916804	-	-
	end	chr1:173935603-173935803	A-rich	Low_complexity
P26	start	chr1:173912060-173912260	-	-
	end	chr1:174154095-174154295	AluJr	SINE/Alu
P29	start	chr1:173909521-173909721	L1MC1	LINE/L1
	end	chr1:173910221-173910421	AluY	SINE/Alu
P31	start	chr1:173906940-173907140	AluSp	SINE/Alu
	end	chr1:173908312-173908512	AluSx1	SINE/Alu
P32	start	chr1:173907223-173907423	-	-
	end	chr1:173907454-173907654	-	-
P33	start	chr1:173906875-173907075	AluSp	SINE/Alu
	end	chr1:173907738-173907938	AluSp	SINE/Alu
P34	start	chr1:173906875-173907075	AluSp	SINE/Alu
	end	chr1:173907738-173907938	AluSp	SINE/Alu
P35	start	chr1:173908260-173908460	AluSx1	SINE/Alu
	end	chr1:173915305-173915505	AluSz	SINE/Alu
P36	start	chr1:173914806-173915006	-	-
	end	chr1:173916822-173917022	-	-
P37	start	chr1:173905822-173906022	L2	LINE/L2
P38	start	chr1:173905822-173906022	L2	LINE/L2
P39	start	chr1:173905822-173906022	L2	LINE/L2

f. Repetitive elements and microhomology at the breakpoints

A deep analysis of the sequences of SV breakpoints obtained by nanopore, LR-PCR, and/or Sanger sequencing (N= 24 SV), or by analysis of CGH array (N= 9 SV) revealed repetitive elements in or close (< than 100 bp) to one of the breakpoints in all except one case. Alu sequences were involved in 34 breakpoints and LINE elements were involved in 14 breakpoints (Table 18 and table 19). Interestingly, some Alu elements are present in recurrent SV. Thus, an AluSp in intron 6 was present in SVs of 3 patients (P34, P33, P31); the same Alu subfamily localized in intron 5 was involved in SVs of cases P34 and P33. Finally, AluSx1 in intron 5 was involved in the breakpoint of 4 SVs (cases P21, P24, P31 and P35).
The observed recurrent breakpoints inter and intragenic fall in regions with higher density of RE (Figure 66 and Figure 67).

g. Recurrent structural variants

Interestingly, four different SVs were identified in at least two unrelated patients with antithrombin deficiency (Figures 60 and 63 and table 17).
• Three patients from 3 countries, Spain, Poland, and Belgium had tandem duplication of exon 6. Two of them (P33 and P34), have duplications with the same size (863 bp) and shared a common breakpoint, however P32, has different size (198 bp) and breakpoints.
• P3 and P4 showed a very similar deletion according with CGH array data and might share a common breakpoint. As P3 was further studied with nanopore sequencing, the nucleotide resolution obtained allowed to design a specific PCR for breakpoint validation. The PCR used to validate the breakpoint in P3 was also applied to P4, confirming that both patients had exactly the same SV. Thus, nucleotide-level resolution revealed the same 20 gene deletion in two unrelated patients (P3 and P4) from Murcia and Granada, towns of the south of Spain separated by 200 Km.
• CGH array also showed very similar data in terms of extension and localization of the deletion in two unrelated Belgian patients: P11 and P12, 534199 bp, and 534213 bp; 1:173884518-174418717 and 1:173884504-174418717, respectively. Unfortunately, no further results were obtained and the absence of nucleotide resolution did not allow confirming or refusing that these patients share the same SV.
• WGS using nanonopore and PCR validation revealed the same insertion of a new retrotransposon with the same TDS in three unrelated Spanish patients, two (P37 and P38) were from Galicia (Lugo and A Coruña, at the north of Spain) and the third (P39), was from Toledo (Center of Spain), but reported ancestors from Galicia.

In order to confirm the recurrent SVs affecting *SERPINC1* and aiming to explore a potential founder effect for these genetic changes, STR analysis of 6 markers that covered 12 Mb close to *SERPINC1* were done in all cases. The analysis of P32, 33 and 34, the cases with exon 6 duplication, revealed that P33 and P34 share common STRs in all positions; however P32, only shares the two closest STRs in the *SERPINC1* telomeric region (D1S2691 and D1S2887) (Table X.6.4). Patients with deletion of 20 genes (P3 and P4) shared common STR alleles till 1 Mb telomeric from *SERPINC1* and the closest STR to *SERPINC1*, D1S2790 (Table X.6.4). For cases with SVA insertion, STR analysis showed common STRs alleles in the three carriers in positions close to *SERPINC1* (up to 1Mb in centromeric region). Patients P37 and P38 share more STR alleles (D1S452, D1S2751 and D1S431). Additionally, phasing analysis of the sequences obtained by WGS with nanopore in two cases, confirmed the linkage of a rare SNP (rs186758691, MAF: 0.0027) to the SVA insertion shared by these patients (Table 20) However, STR analysis did not support a founder origin for the deletion found in P11 and P12 (Table 20)

Table 20: **STR analysis of recurrent SV in patients with Antithrombin deficiency.** Alleles shared by patients carriers of recurrent SV, are shown in bold letter.

					DUPLICATION EXON 6			**DELETION 20 GENES**		**SVA INSERTION**			**DELETION 4 GENES**	
STR	**PCR size**	**REPETITIONS**	**FAM/HEX**	**Distance to SERPINC1**	P32	P33	P34	P3	P4	P37	P38	P39	P11	P12
D1S431	248-292pb	23GT	FAM	-6.052.951	(GT)21	**(GT)17**	**(GT)17**	**(GT)18**	**(GT)18**	(GT)19	(GT)18	(GT)18	(GT)20	(GT)21
					(GT)19	(GT)22	(GT)19	(GT)17	(GT)15	**(GT)21**	**(GT)21**	**(GT)21**		
D1S2799	184-224	21AC	FAM	-5.780.393	(AC)20	**(AC)20**	**(AC)20**	-	(AC)13	(AC)19	(AC)16	(AC)15	(AC)19	(AC)20
									(AC)16			(AC)16	(AC)21	
D1S452	197-225	15GT	HEX	-3.349.071	(GT)12	**(GT)10**	**(GT)10**	(GT)7	(GT)9	(GT)8	(GT)7	-	(GT)10	(GT)9
					(GT)9						(GT)8			
D1S2790	205-247	22TG	FAM	-848.572	(TG)17	**(TG)19**	**(TG)19**	**(TG)18**	**(TG)18**	**(TG)19**	**(TG)19**	**(TG)19**	(TG)20	(TG)22
					(TG)21		(TG)21		(TG)22	(TG)21	(TG)21			
SERPINC1				0										
D1S2691	165-187	22CA	FAM	3.965.256	(CA)20	**(CA)19**	**(CA)19**	**(CA)10**	(CA)19	**(CA)15**	(CA)19	(CA)18	(CA)15	(CA)20
					(CA)19	**(CA)15**	**(CA)15**		**(CA)10**		**(CA)15**	**(CA)15**		
D1S2887	194-215	21TG	HEX	2454572	**(TG)16**	**(TG)16**	**(TG)16**	-	-	**(TG)16**	**(TG)16**	**(TG)16**	**(TG)16**	**(TG)16**
					(TG)18	**(TG)18**	**(TG)18**							
D1S2791	193-214	21AC	FAM	1.112.289	(AC)17	**(AC)21**	(AC)17	**(AC)15**	**(AC)15**	**(AC)17**	**(AC)17**	**(AC)17**	(AC)17	(AC)15
						(AC)18	**(AC)21**	(AC)17					**(AC)18**	**(AC)18**
D1S2751	210-228	18CA	FAM	6.124.702	**(CA)18**	**(CA)18**	**(CA)18**	(CA)15	(CA)18	(CA)18	(CA)18	(CA)15	(CA)19	(CA)18
					(CA)19			(CA)17	(CA)16		(CA)16			

h. MLPA study of INDEL variants

Currently MLPA analysis is only done in cases with negative sequencing data. However, and taking into account the failed results obtained by MLPA in some cases, and indicated before, we speculate that small INDEL might be incorrectly detected by MLPA. Thus, we studied by MLPA 23 cases with different small INDEL from the Spanish cohort. Remarkably, 2 cases (8.7%) had positive findings. Case P40, carrier of a 31 bp in-fame deletion (p.Val407_ Ala416del) was incorrectly diagnosed by MLPA as a deletion of the whole exon 7 (Figure 68). Much more interesting was the result observed in case P41, carrier of a complex INDEL with a predominant duplication of a conserved region, p.Glu241_ Leu242delinsValLeuValLeuValAsnThrArgThr, that resulted in a variant secreted to the plasma (type II deficiency) in which the duplicated region formed a new strand in sheet A, as our structural study confirmed [148] . Surprisingly, MLPA analysis of P41 showed a deletion of exon 4 (Figure 69).

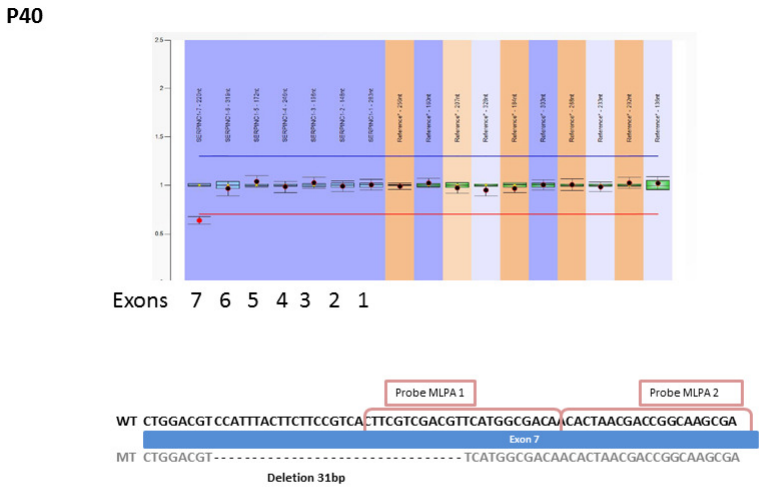


Figure 68: **Erroneous diagnosis of exon 7 deletion by MLPA in a case with antithrombin deficiency caused by a 31 bp in-fame deletion (p.Val407_ Ala416del).** The localization of the deleted region and the sequence recognized by the probes used in the MLPA are also shown.

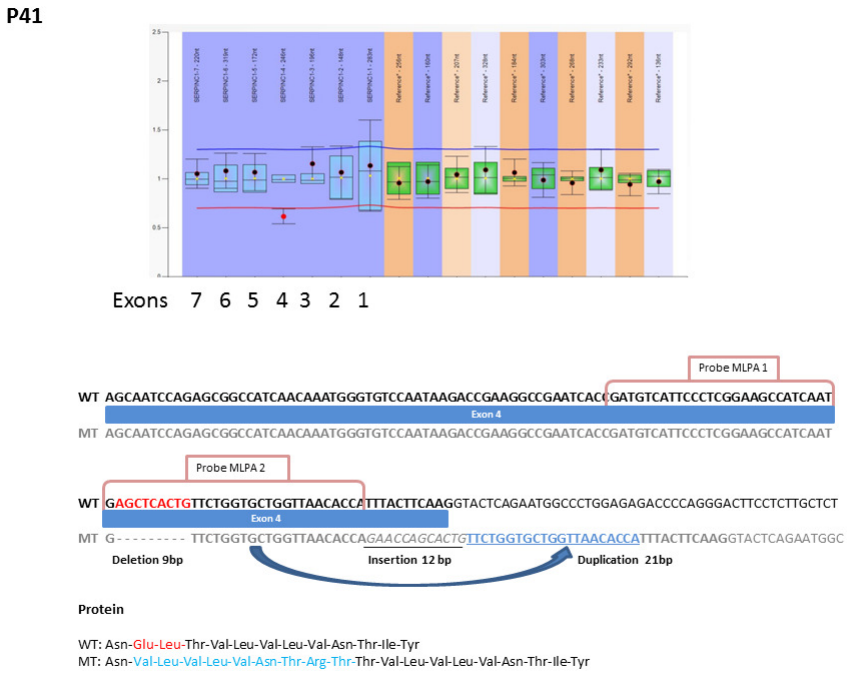


Figure 69: **Erroneous diagnosis of exon 4 deletion by MLPA in a case with antithrombin deficiency caused by a complex INDEL: deletion of 9 bp, duplication of 21 bp and insertion of 12 bp (p.Glu241_Leu242delinsValLeuValLeuValAsnThrArgThr).** The localization of the gene variation and the sequence recognized by the probes used in the MLPA are also shown.

Discussion

5.1 Clinical impact of antithrombin deficiency

Antithrombin deficiency is the strongest thrombophilia and significantly increases the risk of thrombosis. Its diagnosis, usually restricted to functional studies evaluating the anticoagulant activity of this serpin (anti-FXa< 80%), has clinical impact in the management of carriers, leading to a long-life antithrombotic treatment in symptomatic carriers; and prophylaxis under risk conditions and reducing of other thrombotic risk factors in asymptomatic carriers [149] . However, there is a significant clinical heterogeneity among subjects with a diagnosis of antithrombin deficiency. Therefore, the identification and characterization of gene defects causing antithrombin deficiency with different prognosis is relevant.

We explored this objective by two different and antagonist approaches. 1) A retrospective analysis of pediatric thrombosis in probably the world-wide largest cohort of patients with antithrombin deficiency consecutively recruited during more than 20 years in two reference centers from Spain and Belgium. For this objective, we evaluated the impact of different factors such as: age, sex, genetic variants and environmental factors on the risk to suffer pediatric thrombosis. 2) A complete biochemical and genetic characterization of the first recurrent mutation causing antithrombin deficiency in African population

5.1.1 Pediatric thrombosis.

We have performed a retrospective multicentric study, collecting phenotype and genotype data from 968 patients with congenital antithrombin deficiency, to evaluate the risk and clinical features of thrombosis in pediatric age. 73 patients from our cohort developed pediatric thrombosis, which represents the largest cohort of pediatric antithrombin deficient patients world-wide.

The calculation of the thrombotic risk in patients with severe thrombophilic disorders like deficiencies of the natural anticoagulants antithrombin, protein C and protein S is difficult because these are rare disorders. This limitation is even more prominent when considering pediatric thrombosis. In particular for antithrombin deficiency very few data are available about the occurrence of thrombosis in the first two decades and most information results from case reports or small case series [150, 151, 152] , as well as from reports on thrombophilia in large cohorts of pediatric patients [153, 154, 155, 156]

The incidence of pediatric thrombosis among our antithrombin deficient patient cohort was as high as 7.5%. We calculated an incidence of pediatric thrombosis of 0.41%/year among carriers of this severe thrombophilia, which is 300-fold higher than that described in the general population (0.0014%/year) [129] .

We observed more thrombotic complications in males than females (male-to-female ratio 1.2:1), a finding that is in accordance with previous studies in children [157, 153, 156, 158, 159]

Thrombosis in antithrombin deficient children also seems to be age-dependent. In accordance with other studies [160] , we observed a fairly consistent pattern with an initial peak incidence of thrombosis during the neonatal period and a second increment occurring in adolescence (Figure 70) .

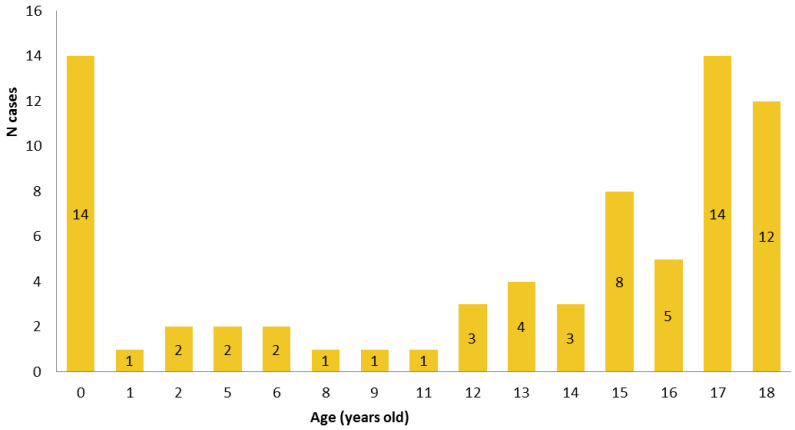


Figure 70: **Representation of the number of cases with pediatric thrombosis and the age at of the event in patients with antithrombin deficiency.**

During the adolescent period, the localization of the thrombosis and the triggering risk factors were similar to those seen in adults, notably in estrogen-related conditions (oral contraceptive pill, pregnancy and puerperium) (Figure 71) .

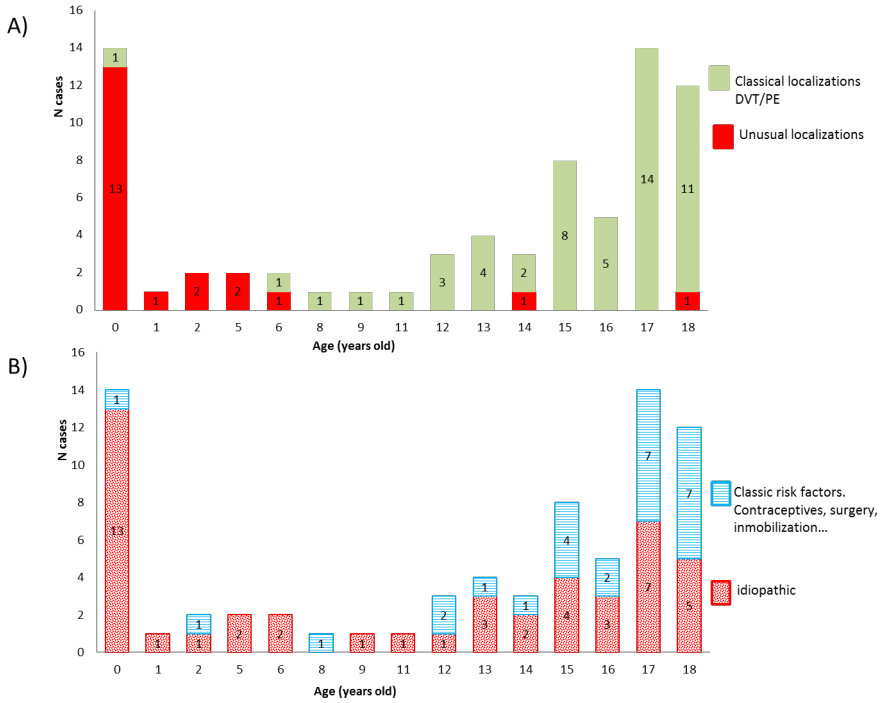


Figure 71: **Representation of A) the localization of the thrombotic event and B) the risk factor associated in each case of patients with antithrombin deficiency.**

The reason for the high incidence of thrombosis in the neonatal period can be attributed to the labile haemostatic system in newborns with reduced levels of many coagulation factors and inhibitors, including antithrombin. Antithrombin levels are known to be 60% reduced at birth and to reach adult values around 3 months of age [161, 162].

We speculate that the physiological antithrombin deficiency at birth is exacerbated by the addition of a congenital defect of this protein, making the neonate most sensitive to any other prothrombotic triggering factor like acidosis, hypoxia, thermal changes, release of tissue factor and a frequent exposure to trauma and manipulation. It is worth mentioning that only one neonate from our cohort the initial thrombosis was catheter-related, while this is the main cause of thrombosis among newborns [163]. We note a strikingly high incidence of cerebral sinovenous thrombosis in our neonatal cohort. It is plausible that the skeletal development of the skull (with fontanels), makes the newborn vulnerable to cerebral thrombosis in this period, particularly if associated with thrombophilia and/or localized trauma [164] . Indeed, three events occurred after assisted delivery, a procedure known to be associated with cerebral thrombosis in 60% of cases [165] .

Of interest is the severe outcome and mortality in our pediatric antithrombin deficient patients. Directly attributable mortality was 4-fold higher compared to children with thrombosis from the Canadian Childhood Thrombophilia registry (8.2% vs 2.2%, respectively) [166] . Two unrelated cases of our six deceased patients were carrying the p.Leu131Phe variant in homozygous state. The occurrence of pediatric thrombosis associated with this variant, antithrombin Budapest III, in homozygosity has been described earlier [151, 152, 132] but, as far as we know, never with fatal outcome. VTE recurrence rate in our antithrombin deficient pediatric cohort was markedly increased compared to that of the overall pediatric population with thrombophilia [154, 155]. All these results support that antithrombin deficiency is the most severe form of thrombophilia, both in childhood and adulthood [30] .

The risk of thrombosis in children with antithrombin type II HBS deficiency in our cohort was low, except when the mutation was present in homozygous state or when combined with other thrombophilic factors. In our whole cohort, 23% of patients were carriers of mutations affecting the heparin-binding site. However, only 3.6% of subjects with type II HBS deficiency developed pediatric thrombosis and most of them presented the deficiency in homozygous state or combined with other thrombophilia, supporting the fact that isolated heparin-binding site deficiency in heterozygous state is less thrombogenic in children than type I or the other type II deficiencies. Interestingly, there was a remarkable incidence of genetic variations associated with formation of disulphide-linked dimers (14/73 or 19.2%). These results strongly support that mutations with conformational consequences have severe clinical implications and might increase the risk of pediatric thrombosis [117, 56] .

It is still a matter of debate whether it is useful to test for thrombophilia in children with a first venous thrombotic event or in asymptomatic children of families with thrombophilia [167, 168] . The identification of an inherited thrombophilic defect does not alter the acute antithrombotic management in children [169] and it is not common practice to administer thromboprophylaxis in children in high-risk situations such as immobility, surgery, or trauma [170] . However, a recent study suggests that thrombophilia care in children should be individualized [171] . The high incidence of severe thrombotic events in children with antithrombin deficiency, most of them in high-risk situations and the high recurrence rate observed in our study support the comment of Nowak-Göttl and coworkers [171]. Thus, we recommend testing for antithrombin deficiency in children of affected families, particularly for those carrying type I deficiency. Carriers might benefit from preventive strategies like prophylaxis in high-risk situations [172, 173, 174] and from counselling concerning risk factors such as oral contraceptive use. Our data also propose to test for antithrombin deficiency in pediatric cases with cerebral sinovenous thrombosis or thrombosis occurring at unusual sites.

5.1.2 Characterization of the first recurrent mutation causing antithrombin deficiency in African population

We have evaluated the consequences and features of the p.Thr147Ala antithrombin variant identified in 12 unrelated patients with antithrombin deficiency of Black-African origin.

The majority of the patients were female (11/12) and suffered from obstetric complications (7/12). Thrombotic events were observed in six of the 12 patients, in half of them in the absence of acquired or transient thrombophilic risk factors. The predominance of the p.Thr147Ala variant in female patients is remarkable as it is not observed in the data from gnomAD, ExAC, and 1000 Genomes. The gender bias observed in our study can be explained by the fact that females, unlike males, are exposed to specific thrombotic risk factors as pregnancy or hormone replacement therapy, among others. Further studies are required to clarify whether the p.Thr147Ala variant is particularly associated with obstetric complications. Familial history of VTE was reported by only one patient, suggesting that this variant is associated with a mild thrombotic phenotype. In all subjects, no other molecular defects were identified in the *SERPINC1* gene. Picard et al reported a female patient with antithrombin deficiency carrying a different mutation on the same amino acid residue, p.Thr147Ile. In this patient, extremely low activity was measured (25%) in combination with an antigen level of 50%. The possible hypothesis of homozygosity or compound heterozygosity could however not be proven. The patient and her twin sister suffered from spontaneous abortion, a clinical presentation that is shared by four of the patients reported in our study [175] . Further studies including large case/control studies of patients with African origin are required to explore the risk of thrombosis associated to this *SERPINC1* mutation, particularly to know if whether mutations affecting this residue could increase the risk of abortion and to know the associated mechanism. In this study we characterize this variant aiming to explain why it is so frequent in a specific population.

Our first concerns were to clarify the pathogenicity of the p.Thr147Ala variant. Our initial approach was bioinformatics. Seven of the 13 in silico prediction tools for pathogenicity classified this variant, also known as SNP rs2227606, as a benign polymorphism. A recent study evaluating different *in silico* methods in predicting pathogenicity of *SERPINC1* variants observed a large variability between several bioinformatics tools. Especially mutations causing type II HBS deficiency were often misclassified as benign by the majority of the *in silico* tools [176] . The most powerful prediction tool in their study appeared to be PolyPhen-2, which classified the p.Thr147Ala variant as "possibly damaging." However, for the same algorithm, the p.Thr147Ala variant was regarded as a "true negative," meaning that data on specificity and negative predictive values should be interpreted with caution. Another study found that predictions from five commonly used algorithms (PolyPhen-2, SIFT, CADD, Provean, and MutationTaster) resulted in 79% concordance for pathogenic variants and only 33% for benign variants [177] . Of these five algorithms, only SIFT predicted our variant to be tolerated.

The second interesting issue of this variant was the conflictive results observed when evaluating the functional consequences of this mutation and to classify the associated deficiency. Antithrombin activity in patient plasma was reduced only when measured with one specific commercial assay, Innovance Antithrombin. Combined with (nearly) normal antigen levels, these results suggest a type II antithrombin deficiency. Our study provides further evidence of the limitations of commercially available methods to diagnose antithrombin deficiency. Indeed, it has already been reported that the sensitivity of certain commercial activity assays, especially thrombin-based ones, tends to be lower toward type II HBS deficiencies [133, 116, 178, 179, 180] . Reduction of incubation time of the plasma with thrombin or FXa can improve the sensitivity [116, 181, 182] . This was also the case in the present study where we observed a significant reduction of antithrombin activity (measured by the anti-Xa assay HemosIL Antithrombin) in the plasma from patients but not in normal controls. Heating of the patient plasma did also induce a reduction of the antithrombin activity when measured with the same assay.

Additional evidence for the pathogenic consequences of p.Thr147Ala was obtained from the studies done on recombinant proteins. The recombinant p.Thr147Ala showed approximately 50% anti-FXa inhibition of the WT antithrombin but without altered heparin affinity, as shown by fluorescence studies. The fact that we could not show any effect on heparin affinity in the recombinant protein could be explained by the use of the β-glycoform background. Two antithrombin glycoforms are found in plasma: first, the α-isoform that represents the main circulating fraction (90%) and contains four N-glycans; second the β-isoform that lacks the N-glycan at position135

is the minor fraction (10%). Of these two glycoforms, β-antithrombin has the higher affinity for heparin and thus functions as the major inhibitor in vivo even though it is the less abundant form [183] . A previous study has shown that the β-isoforms of HBS variants retain their heparin affinity and may also compensate for the effect of mutations causing heparin binding defects [184] . In our study, the recombinant proteins were constructed on the β-isoform, possibly explaining the preserved heparin affinity. Moreover, the anti-Xa activity of the recombinant p.Thr147Ala protein was reduced with both assays, contrasting with the data observed in plasma. When measured with the HemosIL Antithrombin assay, plasma of patients carrying the p.Thr147Ala variant shows normal antithrombin activity, while recombinant, pure β-p.Thr147Ala displays reduced antithrombin activity. This suggests that the effect of the p.Thr147Ala variant on the anticoagulant function of antithrombin is more pronounced on the β-glycoform. Structural modeling revealed that residue Thr147 forms three hydrogen bonds that are all abolished when mutated to Ala. The H-bond with Arg79 is of specific interest as this residue is known to directly interact with heparin and mutations of this residue result in a loss of affinity toward heparin [185, 186]. We speculate that the absence of this hydrogen bond causes a distortion of the HBS and possibly leads to a misalignment of the pentasaccharide. This is further supported by the results of the CIE. Under physiological conditions, no difference is observed between normal and patient plasma. However, under higher ionic strength conditions, an increase in low-affinity variants is observed. Hydrogen bonds are known to be affected by higher ionic strength and the fact that the p.Thr147Ala mutation already induces a loss of three H-bonds only exacerbates this effect. Another stress condition known to disturb hydrogen bonds is temperature. As already discussed, heating of the patient plasma reduced the (apparently normal) activity compared with controls when measured with the HemosIL Antithrombin kit. We speculate that the reaction conditions (pH, incubation time, heparin concentration) of the Innovance Antithrombin kit, mimic a stress condition making it the only commercial assay in this study able to detect the functional (and pathological) consequences of the p.Thr147Ala variant. Moreover, this assay uses human FXa in contrast to the other anti-Xa assays in this study where bovine FXa is used.

Threonine 147 is located at the bottom of helix D, a region highly involved in binding to the pentasaccharide of heparin. Upon binding of antithrombin to heparin, several structural rearrangements occur, like the rotation and extension of helix D [186] . Mutations in amino acids sterically surrounding residue 147 have been proven to highly disturb or even abolish heparin affinity [187, 188] . This effect might be exacerbated in the absence of glycosylation at Asn167, which is the result observed in the recombinant system using a beta background.

Interestingly, the p.Thr147Ala variant was only identified in individuals of Black-African origin. A potential founder effect is usually evaluated by haplotype analysis of carriers of a common mutation. Usually, haplotype studies required from family studies or cloning strategies to strongly determine the haplotype linked to the mutation. Unfortunately, no family studies were possible, and we decided to use a novel approach that might solve the haplotype associated with the p.Thr147Ala variant in all unrelated carriers. LR-PCR and nanopore sequencing, together with a simple bioinformatic tool, allows knowing the genetic variations of each allele. By using this original approach with nanopore sequencing of LR-PCRs we demonstrate a common haplotype in all carriers. This haplotype contains 13 intragenic markers, some of them with very low MAF, strongly supporting a founder effect. So far only two founder mutations have been described in *SERPINC1*, both also responsible for type II HBS deficiencies: Antithrombin Budapest 3 (p.Leu131Phe) in Hungary and antithrombin Basel (p.Pro73Leu) in the Finnish population [132, 133] . We report the first *SERPINC1* founder mutation identified in the Black-African population. Further studies are required to determine at what evolutionary time point this variant, which has a relatively high prevalence in Africans, has occurred. According to the 1000 Genomes data obtained from the African/African-American population dataset of 5,203 individuals (3,315 female and 1,888 male samples) assembled by ExAC, the frequency of the p.Thr147Ala variant in African population is close to 1%, strongly suggesting that it originated a long time ago during evolution and has a wide distribution in this continent. Our own data confirm the prevalence of 1% in individuals of Black-African ancestry. Although the antithrombin p.Thr147Ala variant is predominantly

found in Africans, it was also identified in African-Americans and Europeans, but in much lower numbers. This can be explained by the recent acceleration and geographical diversification (beyond colonial patterns) of migration from Africa to Europe, but increasingly also to North America [189] . Results from a recent meta-analysis show that individuals of African descent appear to have a higher risk for the development of VTE when compared with Caucasians (odds ratio: 1.28, 1.19-1.38 95% confidence interval) [190] . While the factor V Leiden and prothrombin G20210A variant are absent in people of Black-African ancestry, the prevalence of deficiencies of the natural anticoagulants, protein C, protein S, and antithrombin is reported to be equal in both ethnic groups [191, 192, 193] .

It was hypothesized that the higher prevalence of VTE among people of African ancestry might be due to an increased prevalence of acquired risk factors for VTE but this could not be confirmed [191]. The authors of the same study suggest the underlying role of yet unknown genetic components. In the light of our present findings, we presume that the p.Thr147Ala variant in the *SERPINC1* gene could be such a candidate to contribute to the higher risk of thrombosis observed in Africans. Moreover, this variant is hardly detected by commercial assays evaluating antithrombin anticoagulant activity and thus probably under reported. Interestingly, the *PROS1* mutation, p.Val510Met, identified in patient 10, was shown to be a population-specific risk factor for VTE, restricted to individuals of African descent [194] . Analogous to the hypothesis of a positive selection explaining the high prevalence of common prothrombotic polymorphisms like factor V Leiden among Caucasians [195], we speculate that in Africans other mild prothrombotic mutations with possible health benefits might be selected. These polymorphisms, including the p.Thr147Ala variant in *SERPINC1* for the African population, could predispose to different advantageous phenotypes. Beneficial effects of factor V Leiden on hemostasis (less hemorrhagic events during delivery), on inflammation (higher survival rate in and lower susceptibility to severe sepsis), and even on fertility (higher fecundity in both men and women) could underlie the selection [195] . It is hypothesized that these beneficial effects might not be driven by the factor V Leiden mutation itself but rather by (not yet identified) polymorphisms in surrounding genes that are co-inherited with factor V Leiden (a mechanism called genetic hitchhiking) [196] . As *F5* and *SERPINC1* are located in close proximity on the long arm of chromosome 1, it is conceivable that the observations for the factor V Leiden variant can be extrapolated to the antithrombin p.Thr147Ala variant and thus support its positive selection. However, in the modern and aging Western society with increased risk of thrombosis, such as oral contraceptive use or sedentary habits, these originally "positive" defects would show their pathogenic reverse side, revealing new evidence of the complex and sensitive equilibrium of the haemostatic system.

This study has investigated the p.Thr147Ala variation in the *SERPINC1* gene. We could demonstrate that the variant is a pathogenic founder mutation responsible for an inherited type II HBS antithrombin deficiency. As the mutation is restricted to people of African descent, further studies including large case-control studies in populations of African ancestry are required to determine the exact risk of thrombosis and bad pregnancy outcome associated with this variant.

5.1.3 Conclusion from the first objective

The first two chapters of this thesis show new evidence of the clinical heterogeneity of antithrombin deficiency. Some types, such as type I or certain type II deficiencies (homozygous for the Budapest 3 variant, or heterozygous in combination with other prothrombotic risk factors) are very severe prothrombotic risk factors and may cause pediatric thrombosis, in some cases with fatal consequences, affecting unusual localizations, or with recurrence. In contrast, there are mild antithrombin deficiencies, like the p.Thr147Ala variant, with moderate functional consequences that some commercial methods are not able to detect, that may expand in certain populations. For the p.Thr147Ala variant, we have used a novel method based on sequencing large fragments of DNA by nanopore that

is able to define large haplotypes. This method has been useful to determine the founder effect for the p.Thr147Ala variant.

5.2 Structural variants in antithrombin deficiency: new molecular defects involved in this disorder

The key anticoagulant role of antithrombin and the clinical usefulness of its diagnosis have stimulated the analysis of this disorder among patients with thrombophilia. Hundreds of cases with antithrombin deficiency have been identified, and their molecular base studied, mostly focused on the coding gene, *SERPINC1*. Thus, sequencing of exons and flanking regions and MLPA analysis have identified up to 315 different *SERPINC1* gene defects.
As far as we know only 3 regulatory gene variations affecting the promoter or introns have been described [67, 66] . NGS methods covering coding regions or the whole gene, less frequently used, may also identify mutations in *SERPINC1* involved in antithrombin deficiency [72] .
In addition, gross gene rearrangements affecting *SERPINC1* have been identified in cases with antithrombin deficiency, although these gene defects are relatively rare in this disorder [72] . In fact, only 2 duplications and 25 deletions of more than 50 bp are reported at the HGMD (HGMD® Cardiff University 2017). Moreover, SVs involved in antithrombin deficiency have not been deeply characterized. In many cases even the extension of the gross gene defect affecting *SERPINC1* has not been determined, and only few cases have been described at nucleotide level.
It is also important to note that up to 20-30% of cases with antithrombin deficiency have no *SERPINC1* defects detected by these methods (Figure 72). Although other genes may be involved in antithrombin deficiency, as those coding for enzymes that play a role in the N-glycosylation pathway [86] , due to the relevance that this post-translational modification plays in the secretion of this anticoagulant (de la Morena-Barrio et al in preparation), we speculated that other *SERPINC1* defects, such as large gene defects affecting introns, not detected by current routine methods might cause antithrombin deficiency.

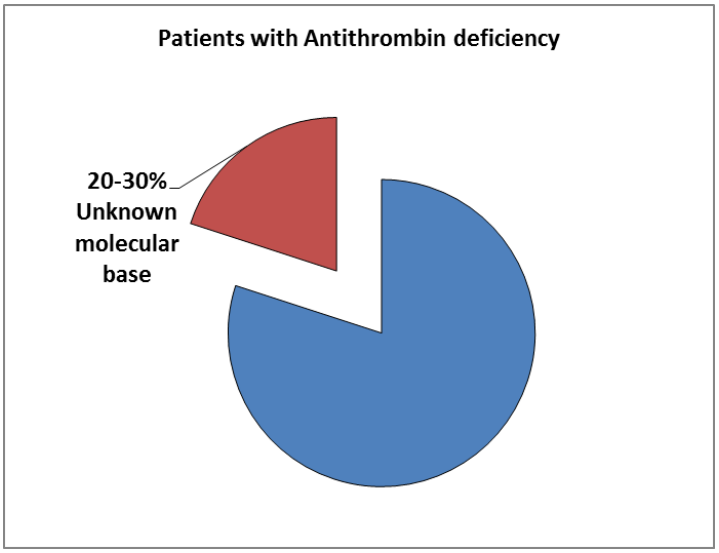


Figure 72: **Graphic representation of patients with antithrombin deficiency with molecular diagnosis (blue) and unknown molecular diagnosis (red)**.

The interest for the optimization of diagnostic methods for SV detection and the discovery of new SV involved in antithrombin deficiency and in any other diseases is because SVs account for a greater number of variable bases than SNVs [197]. Thus, SV has more likely possibility to be pathogenic than a SNV [101] . Indeed, SVs have been associated with many diseases, both germ line and somatic.

5.2.1 Detection of tandem duplication of exon 6 by a specific detection method

The identification of the duplication of exon 6 of *SERPINC1* as a genetic defect involved in antithrombin deficiency was obtained by chance. The modification of the primers used to amplify *SERPINC1* exons was able to detect the first case, and it was possible because the duplication was small (only 193 bp). However, this identification allowed to design a specific method that only detected tandem duplication of exon 6, which allowed to identify the second (and as we will discuss later a third) case, with longer duplication (863 bp). Thus, the study of a large cohort of patients with antithrombin deficiency has revealed that tandem duplication of exon 6, which is hardly detected by current molecular methods, represents one third of gross gene defects associated to antithrombin deficiency, and it is present in up to 1% of patients with antithrombin deficiency. The location of primers used for PCR amplification, usually close to the exons and the length of the duplication may make difficult the detection of duplications involving this exon by classical PCR amplification and sequencing. Moreover, one massive method of sequencing based on small amplicons (Ion Torrent) also failed to detect this genetic defect. Additionally, the small size of this exon (only 65 bp) may also explain the technical difficulties to detect the duplication (and potentially also the deletion) of exon 6 by MLPA.
We have developed in this study a simple method that specifically detects tandem duplications of exon 6 by a single PCR which encourage a new algorithm for the molecular characterization of antithrombin deficiency, as we will discuss later, that guarantee a positive finding in up to 85% of cases and rationalize the cost of this analysis.
The duplication of exon 6 disturbs the genomic organization of *SERPINC1*, potentially modifying the correct splicing process, that could lead to unstable RNA and/or premature stop codons and the associated antithrombin deficiency (Figure 73).

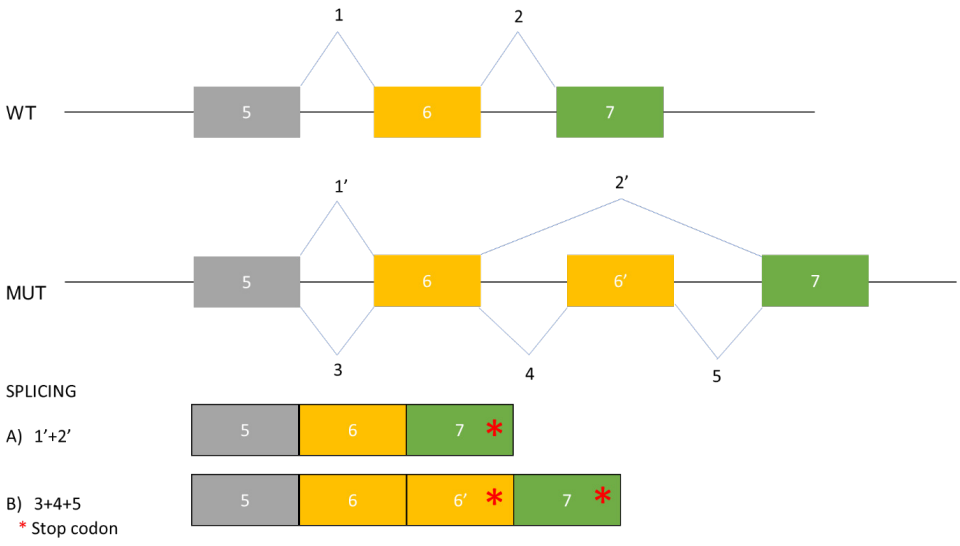


Figure 73: **Schematic representation of *SERPINC1* gene organization of wild type and tandem duplication of exon 6 alleles, showing the predictive splicing that might be generated in each case.** * represents stop codon.

The pathogenic consequences of the duplication of exon 6 were validated by segregation analysis done in the two Spanish families. All 6 carriers had antithrombin deficiency. However, we were not able to verify the potential consequences of this SV on RNA as this gene is of hepatic expression. We treated to amplify *SERPINC1* RNA from peripheral blood, but even using high amounts of cells and fresh RNA, no amplification of wild type *SERPINC1* RNA from healthy blood donors was obtained.

It is possible that certain heterogeneity of the pathogenic consequences of the tandem duplication of exon 6 could be found, particularly for the small duplication, as the duplicated exon 6 might be skipped and a normal RNA might be generated (Figure 73).

Why duplication of exon 6 may be a frequent SV in antithrombin deficiency? The high concentration of Alu sequences surrounding this exon probably explains this possible *hotspot* of SV in *SERPINC1*, as we will discuss latter. Importantly, the finding of the tandem duplication of exon 6 in patients with antithrombin deficiency opened the possibility of other *SERPINC1* defects not detected by current methods to be involved in cases with this disorder and still unknown molecular base, particularly, SVs affecting introns.

5.2.2 Detection of deletion of intron 1 by LR-PCR

According to the conclusion of the previous point, we aimed to detect SVs involving intronic regions by using new methods that might bypass the limitations observed by the use of short amplicons. LR-PCR was a relatively easy cheap method that might be success to identify small SVs affecting introns, especially deletions. Thus, we used LR-PCR in 36 cases with antithrombin deficiency and no genetic defects detected. Interestingly, this approach revealed one case with evident congenital defect due to the family history of antithrombin deficiency that had a 2 Kb deletion affecting intron 1. This result also explains the negative results obtained by MLPA and CGH array. Moreover, this result together with the relative simplicity of LR-PCR of *SERPINC1* (it does only involve 4 amplicons) support the usefulness of this method to be used in the identification of gene variations involved in antithrombin deficiency.

Deletion of intron 1 has been described in other genes with pathological consequences such as osteochondromas, hypertension or alveolar capillary dysplasia, as usually, intron 1 is large and contains regulatory elements. [198, 199, 200, 201] . Additionally, a recent study reported that duplication of promoter, and exon 1 of *F8*, lead to elevation of FVIII levels and severe thrombophilia in two families. [27] . Intron 1 of *SERPINC1* also contains potential regulatory regions, including vitamin D response elements (VDRE), whose distortion causes antithrombin deficiency [66] . Deletion of these regulatory elements might affect the transcription rate of *SERPINC1*. The deletion detected in the present case involves one VDRE element but also covers the acceptor splice site sequence of exon 2 as well as the initial 13 nucleotides of this exon (Figure 74). Therefore, the mechanism explaining type I deficiency associated to this deletion could be an aberrant splicing, a pathogenic mechanism that is probably underestimated [202] . Two possible cryptic acceptor signals in exon 2 could be used and the skipping of exon 2 can also be considered (Figure 73). Independently of the alternative splicing followed, all of them might cause a frameshift and a premature stop codon that could explain the type I deficiency identified in carriers (Figure 73). Unfortunately, due to the nearly null expression of *SERPINC1* in accessible tissues, it was not possible to confirm this aberrant splicing at the mRNA level.

This result shows the first evidence on the existence of gross gene defects affecting introns of *SERPINC1* that are involved in antithrombin deficiency and illustrate the limitation of current methods used to identify the molecular basis of this disorder. Actually, Sanger or NGS sequencing, MLPA, or CGH arrays probably fail to detect structural variants affecting introns.

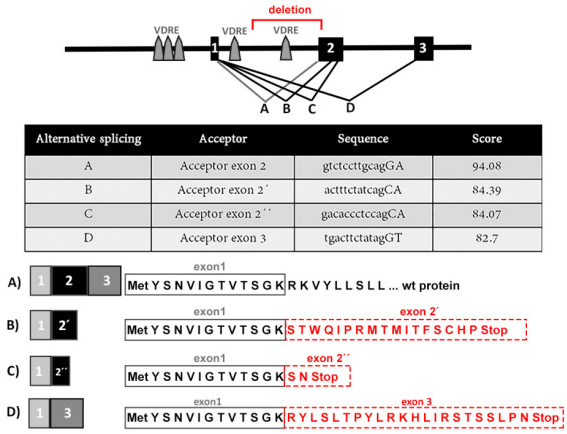


Figure 74: **Schematic representation of *SERPINC1* architecture indicating the localization of the deletion and the VDRE.** The wild type splicing (A) and potential alternative splicing induced by the deletion (B, C, D) are shown. The deletion of the wild type acceptor sequence of exon 2 might favor the use of cryptic acceptor sequences in exon 2 (B, C) or exon 2 skipping (D). The score of each splice site and the resulting aminoacid sequence of each splicing is also shown.

5.2.3 Use of nanopore sequencing for SV detection

However, the difficulties in the detection of SV are not only present in the SV affecting only intronic regions, but also all the spectrum of SVs due to due to the methodology used for their detection and the fact that they are usually involving repetitive elements which encompass over half of the human genome (centromeres, telomeres, and genes with high concentration of repetitive elements) [203] . Recent studies of even highly contiguous de novo assemblies revealed that short-read sequencing misses 40-80% of SVs [140]. Methods both specific of locus or genome wide (MLPA, fluorescence in situ hybridization -FISH-, karyotype, comparative genome hybridization, etc.) may detect SVs but have poor accuracy and do not fully characterize the SV [107] . Recently, long-read sequencing, able to generate reads up to 2 Mb, has emerged as a promising technology that might fully characterize most SVs [204] . For that reason, we used nanopore sequencing for the analysis of SVs involved in antithrombin deficiency. The results obtained by nanopore sequencing show that this technique can be used to resolve SVs causal of antithrombin deficiency, independently of its length (from 2,5 Kb to 968 Kb) or type (deletions, duplications or insertions). It also supplies the exact extension and sequence of the breakpoint of the SVs involved in antithrombin deficiency, and is the method that has clarified the conflictive data obtained by other methods.

5.2.4 Detection of Complex SV

Thus, only nanopore sequencing has been able to detect the first germline complex rearrangement involved in antithrombin deficiency. Complex SVs are genomic rearrangements comprising multiple SVs, typically involving two or more breakpoint junctions. Thus, they have more chance to disturb a gene with pathological consequences. However, because complex SVs are not typically considered in research and clinical diagnostic pipelines, only few complex SVs have been identified as cause of different Mendelian disease [139] . Therefore, complex SVs may be misdiagnosed by other techniques, as we have seen in this case, and may us think that they could be underestimated. The diagnosis of exon 1 deletion obtained by MLPA is incomplete, like the deletion of exons 1 and 2 resulted from

LR-PCR. Only nanopore sequencing was able to show the real *SERPINC1* organization, revealing the final result of the complex SV that was generated by two successive SV: duplication of exons 2 and 3 and deletion of exons 1 and 2. Unfortunately, we only get the final gene architecture resulting from a complex rearrangement, but the sequential steps may only be speculated, and much more importantly, we do not know whether this complex SV was formed by one or two independent mutational events.

5.2.5 Detection of retrotransposon insertion

The most relevant finding of our study was the identification of a new molecular mechanism involved in antithrombin deficiency: the insertion of a large mobile retrotransposon in an intron of SERPINC1 rich in repetitive elements. The most probable pathogenic effect of this insertion is the transcriptional interference of SERPINC1, as reported for many other examples of retrotransposons [205] . Unfortunately, this effect was not demonstrated because the specific hepatic expression of *SERPINC1*. The localization of this insertion also explains why this genetic defect has not been detected using current methods of molecular diagnosis.
An additional bonus of our study was the characterization of a new retrotransposon element. These results highlight the heterogeneous genomic landscape of SVA sequences and underscore the importance of their characterization in order to obtain a reliable catalogue of novel mobile elements to identify and interpret this type of causal variants in other patients and other disorders where retrotransposon insertions might also be involved [145, 206, 69]. This characterization has been historically challenging by the application of classic technologies, but here we show that it can be achieved by *de novo* assembly of long-reads.
This finding also enlarge the panel of pathologies caused by retrotransposon insertion, which together to the difficult diagnosis, strongly suggest that this mechanism may be underestimated and could be involved in many genetic diseases.
Finally, the identification of the same retrotransposon, with the same TSD, in two unrelated families, does not only support the germline transmission of this SV, but also suggest a shared mechanism of formation or a founder effect, which has been confirmed by different strategies (STR and polymorphism analysis), and allowed to design a specific PCR for detection of this retrotransposon insertion, which was discovered in a third patient with antithrombin deficiency of unknown cause by current methods of molecular analysis.

5.2.6 Common mechanism detected in SV causing antithrombin deficiency

A deep characterization of all the SV causing antithrombin deficiency from this cohort was possible by the nucleotide-level resolution achieved nanopore sequencing. The accuracy of the sequence obtained by nanopore was validated by Sanger sequencing of PCRs flanking all breakpoints, whose primers were also designed by the sequence supplied by nanopore. Specific mutational signatures can yield insights into the mechanisms by which the SVs are formed. In these cases we observed that all the breakpoints involved repetitive elements, suggesting a possible common mechanism that was further analysed with the whole cohort of SV.
Thus, encouraged by these results and aiming to explore different methods that have been used to identify and characterize SVs involved in antithrombin deficiency, we extended the study of all SV detected in patients with antithrombin deficiency recruited in our cohort by different techniques. With this study, we also wanted to describe specific and common features of SVs causing antithrombin deficiency.

5.2.7 Dissection of SV in SERPINC1

We have recruited the worldwide largest cohort of patients with antithrombin deficiency caused by SVs that almost double the number of cases described in HGMD (N= 39). Besides, we have used of a wide range of different methods able to detect and characterize SVs, allowing a comparison between them in terms of efficacy and precision

in detecting SVs.
The study of the total cohort of SV affecting *SERPINC1* confirms that SVs affecting *SERPINC1* are all deleterious, causing a type I deficiency that associates with severe clinical features (early and recurrent thrombosis). From a molecular point of view, we confirmed that most SVs causing antithrombin deficiency are deletions affecting exons of *SERPINC1*, but our data remarks that SVs involved in antithrombin deficiency may be highly heterogeneous, as tandem duplications, deletion of introns, or retrotransposon insertions may also be gross genetic defects causing antithrombin deficiency. Despite not being identified in our study, probably by the design used to recruit the cases evaluated, we can expect that other genetic defects rising SVs, such as chromosomal translocations or inversions, might be involved in antithrombin deficiency. Indeed, these SVs are not detected by most methods currently used to molecularly characterize antithrombin deficiency sequencing (Sanger or NGS), MLPA, or CGH array. We speculate that the analysis by nanopore sequencing of new cases with antithrombin deficiency with negative results by Sequencing and MLPA might identify new SVs, inversions, and balanced chromosomal translocations, to be involved in this disorder.

- *Size and type of SV affecting SERPINC1* The size of the SV involved in antithrombin deficiency is also greatly heterogeneous, ranging from 193 bp to 8 Mb, and in many cases of our cohort (21/39, 53.8%) involves neighboring genes. This multilocus affection however does not usually have clinical implications apart from antithrombin deficiency, and its risk of thrombosis. This may be explained because all our patients were selected from haematology or gynecological services, who identified them after thrombophilia screening, in most cases because a personal history of thrombotic events. The best examples are three patients (P1, P2 and P16), who despite having deletions covering more than 20 genes only had thrombosis as the main clinical sign. This may be explained probably because the genes surrounding of *SERPINC1* would require recessive mechanisms to have clinical consequences. Only extremely large deletions affecting *SERPINC1* have syndromic characteristics, like those causing the interstitial 1q deletion syndrome, a rare condition with heterogeneous 1q monosomy (1q21-q25, 1q25-q32 and 1q42qter) that have Dwarfism, severe mental retardation, microcephaly, short neck, and of course, antithrombin deficiency and thrombosis among the associated clinical features [207] .

- *Common mechanism in SV affecting SERPINC1* Our study revealed, as suggested in the previous results, common characteristics of SVs affecting *SERPINC1*: All SVs but one, have repetitive elements and microhomology involved in their breakpoints. These specific mutational signatures can yield insights into the mechanisms by which the SVs are formed.
  Thus, these findings suggests a replication-based mechanism (such as BIR/MMBIR/FoSTeS) underlying the formation of SVs affecting this gene causing antithrombin deficiency. Our study is supported by previous studies evaluating SVs in patients with antithrombin deficiency [208], particularly the excellent study of Picard et al [209], that said that this is the main mechanism of SVs affecting *SERPINC1* associated to antithrombin deficiency. Besides, our study supplies a new evidence for the higher chance of SVs to appear in repetitive rich regions of the genome [210, 211] .
  We observed a non-random formation of SVs driven by the presence of repetitive elements . Inside *SERPINC1* our data point to *hotspots* in introns 6, 5 and 1, corresponding to Alu sequences of the same family (AluSp in intron 6 and 5 and Alusx in intron 5 and 1) (Figure 67 and Table 18). This explains why exon 6 is recurrently involved in SVs causing antithrombin deficiency. Besides, in the total cohort of SVs, Alu element was the most frequently involved in breakpoints. The *SERPINC1* gene is rich in Alu repeated sequences, with nine complete and one partial repeats in its intronic regions that represents about 22% of the *SERPINC1* intron sequence, which is considerably greater than the estimated 5% of the human genome accounted for by Alu repeats [63] . Interestingly, exon 6 is surrounded by five out of nine complete and one partial repeats Alu

sequences (Figure 11) [63] .

Besides, it has been suggested that repetitive elements may provide larger tracks of microhomologies, also termed 'microhomology islands', that could assist strand transfer or stimulate template switching during repair by a replication-based mechanism [212] . These microhomology islands are also observed in intergenic regions of *SERPINC1*, involving 4 genes (*ZBTB37, RABGAPL1, RC3H1* and *DARS2*). These findings highlight the important role that repetitive elements play in the formation of non-recurrent, but non-random, SVs affecting *SERPINC1* and probably other genes rich in repetitive elements (Figure 66).

Additionally, we have observed also a sight tendency to deletions (involving the whole gene or partial *SERPINC1* deletions) going to the telomeric genomic regions (Figure 60 and Table 17). Further studies are required to know the explanation to this finding. Independently of the mechanism, our study has defined a map of regions involved in SVs that may help for the complete characterization of these gross gene defects in patients with antithrombin deficiency. The identification of a complex SV in one case, and two cases with mosaicisms of a large deletion covering a germline deletion, strongly supports multiple recombinant events involving repetitive elements,even in a single patient. This finding, together with the fact that two SVs were de novo, not present in the parents of the probands (data not shown), reinforces our conclusion that this region of the genome is prone to suffer recombinations with pathological consequences. Once generated, the SV may reach germline distribution and thus, it can have mendelian heritage to the following generations (in our cohort at least 26 SVs had family segregation), and they could expand. Actually, we show data (nucleotide sequence, STR and haplotype results) supporting a founder effect for three SVs, the insertion of a retrotransposon, duplication of exon 6, and a 20 genes deletion, causing antithrombin deficiency.

5.2.8 Comparison of different diagnostic methods

Along all the study, different techniques have been used for the detection and characterization of SVs affecting *SERPINC1*. We started with a specific PCR for a precise SV (duplication of exon 6), following with the strategy of 4 long amplicons amplified by LR-PCR and sequenced by Illumina, that allowed the detection of an intronic deletion. Among cases with MLPA positive results , we have used CGH array for validation and knowledge of the SV extension. And finally we used nanopore sequencing for unknown cases and breakpoint characterization of SVs detected by other methods not reaching nucleotide resolution (MLPA or GCHa).

First of all, we must point out that all methods have strengths and limitations. Thus, MLPA may detect defects affecting only one exon, but it does not detect SVs affecting only introns, it does not determine the extension of the SV and does not reach nucleotide resolution, which may lead to some mistakes. CGH array may give an approach to the extension of the SV, but this is not very exact, as we will discuss, it does not reach nucleotide resolution, and finally it does not detect SV smaller than 50 Kbp or SVs not causing gain or loss of genetic material (inversions, balanced chromosomal translocation).

In our experience the strongest method to identify and fully characterize SV affecting *SERPINC1* in cases with antithrombin deficiency, and probably any gene and genetic disease, is nanopore sequencing. Recent studies also support the power of long-read whole genome sequencing for the study of SV [211] . The limitation of this methodology is the high price and low coverage. PromethION devices, which are expense, are required to reach a 15x coverage by using genomic DNA and cost over 2000 €/sample. We are currently working with two procedures of enrichment that might allow a reasonable coverage of selected regions, but using a MinION, a device that may be used in practically any laboratory, and with low cost (<400 €).

Moreover, we have detected some discrepancies between different techniques that have leaded us to an interesting discussion about the most appropriated diagnostic algorithm for SV in *SERPINC1*.

a. Discrepant results between previous techniques (MLPA, LR-PCR, Ion Torrent)

1. Duplication of exon 6 Primers used for routine PCR usually are designed close to the exons to cover the coding and flanking regions. This design has been demonstrated to be unable to detect the tandem duplications of exon 6, which is relatively frequent in antithrombin deficiency. The redesign of the primers used for exon 6 was able to identify one case with a small duplication, but failed to detect the larger duplication. Moreover, one massive method of sequencing based on small amplicons (Ion Torrent) also failed to detect this genetic defect. Additionally, the small size of this exon (only 65 bp) may also explain the technical difficulties to detect the duplication (and potentially also the deletion) of exon 6 by MLPA. We designed a specific primer for detection tandem duplications of exon 6, which allowed the identification and characterization of new cases with this gross gene defect among cases with no previous mutation detected.

2. Deletion of intron 1 This SV was detected by LR-PCR but it is also interesting to point out the slight discrepancy of the result generated by CLC analysis of the sequence of the LR-PCR and the sequence obtained by specific amplification. The simple 1,941 bp deletion predicted was indeed a complex deletion with two breakpoints in intron 1 (Figure 49). Therefore, validation of NGS data by Sanger sequencing is highly recommended.
   We also tested this SV by nanopore sequencing using the MinION device. The results obtained by nanopore sequencing were identical to that obtained by Sanger sequencing (Figure 64).

3. Complex SV In one case, P1 of Chapter 4.2.3, several techniques rendered an incomplete diagnosis of the SV causing antithrombin deficiency: MLPA detected a deletion of exon 1, but LR-PCR followed by NGS found a deletion of exons 1 and 2. Only nanopore sequencing was able to give the real defect as it resolves the genetic architecture of gene at nucleotide resolution confirming a complex SV (Figure 55). CGH array also performed in this case, failed to detect the SV, as its size was smaller than 50Kb.

4. Retrotrasposon insertion Retrotransposon insertions were not detected by Sanger, LR-PCR and Illumina sequencing, MLPA, CGH array, or NGS of the whole gene. Only nanopore sequencing was able to detect this particular SV and this method allowed designing a specific PCR amplification, which was used to validate the results and to screen new cases.

b. Discrepancies of MLPA

Specific discrepancies of MLPA and nanopore sequencing detection.
We present three examples of an incorrect /incomplete diagnosis by MLPA.

1. In case P6 of chapter 4.2.3, MLPA detected a duplication of exons 1, 2 and 4 and a deletion of exon 6. However, nanopore sequencing revealed a tandem duplication of exons 1 to 5, which was confirmed by LR-PCR (Figure 55).

2. In case P18 of chapter 4.2.4, MLPA detected a deletion of exon 1 and 2, but sequencing the LR-PCR revealed that exon 2 was only partially deleted (282 bp).

3. The discrepancy observed in case P36 of chapter 4.2.4 was the contrary: a small portion of exon 2 (13 bp) was deleted together with a high proportion of intron 1, but MLPA did not detect the deletion of exon 2.

The explanation to the results of cases P18 and P36 may be done by effect of the SV on the position of probes used for MLPA. As shown in figure 75, if the deletion affects the hybridization of the MLPA probe, this assay will call it as a complete deletion of the affected exon, but if it does not reach the probes, MLPA does not detect it. This explains also the results observed in the analysis of INDELs affecting exons, when the INDEL is affecting the

hybridization of MLPA probe.

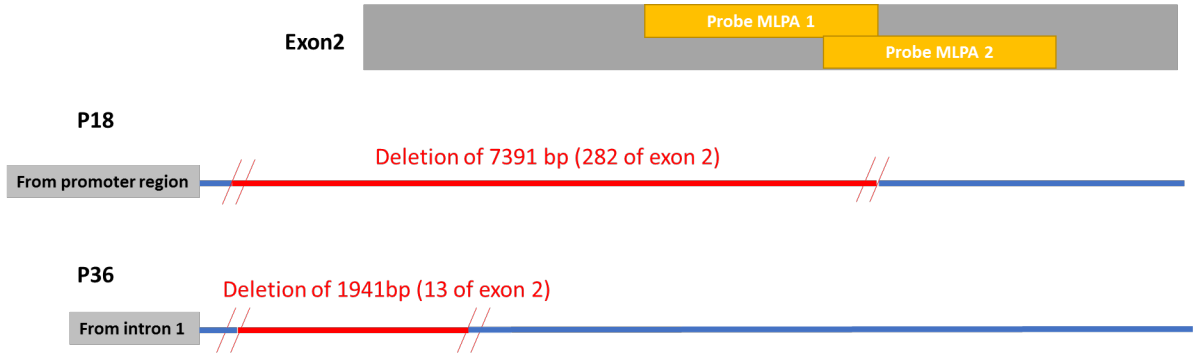


Figure 75: **Relative representation of the extension of SV of P18 and P36 in *SERPINC1* showing the regions of hybridization for MLPA probes**.

c. Discrepant results in coordinates obtained by CGH array vs nanopore sequencing

Finally, comparison of nanopore sequencing and CGH array data revealed that SV coordinates obtained by CGH array were not precise and differed, from 716 to 61316 bp from the obtained by nanopore sequencing (Table 21).

Table 21: **Comparison between genomic coordinates of SV obtained by CGH array or nanopore sequencing in cases where both techniques were done.** Mean and standard deviation are shown for start and end positions.

	aCGH coordinates	Nanopore coordinates	Difference	
			start	end
P2	1:172928626-174796949	1:172987296-174843232	-58670	-46283
P3	1:173786826-174756183	1:173848142 -174816147	-61316	-59964
P24	1:173907795-174100688	1:173908511-174102900	-716	-2212
P26	1:173911377-174138926	1:173912160-174154195	-783	-15269
P9	1:173787033-174225421	1:173787361-174223422	-328	1999
P16	1:173692858-176076493	1:173686375-176083118	6483	-6625
		media	-24362,6	-21392,333
		sd	31712,0	26762,4

5.2.9 Nanopore sequencing

From our experience, nanopore sequencing has been revealed as the most appropriate to identify and characterize at nucleotide level all SVs independently of its length or type. Moreover, it has been able to detect SVs not found by any other techniques.

Additionally, this technology defines the extension and gives the sequence of the breakpoint and neighboring areas, which allow knowing the final genetic architecture after the rearrangement and may assist to define the mechanism involved. This technology has not only detect and characterize SVs, but also has provided the haplotype resolution in cases with p.Thr147Ala. The simple use of two long amplicons obtained by LR-PCR covering the variant of interest and the sequencing by nanopore is enough for the demonstration of the common haplotype for all carriers. This procedure avoids the analysis of relatives, which is sometimes difficult to get, like in our study, or cloning and sequencing with is costly and time consuming. By LR-PCR and nanopore sequencing, we obtain results of

sequencing in 20 minutes applying a specific script for haplotype definition. This procedure can be also applied in other studies, such as the study of biallelic mutations, which impacts on the prognosis and treatment of patients.

5.2.10 Diagnostic algorithm

Our study will finally allow us to suggest a new diagnostic algorithm for the analysis of the molecular base of antithrombin deficiency based on the cost and effectiveness of a positive finding. This analysis may start by sequencing (Sanger sequencing may still be a useful and cheap method) the 7 exons and flanking regions. NGS methods covering the 7 exons or the whole gene may also be useful, although the efficacy of sequencing intronic regions still requires further studies. This approach will solve the defect causing antithrombin deficiency in up to 70-75% of cases.

In cases with negative results, search for hypoglycosylation of antithrombin and other hepatic proteins (antitrypsin or transferrin) must be done by a simple Western Blot or other methods (HPLC). Up to 5% of cases with antithrombin deficiency are explained by defects of glycosylation.

With similar rate of positive findings MLPA is the next method to be used. From 5 to 8% of cases with antithrombin deficiency are explained by SVs detected by MLPA, although the extension of the affected region is not determined. Regarding the frequency of the tandem duplication of exon 6 (1% of cases) and the simple and cheap method we have developed to specifically detect this SV, we propose to include this specific PCR amplification in the molecular diagnostic algorithm.

The complete characterization of the SV detected by MLPA, may be done by nanopore sequencing, and we strongly recommend the use of MinION and an enrichment method. Finally, to cases with unknown defects after running these methods, we suggest WGS by nanopore using promethION (Figure 76).

We are currently developing new methods using nanopore sequencing that may allow a fast, specific and simple sequencing of the *SERPINC1* region joining in a single procedure steps 1, 3, 4, 5 and 6 of the current algorithm, with an expected efficiency rate of positive findings over 85-90% with additional bonus such as haplotype definition and methylation analysis. We hope this new method might be useful in a near future.

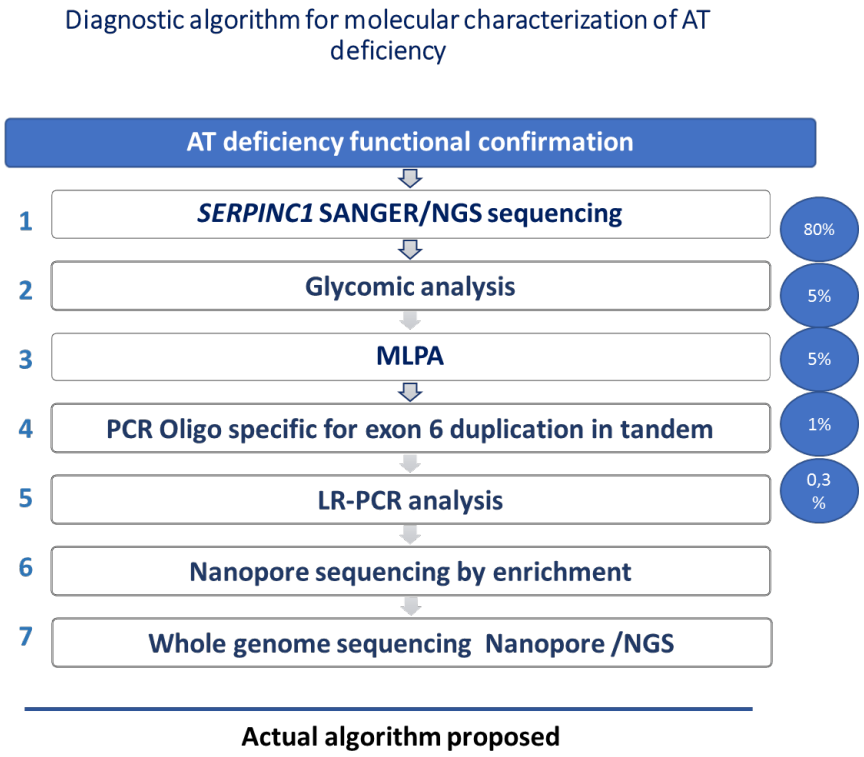


Figure 76: **Diagnostic algorithm for molecular characterization of antithrombin deficiency.** The percentage of patients with antithrombin deficiency characterized by each method in our cohort is also indicated.

5.2.11 Conclusions from the second chapter.

We have done a fine dissection of SVs involved in antithrombin deficiency. We describe the heterogeneity, in size and type, of these SVs, finding the first cases with a retrotransposon insertion in an intron as the mechanism causing antithrombin deficiency. Our picture shows a map with some *hotspots* both intra and intergenic breakpoints that might be useful for the complete characterization of new gross gene defects involving *SERPINC1* and causing antithrombin deficiency. Our study also demonstrates the importance of repetitive elements in the generation of SVs associated to antithrombin deficiency.

Finally, the use of different methods to detect and characterize SVs, LR-PCR, MLPA, CGH array, and nanopore sequencing, support that the new third generation sequencing is the best one since it detects all SVs, independently of its size of type, determines the exact extension and achieve a nucleotide resolution, also at the breakpoint, which facilitate further studies and help to understand the mechanism involved in their generation.

Conclusions

6.1 Conclusions

The results of this study reveal new insights on the spectrum of clinical consequences and new mechanisms involved in antithrombin deficiency with direct impact on diagnosis, prognosis and clinical management of patients with this rare disease.

1. Clinical consequences of antithrombin deficiency.
   1.1. The study of the largest worldwide cohort of subjects with antithrombin deficiency provides strong evidence of the very high risk of pediatric thrombosis associated to this thrombophilia (300-fold compared with the general population).
   1.2. Our results support the screening of antithrombin deficiency, particularly if it is type I, in children of affected families, in order to benefit from preventive strategies like prophylaxis in high-risk situations and from counseling concerning risk factors such as oral contraceptive use.
   1.3. Antithrombin p.Thr147Ala, responsible for a mild antithrombin type II deficiency of difficult diagnosIs, is the first *SERPINC1* founder mutation reported in Africa.

2. The study of the largest cohort of structural variants in antithrombin deficiency has revealed new insights of these gross gene defects.
   2.1. SV affecting *SERPINC1* are heterogeneous in size (193 bp to 1.8 Mb) and type (deletions, duplications, insertions, and complex SV).
   2.2. We have identified for the first time an insertion of a new SVA element causing antithrombin deficiency not detected by any other diagnostic method.
   2.3. Repetitive elements are involved in SV causing antithrombin deficiency suggesting a replication-based mechanism (such as BIR/MMBIR/FoSTeS) for generation of these SVs and defining intra and intergenic hotpots for SV affecting *SERPINC1*.
   2.4. The combination of different genetic diagnostic methods is useful for the characterization of SVs, but Nanopore sequencing is the most suitable method to fully characterize SVs associated with antithrombin deficient, providing the exact extension and genetic architecture, and the breakpoint sequence. This method also allows analysis of haplotypes, which is very useful for identification of mutations with founder effect.

Bibliography

[1] Carlos Bravo-Pérez, Vicente Vicente, and Javier Corral. Management of antithrombin deficiency: an update for clinicians, jun 2019. 4, 20

[2] ISTH Steering Committee for World Thrombosis Day. Thrombosis: a major contributor to the global disease burden. *Journal of Thrombosis and Haemostasis*, 12(10):1580–1590, 2014. 4, 5

[3] Tomasz Urbanek and Nicos Labropoulos. Can we predict and prevent the postthrombotic syndrome?, jan 2021. 5

[4] A. Rabinovich and S. R. Kahn. The postthrombotic syndrome: current evidence and future challenges, feb 2017. 5

[5] Yongping Yu, Li Yang, Yuhai Zhang, Liang Dong, Jingwen Xia, Ning Zhu, Xinpeng Han, Liying Fang, Yaqin Chai, Mengjie Niu, Lingli Liu, Xuemin Yang, Shuoyao Qu, and Shengqing Li. Incidence and risk factors of chronic thromboembolic pulmonary hypertension in patients with diagnosis of pulmonary embolism for the first time in real world. *Clinical Respiratory Journal*, 12(11):2551–2558, nov 2018. 5

[6] Clive Kearon and Elie A. Akl. Duration of anticoagulant therapy for deep vein thrombosis and pulmonary embolism, mar 2014. 5

[7] Andras Ruppert, Thomas Steinle, and Michael Lees. Economic burden of venous thromboembolism: a systematic review., 2011. 5

[8] Samuel Z. Goldhaber and Henri Bounameaux. Pulmonary embolism and deep vein thrombosis. In *The Lancet*, volume 379, pages 1835–1846. Lancet Publishing Group, may 2012. 5

[9] Joseph J. Shatzel, Matthew O'Donnell, Sven R. Olson, Matthew R. Kearney, Molly M. Daughety, Justine Hum, Khanh P. Nguyen, and Thomas G. DeLoughery. Venous thrombosis in unusual sites: A practical review for the hematologist. *European Journal of Haematology*, 102(1):53–62, jan 2019. 5

[10] Mary Cushman. Epidemiology and Risk Factors for Venous Thrombosis. *Seminars in Hematology*, 44(2):62–69, apr 2007. 5, 7

[11] Scott D. Grosse, Richard E. Nelson, Kwame A. Nyarko, Lisa C. Richardson, and Gary E. Raskob. The economic burden of incident venous thromboembolism in the United States: A review of estimated attributable healthcare costs, jan 2016. 5

[12] C. Heleen van Ommen and Ulrike Nowak-Göttl. Inherited Thrombophilia in Pediatric Venous Thromboembolic Disease: Why and Who to Test. *Frontiers in Pediatrics*, 5:50, mar 2017. 5

[13] Astrit Dautaj, Geraldo Krasi, Vilma Bushati, Vincenza Precone, Miriam Gheza, Francesco Fioretti, Marianna Sartori, Alisia Costantini, Sabrina Benedetti, Matteo Bertelli, and Stefano Paolacci. Hereditary thrombophilia, oct 2019. 6

[14] John A. Heit, Frederick A. Spencer, and Richard H. White. The epidemiology of venous thromboembolism. *Journal of Thrombosis and Thrombolysis*, 41(1):3–14, jan 2016. 6, 7, 14

[15] O EGEBERG. THROMBOPHILIA CAUSED BY INHERITABLE DEFICIENCY OF BLOOD ANTITHROMBIN. *Scandinavian journal of clinical and laboratory investigation*, 17(1):92, 1965. 7, 8

[16] Juan Carlos Souto, Laura Almasy, Montserrat Borrell, Francisco Blanco-Vaca, José Mateo, José Manuel Soria, Inma Coll, Rosa Felices, William Stone, Jordi Fontcuberta, and John Blangero. Genetic Susceptibility to Thrombosis and Its Relationship to Physiological Risk Factors: The GAIT Study. *The American Journal of Human Genetics*, 67(6), dec 2000. 7

[17] 7

[18] Poort S R, Rosendaal F R, Reitsma P H, and Bertina R M. A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis - PubMed. *Blood*, 88:3698–703, nov 1996. 7

[19] Sara Lindström, Jennifer A. Brody, Constance Turman, Marine Germain, Traci M. Bartz, Erin N. Smith, Ming-Huei Chen, Marja Puurunen, Daniel Chasman, Jeffrey Hassler, Nathan Pankratz, Saonli Basu, Weihua Guan, Beata Gyorgy, Manal Ibrahim, Jean-Philippe Empana, Robert Olaso, Rebecca Jackson, Sigrid K. Brækkan, Barbara McKnight, Jean-Francois Deleuze, Cristopher J. O'Donnell, Xavier Jouven, Kelly A. Frazer, Bruce M. Psaty, Kerri L. Wiggins, Kent Taylor, Alexander P. Reiner, Susan R. Heckbert, Charles Kooperberg, Paul Ridker, John-Bjarne Hansen, Weihong Tang, Andrew D. Johnson, Pierre-Emmanuel Morange, David A. Trégouët, Peter Kraft, Nicholas L. Smith, and Christopher Kabrhel. A large-scale exome array analysis of venous thromboembolism. *Genetic Epidemiology*, 43(4):449–457, jun 2019. 8

[20] Derek Klarin, Emma Busenkell, Renae Judy, Julie Lynch, Michael Levin, Jeffery Haessler, Krishna Aragam, Mark Chaffin, Mary Haas, Sara Lindström, Themistocles L. Assimes, Jie Huang, Kyung Min Lee, Qing Shao, Jennifer E. Huffman, Christopher Kabrhel, Yunfeng Huang, Yan V. Sun, Marijana Vujkovic, Danish Saleheen, Donald R. Miller, Peter Reaven, Scott DuVall, William E. Boden, Saiju Pyarajan, Alex P. Reiner, David Alexandre Trégouët, Peter Henke, Charles Kooperberg, J. Michael Gaziano, John Concato, Daniel J. Rader, Kelly Cho, Kyong Mi Chang, Peter W.F. Wilson, Nicholas L. Smith, Christopher J. O'Donnell, Philip S. Tsao, Sekar Kathiresan, Andrea Obi, Scott M. Damrauer, and Pradeep Natarajan. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease, nov 2019. 8

[21] Karl C. Desch, Ayse B. Ozel, Matt Halvorsen, Paula M. Jacobi, Krista Golden, Mary Underwood, Marine Germain, David Alexandre Tregouet, Pieter H. Reitsma, Clive Kearon, Lauren Mokry, J. Brent Richards, Frances Williams, Jun Z. Li, David Goldstein, and David Ginsburg. Whole-exome sequencing identifies rare variants in STAB2 associated with venous thromboembolic disease. *Blood*, 136(5):533–541, jul 2020. 8

[22] Paolo Simioni, Daniela Tormene, Giulio Tognin, Sabrina Gavasso, Cristiana Bulato, Nicholas P. Iacobelli, Jonathan D. Finn, Luca Spiezia, Claudia Radu, and Valder R. Arruda. X-Linked Thrombophilia with a Mutant Factor IX (Factor IX Padua). *New England Journal of Medicine*, 361(17), oct 2009. 8

[23] Yuhri Miyawaki, Atsuo Suzuki, Junko Fujita, Asuka Maki, Eriko Okuyama, Moe Murata, Akira Takagi, Takashi Murate, Shinji Kunishima, Michio Sakai, Kohji Okamoto, Tadashi Matsushita, Tomoki Naoe, Hide-

hiko Saito, and Tetsuhito Kojima. Thrombosis from a Prothrombin Mutation Conveying Antithrombin Resistance. *New England Journal of Medicine*, 366(25):2390–2396, jun 2012. 8

[24] René Mulder, Ton Lisman, Joost C.M. Meijers, James A. Huntington, André B. Mulder, and Karina Meijer. Linkage analysis combined with whole-exome sequencing identifies a novel prothrombin (F2) gene mutation in a Dutch Caucasian family with unexplained thrombosis, jul 2020. 8

[25] Yi Tang, Liyang Zhang, Wenlin Xie, Jieyuan Jin, Yujiao Luo, Mingyang Deng, Zhengyu Liu, Hong Wei Pan, Yi Zhang, Zhaofen Zheng, and Liang Liang Fan. A Novel Heterozygous Variant in F2 Gene in a Chinese Patient With Coronary Thrombosis and Acute Myocardial Infarction Leads to Antithrombin Resistance. *Frontiers in Genetics*, 11, mar 2020. 8

[26] V. Djordjevic, M. Kovac, P. Miljic, M. Murata, A. Takagi, I. Pruner, D. Francuski, T. Kojima, and D. Radojkovic. A novel prothrombin mutation in two families with prominent thrombophilia - the first cases of antithrombin resistance in a Caucasian population. *Journal of Thrombosis and Haemostasis*, 11(10):1936–1939, oct 2013. 8

[27] Paolo Simioni, Stefano Cagnin, Francesca Sartorello, Gabriele Sales, Luca Pagani, Cristiana Bulato, Sabrina Gavasso, Francesca Nuzzo, Francesco Chemello, Claudia Maria Radu, Daniela Tormene, Luca Spiezia, Tilman M Hackeng, Elena Campello, and Elisabetta Castoldi. Partial F8 gene duplication (Factor VIII Padua) associated with high factor VIII levels and familial thrombophilia. *Blood*, dec 2020. 8, 119

[28] Bengt Zöller, Peter J. Svensson, Björn Dahlbäck, Christina Lind-Hallden, Christer Hallden, and Johan Elf. Genetic risk factors for venous thromboembolism, sep 2020. 8

[29] Saskia Middeldorp. Inherited thrombophilia: A double-edged sword, 2016. 8, 20

[30] Willem M. Lijfering, Jan Leendert P. Brouwer, Nic J.G.M. Veeger, Ivan Bank, Michiel Coppens, Saskia Middeldorp, Karly Hamulyák, Martin H. Prins, Harry R. Büller, and Jan Van Der Meer. Selective testing for thrombophilia in patients with first venous thrombosis: Results from a retrospective family cohort study on absolute thrombotic risk for currently known thrombophilic defects in 2479 relatives. *Blood*, 113(21):5314–5322, 2009. 8, 16, 113

[31] B. K. Mahmoodi, Jan-Leendert P Brouwer, M. K. Ten Kate, Willem M Lijfering, Nic J G M Veeger, A. B. Mulder, H.C. Kluin-Nelemans, and J. Van der Meer. A prospective cohort study on the absolute risks of venous thromboembolism and predictive value of screening asymptomatic relatives of patients with hereditary deficiencies of protein S, protein C or antithrombin. *Journal of Thrombosis and Haemostasis*, 8(6):1193–1200, 2010. 8

[32] Noelene S. Quinsey, Ainslie L. Greedy, Stephen P. Bottomley, James C. Whisstock, and Robert N. Pike. Antithrombin: In control of coagulation, mar 2004. 8

[33] Juan Carlos Souto, Laura Almasy, Montserrat Borrell, Merce Garí, Elisabet Martínez, José Mateo, William H. Stone, John Blangero, and Jordi Fontcuberta. Genetic determinants of hemostasis phenotypes in Spanish families. *Circulation*, 101(13):1546–1551, apr 2000. 9

[34] Veronique Picard, Eva Ersdal-Badju, and Susan C. Bock. Partial glycosylation of antithrombin III asparagine-135 is caused by the serine in the third position of its N-glycosylation consensus sequence and is responsible for production of the .beta.-antithrombin III isoform with enhanced heparin affinity. *Biochemistry*, 34(26), jul 1995. 9

[35] Beate Luxembourg, Daniel Delev, Christof Geisen, Michael Spannagl, Manuela Krause, Wolfgang Miesbach, Christine Heller, Frauke Bergmann, Ursula Schmeink, Ralf Grossmann, Edelgard Lindhoff-Last, Erhard Seifried, Johannes Oldenburg, and Anna Pavlova. Molecular basis of antithrombin deficiency. *Thrombosis and Haemostasis*, 105(04), nov 2011. 9, 73, 93

[36] Leslie R. Berry, Bruce Thong, and Anthony K.C. Chan. Comparison of recombinant and plasma-derived antithrombin biodistribution in a rabbit model. *Thrombosis and Haemostasis*, 102(2):302–308, aug 2009. 9

[37] Ruby H.P. Law, Qingwei Zhang, Sheena McGowan, Ashley M. Buckle, Gary A. Silverman, Wilson Wong, Carlos J. Rosado, Chris G. Langendorf, Rob N. Pike, Philip I. Bird, and James C. Whisstock. An overview of the serpin superfamily, may 2006. 10

[38] Lois T. Hunt and Margaret O. Dayhoff. A surprising new protein superfamily containing ovalbumin, antithrombin-III, and alpha1-proteinase inhibitor. *Topics in Catalysis*, 95(2):864–871, 1980. 10

[39] P. I. Bird. Serpins and regulation of cell death., 1998. 10

[40] James A. Irving, Robert N. Pike, Arthur M. Lesk, and James C. Whisstock. Phylogeny of the serpin superfamily: Implications of patterns of amino acid conservation for structure and function. *Genome Research*, 10(12):1845–1864, 2000. 10

[41] James C. Whisstock and Stephen P. Bottomley. Molecular gymnastics: serpin structure, folding and misfolding, dec 2006. 10

[42] J. C. Rau, L. M. Beaulieu, J. A. Huntington, and Frank C. Church. Serpins in thrombosis, hemostasis and fibrinolysis, 2007. 10

[43] Gary A. Silverman, Phillip I. Bird, Robin W. Carrell, Frank C. Church, Paul B. Coughlin, Peter G.W. Gettins, James A. Irving, David A. Lomas, Cliff J. Luke, Richard W. Moyer, Philip A. Pemberton, Eileen Remold-O'Donnell, Guy S. Salvesen, James Travis, and James C. Whisstock. The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature, sep 2001. 10

[44] Penelope E. Stein and Robin W. Carrell. What do dysfunctional serpins tell us about molecular mobility and disease? *Nature Structural Biology*, 2(2):96–113, 1995. 10

[45] James Whisstock, Richard Skinner, and Arthur M. Lesk. An atlas of serpin conformations, 1998. 10

[46] Wolfram Bode and Robert Huber. Structural basis of the endoproteinase-protein inhibitor interaction, mar 2000. 10

[47] Mark R. Wardell, Jan Pieter Abrahams, David Bruce, Richard Skinner, and Andrew G.W. Leslie. Crystallization and preliminary X-ray diffraction analysis of two conformations of intact human antithrombin. *Journal of Molecular Biology*, 234(4):1253–1258, dec 1993. 11

[48] Daniel J.D. Johnson, Jonathan Langdown, Wei Li, Stephan A. Luis, Trevor P. Baglin, and James A. Huntington. Crystal structure of monomeric native antithrombin reveals a novel reactive center loop conformation. *Journal of Biological Chemistry*, 281(46):35478–35486, nov 2006. 11, 12

[49] Lei Jin, Jan Pieter Abrahams, Richard Skinner, Maurice Petitou, Robert N. Pike, and Robin W. Carrell. The anticoagulant activation of antithrombin by heparin. *Proceedings of the National Academy of Sciences of the United States of America*, 94(26):14683–14688, dec 1997. 12

[50] J. A. Huntington. Mechanisms of glycosaminoglycan activation of the serpins in hemostasis, jul 2003. 12

[51] Jonathan Langdown, Klara J. Belzar, Wendy J. Savory, Trevor P. Baglin, and James A. Huntington. The Critical Role of Hinge-Region Expulsion in the Induced-Fit Heparin Binding Mechanism of Antithrombin. *Journal of Molecular Biology*, 386(5):1278–1289, mar 2009. 12

[52] Maurice Petitou, Philippe Duchaussoy, Jean Marc Herbert, Gérald Duc, Mohamed El Hajji, Jean François Branellec, François Donat, José Necciari, Roger Cariou, Jean Bouthier, and Eric Garrigou. The synthetic pentasaccharide fondaparinux: First in the class of antithrombotic agents that selectively inhibit coagulation factor Xa, aug 2002. 13

[53] Hana Im, Hee-Young Ahn, and Myeong-Hee Yu. Bypassing the kinetic trap of serpin protein folding by loop extension. *Protein Science*, 9(8):1497–1502, 2000. 13

[54] Javier Corral, Jose Rivera, Constantino Martinez, Rocio Gonzalez-Conejero, Antonia Minano, and Vicente Vicente. Detection of conformational transformation of antithrombin in blood with crossed immunoelectrophoresis: New application for a classical method. *Journal of Laboratory and Clinical Medicine*, 142(5):298–305, 2003. 13, 34

[55] A. Mushunje, G. Evans, S. O. Brennan, R. W. Carrell, and A. Zhou. Latent antithrombin and its detection, formation and turnover in the circulation. *Journal of Thrombosis and Haemostasis*, 2(12):2170–2177, dec 2004. 13, 34

[56] Javier Corral, Vicente Vicente, and Robin W Carrell. Thrombosis as a conformational disease. *Haematologica*, 90(2):238–46, 2005. 14, 113

[57] David Hernández-Espinosa, Adriana Ordóñez, Vicente Vicente, and Javier Corral. Factors with conformational effects on haemostatic serpins: Implications in thrombosis, sep 2007. 14

[58] Michael S. O'Reilly, Steven Pirie-Shepherd, William S. Lane, and Judah Folkman. Antiangiogenic activity of the cleaved conformation of the serpin antithrombin. *Science*, 285(5435):1926–1928, sep 1999. 14

[59] Mohammed Asmal, Michael Seaman, Wenyu Lin, Raymond T. Chung, Norman L. Letvin, and Ralf Geiben-Lynn. Inhibition of HCV by the serpin antithrombin III. *Virology Journal*, 9, 2012. 14

[60] Praveen Papareddy, Madlen Rossnagel, Femke Doreen Hollwedel, Gülcan Kilic, Srinivas Veerla, Clément Naudin, Emanuel Smeds, Johannes Westman, Irene Martinez-Martinez, Arne Egesten, Maria Eugenia de la Morena-Barrio, Javier Corral, Adam Linder, Andrea Artoni, Maria Abbattista, Cristina Novembrino, Cord Herbert Brakebusch, Ida Martinelli, Gopinath Kasetty, and Heiko Herwald. A human antithrombin isoform dampens inflammatory responses and protects from organ damage during bacterial infection. *Nature Microbiology*, 4(12):2442–2455, dec 2019. 14

[61] José A. Guerrero, Raúl Teruel, Constantino Martínez, Isabel Arcas, Irene Martínez-Martínez, María Eugenia De La Morena-Barrio, Vicente Vicente, and Javier Corral. Protective role of antithrombin in mouse models of liver injury. *Journal of Hepatology*, 57(5):980–986, nov 2012. 14

[62] Ginés Luengo-Gil, María Inmaculada Calvo, Ester Martín-Villar, Sonia Águila, Nataliya Bohdan, Ana I. Antón, Salvador Espín, Francisco Ayala De La Peña, Vicente Vicente, Javier Corral, Miguel Quintanilla, and Irene Martínez-Martínez. Antithrombin controls tumor migration, invasion and angiogenesis by inhibition of enteropeptidase. *Scientific Reports*, 6, jun 2016. 14

[63] Robin J. Olds, David A. Lane, Vijoy Chowdhury, Valerio De Stefano, Guiseppe Leone, and Swee Lay Thein. Complete Nucleotide Sequence of the Antithrombin Gene: Evidence for Homologous Recombination Causing Thrombophilia. *Biochemistry*, 32(16):4216–4224, apr 1993. 15, 122, 123

[64] Susan Clark Bock and Diane J. Levitan. Characterization of an unusual DNA length polymorphism 5' to the human antithrombin III gene. *Nucleic Acids Research*, 11(24):8569–8582, dec 1983. 15

[65] Edward V. Prochownik. Relationship between an enhancer element in the human antithrombin III gene and an immunoglobulin light-chain gene enhancer. *Nature*, 316(6031):845–848, 1985. 15

[66] Mara Toderici, María Eugenia De La Morena-Barrio, José Padilla, Antonia Miñano, Ana Isabel Antón, Juan Antonio Iniesta, María Teresa Herranz, Nuria Fernández, Vicente Vicente, and Javier Corral. Identification of Regulatory Mutations in SERPINC1 Affecting Vitamin D Response Elements Associated with Antithrombin Deficiency. *PLoS ONE*, 11(3), mar 2016. 15, 117, 119

[67] María Eugenia de la Morena-Barrio, Ana Isabel Antón, Irene Martínez-Martínez, José Padilla, Antonia Miñano, José Navarro-Fernández, Sonia Águila, María Fernanda López, Jordi Fontcuberta, Vicente Vicente, and Javier Corral. Regulatory regions of SERPINC1 gene: Identification of the first mutation associated with antithrombin deficiency. *Thrombosis and Haemostasis*, 107(3):430–437, mar 2012. 15, 117

[68] Ana I. Antón, Raúl Teruel, Javier Corral, Antonia Miñano, Irene Martánez-Martánez, Adriana Ordóñez, Vicente Vicente, and Beatriz Sánchez-Vega. Functional consequences of the prothrombotic serpinc1 rs2227589 polymorphism on antithrombin levels. *Haematologica*, 94(4):589–592, apr 2009. 15

[69] Haig H. Kazazian and John V. Moran. Mobile DNA in Health and Disease. *New England Journal of Medicine*, 377(4):361–370, jul 2017. 15, 83, 121

[70] P. Bucciarelli, S. M. Passamonti, E. Biguzzi, F. Gianniello, F. Franchi, P. M. Mannucci, and I. Martinelli. Low borderline plasma levels of antithrombin, protein C and protein S are risk factors for venous thromboembolism. *Journal of Thrombosis and Haemostasis*, 10(9):1783–1791, sep 2012. 15

[71] K Ishiguro, T Kojima, K Kadomatsu, Y Nakayama, A Takagi, M Suzuki, N Takeda, M Ito, K Yamamoto, T Matsushita, K Kusugami, T Muramatsu, and H Saito. Complete antithrombin deficiency in mice results in embryonic lethality. *The Journal of clinical investigation*, 106(7):873–8, 2000. 16, 45

[72] Javier Corral, Eugenia De La Morena-Barrio, and Vicente Vicente. The genetics of antithrombin. *Thrombosis Research*, 169:23–29, 2018. 16, 117

[73] C. Bravo-Pérez, M. E. de la Morena-Barrio, A. Palomo, L. Entrena, B. de la Morena-Barrio, J. Padilla, A. Miñano, E. Navarro, R. Cifuentes, J. Corral, and V. Vicente. Genotype–phenotype gradient of SERPINC1 variants in a single family reveals a severe compound antithrombin deficiency in a dead embryo, oct 2020. 16, 18

[74] David Hernández-Espinosa, Antonia Miñano, Constantino Martínez, Elena Pérez-Ceballos, Inmaculada Heras, José L. Fuster, Vicente Vicente, and Javier Corral. L-asparaginase-induced antithrombin type I deficiency: Implications for conformational diseases. *American Journal of Pathology*, 169(1):142–153, 2006. 17, 33

[75] Carlos Bravo-Pérez, Maria E. de la Morena-Barrio, Vicente Vicente, and Javier Corral. Antithrombin deficiency as a still underdiagnosed type of thrombophilia: a primer for internists, oct 2020. 17

[76] Sonia Águila, Gonzalo Izaguirre, Irene Martínez-Martínez, Vicente Vicente, Steven T. Olson, and Javier Corral. Disease-causing mutations in the serpin antithrombin reveal a key domain critical for inhibiting protease activities. *Journal of Biological Chemistry*, 292(40):16513–16520, oct 2017. 17

[77] María de la Morena-Barrio, Edna Sandoval, Pilar Llamas, Ewa Wypasek, Mara Toderici, José Navarro-Fernández, Agustin Rodríguez-Alen, Nuria Revilla, Raquel López-Gálvez, Antonia Miñano, José Padilla,

Belén de la Morena-Barrio, Jorge Cuesta, Javier Corral, and Vicente Vicente. High levels of latent antithrombin in plasma from patients with antithrombin deficiency. *Thrombosis and Haemostasis*, 117(5):880–888, 2017. 17

[78] Vanessa Swoboda, Katharina Zervan, Katharina Thom, Christine Mannhalter, Peter Quehenberger, Ingrid Pabinger, and Christoph Male. Homozygous antithrombin deficiency type II causing neonatal thrombosis. *Thrombosis Research*, 158:134–137, oct 2017. 18

[79] Javier Corral, David Hernandez-Espinosa, Jose Manual Soria, Rocio Gonzalez-Conejero, Adriana Ordonez, Jose Ramon Gonzalez-Porras, Elena Perez-Ceballos, Ramon Lecumberri, Ignacio Sanchez, Vanessa Roldan, Jose Mateo, Antonia Minano, Marcos Gonzalez, Ignacio Alberca, Jordi Fontcuberta, and Vicente Vicente. Antithrombin Cambridge II (A384S): An underestimated genetic risk factor for venous thrombosis. *Blood*, 109(10):4258–4263, may 2007. 18

[80] José Navarro-Fernández, María Eugenia De La Morena-Barrio, José Padilla, Antonia Miñano, Nataliya Bohdan, Sonia Águila, Irene Martínez-Martínez, Teresa S. Sevivas, Carmen De Cos, Nuria Fernández-Mosteirín, Pilar Llamas, Susana Asenjo, Pilar Medina, Carlos C. Souto, Kim Overvad, Søren R. Kristensen, Javier Corral, and Vicente Vicente. Antithrombin Dublin (p.Val30Glu): A relatively common variant with moderate thrombosis risk of causing transient antithrombin deficiency. *Thrombosis and Haemostasis*, 116(1):146–154, jul 2016. 18

[81] Sonia Águila, Irene Martínez-Martínez, Miriam Collado, Pilar Llamas, Ana I. Antón, Consuelo Martínez-Redondo, José Padilla, Antonia Miñano, María E. de la Morena-Barrio, Ángel García-Avello, Vicente Vicente, and Javier Corral. Compound heterozygosity involving Antithrombin Cambridge II (p.Ala416Ser) in antithrombin deficiency, jan 2013. 18

[82] Srikumar M. Raja, Neetu Chhablani, Richard Swanson, Elizabeth Thompson, Mike Laffan, David A. Lane, and Steven T. Olson. Deletion of P1 arginine in a novel antithrombin variant (antithrombin London) abolishes inhibitory activity but enhances heparin affinity and is associated with early onset thrombosis. *Journal of Biological Chemistry*, 278(16):13688–13695, apr 2003. 18

[83] Irene Martínez-Martínez, José Navarro-Fernández, Sonia Aguila, Antonia Miñano, Nataliya Bohdan, María Eugenia De La Morena-Barrio, Adriana Ordóñez, Constantino Martínez, Vicente Vicente, and Javier Corral. The infective polymerization of conformationally unstable antithrombin mutants may play a role in the clinical severity of antithrombin deficiency. *Molecular medicine (Cambridge, Mass.)*, 18(1):762–770, 2012. 19

[84] José Navarro-Fernández, María Eugenia De La Morena-Barrio, José Padilla, Antonia Miñano, Nataliya Bohdan, Sonia Águila, Irene Martínez-Martínez, Teresa S. Sevivas, Carmen De Cos, Nuria Fernández-Mosteirín, Pilar Llamas, Susana Asenjo, Pilar Medina, Carlos C. Souto, Kim Overvad, Søren R. Kristensen, Javier Corral, and Vicente Vicente. Antithrombin Dublin (p.Val30Glu): A relatively common variant with moderate thrombosis risk of causing transient antithrombin deficiency. *Thrombosis and Haemostasis*, 116(1):146–154, jul 2016. 19

[85] José Navarro-Fernández, María Eugenia de la Morena-Barrio, Emma Martínez-Alonso, Ingunn Dybedal, Mara Toderici, Nataliya Bohdan, Antonia Miñano, Ketil Heimdal, Ulrich Abildgaard, José Ángel Martínez-Menárguez, Javier Corral, and Vicente Vicente. Biochemical and cellular consequences of the antithrombin p.Met1? Mutation identified in a severe thrombophilic family. *Oncotarget*, 9(69):33202–33214, sep 2018. 19

[86] M. E. de la Morena-Barrio, I. Martínez-Martínez, C. de Cos, E. Wypasek, V. Roldán, A. Undas, M. van Scherpenzeel, D. J. Lefeber, M. Toderici, T. Sevivas, F. España, J. Jaeken, J. Corral, and V. Vicente. Hypoglycosylation is a common finding in antithrombin deficiency in the absence of a SERPINC1 gene defect. *Journal of Thrombosis and Haemostasis*, 14(8):1549–1560, aug 2016. 19, 64, 117

[87] Javier Corral, María Eugenia de la Morena-Barrio, and Vicente Vicente. The genetics of antithrombin. *Thrombosis research*, 169:23–29, jul 2018. 19

[88] Elizabeth M. Van Cott, Christelle Orlando, Gary W. Moore, Peter C. Cooper, Piet Meijer, and Richard Marlar. Recommendations for clinical laboratory testing for antithrombin deficiency; Communication from the SSC of the ISTH. *Journal of Thrombosis and Haemostasis*, 18(1):17–22, jan 2020. 19

[89] J. D. Watson and F. H.C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953. 20

[90] Ray Wu and A. D. Kaiser. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *Journal of Molecular Biology*, 35(3):523–537, 1968. 20

[91] W. Gilbert and A. Maxam. The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences of the United States of America*, 70(12 (I)):3581–3584, 1973. 20

[92] Jay Shendure, Shankar Balasubramanian, George M. Church, Walter Gilbert, Jane Rogers, Jeffery A. Schloss, and Robert H. Waterston. DNA sequencing at 40: Past, present and future. *Nature*, 550(7676):345–353, oct 2017. 21

[93] P. J. Roberts. Human genome project, 2001. 21

[94] William Waggott. Long Range PCR. In *Clinical Applications of PCR*, volume 16, pages 81–92. Humana Press, New Jersey, 1998. 21

[95] Martin Kircher and Janet Kelso. High-throughput DNA sequencing - Concepts and limitations, jun 2010. 21

[96] Samuel P. Strom, Hane Lee, Kingshuk Das, Eric Vilain, Stanley F. Nelson, Wayne W. Grody, and Joshua L. Deignan. Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. *Genetics in Medicine*, 16(7):510–515, 2014. 22

[97] Tyler F. Beck, James C. Mullikin, and Leslie G. Biesecker. Systematic evaluation of sanger validation of next-generation sequencing variants. *Clinical Chemistry*, 62(4):647–654, apr 2016. 22

[98] Linnea M. Baudhuin, Susan A. Lagerstedt, Eric W. Klee, Numrah Fadra, Devin Oglesbee, and Matthew J. Ferber. Confirming variants in next-generation sequencing panel testing by sanger sequencing. *Journal of Molecular Diagnostics*, 17(4):456–461, jul 2015. 22

[99] Wenbo Mu, Hsiao Mei Lu, Jefferey Chen, Shuwei Li, and Aaron M. Elliott. Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. *Journal of Molecular Diagnostics*, 18(6):923–932, nov 2016. 22

[100] John Huddleston, Mark J.P. Chaisson, Karyn Meltz Steinberg, Wes Warren, Kendra Hoekzema, David Gordon, Tina A. Graves-Lindsay, Katherine M. Munson, Zev N. Kronenberg, Laura Vives, Paul Peluso, Matthew Boitano, Chen Shin Chin, Jonas Korlach, Richard K. Wilson, and Evan E. Eichler. Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Research*, 27(5):677–685, may 2017. 22

[101] Evan E. Eichler. Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. *New England Journal of Medicine*, 381(1):64–74, jul 2019. 22, 118

[102] Thomas S.K. Wan. Cancer cytogenetics: Methodology revisited. *Annals of Laboratory Medicine*, 34(6):413–425, 2014. 23

[103] Thomas S.K. Wan. Cancer cytogenetics: An introduction. *Methods in Molecular Biology*, 1541:1–10, 2017. 23

[104] Antoine M. Snijders, Norma Nowak, Richard Segraves, Stephanie Blackwood, Nils Brown, Jeffrey Conroy, Greg Hamilton, Anna Katherine Hindle, Bing Huey, Karen Kimura, Sindy Law, Ken Myambo, Joel Palmer, Bauke Ylstra, Jingzhu Pearl Yue, Joe W. Gray, Ajay N. Jain, Daniel Pinkel, and Donna G. Albertson. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nature Genetics*, 29(3):263–264, oct 2001. 23

[105] N. De Leeuw, J. Y. Hehir-Kwa, A. Simons, A. Geurts Van Kessel, D. F. Smeets, B. H.W. Faas, and R. Pfundt. SNP array analysis in constitutional and cancer genome diagnostics - Copy number variants, genotyping and quality control. *Cytogenetic and Genome Research*, 135(3-4):212–221, dec 2011. 23

[106] Liborio Stuppia, Ivana Antonucci, Giandomenico Palka, and Valentina Gatta. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases, mar 2012. 23

[107] Daniel R. Garalde, Elizabeth A. Snell, Daniel Jachimowicz, Botond Sipos, Joseph H. Lloyd, Mark Bruce, Nadia Pantic, Tigist Admassu, Phillip James, Anthony Warland, Michael Jordan, Jonah Ciccone, Sabrina Serra, Jemma Keenan, Samuel Martin, Luke McNeill, E. Jayne Wallace, Lakmal Jayasinghe, Chris Wright, Javier Blasco, Stephen Young, Denise Brocklebank, Sissel Juul, James Clarke, Andrew J. Heron, and Daniel J. Turner. Highly parallel direct RN A sequencing on an array of nanopores. *Nature Methods*, 15(3):201–206, mar 2018. 23, 24, 120

[108] Doruk Beyter, Helga Ingimundardottir, Hannes P. Eggertsson, Eythor Bjornsson, Snaedis Kristmundsdottir, Svenja Mehringer, Hakon Jonsson, Marteinn T. Hardarson, Droplaug N. Magnusdottir, Ragnar P. Kristjansson, Sigurjon A. Gudjonsson, Sverrir T. Sverrisson, Guillaume Holley, Gudmundur Eyjolfsson, Isleifur Olafsson, Olof Sigurdardottir, Gisli Masson, Unnur Thorsteinsdottir, Daniel F. Gudbjartsson, Patrick Sulem, Olafur T. Magnusson, Bjarni V. Halldorsson, and Kari Stefansson. Long read sequencing of 1,817 Icelanders provides insight into the role of structural variants in human disease, nov 2019. 23, 73, 75

[109] Mircea Cretu Stancu, Markus J. Van Roosmalen, Ivo Renkens, Marleen M. Nieboer, Sjors Middelkamp, Joep De Ligt, Giulia Pregno, Daniela Giachino, Giorgia Mandrile, Jose Espejo Valle-Inclan, Jerome Korzelius, Ewart De Bruijn, Edwin Cuppen, Michael E. Talkowski, Tobias Marschall, Jeroen De Ridder, and Wigard P. Kloosterman. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nature Communications*, 8(1), dec 2017. 23, 24

[110] Nobuaki Kono and Kazuharu Arakawa. Nanopore sequencing: Review of potential applications in functional genomics. *Development, Growth & Differentiation*, 61(5):316–326, jun 2019. 24

[111] Hengyun Lu, Francesca Giordano, and Zemin Ning. Oxford Nanopore MinION Sequencing and Genome Assembly, oct 2016. 24

[112] Arthur C. Rand, Miten Jain, Jordan M. Eizenga, Audrey Musselman-Brown, Hugh E. Olsen, Mark Akeson, and Benedict Paten. Mapping DNA methylation with high-throughput nanopore sequencing. *Nature Methods*, 14(4):411–413, feb 2017. 24

[113] Jared T. Simpson, Rachael E. Workman, P. C. Zuzarte, Matei David, L. J. Dursi, and Winston Timp. Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods*, 14(4):407–410, feb 2017. 24

[114] Jana Ebler, Marina Haukness, Trevor Pesout, Tobias Marschall, and Benedict Paten. Haplotype-aware diplotyping from noisy long reads. *Genome Biology*, 20(1):116, jun 2019. 25

[115] Sergey Koren, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, Nicholas H. Bergman, and Adam M. Phillippy. Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Research*, 27(5):722–736, may 2017. 25

[116] Christelle Orlando, Olivier Heylen, Willy Lissens, and Kristin Jochmans. Antithrombin heparin binding site deficiency: A challenging diagnosis of a not so benign thrombophilia. *Thrombosis Research*, 135(6):1179–1185, jun 2015. 33, 59, 114

[117] J. Corral, J. A. Huntington, R. González-Conejero, A. Mushunje, M. Navarro, P. Marco, V. Vicente, and R. W. Carrell. Mutations in the shutter region of antithrombin result in formation of disulfide-linked dimers and severe venous thrombosis. *Journal of Thrombosis and Haemostasis*, 2(6):931–939, jun 2004. 34, 113

[118] François Olivier Desmet, Dalil Hamroun, Marine Lalande, Gwenaëlle Collod-Bëroud, Mireille Claustres, and Christophe Béroud. Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*, 37(9), 2009. 35

[119] Belén De La Morena-Barrio, María Eugenia De La Morena-Barrio, José Padilla, Raúl Teruel-Montoya, Susana Asenjo, Ewa Wypasek, Anetta Undas, Antonia Miñano, Vicente Vicente, and Javier Corral. Identification of a New Mechanism of Antithrombin Deficiency Hardly Detected by Current Methods: Duplication of SERPINC1 Exon 6, may 2018. 35, 68, 98

[120] Johannes Köster and Sven Rahmann. Snakemake - A scalable bioinformatics workflow engine, oct 2018. 37

[121] Wouter De Coster, Peter De Rijk, Arne De Roeck, Tim De Pooter, Svenn D'Hert, Mojca Strazisar, Kristel Sleegers, and Christine Van Broeckhoven. Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Research*, 29(7):1178–1187, 2019. 37

[122] Heng Li. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100, sep 2018. 37, 38

[123] Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, aug 2009. 38, 39

[124] James T. Robinson, Helga Thorvaldsdóttir, Aaron M. Wenger, Ahmet Zehir, and Jill P. Mesirov. Variant review with the integrative genomics viewer, nov 2017. 38

[125] Aaron R. Quinlan and Ira M. Hall. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, jan 2010. 39

[126] Tuomo Mantere, Simone Kersten, and Alexander Hoischen. Long-read sequencing emerging in medical genetics, 2019. 40

[127] Elien Roose, Claudia Tersteeg, Ruth Demeersseman, An-Sofie Schelpe, Louis Deforche, Inge Pareyn, Aline Vandenbulcke, Nele Vandeputte, Daan Dierickx, Jan Voorberg, Hans Deckmyn, Simon De Meyer, and Karen Vanhoorelbeke. Anti-ADAMTS13 Antibodies and a Novel Heterozygous p.R1177Q Mutation in a Case of

Pregnancy-Onset Immune-Mediated Thrombotic Thrombocytopenic Purpura. *TH Open*, 02(01):e8–e15, jan 2018. 41

[128] Douglas G., Altman, Chapman, and Hall. Practical statistics for medical research. Douglas G. Altman, Chapman and Hall, London, 1991. No. of pages: 611. *Statistics in Medicine*, 10(10):1635–1636, oct 1991. 42

[129] Elizabeth A. Chalmers. Epidemiology of venous thromboembolism in neonates and children. *Thrombosis Research*, 118(1):3–12, 2006. 45, 111

[130] Guy Young, Manuela Albisetti, Mariana Bonduel, Leonardo Brandao, Anthony Chan, Frauke Friedrichs, Neil A Goldenberg, Eric Grabowski, Christine Heller, Janna Journeycake, Gili Kenet, Anne Krümpel, Karin Kurnik, Aaron Lubetsky, Christoph Male, Marilyn Manco-Johnson, Prasad Mathew, Paul Monagle, Heleen van Ommen, Paolo Simioni, Pavel Svirin, Daniela Tormene, and Ulrike Nowak-Göttl. Impact of inherited thrombophilia on venous thromboembolism in children: a systematic review and meta-analysis of observational studies. *Circulation*, 118(13):1373–82, sep 2008. 45

[131] R. C. Tait, Isobel D. Walker, D. J. Perry, S. I. A. M. Islam, M. E. Daly, F. McCall, J. A. Conkie, and R. W. Carrell. Prevalence of antithrombin deficiency in the healthy population. *British Journal of Haematology*, 87(1):106–112, may 1994. 55

[132] R Gindele, Z Oláh, P Ilonczai, M Speker, Á Udvari, A Selmeczi, G Pfliegler, E Marján, B Kovács, Z Boda, L Muszbek, and Z Bereczky. Founder effect is responsible for the p.Leu131Phe heparin-binding-site antithrombin mutation common in Hungary: phenotype analysis in a large cohort. *Journal of thrombosis and haemostasis : JTH*, 14(4):704–15, 2016. 55, 113, 115

[133] M. Puurunen, P. Salo, S. Engelbarth, K. Javela, and M. Perola. Type II antithrombin deficiency caused by a founder mutation Pro73Leu in the Finnish population: Clinical picture. *Journal of Thrombosis and Haemostasis*, 11(10):1844–1849, oct 2013. 55, 114, 115

[134] Sue Richards, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, Julie Gastier-Foster, Wayne W. Grody, Madhuri Hegde, Elaine Lyon, Elaine Spector, Karl Voelkerding, and Heidi L. Rehm. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5):405–424, may 2015. 57

[135] P. C. Cooper, F. Coath, M. E. Daly, and M. Makris. The phenotypic and genetic assessment of antithrombin deficiency, jun 2011. 59

[136] Michael Caspers, Anna Pavlova, Julia Driesen, Ursula Harbrecht, Robert Klamroth, Janos Kadar, Ronald Fischer, Bettina Kemkes-Matthes, and Johannes Oldenburg. Deficiencies of antithrombin, protein C and protein S - practical experience in genetic analysis of a large patient cohort. *Thrombosis and Haemostasis*, 108(2):247–257, aug 2012. 64, 93

[137] Shogo Tamura, Erika Hashimoto, Nobuaki Suzuki, Misaki Kakihara, Koya Odaira, Yuna Hattori, Mahiru Tokoro, Sachiko Suzuki, Akira Takagi, Akira Katsumi, Fumihiko Hayakawa, Atsuo Suzuki, Shuichi Okamoto, Takeshi Kanematsu, Tadashi Matsushita, and Tetsuhito Kojima. Molecular basis of SERPINC1 mutations in Japanese patients with antithrombin deficiency. *Thrombosis Research*, 178:159–170, jun 2019. 64

[138] Miguel de Sousa Dias, Imma Hernan, Beatriz Pascual, Emma Borràs, Begoña Mañé, Maria José Gamundi, and Miguel Carballo. Detection of novel mutations that cause autosomal dominant retinitis pigmentosa

in candidate genes by long-range PCR amplification and next-generation sequencing. *Molecular Vision*, 19:654–664, mar 2013. 68

[139] Alba Sanchis-Juan, Jonathan Stephens, Courtney E. French, Nicholas Gleadall, Karyn Mégy, Christopher Penkett, Olga Shamardina, Kathleen Stirrups, Isabelle Delon, Eleanor Dewhurst, Helen Dolling, Marie Erwood, Detelina Grozeva, Luca Stefanucci, Gavin Arno, Andrew R. Webster, Trevor Cole, Topun Austin, Ricardo Garcia Branco, Willem H. Ouwehand, F. Lucy Raymond, and Keren J. Carss. Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. *Genome Medicine*, 10(1), dec 2018. 73, 120

[140] Fritz J. Sedlazeck, Hayan Lee, Charlotte A. Darby, and Michael C. Schatz. Piercing the dark matter: Bioinformatics of long-range sequencing and mapping, jun 2018. 73, 120

[141] Ernest Turro, William J. Astle, Karyn Megy, Stefan Gräf, Daniel Greene, Olga Shamardina, Hana Lango Allen, Alba Sanchis-Juan, Mattia Frontini, Chantal Thys, Jonathan Stephens, Rutendo Mapeta, Oliver S. Burren, Kate Downes, Matthias Haimel, Salih Tuna, Sri V.V. Deevi, Timothy J. Aitman, David L. Bennett, Paul Calleja, Keren Carss, Mark J. Caulfield, Patrick F. Chinnery, Peter H. Dixon, Daniel P. Gale, Roger James, Ania Koziell, Michael A. Laffan, Adam P. Levine, Eamonn R. Maher, Hugh S. Markus, Joannella Morales, Nicholas W. Morrell, Andrew D. Mumford, Elizabeth Ormondroyd, Stuart Rankin, Augusto Rendon, Sylvia Richardson, Irene Roberts, Noemi B.A. Roy, Moin A. Saleem, Kenneth G.C. Smith, Hannah Stark, Rhea Y.Y. Tan, Andreas C. Themistocleous, Adrian J. Thrasher, Hugh Watkins, Andrew R. Webster, Martin R. Wilkins, Catherine Williamson, James Whitworth, Sean Humphray, David R. Bentley, Stephen Abbs, Lara Abulhoul, Julian Adlard, Munaza Ahmed, Hana Alachkar, David J. Allsup, Jeff Almeida-King, Philip Ancliff, Richard Antrobus, Ruth Armstrong, Gavin Arno, Sofie Ashford, Anthony Attwood, Paul Aurora, Christian Babbs, Chiara Bacchelli, Tamam Bakchoul, Siddharth Banka, Tadbir Bariana, Julian Barwell, Joana Batista, Helen E. Baxendale, Phil L. Beales, David L. Bennett, David R. Bentley, Agnieszka Bierzynska, Tina Biss, Maria A.K. Bitner-Glindzicz, Graeme C. Black, Marta Bleda, Iulia Blesneac, Detlef Bockenhauer, Harm Bogaard, Christian J. Bourne, Sara Boyce, John R. Bradley, Eugene Bragin, Gerome Breen, Paul Brennan, Carole Brewer, Matthew Brown, Andrew C. Browning, Michael J. Browning, Rachel J. Buchan, Matthew S. Buckland, Teofila Bueser, Carmen Bugarin Diz, John Burn, Siobhan O. Burns, Oliver S. Burren, Nigel Burrows, Carolyn Campbell, Gerald Carr-White, Keren Carss, Ruth Casey, Jenny Chambers, John Chambers, Melanie M.Y. Chan, Calvin Cheah, Floria Cheng, Patrick F. Chinnery, Manali Chitre, Martin T. Christian, Colin Church, Jill Clayton-Smith, Maureen Cleary, Naomi Clements Brod, Gerry Coghlan, Elizabeth Colby, Trevor R.P. Cole, Janine Collins, Peter W. Collins, Camilla Colombo, Cecilia J. Compton, Robin Condliffe, Stuart Cook, H. Terence Cook, Nichola Cooper, Paul A A. Corris, Abigail Furnell, Fiona Cunningham, Nicola S. Curry, Antony J. Cutler, Matthew J. Daniels, Mehul Dattani, Louise C. Daugherty, John Davis, Anthony De Soyza, Sri V.V. Deevi, Timothy Dent, Charu Deshpande, Eleanor F. Dewhurst, Peter H. Dixon, Sofia Douzgou, Kate Downes, Anna M. Drazyk, Elizabeth Drewe, Daniel Duarte, Tina Dutt, J. David M. Edgar, Karen Edwards, William Egner, Melanie N. Ekani, Perry Elliott, Wendy N. Erber, Marie Erwood, Maria C. Estiu, Dafydd Gareth Evans, Gillian Evans, Tamara Everington, Mélanie Eyries, Hiva Fassihi, Remi Favier, Jack Findhammer, Debra Fletcher, Frances A. Flinter, R. Andres Floto, Tom Fowler, James Fox, Amy J. Frary, Courtney E. French, Kathleen Freson, Mattia Frontini, Daniel P. Gale, Henning Gall, Vijeya Ganesan, Michael Gattens, Claire Geoghegan, Terence S.A. Gerighty, Ali G. Gharavi, Stefano Ghio, Hossein Ardeschir Ghofrani, J. Simon R. Gibbs, Kate Gibson, Kimberly C. Gilmour, Barbara Girerd, Nicholas S. Gleadall, Sarah Goddard, David B. Goldstein, Keith Gomez, Pavels Gordins, David Gosal, Stefan Gräf, Jodie Graham, Luigi Grassi, Daniel Greene, Lynn Greenhalgh, Andreas Greinacher, Paolo Gresele, Philip Griffiths, Sofia Grigoriadou, Russell J. Grocock, Detelina Grozeva, Mark Gurnell,

Scott Hackett, Charaka Hadinnapola, William M. Hague, Rosie Hague, Matthias Haimel, Matthew Hall, Helen L. Hanson, Eshika Haque, Kirsty Harkness, Andrew R. Harper, Claire L L. Harris, Daniel Hart, Ahamad Hassan, Grant Hayman, Alex Henderson, Archana Herwadkar, Jonathan Hoffman, Simon Holden, Rita Horvath, Henry Houlden, Arjan C C. Houweling, Luke S. Howard, Fengyuan Hu, Gavin Hudson, Joseph Hughes, Aarnoud P. Huissoon, Marc Humbert, Sean Humphray, Sarah Hunter, Matthew Hurles, Melita Irving, Louise Izatt, Roger James, Sally A. Johnson, Stephen Jolles, Jennifer Jolley, Dragana Josifova, Neringa Jurkute, Tim Karten, Johannes Karten, Mary A. Kasanicki, Hanadi Kazkaz, Rashid Kazmi, Peter Kelleher, Anne M. Kelly, Wilf Kelsall, Carly Kempster, David G. Kiely, Nathalie Kingston, Robert Klima, Nils Koelling, Myrto Kostadima, Gabor Kovacs, Ania Koziell, Roman Kreuzhuber, Taco W. Kuijpers, Ajith Kumar, Dinakantha Kumararatne, Manju A. Kurian, Michael A. Laffan, Fiona Lalloo, Michele Lambert, Allan Lawrie, D. Mark Layton, Nick Lench, Claire Lentaigne, Tracy Lester, Adam P. Levine, Rachel Linger, Hilary Longhurst, Lorena E. Lorenzo, Eleni Louka, Paul A. Lyons, Rajiv D. Machado, Robert V. MacKenzie Ross, Bella Madan, Eamonn R. Maher, Jesmeen Maimaris, Samantha Malka, Sarah Mangles, Rutendo Mapeta, Kevin J. Marchbank, Stephen Marks, Hugh S. Markus, Hanns Ulrich Marschall, Andrew Marshall, Jennifer Martin, Mary Mathias, Emma Matthews, Heather Maxwell, Paul McAlinden, Mark I. McCarthy, Harriet McKinney, Aoife McMahon, Stuart Meacham, Adam J. Mead, Ignacio Medina Castello, Karyn Megy, Sarju G G. Mehta, Michel Michaelides, Carolyn Millar, Shehla N. Mohammed, Shahin Moledina, David Montani, Anthony T. Moore, Joannella Morales, Nicholas W. Morrell, Monika Mozere, Keith W. Muir, Andrew D. Mumford, Andrea H. Nemeth, William G. Newman, Michael Newnham, Sadia Noorani, Paquita Nurden, Jennifer O'Sullivan, Samya Obaji, Chris Odhams, Steven Okoli, Andrea Olschewski, Horst Olschewski, Kai Ren Ong, S. Helen Oram, Elizabeth Ormondroyd, Willem H. Ouwehand, Claire Palles, Sofia Papadia, Soo Mi Park, David Parry, Smita Patel, Joan Paterson, Andrew Peacock, Simon H H. Pearce, John Peden, Kathelijne Peerlinck, Christopher J. Penkett, Joanna Pepke-Zaba, Romina Petersen, Clarissa Pilkington, Kenneth E.S. Poole, Radhika Prathalingam, Bethan Psaila, Angela Pyle, Richard Quinton, Shamima Rahman, Stuart Rankin, Anupama Rao, F. Lucy Raymond, Paula J. Rayner-Matthews, Christine Rees, Augusto Rendon, Tara Renton, Christopher J. Rhodes, Andrew S.C. Rice, Sylvia Richardson, Alex Richter, Leema Robert, Irene Roberts, Anthony Rogers, Sarah J. Rose, Robert Ross-Russell, Catherine Roughley, Noemi B. A Roy, Deborah M. Ruddy, Omid Sadeghi-Alavijeh, Moin A. Saleem, Nilesh Samani, Crina Samarghitean, Alba Sanchis-Juan, Ravishankar B. Sargur, Robert N. Sarkany, Simon Satchell, Sinisa Savic, John A. Sayer, Genevieve Sayer, Laura Scelsi, Andrew M. Schaefer, Sol Schulman, Richard Scott, Marie Scully, Claire Searle, Werner Seeger, Arjune Sen, W. A.Carrock Sewell, Denis Seyres, Neil Shah, Olga Shamardina, Susan E. Shapiro, Adam C. Shaw, Patrick J. Short, Keith Sibson, Lucy Side, Ilenia Simeoni, Michael A A. Simpson, Matthew C. Sims, Suthesh Sivapalaratnam, Damian Smedley, Katherine R. Smith, Kenneth G.C. Smith, Katie Snape, Nicole Soranzo, Florent Soubrier, Laura Southgate, Olivera Spasic-Boskovic, Simon Staines, Emily Staples, Hannah Stark, Jonathan Stephens, Charles Steward, Kathleen E. Stirrups, Alex Stuckey, Jay Suntharalingam, Emilia M. Swietlik, Petros Syrris, R. Campbell Tait, Kate Talks, Rhea Y.Y. Tan, Katie Tate, John M. Taylor, Jenny C. Taylor, James E. Thaventhiran, Andreas C. Themistocleous, Ellen Thomas, David Thomas, Moira J. Thomas, Patrick Thomas, Kate Thomson, Adrian J. Thrasher, Glen Threadgold, Chantal Thys, Tobias Tilly, Marc Tischkowitz, Catherine Titterton, John A. Todd, Cheng Hock Toh, Bas Tolhuis, Ian P. Tomlinson, Mark Toshner, Matthew Traylor, Carmen Treacy, Paul Treadaway, Richard Trembath, Salih Tuna, Wojciech Turek, Ernest Turro, Philip Twiss, Tom Vale, Chris Van Geet, Natalie van Zuydam, Maarten Vandekuilen, Anthony M. Vandersteen, Marta Vazquez-Lopez, Julie von Ziegenweidt, Anton Vonk Noordegraaf, Annette Wagner, Quinten Waisfisz, Suellen M. Walker, Neil Walker, Klaudia Walter, James S. Ware, Hugh Watkins, Christopher Watt, Andrew R. Webster, Lucy Wedderburn, Wei Wei, Steven B. Welch, Julie Wessels, Sarah K. Westbury, John Paul Westwood, John Wharton, Deborah

Whitehorn, James Whitworth, Andrew O. M Wilkie, Martin R. Wilkins, Catherine Williamson, Brian T. Wilson, Edwin K.S. Wong, Nicholas Wood, Yvette Wood, Christopher Geoffrey Woods, Emma R R. Woodward, Stephen J. Wort, Austen Worth, Michael Wright, Katherine Yates, Patrick F.K. Yong, Timothy Young, Ping Yu, Patrick Yu-Wai-Man, Eliska Zlamalova, Nathalie Kingston, Neil Walker, Christopher J. Penkett, Kathleen Freson, Kathleen E. Stirrups, and F. Lucy Raymond. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*, 583(7814):96–102, jul 2020. 73, 75

[142] Courtney E. French, Isabelle Delon, Helen Dolling, Alba Sanchis-Juan, Olga Shamardina, Karyn Mégy, Stephen Abbs, Topun Austin, Sarah Bowdin, Ricardo G. Branco, Helen Firth, Salih Tuna, Timothy J. Aitman, Sofie Ashford, Willian J. Astle, David L. Bennet, Marta Bleda, Keren J. Carss, Patrick F. Chinnery, Sri V.V. Deevi, Debra Fletcher, Daniel P. Gale, Stefan F. Gräf, Fengyuan Hu, Roger James, Mary A. Kasanicki, Nathalie Kingston, Ania B. Koziell, Hana Lango Allen, Eamonn R. Maher, Hugh S. Markus, Stuart Meacham, Nicholas W. Morrell, Christopher J. Penkett, Irene Roberts, Kenneth G.C. Smith, Hannah Stark, Kathleen E. Stirrups, Ernest Turro, Hugh Watkins, Catherine Williamson, Timothy Young, John R. Bradley, Willem H. Ouwehand, F. Lucy Raymond, Shruti Agrawal, Ruth Armstrong, Kathryn Beardsall, Gusztav Belteki, Marion Bohatschek, Susan Broster, Rosalie Campbell, Rajiv Chaudhary, Cristine Costa, Angela D'Amore, Annie Fitzsimmons, Jennifer Hague, Joanne Harley, Shazia Hoodbhoy, Riaz Kayani, Wilf Kelsall, Sarju G. Mehta, Roddy O'Donnell, Samantha O'Hare, Amanda Ogilvy-Stuart, Stergios Papakostas, Soo Mi Park, Alasdair Parker, Nazima Pathan, Matina Prapa, Audrienne Sammut, Richard Sandford, Katherine Schon, Yogen Singh, Kelly Spike, Ana Lisa Taylor Tavares, Doris Wari-Pepple, Hilary S. Wong, C. Geoff Woods, F. Lucy Raymond, David H. Rowitch, and F. Lucy Raymond. Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. *Intensive Care Medicine*, 45(5):627–636, may 2019. 75

[143] Konrad J. Karczewski, Laurent C. Francioli, Grace Tiao, Beryl B. Cummings, Jessica Alföldi, Qingbo Wang, Ryan L. Collins, Kristen M. Laricchia, Andrea Ganna, Daniel P. Birnbaum, Laura D. Gauthier, Harrison Brand, Matthew Solomonson, Nicholas A. Watts, Daniel Rhodes, Moriel Singer-Berk, Eleina M. England, Eleanor G. Seaby, Jack A. Kosmicki, Raymond K. Walters, Katherine Tashman, Yossi Farjoun, Eric Banks, Timothy Poterba, Arcturus Wang, Cotton Seed, Nicola Whiffin, Jessica X. Chong, Kaitlin E. Samocha, Emma Pierce-Hoffman, Zachary Zappala, Anne H. O'Donnell-Luria, Eric Vallabh Minikel, Ben Weisburd, Monkol Lek, James S. Ware, Christopher Vittal, Irina M. Armean, Louis Bergelson, Kristian Cibulskis, Kristen M. Connolly, Miguel Covarrubias, Stacey Donnelly, Steven Ferriera, Stacey Gabriel, Jeff Gentry, Namrata Gupta, Thibault Jeandet, Diane Kaplan, Christopher Llanwarne, Ruchi Munshi, Sam Novod, Nikelle Petrillo, David Roazen, Valentin Ruano-Rubio, Andrea Saltzman, Molly Schleicher, Jose Soto, Kathleen Tibbetts, Charlotte Tolonen, Gordon Wade, Michael E. Talkowski, Carlos A. Aguilar Salinas, Tariq Ahmad, Christine M. Albert, Diego Ardissino, Gil Atzmon, John Barnard, Laurent Beaugerie, Emelia J. Benjamin, Michael Boehnke, Lori L. Bonnycastle, Erwin P. Bottinger, Donald W. Bowden, Matthew J. Bown, John C. Chambers, Juliana C. Chan, Daniel Chasman, Judy Cho, Mina K. Chung, Bruce Cohen, Adolfo Correa, Dana Dabelea, Mark J. Daly, Dawood Darbar, Ravindranath Duggirala, Josée Dupuis, Patrick T. Ellinor, Roberto Elosua, Jeanette Erdmann, Tõnu Esko, Martti Färkkilä, Jose Florez, Andre Franke, Gad Getz, Benjamin Glaser, Stephen J. Glatt, David Goldstein, Clicerio Gonzalez, Leif Groop, Christopher Haiman, Craig Hanis, Matthew Harms, Mikko Hiltunen, Matti M. Holi, Christina M. Hultman, Mikko Kallela, Jaakko Kaprio, Sekar Kathiresan, Bong Jo Kim, Young Jin Kim, George Kirov, Jaspal Kooner, Seppo Koskinen, Harlan M. Krumholz, Subra Kugathasan, Soo Heon Kwak, Markku Laakso, Terho Lehtimäki, Ruth J.F. Loos, Steven A. Lubitz, Ronald C.W. Ma, Daniel G. MacArthur, Jaume Marrugat, Kari M. Mattila, Steven McCarroll, Mark I. McCarthy, Dermot McGovern, Ruth McPherson, James B. Meigs, Olle Melander, An-

dres Metspalu, Benjamin M. Neale, Peter M. Nilsson, Michael C. O'Donovan, Dost Ongur, Lorena Orozco, Michael J. Owen, Colin N.A. Palmer, Aarno Palotie, Kyong Soo Park, Carlos Pato, Ann E. Pulver, Nazneen Rahman, Anne M. Remes, John D. Rioux, Samuli Ripatti, Dan M. Roden, Danish Saleheen, Veikko Salomaa, Nilesh J. Samani, Jeremiah Scharf, Heribert Schunkert, Moore B. Shoemaker, Pamela Sklar, Hilkka Soininen, Harry Sokol, Tim Spector, Patrick F. Sullivan, Jaana Suvisaari, E. Shyong Tai, Yik Ying Teo, Tuomi Tiinamaija, Ming Tsuang, Dan Turner, Teresa Tusie-Luna, Erkki Vartiainen, James S. Ware, Hugh Watkins, Rinse K. Weersma, Maija Wessman, James G. Wilson, Ramnik J. Xavier, Benjamin M. Neale, Mark J. Daly, and Daniel G. MacArthur. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809):434–443, may 2020. 75

[144] Simon N Stacey, Birte Kehr, Julius Gudmundsson, Florian Zink, Aslaug Jonasdottir, Sigurjon A Gudjonsson, Asgeir Sigurdsson, Bjarni V Halldorsson, Bjarni A Agnarsson, Kristrun R Benediktsdottir, Katja K H Aben, Sita H Vermeulen, Ruben G Cremers, Angeles Panadero, Brian T Helfand, Phillip R Cooper, Jenny L Donovan, Freddie C Hamdy, Viorel Jinga, Ichiro Okamoto, Jon G Jonasson, Laufey Tryggvadottir, Hrefna Johannsdottir, Anna M Kristinsdottir, Gisli Masson, Olafur T Magnusson, Paul D Iordache, Agnar Helgason, Hannes Helgason, Patrick Sulem, Daniel F Gudbjartsson, Augustine Kong, Eirikur Jonsson, Rosa B Barkardottir, Gudmundur V Einarsson, Thorunn Rafnar, Unnur Thorsteinsdottir, Ioan N Mates, David E Neal, William J Catalona, José I Mayordomo, Lambertus A Kiemeney, Gudmar Thorleifsson, and Kari Stefansson. Insertion of an SVA-E retrotransposon into the CASP8 gene is associated with protection against prostate cancer. *Human molecular genetics*, 25(5):1008–18, mar 2016. 80

[145] Lindsay M. Payer and Kathleen H. Burns. Transposable elements in human genetic disease, dec 2019. 80, 121

[146] Belén de la Morena-Barrio, Jonathan Stephens, María Eugenia de la Morena-Barrio, Luca Stefanucci, José Padilla, Antonia Miñano, Nicholas Gleadall, Juan Luis García, María Fernanda López-Fernández, Pierre Emmanuel Morange, Marja K. Puurunen, Anetta Undas, Francisco Vidal, F. Lucy Raymond, Vicente Vicente García, Willem H. Ouwehand, Javier Corral, and Alba Sanchis-Juan. Long-read sequencing resolves structural variants in SERPINC1 causing antithrombin deficiency and identifies a complex rearrangement and a retrotransposon insertion not characterized by routine diagnostic methods, aug 2020. 97, 98, 99

[147] Belén De la Morena-Barrio, Nina Borràs, Agustín Rodríguez-Alén, María E. de la Morena-Barrio, Juan L. García-Hernández, José Padilla, Carlos Bravo-Pérez, Antonia Miñano, Noelia Rollón, Javier Corral, Francisco Vidal, and Vicente Vicente. Identification of the first large intronic deletion responsible of type I antithrombin deficiency not detected by routine molecular diagnostic methods, 2019. 99

[148] I. Martínez-Martínez, D. J.D. Johnson, M. Yamasaki, J. Navarro-Fernández, A. Ordóñez, V. Vicente, J. A. Huntington, and J. Corral. Type II antithrombin deficiency caused by a large in-frame insertion: Structural, functional and pathological relevance. *Journal of Thrombosis and Haemostasis*, 10(9):1859–1866, sep 2012. 106

[149] Kenneth A. Bauer, Tam M. Nguyen-Cao, and Jeffrey B. Spears. Issues in the Diagnosis and Management of Hereditary Antithrombin Deficiency, sep 2016. 111

[150] Riten Kumar, Anthony K.C. Chan, Jennifer E. Dawson, Julie D. Forman-Kay, Walter H.A. Kahr, and Suzan Williams. Clinical presentation and molecular basis of congenital antithrombin deficiency in children: A cohort study. *British Journal of Haematology*, 166(1):130–139, 2014. 111

[151] S Kuhle, D A Lane, K Jochmanns, C Male, P Quehenberger, K Lechner, and I Pabinger. Homozygous antithrombin deficiency type II (99 Leu to Phe mutation) and childhood thromboembolism. *Thrombosis and haemostasis*, 86(4):1007–11, 2001. 111, 113

[152] Nazan Sarper, Christelle Orlando, Uğur Demirsoy, Sema A Gelen, and Kristin Jochmans. Homozygous antithrombin deficiency in adolescents presenting with lower extremity thrombosis and renal complications: two case reports from Turkey. *Journal of pediatric hematology/oncology*, 36(3):e190–2, 2014. 111, 113

[153] V Limperger, A Franke, G Kenet, S Holzhauer, V Picard, R Junker, C Heller, C Gille, D Manner, K Kurnik, R Knoefler, R Mesters, S Halimeh, and U Nowak-Göttl. Clinical and laboratory characteristics of paediatric and adolescent index cases with venous thromboembolism and antithrombin deficiency. An observational multicentre cohort study. *Thrombosis and haemostasis*, 112(3):478–85, 2014. 111

[154] Verena Limperger, Gili Kenet, Neil A Goldenberg, Christine Heller, Susanne Holzhauer, Ralf Junker, Ulrich C Klostermeier, Ralf Knoefler, Karin Kurnik, Anne Krümpel, Rolf Mesters, Michael Stach, Guy Young, and Ulrike Nowak-Göttl. Impact of high-risk thrombophilia status on recurrence among children with a first non-central-venous-catheter-associated VTE: an observational multicentre cohort study. *British journal of haematology*, 175(1):133–40, 2016. 111, 113

[155] C H van Ommen, H Heijboer, E J van den Dool, B A Hutten, and M Peters. Pediatric venous thromboembolic disease in one single center: congenital prothrombotic disorders and the clinical outcome. *Journal of thrombosis and haemostasis : JTH*, 1(12):2516–22, dec 2003. 111, 113

[156] B Schmidt and M Andrew. Neonatal thrombosis: report of a prospective Canadian and international registry. *Pediatrics*, 96(5 Pt 1):939–43, 1995. 111

[157] Akira Ishiguro, Chibueze Chioma Ezinne, Nobuaki Michihata, Hisaya Nakadate, Atsushi Manabe, Masashi Taki, and Midori Shima. Pediatric thromboembolism: a national survey in Japan. *International Journal of Hematology*, 105(1):52–58, 2017. 111

[158] U Nowak-Göttl, R von Kries, and U Göbel. Neonatal symptomatic thromboembolism in Germany: two year survey. *Archives of disease in childhood. Fetal and neonatal edition*, 76(3):F163–7, 1997. 111

[159] C H van Ommen, H Heijboer, H R Büller, R A Hirasing, H S Heijmans, and M Peters. Venous thromboembolism in childhood: a prospective two-year registry in The Netherlands. *The Journal of pediatrics*, 139(5):676–81, 2001. 111

[160] Janet Y.K. Yang and Anthony K.C. Chan. Pediatric Thrombophilia. *Pediatric Clinics of North America*, 60(6):1443–1462, 2013. 111

[161] Pierre Toulon, Micheline Berruyer, Marie Brionne-François, François Grand, Dominique Lasne, Caroline Telion, Julien Arcizet, Roberta Giacomello, and Neila De Pooter. Age dependency for coagulation parameters in paediatric populations. Results of a multicentre study aimed at defining the age-specific reference ranges. *Thrombosis and haemostasis*, 116(1):9–16, 2016. 112

[162] M Andrew, B Paes, R Milner, M Johnston, L Mitchell, D M Tollefsen, and P Powers. Development of the human coagulation system in the full-term infant. *Blood*, 70(1):165–72, 1987. 112

[163] A Greenway, M.P Massicotte, and P Monagle. Neonatal thrombosis and its treatment. *Blood Reviews*, 18(2):75–84, 2004. 113

[164] G. Kenet, L. K. Lutkhoff, M. Albisetti, T. Bernard, M. Bonduel, L. Brandao, S. Chabrier, A. Chan, G. DeVeber, B. Fiedler, H. J. Fullerton, N. A. Goldenberg, E. Grabowski, G. Gunther, C. Heller, S. Holzhauer,

A. Iorio, J. Journeycake, R. Junker, F. J. Kirkham, K. Kurnik, J. K. Lynch, C. Male, M. Manco-Johnson, R. Mesters, P. Monagle, C. H. van Ommen, L. Raffini, K. Rostasy, P. Simioni, R. D. Strater, G. Young, and U. Nowak-Gottl. Impact of Thrombophilia on Risk of Arterial Ischemic Stroke or Cerebral Sinovenous Thrombosis in Neonates and Children: A Systematic Review and Meta-Analysis of Observational Studies. *Circulation*, 121(16):1838–1847, 2010. 113

[165] F. J. Berfelo, K. J. Kersbergen, C. H. van Ommen, P. Govaert, H. L. M. van Straaten, B.-T. Poll-The, G. van Wezel-Meijler, R. J. Vermeulen, F. Groenendaal, L. S. de Vries, and T. R. de Haan. Neonatal Cerebral Sinovenous Thrombosis From Symptom to Outcome. *Stroke*, 41(7):1382–1388, 2010. 113

[166] P Monagle, M Adams, M Mahoney, K Ali, D Barnard, M Bernstein, L Brisson, M David, S Desai, M F Scully, J Halton, S Israels, L Jardine, M Leaker, P McCusker, M Silva, J Wu, R Anderson, M Andrew, and M P Massicotte. Outcome of pediatric thromboembolic disease: a report from the Canadian Childhood Thrombophilia Registry. *Pediatric research*, 47(6):763–6, 2000. 113

[167] L. Raffini. Thrombophilia in Children: Who to Test, How, When, and Why? *Hematology*, 2008(1):228–235, 2008. 113

[168] C. H. van Ommen and Saskia Middeldorp. Thrombophilia in Childhood: To Test or Not to Test. *Seminars in Thrombosis and Hemostasis*, 37(07):794–801, 2011. 113

[169] Paul Monagle, Anthony K C Chan, Neil A Goldenberg, Rebecca N Ichord, Janna M Journeycake, Ulrike Nowak-Göttl, and Sara K Vesely. Antithrombotic therapy in neonates and children: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*, 141(2 Suppl):e737S–e801S, 2012. 113

[170] Chakri Gavva, Ravindra Sarode, and Ayesha Zia. A clinical audit of thrombophilia testing in pediatric patients with acute thromboembolic events: impact on management. *Blood advances*, 1(25):2386–2391, 2017. 113

[171] Ulrike Nowak-Göttl, Heleen van Ommen, and Gili Kenet. Thrombophilia testing in children: What and when should be tested? *Thrombosis Research*, 164:75–78, 2018. 113

[172] D. Tormene, Paolo Simioni, Paolo Prandoni, Francesca Franz, Patrizia Zerbinati, Giulio Tognin, and Antonio Girolami. The incidence of venous thromboembolism in thrombophilic children: a prospective cohort study. *Blood*, 100(7):2403–2405, 2002. 113

[173] B. K. Mahmoodi, J. L.P. Brouwer, M. K. Ten Kate, W. M. Lijfering, N. J.G.M. Veeger, A. B. Mulder, H. C. Kluin-Nelemans, and J. van Der Meer. A prospective cohort study on the absolute risks of venous thromboembolism and predictive value of screening asymptomatic relatives of patients with hereditary deficiencies of protein S, protein C or antithrombin. *Journal of Thrombosis and Haemostasis*, 8(6):1193–1200, jun 2010. 113

[174] Frank Rühle and Monika Stoll. Advances in predicting venous thromboembolism risk in children. *British journal of haematology*, 180(5):654–665, 2018. 113

[175] Véronique Picard, Ulrike Nowak-Göttl, Christine Biron-Andreani, Marc Fouassier, Corinne Frere, Michèle Goualt-Heilman, Emmanuel de Maistre, Sandra Regina, Lucia Rugeri, Catherine Ternisien, Catherine Trichet, Christine Vergnes, Martine Aiach, and Martine Alhenc-Gelas. Molecular bases of antithrombin deficiency: twenty-two novel mutations in the antithrombin gene. *Human mutation*, 27(6):600, 2006. 114

[176] Beate Luxembourg, Mathias D'Souza, Stephanie Körber, and Erhard Seifrieda. Prediction of the pathogenicity of antithrombin sequence variations by in silico methods. *Thrombosis Research*, 135(2):404–409, feb 2015. 114

[177] Rajarshi Ghosh, Ninad Oak, and Sharon E. Plon. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biology*, 18(1), nov 2017. 114

[178] Kaija Javela, Sari Engelbarth, Leena Hiltunen, Pirjo Mustonen, and Marja Puurunen. Great discrepancy in antithrombin activity measured using five commercially available functional assays. *Thrombosis Research*, 132(1):132–137, jul 2013. 114

[179] Søren R. Kristensen, B. Rasmussen, S. Pedersen, and L. Bathum. Detecting antithrombin deficiency may be a difficult task - More than one test is necessary [3], mar 2007. 114

[180] M. Merz, M. Böhm-Weigert, S. Braun, P. C. Cooper, R. Fischer, K. Hickey, A. Steffan, B. Kemkes-Matthes, and S. Kitchen. Clinical multicenter evaluation of a new FXa-based Antithrombin assay. *International Journal of Laboratory Hematology*, 33(5):498–506, oct 2011. 114

[181] P. C. Cooper, F. Coath, M. E. Daly, and M. Makris. The phenotypic and genetic assessment of antithrombin deficiency, jun 2011. 114

[182] P. L. Harper, M. Daly, J. Price, P. F. Edgar, and R. W. Carrell. Screening for heparin binding variants of antithrombin. *Journal of Clinical Pathology*, 44(6):477–479, 1991. 114

[183] Airlie J. McCoy, Xue Yuan Pei, Richard Skinner, Jan Pieter Abrahams, and Robin W. Carrell. Structure of β-antithrombin and the effect of glycosylation on antithrombin's heparin affinity and activity. *Journal of Molecular Biology*, 326(3):823–833, feb 2003. 115

[184] Irene Martínez-Martínez, José Navarro-Fernández, Alice Østergaard, Ricardo Gutiérrez-Gallego, José Padilla, Nataliya Bohdan, Antonia Miñano, Cristina Pascual, Constantino Martínez, María Eugenia de la Morena-Barrio, Sonia Aguila, Shona Pedersen, Søren Risom Kristensen, Vicente Vicente, and Javier Corral. Amelioration of the severity of heparin-binding antithrombin mutations by posttranslational mosaicism. *Blood*, 120(4):900–4, 2012. 115

[185] T. Koide, S. Odani, K. Takahashi, T. Ono, and N. Sakuragawa. Antithrombin III Toyama: Replacement of arginine-47 by cysteine in hereditary abnormal antithrombin III that lacks heparin-binding ability. *Proceedings of the National Academy of Sciences of the United States of America*, 81(2 I):289–293, 1984. 115

[186] Steven T. Olson, Benjamin Richard, Gonzalo Izaguirre, Sophia Schedin-Weiss, and Peter G.W. Gettins. Molecular mechanisms of antithrombin-heparin regulation of blood clotting proteinases. A paradigm for understanding proteinase regulation by serpin family protein proteinase inhibitors, nov 2010. 115

[187] K. Okajima, H. Abe, S. Maeda, M. Motomura, M. Tsujihata, S. Nagataki, H. Okabe, and K. Takatsuki. Antithrombin III Nagasaki (Ser116-Pro): A heterozygous variant with defective heparin binding associated with thrombosis. *Blood*, 81(5):1300–1305, mar 1993. 115

[188] Alec Mushunje, Aiwu Zhou, James A. Huntington, Jacqueline Conard, and Robin W. Carrell. Antithrombin 'DREUX' (Lys114Glu): A variant with complete loss of heparin affinity. *Thrombosis and Haemostasis*, 88(3):436–443, sep 2002. 115

[189] Marie Laurence Flahaux and Hein De Haas. African migration: trends, patterns, drivers. *Comparative Migration Studies*, 4(1):1, dec 2016. 116

[190] Liang Tang and Yu Hu. Ethnic diversity in the genetics of venous thromboembolism. *Thrombosis and Haemostasis*, 114(5):901–909, 2015. 116

[191] John A. Heit, Michele G. Beckman, Paula L. Bockenstedt, Althea M. Grant, Nigel S. Key, Roshni Kulkarni, Marilyn J. Manco-Johnson, Stephan Moll, Thomas L. Ortel, and Claire S. Philipp. Comparison of characteristics from White- and Black-Americans with venous thromboembolism: A cross-sectional study. *American Journal of Hematology*, 85(7):467–471, jul 2010. 116

[192] Maurizio Margaglione and Elvira Grandone. Population genetics of venous thromboembolism: A narrative review, feb 2011. 116

[193] Raj K. Patel, Elizabeth Ford, Jill Thumpston, and Roopen Arya. Risk factors for venous thrombosis in the black population. *Thrombosis and Haemostasis*, 90(5):835–838, 2003. 116

[194] Roxana Daneshjou, Larisa H. Cavallari, Peter E. Weeke, Konrad J. Karczewski, Katarzyna Drozda, Minoli A. Perera, Julie A. Johnson, Teri E. Klein, Carlos D. Bustamante, Dan M. Roden, Christian Shaffer, Joshua C. Denny, James L. Zehnder, and Russ B. Altman. Population-specific single-nucleotide polymorphism confers increased risk of venous thromboembolism in African Americans. *Molecular Genetics and Genomic Medicine*, 4(5):513–520, sep 2016. 116

[195] Thijs E. van Mens, Marcel Levi, and Saskia Middeldorp. Evolution of factor V Leiden, 2013. 116

[196] T. E. Van Mens, U. N. Joensen, Z. Bochdanovits, A. Takizawa, J. Peter, N. Jørgensen, P. B. Szecsi, J. C.M. Meijers, H. Weiler, E. Rajpert-De Meyts, S. Repping, and S. Middeldorp. Factor V Leiden is associated with increased sperm count. *Human Reproduction*, 32(11):2332–2339, nov 2017. 116

[197] Peter H. Sudmant, Tobias Rausch, Eugene J. Gardner, Robert E. Handsaker, Alexej Abyzov, John Huddleston, Yan Zhang, Kai Ye, Goo Jun, Markus His Yang Fritz, Miriam K. Konkel, Ankit Malhotra, Adrian M. Stütz, Xinghua Shi, Francesco Paolo Casale, Jieming Chen, Fereydoun Hormozdiari, Gargi Dayama, Ken Chen, Maika Malig, Mark J.P. Chaisson, Klaudia Walter, Sascha Meiers, Seva Kashin, Erik Garrison, Adam Auton, Hugo Y.K. Lam, Xinmeng Jasmine Mu, Can Alkan, Danny Antaki, Taejeong Bae, Eliza Cerveira, Peter Chines, Zechen Chong, Laura Clarke, Elif Dal, Li Ding, Sarah Emery, Xian Fan, Madhusudan Gujral, Fatma Kahveci, Jeffrey M. Kidd, Yu Kong, Eric Wubbo Lameijer, Shane McCarthy, Paul Flicek, Richard A. Gibbs, Gabor Marth, Christopher E. Mason, Androniki Menelaou, Donna M. Muzny, Bradley J. Nelson, Amina Noor, Nicholas F. Parrish, Matthew Pendleton, Andrew Quitadamo, Benjamin Raeder, Eric E. Schadt, Mallory Romanovitch, Andreas Schlattl, Robert Sebra, Andrey A. Shabalin, Andreas Untergasser, Jerilyn A. Walker, Min Wang, Fuli Yu, Chengsheng Zhang, Jing Zhang, Xiangqun Zheng-Bradley, Wanding Zhou, Thomas Zichner, Jonathan Sebat, Mark A. Batzer, Steven A. McCarroll, Ryan E. Mills, Mark B. Gerstein, Ali Bashir, Oliver Stegle, Scott E. Devine, Charles Lee, Evan E. Eichler, and Jan O. Korbel. An integrated map of structural variation in 2,504 human genomes. *Nature*, 526(7571):75–81, sep 2015. 118

[198] Cathelijn J.F. Waaijer, Marcel G.T. Winter, Christianne M.A. Reijnders, Daniëlle de Jong, S. John Ham, Judith V.M.G. Bovée, and Károly Szuhai. Intronic deletion and duplication proximal of the EXT1 gene: A novel causative mechanism for multiple osteochondromas. *Genes Chromosomes and Cancer*, 52(4):431–436, apr 2013. 119

[199] Céline Delaloy, Emilie Elvira-Matelot, Maud Clemessy, Xiao-ou Zhou, Martine Imbert-Teboul, Anne-Marie Houot, Xavier Jeunemaitre, and Juliette Hadchouel. Deletion of WNK1 first intron results in misregulation of both isoforms in renal and extrarenal tissues. *Hypertension (Dallas, Tex. : 1979)*, 52(6):1149–54, dec 2008. 119

[200] Przemyslaw Szafranski, Yaping Yang, Melissa U. Nelson, Matthew J. Bizzarro, Raffaella A. Morotti, Claire Langston, and Paweł Stankiewicz. Novel FOXF1 Deep Intronic Deletion Causes Lethal Lung Developmental Disorder, Alveolar Capillary Dysplasia with Misalignment of Pulmonary Veins. *Human Mutation*, 34(11):1467–1471, nov 2013. 119

[201] S Kaneda, N Horie, K Takeishi, A Takayanagi, T Seno, and D Ayusawa. Regulatory sequences clustered at the 5' end of the first intron of the human thymidylate synthase gene function in cooperation with the promoter region. *Somatic cell and molecular genetics*, 18(5):409–15, sep 1992. 119

[202] Hui Y. Xiong, Babak Alipanahi, Leo J. Lee, Hannes Bretschneider, Daniele Merico, Ryan K.C. Yuen, Yimin Hua, Serge Gueroussov, Hamed S. Najafabadi, Timothy R. Hughes, Quaid Morris, Yoseph Barash, Adrian R. Krainer, Nebojsa Jojic, Stephen W. Scherer, Benjamin J. Blencowe, and Brendan J. Frey. The human splicing code reveals new insights into the genetic determinants of disease. *Science*, 347(6218), jan 2015. 119

[203] A. P.Jason de Koning, Wanjun Gu, Todd A. Castoe, Mark A. Batzer, and David D. Pollock. Repetitive elements may comprise over Two-Thirds of the human genome. *PLoS Genetics*, 7(12), dec 2011. 120

[204] Anbo Zhou, Timothy Lin, and Jinchuan Xing. Evaluating nanopore sequencing data processing pipelines for structural variation identification. *Genome Biology*, 20(1), nov 2019. 120

[205] Cheng Ran Lisa Huang, Kathleen H. Burns, and Jef D. Boeke. Active transposition in genomes. *Annual Review of Genetics*, 46:651–675, 2012. 121

[206] Dustin C. Hancks and Haig H. Kazazian. Roles for retrotransposon insertions in human disease, may 2016. 121

[207] Rosanna Pallotta, Leda Dalprà, Monica Miozzo, Tamara Ehresmann, and Paola Fusilli. A patient defines the interstitial 1q deletion syndrome characterized by antithrombin III deficiency. *American Journal of Medical Genetics*, 104(4):282–286, dec 2001. 122

[208] N. J. Beauchamp, M. Makris, F. E. Preston, I. R. Peake, and M. E. Daly. Major structural defects in the antithrombin gene in four families with type I antithrombin deficiency. Partial/complete deletions and rearrangement of the antithrombin gene. *Thrombosis and Haemostasis*, 83(5):715–721, 2000. 122

[209] Véronique Picard, Jian Min Chen, Brigitte Tardy, Marie Françoise Aillaud, Christine Boiteux-Vergnes, Marie Dreyfus, Joseph Emmerich, Cécile Lavenu-Bombled, Ulrike Nowak-Göttl, Nathalie Trillot, Martine Aiach, and Martine Alhenc-Gelas. Detection and characterisation of large SERPINC1 deletions in type i inherited antithrombin deficiency. *Human Genetics*, 127(1):45–53, jan 2010. 122

[210] Mark J.P. Chaisson, Richard K. Wilson, and Evan E. Eichler. Genetic variation and the de novo assembly of human genomes, nov 2015. 122

[211] Peter Ebert, Peter A. Audano, Qihui Zhu, Bernardo Rodriguez-Martin, David Porubsky, Marc Jan Bonder, Arvis Sulovari, Jana Ebler, Weichen Zhou, Rebecca Serra Mari, Feyza Yilmaz, Xuefang Zhao, PingHsun Hsieh, Joyce Lee, Sushant Kumar, Jiadong Lin, Tobias Rausch, Yu Chen, Jingwen Ren, Martin Santamarina, Wolfram Höps, Hufsah Ashraf, Nelson T. Chuang, Xiaofei Yang, Katherine M. Munson, Alexandra P. Lewis, Susan Fairley, Luke J. Tallon, Wayne E. Clarke, Anna O. Basile, Marta Byrska-Bishop, André Corvelo, Uday S. Evani, Tsung-Yu Lu, Mark J.P. Chaisson, Junjie Chen, Chong Li, Harrison Brand, Aaron M. Wenger, Maryam Ghareghani, William T. Harvey, Benjamin Raeder, Patrick Hasenfeld, Allison A. Regier, Haley J. Abel, Ira M. Hall, Paul Flicek, Oliver Stegle, Mark B. Gerstein, Jose M.C. Tubio, Zepeng Mu, Yang I. Li, Xinghua Shi, Alex R. Hastie, Kai Ye, Zechen Chong, Ashley D. Sanders, Michael C. Zody, Michael E. Talkowski, Ryan E. Mills, Scott E. Devine, Charles Lee, Jan O. Korbel, Tobias Marschall, and Evan E.

Eichler. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science*, page eabf7117, feb 2021. 122, 123

[212] Claudia M.B. Carvalho and James R. Lupski. Mechanisms underlying structural variant formation in genomic disorders, apr 2016. 123