How to cite this thesis / dissertation (APA referencing method):

Surname, Initial(s). (Date). *Title of doctoral thesis* (Doctoral thesis). Retrieved from http://scholar.ufs.ac.za/*rest of thesis URL on KovsieScholar*

Surname, Initial(s). (Date). *Title of master's dissertation* (Master's dissertation). Retrieved from http://scholar.ufs.ac.za/*rest of thesis URL on KovsieScholar*

Development and implementation of a real-time reverse transcriptase polymerase chain reaction assay to detect the intron 22 inversion in haemophilia A patients.

By

Jean Frederick Kloppers

December 2017

Submitted in fulfilment of the requirements for the degree MMedSc with specialisation in Human Molecular Biology

in the

Department of Haematology and Cell Biology, Faculty of Health Sciences, University of the Free State, Bloemfontein, South Africa

Supervisor: Dr W.J. Janse van Rensburg

Co-Supervisor: Dr G.M. Marx

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

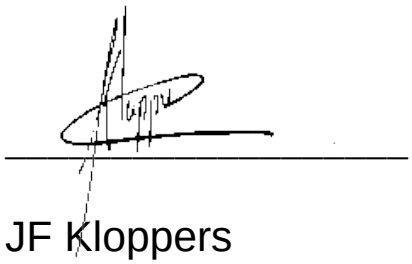
UFS UV

DECLARATION

(i) I, Jean Frederick Kloppers, declare that the master's research dissertation or interrelated, publishable manuscripts / published articles that I herewith submit at the University of the Free State, is my independent work and that I have not previously submitted it for a qualification at another institution of higher education.

(ii) I, Jean Frederick Kloppers, hereby declare that I am aware that the copyright is vested in the University of the Free State.

(iii) I, Jean Frederick Kloppers, hereby declare that all royalties as regards intellectual property that was developed during the course of and/or in connection with the study at the University of the Free State, will accrue to the University.



JF Kloppers

ACKNOWLEDGEMENTS

I would like to thank the following people for making this study possible:

- Doctor Wattie Janse van Rensburg for his exceptional support, leadership and expert advice to plan and complete this study. Thank you for keeping me motivated along the way.
- Doctor Gerda Marx, for her expert advice, motivation and support. Your door was always open, and you always gave sound advice.
- Professor MJ Coetzee and his staff for their expert advice on haemophilia A and his contribution to this study.
- Doctor André de Kock for the advice and use of his laboratory and equipment.
- Professor L Rossetti for her expert advice on the Inverse-PCR method.
- The National Health Laboratory Service for the research grant that funded the study, and the National Research Foundation for the bursary that made it possible for me to complete the study.
- The Department of Haematology and Cell Biology for providing the necessary resources to perform this study.
- Friends and colleagues at the Department of Haematology and Cell Biology for their support and motivation.
- My family and friends for always believing in me and for their continuous support throughout my studies.
- My wife – Ilze – thank you for always supporting me and encouraging me to achieve my goals. You have kept me goal driven, and I dedicate this dissertation to you.
- Lastly, to my Heavenly Father, for always being the rock beneath my feet.

'Ek is tot alles instaat deur Hom wat my krag gee..." Fiippense 4:13

Table of Contents **Page**

List of Figures and Tables

Figure/Table **Page**

List of Abbreviations and Acronyms

%	Percentage
∆Rn	Normalised fluorescent reported value minus the baseline signal
®	Registered
°C	Degrees Celsius
3’	3 prime
5’	5 prime
AD	anno Domini
APC	Activated protein C
aPTT	Activated Partial Thromboplastin time
ATP	Adenosine triphosphate
BcI I	Restriction enzyme that originates from *Bacillus caldolyticus*
bp	Base pair(s)
cDNA	Complimentary DNA
-COOH	Carboxyl group
CpG	Cytosine followed by a Guanine
Ct-value	Cycle threshold value
dL	Decilitre
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
Dup22	Duplication 22
EDTA	Ethylenediaminetetraacetic acid
F8A	Factor VIII-associated gene A
F8B	Factor VIII-associated gene B

FIX	Factor IX
FIX_a	Activated Factor IX
FVIII	Factor VIII
FVIII	Factor VIII gene
FVIII22I	Factor VIII intron 22 transcript
$FVIII_a$	Activated FVIII
$FVIII_B$	Factor VIII associated B gene transcript
$FVIII_{FL}$	Factor VIII full-length transcript
FX	Factor X
FX_a	Activated Factor X
g	Centrifugal force
HIV	Human Immunodeficiency Virus
IgG	Immunoglobulin G
int1	Intron 1
int22	Intron 22
int22h	Intron 22 homologous region
inv22	Intron 22 inversion
I-PCR	Inverse PCR
IS-PCR	Inverse-Shifting PCR
ISTH	International Society on Thrombosis and Haemostasis
IU	Infusion unit
kb	Kilo base pair(s)
LALIGN	Local alignment tool
LD-PCR	Long Distance PCR
LOD	Limit of Detection
M	Molar(s)

Mb	Mega base pair(s)
min	Minutes
MIQE	Minimum Information for Publication of Quantitative Real-Time PCR Experiments
mL	Millilitre(s)
mM	Millimolar(s)
mRNA	Messenger RNA
MT	Mutant-type
N	Total amount
NaCl	Sodium Chloride
NC_	Accession number for genomic references sequences on NCBI
NCBI	National Centre for Biotechnology Information
NEB	New England Biolabs
NH_2-	Amino group
Nm_	Accession number for mRNA reference sequences on NCBI
nt	nucleotides
NTC	Non-Template Control
PCR	Polymerase Chain Reaction
pH	Potential of Hydrogen
PLP	Pre-amplification of LD-PCR
p-value	Calculated probability
rpm	Revolutions per minute
RNA	Ribonucleic acid
-RT	cDNA synthesis control that does not contain reverse transcriptase
RT-PCR	Reverse transcriptase PCR
sec	Seconds

SNP	Single Nucleotide Polymorphism
SOP	Standard Operating Procedure
S-PCR	Sub-cycle PCR
TAE	Tris-acetate-EDTA
TE	Tris-EDTA
Tm	Melting temperature
™	Trademark
U	Units
UKNEQAS	The United Kingdom National External Quality Assessment Service
V	Volt
VNTR	Variable Number of Tandem Repeats
vWF	Von Willebrand Factor
WT	Wild-type
XR_	Accession number for RNA references sequences on NCBI
ZR	Zymo Research
µL	Microlitre(s)
µM	Micromolar(s)

Summary

Keywords: Haemophilia A •*Factor VIII* gene • Intron 22 inversion (Int22) • Inv22 detection methods • Inv22 conventional RT-PCR detection • Inv22 real-time RT-PCR detection method

Background: Haemophilia A is a bleeding disorder with an incidence rate of one in 5,000-10,000 males worldwide. The disorder has a substantial impact on patients and their caregivers and is underdiagnosed in developing countries such as South Africa, where there is a lack of genetic research in the field. Haemophilia A is caused by mutations in the *FVIII* gene, with the most common mutation in severe haemophilia A patients being inv22 (45%). The current inv22 detection methods have disadvantages. The aim of this study was to develop novel inv22 detection methods that could overcome the disadvantages of the current inv22 detection methods.

Methodology: We recruited three controls, including one non-haemophilic volunteer (C1) and two non-related inv22 confirmed severe haemophilia A patients (C2 and C3). The controls were used to evaluate the inv22 I-PCR detection method and to develop a conventional and a real-time RT-PCR inv22 detection method, respectively. The controls were Sanger sequenced to confirm the results of the PCR detection methods. A further 60 participants (35 haemophilia A patients, 18 possible haemophilia A carriers and seven healthy volunteers) from central South Africa were recruited for this study. The newly optimised detection methods were used to screen the 60 participants for the inv22. The PCR results for the 60 participants were confirmed with Sanger sequence analysis. The inv22 results for the conventional RT-PCR, real-time RT-PCR and Sanger sequence analysis in this study were subsequently compared.

Results and discussion: The inv22 I-PCR detection method could not be evaluated after several attempts, and after troubleshooting measures were exhausted. The main reason for this failure was contributed to the fact that the inv22 I-PCR detection method did not have any ‘built-in’ quality control steps that allowed for successful troubleshooting. The need for an uncomplicated inv22 detection method was thus, further motivated. The inv22 conventional and real-time RT-PCR detection methods were successfully optimised. Control C1 was confirmed to be inv22 negative and

controls C2 and C3 to be inv22 positive with Sanger sequence analysis. The novel detection methods overcame some of the disadvantages associated with the previous detection methods, especially referring to cost and turnaround time. The two newly developed methods were used to screen the 60 study participants for the inv22. The results for the 60 study participants were confirmed with Sanger sequence analysis. When the different detection methods' results were compared, only a single difference was found between the two assays. The inv22 was found in 29.41% of our severe haemophilia A population.

Conclusion: Rapid, accurate, reliable and cost-effective inv22 conventional RT-PCR and real-time RT-PCR detection methods were developed and validated. The newly developed inv22 detection methods overcome the disadvantages of the current inv22 detection methods and will allow for the wide-spread screening of haemophilia A patients and possible carriers for the inv22. In this study, the inv22 was found in only 29.41% of severe haemophilia A patients in our population, which was considerably lower than the globally reported 45%, however, this may be attributed to the relatively small sample size of the study.

Opsomming

Sleutelwoorde: Hemofilie A • *Faktor VIII* geen • Intron 22 inversie (int22) • Int22 inversie bepalings metodes • Int22 inversie konvensionele trutranskripsie PKR • Int22 inversie werklike-tyd trutranskripsie PKR.

Agtergrond: Hemofilie A is ‘n oorerflike bloedingsiekte met ‘n insidensie van een in 5,000-10,000 mans wêreldwyd. Die siekte het ‘n groot impak op die pasiënte en hulle versorgers. Studies het bevestig dat hemofilie A ondergediagnoseer is in ontwikkelende lande soos Suid-Afrika, waar daar ‘n groot tekort aan genetiese studies in hierdie spesifieke veld is. Hemofilie A word veroorsaak deur mutasies in die *FVIII* geen en die int22 inversie (inv22) is die bekendste mutasie wat voorkom in pasiënte met erenstige hemofilie A (45%). Die huidige inv22 bepalingsmetodes het verskeie nadele. Die doel van hierdie studie was om nuwe metodes te ontwikkel wat die nadele van die huidige metodes kan oorkom, om sodoende alle moontlike pasiënte en moontlike draers te toets.

Metodologie: Drie kontroles, insluitend een nie-hemofilie vrywilliger (C1) en twee nie-verwante inv22 bevestigde hemofilie pasiënte (C2 en C3) is gewerf. Die kontroles is gebruik om die I-PKR metode te evalueer en die inv22 konvensionele en werklike-tyd trutanskripsie PKR metodes te ontwikkel. Die PKR metodes se resultate is met Sanger volgorde-bepaling bevestig. Daar is ‘n verdere 60 deelnemers (insluitend 35 hemofilie pasiënte, 18 moontlike draers en sewe gesonde vrywilligers) vanuit sentraal Suid-Afrika gewerf. Die geoptimiseerde metodes is gebruik om die 60 deelnemers te toets vir die inv22 en weereens is die resultate met Sanger volgorde-bepaling bevestig. Die voorkoms van die inv22 in sentraal dele van Suid-Afrika is bepaal.

Resultate en bespreking: Die inv22 I-PKR metode kon nie suksesvol geëvalueer word nie na verskeie probeerslae en nadat alle probleemoplossingsmaatreëls uitgeput is. Die onsuksesvolle evaluasie van die metode word toegeskryf daaraan dat die metode nie ingeboude kontrole stappe het nie en gevolglik kon probleemoplossingsmaatreëls nie slaag nie. Dit beklemtoon die behoefte aan ‘n ongekompliseerde inv22 metode. Die inv22 konvensionele en werklike-tyd trutranskripsie PKR metodes was suksesvol ontwikkel en geoptimiseer. Kontrole C1

is as inv22 negatief, en kontroles C2 en C3 is as inv22 positief, met Sanger volgorde-bepaling bevestig. Hierdie nuwe tegnieke elimineer van die nadele wat geassosieer word met die huidige inv22 bepalingsmetodes, veral verwysend na die kostes en die omkeertyd. Die twee nuwe inv22 bepaling metodes is gebruik om vir die inv22 te toets in die 60 studie deelnemers. Weereens is die inv22 resultate vir die 60 studie deelnemers bevestig met Sanger volgorde-bepaling. Die resultate van die nuwe inv22 metodes toon goeie korrelasie met die uitsondering van een resultaat. Gevolglik is die inv22 in 29.41% die van sentraal Suid-Afrikaanse hemofilie A populasie gevind.

Gevolgtrekking: ‘n Bekostigbare, betroubare, akkurate en vinnige inv22 konvensionele trutranskripsie en werklike-tyd trutranskripsie PKR metode is in hierdie studie ontwikkel en gevalideer. Die nuut geontwikkelde inv22 bepalingmetodes elimineer van die nadele wat geassosieer word met die huidige inv22 bepalingmetodes en sal gevolglik lei tot die wyd-verspreide bepaling van die inv22 in hemofilie A pasiënte en moontlike draers. Die inv22 is in slegs 29.41% van die huidige erenstige hemofilie A populasie gevind, dit is heelwat laer as die beskryfde globale 45%, maar kan moontlik aan ons relatief lae studie populasiegrootte toegeskryf word.

Chapter 1: Introduction

Haemostasis refers to the body's physiological response to prevent bleeding. When vascular damage occurs, there is a linked interaction between the vessel wall, circulating platelets and blood coagulation factors. The haemostatic response is a balanced mechanism involving procoagulant and anticoagulant components to prevent bleeding and extensive clot development (Hoffbrand & Moss, 2011a). Bleeding is caused by vascular and platelet abnormalities that disrupt the haemostatic response (Hoffbrand & Moss, 2011b). Several human bleeding disorders have been well described, and for this study, the focus is on Haemophilia A.

Haemophilia A is a bleeding disorder characterised by the deficiency of coagulation factor VIII (FVIII). The disorder is inherited in an X-linked recessive manner that predominately affects male individuals, reportedly affecting one in 5,000 to 10,000 males worldwide (Soucie *et al.,* 1998; Hoffbrand & Moss, 2011a). The prevalence rate of haemophilia is estimated at 105 to 160 per million of the male population (Tezanos Pinto & Ortiz, 2004). In 2017, the World Federation of Hemophilia reported in their 2016 annual global survey that 149,764 individuals were affected by haemophilia A and 29,712 individuals were affected by haemophilia B, respectively (World Federation of Hemophilia, 2017). The survey data on South Africa reported that 1,848 individuals were affected by haemophilia A and 358 individuals were affected by haemophilia B, respectively (World Federation of Hemophilia, 2017). If the incident rate of haemophilia is taken into consideration, more than half of individuals with haemophilia are currently undiagnosed in South Africa, with an estimated population of 56.52 million people (Statistics South Africa, 2017). This problem is, however, not unique to South Africa, as haemophilia remains underdiagnosed in other developing countries as well (Tezanos Pinto & Ortiz, 2004). Currently, in South Africa, there is only one accredited service unit performing genetic testing for haemophilia A. As a result, genetic research is inhibited in this field of study (Mahlangu & Medical and Scientific Council of the South African Haemophilia Federation, 2009).

Haemophilia A is caused by mutations in the *FVIII* gene which is located on the X chromosome. The *FVIII* gene is considered to be large as it consists of 26 exons (Gitschier *et al.,* 1984). There are various unique mutations associated with the disorder (Factor VIII Gene Variant Database, 2017) and the mutational heterogeneity, together with the size of the gene complicates the execution of molecular studies (Green *et al.,* 1991). It has been reported that in approximately 45% of individuals with the severe form of the disorder, there is an inversion present in intron 22 (int22) of the *FVIII* gene (Lakich *et al.,* 1993). The detection of mutations plays an essential role in predicting treatment outcome, as FVIII mutations have been associated with FVIII inhibitor development (Jayandharan *et al.,* 2012; Eckhardt *et al.,* 2013). Furthermore, mutation detection allows for carrier and prenatal diagnosis to be conducted (Kessler *et al.,* 2014). Consequently, molecular studies could assist in the development of gene therapy, currently the only curative prospect for haemophilia A (Rogers & Herzog, 2015).

The int22 inversion (inv22) was first detected by a Southern Blot method (Lakich *et al.,* 1993). This method proves to be accurate but has the disadvantage of being time-consuming, labour intensive and in general an out-dated technique (Kumar *et al.,* 2013). The use of hazardous radioactive reagents is also a prohibitive factor of Southern Blot methods. The Long-distance polymerase chain reaction (LD-PCR) and inverse PCR (I-PCR) have also been applied to detect the inv22 (Liu *et al.,* 1998, Rossetti *et al.,* 2005). However, LD-PCR is difficult to standardise and is sensitive to minor reductions in DNA quality (Polakova *et al.,* 2003, Kumar *et al.,* 2013). Although I-PCR has been reported to be reliable (Rossetti *et al.,* 2005), it remains time-consuming, entails multiple steps and is difficult to standardise (Kloppers & Janse van Rensburg, 2017). The DNA-based analysis of the inversion is complicated by the size of the *FVIII* gene (186 kb), intron 22 (32.4 kb) and the homologous sequence involved in the inv22 (9.5 kb) (Gitschier *et al.,* 1984, Naylor *et al.,* 1995). Therefore the current inv22 detection methods have several disadvantages, and thus, the need to develop novel, reproducible and cost-effective detection methods is paramount.

Chapter 2: Literature review

2. Introduction

This chapter entails the literature review of this study. The history of haemophilia A is presented, and thereafter different aspects of the disorder are discussed such as patterns of inheritance, classification, treatment and diagnosis. Furthermore, an oversight of the genetic components of haemophilia A is given as well as the mutations that cause haemophilia A. The methods that are used in the detection of the inv22 are discussed as well as real-time PCR which is a relevant method in the context of this study.

2.1 The history of haemophilia A

Haemophilia dates back centuries, and the first written reference appeared two years AD. The Jewish writing refers to the third son of a woman being exempted from circumcision by a Rabbi after his brothers died of bleeding complications following circumcision (Rosner, 1969). The continuous studies on haemophilia have contributed to the understanding of haemostasis and allowed the development of a sufficient treatment regime for haemophilia (Ingram, 1976). The first medical description of haemophilia was given by Dr John Otto in 1803 (Lee, 2010). The work of Bulloch and Fildes identified haemophilia cases by sex, inheritance and symptoms (Ingram, 1976). Their casework of 1000 references and 200 pedigrees of haemophilia families made a tremendous impact on what is known about haemophilia today (Lee, 2010).

Haemophilia also had an impact on the royal families of Europe (Stevens, 1999). Queen Victoria of England was a carrier of haemophilia and passed it on to her son Leopold. Two of her daughters were also carriers, and the disorder was passed to a further three grandsons and six great-grandsons. Queen Victoria's daughter Alice, who was a carrier of haemophilia, had a daughter named Alexandra. Alexandra married Tsar Nicholas II of Russia.

Alexandra introduced haemophilia into the Romanov line with the birth of their son, Alexis. It is known that Alexandra placed her trust in Rasputin, a monk that believed in ‘complementary medicine’, to care for Alexis. Her misplaced trust, however, led to the death of both Rasputin and her own family, bringing about the extinction of Queen Victoria’s mutant haemophilia carrying-gene (McKusick, 1965; Ingram, 1976). Therefore, the impact of haemophilia and its genetics have been widely felt in the history of the world.

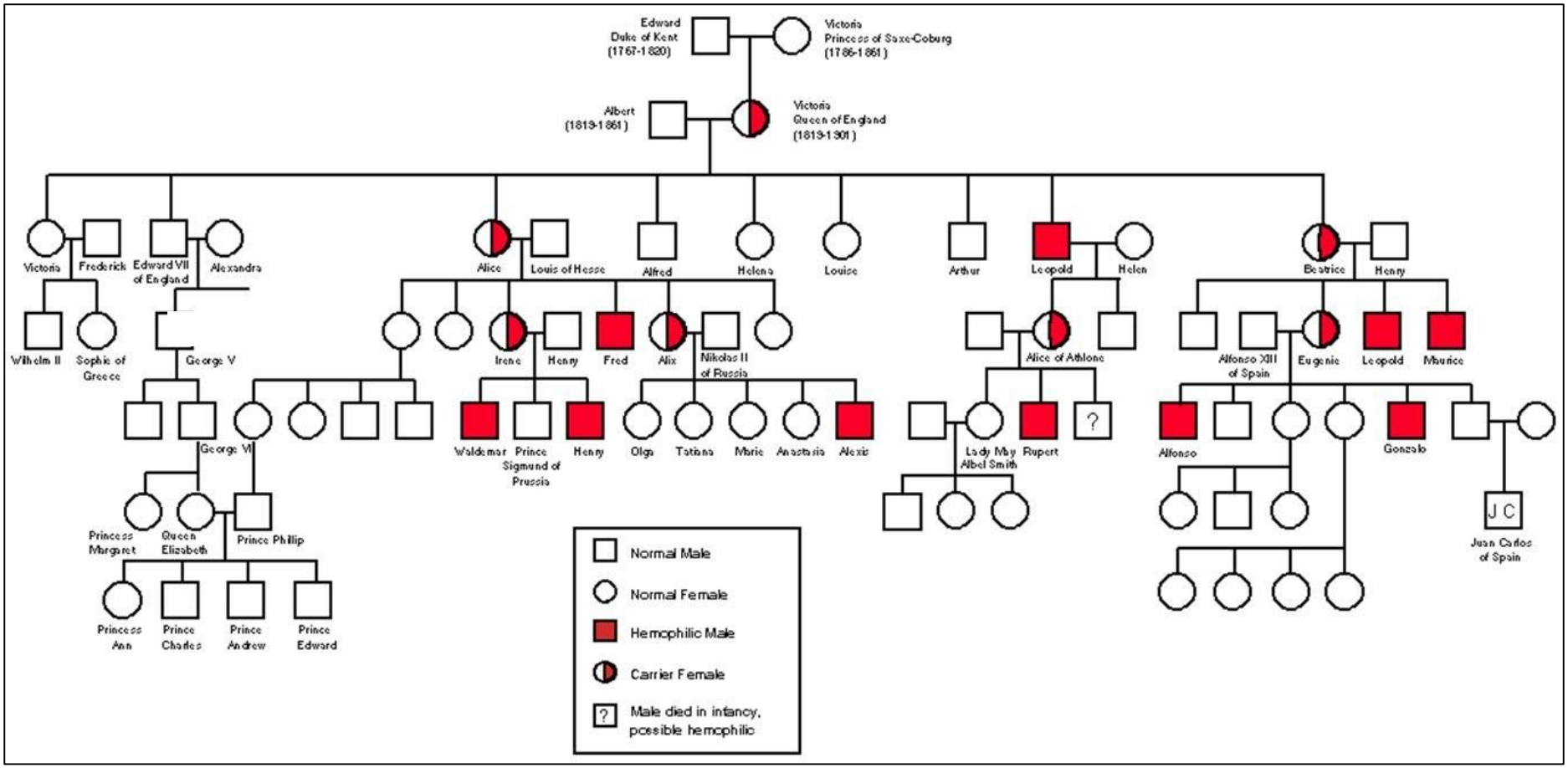


Figure 2.1: Pedigree of Queen Victoria's family and an indication of family members affected by haemophilia (Available: http://www.emersonkent.com/source_documents/victoria_hemophilia.htm, Accessed 20 January 2016).

2.2 Inheritance of haemophilia A

Haemophilia A is an X-linked recessive disorder, predominately affecting male individuals. The disorder is mostly familial but appears to be sporadic in one-third of cases (Hoffbrand & Moss, 2011a). A study of 804 haemophilia pedigrees confirmed this familial and sporadic pattern of inheritance (Kasper & Lin, 2007). The inheritance pattern differs in male and female individuals. A male individual with haemophilia A will carry over the mutant X-chromosome to all of his daughters (100%), but to none of his sons. There is a 50% chance that the children of a female carrier will inherit the mutant X-chromosome. The daughters of the female carrier can, thus, also become a carrier of the disorder, while her sons may be affected with the disorder (Kessler *et al.,* 2014).

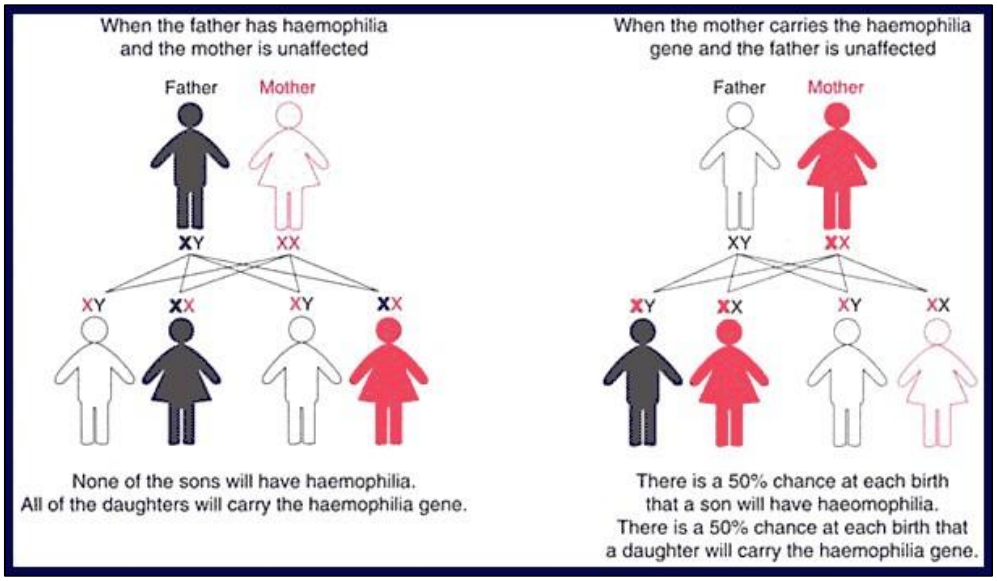


Figure 2.2: Example of an X-linked recessive inheritance pattern as seen in the case of haemophilia A (Available at: https://www.health.qld.gov.au/haemophilia/html/i_bleed_disorders). Accessed: 22 April 2018.

Case reports of females affected by haemophilia A have been described. The genetic mechanisms that are responsible for these females to be affected are extreme lyonization, homozygosity, genetic abnormalities leading to a female inheriting one gene for haemophilia A, and abnormalities involving the X-chromosome (Lusher & Mcmillan, 1978). The extensive inactivation of the X-chromosome, known as skewed X-inactivation, has also been reported to cause haemophilia A in females (Renault *et al.,* 2007; Knobe *et al.,* 2008). In most cases, women are carriers of haemophilia A, with half the amount of normal FVIII present. In cases of extreme lyonization, women tend to have deficient levels of FVIII (Plug *et al.,* 2006).

2.3 Classification and clinical manifestation of haemophilia A

Haemophilia A is classified based on two different parameters according to severity. Firstly, on the clinical bleeding symptoms of the patient and secondly on plasma procoagulant (FVIII) levels (White *et al.,* 2001). The latter is the preferred method, according to the International Society on Thrombosis and Haemostasis (ISTH) and can be classified as follows (White *et al.,* 2001):

Severe: < 1 IU/dL FVIII; Moderate: 1 to 5 IU/dL FVIII and Mild: 5 to 40 IU/dL FVIII.

The clinical manifestation of haemophilia A is correlated to the severity of the disorder (Mahlangu *et al.,* 2008). In cases where haemophilia A is suspected, it is crucial for the physician to inquire about the patient's bleeding history (Zaiden *et al.,* 2014a). Haemophilia A is a systemic haemorrhagic disorder and sites of bleeding vary from life-threating to non-life threating. The life-threating bleeds are usually those including organ bleeds, intracranial bleeds, massive gastrointestinal bleeds, neck and throat bleeds, muscle compartment and genitourinary bleeds. Mucosal bleeding and bleeding into the joints (haemarthroses) and muscle are classified as non-life threating. Haemarthroses occurs more often in haemophilia A patients, while life-threating bleeds are less frequent (Mahlangu *et al.,* 2008).

Severe haemophilia A patients display an anticipated reduction in health-related quality of life in comparison to those that have the disorder in the moderate or mild form (Miners *et al.,* 1999). Severe haemophilia A accounts for between 43% and 70%

of all haemophilia A cases, and it can present in children as young as one year old (Zaiden, 2014b). Recurrent and spontaneous bleeds are most often associated with patients that have the disorder in the severe form (Hoffbrand & Moss, 2011a). A study cohort of 150 patients revealed that patients with severe haemophilia had lower levels of physical function and increased bodily pain. These patients tend to have a substantial amount of bleeding episodes per year (Solovieva, 2001). Life- threating bleeds can occur after injuries, accidents and surgery in patients with severe haemophilia A. In patients with moderate haemophilia A, symptoms include haemarthroses and bleeding into the muscles after minor injuries. However, excessive bleeding occurs after surgery and dental extractions. There is no spontaneous bleeding that occurs in patients with mild haemophilia A. Bleeding can, however, occur after surgery, dental extractions or trauma (Bolton-Maggs & Pasi, 2003).

2.4 Treatment of haemophilia A

There is currently no cure for haemophilia A. Treatment, thus, plays the most prominent role in the management of the disorder. The primary goal of haemophilia A treatment is to replace the deficient FVIII to counteract the associated symptoms, consequently FVIII replacement therapy is currently the primary treatment option (Bolton-Maggs & Pasi, 2003). Although FVIII therapy is central to haemophilia A treatment, a comprehensive treatment plan needs to be considered that involves a multidisciplinary team of caregivers from various fields such as genetics, haematology, psychosocial and physiotherapy. The patients need to be educated on haemophilia A, and it is essential that they are educated about home therapies. Data collection on bleeding episodes and applicable research should be conducted if possible (Srivastava *et al.,* 2013); especially considering the prospect of gene therapy being the only curative prospect for haemophilia A. The goal of gene therapy is to elevate FVIII levels to just above 1 IU/dL in people with severe haemophilia A, in order to decrease mortality rates (Hoffbrand & Moss, 2011a; Sharma *et al*., 2014). Gene therapy is seen as a curative, once off treatment, and the successful implementation thereof will decrease the cost associated with FVIII replacement therapy and will allow a better quality of life as patients are to spend much less time receiving treatment. There is progress in haemophilia A gene therapy. Jobson and Brown (2017) reported

promising early-stage clinical results that show a high amount of circulating FVIII and a reduction of bleeding episodes in severe haemophilia A patients. This was found in severe haemophilia A patients that was treated with an adeno-associated virus vector that encodes for FVIII. The development of FVIII inhibitors remains a risk for haemophilia A gene therapy (Jobson & Brown, 2017). A study by Ragarajan *et al.* (2017) also reported the use of the adeno-associated virus vector that encodes for FVIII to treat for haemophilia A. The results were also promising, with the stabilisation of haemostasis in all of the severe haemophilia A patients that were treated. However, as Sharma *et al.* (2014) previously reported, the adverse side effects and safety of FVIII gene therapy still needs to be carefully evaluated.

2.4.1 FVIII replacement therapy

The primary treatment for haemophilia A is FVIII replacement therapy. The purpose of this treatment is to arrest bleeding by increasing the patient's FVIII levels (Canadian Hemophilia Society, 2016). Before 1964, fresh or fresh-frozen plasma was used as a source of FVIII. In 1964, cryoprecipitated FVIII was introduced and allowed a ten-fold increase in FVIII concentration in comparison to fresh or fresh-frozen plasma. Concentrated FVIII continued in the 1970's, but the widespread use of these FVIII products contributed to hepatitis and HIV epidemics in people with haemophilia. The epidemic occurred due to the transmitted hepatitis and HIV viruses that affected the plasma pool of which the FVIII products were constituted of and subsequently, this led to the purification and removal of viruses from the blood products. The purification techniques included various heat treatments, which was followed by immune-affinity chromatographic techniques that utilised monoclonal antibodies to purify the FVIII products (Roberts, 1991). In 1984, the *FVIII* gene was successfully cloned, which led to the production of recombinant FVIII. Recombinant FVIII is currently the therapeutic substance of choice to treat haemophilia A as it is free of any transmitting viruses (American Society of Hematology, 2008). Currently, there are a few long lasting recombinant FVIII products available. These products include Recombinate®, which is generated from the full *FVIII* gene and consists of human albumin and animal proteins. The addition of Refacto®, which is a B-domain deleted recombinant FVIII product, increased the secretion of the recombinant FVIII from cells. Second-

generation products which include Kogenate® doesn't contain any human albumin, and third-generation recombinant FVIII products, such as Novoeight® and Kovaltry® are free of any human or animal products. Finally, a fourth-generation recombinant FVIII product, Nuwiq®, produced from human embryonic kidney, contains the full post-translational modifications to FVIII and is reported to decrease immunogenicity (Lieuw, 2017). Although new generation FVIII replacement therapies look promising the use of these therapies, both recombinant and plasma derived, still has a significant disadvantage in that it may lead to the development of FVIII inhibitors (Farrugia, 2015). However, reports of studies that found third-generation recombinant FVIII products less prone to FVIII inhibitor development than second-generation products have been described. It remains to be seen if newly developed recombinant FVIII products completely reduce the development of FVIII inhibitors in previously untreated patients (Lieuw, 2017).

2.4.2 FVIII inhibitors

FVIII inhibitors are immunoglobulin G (IgG) antibodies that form against FVIII. The formation of these antibodies leads to ineffective treatment. FVIII inhibitors are either alloantibodies that are developed against FVIII replacement therapy (exogenous FVIII) in inherited haemophilia A, or autoantibodies that are developed against endogenous FVIII in acquired haemophilia A (Shima, 2006). FVIII inhibitors occur in about 30% of patients that have severe haemophilia A (Astermark *et al.,* 2005). The genetic and environmental factors that are associated with inhibitor development have been evaluated, but the results obtained have been inconsistent (Oldenburg *et al.,* 2004). It has, however, been reported that patients with large rearrangements of the *FVIII* gene are associated with a higher risk to develop FVIII inhibitors (Goodeve 2003). In a study where 113 haemophilia A families were evaluated, 74 families tested positive for the intron22 inversion (inv22) and of these 74 families, 40% were associated with the development of FVIII inhibitors (Astermark *et al.,* 2005). Although the inv22 has been considered to be a high-risk mutation for FVIII inhibitor development, it has been reported that in most cohorts the inv22 is only associated with 20% of inhibitor development (Bardi & Astermark, 2015). Contrary, African Americans with haemophilia A have a 3x higher incidence rate for FVIII inhibitor development than

Caucasians, and thus, the association between ethnic origin and inhibitor development is worth investigating (Ragni *et al.,* 2009; Gunasekera *et al.,* 2015). This possible association further highlights the need for genetic screening of people with haemophilia, as the presence of a specific mutation may have far-reaching treatment and management implications.

2.5 The genetics of Factor VIII

FVIII, also known as 'antihaemophilic' factor is part of the coagulation factors involved in the coagulation cascade (Hoffbrand & Moss, 2011b). The *FVIII* gene is located on the long arm of the X-chromosome (Xq.28) and is approximately 186 kb long. The *FVIII* gene consists of 26 exons (**Figure 2.3**) and approximately 9000 bp of coding sequence. The *FVIII* gene has among other, six large introns, with int22 being 32.4 kb in size (Gitschier *et al.,* 1984). Int22 of the *FVIII* gene contains a CpG island that serves as a bidirectional promoter for genes known as Factor VIII associated A (*F8A*) and B (*F8B*) genes, respectively (Levinson *et al.,* 1990; Levinson *et al.,* 1992). The *F8A* gene, nested within int22 of the *FVIII* gene, is intron-less and transcribed in the opposite direction of the *FVIII* gene. Two further copies of the F8A gene were located 1.1 Mb from the *FVIII* gene (Levinson *et al.,* 1990). The *F8B* gene is located 122 bp from the *F8A* gene. As *F8B* is transcribed in the same direction as the *FVIII* gene, the first exon is found in int22, and the transcript is spliced to exons 23 to 26 of the *FVIII* gene (Levinson *et al.,* 1992).

The *FVIII* locus contains two regulated transcription units, allowing the synthesis of two exonic-sequences. The transcript variants are $FVIII_{FL}$ (Full length) and $FVIII_{b}$, respectively (**Figure 2.3**). The $FVIII_{FL}$ transcript comprises of exons 1 to 26 of the *FVIII* gene (Pandey *et al.,* 2013; Sauna *et al.,* 2015). The $FVIII_{b}$ transcript contains 169 bp of the *F8B* gene and a further 2429 bp from the last four exons of the *FVIII* gene, which is identical to exons 23 to 26 of the $FVIII_{FL}$ transcript (Sauna *et al.,* 2015). The FVIII protein is a glycoprotein, synthesised as a single polypeptide chain of 2351 amino acids (Callaghan & Kaufman, 2008). The domain structure of FVIII, (NH_2)-A1-a1-A2-a2-B-a3-A3-C1-C2-(COOH), contains a triplicated A domain, a duplicated C domain and a glycosylated B domain (Takeyama *et al.,* 2013). After post-translational modifications, the FVIII heterodimer is released from the cell, and it then binds to von

Willebrand Factor (vWF) (Callaghan & Kaufman, 2008). The FVIII heterodimer forms a complex with vWF that protects the molecule from degradation (Vlot *et al.,* 1995). The heterodimer comprises a light chain (Domain structure: A3-C1-C2) and a heavy chain (Domain structure: A1-A2). The respective chains bind non-covalently with a metal ion-dependent linkage (Fay, 2004; Shen *et al.,* 2008).

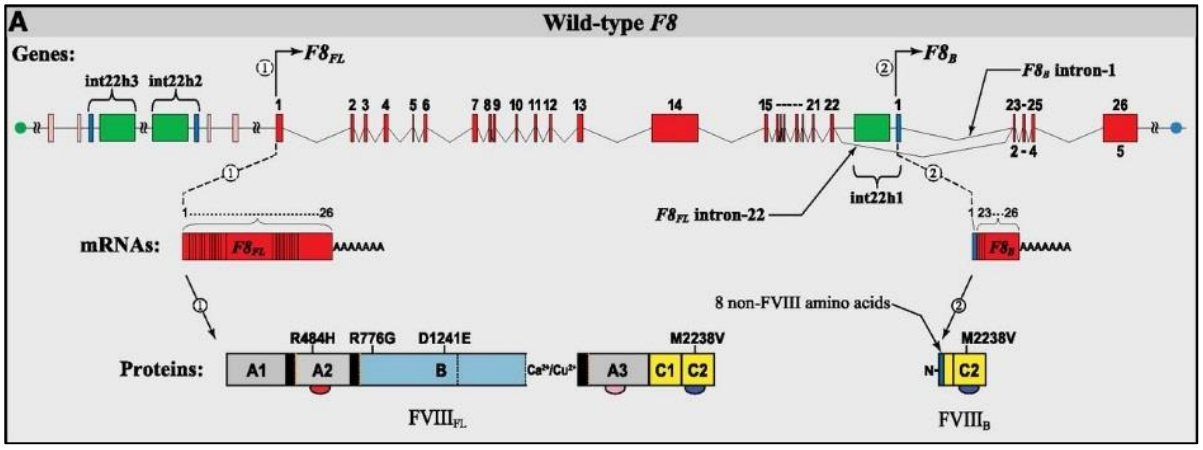


Figure 2.3: Structure of the *FVIII* gene, mRNA and proteins (Copied from Sauna *et al.,* 2015).

2.6 The function of FVIII

Haemostasis consists of two stages, namely primary and secondary haemostasis. Primary haemostasis is the first response to arrest bleeding and platelets play a critical role in forming a thrombus at the site of vascular injury (Clemetson, 2012). Secondary haemostasis involves the stabilization of the thrombus by fibrin through the coagulation cascade. The coagulation cascade is comprised of enzyme complexes, and these enzyme complexes require circulating coagulation factors, which are either enzyme precursors or cofactors to aid in the process of forming a stable haemostatic plug at the site of bleeding (Hoffbrand & Moss, 2011b). Activated FVIII (FVIIIa) is specific to the intrinsic-tenase complex (**Figure 2.4**), where it acts as a cofactor to activated factor IX (FIXa). The binding of FVIIIa to FIXa increases the catalytic efficiency of FIXa. The intrinsic-tenase complex promotes the activation of factor X (FX) to activated factor X (FXa), causing a burst of thrombin to be released, thus,

allowing the conversion of fibrinogen to fibrin. Fibrin is then cross-linked by factor XIII which stabilises the haemostatic plug that has formed at the site of vascular injury (Dahlback, 2000; Hoffman & Monroe, 2007; Ziedens & Mann, 2010). The deficiency of coagulation factors involved in the stabilization of the haemostatic plug subsequently leads to the failure of secondary haemostasis and causes excessive and prolonged bleeding (Bolton-Maggs & Pasi, 2003).

The role of FVIII has been explained in the traditional coagulation cascade, but emphasis needs to be placed on the cell-based model of coagulation as well. The cell-based model of coagulation consists of two linked phases, namely initiation and propagation. In the initiation phase, tissue factor binds to activated factor FVII (FVIIIa) and the complex that is formed activates both FIX and FX. The activated platelet membrane acts as a catalyst and prothrombin is converted to thrombin solely by FXa. Thrombin then activates FV, FVIII and FXI, and the propagation phase then occurs on activated platelets. During this phase, the tenase-complex is formed when FVIIIa binds to FIXa, and FX is more potently activated by this complex. Subsequently, FXa then forms part of the prothrombinase complex and large amounts of thrombin is released. Once more, the deficiency of FVIII will impact the propagation phase of the cell-based model of coagulation and will cause excessive bleeding as inadequate amounts of thrombin is released (Gargin, 2012).

FVIII, bound to vWF, circulates as heterodimers in plasma (Myles *et al.,* 2002). The serine protease known as thrombin is mainly responsible for the proteolytic activation of FVIII to produce a heterotrimer comprising of three subunits, A1, A2 and A3-C1-C2, respectively (Callaghan & Kaufman, 2008; Shen *et al.,* 2008). FVIIIa can then associate to the phospholipid membrane and act as a cofactor to FIX (Myles *et al.,* 2002). The combination of two mechanisms is responsible for the inactivation of FVIIIa. Firstly, as the A2 subunit of the heterotrimer dissociates, FVIIIa is inactivated, and secondly, activated protein C (APC) cleaves the FVIIIa molecule in the A1 and A2 subunit, promoting its inactivation (Gale *et al.,* 2008).

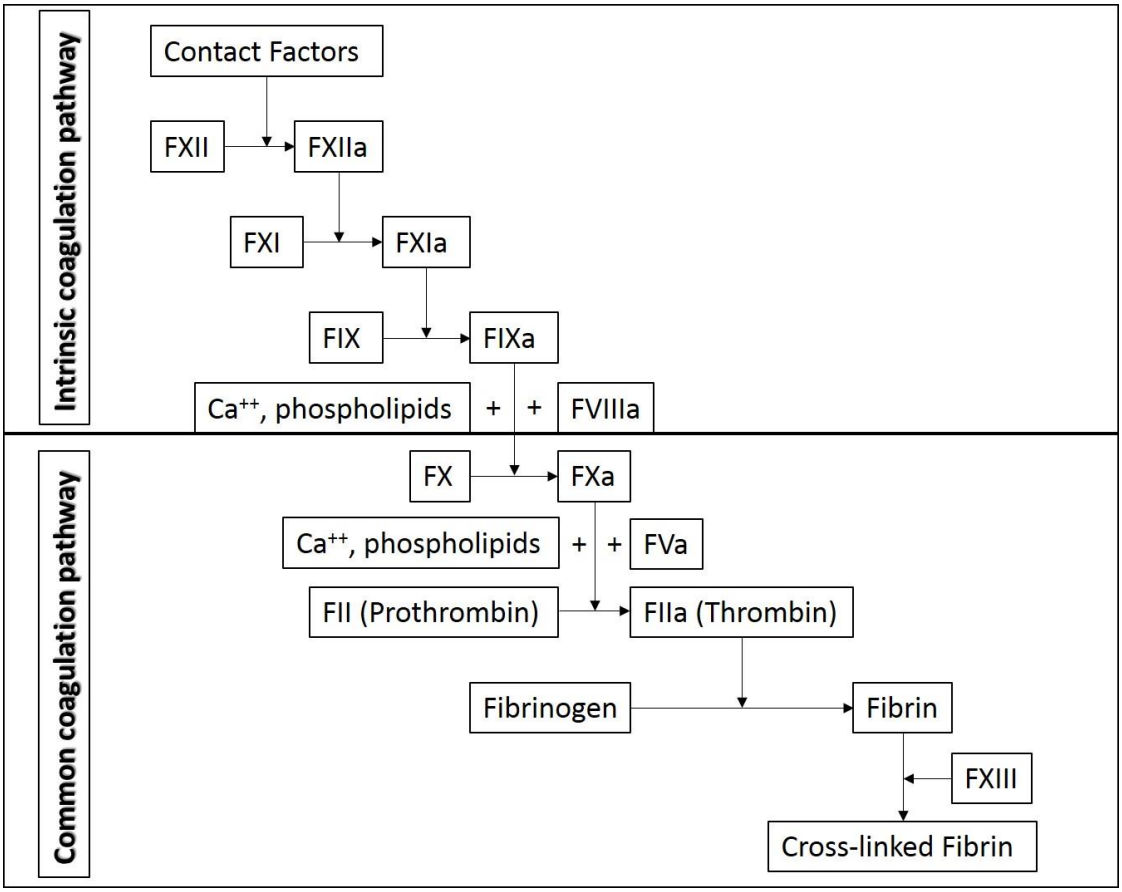


Figure 2.4: Intrinsic and common pathways of coagulation (Copied from Kloppers & Janse van Rensburg, 2015).

2.7 Diagnosis of haemophilia A

The diagnostic approach to haemophilia A consists of the inquisition of a bleeding history, a physical examination, a screening test for bleeding tendency and a confirmatory test (Mahlangu *et al.,* 2008). It is essential that the tests used are accurate to reflect the actual condition of the patient (Peetz, 2007). The diagnostic approach is crucial as it will stipulate the course of therapeutic intervention (Mahlangu *et al.,* 2008). Although laboratory diagnosis and the classification of haemophilia A are important, the causative mutation should be investigated. The direct mutation approaches are very accurate (100%) and allow families to make informed decisions regarding family planning. Furthermore, the direct mutation approaches are used to identify carriers of the disorder, which proves to be useful in genetic counselling (Peyvandi *et al.,* 2006).

The detection of mutations associated with haemophilia A is essential, as laboratory coagulation studies are not sufficient for carrier detection and prenatal diagnosis (Kessler *et al.,* 2014). The confirmation of carrier status allows the affected individual to make an informed decision. As these individuals know their haemophilia A status, it allows them to act with precaution with regards to activities that can lead to severe bleeding and allows physicians to provide them with the necessary care (Habart, 2005; Kessler *et al.,* 2014). Furthermore, prenatal diagnosis of haemophilia A allows the parents of the unborn child to make an informed decision with regards to the future of their child. They can be informed on what treatment options are available, and they can be educated on the life expectancies associated with haemophilia A (Darby *et al.,* 2007; Kessler *et al.,* 2014). Additionally, as stated earlier, it is well known that specific FVIII mutations contribute to the development of FVIII inhibitors and therefore, the detection of these mutations are essential in managing haemophilia A treatment (Schwaab *et al.,* 1995; Eckhardt *et al.,* 2013).

2.7.1 Coagulation studies

The performance of laboratory coagulation studies is an essential diagnostic tool for haemophilia A (Mahlangu *et al.,* 2008). Firstly, a screening test is conducted to assess bleeding tendency, and then a confirmatory test is performed to quantify the concentration of a specific coagulation factor. As haemophilia A is caused by the deficiency of FVIII, a FVIII assay is used to quantify the amount of functional FVIII present in the patient's plasma (Hoffbrand & Moss, 2011b).

The screening test used to assess bleeding tendency in haemophilia A patients is the Activated Partial Thromboplastin Time (aPTT). The aPTT measures the intrinsic pathway of coagulation in which FVIII functions as a cofactor. The aPTT assay aims to measure the time it takes for plasma to clot and in the process indicate whether an intrinsic-tenase complex component is deficient or in excess (Hoffbrand & Moss, 2011a; Hoffbrand & Moss, 2011b). An abnormally prolonged aPTT (reference times differ among laboratories) can be indicative of a clotting factor deficiency or other causes such as von Willebrand disease, liver disease, FVIII inhibitors, heparin administration or lupus anticoagulant to mention a few (Kamal *et al.,* 2007), and therefore, needs to be confirmed with a factor assay (Bashawri & Ahmed, 2007). The

FVIII assay is used as a confirmatory test for FVIII deficiencies. Importantly, the FVIII assay needs to be accurate, sensitive and specific to reflect the actual condition of the patient in question (Peetz, 2007). Although coagulation studies play an integral role in diagnosing haemophilia A, they cannot accurately identify carriers of haemophilia A and are unable to identify FVIII mutations (Olsson *et al.,* 2014). Furthermore, it has been suggested that different gene defects might explain the different phenotypes observed in haemophilia A, and these phenotypes are caused by factors such as gene defects and are independent of the FVIII levels (Van den Berg *et al.,* 2007). Consequently, coagulation studies are essential in the management of haemophilia A, but the detection of mutations are just as important in cases where the specific phenotype is independent of the FVIII levels.

2.7.2 FVIII mutation detection

The rapid detection of the causative mutation is central to managing haemophilia A. Not only does it impact on directing targeted therapy, but also plays a role in tailoring a comprehensive treatment plan. Likewise, mutational analysis affects pharmacogenomics and highlights the importance of causative mutation analysis, and its benefit to the patient's treatment plan (Beastall & Evans, 2012). Molecular diagnostics is complicated by the *FVIII* gene size and the large number of causative mutations (Factor VIII Gene Variant Database, 2017). The two different approaches to genetic diagnosis involve indirect mutation detection (linkage analysis) or direct mutation detection (Peyvandi *et al.,* 2006). The indirect approach to mutation detection involves the analysis of single nucleotide polymorphisms (SNP) or that of microsatellite variable number tandem repeat (VNTR) markers in the *FVIII* gene (Peake *et al.,* 1993). This approach is based on tracking the defective X-chromosome in the family and has a 99% reliability rate, as genetic recombination cannot be excluded. The direct mutation detection method is almost 100% accurate and also the method of choice (Peyvandi *et al.,* 2006). It is recommended that the presence of inv22 and intron 1 (int1) inversions should be eliminated before conducting any other test, especially in patients with the severe disorder. After that, the screening of the other 26 exons for other mutations should be done, and if an abnormal segment is detected, direct sequencing can be applied (Habart, 2005). The molecular analysis of haemophilia A

has led to the development of comprehensive prenatal and carrier detection methods, and the continuous development of these methods allows for more rapid and cost-effective analysis of the genetics associated with haemophilia A (Keeney *et al.,* 2005).

2.8 Mutations associated with haemophilia A

When the size of the *FVIII* gene is considered, it is explanatory why mutational heterogeneity is associated with haemophilia A. There is a reduction in the genetic fitness of males with haemophilia A, and thus, mutations originate independently in families. The gain of new mutations must, therefore, be considered (Tuddenham *et al.,* 1994; Naylor *et al.,* 1995) Furthermore, the possibility of mutations originating during early embryogenesis should not be excluded. Somatic mosaicism is a fairly common event in haemophilia A (Leuer *et al.,* 2001). The prevalence of mutations associated with haemophilia A are listed in **Table 2.1** below. The most common mutations associated with severe haemophilia A are inv22 and the intron 1 inversion (Lackich *et al.,* 1993; Bagnall *et al.,* 2002; Kaufman *et al.,* 2013). A further 2015 unique variants associated with haemophilia A have been reported (Factor VIII Gene Variant Database, 2017).

Table 2.1: Prevalence of mutations associated with haemophilia A

Type of mutation	Prevalence of mutation (%)	References
Intron 22 inversion	45%	(Lakich *et al.,* 1993) (Naylor *et al.,* 1995) (Kaufman *et al.,* 2013) (Johnsen *et al.,* 2017)
Intron 1 inversion	5%	(Brinke *et al.,* 1996) (Bagnall *et al.,* 2002)
Point mutations	66.6%	(Factor VIII Gene Variant Database, 2017)

Deletions	23.2%	(Factor VIII Gene Variant Database, 2017)
Polymorphisms	2.3%	(Factor VIII Gene Variant Database, 2017)
Insertions (Includes 34 mutations)	1.7%	(Factor VIII Gene Variant Database, 2017)
Insertions and Deletions	1.4%	(Factor VIII Gene Variant Database, 2017)

2.8.1 Intron 1 inversion associated with severe haemophilia A

An inversion in int1 of the *FVIII* gene has been reported to be associated with severe haemophilia A. Homologous recombination between a sequence in intron 1 (int1-1) of the *FVIII* gene and a homologous copy (int1-2) also located on the X-chromosome, results in the inversion. It was first found in monozygotic twins with haemophilia A (Brinke *et al.,* 1996). The inversion in int1 results in two hybrid-transcripts being formed. The one transcript is known to contain a promoter and part of exon 1 of the *FVIII* gene, followed by alternative exons and exons of the *VBP1* gene *(*Brinke *et al.,* 1996; Brinke *et al.,* 1997). The other transcript contains all but the penultimate exon of the *C6.1A* gene, followed by alternative exons and finally exons 2 to 26 of the *FVIII* gene (Brinke *et al.,* 1996). The int1 inversion occurs in approximately 5% of individuals with severe haemophilia A (Bagnall *et al.,* 2002; Tizzano *et al.,* 2003; Schroder *et al.,* 2006)

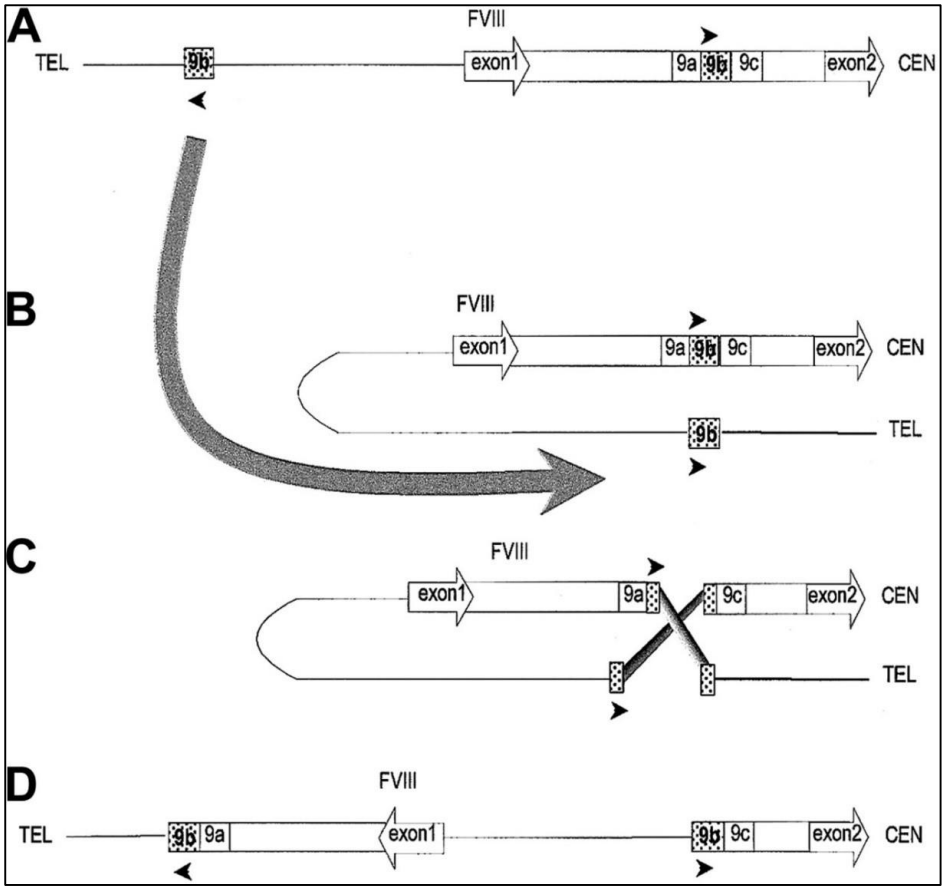


Figure 2.5: The proposed model of the int1 inversion associated with severe haemophilia A (Copied from Bagnall *et al.,* 2002). (A): Indicates the *FVIII* intron 1, with a homologous sequence within intron 1, termed 9b. The 9b sequence is repeated outside the *FVIII* gene. (B, C): Indicates the homologous recombination that occurs between the two 9b sequences, which is ultimately responsible for the inversion occurring. (D): The result of the intron 1 inversion, causing a change in the *FVIII* gene sequence.

2.8.2 Inv22 associated with severe haemophilia A

Higuchi and colleagues (1991), found that half of the severe haemophilia A patients had no mutation in the coding or splice regions of the *FVIII* gene, which suggested that the mutation causing severe haemophilia A could be in DNA sequences that originate outside the *FVIII* gene (Higuchi *et al.,* 1991). It led to the finding that these patients with severe haemophilia A had an inversion disrupting the *FVIII* gene. The inversion was located in int22 of the *FVIII* gene and involves DNA sequences outside

the *FVIII* gene. The inv22 accounts for approximately 45% of all severe haemophilia A cases (Lakich *et al.,* 1993; Kaufman *et al.,* 2013), as well as for the abnormal mRNA found in these patients (Naylor *et al.,* 1993a). The inversion almost exclusively originates in male germ cells, suggesting that in female meiosis the paring of the X chromosomes have an inhibitory effect on the inversion occurring (Rossiter *et al.,* 1994).

As previously described, int22 contains a region, namely the *F8A* gene, which is involved in the inv22. This *F8A* region, known as the int22 homologous region (int22h-1), is 9.5 kb in size and has two copies (int22h-2 and int22h-3) near the telomere of the X chromosome (Xq) (Naylor *et al.,* 1995). The three homologous regions were confirmed by probes for the DXSII5 locus, one region (int22h-1) in int22 and the other copies, int22h-2 (proximal copy) and inth22-3 (distal copy), upstream of the 5' end of the *FVIII* gene (Freije & Schlessinger, 1992). All three respective copies contain a CpG island from which *F8A* can be transcribed (Levinson *et al.,* 1990).

The occurrence of the inv22 involves homologous recombination between int22h-1 and one of the respective copies, either int22h-2 or int22h-3 (Lakich *et al.,* 1993). The similarity of the three copies was investigated, and only three consistent nucleotide base pair differences were found between int22h-1, int22h-2 and int22h-3 (Naylor *et al.,* 1995). Two different inv22 events can occur where a type 1 inv22 involves the cross-over event between int22h-1 and int22h-3, and a type 2 inv22 involves the cross-over event between int22h-1 and int22h-2, respectively. The third type of inv22, type 3A and 3B, have been identified, this type of inversion occurs when an individual has more than two extragenic copies of the int22h region (Rossiter *et al.,* 1994). The type 1 inv22 has a higher occurrence than type 2 (Jenkins *et al.,* 1994; Antonarakis *et al.,* 1995). It was confirmed by Antonarakis and colleagues (1995) who reported that in a total of 2,093 patients with severe haemophilia A, 35% had the type 1 inv22 and 7% had the type 2 inv22, respectively (Antonarakis *et al.,* 1995). No substantial clinical difference has been found between the type 1 and type 2 inv22.

It was originally thought that int22h-2 and int22h-3 copies were similarly orientated and in opposite orientation to int22h-1 (Lakich *et al.,* 1993). However, upon determination of the DNA sequence of the X chromosome, it was reported that int22h-2 and int22h-3 were in fact inversely orientated and int22h-1 and int22h-2 were

similarly orientated (Ross *et al.,* 2005). Considering this orientation, recombination between in22h-1 and int22h-2 would lead to the deletion or duplication of the 400 kb sequence that separates the respective copies from each other (Bagnall *et al.,* 2006). However, Ross *et al.* (2005) reported that int22h-2 and int22h-3 are part of an imperfect palindrome. The arms of the palindrome contain int22h-2 and int22h-3, respectively. The int22h copies, thus, undergo intra-chromosomal recombination that causes int22h-2 and int22h-3 to swap around (**Figure 2.6**)**.** The recombination event allows the orientation of int22h-2 and int22h-3 to change relative to int22h-1, leading to a type 2 inv22 occurring. As int22h-2 and int22h-3 form part of an imperfect palindrome that allows the intra-chromosomal recombination event to occur, the 400 kb deletion or duplication, does not transpire (Bagnall *et al.,* 2006). It may also explain why the type 1 inv22 is more common in severe haemophilia A patients compared to the type 2 inv22 (Jenkins *et al.,* 1994; Antonarakis *et al.,* 1995). The misalignment of int22h repeats can lead to inter-chromosomal recombination between the repeats, resulting in a reciprocal duplication and deletion. A deletion can also occur as a result of intra-chromosomal/chromatid homologous recombination. Int22 deletions and duplications are also known FVIII mutations (Bagnall *et al.,* 2006).

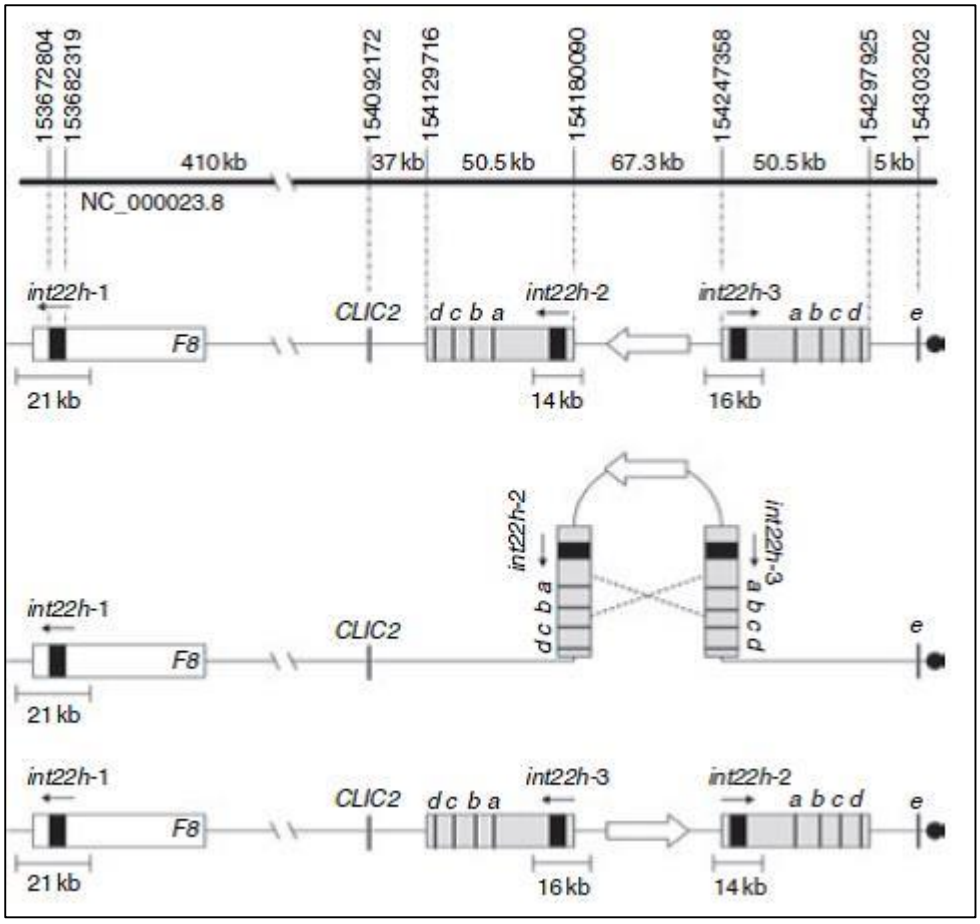


Figure 2.6: The intra-chromosomal recombination event that occurs between int22h-2 and int22h-3 that results in the inv22 (Copied from Bagnall *et al.,* 2006).

2.8.2.1 The effect of the inv22 on the *FVIII* gene

The inv22 causes the *FVIII* gene to be divided into two parts. The first part contains exons 1 to 22 and the second part contains exons 23 to 26. The two-parts are however separated from each other and in opposite orientation (Lakich *et al.,* 1993). It was reported that the amplification between the boundaries of FVIII exons 22 and 23 in the mutant phenotype (inv22) yields no PCR product, as the exons are separated from each other (Naylor *et al.,* 1993b). However, the study by Pandey *et al.* (2013) reported that the inv22 transcript contained an additional exon, termed exon 23c, which is dissimilar from the FVIII wild-type exon 23. Sequence analysis revealed that exon 23c expressed 16 amino acids that were adjacent to the amino acids expressed for by the wild-type exon 22 of the *FVIII* gene. It explains why PCR amplification is unsuccessful

in this region considering that the FVIII wild-type exon 23 has been replaced by the alternative exon 23c (**Figure 2.7**) (Pandey *et al.,* 2013). Haemophilia A patients with the inv22 collectively produce two FVIII exon-containing mRNAs. The two FVIII exon-containing mRNA's (FVIIII22I and FVIIIB) express the full-length FVIII amino acid sequence but are non-secretory polypeptide chains which consequently lead to the deficiency of FVIII (Pandey *et al.,* 2013).

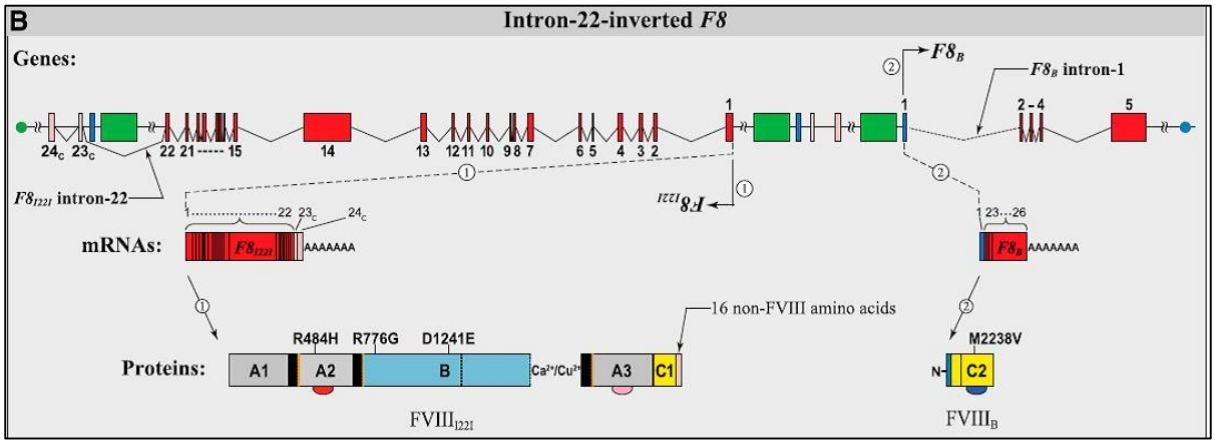


Figure 2.7: The *FVIII* gene, mRNA and protein of the inv22 genotype (Copied from Sauna *et al.,* 2015).

2.9. The detection of the inv22 mutation

Currently, three main methods are used to detect the inv22 associated with severe haemophilia A. These three methods are Southern Blotting, Long-distance PCR (LD-PCR) and Inverse-PCR (I-PCR) (Lakich *et al.,* 1993; Liu & Sommer, 1998; Rossetti *et al.,* 2005). Each of these methods has different advantages and disadvantages which need to be carefully considered when using them to detect the inv22. Additional PCR based methods have been reported to detect the inv22, however, none without notable disadvantages.

2.9.1 The inv22 Southern Blot method

The inv22 Southern Blot method utilises restriction enzymes to digest DNA at specific sequences (Southern, 1975). In **Figure 2.8**, an example of inv22 results as detected by the Southern Blot method is displayed. The inv22 Southern Blot method is based on DNA digestion with endonuclease restriction enzymes specific to the inv22 target site. The inv22 restriction products are then visualised on an agarose electrophoresis gel. The denatured inv22 DNA products are consequently blotted onto a nylon membrane and preserved. The inv22 DNA fragments on the nylon membrane are hybridised to a labelled oligonucleotide probe that is complementary to the inv22 DNA fragments. After the excess probe is washed off, the hybridised probe is visualised with autoradiography (Levinson *et al.,* 1990; Lakich *et al.,* 1993). The inv22 Southern Blot method is reported to be accurate (Lakich *et al.,* 1993), but conversely time-consuming and labour-intensive (Liu & Sommer, 1998; Rossetti *et al.,* 2005; Kumar *et al.,* 2013). In general, the Southern Blot method requires a large quantity of DNA for analysis (He *et al.,* 2002) and tends to have variable sensitivity. These limitations can lead to the misinterpretation of results (Grasso *et al.,* 2014).

2.9.2 The inv22 Long Distance Polymerase Chain Reaction (LD-PCR)

LD-PCR is a method used to amplify long DNA fragments of up to 25 kb (Ohler & Rose, 1992). Subsequently, an inv22 LD-PCR method has been developed as the DNA sequences in the *FVIII* gene that are associated with the inv22 are large (12 kb). The method utilises four primers to amplify the regions in the *FVIII* gene that allows for the differentiation between the different genotypes, namely the FVIII wild-type, inv22 and inv22 carrier. Two of the primers are designed to flank int22h-1 which is located in the *FVIII* gene. The other two primers are designed to flank the int22h-2 and int22h-3 sequences, respectively (Liu *et al.,* 1998). For the wild-type genotype, two fragments will be produced which are 12 kb and 10 kb, respectively. The inv22 genotype will produce two fragments of 11 kb and one fragment of 10 kb. Finally, carriers of the inv22 will produce three fragments, namely 12 kb, 11 kb and 10 kb (**Figure 2.9**) (Liu *et al.,* 1998). The method was modified by the introduction of a sub-cycling PCR (S-PCR) step, which allowed the efficient amplification of all four PCR segments, which was previously inconsistent (Liu & Sommer, 1998). The inv22 LD-

PCR method was then further modified to develop a discriminatory test that allows for the detection of the type 1 and type 2 inv22, respectively (Bagnall *et al.,* 2006). It has been reported that the inv22 LD-PCR method is difficult to standardise and variations occur in laboratory results (Kumar *et al.,* 2013). Furthermore, the quantity and quality of DNA may influence the product yield of the LD-PCR method in general (Liu & Sommer, 1998). Likewise, DNA of good quality and quantity is necessary to obtain optimal results with the inv22 LD-PCR method (Polakova *et al.,* 2003). Consequently, if poor quality and quantity of DNA were originally obtained, DNA will have to be collected from the patient again, which may be inconvenient.

2.9.3 The inv22 Inverse Polymerase Chain Reaction (I-PCR)

The inv22 I-PCR method allows amplification of unknown DNA sequences. To be able to detect unknown DNA sequences, the original DNA sequences need to be rearranged as primers can't amplify unknown sequences. It is accomplished by the utilisation of restriction enzymes to produce DNA fragments for which primers can be designed to allow amplification (Triglia *et al.,* 1988). The inv22 I-PCR method developed by Rossetti *et al.* (2005) can distinguish between the FVIII wild-type and the inv22 genotype and employs three primers to amplify the DNA sequences associated with the genotypes. The PCR products, visualised by gel electrophoresis, differ in size for the respective genotypes allowing them to be distinguished from each other (**Figure 2.10**). Inverse shifting PCR (IS-PCR) is a modified inv22 I-PCR method, with the ability to detect the FVIII wild-type and inv22 genotypes, as well as a duplication of int22 (Dup22) and types 1 and type 2 int22 deletions (Rossetti *et al.,* 2008). The inv22 IS-PCR method consists of two methods, firstly the inv22 diagnostic test for the detection of deleterious mutations and non-deleterious int22 mutations. The second method is an inv22 complementary test that can discriminate the inv22 from the int22 deletion, as well as the int22 duplication from the FVIII wild-type (Rossetti *et al.,* 2008).

The inv22 I-PCR method has several characteristics which makes it a suitable method of choice. These include reproducible results; the ability to detect two genotypes and with further adaption quantification through fluorescent chemistries may be possible (Rossetti *et al.,* 2005). Apart from the modified inv22 IS-PCR method, an inv22 I-PCR

method has been designed with an added capillary gel electrophoresis step. This modified inv22 I-PCR method also can detect a type 1 and 2 inv22, respectively (Pan *et al.,* 2014). However, the use of restriction enzyme (B*cl* I) digestion of specific nucleotide sequences, may interfere with the detection of other point mutations in the *FVIII* gene that causes haemophilia A (Pan *et al.,* 2014).The inv22 I-PCR method is also reported to be time-consuming and contains multiple steps, such as DNA digestion, several DNA purification steps, DNA-ligation, PCR and post-PCR analysis (Kumar *et al.,* 2013). The method lacks 'built in' control measures between its different steps which makes troubleshooting and quality control a challenge. Consequently, the inv22 I-PCR method is difficult to standardise in sub-optimally resourced laboratories (Kloppers & Janse van Rensburg, 2017).

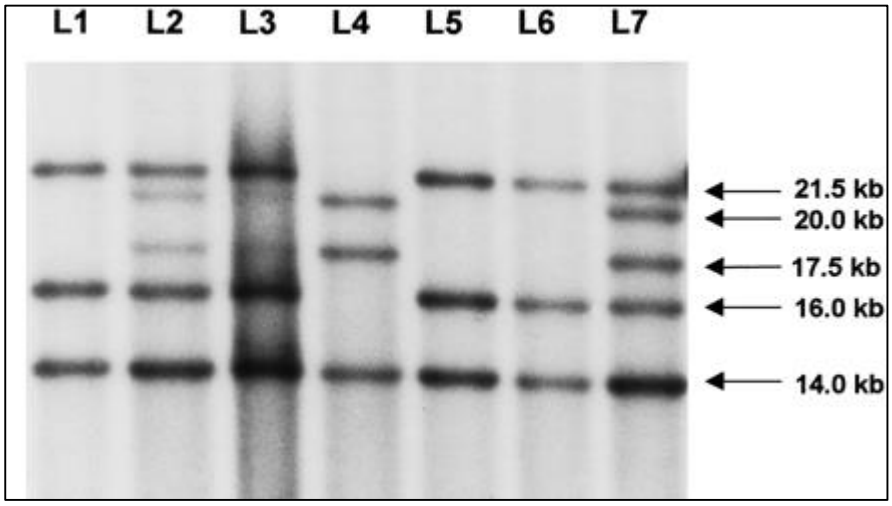


Figure 2.8: Gel electrophoresis image illustrating an example of an inv22 Southern Blot's result. Lanes (L1, L5 and L6): Inv22 negative (21.5 kb, 16 kb and 14 kb). Lanes (L2 and L3): Type 1 inv22 carrier (Somatic Mosaicism) (21.5 kb, 20 kb, 17.5 kb, 16 kb and 14 kb). Lane (L4): Type 1 inv22 (20 kb, 17.5 kb and 14 kb). Lane (L7): Type 1 inv22 carrier (21.5 kb, 20 kb, 17.5 kb, 16 kb and 14 kb) (Copied from Oldenburg *et al.,* 2000).

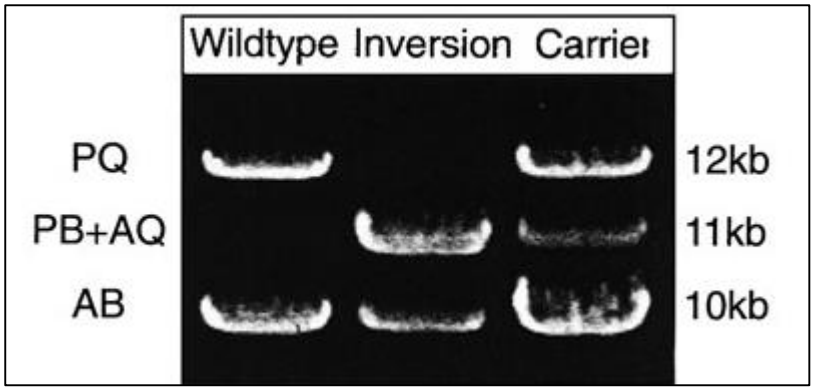


Figure 2.9: Gel electrophoresis image illustrating an example of an inv22 LD-PCR's result. The gel-image indicates a wild-type, inv22 positive patient and a carrier of the inv22 (Copied from Liu & Sommer, 1998).

Figure 2.10: Gel electrophoresis image illustrating an example of an inv22 I-PCR's result. Lane (1): Inv22 negative. Lane (2): Type 1 inv22. Lane (3): Type 1 inv22 carrier. Lane (4): Type 2 inv22. (5): Type 2 inv22 carrier (Copied from Rossetti *et al.*, 2008).

2.9.4 Other PCR-based methods that are used for inv22 detection

Recently, novel approach methods have been developed to detect the inv22. The first of those is a reverse transcriptase real-time PCR method developed by Kumar *et al.* (2013). The method is based on the amplification of four regions in the *FVIII* gene, namely exons 21 to 22 (333 bp fragment), exons 22 to 23 (480 bp fragment) and exons 23 to 24 (393 bp fragment). In a case where the inv22 is present, the method relies on the absence of the 480 bp fragment to diagnose a patient positively with the inv22 (Kumar *et al.,* 2013). This inv22 detection method thus relies on the absence of a PCR-fragment to confirm the inv22 genotype. It raises the question whether this could lead to false-positive inv22 results if the genotype is not confirmed by sequence analysis in each case.

Secondly, Ding *et al.* (2016) modified the original inv22 LD-PCR method. The new method which has two multiplex pre-amplification of LD-PCR (PLP) consists of two combinations of five primers that are used to amplify the wild-type and chimeric int22h alleles, respectively. The method also consists of a carrier mosaicism method which is similar to the multiplex pre-amplification of LD-PCR but has a different combination of primers to amplify only the chimeric int22h alleles. The PLP's are then quantified using AccuCopy (Ding *et al.,* 2016). As this is still an LD-PCR method, the disadvantages of the method, such as the difficulty to standardise in different laboratories (Kumar *et al.,* 2013) is problematic. The yield of poor results due to poor quality or quantity of DNA remains a disadvantage (Polakova *et al.,* 2003). Dutta and colleagues (2016) developed a nested RT-PCR method that can detect the inv22. This method, however, relies on multiple primers and various PCR fragments and consequently will be more challenging to optimise and implement in a routine diagnostic laboratory (Dutta *et al.,* 2016).

Although these PCR based inv22 detection methods have been developed, the three main methods, namely the inv22 Southern Blot, LD-PCR and I-PCR are still the methods of choice. It could be because the PCR based inv22 methods are difficult to standardise in laboratories, quality control procedures or questionability regarding accuracy to detect the inv22.

2.10 Real-time Polymerase Chain Reaction

Real-time PCR also referred to as quantitative PCR (qPCR) allows DNA to be detected and quantified in a single reaction using various fluorescent detection chemistries. Real-time PCR equipment can detect a fluorescent signal emitted when hybridisation occurs between the detection chemistry and the DNA target. The method detects the PCR product during the amplification process, specifically in the exponential phase of PCR when the reaction starts to generate copies of the DNA target. The fluorescent intensity can be correlated to the DNA concentration, allowing quantification of the DNA start copy number. The specific cycle number at which the fluorescent intensity surpasses the background fluorescent emission is known as the Ct-value. The Ct-value is measured in the exponential phase of the PCR reaction and is used to determine the input DNA concentration as the latter is inversely proportional to the Ct-value (Heid, *et al,* 1996; Bustin, 2005).

Real-time reverse transcriptase PCR (Real-time RT-PCR) is a molecular technique that is used when information about RNA is required. RNA, can however not be used as starting material in PCR and therefore, has to be converted to what is known as complementary DNA (cDNA). It is made possible by the enzyme reverse transcriptase, which is an RNA-dependent polymerase that catalyses DNA synthesis using RNA as starting material. The end product is cDNA which can be used as a template for PCR amplification (Farkas & Holland, 2009).

There are various detection chemistries available for the detection of DNA/cDNA in real-time PCR. These chemistries include nonspecific chemistries such as DNA binding/intercalating dyes and target-specific probes labelled with a unique fluorescent labelling molecule. Examples of these include hybridization probes, hydrolysis probes and hairpin probes (Bustin, 2005; Wong & Medrano, 2005). The DNA detection chemistry that is most frequently used is in real-time PCR is SYBR Green, an intercalating dye that is inserted into double-stranded DNA molecules. Advocates of SYBR Green argue that it is more cost-effective than probe-based chemistries and SYBR Green can be used to differentiate DNA by using melt curve analysis (Monis *et al.,* 2005). In this process, double-stranded DNA is melted or denatured, into single-stranded DNA with an increase in temperature, the fluorescent data is collected at each temperature point. As the double-stranded DNA denatures into single-stranded

molecules, the SYBR Green is released, and a dissociation curve can be constructed. Based on the length and unique DNA sequence different amplicons will have different melting temperatures (Tm) (Bustin, 2005).

The introduction of real-time PCR has overcome many of the limitations of conventional PCR and other molecular techniques such as blot analysis. In summary, the advantages of the real-time PCR method includes: (1) Elimination of post-PCR analysis as the real-time PCR method can accurately detect and quantify DNA in a homogeneous reaction without the need for any form of post PCR electrophoresis (Bustin, 2005). (2) The real-time PCR method reduces post-PCR contamination (Heid *et al.,* 1996) and (3) has high throughput ability that includes reproducibility with quality control (Spurgeon *et al*., 2008). (4) The real-time RT-PCR method can accurately quantify subtle changes in gene expression in a mixture of total cellular RNA (Kang *et al.,* 1997; Valasek & Repa, 2005). It can also detect transcripts that are very low in abundance (Yang *et al.,* 2005; Qiu *et al.,* 2013).

Several guidelines need to be taken into consideration in the application of real-time PCR as detection and quantification method. The Minimum Information for Publication of Quantitative Real-Time PCR experiments (MIQE) guidelines outline the important issues that need to be considered (Bustin *et al.,* 2009). Importantly, the method needs to be designed with great caution, and appropriate controls need to be in place for accuracy and reliability (Wong & Medrano, 2005). The application of real-time PCR has been proven to benefit a wide variety of disciplines ranging from virology, microbiology and haematology. It has multiplex capabilities that can discriminate between several genotypes in a single reaction (Bustin, 2005). Real-time PCR can be used for the detection of minimal residual disease, detection of DNA copy number, expression analysis, allelic discrimination, confirmation of microarray data and gene expression studies (Ginzinger, 2002). Consequently, real-time PCR is a molecular technique with several advantages that could be well utilised to advance inv22 detection methods.

2.11 Conclusion

Haemophilia A is one of the most common hereditary bleeding disorders and has a tremendous impact on the lives of affected individuals and their caregivers (Miners *et al.,* 1999; Cassis *et al.,* 2011). In under-resourced countries, haemophilia A is underdiagnosed and insufficient genetic research is being conducted (Peyvandi *et al*., 2006). In order to improve care and education to haemophilia A patients and carriers, genetic research needs to be promoted to ultimately detect the causative mutations of haemophilia A (Kessler *et al.,* 2014). In many cases, carriers are unaware of their haemophilia A status which can be because bleeding episodes have not occurred previously (asymptomatic). Therefore, it is essential that these carriers are identified to allow them to make an informed decision through genetic counselling (Winikoff & Lee, 2010).

The specific mutations which have been associated with FVIII inhibitor development make FVIII mutation detection an essential part of haemophilia A management and care (Jayandharan *et al.,* 2012; Eckhardt *et al.,* 2013). The most common causative mutation of severe haemophilia A is the inv22, which is associated with 45% of all severe cases (Lakich *et al.,* 1993; Kaufman *et al.,* 2013). The current utilised inv22 detection methods namely, Southern Blot, LD-PCR and I-PCR all have prominent disadvantages as discussed in the literature review. Likewise, PCR-based inv22 methods also have several disadvantages that could hamper their introduction as main inv22 detection methods. It presents a need for the development of an inv22 detection method, that is rapid, accurate and cost-effective to overcome the disadvantages associated with the current methods.

Chapter 3: Methodology

3. Introduction

This chapter includes the rationale, aim, objectives, study design, study population and the ethics considerations that were adhered to. The study procedure is outlined, and the methods that were used in the study are described.

3.1 Rationale, aim and objectives

3.1.1 Rationale

The inv22 is reportedly associated with 45% of all severe haemophilia A cases. However, it is currently underdiagnosed in developing countries such as South Africa. It is mainly due to the lack of genetic research, genetic testing facilities and overall limited health care resources in the country. The current inv22 detection methods all have eminent disadvantages and could have a negative impact on the wide-spread screening of FVIII mutations. Since, FVIII mutations, such as the inv22, play a role in haemophilia A treatment, methods to detect these mutations, as well as genetic counselling, should be widely available and easily implementable in sub-optimally resourced laboratories.

3.1.2 Aim

The aim of this study was to develop an inv22 real-time reverse transcriptase polymerase chain reaction (RT-PCR) detection method to overcome the disadvantages of the current detection methods.

3.1.3 Objectives

Objective 1:

The first objective is to evaluate the inv22 using the Inverse PCR (I-PCR) method in the setting of a South African service delivery laboratory, as it is currently one of the main methods used to detect the inv22 associated with severe haemophilia A patients.

Objective 2:

Objective two is to develop an inv22 conventional gel electrophoresis RT-PCR detection method that will allow for sequence analysis to be performed and to confirm the presence or absence of the inv22 in each sample.

Objective 3:

The third objective is to develop an inv22 with a real-time RT-PCR detection method that is repeatable and cost-effective.

Objective 4:

The fourth objective is to draw a comparison between the detection methods developed in the study namely; the conventional gel electrophoresis RT-PCR detection method, the real-time RT-PCR detection method and Sanger sequence analysis.

Objective 5:

The last objective is to determine the presence of the inv22 in the central South African haemophilia A population.

3.2 Study design

A case-control analytical study was conducted. The study was performed in the Department of Haematology and Cell Biology at the Faculty of Health Sciences, University of the Free State, Bloemfontein, South Africa.

3.3 Study population

Sixty-three (63) individuals were recruited to participate in this study. These 63 individuals consisted of three study controls with known inv22 genotypes and 60 study participants with unknown inv22 genotypes, respectively.

The study controls consisted of one healthy non-haemophilic control (C1) and two unrelated haemophilia A patients (C2 and C3). Control C1 was recruited voluntarily from the Department of Haematology and Cell Biology in the Faculty of Health Sciences at the University of the Free State, Bloemfontein. Control C2 and C3 were haemophilia A patients that were treated at the Haematology clinic at Universitas Academic Hospital in Bloemfontein. This two haemophilia A patients were referred for inv22 screening at the National Health Laboratory Service, University of the Witwatersrand, School of Pathology, Division of Human Genetics in Johannesburg, South Africa, where the inv22 Southern Blot detection method was used to confirm inv22 status. Both control C2 and C3 were found to be positive for the inv22 (**Appendix A**).

The 60 study participants consisted of 35 haemophilia A patients, 18 potential haemophilia A carriers and seven non-haemophilic volunteers. The 35 haemophilia A patients and 18 potential haemophilia A carriers were voluntarily recruited from the Haematology clinics at the Universitas Academic Hospital in Bloemfontein and the Kimberley Hospital Complex in Kimberley, South Africa. The seven non-haemophilic volunteers were recruited from the Department of Haematology and Cell Biology in the Faculty of Health Sciences at the University of the Free State, Bloemfontein, South Africa.

3.3.1 Inclusion criteria

- The study included adult and minor participants from the ages of six years to 65 years.
- Haemophilia A patients were included if they have been diagnosed with haemophilia A according to the guidelines by the International Society on Thrombosis and Haemostasis (ISTH) (White *et al.,* 2001).
- Participants that voluntarily gave written informed consent.

- Minors of whom parents/legal guardians gave written informed consent
- Minors that voluntarily could sign an assent form.

3.3.2 Exclusion criteria

Participants were excluded from the study for the following reasons:

- The participants had been diagnosed with haemophilia B.
- Minors younger than six years old that could not sign an assent form were excluded.
- Minors of whom parents/legal guardians refused to give informed consent for the minor/s.
- Participants were pregnant.
- Participants were unable to give informed consent.

3.3.3 Ethics consideration

This study was approved by the Health Sciences Research Ethics Committee (**ECUFS 166/2015**) (**Appendix B**) in the Faculty of Health Sciences at the University of the Free State, Bloemfontein, South Africa. Approval was also obtained from the Free State Department of Health and the Northern Cape Department of Health in South Africa. Each participant willingly signed an informed consent form (**Appendix C**). The parents or legal guardians willingly signed an inform consent form in the case of the participants being a minor (**Appendix D**). Each minor willingly signed an assent form to participate in the study (**Appendix E**). The consent forms were available in English, Afrikaans or Southern Sotho.

3.4 Study procedure

Figure 3.1 is a schematic representation of the study procedure followed in this study. Firstly, ethics approval was obtained, and the study was approved by an evaluation committee in the Department of Haematology and Cell Biology, University of the Free State, Bloemfontein, South Africa.

The study controls C1, C2 and C3 were first recruited. These controls were then subsequently used to evaluate the inv22 I-PCR method and to develop the inv22 conventional gel electrophoresis and real-time RT-PCR detection methods for haemophilia A patients and potential carriers. After that, the study controls were Sanger sequenced to confirm the presence or the absence of the inv22.

The 60 study participants were recruited for the study according to inclusion and exclusion criteria. The study participants were screened with the inv22 conventional gel electrophoresis-based RT-PCR detection method, and with the newly developed inv22 real-time RT-PCR detection method with C1, C2 and C3 serving as controls. The results obtained with the inv22 conventional gel electrophoresis-based RT-PCR and real-time RT-PCR detection methods were confirmed by Sanger Sequencing.

The results obtained for the 60 study participants with the inv22 conventional gel electrophoresis-based RT-PCR detection method, inv22 real-time RT-PCR method and Sanger sequencing were tabulated and compared to each other.

The presence of the inv22 was evaluated in the central South African haemophilia A population, and it was determined which of the study participants with the inv22 had a history of FVIII inhibitor development.

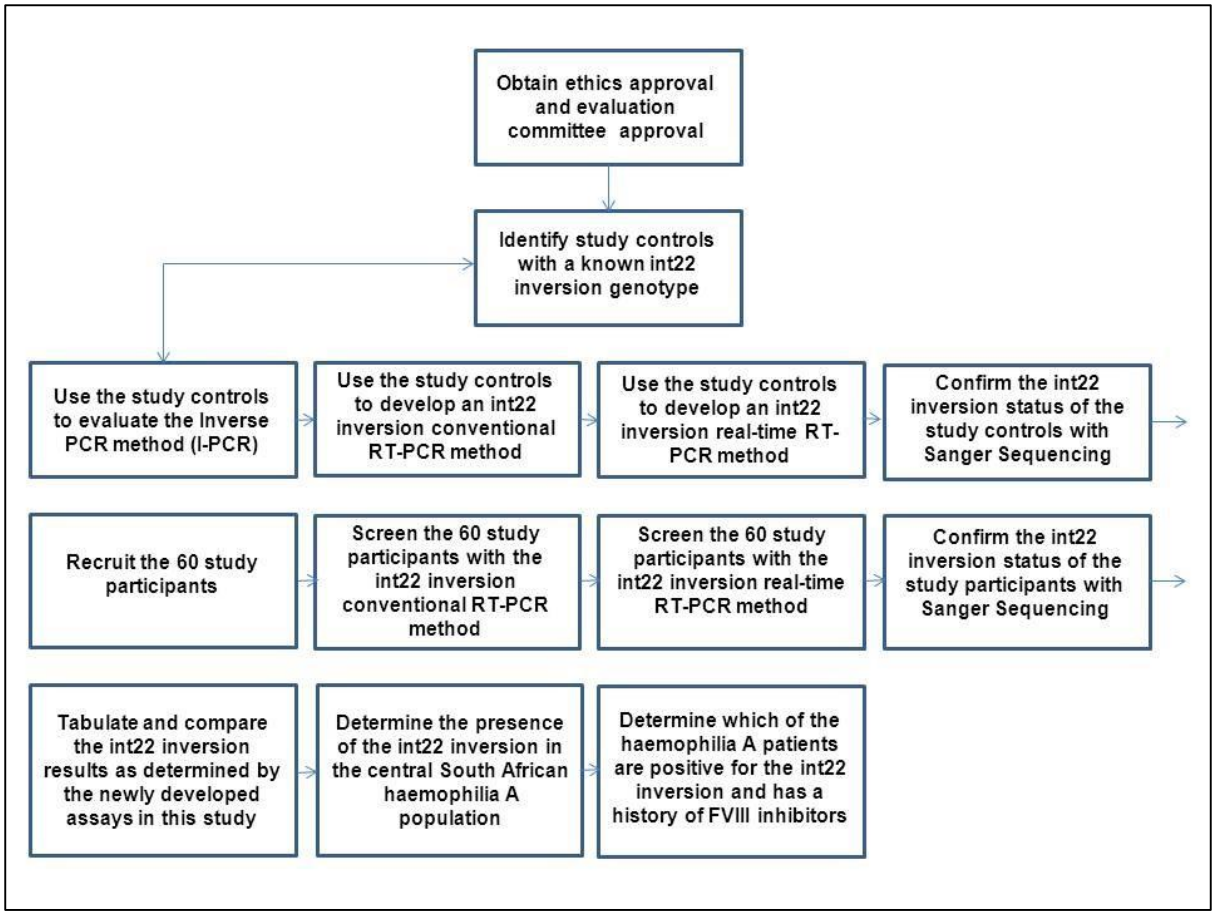


Figure 3.1: Schematic representation of the study procedure

3.5 Sample collection

Fifteen millilitres (15 mL) of venous blood were collected from each study control (C1, C2 and C3) in EDTA containing tubes (BD Vacutainer, Reference no: 368861, Becton Dickinson, South Africa). One EDTA tube of each study control was used for DNA extraction which was used in the evaluation of the inv22 I-PCR method. The remaining blood of each study control was used to extract RNA which was converted to cDNA and used in the development of the inv22 conventional and the real-time RT-PCR detection methods. Fifteen millilitres (15 mL) of venous blood was collected from all 60 study participants in EDTA containing tubes (BD Vacutainer, Reference no: 368861, Becton Dickinson, South Africa). The blood of each sample was used for RNA extraction which was converted to cDNA. The cDNA was then used to screen study participant's samples for the inv22.

3.6 Evaluation of the inv22 I-PCR detection method

The study control samples C1, C2 and C3 were used to evaluate the inv22 I-PCR detection method (Rossetti *et al.,* 2008). The complete protocol used for the method was kindly provided by the author that developed the method (Professor Liliana C Rossetti) (Rossetti *et al.,* 2008). The method procedure for the inv22 is outlined in **Figure 3.2** below.

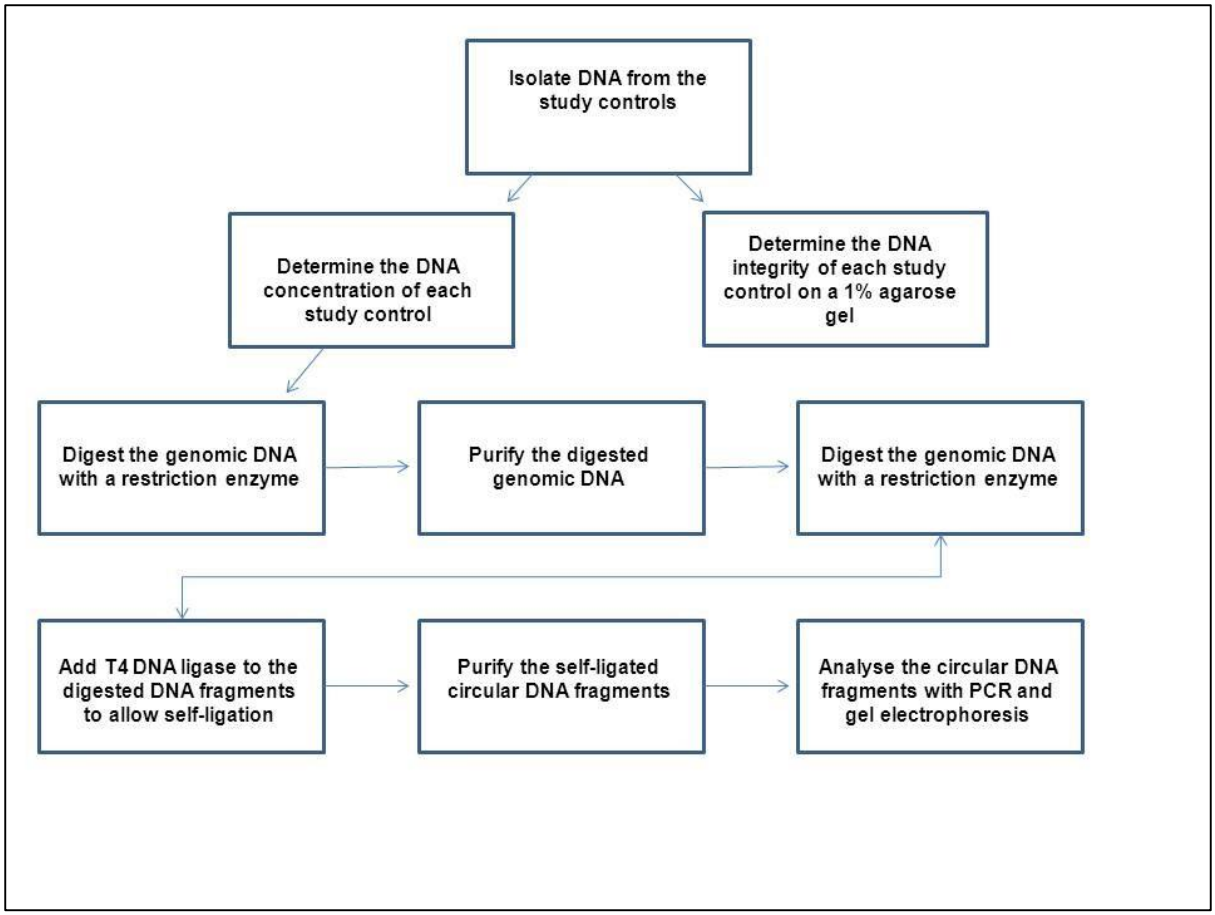


Figure 3.2: Diagram of the I-PCR study procedure

3.6.1 DNA Isolation

DNA was isolated from each of the study control samples (C1, C2 and C3) by using the Wizard® Genomic DNA Purification Kit (Promega, USA). The red blood cells were lysed by adding 1 mL of each sample to 3 mL Cell Lysis Solution. The mixture was

inverted and incubated at room temperature for 10 min. The samples were centrifuged at 2000 x *g* for 10 min. The supernatant was discarded, and the cell pellet was vortexed. In the next step, 1 mL of Nuclei Lysis Solution was added to the cell pellet, and the mixture was inverted. The proteins were precipitated by adding 330 µL of Protein Precipitation Solution to the mixture, and the samples were centrifuged at 2000 x *g* for 10 min. The supernatant of each sample was transferred to a new tube containing 1 mL of isopropanol for DNA precipitation and centrifuged at 2000 x *g* for 10 min. The DNA pellet was purified by adding 1 mL of a 70% ethanol solution to the DNA pellet, which was centrifuged at 2000 x *g* for 10 min. The ethanol was aspirated from the tube and discarded. The DNA was rehydrated with the addition of 100 µL of DNA Rehydration Solution at 65°C for 60 min.

3.6.2 Determination of DNA quantity and integrity

The DNA concentrations of the study controls were measured on the NanoDrop 2000 Spectrophotometer (Thermo Scientific, USA) by pipetting 1 µL of the eluted DNA of each sample onto the measurement pedestal. This step was repeated in triplicate, and the average DNA concentration of each study control sample was determined. Before the measurement, the instrument was blanked by pipetting 1 µL of 0.1X TE Buffer onto the measurement pedestal. The DNA concentrations of the samples were documented, and the DNA was then stored at -20°C until the day of analysis. The integrity of each DNA sample was determined by gel electrophoresis. The DNA of each sample was resolved on a 1% agarose gel, dissolved in 1X TAE buffer (2 M Tris-acetate and 0.05 M EDTA pH8) and stained with ethidium bromide (3 µL). A 1 Kb molecular weight marker (Invitrogen, USA) was also resolved on the 1% agarose gel to confirm the fragment size of the DNA products. The gel was electrophoresed at 80 V for 45 min. Afterwards, the DNA products were visualised under UV light on the Kodak Gel Logic documentation system (Carestream Molecular Imaging, Rochester, NY, USA). The DNA samples that had multiple fragments as seen on the gel image were classified as degraded. It was an important aspect as the genomic DNA input is critical for successful I-PCR (Standard Operating Procedure, Liliana C Rossetti).

3.6.3 *BclI* restriction digestion of genomic DNA

For each of the study control samples, 20 µL of a 100 ng/µL DNA template was added to 5 µL of 10X Buffer G (Thermo Scientific, USA) and 2 µL of *BclI* enzyme (10 U/µg DNA) (Thermo Scientific, USA). The reaction was made up with nuclease-free water to obtain a total reaction volume of 50 µL. The samples were incubated in a thermal cycler (Veriti, Applied Biosystems, USA) at 55°C for 3 hours.

3.6.4 *BclI* restriction fragment DNA purification

The digested genomic DNA of each study control sample was aspirated into a 1.5 mL tube, and 100 µL of nuclease-free water was added to the tube for a total reaction volume of 150 µL. The digested DNA fragments were extracted in a 1:1 volume of phenol-chloroform pH 8.0 (Affymetrix, USA) and chloroform: isoamyl alcohol (24:1) (Sigma, USA). The samples were gently mixed and centrifuged at 13,000 rpm for 5 min. The upper aqueous phase was transferred to a new tube, and the solution was extracted in one volume of chloroform: isoamyl alcohol (24:1) and centrifuged at 13,000 rpm for 5 min. The aqueous phase was transferred to a new tube, and the volume of the DNA solution was measured. The DNA was precipitated in 0.3 M NaCl, and two volumes of absolute ethanol (one volume is equal to one volume of the measured DNA solution) was added. The precipitated DNA was incubated for 30 min at -20°C and then centrifuged at 13,000 rpm for 15 min. The air-dried pellet was suspended in 50 µL of nuclease-free water.

3.6.5 DNA Ligation

A ligation reaction was prepared by adding 40 µL of 10X DNA Ligase Buffer (Thermo Scientific, USA) and 0.6 µL of T4 DNA ligase (5 U/ µL) (Thermo Scientific, USA) which was made up to 350 µL with nuclease-free water. The 50 µL of *BclI* digested DNA from the previous step was added to the 350 µL ligation master mix. The mixture was then incubated in a thermal cycler (G-Storm, Life Science Research, UK) at 15°C overnight.

3.6.6 Purification of DNA Ligation reaction

The ligated DNA of each study control was extracted in one volume (400 µL) of phenol-chloroform pH 8,0 (Affymetrix, USA) and one volume chloroform: isoamyl alcohol (24:1) (Sigma, USA). The two phases were gently mixed and centrifuged at 13,000 rpm for 5 min. The upper aqueous phase was transferred to a new tube, and the solution was extracted in one volume of chloroform: isoamyl alcohol (24:1) (Sigma, USA). The samples were centrifuged at 13,000 rpm for 5 min, and the upper aqueous phase was transferred to a new tube. The DNA solution of each sample was measured, and the DNA was precipitated in 0.3 M NaCl and two volumes of absolute ethanol. The mixed solution was then incubated at -20°C for 30 min before it was centrifuged at 13,000 rpm for 15 min. The supernatant was discarded, and the DNA pellet was air-dried before being resuspended in 40 µL of nuclease-free water.

3.6.7 Analysis of circular DNA with PCR

A PCR reaction mixture was prepared for each of the study control samples by using Q5® High-Fidelity DNA Polymerase (NEB, USA). The master mix contained 5 µL of 5X Q5 Reaction Buffer, 2 µL of a dNTP mix (2.5 mM for each dNTP) (NEB, USA), 0.25 µL of Q5® High-Fidelity DNA Polymerase (0.02 U/ µL), 1.5 µL of primer IU (10 µM), 1.5 µL of primer ID (10 µM), 1.5 µL of primer 2U (10 µM), 1.5 µL of primer 3U (10 µM) and 6.9 µL of nuclease-free water. After that, 5 µL of the DNA-ligated product of each sample was added to the respective PCR mixtures. The primers used in this reaction are summarised in **Table 3.1**. The PCR mixtures were briefly centrifuged and vortexed. The mixtures were subjected to the following PCR cycling conditions in a thermal cycler (Veriti, Applied Biosystems, USA); initial denaturation at 94°C for 2 min, 30 cycles at 94°C for 30 sec, 56°C for 60 sec and 72°C for 90 sec. A final extension step of 72°C was conducted for 5 min. A non-template control (NTC) was included for quality control purposes.

Table 3.1: The primer sets for the inv22 I-PCR detection method. The table indicates the primer name, sequences, accession numbers and expected PCR fragment sizes.

Primer	Sequence 5'-3'	NC_000023.9	Expected PCR fragment sizes
ID	ACATACGGTTTAGTCACAAGT	153758587-608	n/a
IU	CCTTTCAACTCCATCTCCAT	153779730-50	Primer ID and IU: 487 bp
2U	ACGTGTCTTTTGGAGAAGTC	154270775-95	Primer ID and 2U: 385 bp
3U	CTCACATTGTGTTCTTGTAGTC	154333426-48	Primer ID and 3U: 333 bp

*ID/IU/2U/3U: The primer names (Rossetti, *et al.,* 2008). NC_000023.9: NCBI accession number of the reference sequence used to design the primers. bp: base pairs

3.6.8 Analysis of PCR products

The PCR products were evaluated with gel electrophoresis to determine if the expected PCR fragments were obtained. Consequently, the PCR products were resolved on an ethidium bromide (3 μL) stained 1.5% agarose gel in 1X TAE buffer (2 M Tris-acetate and 0.05 M EDTA pH 8,0). The gel was electrophoresed at 100 V for 40 min. A molecular weight marker (NEB, USA), that was also resolved on the gel was used to confirm the specific PCR fragment sizes. The gel image was visualised under UV light on the Kodak Gel Logic documentation system (Carestream Molecular Imaging, Rochester, NY, USA).

3.7 The development of an inv22 conventional and a real-time RT-PCR detection method

The study controls were used to develop an inv22 conventional and a real-time RT-PCR detection method. The RNA of each study control sample was firstly stabilised before RNA was isolated. The RNA concentrations were measured, and cDNA was prepared. The cDNA concentrations of each sample were determined and used to optimise the inv22 conventional and real-time RT-PCR detection methods, respectively.

3.7.1 RNA stabilization

The RNA of the study controls were stabilised before RNA isolation. For each sample, 1.2 mL of RNA*Later*™ solution (Ambion, Life Technologies, USA) was dispensed in a 2 mL microfuge tube. Whole blood of each sample was fractionated by centrifugation at 1500-2000 x *g* for 15 min at room temperature. The plasma was aspirated using a disposable sterile plastic transfer pipet and disposed of. The exposed white blood cells were then aspirated and expelled into the tube containing the RNA*Later*™solution. The white blood cell and RNA*Later*™ solution were mixed thoroughly by inversion and stored at -20°C until the day of RNA isolation.

3.7.2 RNA Isolation

RNA was isolated from each sample using the RiboPure-Blood Kit (Ambion, Life Technologies, USA), according to the manufacturer's instructions. Samples stored in RNA*Later*™ were centrifuged for 1 min at 14,000 rpm. The supernatant was then removed by aspiration, and 800 µL of Lysis Solution and 50 µL of Sodium Acetate Solution were added to the cell pellet of RNA*Later*™-stabilised samples. The samples were then vigorously vortexed to lyse the cells and 500 µL of 1:1 Acid-Phenol: Chloroform solution was added to the cell lysate with an additional 30 sec vortex step. The mixture was left to stand at room temperature for 5 min before it was centrifuged at 14,000 rpm for 1 min at room temperature for phase separation. The aqueous phase was transferred to a microfuge tube with 600 µL of 100% molecular-grade ethanol and mixed by vortexing. Filter cartridges were used to purify the RNA, and 700 µL of the

sample (aqueous phase and ethanol mixture) was added to the filter cartridge assembly. The samples were centrifuged for 10 sec at 14,000 rpm to pass the solution through the cartridge, and the flow-through was discarded. The step was repeated to filter the remainder of the sample. The filter was then washed by adding 700 µL of Wash Solution 1 to the filter cartridge, and the cartridges were centrifuged for 10 sec at 14,000 rpm to pass the wash solution through the filter. The flow-through was discarded before 700 µL of Wash Solution 2/3 was added to the filter cartridge. The cartridge was centrifuged for 10 sec at 14,000 rpm, and this step was repeated. After the flow-through was discarded, the cartridge was centrifuged for 1 min at 14,000 rpm to allow residual fluid to be removed from the filter. The filter cartridge was placed in a clean microfuge collection tube, and 50 µL of Elution buffer was added to the filter and left at room temperature for 20 sec before centrifugation for 30 sec at 14,000 rpm to recover the RNA. A DNase I digestion step was included to remove genomic DNA from the eluted RNA. It was conducted by adding 1/20th volume of a 20X DNase Buffer and 1 µL DNase I (8U/µL) to the eluted RNA. The samples were incubated for 30 min at 37°C on a heating block. Finally, DNase Inactivation reagent equal to 20% of the sample was added, and the samples were mixed thoroughly. RNA samples were stored at -80°C until the day of analysis.

3.7.3 Determination of RNA concentrations

The RNA concentration of each sample was measured on the Qubit® 2.0 Fluorometer. The Qubit® 2.0 Fluorometer utilises specially formulated dyes to bind specifically to RNA. When the formulated dye binds to the specific target molecule, it emits fluorescence that is detected by the instrument which can then be used to determine the RNA concentrations (Invitrogen, USA).

RNA concentrations were then determined in duplicate on the Qubit® 2.0 Fluorometer for each sample. The RNA concentration of each sample was determined immediately after RNA isolation with a Qubit® RNA HS Assay Kit (Invitrogen, USA). Qubit® Working Solution was prepared by diluting the Qubit® RNA Reagent (1:200) in Qubit® RNA Buffer. A 200 µL working solution was prepared for the two calibration standards and each sample. For the two calibration standards, 190 µL of working solution was added to 10 µL of each corresponding calibration standard in a 0.6 mL tube. Each

sample tube was prepared by adding 199 µL of working solution to the tube and 1 µL of each sample (The RNA of each sample was diluted 1:5 by adding 1 µL of RNA to 4 µL of nuclease-free water). The tubes were vortexed for 5 sec and incubated at room temperature for 2 min (Invitrogen, USA).

3.7.4 cDNA synthesis and concentration determination

The High-Capacity RNA-to-cDNA kit (Applied Biosystems, USA) was used to prepare the cDNA for each sample as per manufacturer's instructions. The kit components were thawed on ice before the setup. The reactions were prepared on ice, and for each sample, 10 µL of a 2X RT Buffer, 1 µL of a 20X Enzyme mix and 2µg (2000 ng/ µL) of RNA was added to a 0.2 mL PCR tube. A total reaction volume of 20 µL was obtained by adding nuclease-free water. A reverse transcriptase negative (-RT) control was included in each reaction run. The (-RT) control contained all the components except the 20X Enzyme Mix. The samples were briefly centrifuged and were kept on ice before they were incubated at 37°C for 60 min. The reaction was stopped by heating the samples to 95°C for 5 min. The incubation and heating step was conducted in a thermal cycler (Veriti, Applied Biosystems, USA). The cDNA concentrations of each sample were then determined by using the NanoDrop 2000 Spectrophotometer (Thermo Scientific, USA) by pipetting 1 µL of the eluted cDNA of each sample onto the measurement pedestal. This step was repeated in duplicate, and the average cDNA concentration of each sample was determined. Before the measurement, the instrument was blanked by pipetting 1 µL of 0.1X TE Buffer onto the measurement pedestal (SOP, Quantifying Complementary Deoxyribonucleic Acid).

3.8 The Optimization of the inv22 conventional RT-PCR detection method

3.8.1 Primer design and selection

An inv22 conventional RT-PCR detection method was additionally developed for those laboratories that are not set up for real-time PCR. The primers were designed to amplify the regions in the *FVIII* gene that are associated with the inv22. Therefore, two sets of primers were designed using the online program, IDT PrimerQuest (Integrated

DNA Technologies, https://eu.idtdna.com/Primerquest/Home/Index). The primers were evaluated using the online program, IDT Oligo Analyzer (Integrated DNA Technologies, 2017, Available at https://eu.idtdna.com/calc/analyzer) to verify oligonucleotide parameters such as melting temperatures, hairpin loop formation and primer self-dimerization. The primers were also evaluated using the Basic Alignment Search Tool (BLAST) to ensure that they are specific to their target sequences (https://www.ncbi.nlm.nih.gov/BLAST). The lyophilised primers were reconstituted by adding the manufacturer's prescribed volume of 0.1X TE buffer to each primer. The first set of primers (wild-type primers) was designed to amplify the splice region between the boundaries of exon 22 and exon 23 in the *FVIII* gene, which represented the FVIII wild-type genotype (inv22 negative). The non-haemophilic A volunteer, C1, was used to evaluate the wild-type primer set. It was expected that the wild-type primers would yield a PCR fragment of 260 bp. The wild-type primer set was also tested on study controls C2 and C3 that represented the inv22 genotype (mutant-genotype).

The second set of primers (mutant-type primers) was designed to amplify the splice region between exon 22 and exon 23c, which represented the mutant-type genotype (inv22 positive) as predicted by Pandey *et al,* 2013. The study control samples, C2 and C3, were used to evaluate the mutant-primer set, and it was expected that the mutant-type primers would yield a PCR fragment of 132 bp in size. The mutant-type primer set was also used to test the wild-type study control C1. Consequently, the wild-type and mutant-type primers sets were selected for the inv22 conventional RT-PCR detection method (**Table 3.2**). The specific exon-binding sites of the primers, as tabulated in **Table 3.2**, are illustrated in **Figure 3.3**.

Table 3.2 Wild-type and mutant-type primer sets for the inv22 conventional RT-PCR detection method. The table indicates the primer names, sequences, melting temperatures (Tm), guanine-cytosine content and the expected PCR fragment sizes.

Primers	Sequence 5'-3'	Tm	GC%	Expected PCR fragment sizes
Wild-type F	TATTCACGGCATCAAGACCC	60°C	50%	260 bp
Wild-type R	TCAACTCCATGCGAAGAGTG	60°C	50%	
Mutant-type F	CAGCCTCTACATCTCTCAGTTTATC	62°C	44%	132 bp
Mutant-type R	TAGGAAGGCTCCCAATTCTTTC	62°C	45.5%	

*F: Forward primer. R: Reverse primer. Tm: Melting temperature. GC%: Guanine and Cytosine content. bp: base pairs.

The wild-type primer binding site on mRNA sequence: *FVIII* exon 22 spliced to *FVIII* exon 23.

GGTGGATCTGTTGGCACCAATGATTATTCACGGCATCAAGACCCAGGGTGCCC
GTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTCTTGA
TGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATG:GTCT
TCTTTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAAT
TATTGCTCGATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAGCACTCTT
CGCATGGAGTTGATGGGCTGTGATTTAAATA

Red: Wild-type F primer.

Green: Wild-type R primer.

*The (:) indicates the sequence site where exon 22 is spliced to exon 23

Mutant-type primer binding site on the mRNA sequence: *FVIII* exon 22 spliced to alternative exon 23c.

TTTCAGATTACAGCTTCAGGACAATATGGACAGTGGGCCCCAAAGCTGGCCAG
ACTTCATTATTCCGGATCAATCAATGCCTGGAGCACCAAGGAGCCCTTTTCTTG
GATCAAGGTGGATCTGTTGGCACCAATGATTATTCACGGCATCAAGACCCAGG
GTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAG
TCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAAT
G:CACGGTGTCTTGGAAAATGGAAAGAATTGGGAGCCTTCCTACCGCTGGTGA

Red: Mutant-type F primer.

Green: Mutant-type R primer

*The (:) indicates the sequence site where exon 22 is spliced to exon 23c

Figure 3.3: Illustration of the wild-type and mutant-type primer binding sites. F: Forward primer. R: Reverse primer.

3.8.2 The determination of the optimal annealing temperatures for the inv22 conventional RT-PCR

A temperature gradient was set up for both the wild-type primers set (Wild-type F and R) and the mutant-type primers (Mutant-type F and R) to determine the optimised annealing temperature. The annealing temperatures ranged from 57°C to 62°C, with an increment rate of 1°C. For each study control sample, two separate PCR reactions were prepared, one containing the wild-type primers and the other containing mutant-type primers, respectively. Each PCR mixture contained 5 µL of 5X Q5 Reaction Buffer and 0.25 µL of Q5® High-Fidelity DNA Polymerase (0.02 U/ µL) (NEB, USA). As well as 1 µL of dNTP mix (2.5 mM for each dNTP) (NEB, USA), 1 µL of forward (10 µM) and reverse (10 µM) primer, 1 µL of cDNA (100 ng/µL) and finally nuclease-free water for a total reaction volume of 25 µL. The prepared PCR mixtures were then briefly centrifuged and vortexed, and placed in a thermal cycler (Veriti, Applied Biosystems, USA). The cycling conditions were one cycle at 98°C for 10 min, 40 cycles at 98°C for 30 sec, 57°C to 62°C for 1 min and 72°C for 1 min. The final step was one cycle at 72°C for 5 min. A non-template control (NTC) was included in each run for quality control purposes. A multiplex PCR containing both the wild-type and mutant-type primers were also evaluated using the study control samples C1, C2 and C3. The same PCR conditions that were used for the single primer reactions were used for a multiplex reaction.

The PCR products were resolved on an ethidium bromide (3 µL) stained 2% agarose gel in 1X TAE buffer (2 M Tris-acetate and 0.05 M EDTA pH 8,0). The gel was electrophoresed at 100 V for 40 min. A molecular weight marker (NEB, USA) that was also resolved in the gel was used to confirm the specific PCR fragment sizes. The gel image was visualised under UV light on the Kodak Gel Logic documentation system (Carestream Molecular Imaging, Rochester, NY, USA).

3.8.3 Evaluation of control samples C1, C2 and C3 with the inv22 conventional RT-PCR detection method at the optimal determined annealing temperatures

The optimised inv22 conventional RT-PCR assay was used to evaluate C1, C2 and C3. The PCR was set up as in **section 3.8.2**, except the optimised annealing temperature for the wild and mutant-type reaction, were used. The PCR products were

resolved on an ethidium bromide (3 µL) stained 2% agarose gel in 1X TAE buffer (2 M Tris-acetate and 0.05 M EDTA pH 8,0). The gel was electrophoresed at 100 V for 40 min. A molecular weight marker (Bioline, USA) that was also resolved in the gel was used to confirm the specific PCR fragment sizes. The gel image was visualised under UV light on the Kodak Gel Logic documentation system (Carestream Molecular Imaging, Rochester, NY, USA) to determine whether C1, C2 and C3 are FVIII wild-type positive, inv22 homozygous or inv22 carriers, respectively.

3.8.4 Limit of detection (LOD) for the inv22 conventional RT-PCR

The limit of detection (LOD) for the inv22 conventional RT-PCR was determined by using study control C2. RNA was isolated from C2, and the RNA concentration was determined (**sections 3.7.2 and 3.7.3**). Thereafter, the cDNA was prepared and the known cDNA concentration (**section 3.7.4**) was used to make a range of dilution samples using nuclease-free water: 1000 ng/µL, 500 ng/µL, 100 ng/µL, 50 ng/µL, 10 ng/µL and 0 ng/µL (blank control). Consequently, the inv22 conventional RT-PCR was set up at the optimal annealing temperature (**section 3.8.2**) using the six diluted samples. Only the mutant-type reaction was set up as the LOD for inv22 had to be determined. The PCR products were resolved on an ethidium bromide (3 µL) stained 2% agarose gel in 1X TAE buffer (2 M Tris-acetate and 0.05 M EDTA pH 8,0). The gel was electrophoresed at 100 V for 40 min. A molecular weight marker (NEB, USA) that was also resolved in the gel was used to estimate the specific PCR product sizes. The gel image was visualised under UV light on the Kodak Gel Logic documentation system (Carestream Molecular Imaging, Rochester, NY, USA) to determine the LOD for the inv22 conventional RT-PCR detection method.

3.9 The Optimization of the inv22 real-time RT-PCR detection method

The study controls (C1, C2 and C3) were used in the development of the inv22 real-time RT-PCR detection method. Power SYBR® Green Master Mix (Applied Biosystems, USA) was used as detection chemistry with an additional melt profile analysis to ensure real-time RT-PCR reaction optimization.

3.9.1 Primer selection and cDNA concentrations

The wild-type and mutant-type primer pairs that were optimised and selected for the inv22 conventional RT-PCR detection method were subsequently used in the inv22 real-time RT-PCR detection method (**Table 3.2**, and illustrated in **Figure 3.3**). The optimal cDNA concentrations for the method were determined by evaluating each of the study control samples at 10 ng/µL, 50 ng/µL and 100 ng/µL. The cDNA concentration of each of the study control samples, as measured previously (**section 3.7.4**), were used to make a dilution range (10 ng/µL, 50 ng/µL and 100 ng/µL) using 0.1X TE Buffer.

After that, for each of the study control samples, a wild-type and mutant-type real-time RT-PCR reaction (single-reactions) was prepared using the different cDNA concentrations (10 ng/µL, 50 ng/µL and 100 ng/µL) in duplicate. Each reaction contained 12.5 µL of Power SYBR® Green PCR Master Mix (Applied Biosystems, USA), 1 µL of the respective forward (10 µM) and reverse (10 µM) primers, 1µL of sample cDNA and 4.5 µL of nuclease-free water for a total reaction volume of 20 µL. Each sample was briefly centrifuged and loaded onto the Applied Biosystems™ QuantStudio™ 5, Real-Time PCR System (Applied Biosystems, USA). The pre-programmed, SYBR green Quantification assay for a standard block, 0.2 uL reactions, was selected, with ROX-dye as a passive control. The real-time PCR cycling conditions were as follows: one cycle at 95°C for 10 min followed by 40 cycles at 95°C for 15 sec and 62°C for 1 min. After that melt curve analysis was conducted at 95°C for 1 min, 62°C for 1 min followed by 40 cycles of melting starting at 60°C with 0.5°C increments (10 sec for each increment) of heating.

The real-time RT-PCR amplification curves and melt curves were then analysed to determine the accuracy of the inv22 real-time RT-PCR detection method in determining the FVIII wild-type, inv22 homozygous and inv22 carrier genotypes.

3.10 Confirmation of study controls C1, C2 and C3 with Sanger sequence analysis

Sanger sequence analysis was used to confirm study control C1 as inv22 negative and study controls C2 and C3 as inv22 positive, respectively. The analysis was based

on the presence or absence of exon 23c as described by Pandey, *et al.* (2013). It was expected that C1 would contain the FVIII wild-type exon 23, and conversely, C2 and C3 would contain exon 23c as it is associated with inv22. The primers that were designed for the inv22 conventional and real-time RT-PCR were used in the sequencing reactions (**Table 3.2**).

The PCR product of each study control sample that yielded DNA fragments detectable by gel electrophoretic analysis was purified by using the ExoSAP-IT® *Express* PCR Product Cleanup (Affymetrix, USA). For the purification reaction, 5 µL of PCR product and 2 µL of ExoSAP-IT® *Express* was mixed and incubated at 37°C for 4 minutes to degrade excess primers and nucleotides. After that, the ExoSAP-IT® *Express* enzyme was inactivated at 80°C for 1 min. The purified PCR product was used in the sequencing reaction, and the sequencing reaction was prepared by using the BigDye™ Terminator v3.1 Kit (Applied Biosystems, USA). The forward sequencing reaction for the wild-type and the mutant-type contained 4 µL BigDye™ Terminator v3.1 Ready Reaction Mix, 2 µL BigDye™ Terminator v3.1 Sequencing Buffer, 2 µL of the 3.2 µM forward primer (Wild-type F or Mutant-type F), 2 µL of the purified PCR product and 10 µL of nuclease-free water for a total reaction volume of 20 µL.

The wild-type and mutant-type reverse sequencing reactions contained the same sequencing cocktail. However the wild-type reverse sequence reaction contained 2 µL of the 3.2 µM reverse primer, Wild-type R, and the mutant-type reverse sequencing reaction contained 2 µL of the 3.2 µM reverse primer, Mutant-type R. The sequencing reactions were subjected to the following cycling conditions in the thermal cycler (Veriti, Applied Biosystems, USA). One (1) cycle at 96°C for 1 min, 25 cycles at 96°C for 10 sec, 50°C for 5 sec and 60°C for 4 min.

Subsequently, the sequencing products were purified using the ZR DNA Sequencing Clean-Up Kit™ (Zymo Research, USA). For the purification reaction, 240 µL of Sequencing Binding Buffer was added to the sequencing product, and the mixture was briefly vortexed. The mixture was then transferred to a collection tube and centrifuged at 13,000 rpm for 30 sec. The flow-through was discarded, and 300 µL of Sequencing Wash Buffer was added to the column. The columns were centrifuged at 13,000 rpm for 30 sec with an additional centrifugation step at 13,000 rpm for 1 min to ensure all excess Sequencing Wash Buffer is removed from the filter cartridge. The sequencing

product was eluted by adding 20 µL of nuclease-free water to the column. The columns were centrifuged at 13,000 rpm for 15 sec to collect the purified sequence containing elute.

The sequence raw reads were analysed on the ABI Prism 3130 Genetic Analyzer (Applied Biosystems, USA). The sequencing data were analysed by using the Sequencing Analysis Program v5.3.1 (Applied Biosystems, USA) and Chromas v2.5.1. Alignment comparison was conducted between the sample sequences and the reference sequences using LALIGN online sequence alignment tool (Huang & Miller, 1991). The sequences obtained for the study controls were compared to the exon 22 and exon 23 mRNA reference sequence of the *FVIII* gene (NM_000132.3) and the exon 22 and exon 23c mutant-type mRNA reference sequence as described by Pandey *et al*. (2013) (**Appendix F**).

3.11 Screening of the 60 study participants with the inv22 conventional and the real-time RT-PCR detection methods

RNA was stabilised and isolated from the 60 study participants' samples, and the RNA concentrations were determined. After that cDNA was synthesised and the cDNA concentrations were determined before the cDNA was used for the respective inv22 conventional and real-time RT-PCR detection methods. This method procedure was conducted as described in **section 3.7.1 to 3.7.4**.

3.11.1 The inv22 conventional RT-PCR detection method

The 60 study participants were screened with the inv22 conventional RT-PCR detection method which was optimised as described in **section 3.8**. For each study participant, a wild-type and a mutant-type PCR reaction were prepared. An NTC was included in each run for quality control purposes. The samples were loaded onto a 96-well thin-wall PCR plate and sealed with optical film before being centrifuged briefly. The PCR plate was loaded onto the thermal cycler (Veriti, Applied Biosystems, USA) and the cycling conditions were one cycle at 98°C for 10 min followed by 40 cycles at 98°C for 30 sec, 62°C for 1 min and 72°C for 1 min. The next cycle was at 72°C for 5 min and a hold cycle at 4°C. The PCR products were resolved on an ethidium

bromide (3 µL) stained 2% agarose gel in 1X TAE buffer (2 M Tris-acetate and 0.05 M EDTA pH 8,0). The gel was electrophoresed at 100 V for 40 min. A molecular weight marker (NEB, USA) that was also resolved on the gel was used to confirm the PCR fragments. Afterwards, the gel images were visualised under UV light on the Kodak Gel Logic documentation system (Carestream Molecular Imaging, Rochester, NY, USA) to determine if the study participants had the FVIII wild-type, inv22 homozygous or inv22 carrier genotypes, respectively.

3.11.2 The inv22 real-time RT-PCR detection method

After that, the 60 study participants were screened with the inv22 real-time RT-PCR detection method. The cDNA of each sample was diluted to a 100 ng/µL solution with 0.1X TE buffer. For each sample, two separate real-time RT-PCR reactions were prepared in duplicate, the first reaction contained the wild-type primers (wild-type F and R) and the second reaction contained the mutant-type primers (mutant-type F and R). An NTC was included in each run and ROX-dye as a passive control. Control C1 served as the inv22 negative control, and C2 and C3 served as the inv22 positive controls. The real-time RT-PCR reactions for each sample were prepared in a 96-well thin-wall PCR plate. It was done by adding 12.5 µL of Power SYBR® Green PCR Master Mix (Applied Biosystems, USA), 1 µL of the respective forward (10 µM) and reverse (10 µM) primers, 1µL of sample cDNA (100 ng/µL) and 4.5 µL of nuclease-free water for a total reaction volume of 20 µL to each respective well. The PCR plate was sealed with optical film and was briefly centrifuged. The PCR plate was uploaded onto the Applied Biosystems™ QuantStudio™ 5, Real-Time PCR System (Applied Biosystems, USA). The pre-loaded SYBR Green Quantification Assay program for a standard block, 0.2 uL reactions, was selected. The real-time PCR cycling conditions started with one cycle at 95°C for 10 min followed by 40 cycles at 95°C for 15 sec and 62°C for 1 min.

3.11.3 Sanger sequence analysis

The PCR products of each study participant as determined with the inv22 conventional RT-PCR detection method that yielded DNA fragments detectable by gel

electrophoretic analysis were analysed by Sanger sequencing. The sequence analysis was conducted as described in **section 3.10** above.

3.12 Comparison of the inv22 detection methods developed in this study

The inv22 results obtained for the study controls with the Southern Blot (**Appendix A**), inv22 conventional RT-PCR, the real-time RT-PCR detection method, and Sanger sequence analysis were subsequently tabulated and compared. Likewise, the inv22 results obtained with the conventional RT-PCR method, the real-time RT-PCR method, and Sanger sequence analysis were tabulated and compared for the 60 study participants.

3.13 The presence of the inv22 in the central South African haemophilia A population

The presence or absence of the inv22 in the central South African haemophilia A population was evaluated and expressed as a percentage. The percentage inv22 participants in our study population were then compared to other studies in the world that also determined the presence of the inv22 in their respective study populations. It was furthermore determined which of the study participants who had the inv22, also had a history of FVIII inhibitor development.

Chapter 4: Results and Discussion

4. Introduction:

In this chapter the results of the study are given, data analysis is conducted followed by an in-depth discussion of the results. Furthermore, the recommendations, limitations and the impact of the study are discussed.

4.1. Population group:

The haemophilia A population of South Africa needs to be taken into consideration when the sample size used in this study is evaluated. As previously mentioned, haemophilia A is underdiagnosed in South Africa (Mahlangu & Medical and Scientific Council of the South African Haemophilia Federation, 2009) and therefore, a small, but a comprehensive group (n=63) in central South Africa were included in this study. The majority of the study population consisted of self-proclaimed black African (67.7%), followed by the Caucasian population (20.9%) and the mixed ancestry population (11.3%). The int22 inversion (inv22) status of the study participants was mostly unknown. The possible carriers were all female and either the mother or the sister of a haemophilia A patient, respectively. The ratio of male to female participants was 45:18. The recruitment of the participants was a great challenge in this study. Usually, haemophilia A patients only visited the haematology clinic when they require a new prescription for FVIII concentrates or in cases where those patients had bleeding episodes. Nevertheless, a satisfactory number of haemophilia A patients and study controls could be recruited for this study to achieve the study aim and objectives.

4.2 Evaluation of the inv22 Inverse PCR (I-PCR) detection method

The inv22 I-PCR detection method is one of the main methods currently used globally. The United Kingdom National External Quality Assessment Service (UKNEQAS) indicated that half of the laboratories participating in their external quality assurance program are currently using the I-PCR method (Cutler *et al.,* 2015). Therefore, the method was evaluated in this study.

The evaluation of the inv22 I-PCR detection method was conducted stepwise according to a standard operating procedure (SOP) that is used by the developer of the method (Personal communication) (Rossetti *et al.,* 2008). The SOP has several recommendations that need to be considered when using the inv22 I-PCR as a detection method. Firstly, the quality and the exact quantity of the input genomic DNA is critical to the success of the method. The *BcI* I enzyme used in the DNA digestion step can only be inactivated by using phenol-chloroform treatment and not by heat inactivation, according to the SOP. The *BcI* I enzyme manufacturer (Technical note, Product ER0721, Thermo Scientific, USA) however suggests that the enzyme can be inactivated by using silica column based purification methods. The SOP further indicates that it is crucial to avoid carry-over of chloroform during the phenol-chloroform treatment step. Excess NaCl should not be used when precipitating the DNA as it can inhibit DNA-ligation and PCR-amplification steps. It is essential to take into consideration that the T4 DNA ligase buffer contains ATP and should thus be stored at -20°C with care being taken not to freeze and thaw multiple times. The critical step according to the SOP is to allow DNA-ligation to occur in large volumes such as 400 µL as it favours the formation of DNA circles. All of the above-mentioned recommendations were implemented when the I-PCR was performed in this study.

4.2.1 Determination of DNA quantity and integrity

The DNA concentration of each control sample that was measured in triplicate is summarised in **Table 4.1** below. The genomic DNA for control C1, C2 and C3 were used in the DNA-digestion step and their DNA concentrations were above 250 ng/µL, which were suitable for the digestion reaction, as only 100 ng/µL was required. After the DNA of controls C1, C2 and C3 were extracted, the genomic DNA was evaluated on a 1% agarose gel (**Figure 4.1**). The genomic DNA of each of the controls was of

acceptable integrity for the I-PCR method, as no multiple fragments, indicative of DNA degradation, was observed.

Table 4.1: DNA concentrations measured for study control sample C1, C2 and C3.

Control Sample	DNA concentration: Measurement 1 (ng/µL)	DNA concentration: Measurement 2 (ng/µL)	DNA concentration: Measurement 3 (ng/µL)	Average (ng/µL)
C1	250	256	255	254
C2	321	323	324	323
C3	350	359	357	356

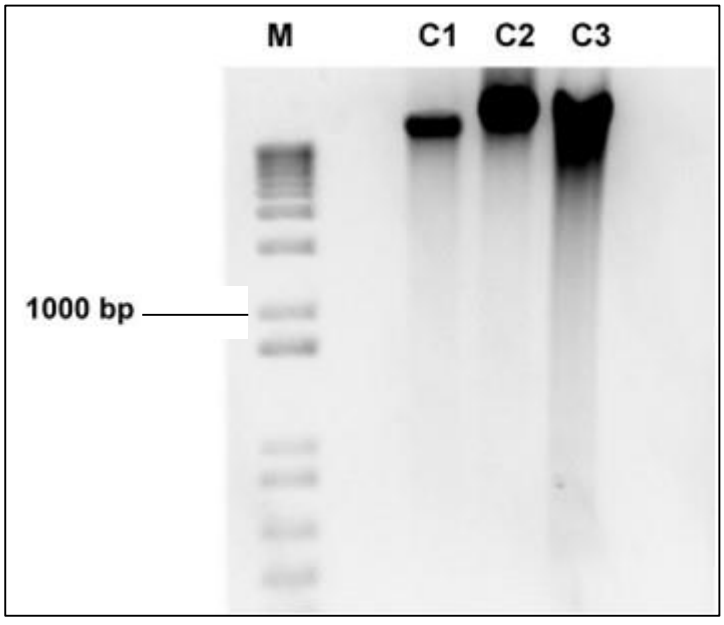


Figure 4.1: A 1% agarose gel image depicting the genomic DNA of the study controls C1, C2 and C3. M: Denotes the 1 Kb molecular weight marker.

4.2.2 Troubleshooting measures implemented for the inv22 I-PCR detection method

The inv22 I-PCR yielded no results when controls C1, C2 and C3 were evaluated, and therefore several troubleshooting measures were attempted. As the method has several steps that need to be conducted before a result can be obtained, the troubleshooting measures were challenging. Consequently, each of the method's steps was evaluated individually to resolve the problem that occurred. The following troubleshooting measures were taken after consultation with the developer of the method (Personal communication):

4.2.2.1 Repeated I-PCR

The first step was to repeat the I-PCR analysis. Subsequently, the nuclease-free water that was used in the reaction was replaced, and the I-PCR analysis was repeated. Post-PCR gel electrophoresis analysis revealed no results.

4.2.2.2 Systematic replacement of PCR reagents

The next step was to replace the PCR reagents that were used systematically. Firstly, the nuclease-free water was replaced, followed by the Taq polymerase, dNTP's, $MgCl_2$ and the PCR buffer. After these measures were taken, there were still no results obtained. After that, a range of different DNA concentrations (100 ng/µL to 1000 ng/µL) were used in the reactions, all of which yielded no results.

4.2.2.3 Re-extraction of genomic DNA

The DNA of study control samples C1, C2 and C3 were re-extracted, and the DNA concentration of each sample was determined. The DNA concentrations that were measured ranged between 221 ng/µL and 280 ng/µL and were, therefore, suitable for the inv22 I-PCR method. All the steps of the inv22 I-PCR detection method were conducted, and post-PCR gel electrophoresis analysis once again revealed no results.

4.2.2.4 Replacement of the *BclI* enzyme for DNA-digestion step

The B*cl* I enzyme that was used in the DNA-digestion step of the method was replaced together with the buffer that was used with the enzyme and the method's steps were repeated. The post-PCR analysis revealed no results.

4.2.2.5 Column-based purification of B*cl* I fragment

The SOP suggests that phenol-chloroform treatment should be used after DNA-digestion to inactivate the B*cl* I enzyme. This treatment was conducted, however, as the reaction, still failed a column-based (GeneJET, Thermo Fisher Scientific, USA) purification reaction was used. The column based purification reaction had no effect as post-PCR analysis revealed that no results were obtained.

4.2.2.6 Replacement of the T4 DNA-ligase enzyme

The T4 DNA-ligase enzyme that is used for DNA-ligation in the method were replaced to determine whether the reaction is failing due to the T4 DNA-ligase being degraded. The new T4 DNA-ligase (Thermo Scientific, USA) was then used in the DNA-ligation step of the inv22 I-PCR detection method, however, after post-PCR analysis, no results were obtained.

The main challenge that was associated with the inv22 I-PCR detection method was that there were no quality controls measures that could be implemented between the different steps. There is thus a lack of checkpoints between DNA-digestion, DNA-ligation and the PCR. The only way to confirm whether DNA-digestion was successful is with gel electrophoresis, which is not an absolute precise method. Gel electrophoresis image results obtained after DNA-digestion in this study produced a faint smear (data not shown) which may or may not be indicating low DNA concentrations influencing the success of the DNA-digestion step.

Similarly, there are no measures in place that can determine if DNA-ligation was successful without continuing with the PCR-analysis. In essence, only after post-PCR gel electrophoresis analysis is conducted can it be determined whether the I-PCR was successful.

Several troubleshooting measures were taken to obtain results with the inv22 I-PCR detection method as described above. These measures were taken with constant advice being provided by the developer of the specific I-PCR method (Personal communication). Consequently, the evaluation of the I-PCR method was deemed unsuccessful. It is unknown whether other laboratories in South Africa found similar challenges in implementing the inv22 I-PCR method and is worth investigating. The disadvantages associated with the method make it difficult to reproduce in a sub-optimally resourced laboratory and therefore highlights the need for a simple and cost-effective method to detect the inv22.

4.3 The development of an inv22 conventional and real-time RT-PCR detection method

4.3.1 RNA and cDNA concentrations

The RNA and cDNA concentrations of the study controls and the study participants ranged between 44 ng/µL to 313 ng/µL and 497 ng/µL to 1996 ng/µL, respectively (**Appendix G**)**.** Considering that for each cDNA reaction prepared a total of 2000 ng/µL total RNA was added, there was a slight decrease in concentration as the cDNA concentrations were lower (< 2000 ng/µL).

4.3.2 The Optimization of the inv22 conventional RT-PCR detection method

The two primer pairs (wild-type and mutant-type) (**Table 3.2**) that were optimised lead to the amplification of PCR fragments that could determine if patients have the wild-type, inv22 homozygous or inv22 carrier genotypes. The study controls C1, C2 and C3 were used in the optimization of the wild-type and the mutant-type primer sets, respectively.

4.3.2.1 Wild-type primer set optimization

A temperature gradient was conducted to determine the optimal annealing temperature for the wild-type primer set. The gel image in **Figure 4.2** below depicts the results. It was determined that study control sample C1 yielded an expected PCR fragment of 260 bp at each of the annealing temperatures (57 °C to 62 °C). Therefore the optimal annealing temperature for the wild-type primer set was 62 °C. When study control C2 and C3 were tested with the wild-type primers, no PCR fragment could be detected, which was expected, as C2 and C3 were negative for the FVIII wild-type genotype as confirmed by Southern Blot analysis (**Appendix A**).

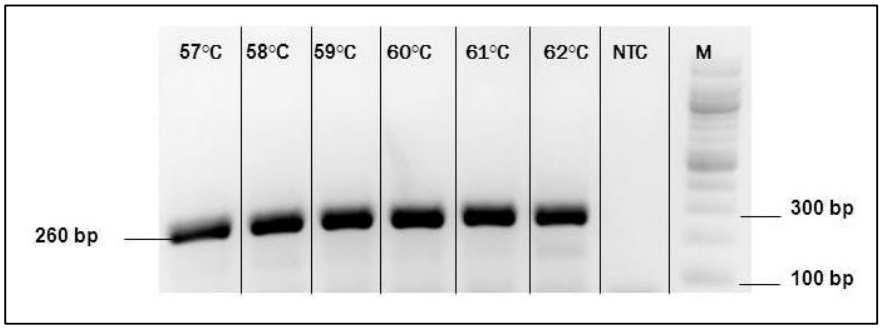


Figure 4.2: Gel electrophoresis image that depicts the results of the temperature gradient conducted in order to evaluate the wild-type primer set. M: Denotes 100 bp molecular weight marker used. NTC: non-template control. bp: base pairs.

4.3.2.2 Mutant-type primer set optimization

A temperature gradient was conducted to evaluate the mutant-type primer set to determine the optimal annealing temperature. The gel images in **Figure 4.3** below depict the results of the conventional RT-PCR temperature gradient that was performed using study control C2 and C3, respectively. C2 and C3 yielded PCR fragments of a 132 bp at 57°C to 62°C which was expected as these controls were positive for inv22. The optimal annealing temperature was determined to be 62°C, thus allowing the wild-type and mutant-type RT-PCR reactions to run concurrently at an annealing temperature of 62 °C. C1 yielded no PCR fragment which was also expected as C1 was negative for the inv22 as the individual was a non-haemophilic volunteer.

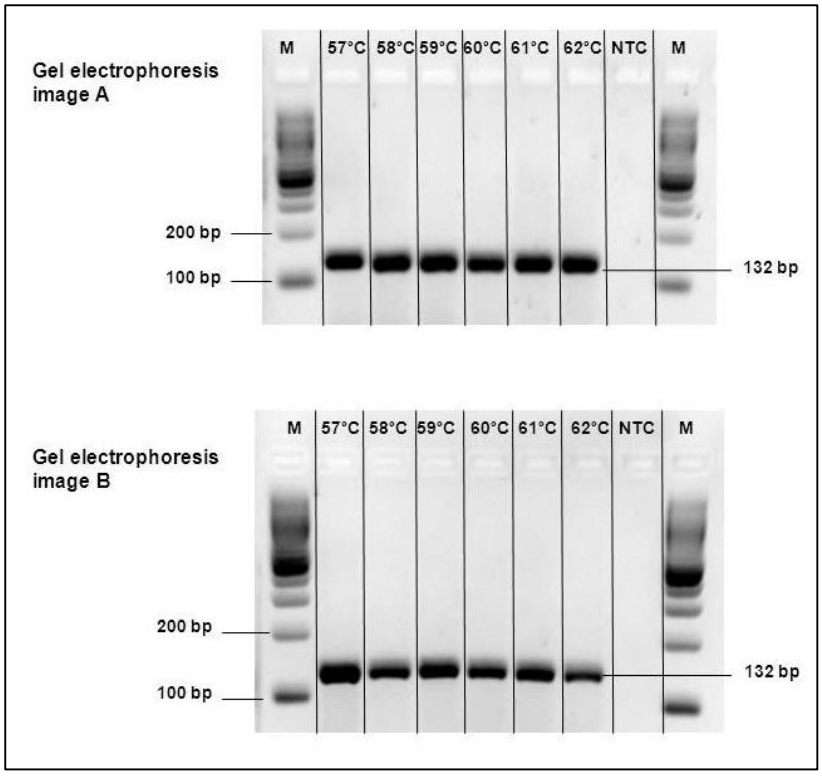


Figure 4.3: Gel electrophoresis image that depicts the results of the temperature gradient conducted in order to evaluate the mutant-type primer set. Gel electrophoresis image A: C2 (132 bp) and Gel electrophoresis image B: C3 (132 bp). M: Denotes 100 bp molecular weight marker used. NTC: non-template control. bp: base pairs.

4.3.2.3 Multiplex of the wild-type and the mutant-type primers

When the wild-type and the mutant-type primers were combined in a multiplex reaction, unfavourable results were found. The sizes of the PCR fragments that were obtained for C1 and C2 were not consistent with the expected PCR fragment sizes, namely 260 bp for C1 and 132 bp for C2, respectively. The results of the temperature gradient multiplex PCR are depicted in **Figure 4.4**. Further optimization attempts should be considered to optimise an inv22 multiplex PCR method, which was not in the scope of this study. It was thus decided to rather set up single PCR reactions, one containing the wild-type primer set and the other the mutant-type primer set, respectively.

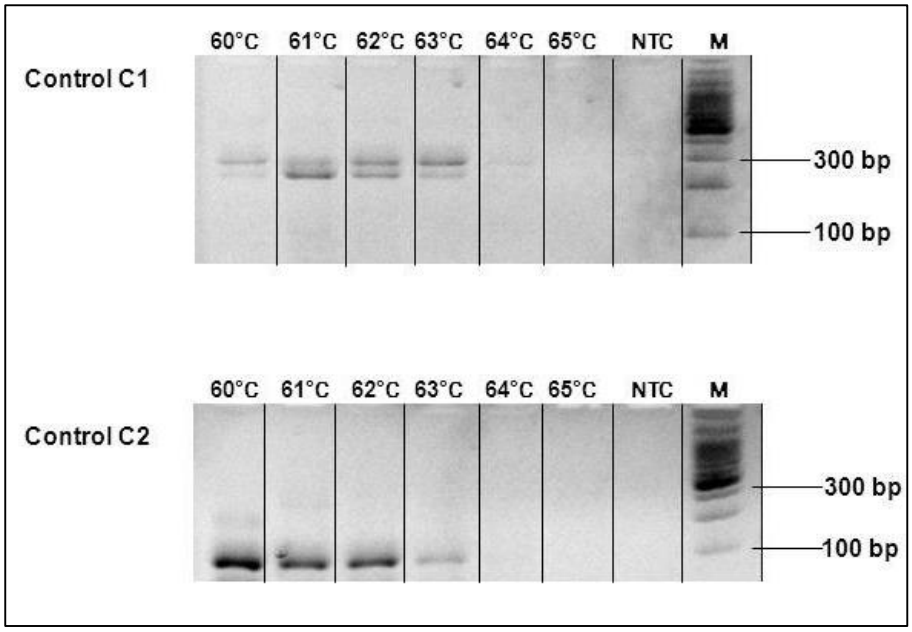


Figure 4.4: Gel electrophoresis images of the multiplex PCR for Control C1 and C2 at annealing temperatures of 60°C to 65°C. M: Denotes the 100 bp molecular weight marker that was used. NTC: non-template control. bp: base pairs.

4.3.2.4 Inv22 conventional RT-PCR results for study control samples C1, C2 and C3

C1 only yielded a PCR fragment of 260 bp, and conversely, C2 and C3 only yielded PCR fragments of 132 bp, respectively (**Figure 4.5**). As described by Naylor *et al.* (1993b), the *FVIII* gene is disrupted by the inv22, and this leads to no PCR amplification between the boundaries of exon 22 and exon 23 of the *FVIII* gene. Consequently, this study results are comparable to Naylor *et al.* (1993b) as C2 and C3, which are positive for the inv22, yielded no PCR fragments when screened with the wild-type primer set, indicating that exon 22 and exon 23 of the *FVIII* gene is indeed disrupted by the presence of the inv22. Conversely, a 132 bp PCR fragment was detected when C2 and C3 were screened with the mutant type primer set, and this confirms that in the case of the inv22 being present, exon 22 is spliced to an alternative exon. It is expected that the alternative exon is exon 23c as was predicted by Pandey *et al.* (2013). The association of exon 23c and the inv22, however, needs to be confirmed by sequence analysis.

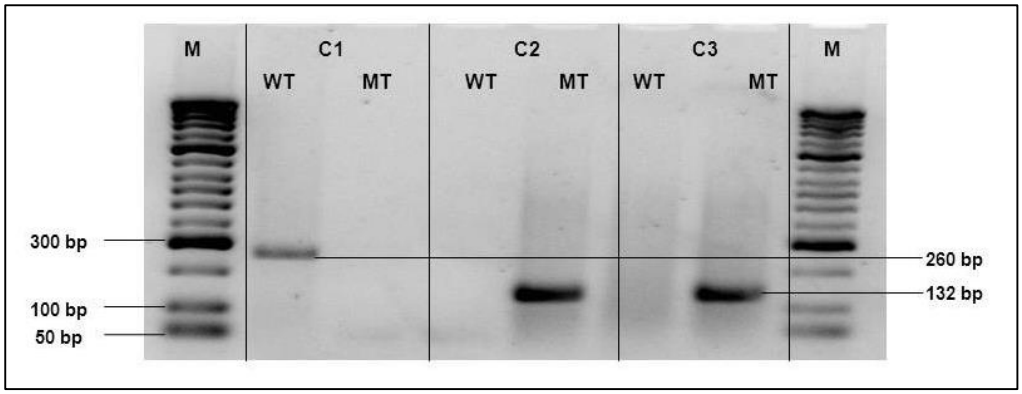


Figure 4.5: Gel electrophoresis image of control C1, C2 and C3, genotype using the inv22 conventional RT-PCR method. M: Denotes the molecular weight marker that was used. bp: base pairs.

4.3.2.5 Limit of detection (LOD) results for the inv22 conventional RT-PCR method

The cDNA of C2 was used to make a dilution range of 1000 ng/µL, 500 ng/µL, 100 ng/µL, 50 ng/µL, 10 ng/µL and 0 ng/µL (blank control). The different diluted samples of C2 were used to analyse the LOD of the inv22 conventional RT-PCR detection method's mutant reaction. The gel-image in **Figure 4.6** depicts the results that were obtained. The inv22 mutation could be detected at 1000 ng/µL, 500 ng/µL, 100 ng/µL, 50 ng/µL and 10 ng/µL. The PCR fragment obtained at a cDNA concentration of 10 ng/ µL was however slightly faint and consequently to obtain optimal results the preferred LOD of the inv22 conventional RT-PCR assay's mutant reaction was determined to be 50 ng/µL. Although the inv22 PCR fragment could be detected at each of the different cDNA concentrations, it would have been desirable to have repeat the assay in triplicate at each of the cDNA concentrations to determine the LOD. This is a limitation of the current study and should be investigated in a validation study should the assay be introduced as a diagnostic test for inv22.

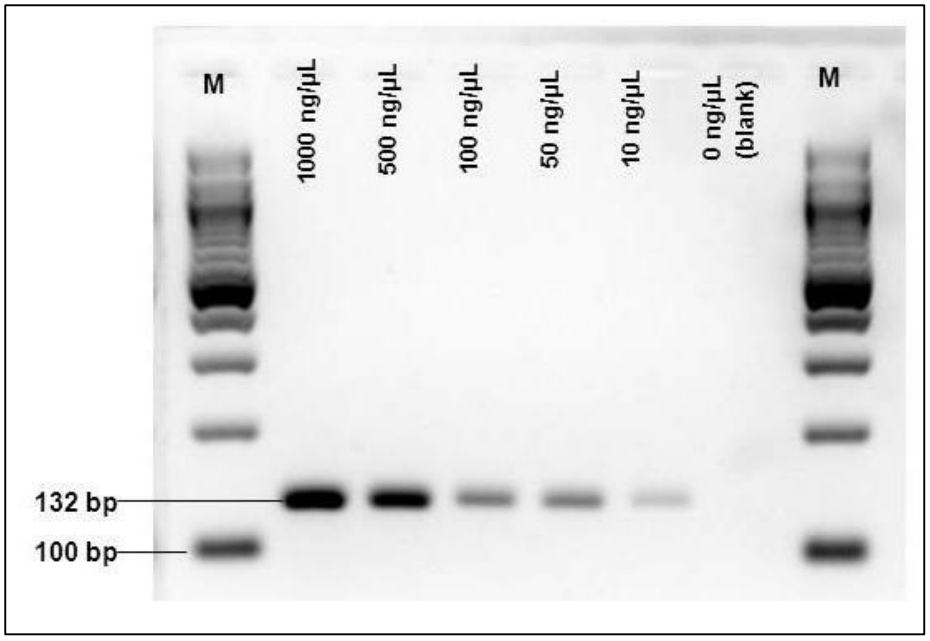


Figure 4.6: Gel electrophoresis image that depicts the LOD analysis conducted for the inv22 conventional RT-PCR detection method using C2 at different cDNA concentrations. M: Denotes the 100 bp molecular weight marker that was used. bp: base pairs.

4.3.3 The Optimization of the inv22 real-time RT-PCR method

Real-time PCR overcomes many of the limitations associated with conventional PCR as discussed in the literature review (Heid *et al.,* 1996; Bustin, 2005; Spurgeon *et al.,* 2008). It proves to be a sensitive method that can detect DNA/cDNA in low quantities (Yang *et al.,* 2005; Qiu *et al.,* 2013).

The inv22 real-time RT-PCR reactions were optimised with SYBR Green intercalating dye chemistry using the primer sets as applied in the conventional PCR (**Table 3.2**). Control C1 served as the inv22 negative control, while control C2 and C3 were used as the inv22 positive control. As the optimal annealing temperature for the wild-type and the mutant-type primer sets were pre-determined with the inv22 conventional RT-PCR detection method, the same annealing temperatures were applied for the real-time method.

The DNA or cDNA concentrations used in real-time PCR play an important role in the success of the method as recommended by the manufacturer of the detection

chemistry (Power SYBR® Green Master Mix, Applied Biosystems, USA). The manufacturer recommended that cDNA concentrations of between 100 ng to 1 pg must be used for optimal real-time PCR (Package insert, Power SYBR® Green Master Mix, Applied Biosystems, USA). No amplification curves were obtained for any of the study control sample reactions with cDNA concentrations of 10 ng/µL or 50 ng/µL, respectively. There was, however, optimal amplification when a cDNA concentration of 100 ng/µL was used.

The reactions were optimised using two different sets of primers in separate tubes. One set of primers to identify the genotype without the inv22 or FVIII wild-type and the other set to identify the inv22 or mutant type. The duplicate reactions containing the wild-type primer set and C1 template produced amplification curves (**Figure 4.7**), but no amplification was detected in reactions with the mutant-type primer set (results not shown). It was expected as C1 is inv22 negative. When C2 and C3 were evaluated with the mutant type primer set, amplification curves were detected (**Figure 4.8**), but not in the wild-type reactions (results not shown). It was once more expected as control C2 and C3 were inv22 positive as previously determined by Southern Blot analysis.

The melt-curve plots as seen in **Figure 4.9** only show a single melt-curve for each of the controls indicating an absence of primer dimer or multiple fragments. **Table 4.2** summarises the results for C1, C2 and C3. The non-template controls (NTC) were negative in the optimization reactions, and the passive ROX-dye that served as an internal control passed the quality control check. Real-time PCR can only be as reliable as the controls and accompanying quality assurance measurements and therefore as these controls passed the inv22 real-time RT-PCR detection method was deemed reliable (Bustin, 2005). In this study, an optimised real-time PCR method was therefore developed that is accurate and sensitive to detect the FVIII wild-type and inv22 genotypes, respectively. The method also consequently can detect the inv22 carrier genotype and can, therefore, be introduced as an inv22 detection method in laboratories that are equipped for real-time PCR.

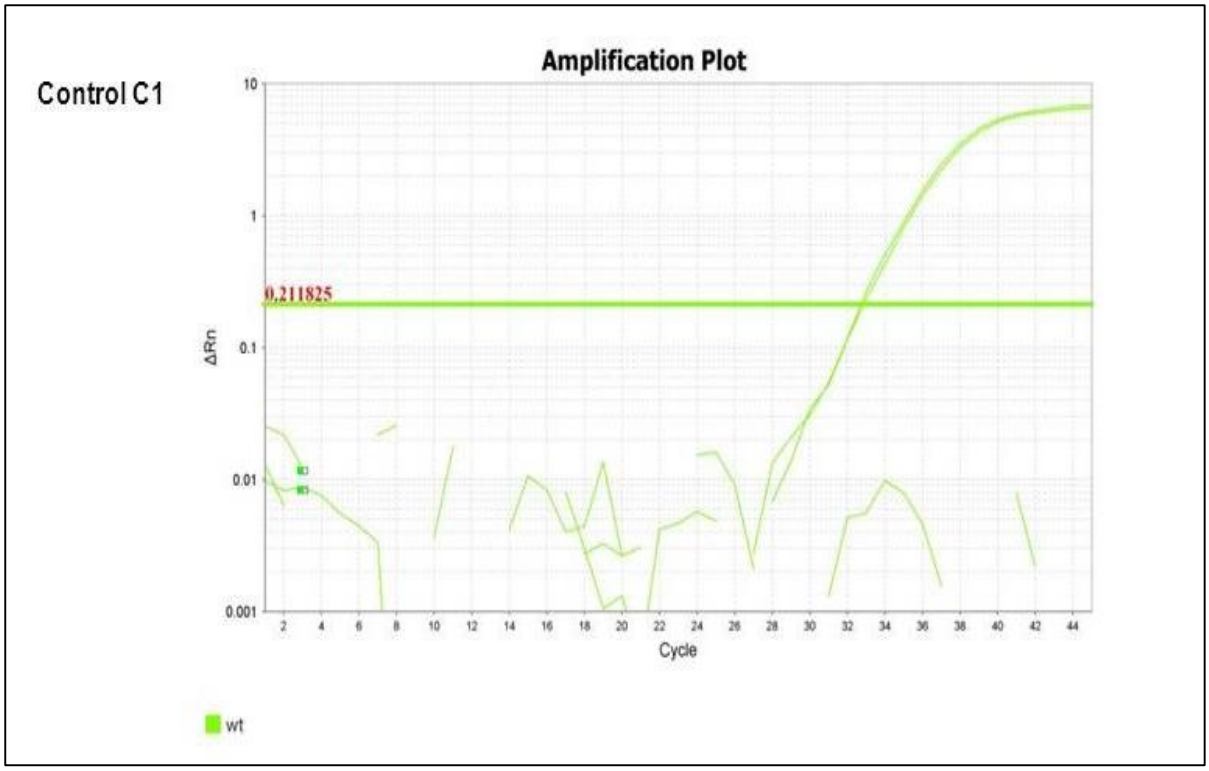


Figure 4.7: Real-time RT-PCR amplification plot for control C1. The plot represents the duplicate reaction with the wild-type primers. WT: Wild-type. ΔRn: Normalised fluorescent reported value minus the baseline signal.

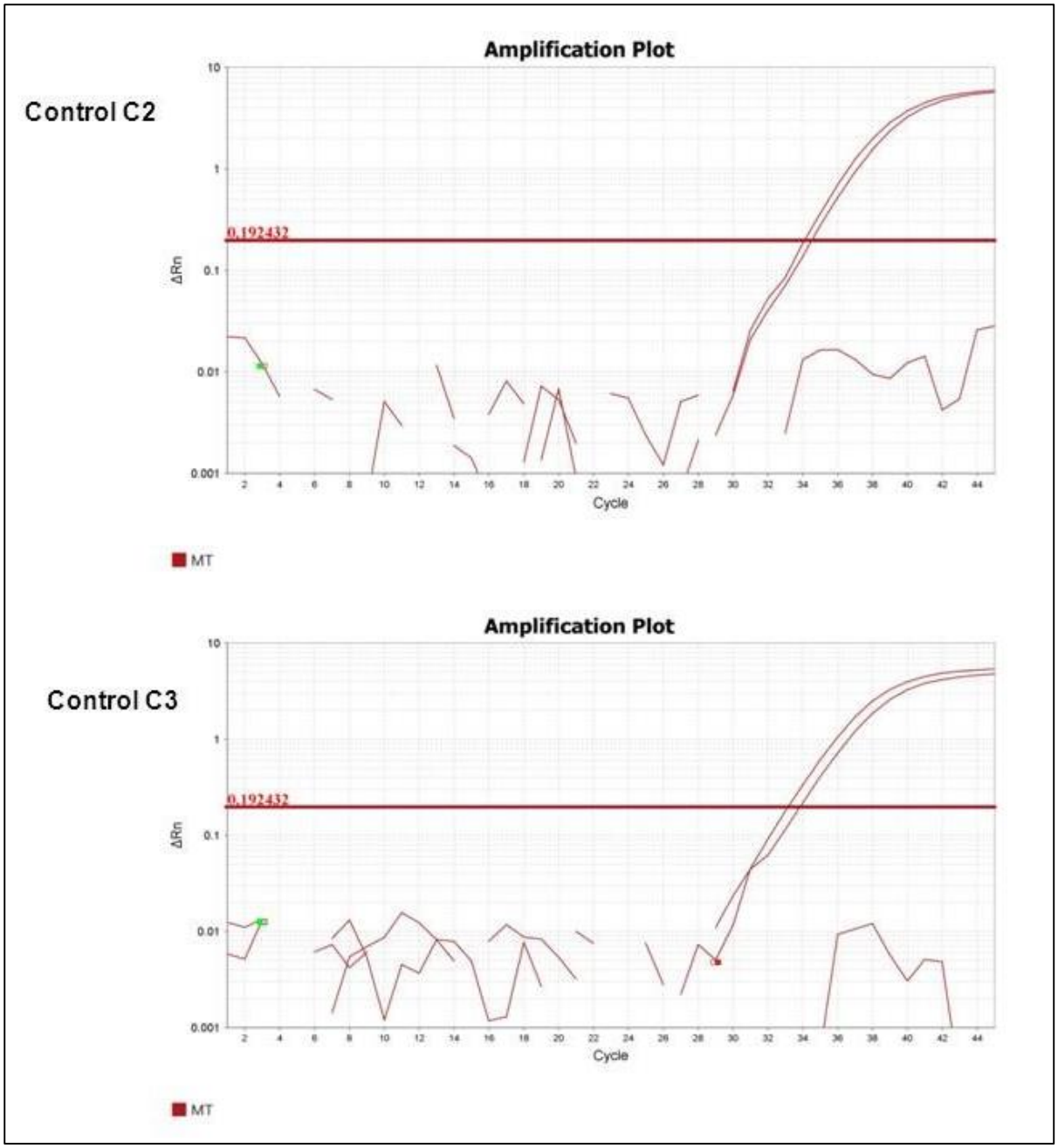


Figure 4.8: Real-time RT-PCR amplification plots of duplication reactions for control C2 and C3, respectively. The plot represents the duplicate reaction with the mutant-type primers. MT: Mutant-type. ΔRn: Normalised fluorescent reported value minus the baseline signal.

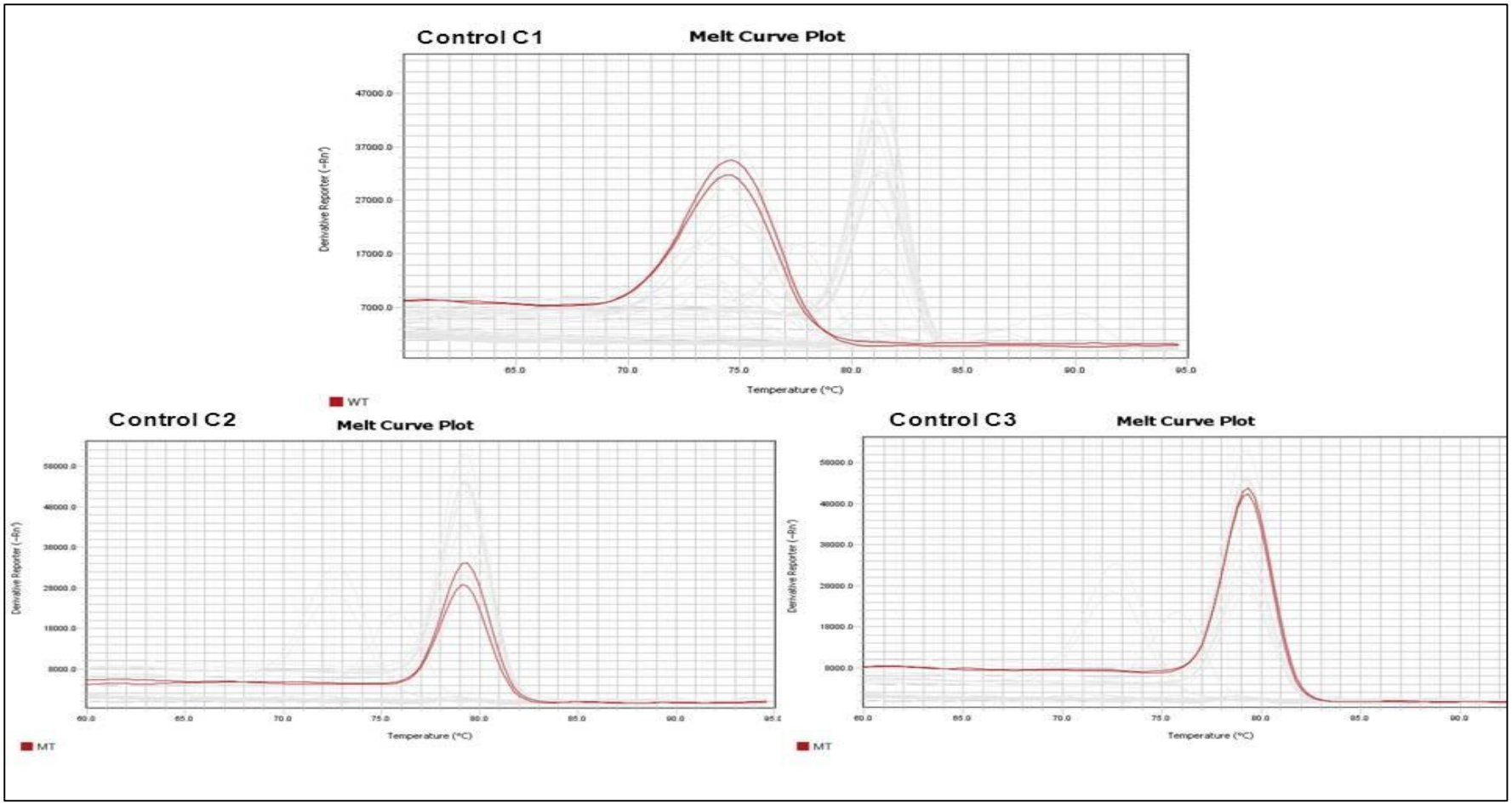


Figure 4.9: Derivative melt curve plots (a negative derivative of fluorescence with respect to temperature) in duplicate for controls C1, C2 and C3, respectively. WT: Wild-type. MT: Mutant-type.

Table 4.2 Real-time RT-PCR results for study control C1, C2 and C3. The table indicates target detection, mean Ct value, standard deviation (SD) and Inv22 status.

Sample	Wild-type target	Mutant type target	Mean Ct-value	SD	Inv22 status
C1	Detected	Not detected	32.924	0.045255	Negative
C2	Not detected	Detected	34.135	0.176777	Positive
C3	Not detected	Detected	33.792	0.313955	Positive

4.4 Confirmation of study controls C1, C2 and C3 with Sanger sequence analysis

The inv22 status of the study controls has been determined with the Southern Blot detection method (**Appendix A**). The study controls were analysed with Sanger sequencing to confirm the presence or the absence of the inv22. Sanger sequencing confirmed C1 as inv22 negative and C2 and C3 as inv22 positive, respectively (**Figure 4.10, Figure 4.11, Figure 4.12**). The Sanger sequencing results that were obtained in this study for C2 and C3 were in concordance with the results that were obtained previously with Southern Blot analysis. Sequence analysis revealed that C1 contained exon 22 spliced to exon 23 and aligned with the wild-type mRNA reference sequence (NM_000132.3) (**Figure 4.13**). C2 and C3 aligned to the mutant mRNA reference sequence and contained exon 22 spliced to exon 23c as predicted by Pandey *et al.* (2013) (**Figure 4.14 and Figure 4.15**). In the study by Pandey *et al.* (2013), the amino acid sequence of exon 23c was predicted and confirmed by the sequencing data of this study (**Appendix F**).

C2 and C3 are severe haemophilia A patients not related to each other, which both tested positive for the inv22. Sequence analysis revealed that these controls had identical sequences, specifically referring to exon 22 spliced to exon 23c (**Figure 4.16**). Consequently, it can be hypothesised that all severe haemophilia A patients,

positive for inv22, will have exon 22 spliced to the alternative exon 23c. It will, however, need to be confirmed in other severe haemophilia A populations.

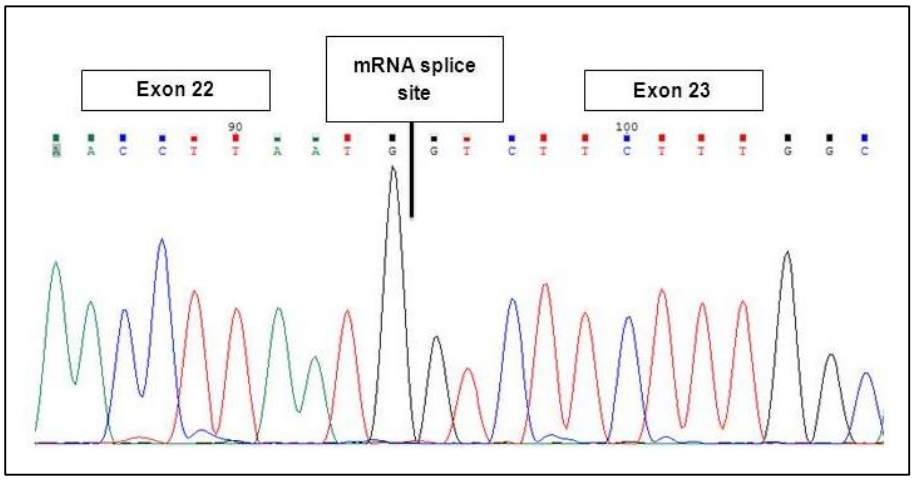


Figure 4.10: A sequence electropherogram depicting the mRNA sequence of control C1 at position 6591-6612 (NM_000132.3). The figure illustrates the FVIII exon 22 spliced to FVIII exon 23.

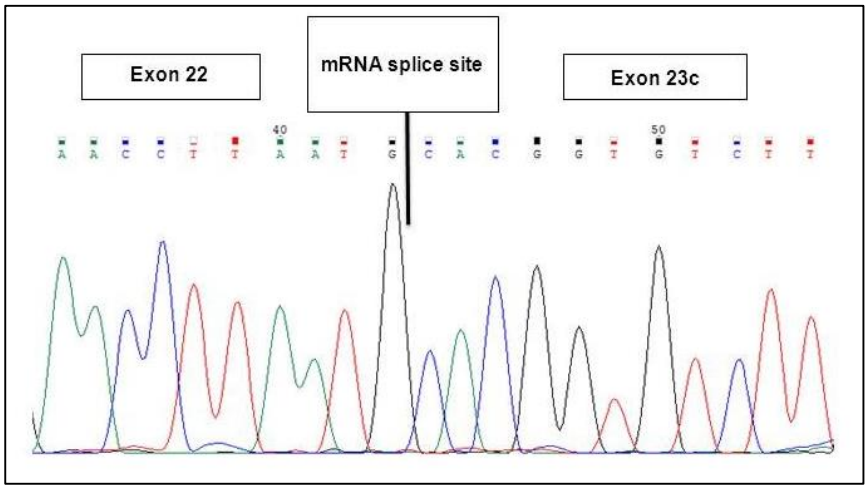


Figure 4.11: A sequence electropherogram depicting the mRNA sequence of control C2 at position 6591 (FVIII exon 22, NM_000132.3) and exon 23c (position 1918 to 1930 XR_001731635.1.

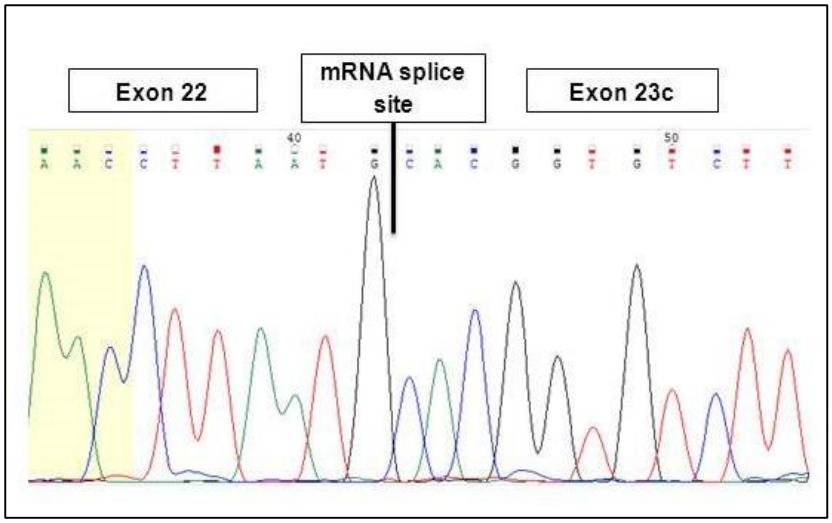


Figure 4.12: A sequence electropherogram depicting the sequence of control C3 at position 6591 (FVIII exon 22, NM_000132.3) and exon 23c (position 1918 to 1930 XR_001731635.1)

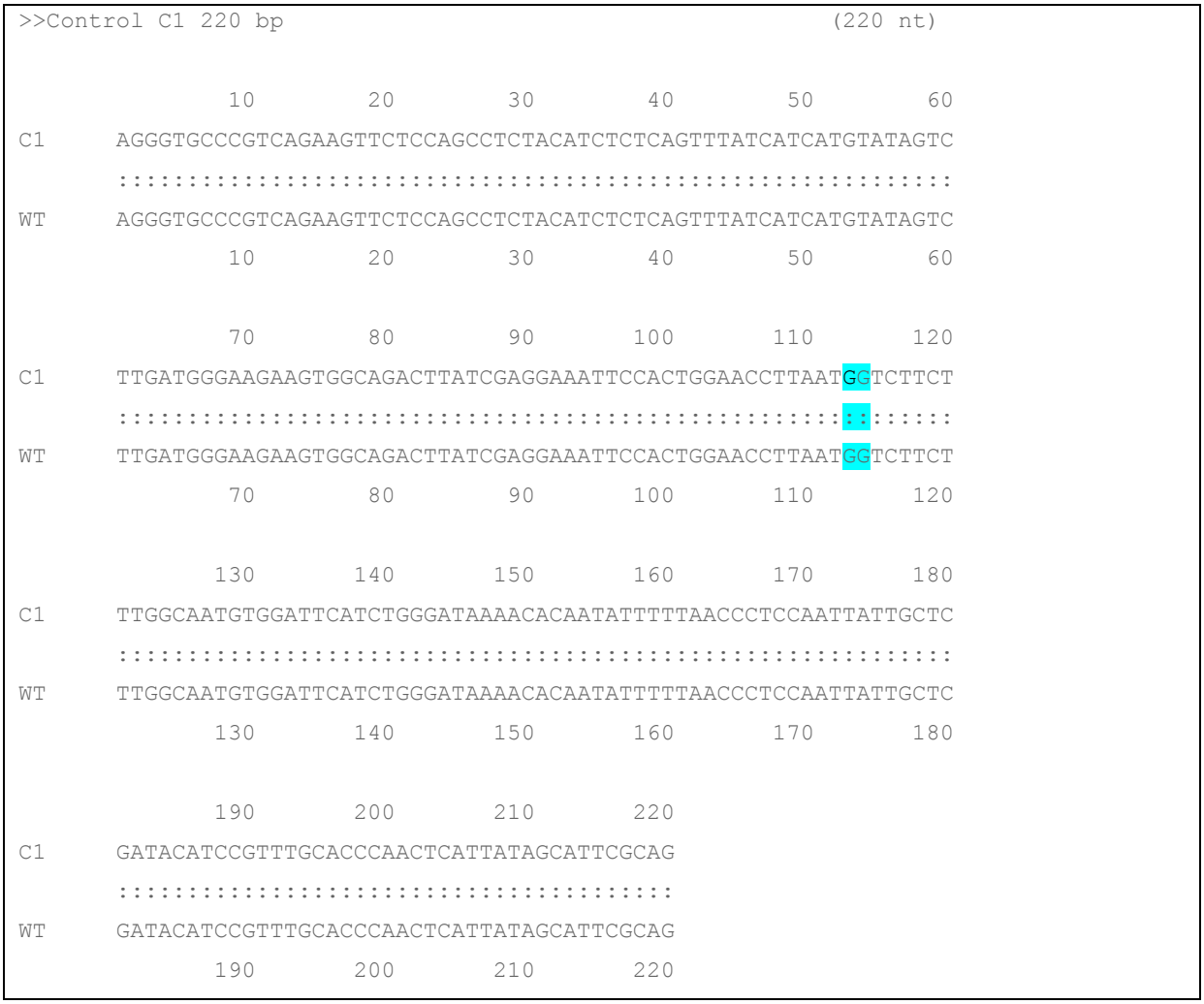


Figure 4.13: Sequence analysis result that reveals control C1 contained the FVIII exon 22 spliced to the FVIII exon 23, which aligned to the FVIII wild-type mRNA reference sequence (NM_000132.3). C1: Control C1. WT: Wild-type. The splice site is highlighted in blue.

```
>>Control C2 85 bp                                        (85 nt)

               10        20        30        40        50        60
C2     ATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Mutant ATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACC
               10        20        30        40        50        60

               70        80
C2     TTAATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::::
Mutant TTAATGCACGGTGTCTTGGAAAATG
               70        80
```

Figure 4.14: Sequence analysis result that reveals control C2 contained the FVIII exon 22 spliced to the alternative exon 23c, which aligned to mRNA mutant reference sequence (Appendix F). C2: Control C2. MT: Mutant-type. bp: base pairs. nt: nucleotides. The splice site is highlighted in blue.

```
>>Control C3 85 bp                                        (85 nt)

               10        20        30        40        50        60
C3     ATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Mutant ATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACC
               10        20        30        40        50        60

               70        80
C3     TTAATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::::
Mutant TTAATGCACGGTGTCTTGGAAAATG
```

Figure 4.15: Sequence analysis result that reveals control C3 contained the FVIII exon 22 spliced to the alternative exon 23c, which aligned to mRNA mutant reference sequence (Appendix F). C2: Control C2. MT: Mutant-type. bp: base pairs. nt: nucleotides. The splice site is highlighted in blue.

```
>>C2 41 bp vs C3 41 bp                                          (41 nt)


               10        20        30        40
C2     CACGGTGTCTTGGAAAATGGAAAGAATTGGGAGCCTTCCTA
       ::::::::::::::::::::::::::::::::::::::::
C3     CACGGTGTCTTGGAAAATGGAAAGAATTGGGAGCCTTCCTA
               10        20        30        40
```

Figure 4.16: Comparison of the exon 23c sequence in controls C2 and C3, respectively. C2: Control C2. C3: Control C3. bp: base pairs. nt: nucleotides.

4.5 Screening the 60 study participants for the inv22 using the conventional and the real-time RT-PCR detection methods

4.5.1 The inv22 conventional RT-PCR results (Appendix H)

The inv22 conventional RT-PCR detection method proved to be accurate and reliable, especially in a laboratory that is under-resourced and does not have the means or expertise to acquire a real-time PCR instrument (Kloppers & Janse van Rensburg, 2017).

At the onset of the study, the study participants had an unknown inv22 status. It was found that of the 60 study participants screened with the inv22 conventional RT-PCR detection method, eight were inv22 positive, four were inv22 carriers and the rest were negative for the inv22, respectively (**Appendix H**). The results from the inv22 negative participants were comparable to that of C1 – a single 260 bp fragment. PCR results from the non-haemophilia volunteers (P54-P60) were negative for the inv22, as expected. The results from the inv22 positive participants were comparable to that of C2 and C3 (known inv22 genotypes) – a single 132 bp fragment. The study participants that were inv22 carriers (inv22 heterozygous) had both the 260 bp fragment and the 132 bp fragment, respectively.

4.5.2 The inv22 real-time RT-PCR results (Appendix I).

The results of the 60 study participant samples that were screened for inv22 indicated that eight were inv22 positive, three were inv22 carriers, and the rest were inv22 negative. Unfortunately, no results could be obtained for sample P14 with the inv22 real-time RT-PCR detection method after several attempts. The stored RNA of sample P14 was used to synthesise cDNA again. However, the results remained inconclusive. It could be attributed to sample-specific degradation, and unfortunately, the study participant could not be reached to obtain another blood sample. Nevertheless, the sample results obtained from the inv22 real-time RT-PCR detection method was found to be accurate, as the study participants that were inv22 negative compared well with C1. Similarly, the inv22 positive study participants compared well to the C2 and C3 (known inv22 genotypes) that were included in each of the real-time RT-PCR runs for quality control purposes.

In **Figures 4.17-4.19,** examples of the amplification curves of some of the study participants that were screened with the real-time method are displayed. Each figure presents two sets of amplification curves, firstly for the wild-type reaction and secondly for the mutant-type reaction.

Figure 4.17 displays an example of the inv22 negative C1 control sample (**Light blue amplification curves**) with one of the study participants, sample P8, that were inv22 negative (**Purple amplification curves**) in the wild-type primer set reaction. The same study participant had no amplification curve in the mutant-type primer set reaction, and only amplification for inv22 positive C2 control (**Light blue amplification curves**) can be observed.

In **Figure 4.18**, only the amplification curves for inv22 negative control sample C1 (**Yellow amplification curves**) can be observed in the wild-type primer set reaction. In the mutant-type primer set reaction amplification curves can be observed for inv22 positive control sample C2 (**Light blue amplification curves**) and study participant (P32) that was inv22 positive (**Pink amplification curves**).

Figure 4.19 displays the amplification curves for one of the inv22 carriers. In the wild-type primer set reaction amplification curves for inv22 negative control sample C1 (**Light blue amplification curves**) and the carrier (P53) (**Green amplification curves**) can be observed. In the mutant-type primer set reaction, amplification curves

can be observed for inv22 positive control sample C2 (**Red amplification curves**) and the carrier (P53) (**Maroon amplification curves**). The NTC's were negative in all the above reactions.

All the study participant samples that were inv22 negative, produced amplification curves in the wild-type primer set reactions. Conversely, study participant samples that were inv22 positive only showed amplification in the mutant-type primer set reactions. Carrier samples for inv22 produced amplification curves in the wild-type, and the mutant-type primer set reactions. Furthermore, the control samples (C1, C2 and C3) with known inv22 genotypes, as confirmed by Southern Blot and sequence analysis, were included with each reaction run to serve as controls for unknown participants samples. The inv22 genotype status was determined by matching the real-time amplification results from the unknown samples to that of the controls.

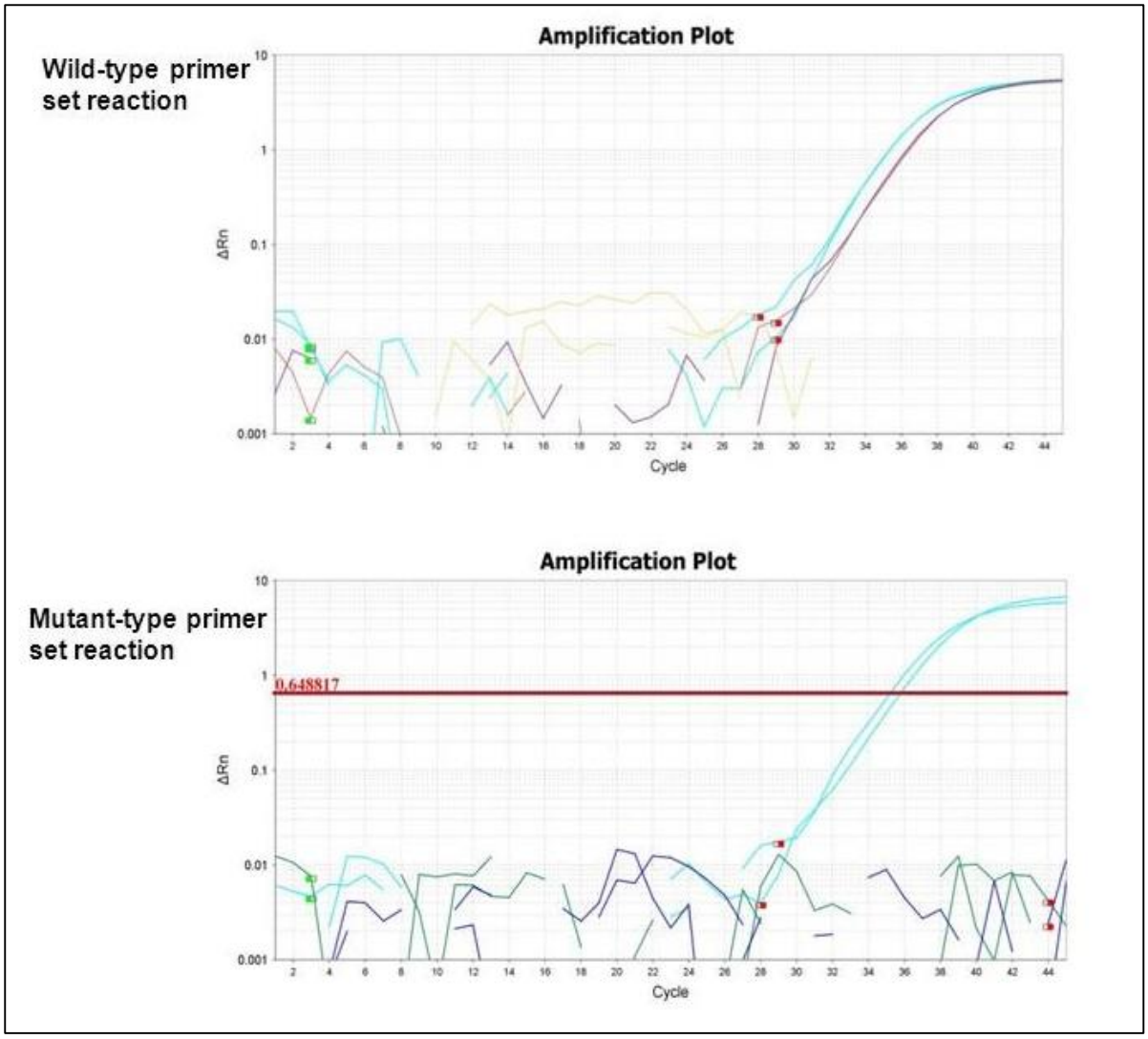


Figure 4.17: An example of amplification plots for study participant P8, an inv22 negative sample using real-time RT-PCR. Wild-type primer set reaction: Control C1 (Light blue amplification curves) and study participant P8 (Purple amplification curves). Mutant-type primer set reaction: Control C2 (Light blue amplification curves) and study participant P8 with no amplification curve (Purple).

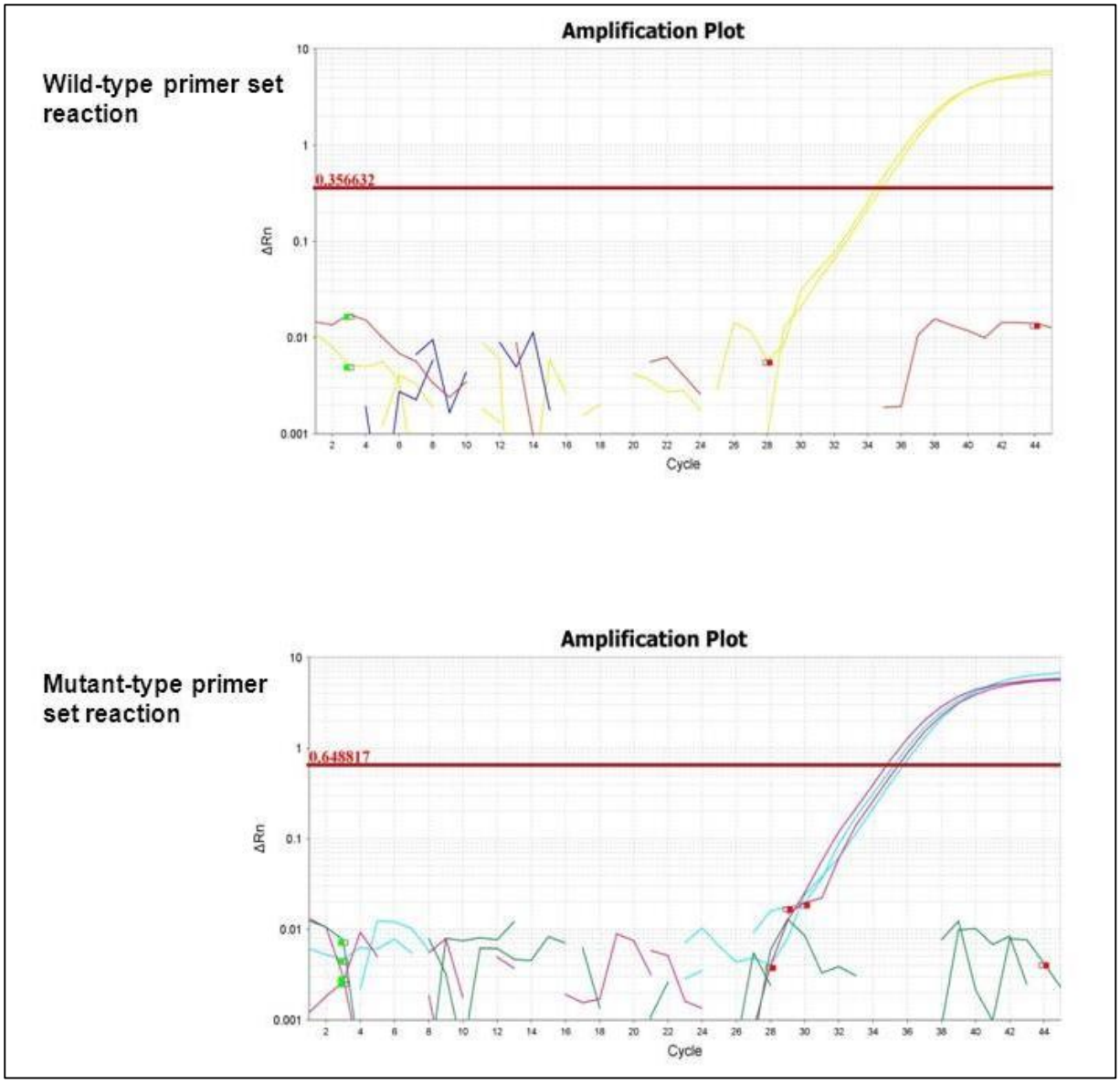


Figure 4.18: An example of amplification plots for study participant P32, an inv22 positive sample using real-time RT-PCR. Wild-type primer set reaction: Control C1 (Yellow amplification curves) and no amplification curve for study participant P32 (Purple). Mutant-type primer set reaction: Control C2 (Light blue amplification curves) and study participant P32 (Pink amplification curves).

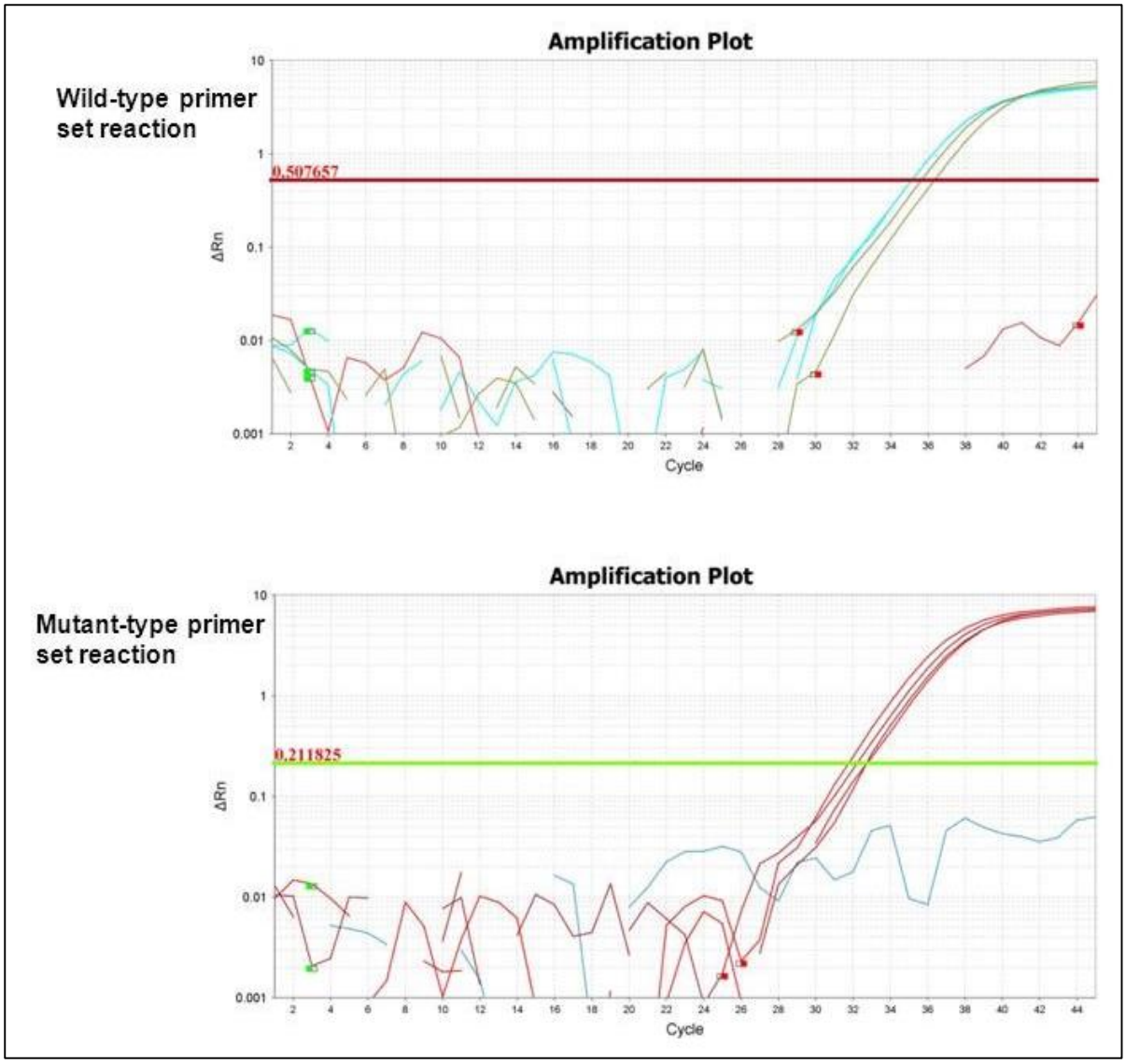


Figure 4.19: An example of amplification plots for study participant P53, an inv22 carrier sample using real-time RT-PCR. Wild-type primer set reaction: Control C1 (Light blue amplification curves) and study participant P53 (Green amplification curves). Mutant-type primer set reaction: Control C2 (Red amplification curves) and study participant P53 (Maroon amplification curves).

The Ct-values obtained in the real-time PCR assay were subsequently high, above 30, as was found in a similar real-time PCR study by Kumar *et al.* (2013). As the cDNA concentrations for each sample was standardised in this study, the Ct-values could not be used to detect the input amount of mRNA and subsequently the higher Ct-values obtained with the real-time PCR assay in this study were not significant.

The developed real-time assay was not used for FVIII quantification purposes, but only to detect the presence or absence of the inv22 mutation. Although FVIII is mainly synthesised in the liver and endothelial cells and not in lymphocytes (Hoffbrand & Moss, 2011a), adequate amounts of FVIII mRNA were present in lymphocytes to warrant the use of the real-time assay to detect inv22. The inv22 real-time RT-PCR method proved to be accurate and sensitive by matching all genotype control samples. Subsequently, the assay has been validated and can be introduced as a standard inv22 detection method in laboratories that are equipped for real-time PCR that relies on methods with a high-throughput.

4.5.3 Confirmation of the inv22 results for the 60 study participants samples with Sanger sequencing (Appendix J)

Sanger sequence analysis was conducted on all the study samples confirming the results obtained with the inv22 conventional and real-time RT-PCR detection method. The sequences of the study samples that resulted in an inv22 negative result aligned with the wild-type mRNA reference sequence (NM_000132.3) that contained exon 22 and exon 23 of the *FVIII* gene. Consequently, the full length ($FVIII_{FL}$) FVIII transcript can be produced and these individuals have functional FVIII proteins (Pandey *et al.,* 2013; Sauna *et al.,* 2015). The study samples that resulted in an inv22 positive result produced sequences that aligned with the mutant mRNA reference sequence (**Appendix F**) and contained exon 22 of the *FVIII* gene, spliced to the alternative exon 23c. Likewise, the samples that resulted in inv22 carriers showed sequences that aligned to the mutant mRNA reference sequence, as well as to the FVIII wild-type mRNA reference sequence.

Pandey *et al.* (2013) predicted in their study that the inv22 transcript would contain 16 amino acids that were derived from the alternative exon 23c. This prediction was confirmed in this study, and it can be concluded that the genotypes of all haemophilia A patients and carriers of the inv22 will contain exon 22 of the *FVIII* gene, spliced to the alternative exon 23c. Furthermore, exon 23c was found to have the same sequence in all cases that are inv22 positive. Consequently, the presence of the FVIII wild-type, inv22 and inv22 carrier genotypes have been confirmed with Sanger sequence analysis for each of the 60 study participants, respectively. This finding allows for the int22 detection methods developed in this study to be used in any

laboratory in the world. The principle on which the method functions applies to any haemophilia A population with inv22, as the alternative exon 23c was found in this study population and the study population of Pandey *et al.* (2013).

4.6 Comparison of the inv22 results between the different inv22 detection methods developed in this study

The results of the respective methods were tabulated and compared. It was found that C1 was inv22 negative with conventional, real-time RT-PCR and Sanger sequence analysis. C2 and C3 were found to be inv22 positive with the conventional, real-time RT-PCR method and Sanger sequence analysis, respectively. Furthermore, C2 and C3 were previously screened positive for the inv22 using the Southern Blot analysis method, which is currently one of the main screening methods used to detect the inv22 (Lakich *et al.,* 1993). Therefore, the inv22 positive controls (C2 and C3) showed a 100% concordance when the Southern Blot, conventional RT-PCR, real-time RT-PCR and sequence analysis results were compared. The comparison between the respective methods is tabulated in **Table 4.3**.

The methods all compared well within the study sample group of unknown genotypes, except for one study participant (P14), for which the real-time RT-PCR method results were inconclusive (**Table 4.4**). It can, therefore, be concluded that the inv22 results obtained with the respective detection methods in this study are comparable regarding accuracy.

Table 4.3: Inv22 results comparison of the study controls with Southern Blot analysis, conventional RT-PCR, real-time RT-PCR and sequence analysis methods.

Control	Southern Blot	Conventional RT-PCR	Real-time RT-PCR	Sequence analysis
C1	n/a	Negative	Negative	Negative
C2	Positive	Positive	Positive	Positive
C3	Positive	Positive	Positive	Positive

Table 4.4 Inv22 results comparison for the study participants with conventional RT-PCR, real-time RT-PCR and sequence analysis methods.

Study participants	Haemophilia A severity	Conventional RT-PCR	Real-time RT-PCR	Sequence analysis
P1	Severe	Negative	Negative	Negative
P2	Possible carrier	Negative	Negative	Negative
P3	Severe	Positive	Positive	Positive
P4	Possible carrier	Negative	Negative	Negative
P5	Severe	Positive	Positive	Positive
P6	Possible carrier	Negative	Negative	Negative
P7	Severe	Negative	Negative	Negative
P8	Severe	Negative	Negative	Negative
P9	Severe	Negative	Negative	Negative
P10	Mild	Negative	Negative	Negative
P11	Severe	Negative	Negative	Negative

P12	Severe	Negative	Negative	Negative
P13	Severe	Negative	Negative	Negative
P14	Possible carrier	Carrier	Unknown	Carrier
P15	Severe	Negative	Negative	Negative
P16	Severe	Negative	Negative	Negative
P17	Possible carrier	Negative	Negative	Negative
P18	Severe	Negative	Negative	Negative
P19	Severe	Positive	Positive	Positive
P20	Possible carrier	Carrier	Carrier	Carrier
P21	Severe	Positive	Positive	Positive
P22	Severe	Positive	Positive	Positive
P23	Severe	Positive	Positive	Positive
P24	Possible carrier	Negative	Negative	Negative
P25	Severe	Negative	Negative	Negative
P26	Possible carrier	Negative	Negative	Negative
P27	Severe	Negative	Negative	Negative
P28	Possible carrier	Negative	Negative	Negative
P29	Severe	Negative	Negative	Negative
P30	Severe	Positive	Positive	Positive
P31	Possible carrier	Negative	Negative	Negative
P32	Severe	Positive	Positive	Positive
P33	Possible carrier	Negative	Negative	Negative
P34	Severe	Negative	Negative	Negative

P35	Possible carrier	Carrier	Carrier	Carrier
P36	Severe	Negative	Negative	Negative
P37	Severe	Negative	Negative	Negative
P38	Possible carrier	Negative	Negative	Negative
P39	Severe	Negative	Negative	Negative
P40	Severe	Negative	Negative	Negative
P41	Severe	Negative	Negative	Negative
P42	Severe	Negative	Negative	Negative
P43	Possible carrier	Negative	Negative	Negative
P44	Possible carrier	Negative	Negative	Negative
P45	Possible carrier	Negative	Negative	Negative
P46	Severe	Negative	Negative	Negative
P47	Possible carrier	Negative	Negative	Negative
P48	Severe	Negative	Negative	Negative
P49	Mild	Negative	Negative	Negative
P50	Severe	Negative	Negative	Negative
P51	Severe	Negative	Negative	Negative
P52	Mild	Negative	Negative	Negative
P53	Possible carrier	Carrier	Carrier	Carrier
P54	Non-haemophilic volunteer	Negative	Negative	Negative
P55	Non-haemophilic volunteer	Negative	Negative	Negative

P56	Non-haemophilic volunteer	Negative	Negative	Negative
P57	Non-haemophilic volunteer	Negative	Negative	Negative
P58	Non-haemophilic volunteer	Negative	Negative	Negative
P59	Non-haemophilic volunteer	Negative	Negative	Negative
P60	Non-haemophilic volunteer	Negative	Negative	Negative

4.7 The presence of the inv22 mutation in the central South African haemophilia A population

The inv22 genotypes of the central South African haemophilia A population is mostly unknown. It was confirmed when study control samples were being recruited for this study at the Haematology clinics at Universitas Academic Hospital and Kimberley Hospital Complex, respectively. There were only two haemophilia A patients with known inv22 genotypes that could be recruited at these respective clinics. The remainder of the haemophilia A patients had an unknown inv22 status. It was, therefore, paramount to determine the inv22 genotypes in this specific haemophilia A and possible carrier population. Consequently, it was found that in total the inv22 was present in 14 of the total study population (study participants and study controls), of which ten of the haemophilia A patients carried a homozygous genotype for inv22. Four of the study participants with no haemophilia A symptoms were inv22 carriers and thus heterozygous for the mutation.

The haemophilia A study participants with the inv22 homozygous genotype suffer from severe haemophilia A. This result corresponds to most of the inv22 studies worldwide (Lakich *et al.,* 1993; Naylor *et al.,* 1993a; Antonarakis *et al.,* 1995; Johnsen *et al.,* 2017). The study population included three mild haemophilia A participants that tested negative for the inv22 and consequently no definitive assumption could be made regarding the presence of the inv22 in mild and moderate haemophilia A patients. **Figure 4.20** below displays the distribution of the results that were found in this study, with the majority of the haemophilia A study participants (including control C1) negative for inv22 genotype. In **Figure 4.21**, the distribution of inv22 carriers in the central South African haemophilia A population is displayed.

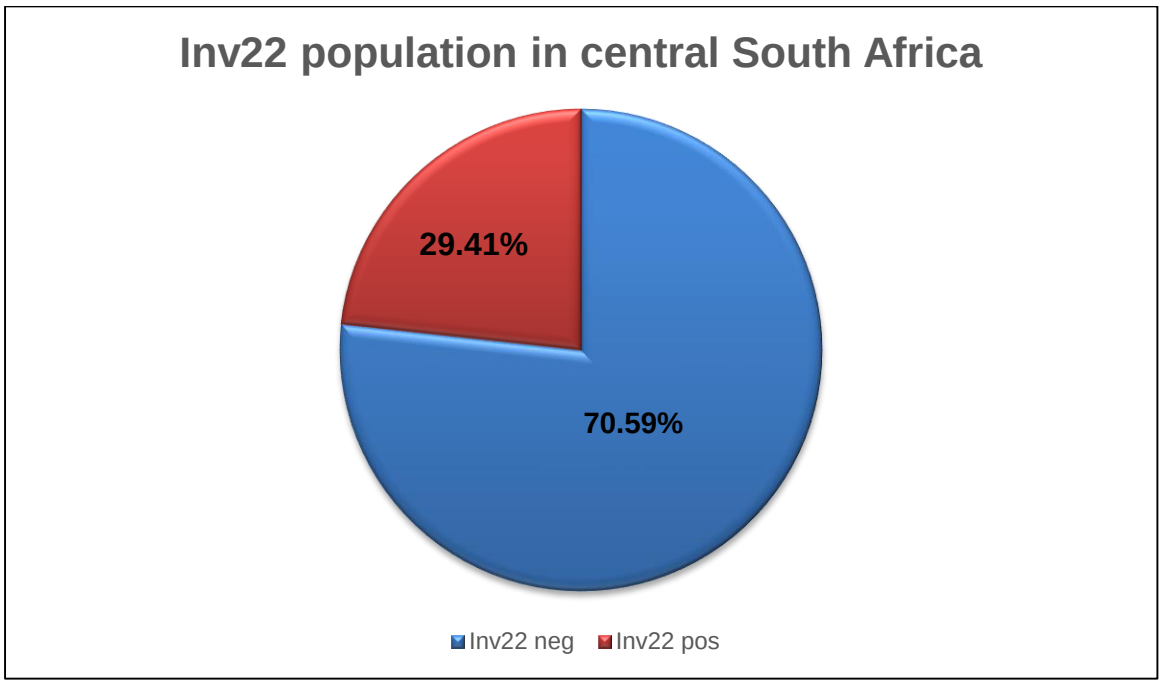


Figure 4.20: Pie chart that displays the distribution of the inv22 results in the central South African haemophilia A population.

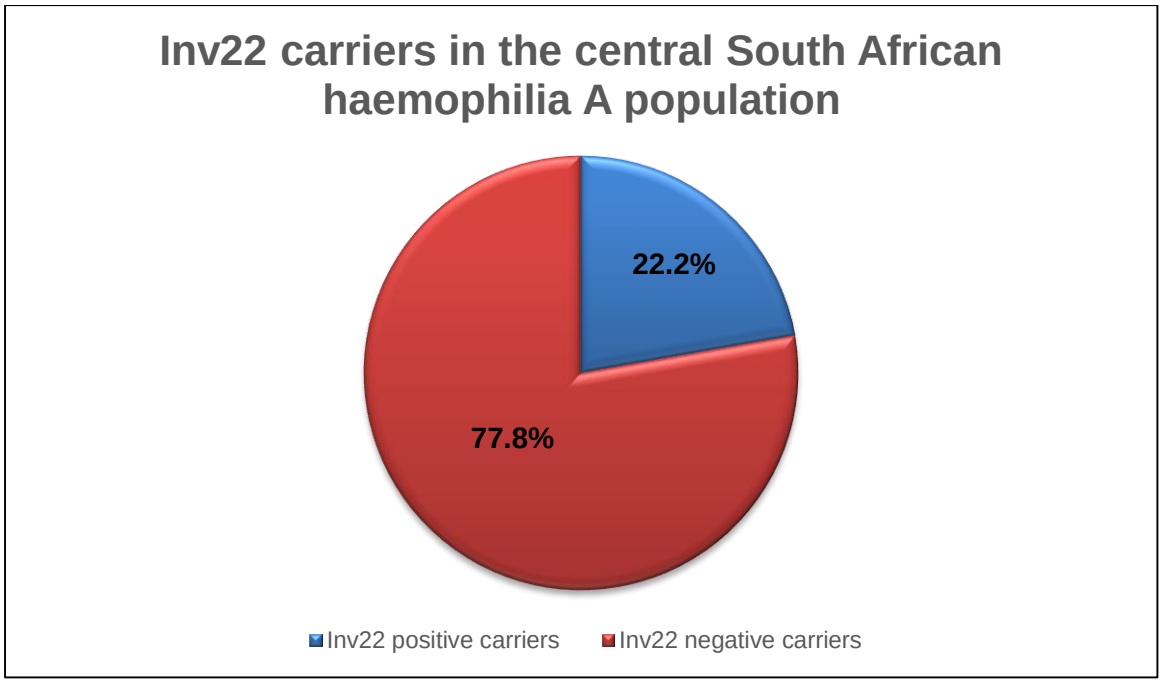


Figure 4.21: Pie chart that displays the distribution of the inv22 carrier results in the central South African haemophilia A population.

The inv22 was found to be present in 29.41% of severe haemophilia A patients (including the two unrelated controls) (**Figure 4.20**) and 22.2% of possible carriers in this study (**Figure 4.21**). In comparison to global studies that reported the inv22 in 45% of all severe haemophilia A cases, the 29.41% found in this study is lower (Lakich *et al.,* 1993; Naylor *et al.,* 1993a; Antonarakis *et al.,* 1995; Johnsen *et al.,* 2017). The lower presence of inv22 in this study could be attributed to the relatively small size of the study population and to confirm this finding a larger study cohort of haemophilia A patients will need to be investigated.

4.7.1 The presence of FVIII inhibitors in central South African haemophilia A patients that are positive for the inv22

It has been reported that FVIII mutations are associated with the development of FVIII inhibitors in patients with haemophilia A (Jayandharan *et al.,* 2012; Eckhardt *et al.,* 2013). FVIII inhibitors occur in 30% of severe haemophilia A patients and as discussed has a tremendous impact on treatment outcomes as FVIII inhibitors form antibodies

against the replacement FVIII therapy (Astermark *et al.,* 2005). Bardi and Astermark (2015) have reported that the inv22 is associated with FVIII inhibitor development in 20% of cases. However, cases of up to 40% have also been reported (Astermark *et al.,* 2005). The FVIII inhibitors history of each of the haemophilia A participants in this study were collected from patient records (**Appendix K**). There were nine out of the 37 haemophilia A patients (35 study participants and two controls) who had a history of FVIII inhibitor development and three of the haemophilia A patients positive for the inv22 were positive for FVIII inhibitors. The size of the study population in this study does not allow for any association to be made with regards to the presence of the inv22 and FVIII inhibitor development. It can, nonetheless, be concluded that there were some of the patient's positive for the inv22 and FVIII inhibitors, respectively. It has been reported that African haemophilia A patients have an increased risk of inhibitor development (Witmer &Young, 2013). Considering that the central South African haemophilia A population mostly consist of self-proclaimed black Africans, 67.7% in this study, it would be worth investigating the co-incidence of inv22 and FVIII inhibitor development in these patients in future studies.

The introduction of the newly developed inv22 detection methods will make screening of the mutation more frequent. Accordingly, association studies can then be conducted to determine whether there is an association between the inv22 and FVIII inhibitor development in the South African haemophilia A population.

4.8 Recommendations and limitations

4.8.1 Study population

The size of the study population was a limiting factor of the study. A larger study population would have allowed a more accurate comparison of the inv22 status in the South African haemophilia A population to that of the haemophilia A populations around the world. Secondly, a larger cohort would allow association studies to be conducted between the presence of the inv22 and FVIII inhibitor development in haemophilia A patients. Considering that haemophilia A is underdiagnosed in South Africa and that there is a single accredited genetic research centre available to screen for FVIII mutations (Mahlangu & Medical and Scientific Council of the South African Haemophilia Federation, 2009), the study population recruited in the study was

sufficient. It allowed for new inv22 detection methods to be developed which can consequently lead to the wide-spread screening of FVIII mutations associated with haemophilia A.

4.8.2 Inv22 conventional RT-PCR detection method

A successful inv22 conventional RT-PCR detection method was additionally developed to the real-time RT-PCR method. We recommend that the conventional method is used in diagnostic and research laboratories that are not equipped for real-time PCR or that do not have the expertise to conduct real-time PCR. This additional inv22 detection method proves to be cheaper in comparison to the Southern Blot method when cost-analysis was conducted (Cost analysis conducted with the National Health Laboratories cost-analysis program, data not shown) (Kloppers & Janse van Rensburg, 2017). Conventional PCR includes post-PCR analysis, such as gel electrophoresis increasing labour. Additionally, it allows for contaminants to influence the PCR results and therefore one should carefully evaluate the PCR results to ensure that it is contaminant free (Bustin, 2005). We would also recommend, for diagnostic purposes, that an internal amplification control is included in each conventional PCR reaction to eliminate false-negative results (Hoorfar *et al.,* 2004; Randall *et al.,* 2010). The internal amplification control could be a well-preserved DNA sequence (housekeeping gene) that is not known to have expression variations in different individuals. When the PCR results are evaluated, the internal control will produce a PCR fragment in each reaction whether the patient is positive or negative for the inv22. The absence of the internal amplification control PCR fragment will consequently then be indicative of a failed PCR reaction. Furthermore, the multiplex reaction could be optimised which will even further reduce the cost of the method.

4.8.3 Inv22 real-time RT-PCR detection method

Real-time PCR is only as accurate and reproducible as the quality controls simultaneously run with unknown samples (Bustin, 2005). We, therefore, recommend that an inv22 positive and a negative control are always included in the newly developed inv22 real-time RT-PCR detection method. In this study, the controls that

were used were patients that were confirmed as inv22 positive or negative by Sanger sequence analysis. Alternatively, controls may be synthesised by using the mRNA reference sequence for the FVIII normal allele (WT) (NM_000132.3) and the inv22 genotype (**Appendix F**).

Future studies can attempt to optimise real-time PCR probes to increase the specificity of the method. The use of probe chemistries, such as Taqman probes, unfortunately, increases the cost of the method in comparison to SYBR green (Gasparic *et al.,* 2010). Furthermore, to increase the turnaround time of the method, cDNA synthesis and real-time PCR can be combined into a one-step real-time RT-PCR. However, such a method will require additional optimization, as the introduction of a reverse-transcription step combined in one-step with real-time PCR may introduce substantial variation within samples without quality steps to standardise concentrations for example (Bustin *et al.,* 2009). It would further be recommendable to also include an internal amplification control in the real-time assay, as was recommended in the conventional assay. The inclusion of an internal amplification control will subsequently detect PCR inhibition and in the process eliminate false-negative results. The real-time RT-PCR method as developed in this study is not able to distinguish between a type one and a type two inv22 as described in **section 2.8.2.** However, since there is no reported clinical difference between the two types of inv22s, it is not considered as a limitation in this study. It must be emphasised, that even though it does not distinguish between the two types, we firmly believe that our results conclusively show that our methods can detect both types.

4.9 The impact of the study

The inv22 mutation is reportedly associated with 45% of all severe haemophilia A case and is the first mutation that is screened for when haemophilia A genotypes are determined. The importance of the inv22 mutation and research associated with the mutation has thus been well established. The main inv22 detection methods, namely Southern Blot, LD-PCR and I-PCR all have eminent disadvantages, and therefore this study aimed to develop new methods. The newly developed, optimised and validated inv22 conventional and real-time RT-PCR detection methods resulting from this study will have a tremendous impact on this research field because some of the disadvantages that are associated with the main methods could be altered. There is a lack of genetic research in haemophilia A in countries such as South Africa, and therefore this study promotes future genetic studies on haemophilia A. Although based on a smaller population group, this study also determined what the presence of the inv22 mutation in the central South African haemophilia A population is, and identified the haemophilia A patients with inhibitors that are positive for inv22. Consequently, future studies can use the newly developed method to screen a large cohort of haemophilia A patients to enable the haemophilia A genetic profile of patients to be documented. The newly developed inv22 methods are also rapid and cost-effective and can, therefore, be implemented in under-resourced laboratories. The study has also determined the importance of inv22 carrier detection and the newly developed methods can be used to detect carriers of inv22. Consequently, these carriers can be genetically counselled. Finally, to aid in the only current prospective cure for haemophilia A, namely gene therapy, basic genetic research in the field of haemophilia A needs to be conducted within all populations.

Chapter 5: Conclusion

Haemophilia A is an inherited bleeding disorder that has a serious impact, causing a reduced quality of life on patients and their families (Cassis *et al.,* 2011). In South Africa, like many other developing countries, the disorder is reported to be underdiagnosed. It can be contributed to the lack of genetic facilities and the availability of genetic test to readily screen haemophilia A patients and possible carriers for FVIII mutations (Tezanos Pinto & Ortiz, 2004; Mahlangu & Medical and Scientific Council of the South African Haemophilia Federation, 2009). FVIII mutation detection has proved to be essential in haemophilia A treatment, genetic counselling and developing a cure for haemophilia A, respectively. There are various mutations in the *FVIII* gene that cause haemophilia A, and of those mutations, it is reported that the inv22 accounts for 45% of all severe cases worldwide (Kaufman *et al.,* 2013; Johnsen *et al.,* 2017).

There are three commonly used methods used to detect the inv22, namely the Southern Blot, Long-distance PCR (LD-PCR) and Inverse PCR (I-PCR). Although other inv22 PCR detection methods are mentioned, they are not discussed as they are not currently utilised as standard inv22 detection methods. The disadvantages of these main three inv22 detection methods have been discussed, and the need for a rapid and cost-effective method was advocated.

Haemophilia A patients that tested positive for the inv22 using the gold standard Southern Blot method were identified as study controls in our study. A non-haemophilic volunteer (C1) and the two haemophilia A patient samples (C2 and C3) were then consequently used in this study to evaluate the inv22 I-PCR detection method. The same controls were used to develop and optimise an inv22 conventional and a real-time RT-PCR detection method that can overcome some of the disadvantages of the current detection methods.

The evaluation of the inv22 I-PCR detection method yielded unfavourable results. The method was conducted as per the SOP provided by the developer of the method. After several attempts and troubleshooting measures, the evaluation of the inv22 I-PCR detection method failed. The method proved challenging to implement due to the lack

of quality control measures in between the different steps. Due to resource constraints, it was decided to abandon the I-PCR method, after numerous troubleshooting steps were followed. It was a further indication that there was a need for an uncomplicated inv22 detection method that could be utilised in sub-optimally resourced laboratories.

An inv22 conventional RT-PCR detection method was developed using the inv22 genotype confirmed controls, with the ability to detect the FVIII wild-type, inv22 and inv22 carrier genotypes, respectively. Although the main aim of the study was to develop a real-time RT-PCR method, the conventional method provided to be a good alternative for laboratories that are not set up for real-time PCR. The inv22 conventional RT-PCR detection method proved to be rapid, sensitive (Limit of detection 50 ng/µL) and cost-effective, especially considering that it is cheaper than the current inv22 Southern Blot method.

The inv22 real-time RT-PCR detection method developed in this study accurately detected the presence of the FVIII wild-type, inv22 and inv22 carrier genotypes, respectively. The real-time method could be suitable for use in laboratories that are equipped for real-time PCR and will allow a high throughput of samples, which will be advantageous in laboratories with high sample numbers. The inv22 real-time RT-PCR detection method can however only be as reliable as the accompanying controls and the quality assurance measures implemented by the laboratory. It needs to be emphasised that one major advantage of the inv22 real-time RT-PCR detection method is that it does overcome the limitation of post-PCR analysis that is associated with conventional PCR.

Finally, the 60 study participants with unknown inv22 statuses were genotyped with the conventional RT-PCR method, the real-time RT-PCR method and Sanger sequencing. Consequently, eight study participants tested positive for inv22, four were found to be inv22 carriers, and the remainder of the study participants were all inv22 negative, as genotyped using the conventional RT-PCR method. The real-time RT-PCR method had similar results to the conventional RT-PCR method, with a single difference as study participant P14 yielded no results with the real-time method. A comparison between the results obtained with the inv22 conventional and real-time RT-PCR detection methods developed in this study showed a good concordance.

Sanger sequence analysis confirmed the inv22 genotype for each of the study participants, and in the process, the detection methods were consequently validated.

The inv22 conventional and the real-time RT-PCR detection methods overcame many of the disadvantages that are associated with the current inv22 detection methods. In comparison to the inv22 Southern Blot method, the methods developed in this study are cheaper, less laborious and has a turnaround time of 8 hours in comparison to approximately 168 hours. In comparison to the inv22 LD-PCR method, the new detection methods are easy to implement and to standardise. The implementation of the inv22 I-PCR method proved to be too challenging in our laboratory and in comparison to the new detection methods are less complicated and have much fewer steps than the I-PCR method. Furthermore, the new detection methods rely on FVIII mRNA analysis rather than DNA analysis which are used for I-PCR. DNA analysis of the inv22 is complicated by the size of the *FVIII* gene, intron 22 and the homologous sequence within intron 22 that is involved in the inversion (Gitschier *et al.,* 1984, Naylor *et al.,* 1995), whereas the mRNA provides a less complicated target molecule. Accordingly, in comparison, the inv22 conventional and real-time RT-PCR detection methods eliminate some of the disadvantages associated with each of the three main inv22 detection methods.

Global studies reported that the inv22 accounts for 45% of all severe haemophilia A cases (Lakich *et al.,* 1993; Naylor *et al.,* 1993a; Antonarakis *et al.,* 1995; Johnsen *et al.,* 2017). The inv22 was found in only 29.41% of our study population which was substantially lower than 45%. The small population size in this study, however, needs to be taken into consideration when the data is evaluated, and a larger cohort needs to be evaluated to determine if the inv22 indeed has a lower presence in the South African haemophilia A population. Three of the nine haemophilia A patients with FVIII inhibitors were also positive for the inv22. However, no associations could be made due to the small size of the study population.

The development of the inv22 conventional and real-time RT-PCR detection methods will allow for more haemophilia A families to be screened for the mutation as the methods are rapid, uncomplicated and cost-effective. It will promote genetic research in haemophilia A in South Africa considering that the inv22 remains the foremost mutation that results in severe haemophilia A. The consistent advancement of

laboratory test that is easily implementable will make the wide-spread screening for FVIII mutations that cause haemophilia A more of a reality. Consequently, it will resolve some of the issues that are associated with haemophilia A diagnostics and management in developing countries as reported by Tezanos Pinto & Ortiz (2004) and Mahlangu *et al.* (2009).

References

- American Society of Hematology, 2008. *Hemophilia: From Plasma to Recombinant Factors.* [Online] Available at: http://www.hematology.org/About/History/50-Years/1524.aspx [Accessed 04 October 2016].

- Antonarakis, S. E., Rossiter, J. P., Young, M., Horst, J., De Moerloose, P., Sommer, S. S., Ketterling, R. P., Kazazian, H. H., Jr., Negrier, C., Vinciguerra, C., Gitschier, J., Goossens, M., Girodon, E., Ghanem, N., Plassa, F., Lavergne, J. M., Vidaud, M., Costa, J. M., Laurian, Y., Lin, S. W., Lin, S. R., Shen, M. C., Lillicrap, D., Taylor, S. A., Windsor, S., Valleix, S. V., Nafa, K., Sultan, Y., Delpech, M., Vnencak-Jones, C. L., Phillips, J. A., 3rd, Ljung, R. C., Koumbarelis, E., Gialeraki, A., Mandalaki, T., Jenkins, P. V., Collins, P. W., Pasi, K. J., Goodeve, A., Peake, I., Preston, F. E., Schwartz, M., Scheibel, E., Ingerslev, J., Cooper, D. N., Millar, D. S., Kakkar, V. V., Giannelli, F., Naylor, J. A., Tizzano, E. F., Baiget, M., Domenech, M., Altisent, C., Tusell, J., Beneyto, M., Lorenzo, J. I., Gaucher, C., Mazurier, C., Peerlinck, K., Matthijs, G., Cassiman, J. J., Vermylen, J., Mori, P. G., Acquila, M., Caprino, D. & Inaba, H., 1995. Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. *Blood,* Volume 86, pp. 2206-2212.

- Astermark, J., Oldenburg, J., Escobar, M., White II, G., Berntrop, E. & MIBS group., 2005. The Malmö International Brother Study (MIBS). *Haematologica,* Volume 90, pp. 924-930.

- Bagnall, R. D., Giannelli, F. & Green, P. M., 2006. Int22h-related inversions causing hemophilia A: a novel insight into their origin and a new more discriminant PCR test for their detection. *Journal of Thrombosis and Haemostasis,* Volume 4, pp. 591-598.

- Bagnall, R. D., Waseem, N., Green, P. M. & Giannelli, F., 2002. Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A. *Blood,* Volume 99, pp. 168-174.

- Bardi, E. & Astermark, J., 2015. Genetic risk factors for inhibitors in haemophilia A. *European Journal of Haematology,* Volume 94(77), pp. 7-10.
- Bashawri, L.A.M. & Ahmed, M.A., 2007. The approach to a patient with a bleeding episode: For the primary care physician. *Journal of Family and Community Medicine*, Volume 14(2), pp. 53-58.
- Beastall, G. H. & Evans, S. A., 2012. Foreword. In: C. A. Burtis, E. R. Ashwood & D. E. Bruns, eds. *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics.* S .l.: Elsevier Inc.
- Bolton-Maggs, P. & Pasi, K., 2003. Haemophilias A and B. *Lancet,* Volume 24;361(9371), pp. 1801-1809.
- Brinke, A., Green, P. M. & Giannelli, F., 1997. Characterization of the gene (VBP1) and transcript for the von Hippel-Lindau binding protein and isolation of the highly conserved murine homologue. *Genomics,* Volume 45, pp. 105-112.
- Brinke, A., Tagliavacca, L., Naylor, J., Green, P., Giangrande, P. & Giannelli, F., 1996. Two chimaeric transcription units result from an inversion breaking intron 1 of the factor VIII gene and a region reportedly affected by reciprocal translocations in T-cell leukaemia. *Human Molecular Genetics,* Volume 5, pp. 1945-1951.
- Bustin, S., 2005. Real-Time PCR. In: J. Fuchs & M. Podda, eds. *Encyclopedia of Diagnostic Genomics and Proteonomics.* S .l.: Marcel Decker Inc., pp. 1117-1125.
- Bustin, S. A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M., Shipley, G., Vandesompele, J. & Wittwer, C.T., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry,* Volume 55, pp. 611-622.
- Callaghan, M. & Kaufman, R., 2008. Synthesis and Secretion of Coagulation Factor VIII. In: K. Tanaka *et al*. eds. *Recent Advances in Thrombosis and Hemostasis 2008.* S.l.: Springer Japan, pp. 45-67.

- Canadian Hemophilia Society, 2016. *Canadian Hemophilia Society: Factor VIII replacement therapy.* [Online]
  Available at: http://www.hemophilia.ca/en/bleeding-disorders/hemophilia-a-and-b/the-treatment-of-hemophilia/factor-replacement-therapy/
  [Accessed 04 September 2016].

- Cassis, F., Querol, F., Forsyth, A. & Iorio, A., 2011. Psychosocial aspects of haemophilia: a systematic review of methodologies and findings. *Haemophilia,* Volume 18(3), pp. e101-14.

- Clemetson, K., 2012. Platelets and primary haemostasis. *Thrombosis Research,* Volume 129(3), pp. 220-224.

- Cutler, J.; McGaffin, G.; White, A. & Mitchell, M., 2015. Rare alternate presentations of the F8 intron 22 inversion mutation complicate detection and may lead to misdiagnosis. *International Society of Thrombosis and Haemostasis,* Volume 13(2), p827.

- Dahlback, B. 2., 2000. Blood coagulation. *Lancet,* Volume 355, pp. 1627-1632.

- Darby, S. C., Kan, S. W., Spooner, R. J., Giangrande, P. L., Hill, F. G., Hay, C. R., Lee, C. A., Ludlam, C. A. & Williams, M., 2007. Mortality rates, life expectancy, and causes of death in people with hemophilia A or B in the United Kingdom who were not infected with HIV. *Blood,* Volume 110, pp. 815-825.

- Ding, Q., Wu, X., Lu, Y., Chen, C., Shen, R., Zhang, X., Jiang, Z. & Wang, X., 2016. AccuCopy quantification combined with pre-amplification of long-distance PCR for fast analysis of intron 22 inversion in haemophilia A. *Clinica Chimica Acta,* Volume 458, pp. 78-83.

- Dutta, D., Gunasekera, D., Ragni, M. & Pratt, K., 2016. Accurate, simple, and inexpensive assays to diagnose F8 gene inversion mutations in haemophilia A patients and carriers. *Blood,* Volume 1(3), pp. 231-239.

- Eckhardt, C. L., van Velzen, A.S., Peters, M., Astermark, J., Brons P.P., Castaman, G., Cnossen, M.H., Dors, N., Escuriola-Ettingshausen,

C., Hamulyak, K., Hart, D.P., Hay, C.R., Haya, S., van Heerde, W.L., Hermans, C., Holmström, M., Jimenez-Yuste, V., Keenan, R.D., Klamroth, R., Laros-van Gorkom, B.A., Leebeek, F.W., Liesner, R., Mäkipernaa, A., Male, C., Mauser-Bunschoten, E., Mazzucconi, M.G., McRae, S., Meijer, K., Mitchell, M., Morfini, M., Nijziel, M., Oldenburg, J., Peerlinck, K., Petrini, P., Platokouki, H., Reitter-Pfoertner, S.E., Santagostino, E., Schinco, P., Smiers, F.J, Siegmund, B., Tagliaferri, A., Yee, T.T., Kamphuisen, P.W., van der Bom, J.G., Fijnvandraat, K. & INSIGHT Study Group., 2013. Factor VIII gene (F8) mutation and risk of inhibitor development in nonsevere hemophilia A. *Blood,* Volume 122, pp. 1954-1962.

- Factor VIII Gene Variant Database, 2017. *Factor VIII Gene Variant Database.* [Online]
  Available at: http://www.factorviii-db.org/
  [Accessed 03 December 2017].

- Farkas, D. & Holland, C., 2009. Overview of Molecular Diagnostic Techniques and Instrumentation. In: R. Tubbs & M. Stoler, eds. *Cell and Tissue Based Molecular Pathology.* S .l.: Elsevier Inc., pp. 19-32.

- Farrugia, A., 2015. A new recombinant FVIII from a human cell line. *Blood Transfusion,* Volume 13, pp. s555-s559.

- Fay, P. J., 2004. Activation of factor VIII and mechanisms of cofactor action. *Blood review,* Volume 18, pp. 1-15.

- Freije, D. & Schlessinger, D., 1992. A 1.6-Mb contig of yeast artificial chromosomes around the human factor VIII gene reveals three regions homologous to probes for the DXS115 locus and two for the DXYS64 locus. *American Journal of Human Genetics,* Volume 51, pp. 66-80.

- Gale, A. J., Cramer, T. J., Rozenshteyn, D. & Cruz, J. R., 2008. Detailed mechanisms of the inactivation of factor VIIIa by activated protein C in the presence of its cofactors, protein S and factor V. *The Journal of Biological Chemistry,* Volume 283, pp. 16355-16362.

- Gargani, Y., 2012. Haemostasis. In: Horton-Szar, D, eds. *Crash Course: Haematology and Immunology.* S.I.: Mosby Elsevier Ltd, pp. 57-59.
- Gasparic, M., Tengs, T., La Paz, J., Pla, M., Esteve, T., Zel, J. & Gruden, K., 2010. Comparison of nine different real-time PCR chemistries for qualitative and quantitative applications in GMO detection. *Analytical and Bioanalytical Chemistry,* Volume 396, pp. 2023–2029.
- Ginzinger, D., 2002. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Experimental Haematology,* Volume 30, pp. 503-512.
- Gitschier, J., Wood, W. I., Goralka, T. M., Wion, K. L., Chen, E. Y., Eaton, D. H., Vehar, G. A., Capon, D. J. & Lawn, R. M., 1984. Characterization of the human factor VIII gene. *Nature,* Volume 312, pp. 326-330.
- Goodeve, A., 2003. Genetic determinants of inhibitor formation in patients with hemophilia. *Haematologica,* Volume 88(12), pp. 2-3.
- Grasso, M., Boon, E. M. J., Filipovic-Sadic, S., Van Bunderen, P. A., Gennaro, E., Cao, R., Latham, G. J., Hadd, A. G. & Coviello, D. A., 2014. A Novel Methylation PCR that Offers Standardized Determination of FMR1 Methylation and CGG Repeat Length without Southern Blot Analysis. *The Journal of Molecular Diagnostics,* Volume 16, pp. 23-31.
- Green, P. M., Montandon, A. J., Bentley, D. R. & Giannelli, F., 1991. Genetics and molecular biology of haemophilias A and B. *Blood Coagulation and Fibrinolysis,* Volume 2, pp. 539-565.
- Gunasekera, D., Ettinger, R., Fletcher, S., James, E., Liu, M., Barett, J., Withycombe, J., Matthews, D., Epstein, M., Hughes, R. & Pratt, K., 2015. Factor VIII gene variants and inhibitor risk in African American hemophilia A patients. *Blood,* Volume 126, pp. 895-904.
- Habart, D., 2005. Molecular diagnosis of haemophilia A in clinical practice. *Casopis Lekaru Ceskych,* Volume 144, pp. 795-800.
- Heid, C., Stevens, J., Livak, K. & Williams, P., 1996. Real time quantitative PCR. *Genome Research,* Volume 6, pp. 986-994.

- He, L., Chinnery, P. F., Durham, S. E., Blakely, E. L., Wardell, T. M., Borthwick, G. M., Taylor, R. W. & Turnbull, D. M., 2002. Detection and quantification of mitochondrial DNA deletions in individual cells by real‐time PCR. *Nuclei acid research,* Volume 30(e68), pp. 1-6.
- Higuchi, M., Kazazian, H. H., Jr., Kasch, L., Warren, T. C., Mcginniss, M. J., Phillips, J. A., 3rd, Kasper, C., Janco, R. & Antonarakis, S. E., 1991. Molecular characterization of severe hemophilia A suggests that about half the mutations are not within the coding regions and splice junctions of the factor VIII gene. *Proceedings of the National Academy of Sciences of the United States of America,* Volume 88, pp. 7405-7409.
- Hoffbrand, A. & Moss, V., 2011a. Coagulation disorders. In: *Essential Haematology.* S .l.: Wiley-Blackwell Publishers, pp. 346-352.
- Hoffbrand, A. & Moss, V., 2011b. Platelets, blood coagulation and haemostasis. In: *Essential Haematology.* S .l.: Wiley-Blackwell Publishers, pp. 315-323.
- Hoffman, M. & Monroe, M., 2007. Coagulation 2006: A modern view of Hemostasis. *Haematology and Oncology Clinics of North America,* Volume 21, pp. 1-11.
- Hoorfar, J., Malorny, B., Abdulmawjood, A., Cook, N., Wagner, M. & Fach, P., 2004. Practical Considerations in Design of Internal Amplification Controls for Diagnostic PCR Assays. *Journal of Clinical Microbiology,* Volume 42(5), pp. 1863-1868.
- Huang, X. & Miller, W., 1991. A time-efficient, linear-space local similarity algorithm. *Advances in Applied Mathematics,* Volume 12(3), pp. 337-357.
- Ingram, G., 1976. The history of haemophilia. *Journal of Clinical Pathology,* Volume 29(6), pp. 469-479.
- Jayandharan, G. R., Srivastava, A. & Srivastava, A., 2012. Role of molecular genetics in hemophilia: from diagnosis to therapy. *Seminar in Thrombosis and Haemostasis,* Volume 38, pp. 64-78.

- Jenkins, P. V., Collins, P. W., Goldman, E., McCraw, A., Riddell, A., Lee, C. A. & Pasi, K. J., 1994. Analysis of intron 22 inversions of the factor VIII gene in severe hemophilia A: implications for genetic counselling. *Blood,* Volume 84, pp. 2197-2201.

- Jobson, J. & Brown, B.D., 2017. Micromanaging Tolerance in Hemophilia A Gene Therapy. *Molecular Therapy,* Volume 25(8), pp. 1739-1740.

- Johnsen, J., Fletcher, S., Huston, H., Roberge, S., Martin, B., Kircher, M., Josephson, N., Shendure, J., Ruuska, S., Koerper, M., Morales, J., Pierce, G., Aschman, D. & Konkle, B., 2017. Novel approach to genetic analysis and results in 3000 hemophilia patients enrolled in the My Life, Our Future initiative. *Blood Advances,* Volume 1, pp. 824-834.

- Kamal, A.H., Tefferi, A. & Pruthi, R.K., 2007. How to Interpret and Pursue an Abnormal Prothrombin Time, Activated Partial Thromboplastin Time, and Bleeding Time in Adults. *Mayo Clinic Proceedings*, Volume 82(7), pp. 864-873.

- Kang, J., Kaysen, G., Jones Jr, H. & Holland, M., 1997. Rat Liver Transcript Profiling in Normal and Disease States Using a Kinetic Polymerase Chain Reaction Assay. *Methods,* Volume 13, pp. 437-443.

- Kasper, C. & Lin, J., 2007. Prevalence of sporadic and familial haemophilia. *Haemophilia,* Volume 13, pp. 90-92.

- Kaufman, R.J., Fay, P.J., Popolo, L. & Ortel, T.L., 2013. Factor V and factor VIII. In: Marder, V.J., Aird, W.C., Bennett, J.S., Schulman, S. & White, G.C, eds. *Haemostasis and Thrombosis: Basic Principles and Clinical Practice*. 6 ed. S.I.: Lippincott Williams and Wilkins, pp.179-196.

- Keeney, S., Mitchell, M. & A., G., 2005. The molecular analysis of haemophilia A: a guideline from the UK haemophilia centre doctors' organization haemophilia genetics laboratory network. *Heamophilia,* Volume 11, pp. 387-397.

- Kessler, L., Adams, R., Mighion, L., Walther, S. & Ganguly, A., 2014. Prenatal diagnosis in haemophilia A: experience of the genetic diagnostic laboratory. *Haemophilia,* Volume 20, pp. e384-391.
- Kloppers, J. & Janse van Rensburg, W., 2015. Evaluation of a Chromogenic Factor VIII Assay on the SYSMEX CS 2100I Instrument. *Medical Technology South Africa,* pp. 27-30.
- Kloppers, J. & Janse van Rensburg, W., 2017. Rapid identification of the intron 22 inversion in haemophilia A. *Hemophilia,* Volume 23, pp e55-e57.
- Knobe, K. E., Sjorin, E., Soller, M. J., Liljebjorn, H. & Ljung, R. C., 2008. Female haemophilia A caused by skewed X inactivation. *Haemophilia,* Volume 14, pp. 846-848.
- Kumar, P., Husain, N., Soni, P., Faridi, N. J. & Goel, S. K., 2013. Protocol for Detection of Intron 22 Inversion Mutation From Cases With Hemophilia A. *Clinical and Applied Thrombosis and Haemostasis,* Volume 00, pp. 1-5.
- Lakich, D., Kazazian, H. H. J., Antonarakis, S. E. & Gitschier, J., 1993. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nature Genetics,* Volume 5, pp. 236-241.
- Lee, C., 2010. Historical introduction. In: C. Lee & W. Hoots, eds. *Textbook of Hemophilia.* S .l.: Wiley-Blackwell Publishers, pp. 1-4.
- Leuer, M., Oldenburg, J., Lavergne, J., Ludwig, M., Fregin, A., Eigel, A., Ljung, R., Goodeve, A., Peake, I. & Olek, K., 2001. Somatic mosaicism in hemophilia A: a fairly common event. *American Journal of Human Genetics,* Volume 69, pp. 75-87.
- Levinson, B., Kenwrick, S., Gamel, P., Fisher, K. & Gitschier, J., 1992. Evidence for a third transcript from the human factor VIII gene. *Genomics,* Volume 14, pp. 585-589.
- Levinson, B., Kenwrick, S., Lakich, D., Hammonds, G., Jr. & Gitschier, J., 1990. A transcribed gene in an intron of the human factor VIII gene. *Genomics,* Volume 7, pp. 1-11.

- Lieuw, K., 2017. Many factor VIII products available in the treatment of hemophilia A: an embarrassment of the riches? *Journal of Blood Medicine*, Volume 8, pp. 67-73.
- Liu, Q., Nozari, G. & Sommer, S. S., 1998. Single-tube polymerase chain reaction for rapid diagnosis of the inversion hotspot of mutation in hemophilia A. *Blood,* Volume 92, pp. 1458-1459.
- Liu, Q. & Sommer, S. S., 1998. Subcycling-PCR for multiplex long-distance amplification of regions with high and low GC content: application to the inversion hotspot in the factor VIII gene. *Biotechniques,* Volume 25, pp. 1022-1028.
- Lusher, J. M. & Mcmillan, C. W., 1978. Severe factor VIII and factor IX deficiency in females. *American Journal of Medicine,* Volume 65, pp. 637-648.
- Mahlangu, J. N., Gilham, A. & Medical & Scientific Advisory Council of the South African Haemophilia, F., 2008. Guideline for the treatment of haemophilia in South Africa. *South African Medical Journal,* Volume 98, pp. 126-140.
- Mahlangu, J. N. & Medical and Scientific Council of the South African Haemophilia Federation, 2009. Haemophilia care in South Africa: 2004-2007 look back. *Haemophilia,* Volume 15(1), pp. 135-141.
- McKusick, V., 1965. THE ROYAL HEMOPHILIA. *Scientific American,* 213(2), pp. 88-95.
- Miners, A. H., Sabin, C. A., Tolley, K. H., Jenkinson, C., Kind, P. & Lee, C. A., 1999. Assessing health-related quality-of-life in individuals with haemophilia. *Haemophilia,* Volume 5, pp. 378-385.
- Monis, P., Giglio, S. & Saint, C., 2005. Comparison of SYTO9 and SYBR Green I for a real-time polymerase chain reaction and investigation of the effect of dye concentration on amplification and DNA melting curve analysis. *Analytical Biochemistry,* Volume 340, pp. 24-34.
- Myles, T., Yun, T. & Leung, L., 2002. Structural requirements for the activation of human factor VIII by thrombin. *Blood,* Volume 100, pp. 2820-2826.

- Naylor, J. A., Buck, D., Green, P., Williamson, H., Bentley, D. & Giannelli, F. 1995. Investigation of the factor VIII intron 22 repeated region (int22h) and the associated inversion junctions. *Human Molecular Genetics,* Volume 4, pp. 1217-1224.
- Naylor, J., Brinke, A., Hassock, S., Green, P.M. & Giannelli, F., 1993a. Characteristic mRNA abnormality found in half the patients with severe haemophilia A is due to large DNA inversions. *Human Molecular Genetics,* Volume 2, pp. 1773-1778.
- Naylor, J. A., Green, P. M., Rizza, C. R. & Giannelli, F., 1993b. Analysis of factor VIII mRNA reveals defects in every one of 28 haemophilia A patients. *Human Molecular Genetics,* Volume 2, pp. 11-17.
- Ohler, L. D. & Rose, E. A., 1992. Optimization of long-distance PCR using a transposon-based model system. *PCR Methods Applied,* Volume 2, pp. 51-59.
- Oldenburg, J., Rost, S., El-Maarri, O., Leuer, M., Olek, K., Müller, C., Schwaab, R., 2000. De novo *factor VIII* gene intron 22 inversion in a female carrier presents as a somatic mosaicism. *Blood*, Volume 96, pp. 2905-2906.
- Oldenburg, J., Schröder, H., Brackmann, C., Muller-Reible, R., Schwaab, E. & Tuddenham, E., 2004. Environmental and genetic factors influencing inhibitor development. *Seminars in Haematology,* Volume 41(1), pp. 82-88.
- Olsson, A., Margeretha, H., Berntrop, E., Ljung, R. & Baghoei, F., 2014. Clotting factor level is not a good predictor of bleeding in carriers of haemophilia A and B. *Blood coagulation and Fibrinolysis,* Volume 25(5), pp. 471-475.
- Package insert. Power SYBR® Green Master Mix, Applied Biosystems, USA. [Online]. Available at: https://www.thermofisher.com/za/en/home/life-science/pcr/real-time-pcr/real-time-pcr-reagents/sybr-green-real-time-master-mixes/powerup-sybr-green-master-mix.html. [Accessed: 20 January 2017].
- Pandey, G. S., Yanover, C., Miller-Jenkins, L. M., Garfield, S., Cole, S. A., Curran, J. E., Moses, E. K., Rydz, N., Simhadri, V., Kimchi-Sarfaty, C.,

Lillicrap, D., Viel, K. R., Przytycka, T. M., Pierce, G. F., Howard, T. E., Sauna, Z. E. & Investigators, P. S., 2013. Endogenous factor VIII synthesis from the intron 22-inverted F8 locus may modulate the immunogenicity of replacement therapy for hemophilia A. *Nature Medicine,* Volume 19, pp. 1318-1324.

- Pan, T. Y., Chiou, S. S., Wang, C. C. & Wu, S. M., 2014. Separation of intron 22 inversion type 1 and 2 of hemophilia A by a modified inverse-shifting polymerase chain reaction and capillary gel electrophoresis. *Talanta,* Volume 130, pp. 328-335.

- Peake, I. R., Lillicrap, D. P., Boulyjenkov, V., Briet, E., Chan, V., Ginter, E. K., Kraus, E. M., Ljung, R., Mannucci, P. M., Nicolaides, K. & Et Al., 1993. Haemophilia: strategies for carrier detection and prenatal diagnosis. *Bulletin of the World Health Organisation,* Volume 71, pp. 429-458.

- Peetz, D., 2007. Factor VIII Methods: Which Assay Principle for which Indication?. In: I. Scharrer & W. Schramm, eds. *36th Hemophilia Symposium Hamburg 2005.* S .l.: Springer Berlin Heidelberg, pp. 71-74.

- Peyvandi, F., Jayandharan, G., Chandy, M., Srivastava, A., Nakaya, S. M., Johnson, M. J., Thompson, A. R., Goodeve, A., Garagiola, I., Lavoretano, S., Menegatti, M., Palla, R., Spreafico, M., Tagliabue, L., Asselta, R., Duga, S. & Mannucci, P. M., 2006. Genetic diagnosis of haemophilia and other inherited bleeding disorders. *Haemophilia,* Volume 12(3), pp. 82-89.

- Plug, I., Mauser-Bunschoten, E. P., Brocker-Vriends, A. H., Van Amstel, H. K., Van Der Bom, J. G., Van Diemen-Homan, J. E., Willemse, J. & Rosendaal, F. R., 2006. Bleeding in carriers of hemophilia. *Blood,* Volume 108, pp. 52-56.

- Polakova, H., Zmetakova, I. & Kadasi, L., 2003. Long distance PCR in detection of inversion mutations of F8C gene in hemophilia A patients. *General Physiology and Biophysics,* Volume 22, pp. 243-253.

- Qiu, F., Zheng, H., Yi, Y., Jia, Z., Cao, J. & Bi, S., 2013. Comparative evaluation of a novel TaqMan real-time reverse transcription–polymerase chain reaction assay for hepatitis A virus detection. *Journal of International Medical Research,* Volume 41, pp. 427-434.

- Ragni, M., Ojeifo, O., Feng, J., Yan, J., Hill, K., Sommer, S., Trucco, M., Brambilla, D. & Haemophilia Inhibitor Study group., 2009. Risk Factors for Inhibitor Formation in Hemophilia: A Prevalent Case-Control Study. *Haemophilia,* Volume 15(5), pp. 1074-1082.

- Randall, L., Lemma, F., Rodgers, J., Vidal, A. & Clifton-Hudley, F., 2010. Development and evaluation of internal amplification controls for use in a real-time duplex PCR assay for detection of Campylobacter coli and Campylobacter jejuni. *Journal of Medical Microbiology,* Volume 59, pp. 172-178.

- Rangarajan, S., Walsh, L., Lester, W., Perry, D., Madan, B., Laffan, M., Yu, H., Vettermann, C., Pierce, G.F., Wong, W. & Pasi, K.J., 2017. AAV5-Factor VIII Gene Transfer in Severe Hemophilia A. The *New England Journal of Medicine,* Volume 377, pp. 2519-2530.

- Renault, N. K., Dyack, S., Dobson, M. J., Costa, T., Lam, W. L. & Greer, W. L., 2007. Heritable skewed X-chromosome inactivation leads to haemophilia A expression in heterozygous females. *European Journal of Human Genetics,* Volume 15, pp. 628-637.

- Roberts, H., 1991. Factor VIII Replacement Therapy: Issues and Future Prospects. *Annals of the New York Academy of Science,* Volume 614, pp. 106-113.

- Rogers, G. L. & Herzog, R. W., 2015. Gene therapy for hemophilia. *Frontiers in bioscience (Landmark edition),* Volume 20, pp. 556-603.

- Rosner, F., 1969. Hemophilia in the Talmud and rabbinic writings. *Annals of Internal Medicine,* Volume 70(4), pp. 833-837.

- Rossetti, L. C., Radic, C. P., Larripa, I. B. & De Brasi, C. D., 2005. Genotyping the hemophilia inversion hotspot by use of inverse PCR. *Clinical Chemistry,* Volume 51, pp. 1154-1158.

- Rossetti, L. C., Radic, C. P., Larripa, I. B. & De Brasi, C. D., 2008. Developing a new generation of tests for genotyping hemophilia-causative

rearrangements involving int22h and int1h hotspots in the factor VIII gene. *Journal of Thrombosis and Haemostasis,* Volume 6, pp. 830-836.

- Rossiter, J. P., Young, M., Kimberland, M. L., Hutter, P., Ketterling, R. P., Gitschier, J., Horst, J., Morris, M. A., Schaid, D. J., De Moerloose, P. & Et Al., 1994. Factor VIII gene inversions causing severe hemophilia A originate almost exclusively in male germ cells. *Human Molecular Genetics,* Volume 3, pp. 1035-1039.
- Ross, M. T., Grafham, D., Coffey, A., Scherer, S., Mclay, K., Muzny, D., Platzer, M., Howell, G., Burrows, C., Bird, C. & Et. Al., 2005. The DNA sequence of the human X chromosome. *Nature,* Volume 434, pp. 325-337.
- Sauna, Z. E., Lozier, J. N., Kasper, C. K., Yanover, C., Nichols, T. & Howard, T. E., 2015. The intron-22-inverted F8 locus permits factor VIII synthesis: an explanation for low inhibitor risk and a role for pharmacogenomics. *Blood,* Volume 125, pp. 223-228.
- Schroder, J., El-Maarri, O., Schwaab, R., Muller, C.R. & Oldenburg, J., 2006. Factor VIII intron-1 inversion: frequency and inhibitor prevalence. *Journal of Haemostasis and Thrombosis,* Volume 4, pp. 1141-1143.
- Schwaab, R., Brackmann, H.H., Meyer, C., Seehafer, J., Kirchgesser, M., Haack, A., Olek, K., Tuddenham, E. G. & Oldenburg, J., 1995. Haemophilia A: mutation type determines the risk of inhibitor formation. *Thrombosis and Haemostasis,* Volume 74, pp. 1402-1406.
- Sharma, A., Easow Mathew, M., Sriganesh, V., Neely, J.A. & Kalipatnapu, S., 2014. Gene therapy for haemophilia. *Cochrane Database System Review,* Volume 11, p. CD010822.
- Shen, B., Spigel, P., Chang, C., Huh, J., Lee, J., Kim, J., Kim, Y. & Stoddard, B., 2008. The tertiary structure and domain organization of coagulation factor VIII. *Blood,* Volume 111(3), pp. 1240-1247.
- Shima, M., 2006. Characterization of Factor VIII Inhibitors. *International Journal of Hematology,* Volume 83, pp. 109-118.

- Solovieva, S., 2001. Clinical severity of disease, functional disability and health-related quality of life. Three-year follow-up study of 150 Finnish patients with coagulation disorders. *Haemophilia,* Volume 7, pp. 53-63.
- Soucie, J. M., Evatt, B. & Jackson, D., 1998. The occurrence of hemophilia in the United States: The Hemophilia Surveillance System Project Investigators. *American Journal of Heamatology,* Volume 59(4), pp. 288-294.
- Southern, E. M., 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology,* Volume 98**,** pp. 503-517.
- Spurgeon, S., Jones, R. & Ramakrishnan, R., 2008. High throughput gene expression measurement with real-time PCR in a microfluidic dynamic array. *PloS one,* Volume 3, p. e1662.
- Srivastava, A., Brewer, A. K., Mauser-Bunschoten, E. P., Key, N. S., Kitchen, S., Llinas, A., Ludlam, C. A., Mahlangu, J. N., Mulder, K., Poon, M. C., Street, A. & Treatment Guidelines Working Group on Behalf of the World Federation Of, H., 2013. Guidelines for the management of hemophilia. *Haemophilia,* Volume 19, pp. e1-47.
- Standard Operating Procedure. Quantifying Complementary Deoxyribonucleic Acid. Massachusetts Human Stem Cell Bank. [Online]. Available at: https://www.google.co.za/search?q=Quantifying+Complementary+Deoxyribonucleic+Acid.&rlz=1C1AOHY_enZA718ZA718&oq=Quantifying+Complementary+Deoxyribonucleic+Acid.&aqs=chrome.69i57.926j0j8&sourceid=chrome&ie=UTF-8. [Accessed: 20 July 2016].
- Statistics South Africa, 2017. *Statistical release P0302: Mid-year population estimates 2017 South Africa,* s.l.: Statistics South Africa. [Online] Available at: www.statssa.gov.za/publications/P0302/P03022017.pdf.[Accessed: 22 September 2017].
- Stevens, R., 1999. Historical review: The history of haemophilia in the royal families of Europe. *British Journal of Haematology,* Volume 105, pp. 25-32.

- Takeyama, M., Wakabayashi, H. & Fay, P., 2013. The contribution of Factor VIII Light Chain Residues 2007–2016 to an Activated Protein C-Interactive Site. *Thrombosis and Haemostasis,* Volume 109(2), pp. 187-198.
- Tezanos Pinto, M. & Ortiz, Z., 2004. Haemophilia in the developing world: successes, frustrations and opportunities. *Haemophilia,* Volume 10(4), pp. 14-19.
- Technical note. Product ER7021. Frequently asked questions. Thermo Scientific. USA. [Online]. Available at: https://www.thermofisher.com/search/results?query=ER0721&persona=DocSupport&type=Product+FAQs. [Accessed 3 February 2017].
- Tizzano, E. F., Cornet, M. & Baiget, M., 2003. Inversion of intron 1 of the factor VIII gene for direct molecular diagnosis of hemophilia A. *Haematologica,* Volume 88, pp. 118-120.
- Triglia, T., Peterson, M. G. & Kemp, D. J., 1988. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. *Nuclei Acid Research,* Volume 16, p. 8186.
- Tuddenham, E. G., Schwaab, R., Seehafer, J., Millar, D., Gitschier, J., Higuchi, M., Bidchandani, S., Connor, J., Hoyer, L., Yoshioka, A. & Et. Al., 1994. Haemophilia A: database of nucleotide substitutions, deletions, insertions and rearrangements of the FVIII gene. *Nuclei Acids Research,* Volume 22, pp. 4851-4868.
- Valasek, M. & Repa, J., 2005. The power of real-time PCR. *Advances in physiology education,* Volume 29, pp. 151-159.
- Van den Berg, H., De Groot, P. & Fischer, K., 2007. Phenotypic heterogeneity in severe hemophilia. *Journal of Thrombosis and Haemostasis,* Volume 5(1), pp. 151-156.
- Vlot, A. J., Koppelman, S. J., Van Den Berg, M. H., Bouma, B. N. & Sixma, J. J., 1995. The affinity and stoichiometry of binding of human factor VIII to von Willebrand factor. *Blood,* Volume 85, pp. 3150-3157.

- White, G.C., 2nd, Rosendaal, F., Aledort, L.M., Lusher, J.M., Rothschild, C., Ingerslev, J., Factor, V. & Factor, I. X. S., 2001. Definitions in hemophilia. Recommendation of the scientific subcommittee on factor VIII and factor IX of the scientific and standardization committee of the International Society on Thrombosis and Haemostasis. *Thrombosis and Haemostasis,* Volume 85, p. 560.

- Winikoff, R. & Lee, C., 2010. Hemophilia Carrier Status and Counseling the Symptomatic and Asymptomatic Adolescent. *Journal of Pediatric and Adolescent Gynecology,* Volume 23(6), pp. S43-S47.

- Witmer, C. & Young, G., 2013. Factor VIII inhibitors in hemophilia A: rationale and latest evidence. *Therapeutic Advances in Hematology,* Volume 4(1), pp 59-72.

- Wong, M. & Medrano, J., 2005. Real-time PCR for mRNA quantitation. *Biotechniques,* Volume 39, pp. 75-85.

- World Federation of Hemophilia, 2017. World Federation of Hemophilia report on the annual global survey 2016*,* Canada: *World Federation of Hemophilia.* [Online] Available at: https://www1.wfh.org/publication/files/pdf-1690.pdf. [Accessed: 30 November 2017].

- Yang, L., Ding, J., Zhang, C., Jia, J., Weng, H., Liu, W., Zhang, D., 2005. Estimating the copy number of transgenes in transformed rice by real-time quantitative PCR. *Plant cell reports,* Volume 23, pp. 759-763.

- Zaiden, A. R., Furlong, M. A., Crouch, G. D. & Besa, E. C., 2014a. *Medscape.* [Online]
  Available at: http://emedicine.medscape.com/article/779322-clinical
  [Accessed 16 February 2015].

- Zaiden, A. R., 2014b. *Medscape.* [Online]
  Available at: http://emedicine.medscape.com/article/779322-overview#a0101
  [Accessed 17 February 2015].

- Ziedens, K. B. & Mann, K. G., 2010. Overview of hemostasis. In: C. Lee, E. Berntorp & W. K. Hoots, eds. *Textbook of Hemophilia..* S.l.: Wiley-Blackwell

Publishers. [Online] Available at: https://books.google.co.za/books?hl=en&lr=&id=DkvmZbofLPQC&oi=fnd&pg=PT34&dq=Overview+of+haemostasis+Textbook+of+Haemophilia+second+edition&ots=3VbJYS6Vlf&sig=qwvSJ3FUTldCbsZgulQos7y_-_E#v=onepage&q&f=false. [Accessed 21 October 2015].

Appendix A: Inv22 detection results for Control C2 and C3 with the Southern Blot method

NATIONAL HEALTH
LABORATORY SERVICE
University of the Witwatersrand, School of Pathology
Division Of Human Genetics
Practice No.: 0052002 96

Hospital Street, Johannesburg, 2001
Telephone: +27-11-489-9224/9223/9211
PO Box 1038, Johannesburg, 2000
Fax: +27-11-489-9226 or +27-11-489-9209
Prof A Christianson 011 489-9211/9239 • Prof A Krause 011 489-9219 • Prof M Ramsay 011 489-9214
Prof H Soodyall 011 489-9208 • Dr T Lane 011 489-9221

GENETICS DEPARTMENT Tel (011) 489-9000 Fax (011) 489-9001 Page 1 of 2

[redacted]

Patient
Address
Age(Sex) DoH
Ref Dr
Ward-Hosp
Hosp No
Taken
Report

[redacted]

LABORATORY REPORT

Clinical data Genetic testing
HAEMOPHILIA A
Specimen Blood
Tests ordered HPA

DNA TEST: HAEMOPHILIA A

AIM
To identify or exclude the common Haemophilia A (HPA) intron 22 inversion mutation.

BACKGROUND
CONTROL 2 is affected with HPA. Linked marker analyis was d previously and the high-risk X chromosome was determined (CGC0906432, reported 05/10/2009). DNA testing was undertaken to determine if the disease-causing mutation is the inversion mutation in intron 22 of the Factor VIII gene. The inversion mutation accounts for 45% of severe Haemophilia A chromosomes.

RESULTS
CONTROL 2 tested POSITIVE for the intron 22 distal inversion mutation. esult can be given with 99% accuracy.

NATIONAL HEALTH LABORATORY SERVICE

University of the Witwatersrand, School of Pathology

DIVISION OF HUMAN GENETICS

Hospital Street, Johannesburg, 2001
Telephone: +27-11-489-9224/9223/9211

PO Box 1038, Johannesburg, 2000
Fax: +27-11-489-9226/9209/9286

Professor D Viljoen 489-9239 • Prof A Krause 489-9219 • Prof A Christianson 489-9212 • Prof M Ramsay 489-9214
Dr H Soodyall 489-9208 • Dr AB Lane 489-9221 • Ms T Wessels 489-9243

MOLECULAR GENETICS

Tel (011) 489-9000 Fax (011) 489-9001

[redacted]

Labno :
Patient :
Address :

Tel No :
Gender :
Med Aid :
Member :

Ref Dr :
Practice :
Taken :
Reported :
Copies :

[redacted]

LABORATORY REPORT

SPECIMEN : Blood
TESTS ORDERED: HPA

DNA TEST: HAEMOPHILIA A

AIM

To identify or exclude the common Haemophilia A (HPA) intron 22 inversion mutation and to perform linked marker analysis to track the mutant HPA gene in the family.

BACKGROUND

[redacted] father, [redacted] and son, [redacted], are affected with HPA. She is therefore an obligate carrier of HPA. DNA testing was undertaken to determine whether the disease-causing mutation is the inversion mutation in intron 22 of the Factor VIII gene. This mutation accounts for 45% of severe HPA chromosome. Linked marker analysis, using markers in intron 13 and intron 22 of the Factor VIII gene and an extragenic marker, p39(AC)n, was used to track the high-risk X chromosome in the family.

RESULTS

	INTRON 13	INTRON 22	P39	INVERSION
Possible carrier	20/20	26/28	158/158	UNSUCCESSFUL
CONTROL 3	20	26	158	POSITIVE

This result can be given with 99% accuracy.

Appendix B: Health Science Research Ethics Committee report

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

UFS·UV
HEALTH SCIENCES
GESONDHEIDSWETENSKAPPE

IRB nr 00006240
REC Reference nr 230408-011
IORG0005187
FWA00012784

02 December 2015

MR JF KLOPPERS
DEPARTMENT OF HAEMATOLOGY AND CELL BIOLOGY
FACULTY OF HEALTH SCIENCES
UFS

Dear Mr Kloppers

ECUFS NR 166/2015
PROJECT TITLE: DEVELOPMENT AND IMPLEMENTATION OF A REAL-TIME REVERSE TRANSCRIPTASE POLYMERASE CHAIN REACTION TO DETECT THE INTRON 22 INVERSION IN HAEMOPHILIA A PATIENTS.

1. You are hereby kindly informed that, at the meeting held on 01 December 2015, the Ethics Committee approved the following project after all conditions have been met, when the signed permission letter from the Free State Department of Health was submitted.

2. The Committee must be informed of any serious adverse event and/or termination of the study.

3. Any amendment, extension or other modifications to the protocol must be submitted to the Ethics Committee for approval.

4. A progress report should be submitted within one year of approval of long term studies and a final report at completion of both short term and long term studies.

5. Kindly use the ECUFS NR as reference in correspondence to the Ethics Committee Secretariat.

6. The Ethics Committee functions in compliance with, but not limited to, the following documents and guidelines: The SA National Health Act. No. 61 of 2003; Ethics in Health Research: Principles, Structures and Processes (2015); SA GCP(2006); Declaration of Helsinki; The Belmont Report; The US Office of Human Research Protections 45 CFR 461 (for non-exempt research with human participants conducted or supported by the US Department of Health and Human Services- (HHS), 21 CFR 50, 21 CFR 56; CIOMS; ICH-GCP-E6 Sections 1-4; The International Conference on Harmonization and Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH Tripartite), Guidelines of the SA Medicines Control Council as well as Laws and Regulations with regard to the Control of Medicines, Constitution of the Ethics Committee of the Faculty of Health Sciences.

Yours faithfully

.....
DR SM LE GRANGE
CHAIR: ETHICS COMMITTEE

Cc: Mr WJ Janse van Rensburg

Ethics Committee
Office of the Dean: Health Sciences
T: +27 (0)51 401 7795/7794 | F: +27 (0)51 444 4359 | E: ethicsfhs@ufs.ac.za
Block D, Dean's Division, Room D104 | P.O. Box/Posbus 339 (Internal Post Box G40) | Bloemfontein 9300 | South Africa
www.ufs.ac.za

Appendix C: Example of the informed consent form

Project Title: Development of a real-time reverse transcriptase polymerase chain reaction assay to detect the intron 22 inversion in haemophilia A patients.

To Project Participant:

You are invited to take part in a research project conducted in the Department of Haematology & Cell Biology at the University of the Free State, Bloemfontein, South Africa. In our study we will develop and implement new tests to diagnose mutations in people suffering from haemophilia A. Haemophilia A is a bleeding disorder, caused by the deficiency of a protein that causes a person's blot to clot, namely coagulation factor VIII (FVIII). It is important that these mutations are diagnosed as they play a role in the treatment of haemophilia A.

We need (15 ml) of blood and your time involvement will approximately be five (5) minutes as the blood is drawn. There will be no further engagements, remunerations or costs involved. The risks should not exceed those that are normally expected in donating blood samples for scientific research. You may experience a slight degree of pain and discomfort during the procedure and adverse effects like a slight bruise or swelling at the site where the blood was drawn.

We do not expect any of these adverse medical effects to occur. However, if you do experience any medical problems as a result of your participation, the necessary medical treatment within the ability of the Haematology Clinic will be provided.

Reports resulting from this study might in future be used for scientific publication purposes, but will not identify you as a participant. All information gathered in this study will remain confidential and be given out only with your permission or as required by law. If you give us permission by signing this consent form, we will protect your confidentiality.

If you have any questions about this research at any time, please call Jean Kloppers (0514053116). You may contact the Secretariat of the Ethics Committee of the Faculty of Health Sciences, University of the Free State, at telephone number (051) 4052812 if you have questions regarding your right as a participant in this research project.

By signing this consent form, you indicate that you have read the form and agree voluntarily to participate in the study. If you choose not to take part, there will be no penalty or loss of benefits to which you are entitled. If you agree to take part, you are free to withdraw from it at any time. Likewise, no penalty or loss of benefits to which you are otherwise entitled will occur.

The research study, including the above-mentioned information, has been verbally described to me. I agree to participate in the study, as set out above voluntarily.

…… ……..

Signature Date

Information document for genetic research

Project title: Development of a real-time reverse transcriptase polymerase chain reaction assay to detect the intron 22 inversion in haemophilia A patients.

We are planning to conduct a research project wherein we will attempt to develop new diagnostic tests to detect mutations involved in the serious form of haemophilia A. Haemophilia A is a bleeding disorder where a person's blood can't clot due to a deficiency of a protein known as coagulation factor VIII (FVIII). The mutation is found in the FVIII gene of the human body. Genes are what you inherit from your parents. They are found in every part of your body, and therefore they will be present in blood.

We request your permission to draw blood (15 mL) and to use your DNA (blood) and RNA (blood) for future laboratory analysis.

The findings of this study will not have a direct bearing on your management of this disorder. However, it will allow the diagnosis of mutations that cause haemophilia A to be detected. These mutations might play a role in how haemophilia A is treated.

You are free to refuse consent, and you do not have to give reasons for doing so.

Privacy and Confidentiality

The following arrangements have been made to ensure privacy and confidentiality of your genetic information:

- Your blood sample will be marked with a code and not your name. Researchers will, therefore, be able to identify the sample, but not technicians working with the sample.

Results of research

If this research generates information about you which may be of relevance to the health of other family members, your consent will be sought before offering to disclose such information to the family members concerned, except if such disclosure is compulsory as determined by the law.

Storage of blood, RNA and cDNA

We would like to retain your blood, RNA and cDNA (blood) for possible future research.

The duration of storage will be maximum fifteen years.

Your blood, RNA and cDNA sample will be marked with a code to protect your identity

If you are unhappy to have your blood, RNA and cDNA stored for future research, your genetic material and information will be disposed of at the end of this study, once the sample storage and record-keeping requirements of good research practice have been met.

Do you have any sensitivity on how your blood should be disposed of? If so, what are they?

_____

These will be recorded and taken into account at the time of disposal.

We can dispose of your genetic material even after the research has started since the samples are stored in an identifiable form (unique number).

Voluntary Participation

You do not have to agree to take part in this research, and you are free to withdraw from the research at any time. Your routine medical treatment will not be compromised in any way if you do not participate.

---	---
Signature of participant:	Signature of researcher
Name:	Name:
Date:	Date

Appendix D: Example of the informed consent form signed by parents or legal guardians in case of minor participants

Informed Consent: Parents/Legal guardians of minors:

Project Title: Development and implementation of a real-time reverse transcriptase polymerase chain reaction assay to detect the intron 22 inversion in haemophilia A patients.

To Project Participant:

Your child is invited to take part in a research project conducted in the Department of Haematology & Cell Biology at the University of the Free State, Bloemfontein, South Africa. In our study we will develop and implement new tests to diagnose mutations in people suffering from haemophilia A. Haemophilia A is a bleeding disorder, caused by the deficiency of a protein that causes a person's blood to clot, namely coagulation factor VIII (FVIII). It is important that these mutations are diagnosed as they play a role in the treatment of haemophilia A.

We need (15 ml) of blood and your child's time involvement will approximately be five (5) minutes as the blood is drawn. There will be no further engagements, remunerations or costs involved. The risks should not exceed those that are normally expected in donating blood samples for scientific research. Your child may experience a slight degree of pain and discomfort during the procedure and adverse effects like a slight bruise or swelling at the site where the blood was drawn.

We do not expect any of these adverse medical effects to occur. However, if your child does experience any medical problems as a result of his or her participation, the necessary medical treatment within the ability of the Haematology Clinic will be provided.

Reports resulting from this study might in future be used for scientific publication purposes, but will not identify your child as a participant. All information gathered in this study will remain confidential and be given out only with your permission or as required by law. If you give us permission by signing this consent form, we will protect your child's confidentiality.

If you or your child have any questions about this research at any time, please call Jean Kloppers (0514053116). You may contact the Secretariat of the Ethics Committee of the Faculty of Health Sciences, University of the Free State, at telephone number (051) 4052812 if you have questions regarding your child's right as a participant in this research project.

By signing this consent form, you indicate that you have read the form and agree that your child can voluntarily participate in the study. If you choose not to allow your child to participate in this study, there will be no penalty or loss of benefits to which you or your child are entitled to. If you agree to let your child take part, you are free to withdraw your child from it at any time. Likewise, no penalty or loss of benefits to which you or your child are otherwise entitled to will occur.

The research study, including the above-mentioned information, has been verbally described to me. I agree to let my child voluntarily participate in the study, as set out above.

……. ……..

Signature of parent or legal guardian Date

Appendix E: Example of the assent form signed by minor participants

PARTICIPANT INFORMATION LEAFLET AND ASSENT FORM

TITLE OF THE RESEARCH PROJECT: Development of a new test to detect a mutation in people with haemophilia A

RESEARCHERS NAME(S): Mr JF Kloppers, Dr. WJ Janse van Rensburg and Dr. GM Marx

ADDRESS: Department of Haematology and Cell Biology

Francois Retief Building, Second Floor, Room 404

University of the Free State

DF Malherbe avenue

Bloemfontein

CONTACT NUMBER: 051 405 3116/051 405 3098/ 079 343 9870

What is RESEARCH?

Research is something we do to find new knowledge about the way things (and people) work. We use research projects or studies to help us find out more

about disease or illness. Research also helps us to find better ways of helping, or treating children who are sick.

What is this research project all about?

In our research project we look at the disease known as Haemophilia A. We are looking to develop a new test that will be able to find a specific mutation (fault) that is the reason that you have haemophilia A. This new test will be much faster and will be much easier to do than other tests that are currently used.

Why have I been invited to take part in this research project?

You have been invited to participate in this project because you have haemophilia A

Who is doing the research?

The research is being done by Jean Kloppers, who is a student at the medical school at the University of the Free State. The research project is being supervised by Dr Janse van Rensburg and Dr Marx who is also working for the University of the Free State.

What will happen to me in this study?

A doctor will draw a little blood from you. This is all that will happen to you in this study

Can anything bad happen to me?

There is a risk that you might feel a bit of pain where the blood is drawn, and it might cause a bruise. If you feel any pain or discomfort during or after blood is drawn you should tell the doctor or your parents about it immediately.

Can anything good happen to me?

There is no direct benefit for you in this study. Your participation will, however, help to get a better understanding about haemophilia A and in future help to develop a possible cure.

Will anyone know I am in the study?

Only you, your parents and the researchers will know that you are in this study. Nobody else will know that you are in the study unless you tell them.

Who can I talk to about the study?

If you have any questions about this study, you can contact Jean Kloppers directly on 051 405 3116 or 079 343 9870. Emails can be sent to jeanklpprs@hotmail.com. Alternatively, contact Dr Janse van Rensburg on 051 405 3098.

What if I do not want to do this?

You can refuse to take part in this study, even if your parents agree that you can participate. You can also withdraw from the study at any time you want, and you will not get in trouble for doing so.

Do you understand this research study and are you willing to take part in it?

YES	NO

Has the researcher answered all your questions?

YES	NO

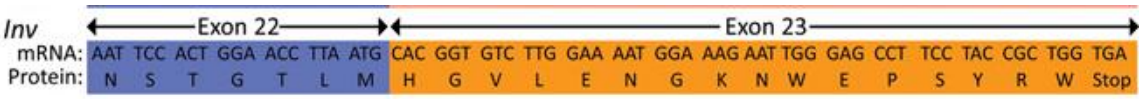
Do you understand that you can pull out of the study at any time?

YES	NO

---------- ----------

Signature of Child Date

Appendix F: Inv22 mRNA reference sequence



The mRNA sequence for the inv22 genotype as predicted by Pandey *et al.* (2013) as translated from the inv22 protein sequence.

5'-
TCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTCTTGATGGGAAGAAGT
GGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGCACGGTGTCTTGGAAA
ATGGAAAGAATTGGGAGCCTTCCTA-3'

The mRNA reference sequence for the inv22 genotype as determined in this study. The highlighted sequence matches the sequence as predicted by Pandey *et al.* (2013).

Appendix G: RNA and cDNA concentrations

RNA concentrations

Sample	RNA concentration 1 (ng/µL)	RNA concentration 2 (ng/µL)	Average RNA concentration
C1	215	220	218
C2	260	264	262
C3	245	244	245
P1	203	208	206
P2	220	221	221
P3	255	259	257
P4	205	208	207
P5	251	254	253
P6	201	206	204
P7	249	248	249
P8	239	241	240
P9	223	225	224
P10	220	215	218
P11	248	250	249
P12	181	178	180
P13	250	245	248
P14	50	46	48
P15	187	188	188
P16	210	210	210

P17	236	231	234
P18	222	226	224
P19	310	315	313
P20	240	242	241
P21	131	132	132
P22	229	230	230
P23	121	125	123
P24	225	227	226
P25	234	235	235
P26	239	238	239
P27	179	181	180
P28	225	228	227
P29	285	291	288
P30	171	171	171
P31	231	233	232
P32	176	179	176
P33	165	160	163
P34	227	225	226
P35	179	174	177
P36	227	225	226
P37	229	228	229
P38	220	221	221
P39	230	234	232

P40	221	224	223
P41	151	148	150
P42	152	150	151
P43	101	103	102
P44	222	225	224
P45	281	286	284
P46	260	262	261
P47	235	235	235
P48	181	181	181
P49	180	175	178
P50	203	205	204
P51	210	211	211
P52	179	175	177
P53	121	125	123
P54	200	201	201
P55	146	148	147
P56	160	164	162
P57	168	170	169
P58	201	200	201
P59	221	222	222
P60	146	143	145

cDNA concentrations:

Sample	cDNA concentration 1 (ng/µL)	cDNA concentration 2 (ng/µL)	Average cDNA concentration
C1	1857	1856	1857
C2	1985	1988	1987
C3	1998	1994	1996
P1	1096	1100	1098
P2	1400	1398	1399
P3	1358	1358	1358
P4	1600	1658	1629
P5	1402	1410	1406
P6	1102	1110	1106
P7	1050	1055	1053
P8	1802	1812	1807
P9	1405	1409	1407
P10	1659	1657	1658
P11	1987	1990	1989
P12	1966	1961	1964
P13	1708	1700	1704
P14	500	494	497
P15	1986	1980	1983
P16	1875	1879	1877
P17	1903	1894	1899

P18	1405	1399	1402
P19	1874	1871	1872.5
P20	1989	1987	1988
P21	1965	1961	1963
P22	1655	1657	1656
P23	1290	1286	1288
P24	1987	1988	1988
P25	1603	1608	1606
P26	1988	1990	1989
P27	1400	1407	1404
P28	1927	1921	1924
P29	1155	1174	1165
P30	1803	1816	1810
P31	1987	1996	1992
P32	1380	1388	1384
P33	1365	1367	1366
P34	1835	1826	1831
P35	1407	1399	1403
P36	1126	1120	1123
P37	1654	1657	1656
P38	1889	1888	1889
P39	1875	1876	1876
P40	534	560	547

P41	1291	1284	1288
P42	1835	1827	1831
P43	1525	1524	1525
P44	1698	1691	1695
P45	1204	1218	1211
P46	1254	1249	1252
P47	1544	1550	1547
P48	1499	1490	1495
P49	1500	1492	1496
P50	1878	1878	1878
P51	1392	1403	1398
P52	1048	1051	1050
P53	1989	1988	1989
P54	1654	1658	1656
P55	1452	1460	1456
P56	1986	1992	1989
P57	1966	1978	1972
P58	1698	1690	1694
P59	1887	1880	1884
P60	1455	1449	1452

Appendix H: Inv22 conventional RT-PCR results for the 60 study participants. The table includes the haemophilia A status, inv22 detection and inv22 status.

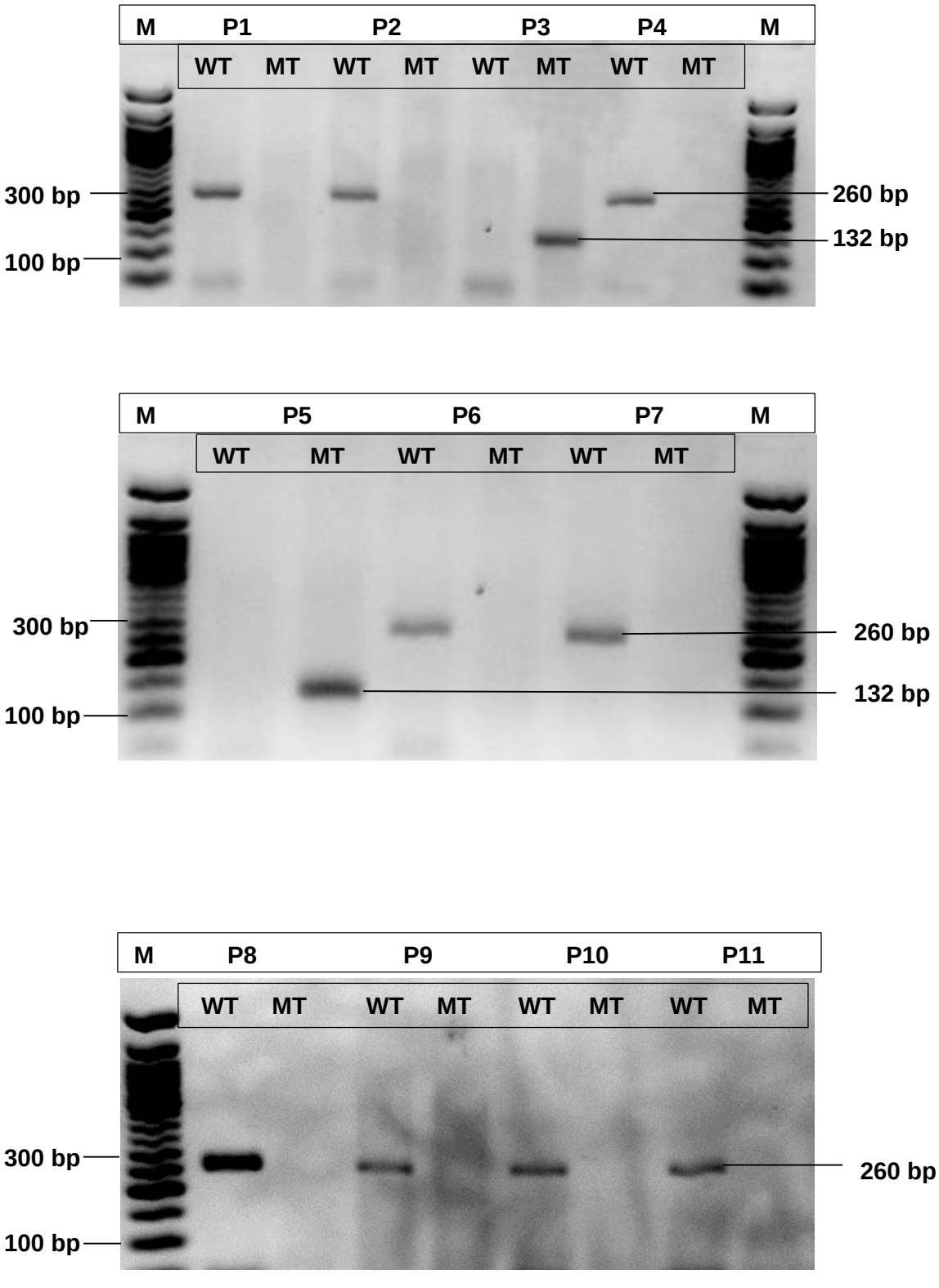
Study participant	Haemophilia A status	Wild-type detected	Mutant-type detected	Inv22 status
P1	Severe	Detected	Not detected	Negative
P2	Possible carrier	Detected	Not detected	Negative
P3	Severe	Not detected	Detected	Positive
P4	Possible carrier	Detected	Not detected	Negative
P5	Severe	Not detected	Detected	Positive
P6	Possible carrier	Detected	Not detected	Negative
P7	Severe	Detected	Not detected	Negative
P8	Severe	Detected	Not detected	Negative
P9	Severe	Detected	Not detected	Negative
P10	Mild	Detected	Not detected	Negative
P11	Severe	Detected	Not detected	Negative
P12	Severe	Detected	Not detected	Negative
P13	Severe	Detected	Not detected	Negative
P14	Possible carrier	Detected	Detected	Carrier
P15	Severe	Detected	Not detected	Negative
P16	Severe	Detected	Not detected	Negative
P17	Possible carrier	Detected	Not detected	Negative
P18	Severe	Detected	Not detected	Negative
P19	Severe	Not detected	Detected	Positive
P20	Possible carrier	Detected	Detected	Carrier

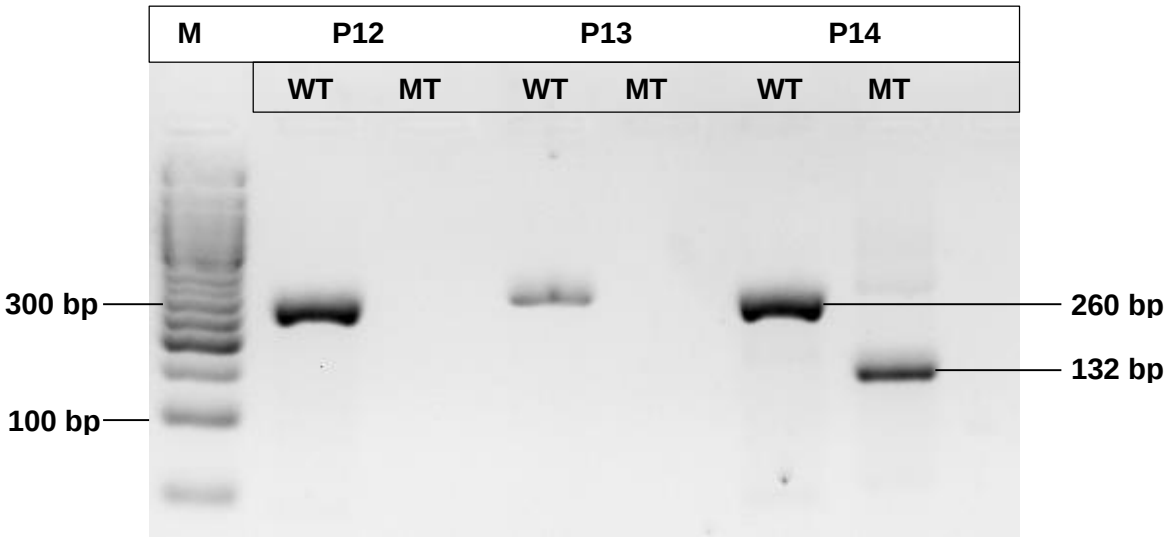
P21	Severe	Not detected	Detected	Positive
P22	Severe	Not detected	Detected	Positive
P23	Severe	Not detected	Detected	Positive
P24	Possible carrier	Detected	Not detected	Negative
P25	Severe	Detected	Not detected	Negative
P26	Possible carrier	Detected	Not detected	Negative
P27	Severe	Detected	Not detected	Negative
P28	Possible carrier	Detected	Not detected	Negative
P29	Severe	Detected	Not detected	Negative
P30	Severe	Not detected	Detected	Positive
P31	Possible carrier	Detected	Not detected	Negative
P32	Severe	Not detected	Detected	Positive
P33	Possible carrier	Detected	Not detected	Negative
P34	Severe	Detected	Not detected	Negative
P35	Possible carrier	Detected	Detected	Carrier
P36	Severe	Detected	Not detected	Negative
P37	Severe	Detected	Not detected	Negative
P38	Possible carrier	Detected	Not detected	Negative
P39	Severe	Detected	Not detected	Negative
P40	Severe	Detected	Not detected	Negative
P41	Severe	Detected	Not detected	Negative
P42	Severe	Detected	Not detected	Negative
P43	Possible carrier	Detected	Not detected	Negative

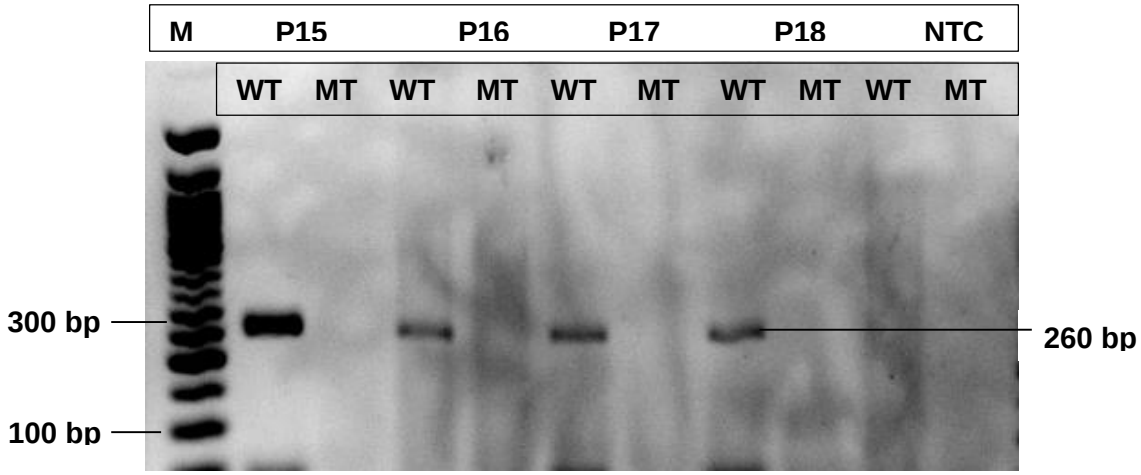
P44	Possible carrier	Detected	Not detected	Negative
P45	Possible carrier	Detected	Not detected	Negative
P46	Severe	Detected	Not detected	Negative
P47	Possible carrier	Detected	Not detected	Negative
P48	Severe	Detected	Not detected	Negative
P49	Mild	Detected	Not detected	Negative
P50	Severe	Detected	Not detected	Negative
P51	Severe	Detected	Not detected	Negative
P52	Mild	Detected	Not detected	Negative
P53	Possible carrier	Detected	Detected	Carrier
P54	Non-haemophilic volunteer	Detected	Not detected	Negative
P55	Non-haemophilic volunteer	Detected	Not detected	Negative
P56	Non-haemophilic volunteer	Detected	Not detected	Negative
P57	Non-haemophilic volunteer	Detected	Not detected	Negative
P58	Non-haemophilic volunteer	Detected	Not detected	Negative

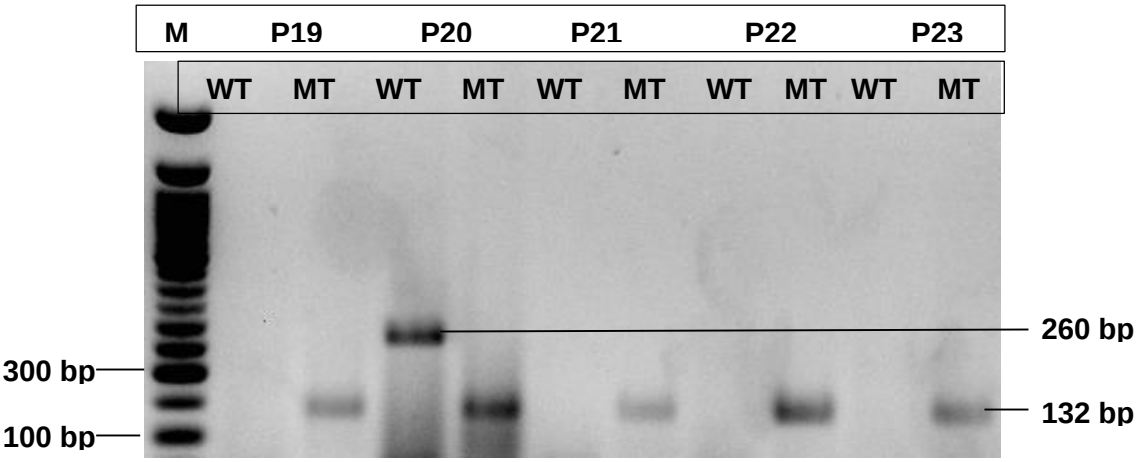
P59	Non-haemophilic volunteer	Detected	Not detected	Negative
P60	Non-haemophilic volunteer	Detected	Not detected	Negative

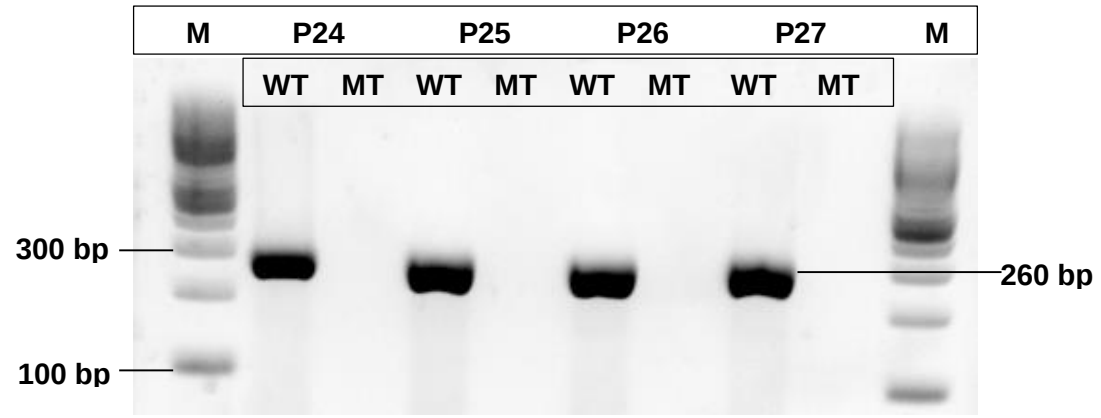
Gel electrophoresis images of the 60 study participants screened with the inv22 conventional RT-PCR detection method.

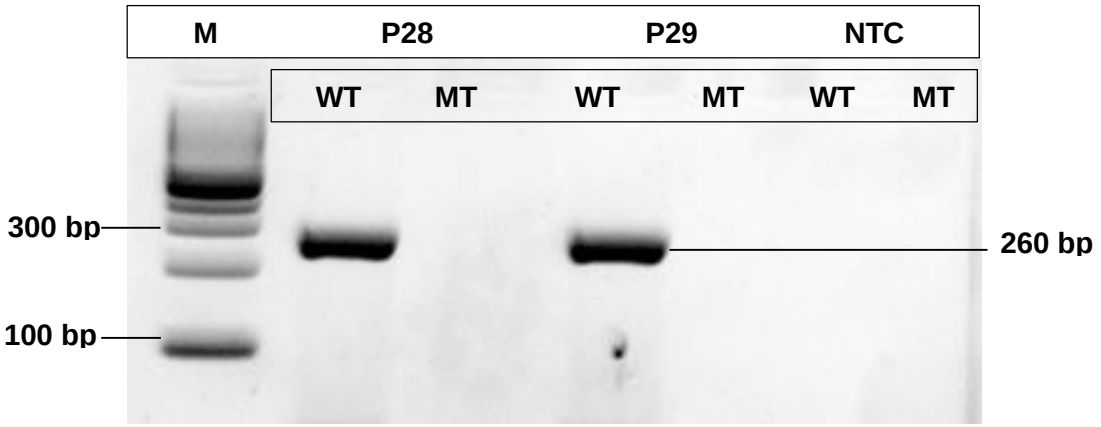


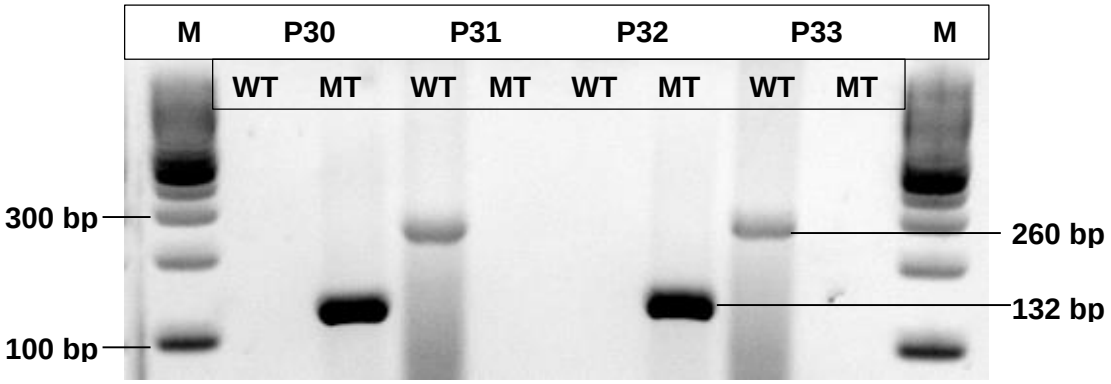

M
P12
P13
P14
WT
MT
WT
MT
WT
MT
300 bp
100 bp
260 bp
132 bp

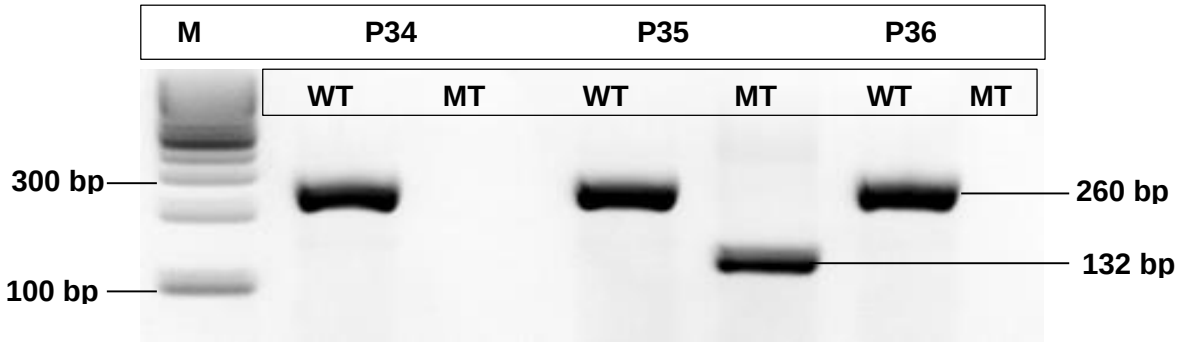

M
P15
P16
P17
P18
NTC
WT
MT
WT
MT
WT
MT
WT
MT
WT
MT
300 bp
100 bp
260 bp

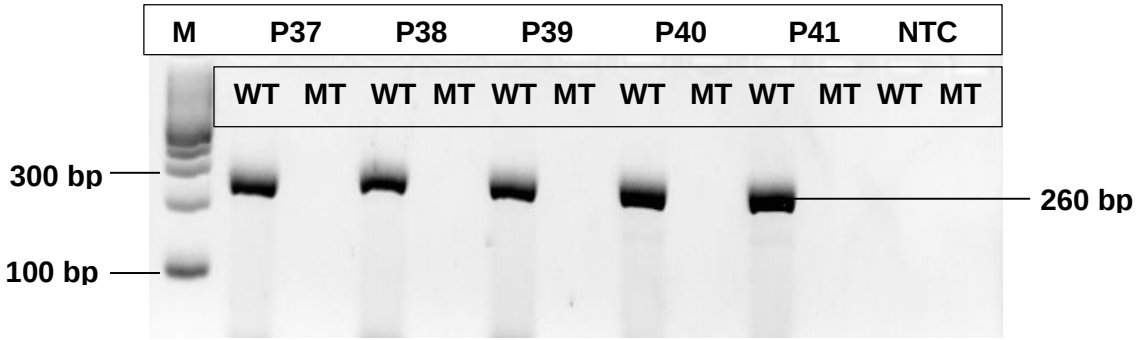

M
P19
P20
P21
P22
P23
WT
MT
WT
MT
WT
MT
WT
MT
WT
MT
300 bp
100 bp
260 bp
132 bp

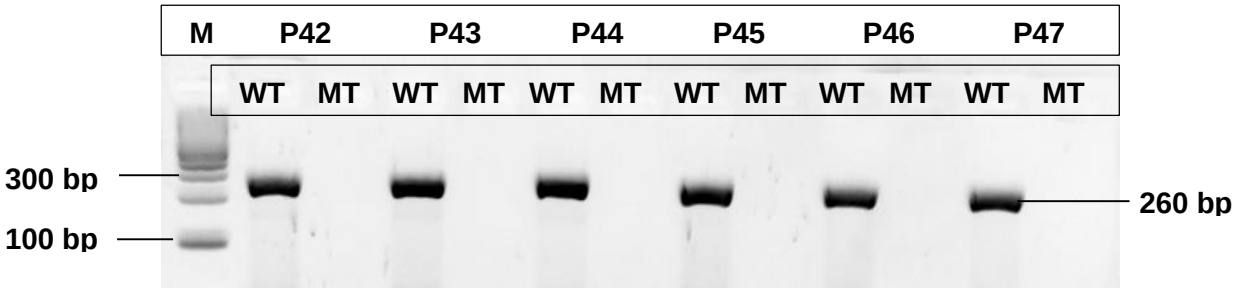

M
P24
P25
P26
P27
M
WT
MT
WT
MT
WT
MT
WT
MT
300 bp
100 bp
260 bp

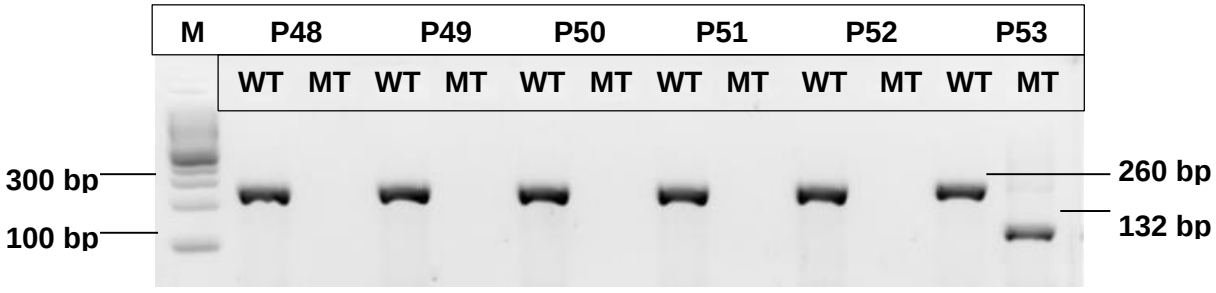

M
P28
P29
NTC
WT
MT
WT
MT
WT
MT
300 bp
100 bp
260 bp

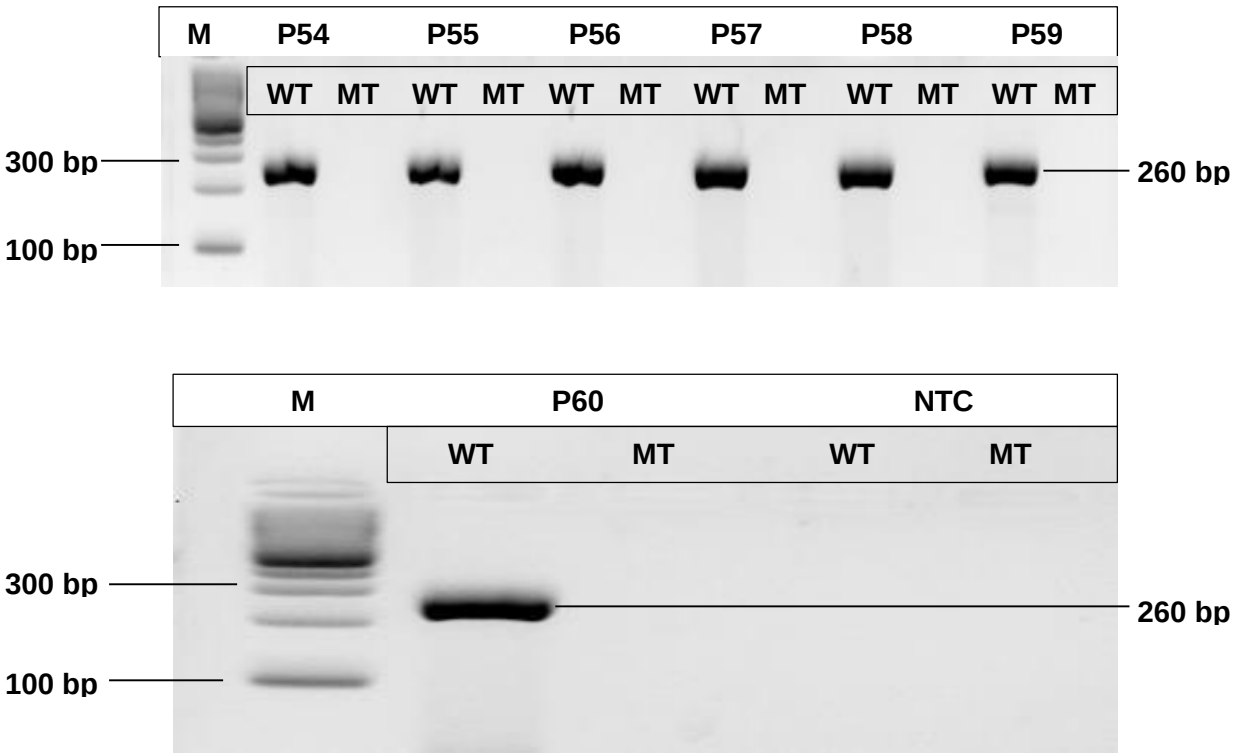

M
P30
P31
P32
P33
M
WT
MT
WT
MT
WT
MT
WT
MT
300 bp
100 bp
260 bp
132 bp


M
P34
P35
P36
WT
MT
WT
MT
WT
MT
300 bp
100 bp
260 bp
132 bp


M
P37
P38
P39
P40
P41
NTC
WT MT WT MT WT MT WT MT WT MT WT MT
300 bp
100 bp
260 bp


M
P42
P43
P44
P45
P46
P47
WT MT WT MT WT MT WT MT WT MT WT MT
300 bp
100 bp
260 bp


M
P48
P49
P50
P51
P52
P53
WT MT WT MT WT MT WT MT WT MT WT MT
300 bp
100 bp
260 bp
132 bp


M
P54
P55
P56
P57
P58
P59
WT MT WT MT WT MT WT MT WT MT WT MT
300 bp
100 bp
260 bp
M
P60
NTC
WT
MT
WT
MT
300 bp
100 bp
260 bp

Appendix I: Inv22 real-time RT-PCR results for the 60 study participants.

Study participant	Haemophilia A status	Wild-type detected	Mutant-type detected	Average Ct value	SD	Inv22 status
P1	Severe	Detected	Not detected	35.269	1.242	Negative
P2	Possible carrier	Detected	Not detected	34.863	0.062	Negative
P3	Severe	Not detected	Detected	33.498	0.467	Positive
P4	Possible carrier	Detected	Not detected	35.054	0.101	Negative
P5	Severe	Not detected	Detected	33.832	0.303	Positive
P6	Possible carrier	Detected	Not detected	34.771	0.774	Negative
P7	Severe	Detected	Not detected	33.685	0.401	Negative
P8	Severe	Detected	Not detected	36.0125	0.007	Negative
P9	Severe	Detected	Not detected	34.872	0.128	Negative
P10	Mild	Detected	Not detected	33.733	0.172	Negative
P11	Severe	Detected	Not detected	34.738	0.470	Negative
P12	Severe	Detected	Not detected	34.318	1.135	Negative
P13	Severe	Detected	Not detected	34.211	0.310	Negative
P14	Possible carrier	Not detected	Not detected	-	-	Unknown

P15	Severe	Detected	Not detected	35.299	0.147	Negative
P16	Severe	Detected	Not detected	34.724	0.810	Negative
P17	Possible carrier	Detected	Not detected	34.603	0.375	Negative
P18	Severe	Detected	Not detected	34.467	1.059	Negative
P19	Severe	Not detected	Detected	32.882	0.668	Positive
P20	Possible carrier	Detected	Detected	32.722/34.338	0.052/0.601	Carrier
P21	Severe	Not detected	Detected	36.095	0.332	Positive
P22	Severe	Not detected	Detected	34.663	0.311	Positive
P23	Severe	Not detected	Detected	36.417	0.615	Positive
P24	Possible carrier	Detected	Not detected	34.327	0.200	Negative
P25	Severe	Detected	Not detected	34.256	1.085	Negative
P26	Possible carrier	Detected	Not detected	33.9875	0.650	Negative
P27	Severe	Detected	Not detected	34.431	0.008	Negative
P28	Possible carrier	Detected	Not detected	34.317	0.260	Negative
P29	Severe	Detected	Not detected	31.730	0.679	Negative
P30	Severe	Not detected	Detected	36.250	0.240	Positive

P31	Possible carrier	Detected	Not detected	34.438	0.455	Negative
P32	Severe	Not detected	Detected	35.436	0.581	Positive
P33	Possible carrier	Detected	Not detected	36.650	0.212	Negative
P34	Severe	Detected	Not detected	34.485	0.257	Negative
P35	Possible carrier	Detected	Detected	36.585/33.650	0.001/1.037	Carrier
P36	Severe	Detected	Not detected	33.809	0.846	Negative
P37	Severe	Detected	Not detected	34.045	0.340	Negative
P38	Possible carrier	Detected	Not detected	34.582	0.674	Negative
P39	Severe	Detected	Not detected	34.536	0.478	Negative
P40	Severe	Detected	Not detected	34.053	0.095	Negative
P41	Severe	Detected	Not detected	34.095	0.562	Negative
P42	Severe	Detected	Not detected	37.764	0.313	Negative
P43	Possible carrier	Detected	Not detected	38.685	0.194	Negative
P44	Possible carrier	Detected	Not detected	34.665	0.532	Negative
P45	Possible carrier	Detected	Not detected	31.100	0.212	Negative
P46	Severe	Detected	Not detected	33.926	0.390	Negative

P47	Possible carrier	Detected	Not detected	34.660	0.196	Negative
P48	Severe	Detected	Not detected	35.810	0.410	Negative
P49	Mild	Detected	Not detected	35.075	1.014	Negative
P50	Severe	Detected	Not detected	33.953	0.186	Negative
P51	Severe	Detected	Not detected	35.650	0.283	Negative
P52	Mild	Detected	Not detected	35.095	0.163	Negative
P53	Possible carrier	Detected	Detected	35.913/32.503	0.337/0.639	Carrier
P54	Non-haemophilic volunteer	Detected	Not detected	34.2365	0.013	Negative
P55	Non-haemophilic volunteer	Detected	Not detected	33.063	0.042	Negative
P56	Non-haemophilic volunteer	Detected	Not detected	33.665	0.478	Negative
P57	Non-haemophilic volunteer	Detected	Not detected	33.127	0.866	Negative
P58	Non-haemophilic volunteer	Detected	Not detected	34.009	0.130	Negative
P59	Non-haemophilic volunteer	Detected	Not detected	32.476	0.317	Negative

P60	Non-haemophilic volunteer	Detected	Not detected	34.559	0.014	Negative

Appendix J: Sanger sequence analysis (LALIGN) results for the 60 study participants

Inv22 negative study participants have a sequence that aligns with the wild-type mRNA reference sequence (NM_000132.3) and inv22 positive study participants have a sequence that aligns with the mutant mRNA reference sequence **(Appendix F)**. Carriers of the inv22 have both the wild-type and the mutant-type sequences, respectively.

```
>>P1 205 bp                                                (205 nt)

               10        20        30        40        50        60
Wild   TATTCACGGCATCAAGACCCAGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P1     TATTCACGGCATCAAGACCCAGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCA
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   GTTTATCATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCAC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P1     GTTTATCATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCAC
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TGGAACCTTAATGGTCTTCTTTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P1     TGGAACCTTAATGGTCTTCTTTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTT
              130       140       150       160       170       180

              190       200
Wild   TAACCCTCCAATTATTGCTCGATAC
       :::::::::::::::::::::::::
P1     TAACCCTCCAATTATTGCTCGATAC
```

```
>>P2 220 bp                                                    (220 nt)
               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P2     AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P2     TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P2     TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       :::::::::::::::::::::::::::::::::::::::
P2     GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
            190       200       210       220
```

```
>P3 86 bp                                                      (86 nt)

             10        20        30        40        50        60
Mutant CATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P3     CATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTT
            10        20        30        40        50        60

             70        80
Mutant AATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::
P3     AATGCACGGTGTCTTGGAAAATG
            70        80
```

```
>>P4 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P4     AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P4     TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P4     TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       :::::::::::::::::::::::::::::::::::::::
P4     GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>P5 86 bp                                                      (86 nt)

             10        20        30        40        50        60
Mutant CATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P5     CATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTT
            10        20        30        40        50        60

             70        80
Mutant AATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::
P5     AATGCACGGTGTCTTGGAAAATG
            70        80
```

```
>>P6 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P6     AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P6     TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P6     TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       :::::::::::::::::::::::::::::::::::::::
P6     GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>P7 116 bp                                                     (116 nt)

               10        20        30        40        50        60
Wild   TATTCACGGCATCAAGACCCAGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P7     TATTCACGGCATCAAGACCCAGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCA
               10        20        30        40        50        60

               70        80        90       100       110
Wild   GTTTATCATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATT
       :::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P7     GTTTATCATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATT
               70        80        90       100       110
```

```
>>P8 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P8     AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P8     TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P8     TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

                190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
     ::::::::::::::::::::::::::::::::::::::::
P8     GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>P9 202 bp                                                    (202 nt)

               10        20        30        40        50        60
Wild   TATTCACGGCATCAAGACCCAGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P9     TATTCACGGCATCAAGACCCAGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCA
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   GTTTATCATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCAC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P9     GTTTATCATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCAC
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TGGAACCTTAATGGTCTTCTTTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P9     TGGAACCTTAATGGTCTTCTTTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTT
              130       140       150       160       170       180

              190       200
Wild   TAACCCTCCAATTATTGCTCGA
       ::::::::::::::::::::::
P9     TAACCCTCCAATTATTGCTCGA
              190       200
```

```
>>P10 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P10    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P10    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P10    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P10    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>P11 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P11    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P11    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P11    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P11    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>P12 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P12    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P12    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P12    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P12    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P13 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P13    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P13    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P13    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P13    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P14 220 bp                                                    (220 nt)


              10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P14    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
              10        20        30        40        50        60

              70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P14    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
              70        80        90       100       110       120

             130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P14    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
             130       140       150       160       170       180

             190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P14    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
             190       200       210       220
```

```
>P14 86 bp                                                      (86 nt)

             10        20        30        40        50        60
Mutant CATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P14    CATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTT
            10        20        30        40        50        60

             70        80
Mutant AATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::
P14    AATGCACGGTGTCTTGGAAAATG
            70        80
```

```
>>P15 220 bp                                                    (220 nt)


               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P15    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P15    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P15    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P15    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P16 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P16    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P16    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P16    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P16    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>P17 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P17    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P17    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P17    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P17    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>P18 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P18    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P18    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P18    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       :::::::::::::::::::::::::::::::::::::::
P18    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P19 85 bp                                                    (85 nt)

               10        20        30        40        50        60
Mutant ATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P19    ATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACC
               10        20        30        40        50        60

               70        80
Mutant TTAATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::::
P19    TTAATGCACGGTGTCTTGGAAAATG
               70        80
```

```
>>P20 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P20    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P20    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P20    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       :::::::::::::::::::::::::::::::::::::::
P20    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P20 85 bp                                                     (85 nt)

               10        20        30        40        50        60
Mutant ATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P20    ATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACC
               10        20        30        40        50        60

               70        80
Mutant TTAATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::::
P20    TTAATGCACGGTGTCTTGGAAAATG
               70        80
```

```
>>P21 85 bp                                                (85 nt)

               10        20        30        40        50        60
Mutant ATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P21    ATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACC
               10        20        30        40        50        60

               70        80
Mutant TTAATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::::
P21    TTAATGCACGGTGTCTTGGAAAATG
               70        80
```

```
>>P22 85 bp                                                (85 nt)

               10        20        30        40        50        60
Mutant ATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P22    ATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACC
               10        20        30        40        50        60

               70        80
Mutant TTAATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::::
P22    TTAATGCACGGTGTCTTGGAAAATG
               70        80
```

```
>>P23 86 bp                                                (86 nt)

            10        20        30        40        50        60
Mutant ATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P23     ATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTA
           10        20        30        40        50        60

            70        80
Mutant ATGCACGGTGTCTTGGAAAATG
       ::::::::::::::::::::::
P23     ATGCACGGTGTCTTGGAAAATG
           70        80
```

```
>>P24 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P24    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P24    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P24    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P24    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P25 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P25    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P25    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P25    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P25    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P26 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P26    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P26    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P26    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P26    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P27 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P27    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P27    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P27    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P27    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P28 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P28    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P28    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P28    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P28    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P29 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P29    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P29    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P29    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       :::::::::::::::::::::::::::::::::::::::
P29    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P30 86 bp                                                     (86 nt)

            10        20        30        40        50        60
Mutant ATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P30    ATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTA
           10        20        30        40        50        60

            70        80
Mutant ATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::
P30    ATGCACGGTGTCTTGGAAAATG
           70        80
```

```
>>P31 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P31    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P31    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P31    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       :::::::::::::::::::::::::::::::::::::::
P31    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P32 86 bp                                                     (86 nt)

            10        20        30        40        50        60
Mutant ATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P32    ATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTA
           10        20        30        40        50        60

             70        80
Mutant ATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::
P32    ATGCACGGTGTCTTGGAAAATG
           70        80
```

```
>>p33 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P33    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P33    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P33    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P33    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p34 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P34    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P34    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P34    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       :::::::::::::::::::::::::::::::::::::::
P34    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p35 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P35    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P35    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P35    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       :::::::::::::::::::::::::::::::::::::::
P35    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>P35 86 bp                                                    (86 nt)

           10        20        30        40        50        60
Mutant ATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P35    ATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTA
           10        20        30        40        50        60

            70        80
Mutant ATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::
P35    ATGCACGGTGTCTTGGAAAATG
           70        80
```

```
>>p36 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P36    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P36    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P36    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P36    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p37 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P37    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P37    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P37    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P37    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p38 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P38    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P38    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P38    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P38    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p39 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P39    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P39    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P39    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P39    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p40 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P40    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P40    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P40    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P40    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p41 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P41    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P41    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P41    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P41    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p42 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P42    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P42    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P42    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P42    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p43 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P43    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P43    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P43    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P43    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p44 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P44    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P44    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P44    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P44    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p45 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P45    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P45    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P45    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P45    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p46 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P46    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P46    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P46    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P46    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p47 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P47    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P47    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P47    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P47    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p48 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P48    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P48    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P48    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P48    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p49 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P49    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P49    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P49    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P49    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p50 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P50    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P50    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P50    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P50    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p51 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P51    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P51    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P51    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       :::::::::::::::::::::::::::::::::::::::
P51    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p52 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P52    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P52    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P52    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P52    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p53 220 bp                                                    (220 nt)


              10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P53    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
              10        20        30        40        50        60

              70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P53    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
              70        80        90       100       110       120

             130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P53    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
             130       140       150       160       170       180

             190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P53    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
             190       200       210       220
```

```
>>P53 86 bp                                                     (86 nt)

            10        20        30        40        50        60
Mutant ATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P53    ATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTA
           10        20        30        40        50        60

            70        80
Mutant ATGCACGGTGTCTTGGAAAATG
       :::::::::::::::::::::::
P53    ATGCACGGTGTCTTGGAAAATG
           70        80
```

```
>>p54 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P53    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P54    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P54    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P54    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p55 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P55    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P55    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P55    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P55    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p56 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P56    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P56    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P56    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P56    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p57 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P57    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P57    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P57    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P57    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p58 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P58    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P58    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P58    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P58    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p59 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P59    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P59    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P59    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P59    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

```
>>p60 220 bp                                                    (220 nt)

               10        20        30        40        50        60
Wild   AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P60    AGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTC
               10        20        30        40        50        60

               70        80        90       100       110       120
Wild   TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P60    TTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCT
               70        80        90       100       110       120

              130       140       150       160       170       180
Wild   TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P60    TTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTC
              130       140       150       160       170       180

              190       200       210       220
Wild   GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
       ::::::::::::::::::::::::::::::::::::::::
P60    GATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAG
              190       200       210       220
```

LALIGN figure keys: p: participant. nt: nucleotides. Wild: Wild-type reference sequence. Mutant: Mutant-type reference sequence.

Appendix K: The presence of FVIII inhibitors in the central South African haemophilia A population

Haemophilia A patient	Haemophilia A status	Inv22 status	History of FVIII inhibitor development
C2	Severe	Positive	Yes
C3	Severe	Positive	No
P1	Severe	Negative	No
P3	Severe	Positive	No
P5	Severe	Positive	Yes
P7	Severe	Negative	No
P8	Severe	Negative	Yes
P9	Severe	Negative	No
P10	Mild	Negative	No
P11	Severe	Negative	Yes
P12	Severe	Negative	No
P13	Severe	Negative	No
P15	Severe	Negative	No
P16	Severe	Negative	No
P18	Severe	Negative	No
P19	Severe	Positive	No
P21	Severe	Positive	No
P22	Severe	Positive	No
P23	Severe	Positive	No

P25	Severe	Negative	No
P27	Severe	Negative	Yes
P29	Severe	Negative	Yes
P30	Severe	Positive	Yes
P32	Severe	Positive	No
P34	Severe	Negative	No
P36	Severe	Negative	No
P37	Severe	Negative	No
P39	Severe	Negative	No
P40	Severe	Negative	No
P41	Severe	Negative	Yes
P42	Severe	Negative	No
P46	Severe	Negative	No
P48	Severe	Negative	Yes
P49	Mild	Negative	No
P50	Severe	Negative	No
P51	Severe	Negative	No
P52	Mild	Negative	No

Appendix L: Publication: Rapid identification of the intron 22 inversion in haemophilia A (Kloppers and Janse van Rensburg, 2017).

Haemophilia
The Official Journal of the World Federation of Hemophilia,
European Association for Haemophilia and Allied Disorders and
the Hemostasis & Thrombosis Research Society

Rapid identification of the intron 22 inversion in haemophilia A

JEAN F. KLOPPERS and WALTER J. JANSE VAN RENSBURG
Department of Haematology and Cell Biology, Faculty of Health Sciences, University of the Free State, Bloemfontein, South Africa

Various mutations are associated with haemophilia A. The mutational heterogeneity and the size of the gene complicate molecular studies [1]. It has been reported that approximately 45% of individuals with severe haemophilia A has an inversion in intron 22 of the FVIII gene [2]. The detection of mutations plays an important part in predicting treatment outcome, as FVIII mutations have been associated with FVIII inhibitor development [3,4]. Furthermore, mutation detection allows for carrier and prenatal diagnosis to be conducted [5].

The intron 22 inversion (inv22) was first detected by means of a Southern blot assay [2]. This technique proves to be accurate, but has the disadvantage of being time-consuming and labour intensive [6]. The availability of Southern blot reagents has also become a major obstacle. Newer methods such as long-distance polymerase chain reaction (LD-PCR) and inverse PCR (I-PCR) have been described to detect the inversion [7,8]. However, LD-PCR is reportedly difficult to standardize and is sensitive to reductions in DNA quality [6,9]. I-PCR involves three steps including restriction, self-ligation and standard PCR analysis. Although I-PCR seems to be reliable, it is time-consuming and requires multiple steps [6]. We had trouble implementing the I-PCR method in our laboratory at the Universitas Academic Hospital in Bloemfontein, South Africa. With the multiple steps involved, without a quality check step built into the method, troubleshooting proved to be almost impossible in our suboptimally resourced setting. DNA-based analysis of the inversion is also complicated by the size of the FVIII gene (186 kb), intron 22 (32.4 kb) and the homologous sequence involved in the inversion (9.5 kb) [10,11]. Therefore, our aim was to develop a method that is not only effective in detecting inv22 but also more rapid and cost-effective.

It has been reported that individuals with the intron 22 inversion produce two FVIII exon-containing mRNAs, namely FVIII22I and FVIIIB respectively. These two FVIII exon-containing mRNAs express the full-length FVIII amino acid sequence, but are non-secretory polypeptide chains [12]. The amplification between the boundary of exons 22 and 23 of the FVIII gene in the mutant phenotype (intron 22 inversion) yields no product, as it is separated from each other [13]. However, it was found that the FVIII22I transcript contained an additional exon, termed exon 23c. Sequence analysis revealed that exon 23c expressed 16 amino acids that were adjacent to the amino acids expressed for by exon 22 of the wild-type FVIII gene [12]. Subsequently, using this sequence, we designed primers that amplify a region that cover the splice site between the wild-type exon 22 and exon 23, as well as primers that amplify the splice site between the wild-type exon 22 and mutant exon 23c (Table 1).

We selected two non-related severe haemophilia A patients (one black African and one Caucasian) who have been confirmed to be inv22 positive with Southern blot to set up the method. A normal healthy volunteer was used as control. Subsequently, we tested 10 severe haemophilia A patients with an unknown inv22 status, as well as two putative carriers also with an unknown status (mothers of patient 3 and patient 6 respectively). The fractionated white blood cells of each sample was stored in RNAlater (Ambion, Life Technologies, Waltham, Massachusetts, USA) solution and RNA was extracted using the RiboPure™-Blood kit (Ambion, Life Technologies, Waltham, Massachusetts, USA). The mRNA was converted into cDNA using the High Capacity RNA-to-cDNA Kit (Applied Biosystems, Foster City, California, USA). Two separate PCR reactions were set up for each sample, one using wild-type exon 22 forward and exon 23 reverse primers, and one using a wild-type exon 22 forward and a mutant exon 23c reverse primer. A temperature gradient PCR was performed in order to determine the optimal annealing temperature (Data S1). The PCR product was loaded on a 2% agarose gel containing ethidium bromide for electrophoresis. After electrophoresis, the fragments were visualized under UV light using a Kodak Gel Logic 212 gel documentation system (Carestream Molecular Imaging, Rochester, NY, USA).

Correspondence: Walter Janse van Rensburg, Department of Haematology and Cell Biology (G2), Faculty of Health Sciences, University of the Free State, 205 Nelson Mandela Drive, 9301 Bloemfontein, South Africa.
Tel.: +27 51 405 3098; fax: +27 51 444 1036; e-mail: jansevrwj@ufs.ac.za

Accepted after revision 7 November 2016

DOI: 10.1111/hae.13142