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*

MUTATION DETECTION IN THE ENDOGLIN GENE IN A FAMILY WITH HEREDITARY HAEMORRHAGIC TELANGIECTASIA

K.T. PETA

***MAGISTER SCIENTIAE*: HUMAN GENETICS**

MUTATION DETECTION IN THE ENDOGLIN GENE IN A FAMILY WITH HEREDITARY HAEMORRHAGIC TELANGIECTASIA

by

Kimberly Thando Peta

Submitted in fulfilment of the requirements for the degree
***MAGISTER SCIENTIAE*: HUMAN GENETICS**

In the Department of Genetics
in the Faculty of Natural and Agricultural Sciences
at the University of the Free state

December 2017

Supervisor: Dr. Gerda Marx
Co-supervisor: Prof. MJ Coetzee

ACKNOWLEGEMENTS

I first thank God for giving me health and all that was necessary for me to complete this degree. I thank my mum for being supportive, loving, caring and understanding throughout this journey. I thank my family for their support. I thank Dr. Gerda Marx for her supervision, intellect, advice, her hands on approach in this project and making it fun. I thank Prof. Marius Coetzee for his large contribution to this project and giving us much priority despite his busy schedule. I thank the National Research Foundation (NRF) and the Deans Merit Bursary for their financial assistance for the past two years. Lastly, I thank Dr. Martin Nyanga for his assistance on the practical part of the project.

TABLE OF CONTENTS

Abbreviations

A	Adenine
ACB	African caribbean
aCGH	Comparative genomic hybridization array
ACVRL1	*Activin A receptor, type IIlike kinase 1*
AD	Autosomal dominant
AJs	Adherens junctions
ALK1	*Activin receptorlike kinase 1*
Ang 1	Angiopoietin 1
AVMs	Arteriovenous malformations
bFGF	basic Fibroblast growth factor
BMP	Bone morphogenetic protein
BMP9	*Bone morphogenetic 9*
BMPR1	Bone morphogenetic protein receptor type 1
BMPR2	Bone morphogenetic protein receptor type 2
C	Cytosine
CAVM	Cerebral arteriovenous malformation
CCT	Cardiovascular computed tomography
CNS	Central nervous system
CTAB	Cetyl trimethylammonium bromide
CXCR3	Chemokine receptor 3
CYT	Cytoplasmic
dbSNP	Single nucleotide polymorphism database
EC	Endothelial cell
ECs	Endothelial cells
ECM	Extracellular matrix

ENG	*Endoglin*
eNOS	endothelial Nitric oxide synthase
ESN	Esan
FN	Fibronectin
FSS	Fluid shear stress
G	Guanine
GDF2	Growth differentiation factor 2
GI	Gastrointestinal tract
GJs	Gap junctions
H2O	Water
HAVM	Hepatic arteriovenous malformations
HHT	Hereditary haemorrhagic telangiectasia
HHT1	Hereditary haemorrhagic telangiectasia subtype 1
HHT2	Hereditary haemorrhagic telangiectasia subtype 2
HSREC	Health sciences research ethics committee
IPAH	Idiopathic pulmonary arterial hypertension
JAM	Junction adhesion molecules
JP	Juvenile polyposis
JPHHT	Juvenile polyposis hereditary haemorrhagic telangiectasia
L	Long
L*ENG*	Long *endoglin*
LOH	Loss of heterozygosity
LWK	Luhya
MDCT	Multi detector computed tomography
MEGA	Molecular evolutionary genetic analysis
MMPs	Membrane metalloproteases
MT11	Membrane type metalloprotease1

MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
MSL	Mende
NCBI	National center for biotechnology information
NGS	Next generation sequencing
NO	Nitric oxide
NOS	Nitric oxide synthase
NOS1	endothelial Nitric oxide synthase 1
NOS2	endothelial Nitric oxide synthase 2
NOS3	endothelial Nitric oxide synthase 3
OMIM	Online mendelian inheritance in man
PAVM	Pulmonary arteriovenous malformations
PCR	Polymerase chain reaction
PDGF	Platelet derived growth factor
PDZ	Postsynaptic density 95/Drosophila disk large/zonula occludens1
PH	Pulmonary hypertension
PPH	Primary pulmonary hypertension
RASA1	*RAS p21 protein activator 1*
RCA	Right coronary artery
RGD	ArgGlyAsp
RSMAD	Receptor SMAD
RSPO3	RSpondin 3
S	Short
S*ENG*	Short *endoglin*
s*ENG*	Soluble *endoglin*
Ser	Serine
SIFT	Sorting intolerant from tolerant

SMAD4	SMAD family member 4
SMCs	Smooth muscle cells
TAE	Tris base, acetic acid and EDTA
TGFβ	Transforming growth factor beta
TGFβI	Transforming growth factor beta type 1
TGFβII	Transforming growth factor beta type 2
TGF βR	Transforming growth factor beta receptor
TGFβR I	Transforming growth factor beta receptor type 1
TGFβR II	Transforming growth factor beta receptor type 2
T	Thymine
Thr	Threonine
TJs	Tight junctions
TM	Transmembrane
VCE	Video capsule endoscopy
VEcadherin	Vascular endothelial cadherin
VEGF	Vascular endothelial growth factor
VN	Vitronectin
VSMCs	Vascular smooth muscular cell
YRI	Yoruba
ZO	Zona occludens
ZP	Zona pellucida

Abstract

Introduction: Hereditary haemorrhagic telangiectasia (HHT) is a rare autosomal dominant bleeding disorder. It is characterised by the presence of mucocutaneous telangiectases, visceral arteriovenous malformations and epistaxis. The phenotype is diagnosed according to the Curaçao criteria. At a molecular level, HHT has been linked to the *Endoglin, Activin kinase 1 (ALK1)* and *SMAD4* genes in numerous studies. The majority of studies are centred on European and American population groups. There are few publications of HHT on people of African descent, none of which are family based studies. To our knowledge, this is the first study presenting mRNA expression sequence data of the *Endoglin (ENG)* gene in a population of African descent. The aim of the study is to detect splice site and exon region mutations present in the *ENG* gene of the family members affected with HHT. **Methodology**: RNA was isolated from blood, stabilised in RNA*later* and stored at -20ºC using the Ribopure blood kit and TRizol® methods. The RNA was converted to cDNA that served as a template molecule for sequence mutation detection. In total 11 primer pairs were designed and used to amplify 15 exon regions of the *endoglin* gene. Sanger sequencing was employed to determine *ENG* mutations. **Results**: Four mutations were identified, two in exon 1, namely c.-324A>G and c.-207G>A and two missense mutations c.640G>A and c.1510G>A located in exon 5 and exon 11 were identified, respectively. The exon 1 mutation c.-324A>G is a population variant. The c.-207G>A exon 1 mutation was concluded to have no effect on the resulting amino acid and only one HHT individual harboured this mutation. The missense mutations c.640G>A and c.1510G>A were previously described in participants with and without HHT in literature, resulting in conflicting interpretations regarding HHT causality. Results from this study indicate that the latter mutations occurred in two different individuals that have been diagnosed with HHT. **Conclusion**: The identified mutations were present in individuals who were formerly diagnosed with HHT but none of them can be proven to be pathogenic, since it was not present in all the HHT affected family members. Future studies should focus on the mutation detection of other HHT associated genes such as *SMAD4, ALK1, BMP9* and *RASA1* genes in this family to decipher HHT pathogenesis in a family from African descent.

Key words: Genetics, splice site, mutations, HHT, vasculogenesis, angiogenesis, haemorrhagic, telangiectases, AVM, Africa.

CHAPTER 1

INTRODUCTION

There are numerous disorders and diseases that occur due to shared phenotypic characteristics and shared molecular causes. Vascular disorders such as cardiovascular disease and hereditary haemorrhagic telangiectasia (HHT) share symptomatic characteristics while the pathophysiology differs. The *endoglin* (*ENG*) gene seems to be involved in several diseases such as atherosclerosis, diabetes, systemic sclerosis, sickle cell anaemia and cancer (López-Novoa & Bernabeu, 2010).

Mutations in the *endoglin* gene have proven to be involved in the pathogenesis of the aforementioned conditions. One of the disorders that arise from mutations in the *ENG* gene is hereditary haemorrhagic telangiectasia (HHT), a rare autosomal dominant vascular disorder. It is characterised by the presence of telangiectases and arteriovenous malformations (AVM). Patients suffering from HHT bleed due to the weakening of the vascular system where the veins and arteries are directly connected without the presence of capillaries. The bleeding leads to additional complications such as anaemia.

Numerous studies around the world have found various mutations in the *ENG* gene that contribute to the pathogenesis of HHT. Other genes have been described that are involved in the pathogenesis of HHT such as *activin receptor kinase 1 (ALK1*) and SMAD *family member 4* (*SMAD4*). All these genes are involved in the transforming growth factor β (TGF-β) pathway that controls the angiogenesis or anti-angiogenesis of endothelial cells.

HHT occurs in all races and ethnic groups, but most of the research has been reported on Caucasian individuals living in first world countries such as United States of America and Europe. Therefore information about the genetic causes of HHT in

individuals of African descent is limited. HHT is a rare vascular disorder that is often misdiagnosed or under-diagnosed, especially in the African continent with limited health resources, consequently very little research is performed on HHT in Africa.

A study on the genetic mutations associated with HHT on African people is thus crucial to add to the body of knowledge about HHT to enable the screening of potential pathogenic HHT manifestations. With this study the aim is to improve our understanding of HHT in general, to determine if *ENG* gene mutations identified in Caucasian populations are causative in a South African family and possibly find novel pathogenic mutations unique to HHT patients from African descent. Research on the *ENG* gene might not only benefit HHT patients but also cancer patients that receive radiotherapy presenting with similar vascular damage (Scharpfenecker *et al*., 2009).

In time, it will aid in genetic screening of families with subclinical phenotypes, to enable earlier treatment intervention and a reduction in patient discomfort. Knowledge about causal HHT gene mutations could further be incorporated into genetic counselling. Genetic counselling helps those affected and their families understand the disease pathology in regards to how the possible mutations cause the disorder, as well as to enable informed decisions regarding family planning.

CHAPTER 2

LITERATURE

STUDY

Introduction

Hereditary haemorrhagic telangiectasia (HHT) also known as Osler-Weber-Rendu syndrome, is a rare autosomal dominant bleeding disorder (Guttmacher *et al.*, 1995; Gallione *et al*., 2000; Westermann *et al*., 2003; Bayrak-Toydemir *et al*., 2004; Chen *et al*., 2013). It is characterised by the presence of mucocutaneous telangiectases, visceral arteriovenous malformations and epistaxis. Males and females are affected equally (Shovlin, 2010). The autosomal dominant pattern of inheritance, implies that an affected parent has a 50% chance of passing on the disorder to each of his/her children (Shovlin, 2010). HHT is currently linked to two genes namely *endoglin* (*ENG*) and *activin receptor-like kinase 1* (*ALK1*), also known as *activin A receptor, type II-like kinase 1* (*ACVRL1*) (Canzonieri *et al*., 2014a). However, there are other genes that play a significant role in the aetiology of telangiectasia such as SMAD family member 4 (*SMAD4*). Mutations in the *ENG* gene cause HHT1 (Online Mendelian Inheritance in Man (OMIM) 187300) and mutations in the *ALK1* gene cause HHT2 (OMIM 600376). Bayrak-Toydemir *et al.* (2004) identified *ENG, ALK1* and *SMAD4* mutations in many patients from various racial and ethnic groups from different locations around the world (Gallione *et al*., 2000). These genes mediate signalling in the transforming growth factor-β (TGF-β) pathway of vascular endothelial cells (Canzonieri *et al*., 2014a).

History of HHT

HHT was first described more than a century ago by Henry Gawen Sutton (1864). In 1896, Henri Jules Rendu recognized a combination of telangiectases and hereditary epistaxis. A decade later, William Osler and Frederic Parkes Weber produced numerous case reports on the condition. Hence the eponymous name of the syndrome is Osler-Weber-Rendu syndrome. In 1909 Hanes coined the term

hereditary haemorrhagic telangiectasia because of the familial nature of the disease, the characteristic telangiectases and haemorrhage observed in individuals affected (Guttmacher *et al*., 1995).

Prevalence

HHT is widely distributed and occurs more often than previously thought. Guttmacher *et al*. (1995) reported a prevalence of 1 in 2 351 people in France. The prevalence in Denmark is 1 in 6 410 (Westermann *et al*., 2003), rising to 1 in 3 500 on the island of Funen. The prevalence in Vermont (USA) was found to be 1 in 16 500, on Leeward Islands in the Caribbean it is 1 in 5 155, and in northern England 1 in 39 213 (Guttmacher *et al.*, 1995). The world-wide estimated prevalence is 1 or 2 per 100 000 individuals, but it has been suggested that the prevalence of HHT could be much higher than stated (Gallione *et al*., 2000; Canzonieri *et al*., 2014a; Tual-Chalot *et al*., 2015). Canzonieri *et al*. (2014a) and Tual-Chalot *et al.* (2015) actually suggested a prevalence of 1 per 5000 – 8000 individuals due to founder effects.

More than 80% of HHT cases have been diagnosed with either HHT1 or HHT2 based on observed *ENG* and *ALK1* gene mutations (Tual-Chalot *et al*., 2015).

Diagnosis

Clinical presentation of HHT can be heterogeneous because it tends to affect numerous organs (Guttmacher *et al*., 1995; Gallione *et al*., 2000). HHT is diagnosed using the Curaçao criteria, which incorporates the common symptoms associated with this disorder (Table 1) (Sharathkumar & Shapiro 2008). Since HHT involves multiple organ systems, it requires clinical and scientific professionals from various disciplines to manage the disorder (Shovlin, 2010) including geneticists, radiologists,

otolaryngologists, haematologists, dermatologists, pulmonologists, psychiatrists, neurologists and cardiologists (Sabba, 2005).

The presence of telangiectases and arteriovenous malformations (described later in the chapter) causes various organs such as the lungs, brain and liver to develop haemorrhages. For an individual to be diagnosed with HHT they have to meet three symptoms in the Curaçao criteria (Table 1). Should an individual's symptoms meet only two criteria, further investigations are warranted before the diagnosis can be made. If only one criterion is met, it is unlikely that the individual tested will be diagnosed with HHT. Given the autosomal dominant pattern of inheritance, the family history also needs to be investigated, hence being the fourth criterion (Grand'Maison, 2009).

TABLE 1: CURAÇAO CRITERIA FOR HEREDITARY HAEMORRHAGIC TELANGIECTASIA DIAGNOSES

Criteria	Description
Epistaxis	Spontaneous and recurrent
Telangiectases	Multiple, at characteristic sites: lips, oral cavity, fingers, nose
Visceral lesions	Gastrointestinal telangiectasia, pulmonary, hepatic, cerebral or spinal arteriovenous malformations
Family history	A first-degree relative with HHT according to these criteria

HHT diagnosis

Definite: 3 or more criteria present

Possible: 2 criteria present

Unlikely: < 2 criteria present

Source: Sharathkumar & Shapiro 2008

Vascular biology

Vascular structure formation

In order to understand vascular pathology involved with HHT, it is essential to first explain the vascular structure formation. The development of endothelial cells (ECs), depends on continuous differentiation of mesodermal cells into haemangioblasts. Haemagioblasts then lead to the formation of vascular structures called primitive blood islands. In the centre of the islands, haemangioblasts differentiate into haematopoietic stem cells, while EC precursors (angioblasts) differentiate from peripheral haemangioblasts (Lamalice *et al.,* 2007; Goumans *et al*., 2009).

Newly formed ECs migrate on a matrix made of hyaluronan and collagen, which allows the merging of blood islands. These coalesce leading to the remodelling of tubular structures, and the formation of the first primitive vascular plexus. The tubules remodel into large vessels through vasculogenesis, which then leads to embryo vascularization. Vasculogenesis is defined as the differentiation of angioblasts into ECs and the formation of a novel vascular network (Lamalice *et al.,* 2007; Goumans *et al*., 2009).

During vascular sprouting, when ECs migrate and proliferate, junctions are partly disordered enabling vascular permeability (Figure 1a). After interaction with pericytes, the vessels become stabilized and junction integrity is revived and permeability is tightly controlled (Figure 1b). Subsequently, apoptosis and cell proliferation are inhibited (Dejana, 2004). EC migration depends on firmly regulated signalling cascades that are activated by several stimuli (Sáez *et al.,* 2014) such as injury (Ammann *et al.,* 2015). During angiogenesis, sufficient signalling events are necessary for appropriate vessels remodelling. Vascular dysfunction would result

from slanted intracellular cross-talk in the involved pathway (Sáez *et al.,* 2014) such a pathway include the transforming growth factor beta (TGF-β) pathway that can either inhibit or promote angiogenesis (Goumans *et al*., 2009).

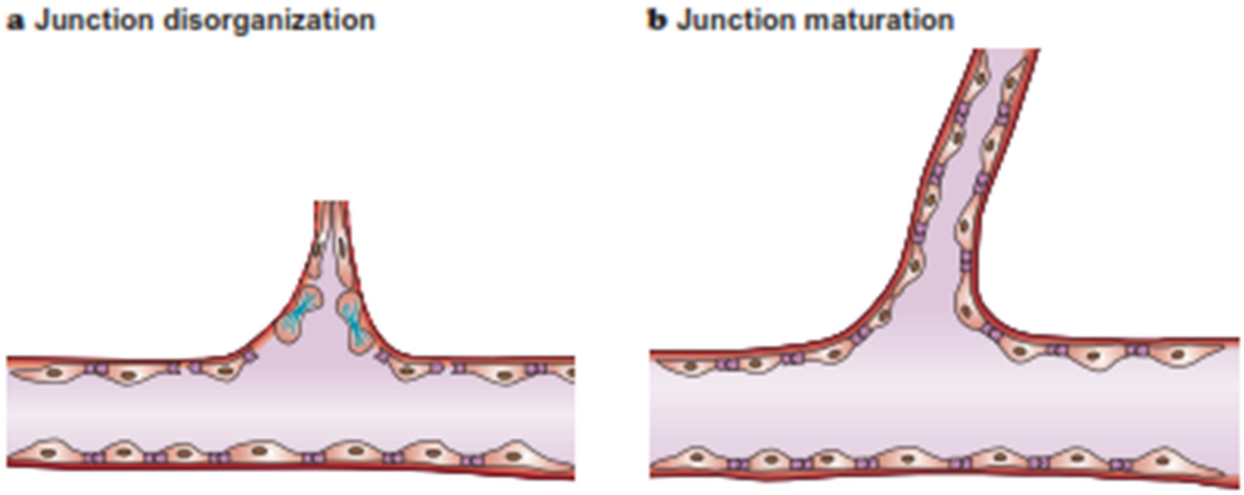


Figure 1: Modulation of junctions in angiogenesis. (a) During vascular sprouting, junctions are partly disordered enabling vascular permeability when ECs migrate and proliferate. (b) After interaction with pericytes, the vessels become stabilized and junction integrity is revived and permeability is tightly controlled. Subsequently, apoptosis and cell proliferation are inhibited (Dejana, 2004).

Angiogenesis

Angiogenesis is the production of new blood vessels from pre-existing blood vessels (Figure 2). The production of angiogenic growth factors such as placental growth factor, vascular endothelial growth factor (VEGF), angiopoietin-1 and inhibitors of differentiation cytokines and proteins, initiates the formation of vessels. Following the binding of these factors to their specific receptors on ECs, EC migration, proliferation and capillary morphogenesis is promoted. These processes are stabilized by the interaction and recruitment with smooth muscle cells (SMCs) and pericytes (López-Novoa & Bernabeu 2010; Park *et al.,* 2015) forming strong vessel walls that constitutes large vessels (Goumans *et al*., 2009). The interaction between EC

surface integrins and extracellular matrix (ECM) proteins (that fill up the extracellular matrix) and cells such as pericytes and SMCs regulates cell migration. The process of forming tube-like networks between ECs is called capillary morphogenesis. Both cell migration and capillary morphogenesis play an important role in vessel formation. ECs migrate to the origin of promigratory signals by means of actin filament formation and focal adhesions (López-Novoa & Bernabeu 2010; Park *et al.,* 2015). When the endothelial cells have mutations in the *ENG* gene that affect protein function, angiogenesis is disrupted causing abnormal tube formation. (Fernedez *et al.,* 2006).

When injury occurs, the endothelial cells detach from their basement membrane, new micro vessels are formed which migrate and proliferate in the interstitial stroma. Three dimensional gels show that endothelial cells grown *in vitro* in the presence of TGF-β, form tube-like cellular aggregates with tight junctions and lumen, which mimic vessel formation. Cell surface integrins and the surrounding extracellular matrix influence the response of endothelial cells to TGF-β (McAllister *et al*., 1994).

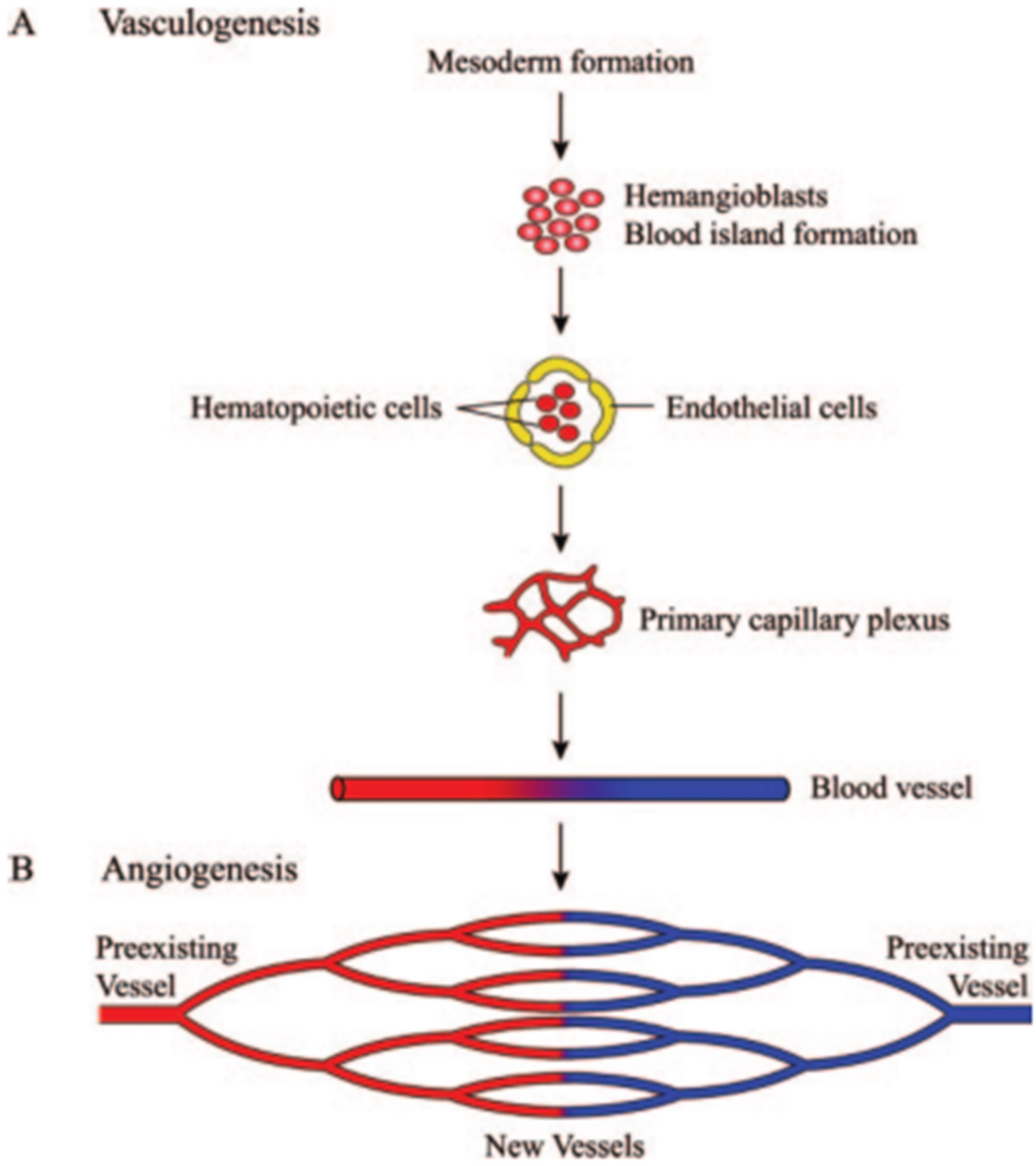


Figure 2: Genesis of the vascular system. A: Primitive blood islands are formed from mesodermal cells differentiating into haemangioblasts. Subsequently, angioblasts (precursors of ECs) are formed from the differentiation of peripheral haemangioblasts. After chemotactic and hypotactic activation, ECs migrate and merge to blood islands. These coalesce leading to the remodelling of tubular structures and formation of first primitive vascular plexus. The tubules remodel into large vessels through vasculogenesis leads to embryo vascularization. B: Angiogenesis is the production of new blood vessels from pre-existing blood vessels (Lamalice *et al*., 2007).

The regulation of Nitric Oxide production

Nitric oxide (NO) a signalling molecule is involved in the regulation of angiogenesis, endothelial permeability, vascular remodelling, vascular tone (Park *et al*., 2015) and cell migration (Lamalice *et al*., 2007). NO is produced from L-arginine by nitric oxide synthase (NOS). There are three major NOS isoforms; neuronal NOS (NOS1), inducible NOS (NOS2) and most abundantly endothelial NOS (eNOS or NOS3). NO plays a part in the regulation physiological processes during angiogenesis such as proliferation, cell survival and migration. The activation of eNOS is regulated by the phosphorylation of serine (Ser) and threonine (Thr) (Park *et al*., 2015). In ENG deficient cells, eNOS is decreased although it is increased in cells overexpressing ENG (Bernabeu *et al.,* 2007).

Fluid shear stress

Endothelial cells are exposed to fluid shear stress (FSS) of the blood stream due to blood pressure and flow, since they line the interior blood vessels (Figure 3). FSS influence EC physiology, morphology, gene expression as well as migratory pathways (Lamalice *et al*., 2007; Chiu & Chien, 2011; Park *et al*., 2015). The shear stress in the venous system of humans ranges from 1 to 6 dyn/cm² and 10 -70 dyn/cm² in the arteries. Normal stress is perpendicular to ECs surface, while shear stress is parallel to EC surface (Chiu & Chien, 2011). Shear stress causes the disassociation on cell-cell contacts and the activation of actin cytoskeleton remodelling through numerous signalling pathways that trigger EC migration (Lamalice *et al*., 2007). In the endothelium, FSS is the most important physiological stimulus for the activation of eNOS and the production of NO (Park *et al*., 2015).

Damage to the ECs contributes to the pathogenesis of vascular disease, atherosclerosis (Sreejayan & Ren, 2007), cancer and HHT (Dallas *et al*., 2008).

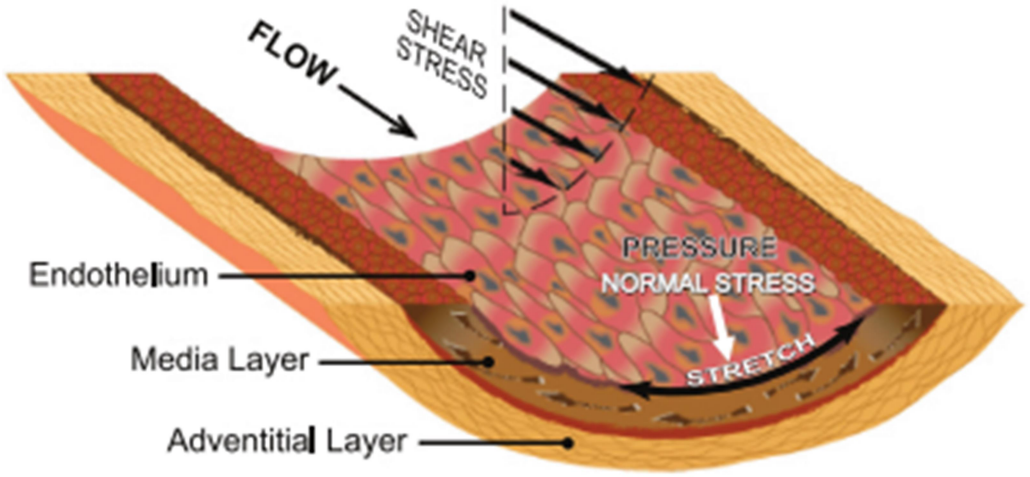


Figure 3: Schematic diagram showing the creation of shear stress (parallel to ECs surface) by blood flow and the creation of normal stress (perpendicular to ECs surface) as well as the action of pressure as a result circumferential stress (Chiu & Chien, 2011).

HHT Manifestations and Treatment

Epistaxis

Epistaxis or nosebleeds, is the most frequent and bothersome symptom of HHT (Guttmacher *et al*., 1995; Bayrak-Toydemir *et al*., 2004; Al-Deen & Bachmann-Harildstad, 2008). Recurrent nosebleeds occur in about 95% of HHT affected people (Bayrak-Toydemir *et al*., 2004) and appears to be more frequent in women than in men with a ratio of 5:1 (Al-Deen & Bachmann-Harildstad, 2008). Nosebleeds may begin at the age of 10 years, however most patients present with epistaxis at the age of 21 years (Guttmacher *et al*., 1995) or 30 years (Al-Deen & Bachmann-Harildstad, 2008). Epistaxis are usually spontaneous and as one increases with age, the more severe and regular nosebleeds occur (Guttmacher *et al*., 1995; Bayrak-Toydemir *et al*., 2004).

The severity of epistaxis varies, and can be classified (Table 2) (Reibez *et al.,* 1995; Al-Deen & Bachmann-Harildstad, 2008). Table 2 illustrates the grading system for epistaxis in HHT that separates the degree of nosebleeds into mild, moderate and severe. It is a single multi-system scale that combines the duration and frequency of a single item and the requirement of a blood transfusion as another item (Al-Deen & Bachmann-Harildstad, 2008). There are similar epistaxis grading systems to the one mentioned above and they are all used currently (Al-Deen & Bachmann-Harildstad 2008). The internationally accepted grading of epistaxis in HHT is the Epistaxis Severity Scoring tool (Hoag *et al.,* 2010). This takes into account frequency, duration of epistaxis, intensity, the need for medical attention, the presence of anaemia, and the need for blood transfusions over preceding three months. Additionally, this grading system is the most commonly used.

Epistaxis leads to chronic iron deficiency anaemia (Bayrak-Toydemir *et al*., 2004) and intolerance to physical activity. Some of the patients with severe anaemia (haemoglobin <7 g/dL) due to severe epistaxis may require blood transfusions (Bayrak-Toydemir *et al*., 2004). They also benefit from intravenous iron dosing (e.g. iron dextran). Fortunately, the severity of epistaxis is not severe enough to warrant medical treatment to treat anaemia (Bayrak-Toydemir *et al*., 2004).

TABLE 2: GRADING OF EPISTAXIS

Severity of epistaxis	Epistaxis frequency	Number of transfusions
Mild	Few episodes per week	None
Moderate	1-2 time per day	<10/ lifetime
Severe	Daily epistaxis lasting greater than 30 min	>10/ lifetime

Copied from (Al-Deen & Bachmann-Harildstad 2008).

Telangiectases

Telangiectases are small fragile convoluted vascular lesions that are found mainly in the nasal cavity, lips (Figure 4C), nose, conjunctivas (Figure 4A), fingertips, tongue (Figure 4B) and the mucosa of gastrointestinal tract, palate, face, trunk, and arms (Bayrak–Toydemir *et al*., 2004). Telangiectases typically present later than epistaxis (Guttmacher *et al*., 1995). These lesions are friable and they rupture, eventually leading to anaemia due to continuous bleeding. Numerous HHT patients suffer all their life with this dreadful disease symptom (Tual-Chalot *et al.,*2015). According to Bayrak-Toydemir *et al*. (2004), it may take 5 to 30 years after epistaxis first occurred for numerous telangiectases to appear on the face, oral cavity and hands.

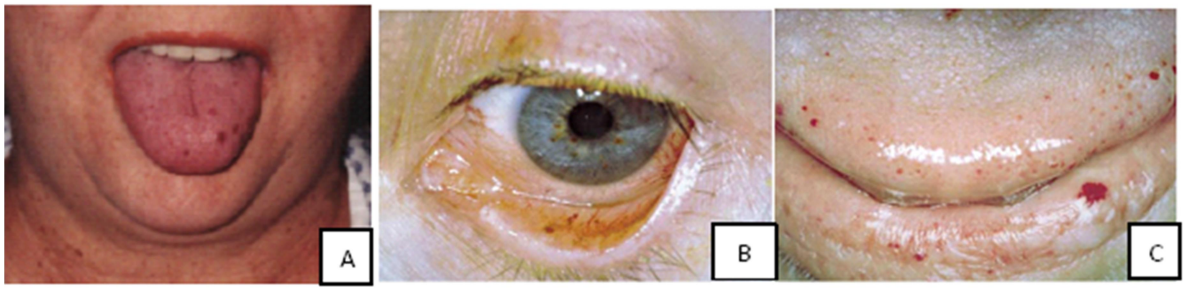


Figure 4: Telangiectases in Oral cavity (A), Conjunctivas (B) and Lips (C) from various patients (Guttmacher *et al*., 1995; Bayrak-Toydemir *et al*., 2004).

Treatment of epistaxis and telangiectases

Nasal lubricants have been found to be helpful for individuals experiencing mild epistaxis (Bayrak-Toydemir *et al*., 2004). Laser ablation was suggested to be the most effective treatment for moderate epistaxis, but cauterisation is a debilitating procedure. No treatment is usually required for skin lesions. However if they do bleed laser treatment is an option (Guttmacher *et al*., 1995; Bayrak-Toydemir *et al*., 2004). Anaemia that resulted from epistaxis or gastrointestinal bleeding can be treated with parenteral or oral iron. Medications such as aspirin and ibuprofen that inhibit blood coagulation should be avoided (Bayrak-Toydemir *et al*., 2004).

Gastrointestinal tract telangiectases

Gastrointestinal (GI) tract telangiectases (Figure 5) is the second most common site of bleeding in patients with HHT after epistaxis (Canzonieri, *et al*., 2014a). Endoscopic techniques may reveal telangiectases in the stomach, duodenum, colon or small intestine that are similar in appearance and size as those found in the nasal mucosa. However, GI bleeding mostly occurs in the stomach and upper duodenum in sites where telangiectases are present (Bayrak-Toydemir *et al*., 2004). GI bleeding occurs in 13-30% of patients suffering from telangiectases (Canzonieri *et al*., 2014a) which begins after the fifth or sixth decades of life (Guttmacher *et al*., 1995).

Canzonieri *et al*. (2014a) used video capsule endoscopy (VCE) and reported that 56-86% of HHT individuals have evenly distributed telangiectases in the entire small bowel (Canzonieri, *et al*., 2014a). GI bleeding in approximately 25% of individuals over 60 years old is associated with anaemia (Bayrak-Toydemir *et al*., 2004). GI bleeding may eventually necessitate transfusion of more than a 100 units of blood over a period of 1 to 3 years (Van Cutsem *et al.,* 1990; Guttmacher *et al*., 1995). The bleeding is gradual but persistent and may escalate with severity and increase with age (Bayrak-Toydemir *et al*., 2004; Canzonieri, *et al*., 2014a). Laser therapy and oestrogen-progesterone treatments are used to treat severe GI bleeding (Kjeldsen *et al.,* 1999).

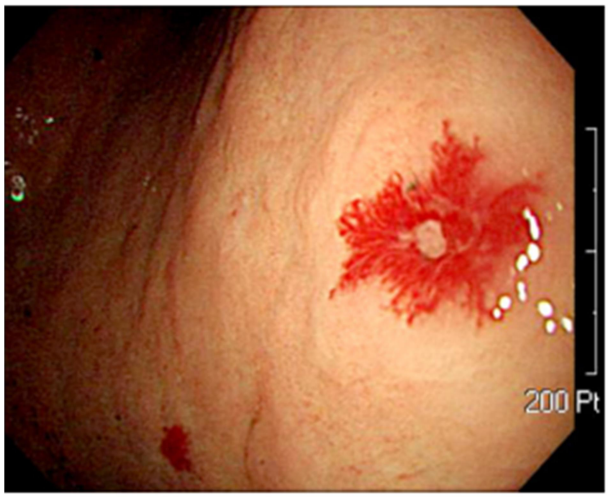
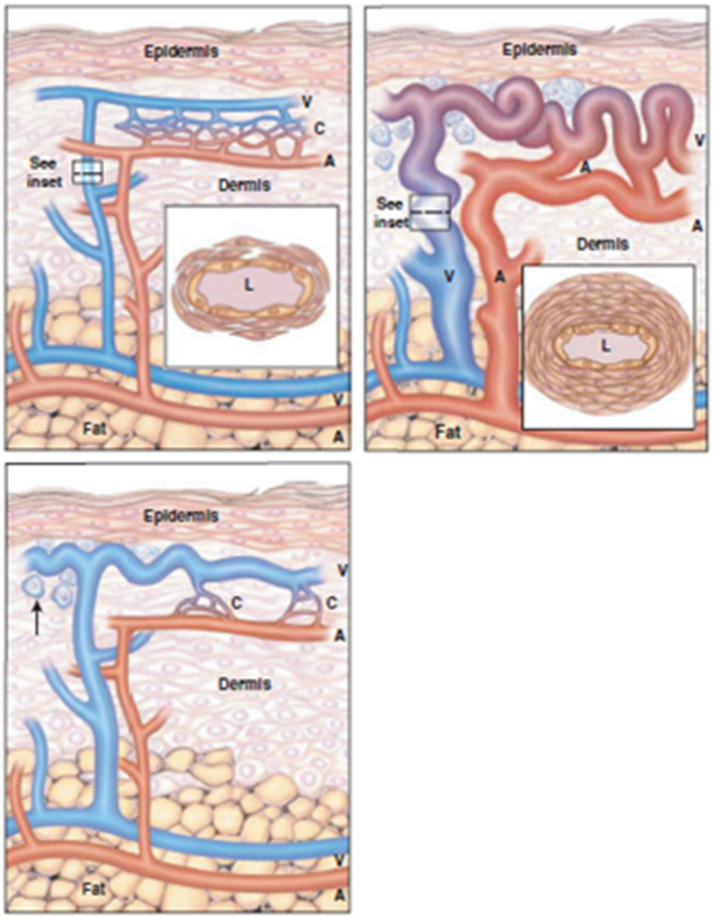


Figure 5: The manifestation of HHT in the GI tract with angiodysplasia in the fundus of the stomach using an endoscopy (Lee *et al*., 2009).

Arteriovenous malformations

Arteriovenous malformations (AVMs) are large connections between arteries and veins that lead to shunting and haemorrhage. AVMs are frequently situated in major organs (Tual-Chalot *et al.,* 2015) such as the lungs, liver, central nervous system (CNS) and upper gastrointestinal tract (Fulbright *et al*., 1998; Bayrak-Toydemir *et al*., 2004). Approximately 10% of patients suffer severe morbidity and die prematurely from AVMs formed in the gastrointestinal tract, CNS and lungs (Fulbright *et al.*, 1998). When formed, the initial phase of the formation of large telangiectases is dilated post-capillary venules (Figure 6). In due course the dilated venules connect to enlarging arterioles through capillary segments. Then the capillary segments disappear allowing the direct connection of the venules and arterioles (McAllister *et al*., 1994; Guttmacher *et al*., 1995) resulting in direct connections between veins and arteries forming the AVMs observed in HHT (Guttmacher *et al*., 1995). Generally AVMs primarily associate with haemorrhage. Thrombi (blood clots) can occur and some of these can embolise to other parts of the body which may cause other secondary complications such as stroke. AVMs usually result from shunting of blood and haemorrhage consequently, causing the key symptoms observed in HHT patients (Bayrak-Toydemir *et al*., 2004).



In normal skin (top panel), arterioles (A) in the papillary dermis are connected to venules (V) through multiple capillaries (C). These vessels arise from larger arterioles and venules at the junction of the dermis and fat. The ultrastructure of a normal postcapillary venule (shown in cross section in the inset) includes the lumen (L), endothelial cells, and two to three layers of surrounding pericytes. In the earliest stage of cutaneous telangiectasia (middle panel), a single venule becomes dilated, but it is still connected to an arteriole through one or more capillaries. A perivascular lymphocytic infiltrate is apparent (arrow). In a fully developed cutaneous telangiectasis (bottom panel), the venule and its branches have become markedly dilated, elongated, and convoluted throughout the dermis. The connecting arterioles have also become dilated and communicate directly with the venules without intervening capillaries. The perivascular infiltrate is still present. The thickened wall of the dilated descending limb (shown in cross section in the inset) contains as many as 11 layers of smooth-muscle cells.

Figure 6: Images adapted from Braverman, Keh, & Jacobson 1990 and repoduced from Guttmacher *et al.,* 1995 and Sharathkumar & Shapiro, 2008, depicting the mechanism that leads to AVM formation.

Central Nervous System Arteriovenous Malformations

Cerebral arteriovenous malformations (CAVMs) (Figure 7) are considered to also be congenital. Approximately 10% of individuals with HHT have CAVMs and these may present at any age with symptoms such as seizures, intracranial haemorrhage, stroke, headache, brain abscess, intracerebral and subarachnoid haemorrhage and transient ischemic attack (Guttmacher *et al*., 1995; Bayrak-Toydemir *et al*., 2004). These symptoms are more common in HHT individuals with a family or personal history of pulmonary arteriovenous malformations (PAVMs). PAVMs are the source of the neurological symptoms experienced in two thirds of the people diagnosed with HHT. In the remaining third, spinal and cerebral AVMs cause seizure, subarachnoid haemorrhage or the less frequent paraparesis. Transient ischemic attack is a neurological dysfunction caused by the loss of blood flow to brain or spinal cord, can progress to ischaemic stroke. PAVMs can result in brain abscesses. The reason for this is that individuals with PAVMs have a right to left shunt that facilitates the passage of bland and septic emboli into the cerebral circulation (Guttmacher *et al.,* 1995). These may be the first symptoms of PAVMs in HHT (Guttmacher *et al*., 1995). The frequency of CAVMs may be underestimated as people who have fatal intracranial haemorrhage may not be diagnosed with HHT (Bayrak-Toydemir *et al*., 2004).

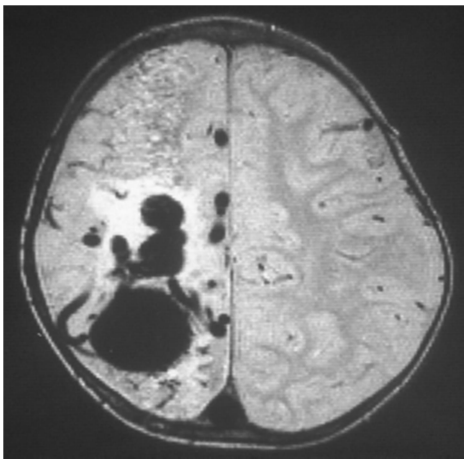


Figure 7: Brain magnetic resonance imaging (MRI) depicting a CAVM of a 9 month old child (Bayrak-Toydemir *et al*., 2004).

The sensitive techniques used to diagnose CAVMs are magnetic resonance imaging (MRI) and angiography. Embolotherapy, neurovascular surgery and stereotactic radiosurgery are all used to treat CAVMs (Guttmacher *et al*., 1995). Transcatheter embolization, stereotactic radiosurgery and resection are used to treat central nervous system (CNS) AVMs frequently in combination. Liver transplantation is the primary form of treatment in symptomatic liver involvement (Bayrak-Toydemir *et al*., 2004).

Pulmonary Arteriovenous Malformations

Pulmonary arteriovenous malformations (PAVMs) are located in the lungs and have significant complications (Guttmacher *et al.*, 1995) due to lack of capillaries (Shin *et al.,* 2010). These complications (mentioned later) are thought to be congenital but may expand after a while. About 70-95% of people with PAVMs also have HHT (Shin *et al.,* 2010). Gallione *et al*. (1998) and Tual-Chalot *et al*. (2015) report that PAVMs are more prevalent in HHT1 than HHT2, occurring in 45% and 8% respectively. *ENG and ALK1* gene products are equally expressed only in the capillaries, distal veins and distal arteries which are consistent with what is observed in PAVMs. However, there is distinct expression in the pulmonary vasculature. The

mechanism by which these lesions form is not properly understood (López-Novoa & Bernabeu, 2010).

PAVMs (Figure 8) may be asymptomatic or may be detected following a dramatic incident such as pulmonary haemorrhage or during exercise, migraine headaches, polycythaemia (the abnormal increase of haemoglobin concentration in the blood) (Guttmacher *et al*., 1995), haemoptysis (blood in cough or in mucus) and haemothorax (collection of blood between lung and chest wall) (Shin *et al.,* 2010). PAVMs are usually multiple and occur in both lungs (Guttmacher *et al*., 1995). Approximately 30% to 40% of individuals with PAVMs will manifest with central nervous system (CNS) symptoms such as brain abscesses (Guttmacher *et al*., 1995; Bayrak-Toydemir *et al*., 2004; Sankelo *et al*., 2008), transient ischaemic attack (interruption of blood supply in the brain) (Shin *et al.,* 2010) which may be a precursor of a stroke (Bayrak-Toydemir *et al*., 2004). About 36% of patients with PAVMs present with cerebral thromboembolic complications (Berg *et al*., 1996).

Pulmonary AVMs do not have a filtering capillary bed (Figure 8) that can eradicate septic or thrombotic emboli. The emboli are free to reach the systemic circulation where they occlude other capillary beds and may cause a disease especially when they reach the cerebral circulation (Guttmacher *et al*., 1995; Sankelo *et al.,* 2008). Furthermore, pulmonary arterial blood passing from the right to the left shunts cannot be oxygenated. This may consequently lead to momentous hypoxemia (Sankelo *et al*., 2008). Fragile vessels may cause haemorrhage into plural cavity or bronchus. Complications from haemorrhaging or shunting may be tragic. However, complications can be avoided when AVMs are diagnosed early (Bossler *et al.,* 2006).

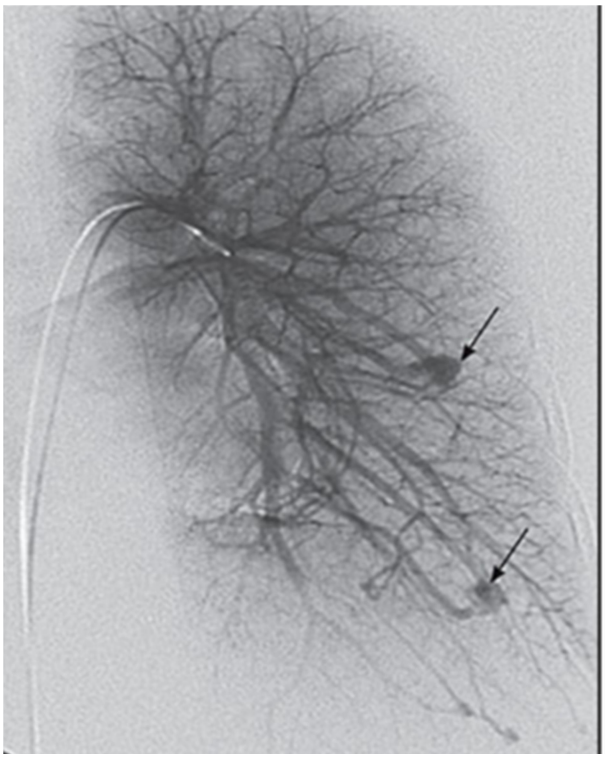


Figure 8: Angiogram of Pulmonary arteriovenous malformations as shown with the arrows (Olitsky, 2010).

Various techniques are used to detect PAVMs such as high-resolution computed tomographic scanning, chest radiography, finger oxymetry and arterial-blood gas measurements. The treatment or surgical management of PAVMs has evolved from lobectomy to wedge resection to arterial supply ligation of the malformation. Other tools used to close the malformations are transcatheter embolotherapy with detachable balloons and stainless-steel coils (Guttmacher *et al*., 1995; Bayrak-Toydemir *et al*., 2004).

Other Arteriovenous Malformations

About 1% of individuals with HHT have spinal AVMs (Figure 9A). Progressive myelopathy, radicular pain and subarachnoid haemorrhage are some of the symptoms. From the time symptoms begin to the time of diagnoses, AVMs are frequently unsuspected. Other AVMs have been located in the vessels of the eye,

coronary arteries (Figure 9B), vagina, urinary tract and spleen (Bayrak-Toydemir *et al*., 2004).

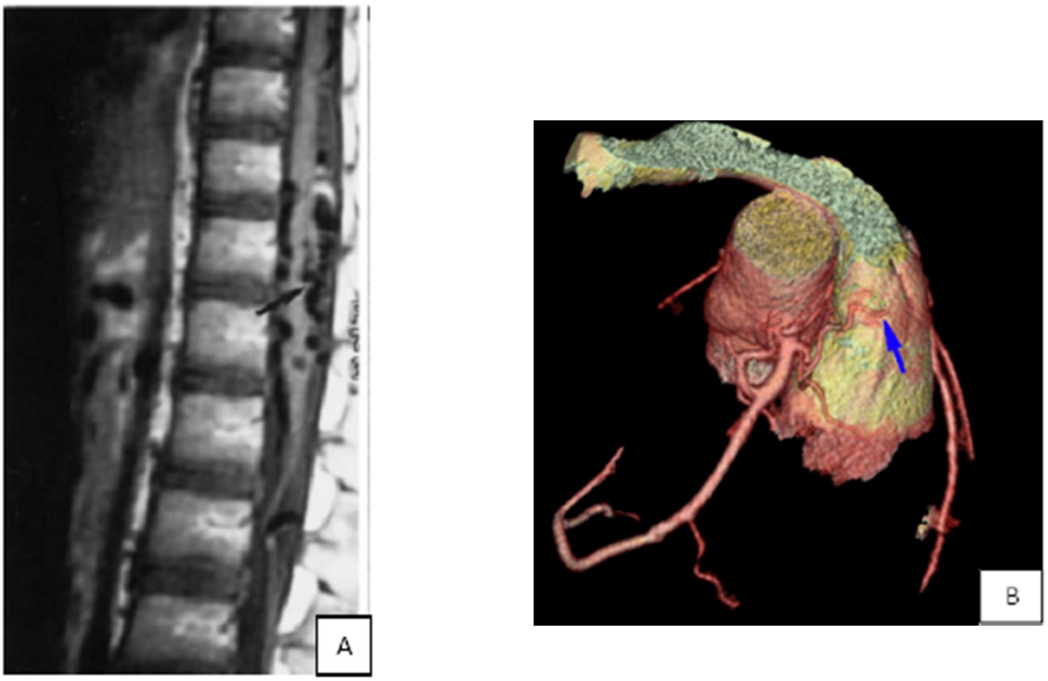


Figure 9: MR image shows enlarged blood vessels along the spinal cord conus (A) (Mandzia *et al*,. 1999) and a cardiovascular computed tomography (CCT) high resolution of a 34 year old female showing a malformation from the right coronary artery (RCA) into the pulmonary artery (B) (Khachatryan *et al*,. 2010).

Reasons for the under and misdiagnosis of HHT

HHT is often misdiagnosed or unnoticed entirely until it becomes a life threatening condition (Guttmacher *et al*., 1994). The reason for this is mainly because it is a rare disease and many physicians are not familiar with the symptoms or manifestations of the disorder (Sabba, 2005). Furthermore, symptoms may occur at a later stage, mainly when the affected individual is an adult and diagnosis of HHT relies on the clinical symptoms. To confirm the diagnoses, genetic testing is done (Mcdonald *et al.*, 2011). However, such testing is not performed or readily available in developing countries.

Molecular Genetics of HHT

Several genes have been identified in literature that are involved in vessel integrity and vascular remodelling such as platelet derived growth factor (PDGF), angiopoietin 1 (Ang1), R-Spondin 3 (RSPO3), and chemokine receptor 3 (CXCR3) (Murakami & Simons, 2009; Mukwaya *et al.,* 2016) amongst others. However, a different set of genes are involved in hereditary haemorrhagic telangiectasia. These genes include *endoglin (ENG), activin receptor-like kinase 1 (ALK1)* and SMAD family member 4 (*SMAD4*). About 85% of individuals affected with HHT have mutations either in the *ENG* or *ALK1* genes with another 2% having mutations in the *SMAD4* gene. Mutations in both *ENG* and *ALK1* genes have been identified in various cohorts of different countries (Canzonieri *et al*., 2014a). *ENG* mutations have been described to be most prevalent in North America and Northern Europe, while the *ALK1* mutations are most prevalent in Southern Europe (McDonald *et al.,* 2011). Mutations within these genes cause disruption of the cell-cell junctions of the EC monolayers consequently altering vascular fragility and permeability. The down-regulation or a deletion of these genes can cause defective vascular developments and may result in embryonic death (Park *et al*., 2015). The mutations include splice variants, indels, exonic and missense mutations (Bayrak-Toydemir *et al*, 2004).

Endoglin (ENG) Gene

The *Endoglin* gene structure

The *endoglin* (*ENG*) gene (MIM 131195) is located on chromosome 9 (9q33-q34.1) and consists of 14 exons where exon 9 is divided into two (9A and 9B). It is 30 kb long (Guttmacher *et al*., 1995; Gallione *et al.,* 2000; Bayrak-Toydemir *et al.*, 2004; Sharathkumar & Shapiro 2008) and encodes for an integral membrane homodimeric

transmembrane glycoprotein that is found in abundance in all tissues containing vascular endothelial cells such as venules, arterioles and capillaries (McAllister *et al*., 1994; Guttmacher *et al*., 1995; Bayrak-Toydemir *et al*., 2004). This gene functions as an auxiliary receptor in the transforming growth factor β (TGF-β) signalling (Park *et al.,* 2015).

The messenger RNA (mRNA) product resulting from gene transcription is 3.4 kb long (Dallas *et al,*. 2008) indicating that the exons of this gene are relatively small (McAllister *et al*., 1994). The extracellular domain is encoded by exons 1 to 12, the transmembrane domain is encoded by exon 13 and the cytoplasmic domain is encoded by exon 14 (Dallas *et al.,* 2008). Exon 12 consists of the smallest number of base pairs, 55 in total and contains a membrane spanning domain. Exon 11 is the longest exon containing 258 base pairs (bp) (Figure 10) (McAllister *et al*., 1994).

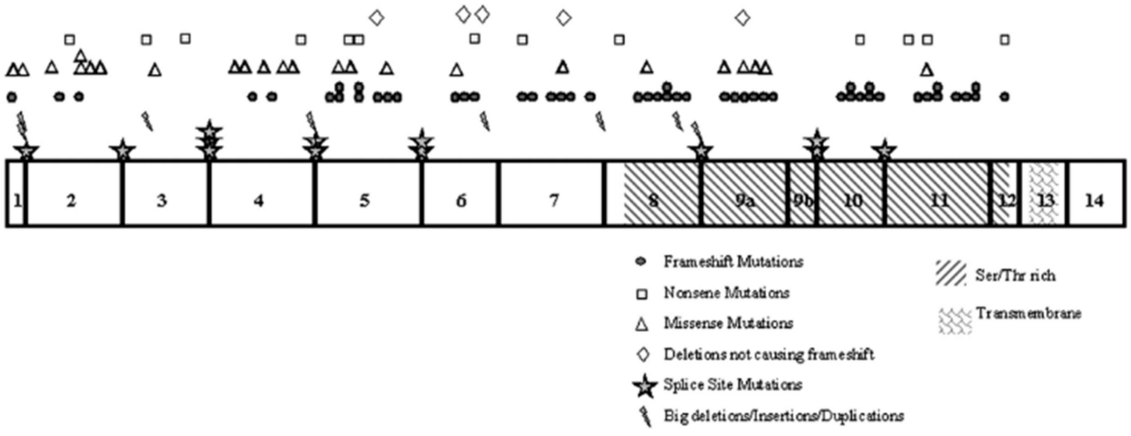


Figure 10: A schematic presentation (not drawn to scale) of the *ENG* gene with 15 exons and the different symbols indicates the type of mutations associated with HHT1. The extracellular, transmembrane and cytoplasmic domains are also depicted (Bayrak-Toydemir *et al*., 2004).

The Endoglin protein structure

The ENG protein contains an enormous extracellular domain that comprises of 561 amino acids, a short cytosolic domain that is constitutively phosphorylated by serine threonine kinases and one hydrophobic transmembrane domain. ENG is expressed as a 180-kDa disulphide linked homodimer that is extremely glycosylated (López-Novoa & Bernabeu, 2010; Park *et al.,* 2015). The principal structure suggests that the NH_2-terminal domain has five N-linked glycosylation sites and a possible O-glycan domain rich with serine and threonine residues close to the membrane-spanning domain. ENG also contains a cell recognition site for many adhesive proteins existing in the extracellular matrix (ECM). The cell recognition site for adhesive proteins in the ECM consists of an Arg-Gly-Asp (RGD) peptide sequence. However, this site does not exist on the ENG protein motifs found in animals like the mouse, dog and rat. ENG belongs to the zona pellucida (ZP) protein family as it shares the ZP domain in the extracellular region that consists of 260 amino acid residues. The orphan domain (NH_2-terminal) is dissimilar in homology to any protein family or domain. The ENG cytosolic domain can be targeted by serine and threonine kinases such as the TGF-β type I and type II receptors, therefore it is constitutively phosphorylated. ENG contains a consensus postsynaptic density 95/Drosophila disk large/zonula occludens-1 (PDZ) sequence binding motif (Ser-Ser-Met-Ala) located at the cytoplasmic domain on the carboxyl terminus. This terminus facilitates the interaction between ENG with several PDZ domain encompassing proteins and ENG distal threonine residues by phosphorylation (Figure 11A) (López-Novoa & Bernabeu, 2010).

The three dimensional structural model of ENG predicted *in silico* has three subdomains. The orphan and ZP domains contain the amino acid residues Glu26-

Ile359 and Gln360-Gly586 respectively. The ZP domain contains the ZP-C and ZP-N subdomains (Figure 11B) (López-Novoa & Bernabeu, 2010).

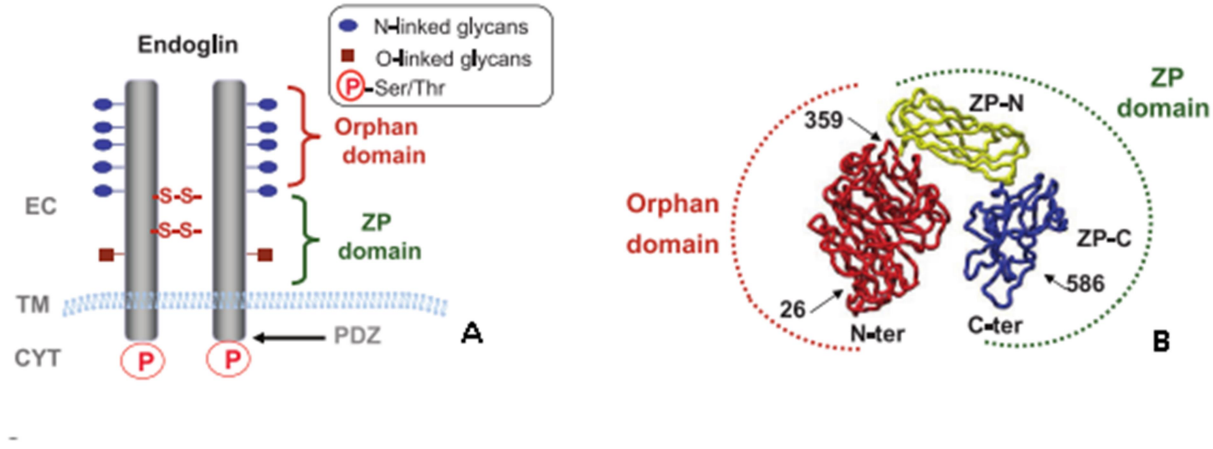


Figure 11: (A) Endoglin membrane protein with an enormous extracellular domain that contains an NH_2-terminal orphan domain and a zona pellucida (ZP) domain comprising of 260 amino acids in the juxtamembrane region. Disulphide bridges link analogous *ENG* monomers (dimers). Consensus motifs that attach to N-linked and O-linked glycans to extracellular domain have been known. The postsynaptic density 95/Drosophila disk large/zonula occludens-1 (PDZ) domain at the cytoplasmic domain contains Ser/Thr residues that are phosphorylated at the carboxyl terminus. The extracellular (EC), transmembrane (TM) and cytoplasmic (CYT) protein domains are indicated.(B) 3 dimensional model of ENG predicted by silico shows 3 various subdomains in yellow, red and blue. The orphan domain (red) contains the amino acid residues Glu26-Ile359 and the ZPdomain comprise of amino acid residues Gln360-Gly586. The ZP-C and ZP-N subdomains are coloured in blue and yellow respectively. The amino acid numbers corresponding to globular domains boarder regions are shown (Image adapted from Llorca *et al.,* 2007 and reproduced from López-Novoa & Bernabeu, 2010).

In the tissues of humans and mice the gene has two alternatively spliced isoforms, the short (S) ENG (S-ENG) and the long (L) ENG (L-ENG) (Figure12). The difference between these protein isoforms is the number of amino acids in their cytoplasmic tails. The S-ENG and the L-ENG contain 14 and 47 amino acids respectively, with a sequence of 7 residues unique to S-ENG which are EYPRPPQ (Figure 12). The mainly expressed isoform is the L-ENG that gives rise to the S-ENG through splicing (López-Novoa & Bernabeu, 2010).

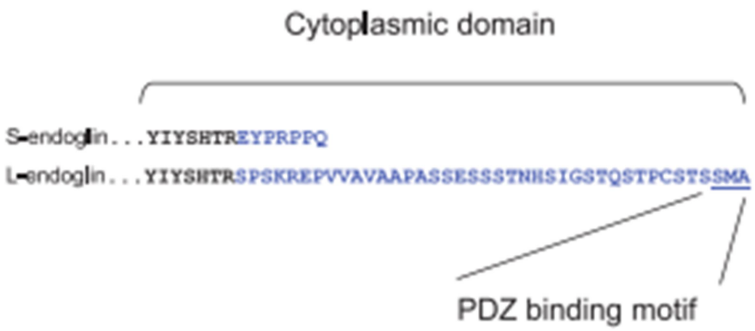


Figure 12: The short (S) and long (L) amino acid sequences of the ENG isoforms. The 7 residues unique to S-ENG are EYPRPPQ (blue highlight). The amino acid differences of the two isoforms are highlighted in blue (Image adapted from Llorca *et al.,* 2007 and reproduced from López-Novoa & Bernabeu, 2010).

In humans, the L-ENG protein has a cytoplasmic domain that is 33 amino acids longer than the S-ENG cytoplasmic domain. Both of these isoform have distinct functions. The S-ENG appears to have an antiangiogenic effect, while L-ENG has a proangiogenic effect. The S-ENG isoform is convoluted with the senescence of ECs and could contribute to age-dependent vascular pathology (López-Novoa & Bernabeu, 2010).

Predominant membrane–bound L-ENG plays an opposing role to that of S-ENG and soluble ENG (sENG) (Figure 13) that are involved in numerous pathological conditions such as cancer, brain AVMs and preeclampsia (López-Novoa & Bernabeu, 2010). Soluble ENG is an antiangiogenic protein (Levine *et al*., 2006) that is a consequence of partial shedding of membrane-bound ENG by membrane-type metalloprotease-1 (MT1-MMP) (Castonguay *et al*., 2011; López-Novoa & Bernabeu 2010).

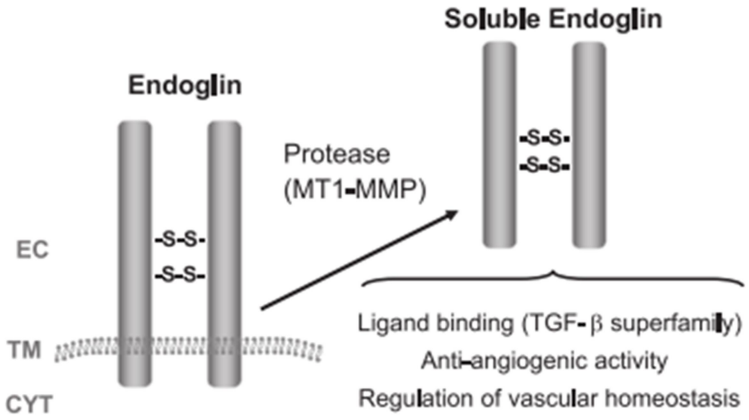


Figure 13: Soluble ENG is generated by membrane bound ENG proteolytic processing. Acting on the juxtamembrane region is a great endoglin-shedding protease called membrane-type metalloprotease-1 (MT1-MMP) that leads to ENG large ectodomain secretion. Some of the functions of soluble ENG include ligand binding of TGF-β family, anti-angiogenic activity and regulation of vascular homeostasis (López-Novoa & Bernabeu, 2010).

Patients with brain AVMs have higher mean levels of s*ENG* than controls. Soluble *ENG* has been linked to other diseases including brain AVMs and preeclampsia as well as hepatitis, diabetes, systemic sclerosis, sickle cell anaemia, cancer, coronary heart disease, malaria and atherosclerosis. The precise mechanism of ENG in these diseases remains unknown (López-Novoa & Bernabeu, 2010).

The distribution of *endoglin* gene expression

The function of this gene is to regulate cell survival and proliferation and it is upregulated in vessel repair sites (Scharpfenecker *et al*., 2009). The distribution of ENG in tissues and cells indicates the role it plays in angiogenesis, vascular development and vascular homeostasis. Low levels of ENG are expressed in stable ECs. However, high expression of ENG is observed in areas where the endothelial sheet is disrupted. These areas include; vascular ECs locations of active angiogenesis during embryogenesis, in healing wounds, inflamed tissues, vascular

injury and in vessels of a tumour. Moreover in the kidney and heart, ENG is overexpressed after ischemia and reperfusion (López-Novoa & Bernabeu, 2010).

ENG is also expressed in other cell types, such as the normal smooth muscle cells (SMCs), where it is expressed at low levels, and vascular smooth muscle cells (VSMCs) of atherosclerotic plaques, where its expression is upregulated. An elevated expression of ENG is also found in activated vessels and possibly induced by hypoxia, cytokines and vascular injury (López-Novoa & Bernabeu, 2010).

The role of e*ndoglin* in vascular pathology

In capillary ECs *ENG* function is required for survival, since extensive apoptosis occurs as a consequence of *ENG* haploinsufficiency. When apoptosis of ECs occurs, an arteriovenous shunt is formed when the capillary network steadily fades (López-Novoa & Bernabeu, 2010) (Figure 14).

When the capillary network fades, the process of angiogenesis is crucial to ensure that new capillary networks are formed. The process of angiogenesis consists of a succession of EC responses to angiogenic stimulation such as ECM degradation, migration, proliferation, budding, tube formation, maturation and dormant endothelium maintenance. These processes are controlled by a compound of various growth factor interactions such as vascular endothelial growth factor (VEGF), TGF-β (and their distinct receptors) and basic fibroblast growth factor (bFGF). These growth factors interact with other angiogenic stimuli like hypoxia (López-Novoa & Bernabeu, 2010).

One of the growth factors, transforming growth factor β (TGF-β) has both stimulatory and inhibitory effects on angiogenesis but both of these processes are subject to

environmental conditions. For example in a healthy individual, hypoxia causes apopstosis to occur in ECs. Consequently, the ENG is expressed preventing further ECs apopstosis and promoting the repair mechanism. However, in a HHT individual, apoptosis continues to occur due to haploinsufficiency of ENG causing capillary regression. This happens because the maximum number of ENG proteins present (threshold) are not sufficient for the repair mechanism to occur. Both of these processes (repair mechanism and capillary regression) occur as a response to the presence or absence of ENG in the TGF-β pathway. This is probably crucial in the development of the vascular lesion (Figure 14) (López-Novoa & Bernabeu, 2010).

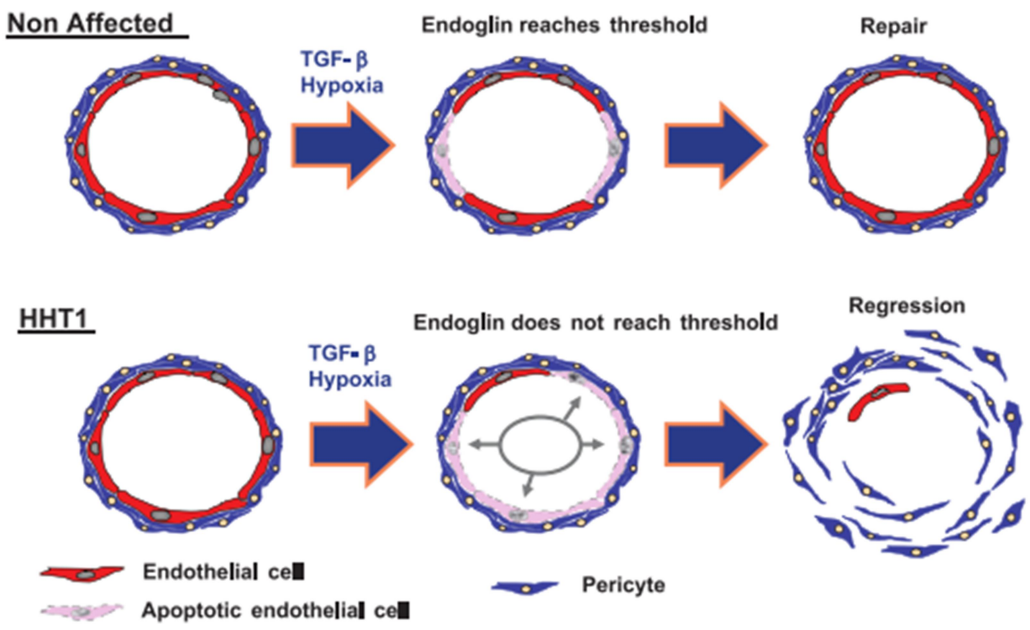


Figure 14: A theoretical representation of the function of ENG in endothelial cells (ECs) apoptosis and its significance in vascular lesions in HHT. Under precise conditions such as hypoxia and TGF-β, the promotion of ENG expression prevents ECs apoptosis in healthy individuals. In HHT patients, *ENG* haploinsufficiency may lead to tremendous apoptosis in ECs where the functioning of ENG is necessary for survival hence leading to capillary regression hence permitting AVM formation. The image depicts a cross section of an individual vessel with ECs in red, apoptotic ECs in pink and pericytes in blue (López-Novoa & Bernabeu, 2010).

Consequences of *ENG* Mutations

Because ENG is a homodimeric transmembrane glycoprotein that is widely expressed in endothelial cells (ECs) (Bayrak-Toydemir *et al*., 2004; Tual-Chalot *et al*., 2015), mutations in this gene can impair the function of ECs (Guttmacher *et al*., 1995). HHT is not caused by a specific mutation, but by any mutation of the *ENG* gene that affects protein function (Gallione *et al*., 1998). For example, a truncation in this gene may lead to a loss of function of the proteins that are necessary for vessel repair (dominant negative mutation) (Guttmacher *et al*., 1995).

Many different mutations on the *ENG* gene have been linked to HHT1 (Table 3), (McAllister *et al*., 1994; Bayrak-Toydemir *et al*., 2004). They are also associated with PAVMs and brain AVMs (Gallione *et al.,* 2000; Chen *et al*., 2013; Tual-Chalot *et al*., 2015). Every deleted copy of the *ENG* gene induces the formation of dysplastic vessels. There are no mutation “hot spot” since the mutations are spread over the entire gene (Bayrak-Toydemir *et al*., 2004) as shown on Figure 10.

McAllister *et al.* (1994) reported premature termination mutation clusters (frameshift or nonsense) that occurred in exons 5 through 11 in the *ENG* gene. Should this mutation be translated, the resulting protein may still have ligand binding activity. This suggests a model of dominant-negative effect at the TGF-β receptor complex involving a truncated protein (Gallione *et al*., 1998).

The g.IVS+IG>A and c.1238G>T mutations (Table 3) are frequent in the small Islands of the Dutch Caribbean located in Netherlands Antilles (of which Curaçao is an island) where HHT has a prevalence of 1/1300. The g.IVS+IG>A is a splice site mutation on exon 1 and present in seven of ten families tested. The missense mutation c.1238G>T found in exon 9a, was identified in two Dutch Caribbean

families of possibly African ancestry and one Dutch family. These findings suggest that these mutations were introduced into the African slave population of the islands by the Dutch colonists as a result of a “founder effect” (Gallione *et al.,* 2000; Bayrak-Toydemir *et al*., 2004).

In two large families with HHT1, two mutations that interfered with recognition sites of the splice-junction and mRNA processing were identified (Gallione *et al*., 1998). In one family, an intron 4 base substitution of A>G terminates the intronic ag/splice acceptor sequence consequently leading to the disease phenotype. (Gallione *et al*., 1998) found another mutation the G1311C mutation in exon 9B which was observed in 25 affected members of the same family but was not identified in 115 individuals of the control group. However, this mutation did not alter the amino acid sequence but created an unfavourable splice-donor sequence by replacing the G to a rare C as the last nucleotide of exon 9B. Therefore Gallione *et al*. (1998) concluded that this allele is a disease phenotype because of its exclusive presence in the affected individuals.

It has been observed that mutations resulting in splicing errors and premature termination generally result in reduced expression of mutant alleles and message stability. Moreover, expression analysis of mutant mRNA shows that mutant transcripts of premature termination codons are degraded as they are barely detected (Gallione *et al*., 1998).

Table 3 represents the numerous splice and exonic mutations found on the *ENG* gene, some of the mutations are association to PAVMs, CAVMS and one HAVM. Some of these mutations have proven to be pathogenic while others pathogenicity remains unknown.

Mutation nomenclature is crucial to understand the mutations listed on the table. Most of the mutations listed on the table begin with a “c” which means that these mutations were identified in cDNA. There are other mutations that are identified in genomic DNA “g”, mitochondrial sequence “m” and protein sequence “p”. The standard format of writing a mutation is as follow c.523G>C; this means the mutation was identified in a cDNA sequence where there was a change of guanine to cytosine at position 523. Other mutations are deletions such as c.1089_1090delTG that shows a 2 nucleotide deletion at positions 1089 and 1090.

TABLE 3: SPLICE AND EXONIC MUTATIONS OF THE *ENG* GENE

	LOCATION	NUCLEOTIDE	CASES REPORTED	AVM	CLASSIFICATION	ARTICLE FOUND
1.	Intron 1	c.68G>A(c.68-1G>A)	2	NA	Pathogenic	(Lesca *et al.*, 2004; Bossler *et al.,* 2006)
2.	Intron 1	c.67+1G>A	7	PAVM	Pathogenic	(Gallione *et al.*, 2000)
3.	Exon 1	g.IVS+IG>A	7	NA	Pathogenic	(Gallione *et al.,* 2000; Bayrak-Toydemir *et al.*, 2004)
4.	Intron 2	c.219G>A	1	PAVM	Unknown	(Cymerman *et al.,* 2003)
5.	Intron 3	c.360+1G>A	14	PAVM	Pathogenic	(Pece *et al.*, 1997; Cymerman *et al.*, 2000)
6.	Intron 3	c.360+1G>C	2	PAVM HAVM	Pathogenic	(Dakeishi *et al.*, 2002)

7.	Intron 3	c.360+4A>G	1	PAVM	Pathogenic	(Shovlin, *et al.,*1997)
8.	Exon 4 missense mutation	c.523G>C	1	PAVM	Pathogenic	(Cymerman *et al.*, 2003)
9.	Intron 5	c.524-2A>G	5	NA	Pathogenic	(Gallione *et al.*, 1998)
10.	Intron 5	c.689+2T>C	1	NA	Pathogenic	(Lesca *et al.*, 2004)
11.	Exon 5 (Deletion/ Frameshift)	c.1089_1090delTG	1	PAVM CAVM	Pathogenic	(Cymerman *et al.*, 2000)
12.	Intron 8 (Intronic deletion)	c.1133+3_1103+8del AAGGGA	1	NA	Unknown	(Lesca *et al.*, 2004)
13.	Exon 9-13 Large deletion	EX9_13del	2	PAVM	Pathogenic	(Cymerman *et al.*, 2003; Shovlin *et al.*, 1997)
14.	Intron 9	c.173-1G>C	1	NA	Pathogenic	(Bossler *et al* 2006)
15.	Intron 10	c.1311+2T>A	1	PAVM	Pathogenic	(Cymerman *et al.*, 2000)
16.	Exon 11 Splice site/ Silent	c.1428G>A	1	NA	Unknown	(Lesca *et al.*, 2004)

PAVM: Pulmonary arteriovenous malformation CAVM: Cerebral arteriovenous malformation

HAVM: Hepatic arteriovenous malformation NA: Not available

Activin receptor-like kinase 1 (ALK1) Gene

Activin receptor-like kinase 1 (*ALK1*) (MIM 601284) also known as *activin A receptor type-II-like kinase 1* (*ACVRL1*) is located on chromosome 12 (12q11-q14), it consists of 10 exons, is 15 kb long and encodes a TGF-β type 1 receptor expressed mainly in vascularized tissue e.g. lungs and in endothelial cells (Sharathkumar & Shapiro 2008). The gene product aids in the formation of blood vessels by mediating signals in the transforming growth factor β pathway. Mutations in this gene may cause protein alternations that do not bind to a ligand, therefore adversely affecting the formation of blood vessels. *ALK1* mutations are associated with HHT type 2 (HHT2) (Bayrak-Toydemir *et al.*, 2004) and liver AVMs (Chen *et al.*, 2013). A schematic presentation of the *ALK1* gene (Figure 15) illustrates the10 exons linked together. The different symbols indicate the type of mutations associated with HHT2. Missense mutations account for 55% of the recurrent *ALK1* mutations. Splice site mutations are rare on *ALK1* but on the *ENG* gene these mutations account for 12% (Bayrak-Toydemir *et al.*, 2004).

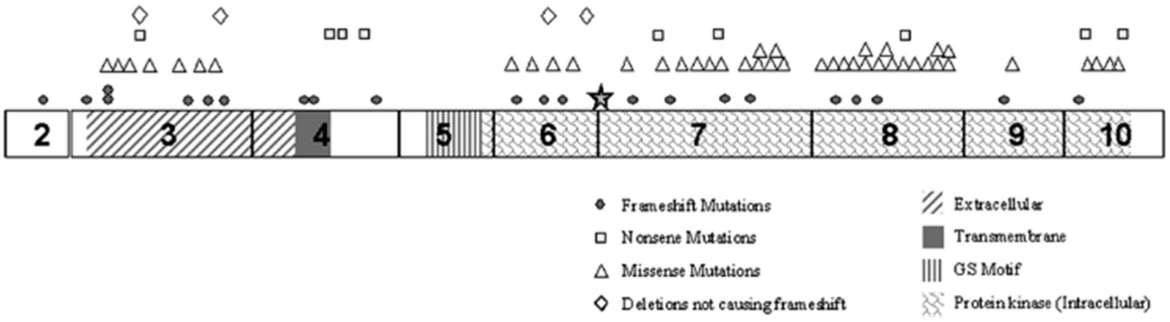


Figure 15: The schematic representation of the *ALK1* gene with 10 exons, the locations of various mutations and various regions of the gene (Bayrak-Toydemir *et al.*, 2004).

SMAD4

SMAD family member 4 (*SMAD4*) (OMIM 175050) mutations cause a rare form of HHT where the vascular lesions occur together with juvenile polyposis (JP) (López-Novoa & Bernabeu 2010; Tual-Chalot *et al.,* 2015; Canzonieri *et al*., 2014a).The combination of both conditions JP-HHT occurs in only 1-2% of patients diagnosed clinically with HHT as observed in *SMAD4* mutations (López-Novoa & Bernabeu, 2010). It is a severe condition as the adenomas of juvenile polyposis often transform into colorectal cancers. Patients often develop aortic dilatation (Teekarikirikul *et al.,* 2013). *SMAD4* transcription factors stimulate the expression of ENG during TGF-β signalling (López-Novoa & Bernabeu, 2010).

Other genes mutations associated with HHT

Mutations in the *Bone morphogenetic 9* (*BMP9*) gene were identified in two unrelated HHT individuals who tested negative for the three main HHT associated gene *ENG, ALK1* and *SMAD4* (Wooderchak-Donahue *et al*., 2013). *BMP9* is also known as growth differentiation factor 2 (GDF2), is located in chromosome 10q11 and it is one of the members of the TGF-β superfamily (McDonald *et al.,* 2015). *BMP9* is involved in the TGF-β pathway by binding to the surface receptors of the ECs including *ENG,* serine and threonine kinases and *ALK1.* The BMP9-dependent TGF-β activates ALK1 which then phosphorylates SMAD1, SMAD5 and SMAD8 to form a complex which then translocates into the nucleus, allowing the regulation of ECs (Figure 16) (McDonald *et al.,* 2015). Mutations in this gene contributed approximately <1% overall for HHT (Wooderchak-Donahue *et al*., 2013).

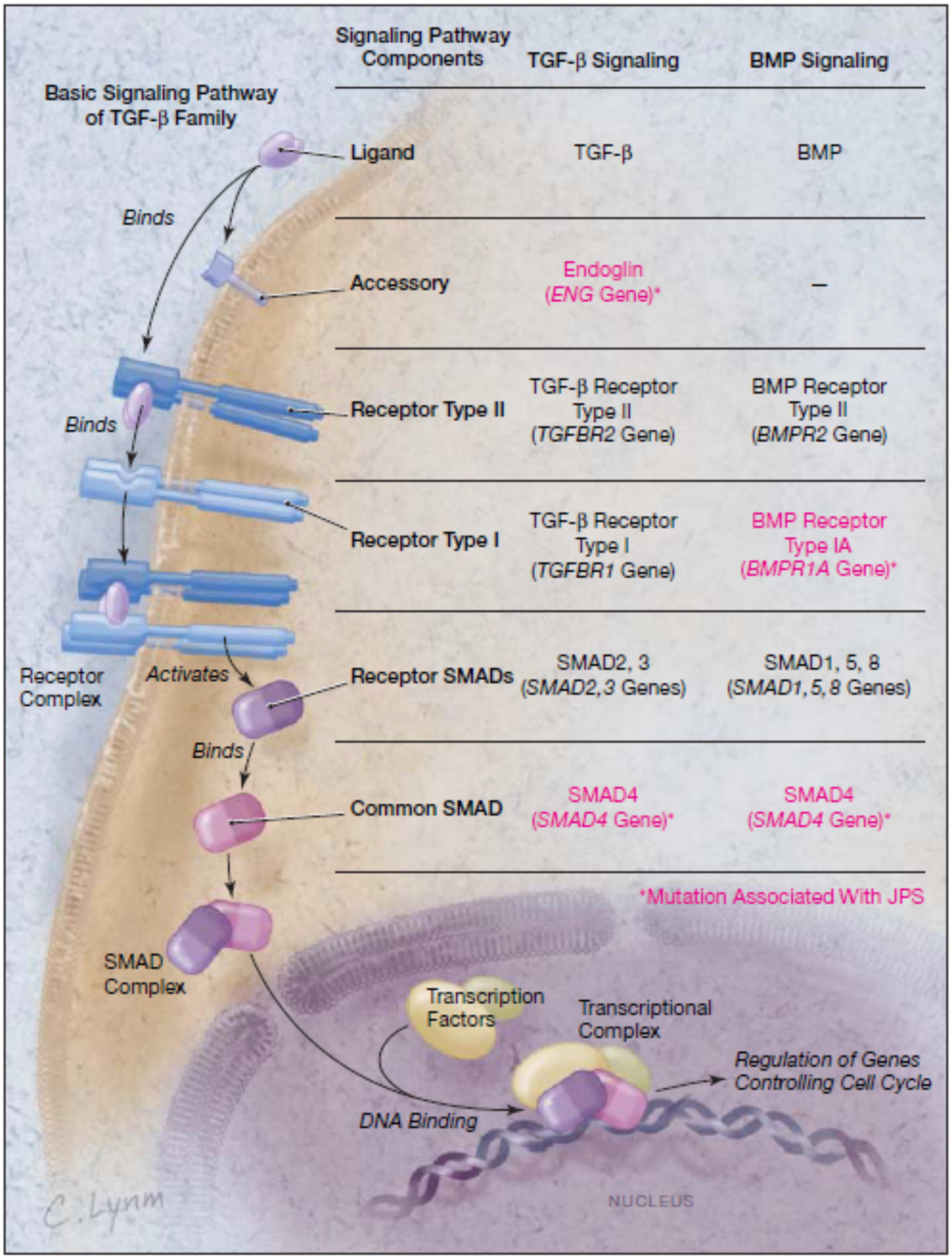


Figure 16: The TGF-β pathway is closely related to the BMP pathway due to the shared SMAD4. Germline mutations of BMPR1A and SMAD4 are associated with heritable JPS. ENG is an accessory protein found only in the TGF-β pathway (Image copied from Sweet *et al.,* 2005).

RAS p21 protein activator 1 (*RASA1*) mutations have also been identified in individuals who are suspected to have HHT. This gene is located in chromosome 5q13.3. The cutaneous capillary malformations observed in *RASA1* related disorder are different from those in HHT but there is ample similarity making it difficult for

medical professionals to distinguish the vascular malformations. Therefore, to rule out the *RASA 1* related disorder, the *RASA 1* has to be screened as well (McDonald *et al.,* 2015).

Transforming growth factor β (TGF-β) Pathway

Endoglin and activin receptor-like kinase 1 are receptor proteins included in the Transforming growth factor β (TGF-β) superfamily (Bayrak-Toydemir *et al*, 2004). TGF-β is a cytokine (Park *et al*., 2015) that consists of a family of at least 25 growth factors that regulate numerous endothelial cell processes such as migration, adhesion, proliferation and organisation and composition of the extracellular matrix (McAllister *et al*., 1994; Guttmacher *et al*., 1995; Park *et al*., 2015) as previously mentioned.

Some of the functions of TGF-β include the regulation of growth, motility, differentiation, wound repair, tissue remodelling and apoptosis in numerous cell types (McAllister *et al*., 1994; Park *et al*., 2015). TGF-β also regulates the production of smooth muscle cells pericytes, matrix proteins by stromal interstitial cells and endothelial cells (McAllister *et al*., 1994). The TGF-β pathway also regulates other biological processes such as embryogenesis, cell cycle control, angiogenesis, growth, development and differentiation of cell types (Sharathkumar & Shapiro 2008). *In vivo* TGF-β is an effective angiogenic factor and a vascular remodelling mediator because it controls production of extracellular matrix by smooth muscle cells, pericytes and endothelial cells (McAllister *et al*., 1994). TGF-β displays both inhibitory and stimulatory angiogenesis in exploratory conditions (Park *et al*., 2015).

Endoglin, activin kinase receptor-like kinase 1, SMAD4 and *Bone morphogenetic protein receptor type 2* genes are all involved in the pathology of HHT due to their interaction in the TGF-β pathway.

TGF-β pathway proteins

The primary TGF-β signalling pathway encompasses at least seven *ALK* type 1 receptors, and five type II receptors that form heterodimeric complex when type 1 reacts with type II allowing a signal to continue downstream. The heterodimeric formation regulates the specificity of the signalling of ligands. Type 1 receptors include activin (*ALK1, ALK2 and ALK4*), BMP (*ALK1, ALK2, ALK3 and ALK6*) and TGF-β (*ALK1, ALK2, ALK5 and ALK7*) receptors. Other molecules that are crucial for the pathway are R-Smads which comprise *SMAD1, SMAD2, SMAD3, SMAD5 and SMAD8.* After phosphorylation these proteins carry the signal downstream to Co-Smad (*SMAD4*) which then continues to carry the signal downstream until the target gene is transcribed (López-Novoa & Bernabeu, 2010).

On the cell surface, tyrosine/threonine kinase receptors mediate the functions of TGF-β. Such receptors include TGF-β receptor type II (TGF-βR II), TGF-β receptor type I (TGF-βR I) and ENG. There are two diverse types of TGF-βR I which includes *ALK1* and *ALK5*. *ALK1* is expressed only in ECs while *ALK5* is expressed ubiquitously. TGF-β participates in angiogenesis by inhibiting or stimulating ECs activation through the balance of *ALK1* and *ALK5* signalling (Park *et al*., 2015). *ENG* and *ALK1* genes are ligands for the TGF-β superfamily (Gallione *et al*., 2000).

The TGF-β pathway function

Figure 17, is a schematic illustration of the TGF-β pathway. *ENG* acts as an accessory protein by being cleaved on the cell surface through proteolysis and sinks by binding specific circulating TGF-βI proteins (*ALK1* and *ALK5*) to form a complex (López-Novoa & Bernabeu, 2010;Park *et al*., 2015; Tual-Chalot *et al.*, 2015). The TGF-β ligand binds to TGF-βR II and TGF-βI dimer binds to two TGF-βR II receptors inducing a heteromeric complex formation on the cell membrane with TGF-βR I receptor, thus activating downstream TGF-β pathway in the cytoplasm (Sharathkumar & Shapiro 2008; López-Novoa & Bernabeu, 2010; Park *et al*., 2015). Both TGF-β type II and type I receptors are glycine/serine kinases that have a catalytic portion that phosphorylates. TGF-βII receptor phosphorylates the Gly/Ser-rich domain of TGF-βI receptor, which then phosphorylates R-SMAD. The pathway splits into two different branches downstream of TGF-βR I (*ALK4, ALK5* and *ALK7*) phosphorylates only *SMAD2* or *SMAD3*. The other type I receptors (*ALK1, ALK2, ALK3* and *ALK7*) phosphorylates only *SMAD1, SMAD5* and *SMAD8.* When R-SMAD (receptor-SMAD) is phosphorylated, it changes conformation that enables it to cleave away from TGF-β type I receptor. The phosphorylated R-SMAD dimerizes with cooperating homolog called Co-SMAD (*SMAD4*) forming a complex that acts like a transcription factor in the cytoplasm. R-SMAD/Co-SMAD complex translocate into the nucleus to induce transcription of target gene (Sharathkumar & Shapiro 2008; López-Novoa & Bernabeu, 2010). R-SMAD/Co-SMAD complex binds to promoter boxes on the promoter region of a target gene increasing affinity of RNA polymerase to bind to target gene. This process is called TGF-β response elements or SMAD response elements. An interruption of any of these processes may cause vascular dysplasia, suggesting that physiological, genetic, mechanical barriers are

essential for the development of each lesion (Guttmacher *et al*., 1995). This pathway regulates the progression or regression of angiogenesis and stimulates *ENG* expression (Figure17) (Sharathkumar & Shapiro 2008;López-Novoa & Bernabeu, 2010). Any defect in the pathway can lead to HHT1, HHT2, PPH and HHT or juvenile polyposis but the exact mechanism still remains unclear (Bayrak-Toydemir et al., 2004; Sharathkumar & Shapiro 2008). Furthermore, besides PPH associated with HHT, mental disorders, cancer, auto-immune disease and fibrosis can also occur as a result of incorrect signalling in the TGF-β pathway (Ten Dijke & Hill, 2004).

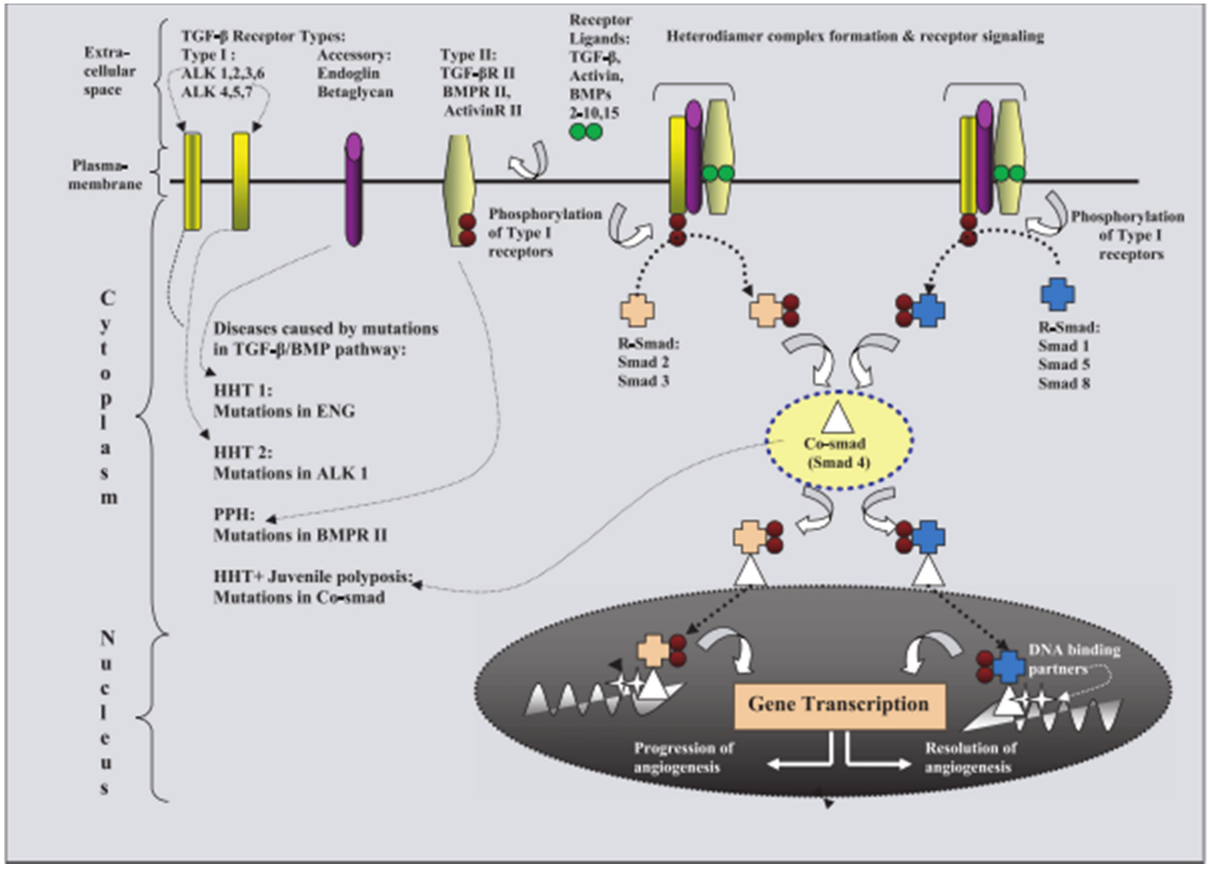


Figure 17: Transforming growth factor β (TGF-β) signalling pathway superfamily illustrating the relationship between HHT1, HHT2, PPH and HHT and Juvenile Polyposis.

In the extracellular matrix, ligands such as TGF-β bind to TGF-βR II on the cell membrane. The accessory receptor (*ENG*) can modulate signalling through receptors TGF-βR II and TGF-βR I. A heteromeric complex is formed when the binding of ligands to TGF-βRII leads to the binding TGF-βR I at the cell surface. The kinase domain of TGF-βR I is phosphorylated and activated initiating SMAD (R-SMAD) proteins phosphorylation in the cytoplasm. The pathway splits into two different branches; TGF-βR I (ALK4, ALK5 and ALK7) phosphorylates only Smad2/3 and ALK1, ALK2, ALK3 and ALK7 phosphorylates Smad1/5/8. The phosphorylated R-Smad binds to Co-Smad (*SMAD4*) forming a complex in the cytoplasm. The complex then translocate into the nucleus to induce transcription of target gene (Sharathkumar & Shapiro 2008).

Animal studies

More than 80% of HHT patients have either HHT1 or HHT 2 and the other 20% are HHT patients who have mutations that derive from *SMAD* 4 genes (juvenile polyposis and HHT). The mouse models carrying mutations in the respective genes have been used to investigate HHT on an experimental level (Tual-Chalot *et al*., 2015). Such animal studies enable researchers to better understand the disease mechanism (Rubio-Ruiz *et al*., 2015). Besides understanding HHT disease mechanism, mouse models are also crucial for the development of new therapies (Tual-Chalot *et al.,* 2015).

Mice heterozygous for *Eng* mutations

Heterozygous ($ENG^{+/-}$) *or* $Acvrl1^{+/-}$ mice were reported to have little phenotypic HHT features. However, between 3 months and 2 years, these mice developed only a few HHT-like symptoms such as ear telangiectases, vessel dilation, epistaxis and vascular lesions, suggesting that additional triggers are essential for the development of HHT (Tual-Chalot *et al.,* 2015).

Heterozygous mice developed normal blood vessels suggesting that angiogenesis is not affected because there are no major vascular defects. It was observed that the brain vasculature in heterozygous mice developed abnormal vascular responses only when exposed to pro-angiogenic stimulations. Therefore, angiogenic stimulation is necessary for the development of vascular dysplasia in heterozygous mouse models. For normal angiogenic responses, ENG needs to be present at physiological levels. It is thought that the abnormal formations of blood vessels in HHT are probably triggered by angiogenesis. However, heterozygous *ENG* mutations and angiogenic triggers together are not adequate to initiate AVM formation suggesting

that another event is essential for AVMs formation (Tual-Chalot *et al.,* 2015). According to the three event hypothesis (Figure 18) two genetic hits which are the loss of both *ENG* alleles as well as an environmental trigger has to occur for the formation of AVMs. These triggers include infection, inflammation, vascular injury or trauma that act in conjunction with *ENG* haploinsufficiency to produce the lesion (López-Novoa & Bernabeu 2010; Tual-Chalot *et al.,* 2015). Inflammatory stimuli such as wounds may lead to the development of AVMs near the site of the wound (Park *et al*. 2009). This was proven when wound-induced AVMs appeared in ENG deficient mice (Choi *et al*., 2014).

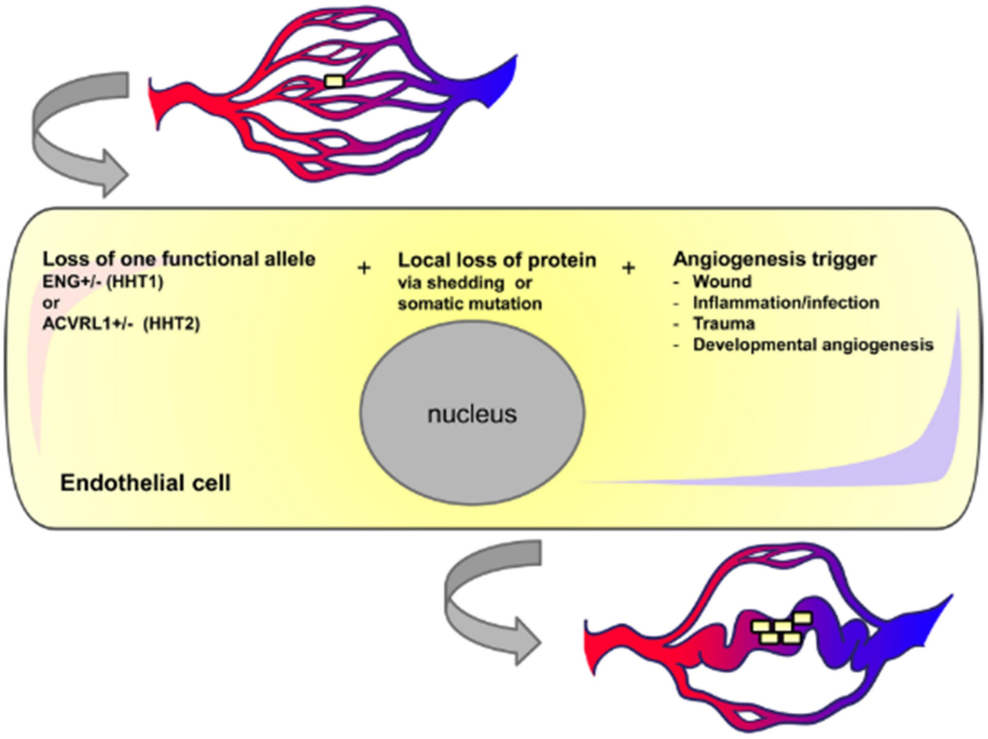


Figure 18: The three event hypothesis for the formation of AVM in HHT. ECs in HHT individuals may undergo the loss of *ENG* or *ALK1* protein either due to somatic mutation or shedding. If vessels containing these permanent or temporary null cells are exposed to pro-angiogenic triggers, the cells may undergo further proliferation and cause the enlargement of vessels. It has been suggested that once blood flow accumulates in the enlarged vessel, further cell proliferation and vessel enlargement will may occur due to a positive feedback loop (Tual-Chalot *et al.,* 2015).

Analysed mouse models of HHT suggest that prevention of angiogenesis or inflammation would decrease the risk of the formation of AVMs while agents that cause vascular stabilization would decrease the risk of haemorrhage. Heterozygosity of *ENG* mutations is similar to the baseline state of HHT1 (Tual-chalot *et al.,* 2015).

Mice homozygous for *Eng* null mutations

Homozygous null (**$Eng^{-/-}$**) mice lacked a functional *ENG* protein and therefore there was no *ENG* expression. This led to cardiovascular defects and by mid-gestation, embryonic death. However, vasculogenesis (which is defined as the differentiation of angioblasts into ECs) and the novel formation of vascular network were not affected (Tual-Chalot *et al*., 2015). Hence Tual-Chalot *et al*. (2015) suggested that *ENG* is not essential for the formation of the vascular plexus or the initial differentiation of ECs. Therefore, the loss of *ENG* in VSMCs (as well as other cells such as pericytes or macrophages) does not lead to the development of HHT lesions. Yolk sacs that lacked *ENG* function had enlarged vessels, while the embryo showed cardiac cushion defects and major vessels showed delayed maturation, with decreased vascular smooth muscle cell (VSMC) in their walls. Given these findings, Tual-Chalot *et al*. (2015) suggest that *ENG* is important for the formation of primitive cardiac cushions and for angiogenesis. Such findings have led to the idea that AVMs develop by the merging of veins and arteries because of the loss of their molecular identity (López-Novoa & Bernabeu, 2010; Tual-Chalot *et al*., 2015).

The cause of enlarged fragile vessels observed in yolk sacs with homozygous *ENG* null mutations remains unknown. These vessels are probably a suitable model of HHT vessels (Tual-Chalot *et al.,* 2015). One of the experiments showed that ENG is needed by ECs for the production of a locally active TGF-β1 protein. The EC-derived

TGF-β1 protein is crucial for maturation and differentiation of the contiguous VSMCs. The presence of fragile vessels in HHT may be explained by the reduced TGF-β1 protein. Stated simply, if ENG is not present, there may not be adequate TGF-β1 protein in the ECs to signal the adjacent VSMCs and stimulate their maturation (Carvalho *et al.,* 2004).

AVMs and loss of heterozygosity (LOH) in HHT patients

As previously mentioned both *ENG* alleles need to be non-functional for the AVMs to form. So how exactly do these AVMs form in HHT patients who are heterozygotes? One possibility is the loss of heterozygosity (LOH) in somatic cells. LOH leads to the loss of both alleles. LOH might be the somatic second hit that triggers the pathology in this syndrome. LOH explains the observation that as *ENG* somatic mutations accumulate with age, so do the clinical symptoms of patients with HHT worsen (Tual-Chalo *et al.,* 2015).

Berg *et al*. (1997) suggested that LOH may be necessary for the development of vascular lesions. In this way the somatic mutations would increase the number $Eng^{-/-}$ ECs that would be susceptible to subsequent exposure to inflammatory and angiogenic signals. The $Eng^{-/-}$ ECs that that lack ENG would become the sites of vascular malformations (Tual-Chalot *et al.,* 2015).

Contrarily, a case report of Bourdeau *et al*. (2000) does not support the above theory. They reported that there was no loss of ENG protein expression in ECs inside one PAVM and one CAVM. Bourdeau *et al*. (2000) concluded that the loss of the second ENG allele was unlikely to be the cause of the formation AVMs. Nevertheless, some studies on pathogenesis of many familial vascular

malformations such as CAVM, glomuvenous malformation and venous malformation observed that some cells contributing to the vascular lesions have LOH mutations. Therefore, the loss of ENG in ECs does not have to be present in all the ECs of an AVM, and AVMs may be ENG mosaics (Tual-Chalot *et al.,* 2015).

New treatment suggested from experiments on animal models

Decreased vascular smooth muscle cells (VSMC) coverage of vessels weakens them and makes them more susceptible to breaking and bleeding. Lebrin *et al*. (2010) studied the use of thalidomide (a vascular maturation factor) as a viable treatment for HHT. This study used a heterozygous *ENG* mouse model to test this concept. Treatment with thalidomide showed that it restored decreased VSMC coverage of vessels on mice. They also showed that epistaxis was decreased. Biopsies from HHT patients who were treated for 3 months showed improved coverage with VSMCs in nasal mucocutaneous vessels (Lebrin *et al*., 2010). Even though these findings support the use of thalidomide for the treatment of HHT, little is known about its long term benefits or side effects (Tual-Chalot *et al.,* 2015). However, the use of thalidomide is highly regulated as it may not be used in pregnancy. It was associated with foetal phocomelia when it was used for morning sickness in the 1950s. Both males and females taking it must use contraception, irrespective of whether their sexual partners are using contraception (Carey *et al.,* 2017). Derivatives of thalidomide are used to treat certain vascular cancers since it induces vascular maturation (Wang *et al.,* 2016). Safer drugs need to be investigated since this drug has adverse effects that are serious such as peripheral neuropathy (Dimopoulos *et al.,* 2004).

Tranexamic acid is an antifibrinolytic drug that prevents the breakdown of blood clots. Gaillard *et al*. (2014) showed that the use of 3 g of tranexamic acid per day reduced the length of epistaxis by 17.3% per month. Additionally, Dheyauldeen *et al.* (2012) showed that a set of intranasal submucous injections of bevacizumab (an anti-VEGF agent) reduced bleeding in patients with HHT. To find better treatment of HHT, mouse models can be used given that the main genetic causal genes are known. However, numerous investigations over a long period of time in mice may ensure that the long-term adverse effects of drugs are known.

HHT and people of African ancestry

HHT has mostly been reported in individuals of European ancestry and rarely in African people (Westermann *et al*., 2003). The disease has also been reported in other ethnic populations such as Japanese, Chinese and people living in the Sahara desert (Westermann *et al*., 2003). Few studies have identified HHT in populations of African ancestry. For example in the United States, with an estimated 20 million African-Americans citizens, only 14 patients have been identified. In sub-Saharan Africa, only one patient from Kenya has been confirmed genetically to suffer from HHT (Westermann *et al*., 2003). A missense mutation was identified in this Kenyan patient. It is located on exon 7 of the *ALK1* gene (c.818T>C, p.L273P)) (Canzonieri *et al*., 2014b). The physical symptoms he suffered from was epistaxis which caused severe iron deficiency; consequently, he was put on permanent iron supplementation. Additionally, he had a positive family history and he had multiple telangiectases on the tongue and lower lip (Kitonyi *et al.,* 2008). The apparent rarity of HHT in Africans was refuted when six patients with HHT and CAVMs in Netherland Antilles were discovered (Westermann *et al*., 2003). The highest prevalence of HHT in the world thus far is the Afro-Caribbean population of the

Dutch Caribbean with a prevalence of 1 in 1331 as a result of a founder effect (Gallione *et al.,* 2000; Westermann *et al*., 2003).

Gallione *et al.* (2000) did a study on HHT in the Dutch Caribbean population of Bonaire and Curaçao, which are predominantly of African ancestry, their results indicate that mutations within the population were not family specific. When a mutation is not family specific it means that the mutations identified within the family varied. Of the two mutations found in the *ENG* gene namely g.IVS+IG>A and c.1238G>T the latter missense mutation was also found in a Dutch family suggesting that there was a shared common Dutch ancestor in some families (Gallione *et al.,* 2000). This shared mutation was transmitted to the African population when the Dutch colonized the Dutch Caribbean and settled in Curaçao during the 1600s. The settlement in Curaçao became a centre of slave trade in the Caribbean for almost 200 years. When the slave trade ended, some of the slaves continued to stay in Bonaire and Curaçao and to this day many of the individuals there are of African ancestry. This mode of transmission is known as a founder effect (Gallione *et al.,* 2000).

In another study done in Italy, two African men (one from Morocco and the other from Senegal) had two and three out of four HHT symptoms respectively according to the Curaçao criteria. Molecular analysis identified two mutations including a splice site mutation in intron 2 of the *ENG* gene (c.220-1G>C,) and a missense mutation in exon 8 of the *ALK1* gene (c.1232G>A, p.R411Q). The mutations were identified on the Moroccan and Senegalese man respectively (Canzonieri *et al*., 2014b). To date there are not many studies done on black African people not only because it is rare and also because it may be under-reported or misdiagnosed.

The effect of HHT on lifestyle

Lund *et al.* (1999) developed a grading system (Figure 19) for epistaxis in HHT by using disease questionnaires. Data of other treatments such as blood transfusion and the use of visual analogue scale was used to develop the scale as shown below (Al-Deen & Bachmann-Harildstad, 2008). This scale indicates that individuals that do not need blood transfusion can either have a poor quality of life or reasonable quality of life that determines whether their nosebleeds are moderate or mild.

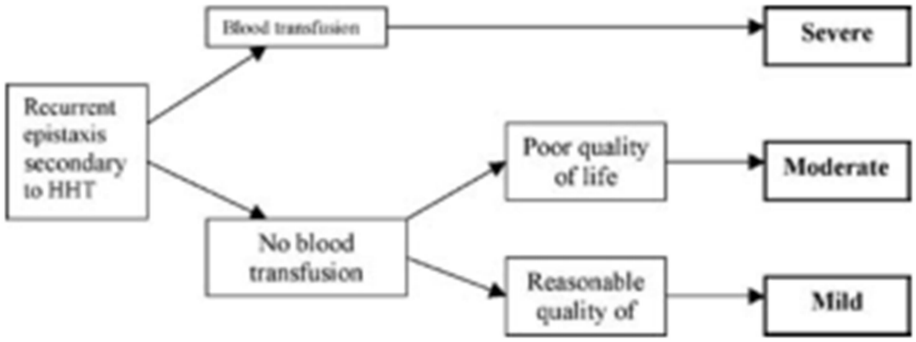


Figure 19: The grading of epistaxis using treatment information, visual analogue scale and questionnaire (Al-Deen & Bachmann-Harildstad, 2008).

The poor quality of life that HHT patients suffer from is broader than just bleeding due to epistaxis; it also has a social implications. Due to presence of AVMs and telangiectases, haemorrhage amongst other symptoms places a severe burden on patients in everyday life. Haemorrhage usually causes anaemia and consequently; affected individuals suffer from fatigue and iron deficiency. Fortunately, the iron deficiency can be restored by the use of oral iron supplements or intravenous iron infusion. However, due to constant haemorrhage and fatigue, these individuals find it difficult to keep a job that requires physical activity such as farming and construction work. As a result their finances are affected and buying nutritious foods becomes a

challenge. Consequently, such individuals apply for a disability grant to enable them with their finances, however the grant is usually not enough but admittedly helpful as some of the individuals in this study cohort reported. These social pressures in addition to HHT symptoms may lead HHT patients feeling depressed and inadequate which in turn may lead to other health issues.

Although there are not many published articles on mortality of HHT, many researchers have concluded that life expectancy in HHT is decreased due to the presence of AVMs, especially cerebral AVMs in children and young adults. Moreover, there is an increased mortality from maternal deaths associated with pregnancy (Shovlin, 2010). The strongest mortality predictor is the severity of gastrointestinal and nasal haemorrhage. It was suggested based on reports that HHT1 patients have a more severe disease phenotype than those with HHT2 (Kjeldsen *et al*., 1999).

Unfortunately there is no treatment at present that can prevent the development of telangiectases and AVMs. However, through early detection, serious complications that occur due to the presence of PAVM and CAVM can be prevented (Kjeldsen *et al*., 1999). Depending on their severity HHT symptoms can impact negatively on the patient's lifestyle. It is therefore important to not only look at HHT patients from the medical point of view, but also the social point of view to enable the affected individuals cope with HHT in a positive way. Part of coping with HHT is to understand HHT physiologically and genetically and this is achieved by a genetic counsellor.

THE AIM OF THE STUDY

HHT is an autosomal dominant genetic disorder, information at molecular level is thus vital to diagnose and manage patients. It has been extensively reported that mutations of the *ENG* gene (Gallione *et al*., 1998; Gallione *et al.,* 2000; Chen *et al*., 2013; Tual-Chalot *et al*., 2015) cause severe life threatening symptoms, such as the presence of AVMs which are more common in patients with *ENG gene* mutations than in the other HHT associated genes. Therefore this study will be focusing on the *ENG* gene. Genetic studies on HHT patients in Africa (and especially on Africans) are few, because it is a rare disorder and mostly under-diagnosed (Gallione *et al.,* 2000 and Westermann *et al*., 2003). Mutations in exon regions of the *ENG* gene have been abundantly documented (Table 3). Therefore, the aim of the study **is to detect splice site and exon region mutations present in the *ENG* gene of the family members affected with HHT. To determine the disease causality, the presence of detected *ENG* gene mutations will be compared between HHT affected and unaffected family members.**

CHAPTER 3

RESEARCH

METHODOLOGY

Study design

Hereditary haemorrhagic telangiectasia (HHT) is an autosomal dominant bleeding disorder. Given that this disorder shows a known pattern of inheritance within a family, this research study is a descriptive family study. This study will take into account the physical HHT associated characteristics as well as demographic information such as age and gender of the participants. The studied family consisted of 11 participating family members with seven members affected with HHT and four were unaffected (Table 4). Family studies help identify the genetic makeup of selected individuals with shared genetic backgrounds permitting a more accurate identification of possible inherited mutations.

TABLE 4: PARTIPANTS DEMOGRAPHICS INFORMATION AND HHT SYMPTOMS

PARTICIPANT NUMBER	GENDER	ETHNICITY	AGE	HHT STATUS	EPITAXIS (Spont. Or Reccur)	TELANGIECTASES	INTERNAL LESIONS (AVMS)
2	M	C	36	Pos	Severe, 2 times weekly	Oral Cavity, Nose	None
3	M	C	18	Pos	Severe, 2 times weekly	Nose	None
4	F	C	54	Neg	None	None	None
5	F	C	42	Pos	Severe, 2 times weekly	Lips, Oral Cavity, Nose	HAVM
7	F	C	31	Neg	None	None	None
8	M	C	36	Neg	None	None	None
9	M	C	62	Pos	Severe, 2 times daily	Lips, Oral Cavity, Nose	PAVM
11	M	C	40	Neg	Severe, 2 times weekly	Nose	None
12	F	C	9	Pos	Severe nose bleeds		
13	F	C	20	Pos	Severe, 2 times a day	Nose	None
15	M	C	45	Pos	Mild, once a month for 1-2 hrs	None	GI (with or without bleeding)

C: Coloured (mixed ancestry) M: Male F: Female Spont: Spontaneous Reccur: Recurrence

Pos: Positive Neg: Negative PAVM: Pulmonary arteriovenous malformation GI: Gastrointestinal

Sample group

The sample group consisted of a South African family who are of mixed ancestry or are referred to as coloured in South Africa (Appendix 3). The coloured population of the Free State is 3.1% of the total Free State population, and 8.9% of the total South African population (Statistics South Africa, 2012).

The sample group included nuclear family members (first degree relatives) in order for a pattern of inheritance to be established in terms of the inheritance of possible mutation(s). It is a five generational family living in the area of Koffiefontein in the Free State, where the members affected with HHT portray an autosomal dominant pattern of inheritance. Participant 9 is the proband of the family because he was the first individual to be diagnosed with HHT within this family (Figure 20). The affected family members have been formerly diagnosed with the disorder based on clinical symptoms by a haematologist using the Curaçao criteria and most have been attending the haematology clinic at Univesitas Hospital for years before the diagnoses was made.

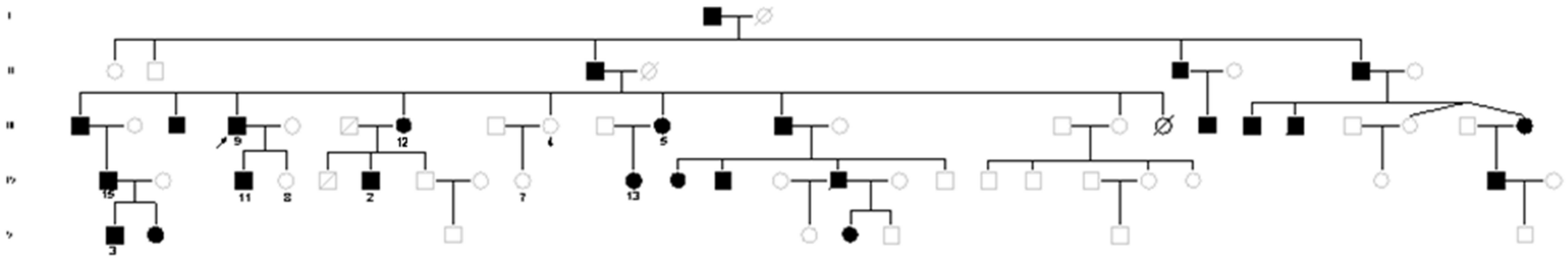


Figure 20: A five generation family pedigree chart showing the autosomal dominant pattern of inheritance. The participants that gave informed consent and donated blood for the study are indicated with study numbers, family members where numbers have been omitted did not donate blood (PedigreeXP (version 2.1.0.174)). The black squares and circles represent affected males and females respectively. The white squares and circles represent the unaffected males and females respectively. The arrow indicates that the proband is participant 9.

Ethical Aspects

Ethical approval for this study was granted by the Health Sciences Research Ethics Committee (HSREC) at the University of the Free on 25th May 2016 (Appendix 1). The ethics approval number for the study is HSREC 79/2016. All samples were originally collected in 2015 as part of a study titled “A clinical and genetic analysis of hereditary haemorrhagic telangiectasia in the Free State and Northern Cape” with the ethics number ECUFS 132/2011. The blood samples were stabilized in RNA*later* but no Nucleic acid isolation was performed in the 2015 study. For the purpose of this thesis, archived, -20ºC RNA*later* stored samples labelled with study numbers and not names were received, delinked from participants’ identities and personal information. Feedback will only be given to the participants at the end of the study if they request it, since the results will not change the treatment or disease outcome of the participants. All the participating individuals voluntarily signed informed consent forms after fully understanding what the project entails. This study was performed only for research purposes and not as a diagnostic service.

Summary of Study Procedure

Blood samples were collected in EDTA tubes by the principle investigator of the previous study a study titled “A clinical and genetic analysis of hereditary haemorrhagic telangiectasia in the Free State and Northern Cape”. Peripheral blood for each participant was transferred to a micro-centrifuge tube with RNA*later* and stored in a -20ºC freezer at the University of the Free State, in the Department of Haematology and Cell Biology. RNA*later* is an aqueous, nontoxic tissue storage reagent that rapidly permeates tissues to stabilize and protect cellular RNA. RNA*later* solution minimizes the need to immediately process tissue samples or to freeze samples in liquid nitrogen for later processing. The sample collection initially

occurred in 2015. All 11 participants signed an informed consent form; giving permission for their genetic material to be stored for future research on HHT (Figure 21). A formal application for this study to take place was sent to the Ethics Committee of the Faculty of Health Sciences at the University of the Free State (ECUFS) was filed and ethical approval was granted for this study (HSREC 79/2016).

For the purpose of this study the principle investigator (MSc candidate) performed all the techniques and methods to follow. Ribonucleic acid (RNA) was isolated and synthesised into cDNA. Primers were designed and used to amplify sections of the entire *ENG* gene using polymerase chain reaction (PCR) after optimization. Agarose gel electrophoresis was used to examine the quality and integrity of the resulting PCR product. The PCR product was then sequenced using Sanger sequencing and the resulting data was analysed using LALIGN, MEGA, ensembl and other software (mentioned later).

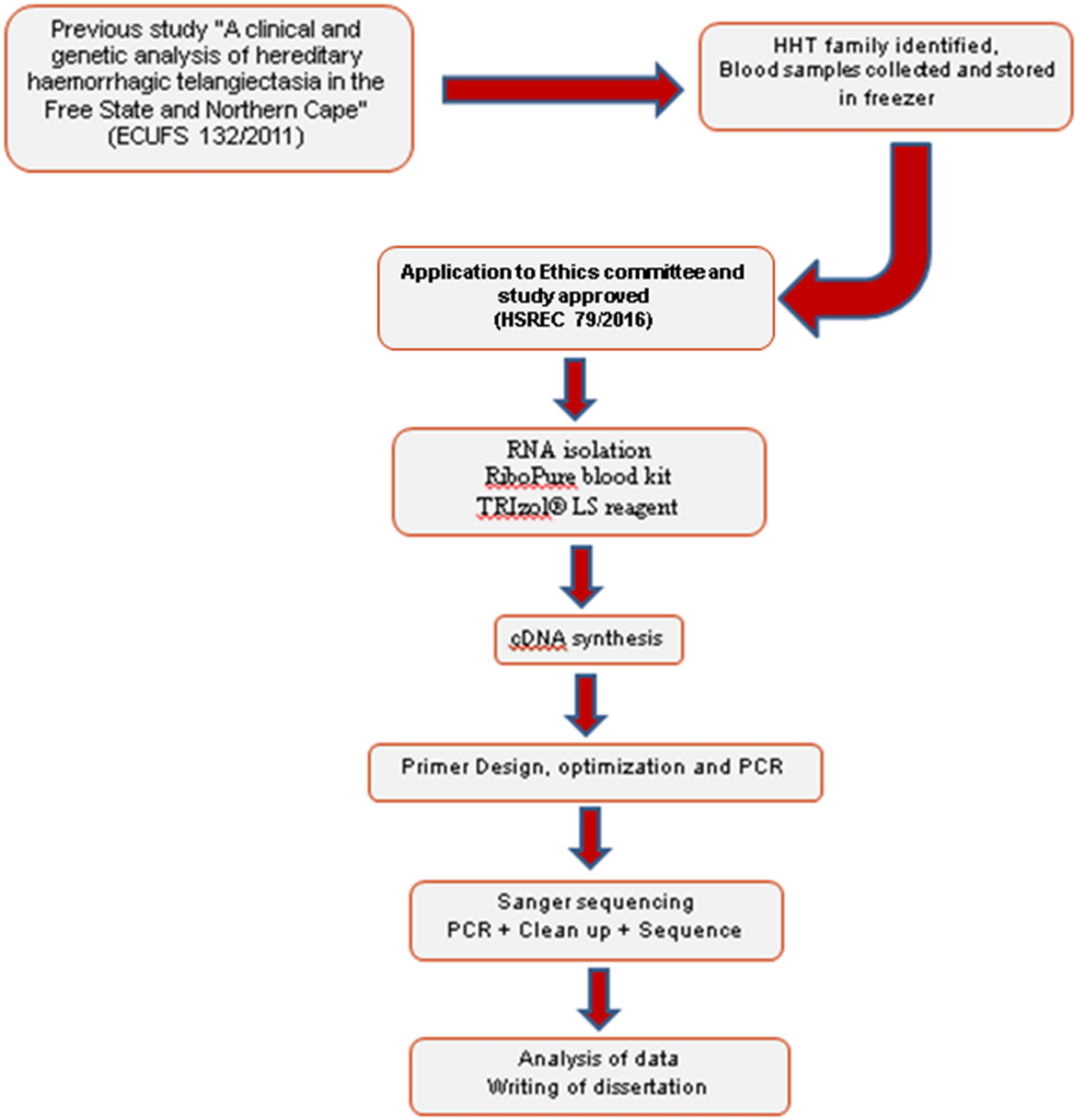


Figure 21: Flow chart demonstrating the various steps of the study procedures. The principle investigator (MSc candidate) of this study performed all the processes following the ethics application for study HSREC 79/2016. This includes RNA isolation, cDNA synthesis, primer design and optimization, sequencing, data analysis and writing of the dissertation.

Measurements

All the laboratory work was performed in the Human Genetics Laboratory in the Department of Genetics, Faculty of Natural and Agricultural sciences of the University of the Free State.

Pedigree analysis

A family tree was generated using GenoPro 2016 (version 3.0.0.7) software. It was later transferred to PedigreeXP (version 2.1.0.174) because this programme follows the international criteria for medical pedigrees (Bennett *et al*., 2008). The clinical information of HHT participants and the demographic information of all the participants has been taken into account while drawing the family tree, only the names have been omitted to ensure participant confidentiality as is ethically required.

Sampling

Blood samples of about 5 to 10 mL were collected in EDTA tubes (BD Vacutainer, Reference no: 368861, Becton Dickinson, South Africa) from 11 participating family members in 2015. One tube of blood, were centrifuged at high speed (16000 rpm) at room temperature (RM) for 15 minutes to separate the whole blood components to upper plasma layer, white blood cells (buffy coat) in the middle and a lower red blood cells layer. The upper plasma layer of each sample was pipetted off and discarded. The buffy coat of about 500 µL was then collected using a pipette and mixed with 1.2 mL RNA*later*® Solution (LifeTechnologies, ThermoFisher, US) in a 2 mL micro-centrifuge tube and stored at -20ºC in the freezer. RNA*later*® stabilizes RNA until it is extracted. This procedure was performed under the supervision of the Principal

Investigator if the ECUFS 132/2011 study in 2015.The RNA*later*® Solution samples were stored at -20º C and subsequently used for RNA isolation and converted to cDNA. The cDNA was then used to screen study participant's exon-exon splice site spanning regions.

RNA extraction

RNA was isolated from the RNA*later*® stabilised samples using two methods. Both the RiboPure™ Blood Kit (Ambion, LifeTechnologies, Thermo Fisher Scientific) and TRIzol® LS Reagent (Invitrogen, Thermo Fisher Scientific) were used to isolate RNA. The results yielded from the two methods were compared for quantity and quality of RNA as well as for sequencing quality.

Method using the RiboPure™ Blood Kit

The RiboPure™ Blood Kit (Ambion, LifeTechnologies, Thermo Fisher Scientific) was used to isolate RNA according to manufacturer's instructions. Before extraction all areas of the lab as well as equipment were cleaned with 70% ethanol and then with RNase Zap™ (Invitrogen, Thermo Fisher Scientific) to eliminate contaminants and RNA on the surface area and on the pipettes. Sterilized filtered tips were used and all the micro-centrifuge tubes were steam autoclaved at 121ºC for 20 minutes.

The RNA*later*® Solution samples were centrifuged for 1 min at 15 000 rpm in a micro-centrifuge. The blood cells and plasma proteins formed a reddish-brown pellet. The supernatant was discarded. Remaining blood cells were lysed by adding 800 µL of lysis solution and 50 µL sodium acetate solution to the cell pellet that originally contained about to 300–500 µL of whole blood. The samples were vortexed vigorously to lyse the blood cells. A volume of 500 µL of acid-phenol:chloroform was

added to the cell lysate and vortexed for 30 sec. The mixture was then incubated at for 5 min and centrifuged for 1 min at 12 000 rpm to separate the aqueous and organic phases. The aqueous upper phase containing the RNA was transferred to a new 2 mL tube with 600 µL of 100% cold ethanol and vortexed thoroughly. An amount of 700 µL of each sample was applied to the labelled filter cartridge assembly (collection tube and filter cartridge) and centrifuged for 5-10 sec each time until the complete sample mixture had passed the column, and the flow through was discarded. A volume of 700 µL of Wash Solution 1 was added to the filter cartridge assemblies, centrifuged for 5-10 sec each time until all the sample mixture had passed the column and the flow through was discarded. Subsequently, 1400 µL (700 µL X 2) of Wash Solution 2 consisting of 2:3 volume Wash Solution 2 and 1:3 100% ethanol was added to the filter cartridge assemblies, centrifuged for 5-10 sec each time until all the sample mixture had passed the column and the flow through was discarded. The filter cartridge for each sample was transferred to a new labelled collection tube and eluted with 50 µL of elution solution preheated to 75º C and left at RM for 20 sec before the centrifugation of 30 sec at 16000 rpm to recover the RNA. An additional 50 µL of preheated elution solution was added to the filter cartridge and centrifuged for 1 min at 16000 rpm to recover remaining RNA and elution solution in the collection tube.

A DNase I digestion step was included to remove genomic DNA from the eluted RNA. A volume of 5µL 20 X DNase Buffer and 1 µL DNase I was added to the eluted RNA and incubated for 30 min 37ºC. Subsequently, 20 µL of DNase inactivation reagent was added, vortexed briefly and stored at RM for 2 min. The samples were then vortexed once during the 2 min step to re-suspend the DNase inactivation

reagent. After that, the samples were centrifuged for 1 min at 15 000 rpm in order to pellet the DNase inactivation reagent and the RNA solution was transferred to a new tube.

Method using the TRIzol® LS Reagent

The TRIzol® LS Reagent manual was followed for the following procedure on the ThermoFisher website (https://www.thermofisher.com/order/catalog/product/10296028). The samples containing RNA*later*® Solution were centrifuged for 1 min at 16000 rpm and a pellet was formed at the bottom of the tube. A volume of 750 µL of TRIzol® LS Reagent (Invitrogen, Thermo Fisher Scientific) was added to the cell pellet. Each sample was then homogenised by pipetting the suspension up and down several times. The homogenised samples were incubated for 5 min at RM to permit complete dissociation of the nucleoprotein complex. There after 200 µL of chloroform was added to the 750 µL of TRIzol® LS Reagent and samples were vigorously shaken for 15 sec before incubation of 10 min at RM. A centrifugation step followed at 12 000 rpm for 15 min at 4ºC separating the sample into different phases, the aqueous phase, interphase and lower red phenol-chloroform phase. The aqueous phase was removed by pipetting and placed into a new 2 mL tube. The remainder of the sample was discarded. After the addition of 500 µL of 100% isopropanol to the aqueous phase the samples were incubated at RM for 10 min. The samples were centrifuged at 12 000 rpm for 10 minutes at 4°C and the supernatant was discarded. The RNA pellet was washed in 1000 µL of 75% ethanol and vortexed. The sample was centrifuged at 7500 rpm for 5 min at 4ºC and the supernatant was discarded. The RNA pellet was left to air dry for 10 min and re-suspended in 30 µL RNase-free water. Following re-suspension the samples were

incubated in a heating block at 60ºC for 12 min, the samples' RNA concentration was measured and stored at -20ºC.

RNA clean up

The Trizol method extracted samples were treated with (cetyl trimethylammonium bromide) CTAB (Sigma-Aldrich catalog Number 57-09-0) to rid the samples of any DNA left in the RNA samples. A volume of 800 µL of CTAB buffer (100 mg/mL) was added to the samples and incubated for 10 min at 65ºC. Subsequently, 333 µL of a 6 M potassium acetate solution was added before incubation on ice for 30 min. A centrifugation step followed at 16 000 rpm for 10 min and the supernatant was collected and placed in a new tube and the rest of the sample was discarded. To precipitate the RNA 500 µL isopropanol was added to the sample and incubated for an hour before centrifugation at 16 000 rpm for 10 min. The samples were washed with 1 mL of 70% of ethanol twice, incubated for 2 min and centrifuged at 12 000 rpm each time. The pellet was then re-dissolved in 50 µL TE buffer.

RNA concentration determination

The concentration of RNA was assessed using the Qubit™3.0 Fluorometer (Invitrogen, Life Technologies). The Qubit™ 3.0 Fluorometer utilizes fluorescent dyes that emit a light signal only when bound to a molecule such as RNA or DNA. The raw fluorescence (RFU) values of diluted standard 1 and standard 2 was also measured to enable a standard curve to be plotted. The concentrations of the other samples were read from this standard curve. The Qubit™3.0 Fluorometer measures only intact RNA and DNA.

According to the manufacturer's instructions, 597 µL (199 µL X 3) of Qubit® RNA HS buffer and 3 µL (1 µL X 3) of Qubit® RNA HS reagent were used to make a working solution for each sample and standards. In using the Qubit® RNA HS Assay kit 10 µL of Standards 1 and 2 were each diluted with 190 µL of the working solution and then measured. A volume of 190 µL of the working solution was added to 10 µL of the sample (each sample was measured twice). Measurements were taken in two categories namely the RNA high sensitivity and RNA high broad range. Both measurements were recorded and compared.

RNA transcription to cDNA

For good quality cDNA synthesis, 10-15 µg of RNA is needed. Too much or too little cDNA concentrations are known to affect downstream applications negatively such as Sanger sequencing where too much background noise may occur (Williams *et al.,* 1992). Additionally, according to RNA Isolation and RT-PCR Analysis protocol, 1 µL of RNA should have 5-500 ng/µL for quality cDNA synthesis. All the RNA samples of the participants were measured and the concentrations were between 100-155 ng/µL.

The high capacity cDNA Reverse Transcription kit (Applied Biosystems, Thermo Fisher Scientific), was used to transcribe the RNA into complementary DNA (cDNA). For a 20 µL reaction, 2 µL of 10X buffer, 0.8 µL of 25X dNTP mix, 2 µL of 10X Random hexamer primers, 1 µL of MultiScribe™ reverse transcriptase, 1 µL of RNase inhibitor and 3.2 µL of nuclease free water (Gibco®, Life Technologies) were used for each sample. The 10 µL master mix was then added to 10 µL to the RNA sample (all above 100 ng/µL) in a 0.5 micro-centrifuge tube. The mixture was pipetted up and down to mix and briefly centrifuged. The samples were placed in a

SimpliAmp™ Thermocycler (Applied Biosystems®, Life Technologies). The following cycling conditions were applied: 25ºC for 10 min, 37ºC for 120 min, 85ºC for 5 min and 4ºC holding temperature.

To determine an estimate of the quality of the cDNA of each patient, a 2% (w/v) agarose gel using 50X TAE (Tris base, acetic acid and EDTA; 40 mM Tris 20 mM acetic acid, 1 mM EDTA, pH 8) buffer stained with GR Green Nucleic Acid dye (Excellegen) was used in electrophoresis. The samples were mixed with Blue/Orange loading dye (6X) (Promega catalog no. G1881) and loaded in the electrophoresis tank with a GeneRuler 100 bp DNA Ladder (Thermo Scientific, Catalog number: SM0241). Electrophoresis was performed for 30 min at 100 V. Samples are then diluted to a concentration of 15-45 ng/µL with nuclease free water.

cDNA sequencing primer design

cDNA sequence Primers (Table 5) were designed with the online Integrated DNA Technologies (IDT, USA) website to span exon splice site regions with additional padding of 500bp. The primers were designed within a boundary of certain parameters as follows. The melting temperature (Tm) had to be between 55ºC and 65ºC and the GC content had to be between 40 to 60%. The homodimer and heterodimer formed had to be less than or equal to 4 base pairs (bp) with exceptions to a few primers. The lengths of the primers were between 17 to 24 bp (Table 5). All the above parameters were checked using IDT OligoAnalyzer Tool version 3.1 (www.idtdna.com) and the Tm was determined. Primers selecting for splice site boundaries of exon regions 2, 3, 4, 6, 7, 8 and 12 were used as described by Dekeishi *et al*. (2002). The standard nucleotide BLAST® webpage of the National

Center for Biotechnology Information (NCBI) was used to ensure that the selected primers were specific for the human *ENG* gene and that the primer pairs were specific for the selected regions. The product size as well as the Tm was determined using this tool. All the primers were initially rehydrated to 100 µM, and then diluted to 10 µM. The latter concentration was used for polymerase chain reaction (PCR) amplification. For sequencing, the primers were diluted to 3.2 µM.

TABLE 5: PRIMER SEQUENCES. The exons spanning splice region, sequence, length, Tm, heterodimer, homodimer and GC content.

EXON	PRIMERS	SEQUENCE	LENGTH (15-24)	Tm (55-65)	HETERODIMER	HOMODIMER	GC% (40-60)
Exon 1	FOWARD	GTGCTTGGGGAGACAAG	17	56	3	4	58.8
	REVERSE	GGCTTTCTTTCAACACTGA	19			3	42.7
Exon 2	FOWARD	ACGTTTGGAAAGTAGGAGTC	20	56	3	4	45
	REVERSE	AAATGCCACCTCTTATGG	18			3	44.4
Exon 3	FOWARD	TGGGTGGCACAACCTAT	17	56	4	3	52.9
	REVERSE	CAGAGATGGACAGTAGGGA	19			2	52.6
Exon 4	FOWARD	CAAATTACTTCCTGACCTCC	20	56	3	4	45
	REVERSE	CAGAACCTGGCATATTCC	18			4	50
Exon 5	FOWARD	CTATCTTTGGCTGTGGGTGAG	21	58	4	2	50
	REVERSE	TAGGAGAAAGCGACTGTGCT	20			3	50
Exon 6	FOWARD	TCCCATAAACCCACACCTG	19	56	4	2	52.6
	REVERSE	GGAAACTTCCCTGATCCAGAGGTT	24			4	50
Exon 7-8	FOWARD	CTGTGGCACAGACTGTGT	18	56	5	5	55.6
	REVERSE	CTAGGACCCCAAGAGTCTT	19			4	52.6
Exon 9A-9B	FOWARD	GTTGTGGTCAGTCCTTGGT	19	58	4	2	52.6
	REVERSE	TGGGGAGGAACAGGCATCAT	20			2	55
Exon 10-11	FOWARD	GAGAGTCAGGCAACTCCACA	20	56	4	4	55
	REVERSE	AAACCACAGACCTGGAAGCTC	21			4	52.4
Exon 12	FOWARD	GATCTTCCAGGACTCACC	18	56	3	3	55.6
	REVERSE	CCCCTTGCCATGTGCTA	17			3	58.8
Exon 13-14	FOWARD	ATGGCTGAGCAGAGCTG	17	60	4	4	58.8
	REVERSE	TGGCTTTGAATGTCACTGCC	20			3	50

Conventional polymerase chain reaction (PCR)

AmpliTaq Gold® 360 Master Mix (Applied Biosystems, Life Technologies) was used to prepare PCR reactions. The PCR reaction consisted of: 12.5 µL of AmpliTaq Gold® master mix, 2 µL of the forward and reverse primers (concentration of primers 10 pmol/µL), 6.5 µL of PCR grade water and 2 µL of cDNA template (15 ng/µL concentration) with a total of 25 µL reaction per individual for each exon. The following PCR conditions were utilized: the initial denaturation temperature was 95ºC for 10 min, 95ºC for 30 sec, and the annealing temperature varies in each exon (Table 5) and cooled at 72ºC for 1 min, process ran for 35 cycles. The final extension temperature was 72ºC for 7 min and the final hold temperature was 4ºC. The process was run on a 96 well plate. Electrophoresis was performed using a 2% (w/v) agarose gel (according to the described method on page 75).

PCR Clean Up

An illustra™ ExoStar 1-Step® (GE Healthcare) enzymatic PCR and sequencing clean-up was performed according to the manufacturer's instructions on the PCR reactions. After the illustra™ ExoStar 1-Step enzyme were thawed, 2 µL was added to 5 µL of the completed PCR reaction and incubated at 37ºC for 15 min and at 80ºC for 15 min, in order for the enzymes to deactivate.

Sequencing reactions

The BigDye® Terminator v3.1 Sequencing Standard Kit (Applied Biosystems, Life Technologies) was used for sequencing. A ⅛ reaction mix was used that consisted of 1 µL premix, 1 µL primer, 2 µL 5x sequencing buffer and 4 µL PCR grade water.

The reaction mix was added to 2 µL PCR product making a total volume of 10 µL. The second PCR regimen was as follows: an initial denaturation of 96ºC for 1 min, 96ºC for 10 sec, the annealing temperatures were different for each exon and 60ºC for 4 min. This process runs for 25 cycles and the holding temperature was 4ºC. The whole process was run on a 96 well plate.

Sequence reaction preparation and loading onto sequencer

A volume of 5 µL from a 125 mM EDTA and 60 µL of 100% ethanol solution was added to each sample, slightly vortexed and centrifuged for 10 sec. The samples were incubated at RM for 15 min in the dark. Thereafter, the samples were centrifuged at 15 000 rpm for 30 min and then the supernatant was removed using a pipette. As a wash step, 200 µL 70% ethanol was added and vortexed before being centrifuged for 20 min. The supernatant was removed using a pipette and samples were left to dry overnight in the dark. The following day, 30 µL of Hi Di™ injection buffer (Thermo Fisher) was added and mixed by pipetting, and then centrifuged briefly. Samples were denatured for 5 min at 94ºC and snap cooled for 5 min in a cold block. Sequencing of all samples were performed on the ABI sequencing machine 3130 Genetic Analyzer (Applied Biosystems, Life Technologies).

Sequence data analysis

The sequences were analysed using LALIGN server algorithm (https://embnet.vital-itch/software/LALIGN_form.html). This server implements the algorithm described by Huang & Miller, (1991).The programme requires the insertion of the reference sequence which is compared against sample sequence (both forward sequences), Mutations are found when a mismatch is observed.

Ensembl is a genome browser used to identify sequence variation and predicts regulatory function (www.ensembl.org). When mutations are identified, ensemble was used to find the type of mutation it was and if the mutation was novel or had been detected previously. SIFT (sorting intolerant from tolerant) (http://sift.jvci.org/) and Polyphen scores (http://genetics.bwh.harvard.edu/pph2/) were included in this programme and were used to identify if the mutations were benign or pathogenic. Citations that were available in this website allowed investigation on the functional effect of the resulting protein.

Clinvar is a bio-informatic tool that merges a mutation to a phenotype (www.ncbi.nhlm.gov/clinvar/intro/) was used after specific mutations were identified. This tool incorporates articles on the mutation/s and the phenotype it is associated to. Additionally, the number of times the mutation/s has been identified and in which population group the mutations have been reported is also provided.

Method Optimization

RNA extraction methods and concentrations

RiboPure™ Blood Kit and TRIzol® LS Reagent were used to extract RNA from stored RNA*later*® Solution samples. Study sample K, was used for optimization purposes. Several blood samples were obtained from a non HHT, non- study family related individual giving informed consent to participate in the study, following all ethical requirements. The blood samples were stored in RNA*later*® Solution as described on page 75. The RNA samples were measured using the high sensitivity and broad ranges in ng/µL but yielded very low concentration results for both methods. To ensure the validity of the extraction method to be used, both extraction

methods were repeated in triplicates and the readings for each sample was taken three times and the average was recorded.

Based on the initial results, the Trizol method of extraction yielded a higher RNA concentration of 60.8 ng/μL. The higher concentration of the Trizol was expected due to the simplicity of the extraction method and the lack of using spin columns that select for longer strands of RNA. Another factor considered to account on the differences between the two methods is re-suspension of RNA and volume used. In RiboPure™ Blood Kit, the initial sample was eluted with 100 μL elution solution yielding a concentration of 52.8 ng/μL for sample one. This is a low concentration compared to the optimal RNA concentration of at least 100 ng/μL required for quality cDNA synthesis. In the attempt to increase the RNA concentration, an overall volume of 30 μL of elution solution was added to elute the RNA ensuring the resulting RNA was less diluted. But the RNA concentration for the last two attempts was less than the first one even after reducing the volume of elution solution added to 30 μL.

Three more attempts for each extraction method to validate the previous results were done to determine the method of choice for the patients' samples. Quality (measured according to cDNA PCR electrophoresis results) and quantity of the resulting RNA is pivotal to ensure that the downstream applications and ultimately sequencing of the cDNA are not hindered due to poor RNA quality. The quantity of the RNA is not as vital as the quality, but essential to complete the molecular analysis for each participant. Only a single change was made to improve the RNA concentration using the RiboPure™ Blood Kit method (Table 6). During the addition of the elution solution, the samples were centrifuged for 1 min twice instead of 30 sec the first time and 1 min the second time to ensure that all the RNA in the spin

column was in the eluted. The volume of the elution solution added varied in the first and the last two samples so as to compare the RNA concentration. The Trizol method remained unchanged except for the usage of cold ethanol. The RNA concentrations (Table 6) of the RiboPure™ Blood Kit improved and were better than the results of the Trizol method. Readings for each sample was taken three times and the average was recorded.

TABLE 6: RNA EXTRACTIONS CONCENTRATIONS, BASED ON AN AVERAGE OF THREE READINGS PER SAMPLE.

RNA extraction Methods	Samples	Elution solution or nuclease free H_2O* volume (µL)	Broad range (ng/µL)	High sensitivity (ng/µL)	Volume measured (µL)
RiboPure™ Blood Kit	K2	15	140	13.3	2
	K5	100	79.9	6.77	2
	K6	100	187	17.4	2
Mean			135.6	12.5	
STDEV			53.68	5.36	
S.E.M			30.99	3.09	
TRIzol® LS Reagent	K1	30*	4.24	Out of range	10
	K2T	20*	13.3	0.646	10
	K3T	30*	23.1	0.698	10
Mean			13.5	0.43	
STDEV			9.43	0.37	

S.E.M			5.45	0.21	

RNA concentration results of the second phase attempts in triplicates for each extraction method.

*Nuclease free H_2O was used dilute samples that were extracted with TRIzol® LS Reagent.

STDEV: standard deviation S.E.M: standard error of the mean

There are various reasons why the TRIzol® LS Reagent yielded less RNA in the second round. Firstly, as mentioned before the Fluorometer measures only intact RNA and not degraded RNA. Degraded RNA is mainly caused by sample manipulation before the addition of Trizol of which was unlikely given all samples stored in RNA*later*® Solution were processed immediately after being taken out of storage (-20ºC freezer). Secondly, samples of the second phase may have not have homogenised completely with the RNase-free water and therefore more time was needed during the incubation period in the heating block. Thirdly, the RiboPure™ Blood Kit was developed and optimized to isolate the highest quality and quantity of RNA from samples stored in RNA*later*® Solution.

The buffy coat from blood samples that were stored in RNA*later* in relation to RiboPure™ Blood kit, was expected to work better as they optimized by the same company to efficiently and effectively yield quality RNA for downstream applications. Schwochow *et al.* (2012) compared four RNA extraction methods including the RiboPure™ Blood Kit and TRIzol® LS Reagent in terms of RNA concentration and quality. They concluded that RiboPure™ Blood had a significantly higher concentration than TRIzol® LS however, the quality (260/280) scored the same. Under oxidising conditions, organic compounds such as phenol may cross-link with RNA hence limiting RNA isolation (Schrader *et al.,* 2012). This may serve as an explanation for the low RNA concentration observed in TRIzol® LS Reagent. In

conclusion, the low yield of RNA attributed to TRIzol® LS Reagent may be due to a number of factors previously mentioned.

cDNA synthesis

Complementary DNA (cDNA) synthesis with or without RNase inhibitor had no effect on the yield and quality of the RNA. However, RNase inhibitor is essential for reducing degradation of RNA. Secondly, it is important for destroying RNases present on tips, tubes that may influence downstream applications. Therefore, the use of RNase inhibitor is a precautionary measure to inhibit and control RNA enzymatic manipulations.

DNA contamination prevention

Prior to cDNA synthesis, method optimization samples (K2, K3T, K4T, K5 and K6) were treated with DNase treatment and CTAB. DNase treatment and CTAB are methods containing reagents that digest contaminating DNA present in extracted RNA using the RiboPure™ Blood kit and TRIzol® LS Reagent respectively. The use of these DNA digesting reagents were optional and weren't deemed necessary according to the manufacturer's instructions. In a study performed by Schwochow *et al.* (2012) showed that the RiboPure™ Blood kit method yielded better quality RNA, yet it proved to contain higher DNA contamination than TRIzol® LS Reagent method isolated samples after PCR. Consequently, less products were amplified compared to TRIzol® LS Reagent method samples (Schwochow *et al.,* 2012). Because of literature reporting on higher DNA contamination to combat contamination in these samples, DNase treatment and CTAB were used after extraction depending on the extraction method used. PCR was performed on the cDNA that resulted from the RNA of optimization samples K3T and K4T, extracted from Trizol did not amplify.

This may be a result of inhibitors or the low concentration of RNA present in the samples before cDNA amplification and not necessarily due to DNA digesting chemicals incorporated in the CTAB method.

After cDNA synthesis the differently treated test samples (K2, K5 and K6) were subjected to conventional PCR using the primer pair for exon 6. The DNase treated sample K6 were compared to the samples that were not treated, K2 and K5, and showed a difference in the intensity of the band on the agarose gel (Figure 22). Samples K2 and K5 bands were much brighter than K6. Based on this result the untreated samples produced a more desirable amplification product than the treated samples. The comparison of the untreated samples sequence results to the reference sequence, resulted in perfect alignment. More untreated samples from all the exons were sequenced and the target regions were clearly amplified and analysed against the reference sequence. Therefore, the use of DNase treatment was not used, as it was found to be unnecessary.

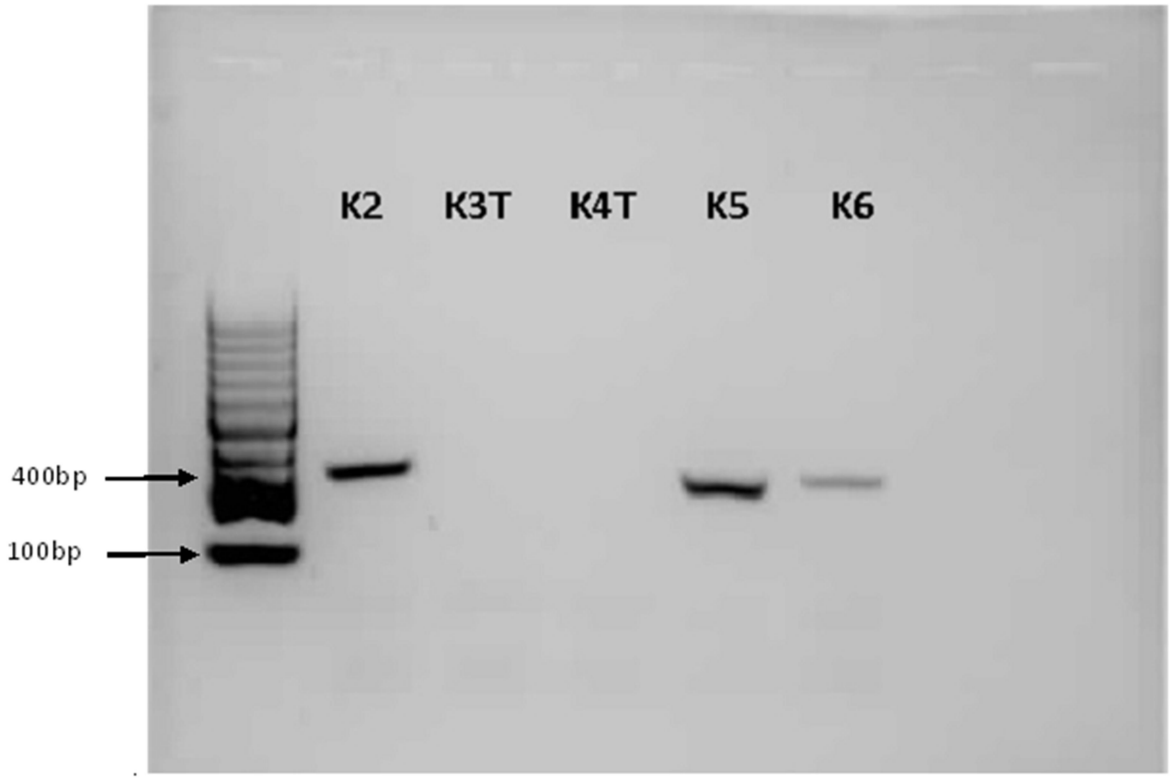
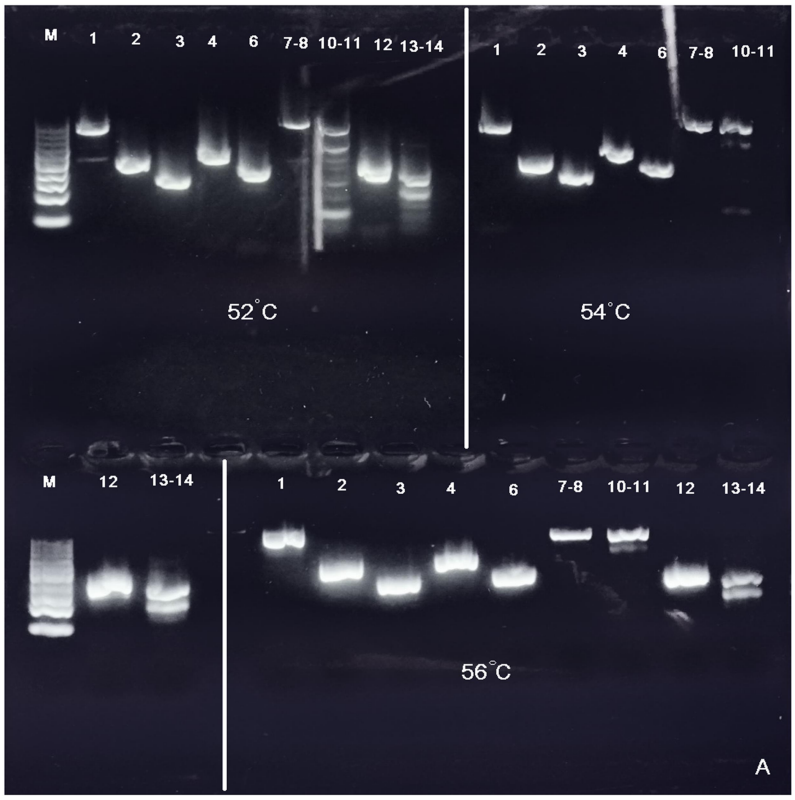


Figure 22: Agarose gel presenting samples treated with reagents within DNase treatment and CTAB methods that digest contaminating DNA present in extracted RNA from using RiboPure™ Blood kit (K2, K5 and K6) and TRIzol® LS Reagent (K3T and K4T) respectively. Marker, Lane 1: sample K2 (400bp), Lane 2: sample K3T, Lane 3: K4T, Lane 4: K5 (350) and Lane 5: K6 (355). None of the samples (K3T and K4T) whose RNA was extracted from Trizol amplified. Untreated samples K2 and K6 bands are brighter than DNase treated sample K6 band.

PCR optimization

Annealing PCR reaction temperatures were optimised using the splice site boundary primers for all the exons together with cDNA by means of a temperature gradient. The following are the temperatures used 52ºC, 54ºC, 56ºC, 58ºC and 60ºC (Figure 23A). Splice site boundary regions of exons 1, 2, 3, 4, 6, 7-8, 10-11 and 12 all amplified optimally at 56ºC and 13-14 at 60ºC. The temperature was chosen based on fragment brightness and the highest temperature producing a fragment, for high specificity of the resulting amplicons. Splice site boundary region for exon 5 produced primer dimers and after several attempts another reverse primer was designed that enabled amplification at 58ºC. However, the resulting fragments from

the amplification were not of good quality. Consequently, the sequence reactions failed. This may have occurred due low sample amplification that resulted from the inefficiency of primers binding at the target region, especially the reverse primer. To solve this problem, the exon 5 site was amplified using the exon 5 forward primer and exon 6 reverse primer, resulting in a longer fragment, which amplified at 58ºC with a better quality fragment. The resulting sequences using the same primer pair revealed good sequences in the exon 5 forward end but the exon 6 reverse primer only presented exon 6 reverse end that did not extend to exon 5 reverse. Since exon 6 reverse did show exon 5 reverse nucleotides, exon 5 reverse primer was used to sequence and resulting sequences are not optimal but analysable (Figure 24). Another primer pair that did not give initial results is exon 9 spanning region of which reverse end is homogenous to many sections within the gene. A second reverse primer was designed and it amplified at 58ºC. However, the resulting bands were of different lengths for each patient which may be accounted to non-specific binding as a result of the homogenous nature of nucleotide bases towards the end of exon 9 (Figure 23B). Figure 23B exon 1 band is not visible but the number of base pairs was established using other agarose gels. Nevertheless, good quality sequences could be obtained and aligned to the reference sequence.



Exon	Base Pairs (bp)
1	900
2	500
3	350
4	680
5	500
6	400
7-8	1000
9A-9B	1600
10-11	1200
12	400
13-14	1350

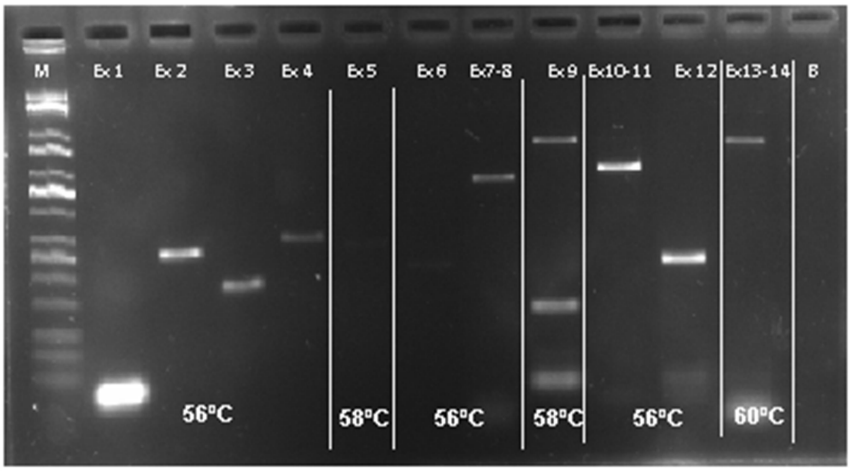


Figure 23A: The temperatures used for PCR optimization of the different exons. A 2% agarose gel depicting the primer pair optimization of splice site boundary exon regions 1, 2, 3, 4, 6, 7-8, 10-11, 12 and 13-14 using gradient temperatures as indicated. The table shows the approximate number of base pairs per exon. B: Exon 1-14 and the number of base pairs of each exon. Lane 1: Marker, Lane 2: exon 1 (band not visible), Lane 3: exon 2, Lane 4: exon 3, Lane 5: exon 4, Lane 6: exon 5, Lane 7: exon 6, Lane 8: exon 7-8, Lane 9: exon 9A-9B, Lane 10: exon 10-11, Lane 11: exon 12, Lane 12: exon 13-14 and Lane 13: Blank.

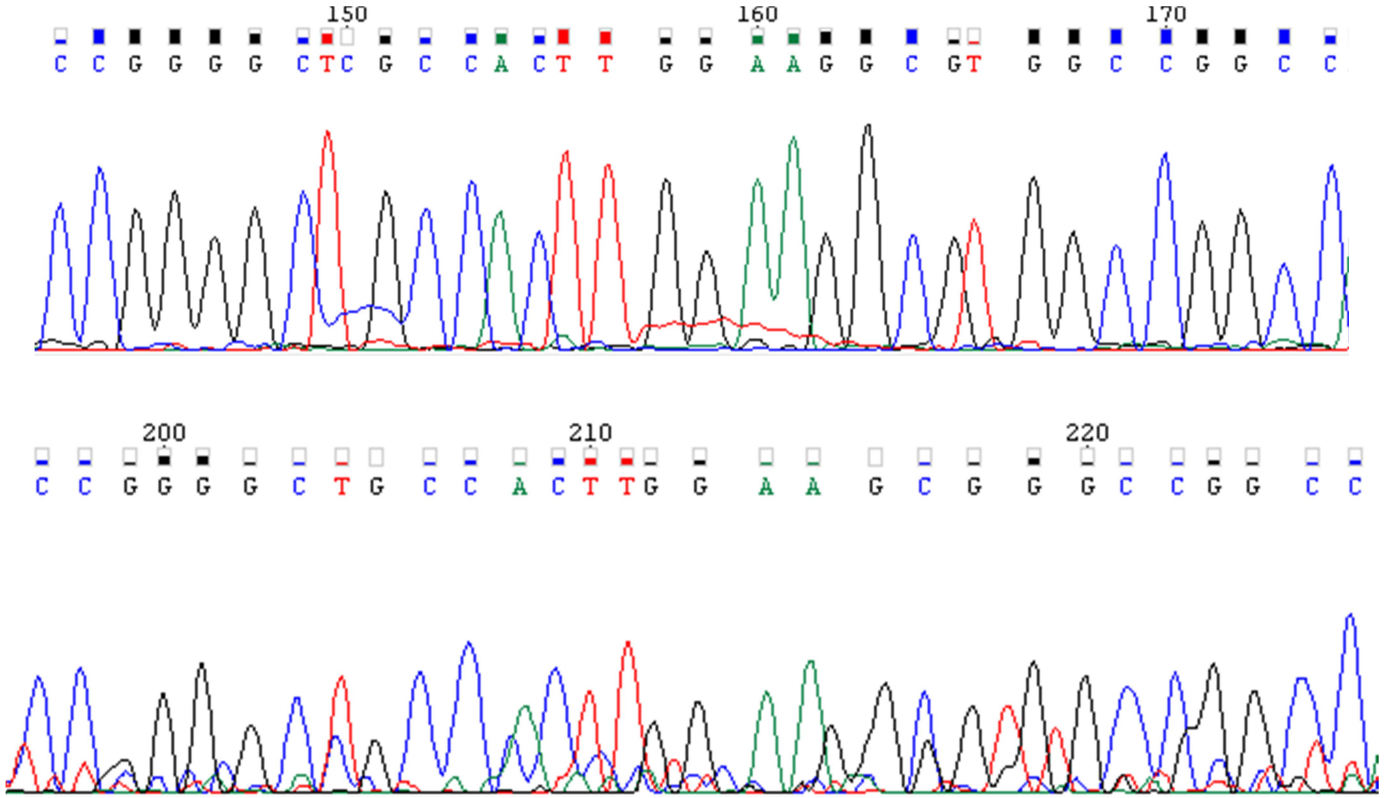


Figure 24:Patient 7 exon 5 forward and reverse sequences. Exon 5 forward (top) sequence was easier to analyse but exon 5 reverse (bottom) was not optimal but analysable.

Conclusion and final methods incorporated

The extraction method chosen to isolate RNA from study samples is the RiboPure™ Blood Kit. This method (Table 6) yielded a high concentration of RNA of more than 100 ng/µL is the required amount for optimal cDNA synthesis. Furthermore, DNase treatment was not used for any of the participant samples because of the low quality fragments that appeared on the agarose gel. Moreover, the use of this treatment was deemed unnecessary because the untreated samples' sequences showed that the target regions amplified aligned with the reference sequence.

CHAPTER 4

RESULTS AND DISCUSSION

Introduction

There are only a few studies that have reported on HHT mutations in the African population. These include Canzonieri *et al.,* (2014b) who described mutations from two African men and Westerman *et al* (2003) that reported on a mutation identified in a Kenyan male. HHT is a rare bleeding disorder and limited data exist on the condition from the African continent. Additionally, because of its rarity, the disorder may be under-reported or misdiagnosed.

Most of the mutations that have been reported on the *ENG* gene are within the exonic and splice site regions. Some have proved to be pathogenic while others benign. Mutations that are pathogenic have been proven to be familial; however, benign mutations that may be linked to HHT have no familial connections.

Pedigree analysis

Pedigree analysis was performed using the GenoPro 2016 (version 3.0.0.7) and PedigreeXP (version 2.1.0.174) software. Based on the pedigree chart (see Figure 22), an autosomal dominant pattern of inheritance is established. Not all family members depicted in the pedigree could be reached and recruited; consequently these individuals were not included as participants in this study. Participants are highlighted with the black numerals. The participants that gave informed consent and donated blood for the study are indicated with study numbers. Family members where numbers have been omitted did not donate blood, either because they are deceased or could not attend the clinic on the days the study recruitment took place. Based on the knowledge of one of the elder family members and the Koffiefontein clinic nurse a comprehensive pedigree chart could be constructed.

Splice site mutations within the study family

Sequence analysis was performed on expressed RNA sequences for splice site boundaries of exon regions 1, 2, 3, 4, 5, 6, 7-8, 9A-9B, 10-11, 12 and 3-14 of the *ENG* gene totalling fourteen mRNA splice site regions. To our knowledge, a study presenting sequence data on splice site regions of the *ENG* gene in a population of African descent has not been attempted or published. Sequencing results from this study resulted in no splice site mutations in the complete *ENG* gene ranging from exon 1 to 14. Raw sequence data was analysed using MEGA (molecular evolutionary genetic analysis) software and can be viewed in Appendix 2. As an example to show the absence of mutations Figure 25, illustrate the exon 3 consensus sequences from all the family members aligned. Additionally to exhibit the electropherogram sequences, Figure 27 is an example of exon 3 of only the forward sequences of two the family members. The splice site region alignment sequence for exon 1 was constructed using LALIGN, between the reference sequence and patient seven (Figure 26).

McDonald *et al.,* (2009) reported that splice site mutations account for approximately 12% of mutations in the *ENG* gene present in patients with HHT. Splice site mutations can be pathogenic and cause major deletions and insertions at the cDNA level. Lesca *et al.* (2004) identified three splice site mutations in French patients; c.689T>C that caused a deletion of exon 5. The second splice mutation, c.1103+3_1103+8 del led to the activation of a cryptic splice site 22 bp upstream in intron 8 consequently causing a frameshift to occur. The third mutation c.1428G>A led to two abnormal transcripts of exon 10 splicing, one excluding exon 10 and the other including intron 11 (Lesca *et al.,* 2004). In another study, 7 splice site mutations were identified in a HHT Dutch population. These mutations are

c.67+1G>A, c.360+1G>A, c.360+5G>A, c.689+2T>C and c.1310delG, located in exons 1, 3, 5 and 8 respectively (Letterboer *et al.,* 2005). Splice mutations are pathogenic and can cause HHT as shown in the studies from Caucasian patients. It is important to note that the lack of splice mutations within this family does not represent all HHT families of African descent.

The absence of splice site mutations in the *ENG* gene as presented in this study is an important result, since the *ENG* splice regions has been associated with HHT in European populations and is also regarded as causal to HHT. This result furthermore highlights the difference in disease causing mechanisms between populations with different ancestry as well as between different families within a population. The complex pathogenesis for HHT such as age dependent expression of HHT and locus heterogeneity as well as several phenotypic mechanisms, including telangiectasia's and AVM's may also explain the discrepancy in mutation analysis between this study and published literature (Bayrak-Toydemir *et al.,* 2004; Abdalla & Letarte, 2006).

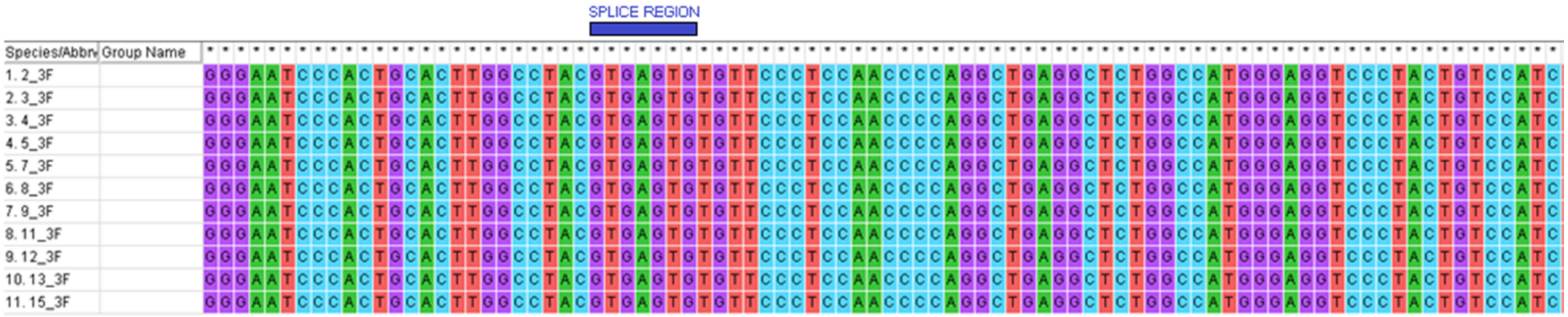


Figure 25: Aligned sequences of splice site boundaries of exon 3 from all the study participants, created using MEGA. No mutations were detected in this region. The splice region is indicated with a blue bar sign at the top. In the species/abbreviation column the first numeral indicates the line number eg.1; the numeral in front of the 'underscore' sign indicates the sample number; and the '_3F' indicates the exon 3 forward strand.

Figure 26: Splice site boundary region for exon 1, illustrating the alignment of the reference sample and patient sample 7 using LALIGN (https://embnet.vital-itch/software/LALIGN_form.html). The highlighted area indicates the splice site of reference (orange) and patient (yellow) samples.

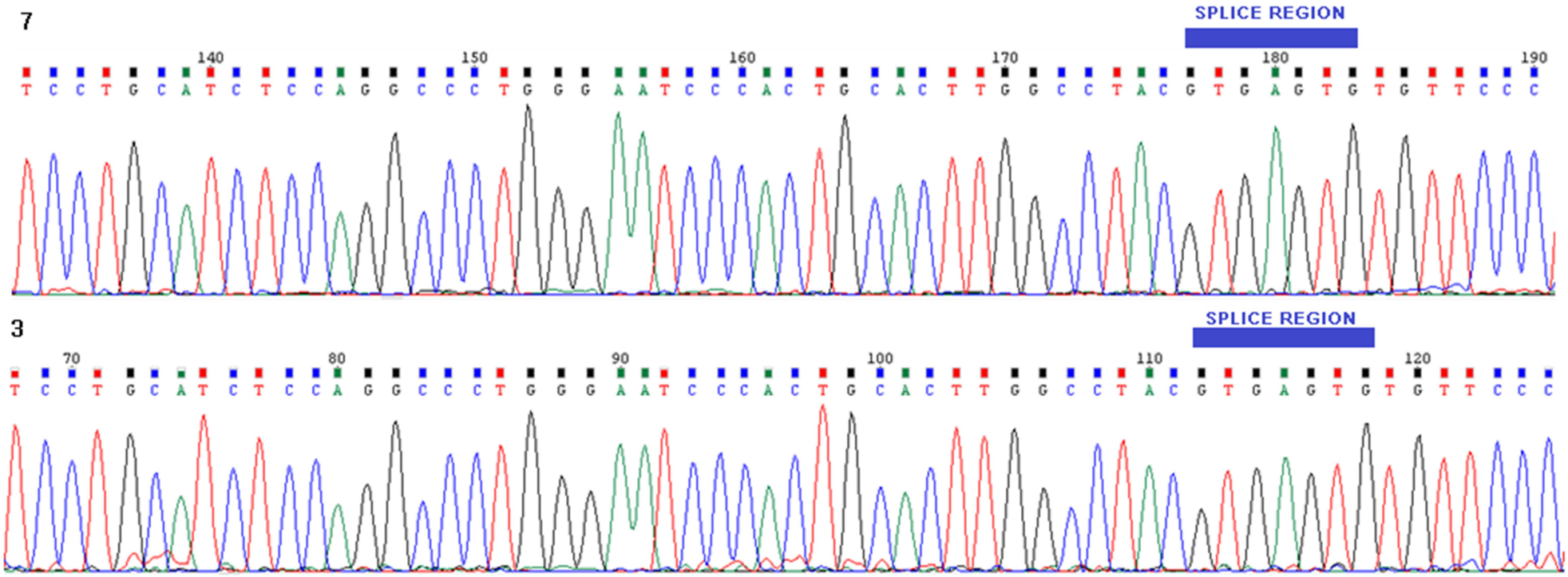


Figure 27: Electropherogram of the splice site boundary region of exon 3 from patient samples 7 and 3. No mutation is present. The splice site regions are indicated with a blue bar at the top. Patient 7 is not affected with HHT while patient 3 is affected.

Single Nucleotide Polymorphisms (SNPs) found in the study family

Results from this study detected SNPs only in exons one, five and eleven. Numerous studies have reported SNPs within exon regions two to four, six to ten and twelve to fourteen, to be associated with HHT (Table 3). For example Lesca *et al.,* (2004) identified numerous mutations in most of the exons in the *ENG* gene. One of the mutation identified was a missense variant c.1096G>C in exon 8 (Li *et al.,* 1999). The amino acid change of this mutation is from an asparagine to a histidine (p.Asp366His) and is described to be likely benign. Letterboer *et al.,* (2005) identified 13 missense mutations, which were found in numerous exons within the *ENG* gene. These mutations include; c.392C>T (exon 4), c.781T>C (exon 6) c.790G>A (exon 7) and c.1238G>T (exon 9A) all of which are benign or likely benign.

McDonald *et al.,* (2009) reported that the majority of mutations present in the *ENG* gene are missense mutations accounting for 27% of the overall mutations. However, the mechanisms of the mutational cause of HHT largely remains unknown (Waite and Eng, 2003) due to the heterogeneity and complexity of the disorder. Consequently, the use of scoring systems such as SIFT and Polyphen aid in classifying the disorder's pathogenicity which enables the investigator to deduce suggestions as to how the mutation cause HHT. It should be noted however that such suggestions warrant further investigations.

Results from this study identified four SNPs, two SNPs in exon one, one SNP in exon 5 and another one in exon 11. None of the SNPs are novel mutations and all have been previously described in the dbSNP database. The two SNPs in exon one have been allocated with accession numbers, rs7033891 and rs1002959572 and the SNP in exon 5 as rs150932144 and exon 11 as rs116330805 (Table 7). Details on each SNP identified from the study population are described below:

Exon 1 SNPs

rs7033891

The rs7033891 SNP in exon 1, position 9:127854679 results in a change of an adenine to a guanine (c.-324A>G) (Figure 28) and occurred in all the individuals in the study group, both in those with and without HHT. Sequence results indicate that the homozygous GG genotype is present in all the studied participants. This is a conserved mutation and therefore has little to no effect on the resulting protein as it is found in the 5' UTR region where it is transcribed but not translated.

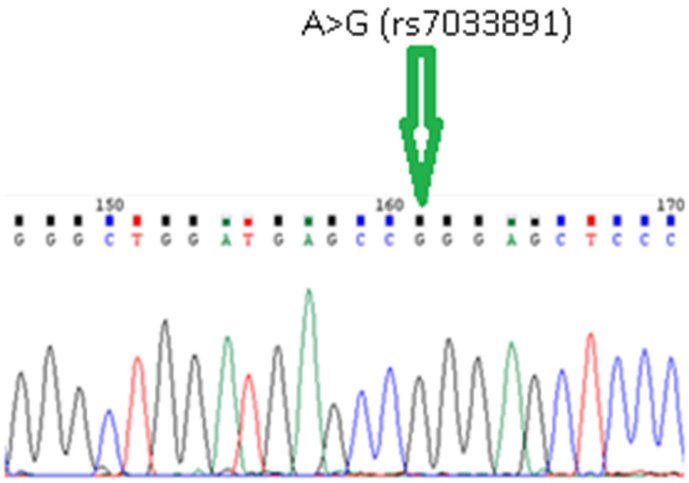


Figure 28: 5' UTR mutation exon 1; c.-324A>G (rs7033891), shown as an example from patient sample 7.

Figure 28 displays the forward strand of the sequence, indicating the change of an A>G. The reference sequence from the ensembl website (www.ensembl.org) indicate a change of a T>C. Based on population allele frequencies data from the 1000 Genomes Project Phase 3 that sampled 26 different populations of which contained 100 to 280 individuals per population group (www.internationalgenome.org), the overall frequency of the C allele is 99% and T allele is 1%. In the African population the C allele accounts for 98% of the tested

African population and only 2% have the T allele. The African populations tested were the Mende in Sierra Leone, Esan and Yoruba in Nigeria, Luhya in Kenya and Mandinka, Fula, Wolof and Jola in the Gambia (Zhang & Dolan, 2010). Other population groups included in the African subpopulations were the African Caribbean in Barbados and African ancestry in Southwest in the United States (Zhang & Dolan, 2010). The T allele is only found in the African ancestry sub-population groups (T=0.023) and not in the other sub-population groups including the Americans, East Asian, European and South Asian (Figure 29A) (www.ensemble.org). The T allele frequency reported in the ethnic groups are as follows: the African Caribbean (ACB) (T=0.021), Esan (ESN) (0.035), Luhya (LWK) (0.020), Mende (MSL) (0.018) and Yoruba (YRI) (0.056) (Figure 29B). Furthermore, based on the genotypic frequencies, the T allele is heterozygous and not homozygous in any of the individuals that possess this variant. Therefore, the reference sequence from ensembl containing the T allele (A allele) is not the ancestral C allele (G) present in most population groups around the world.

Since there was no differentiation between the HHT affected and unaffected individuals, it is concluded that this SNP is a population variant rather than a disease causing mutation.

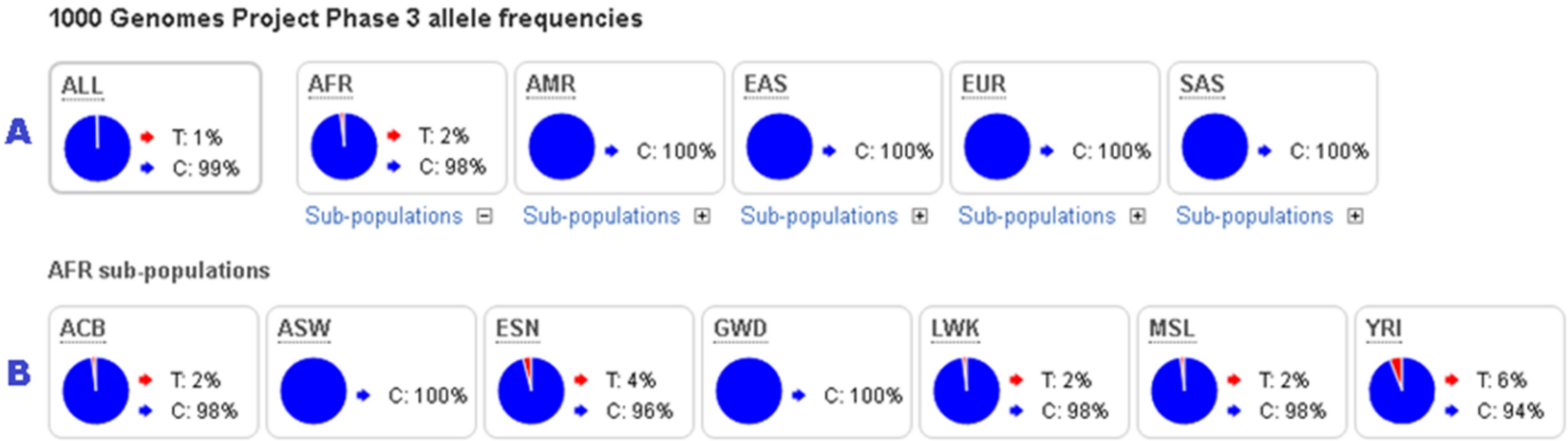


Figure 29: 1000 genomes phase 3 project allele frequencies data on rs7033891 (c.-324A>G): (A) The C allele (ancestral allele) is the only present allele in all population groups except for African population with the T and C alleles present in 1% and 99% in the respectively. (B) The T allele was found in these ethnic groups, the African Caribbean (ACB), Esan (ESN), Luhya (LWK), Mende (MSL) and Yoruba (YRI) in varying frequencies (www.ensembl.org).

rs1002959572

This exon 1 variant is present on position 9:127854562 in the 5' UTR region with a change of a G allele to an A allele (c.-207G>A) (Figure 30). The ancestral allele is the C (G) allele and the T (A) allele is found at 1% only in the African population according to the 1000 Genome Project Phase 3. Very little information has been published and it has not been listed in the HHT mutation database (www.arup.utah.edu). Based on population genetics onthe ensembl.org website, this mutation is not novel. The heterozygous G/A genotype was detected in patients 9, 11, 12 and 13 of the studied participants (Figure 31). All the individuals with the mutation have been diagnosed with HHT. Individuals 9 and 11 are father and son and the father passed the mutation to the son. However, the mother is not a participant of this study consequently her genotype could not be established. Patient 12 parents were deceased at the time of the study, therefore remains unknown from whom this mutation was inherited from. Patient 13's mother (patient 5) tested negative for the rs1002959572 (c.-207G>A) SNP allele, yet again no sample was taken from her father and she could have inherited it from him (Figure 35).

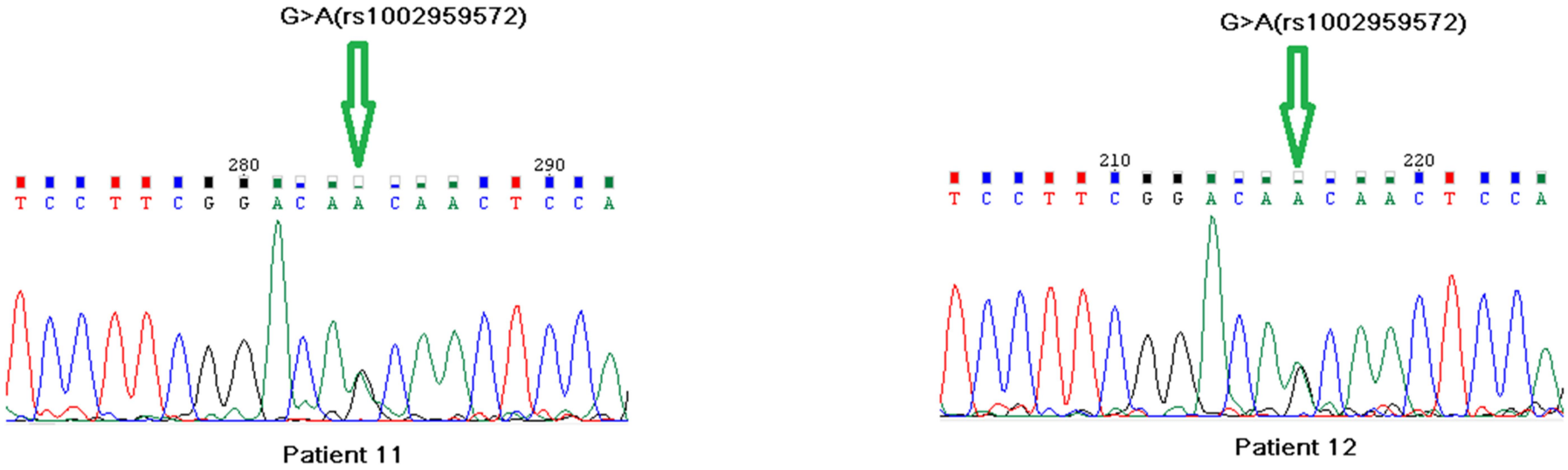


Figure 30: Electrographs illustrating the 5' UTR heterozygous mutation in exon 1; c. -207G>A (rs1002959572) of patients 11 and 12.

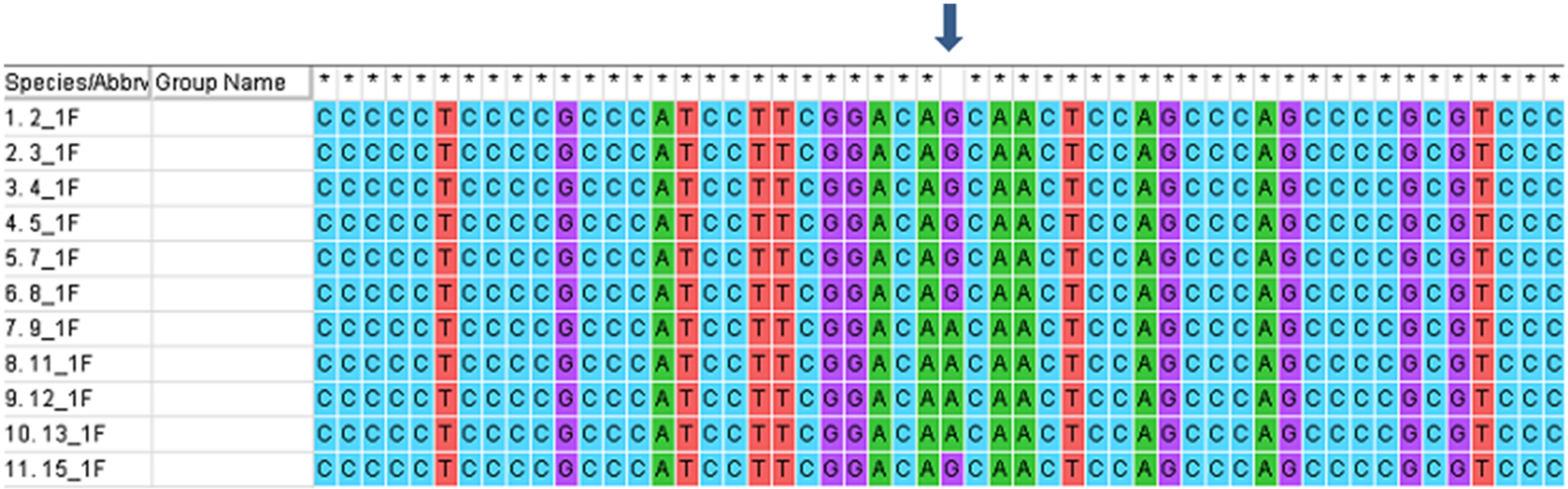


Figure 31: The heterozygous G/A genotype was detected in individuals 9, 11, 12 and 13 and the others have the homozygous G/G genotype. The SNP is indicated with the blue arrow. In the species/abbreviation column the first numeral indicates the line number eg.1; the numeral in front of the 'underscore' sign indicates the sample number; and the '_1F' indicates the exon 1 forward strand.

This mutation has been identified as a regulatory variant based on the data found ensembl.org. Moreover, c.-207G>A has only been observed in the African population with a low frequency of 1.15048e-04 and is not in any other population groups (www.ensembl.org) suggesting this mutation is probably be a population specific variant. Confirming the African ancestry of the family studied. The rarity of this mutation in the African population and that it was only observed in people who are diagnosed with HHT; may suggest that this mutation may be linked to HHT. On the contrary, since the SNP is assumed to be inherited from parents that have not been diagnosed with HHT (parents of individuals 12 and 13) it is unlikely to play a significant role in HHT. Because of this, further investigation on this SNP's role in HHT is warranted, on a more inclusive group of the studied family as well as on non-related HHT families of African descent.

A mutation that has been documented close to this variant is the c.-205A>C SNP (Damjanovich *et al.,* 2011). This mutation was found in one HHT patient in a heterozygous state. Neither the brother nor the father of the mutant SNP carrier had the mutation. It is assumed that he inherited it from his asymptomatic mother that tested positive for the mutant C allele. This variant was predicted not to have a significant effect on the regulation of transcription and translation based on *in silico* analysis (Damjanovich *et al.,* 2011). In expression analysis, the wild type and mutant construct of c.-205A>C showed similar protein levels and it did not affect initiation of ATG translation. Therefore it was concluded that this variant is benign.

Given that the identified variant c. -207G>A is only 2 bases from c.-205A>C and are both situated in the 5' UTR region. The latter variant did not affect initiation of ATG translation; it is possible that the former variant also does not affect ATG initiation since it is further upstream. Additionally, the A allele for SNP -207G>A is present in

only four family members who are diagnosed with HHT and not in the other three patients with HHT (Figure 33) indicating that this mutation is not associated with HHT and the mutation is likely benign.

Exon 5 SNP

rs150932144

This missense mutation (rs150932144) is situated in exon 5 (Figure 32) displaying a G>A (c.640G>A) change, consequently causing an amino acid change from a glycine to a serine (p.Gly214Ser). Mutation rs150932144 was identified in individual 12 who has been positively diagnosed with HHT and it is located in position 9:127825744. This mutation occurs in the zona pellucida-like domain in the ENG protein at the N-terminal (Pfarr *et al.,* 2013).

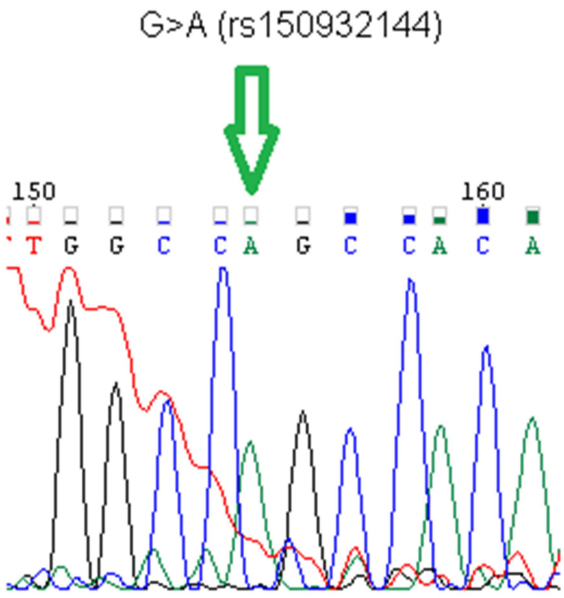


Figure 32: Electropherogram illustrating missense mutation in exon 5; c.640 G>A (rs150932144) of HHT patient 12.

This mutation has been identified and considered pathogenic by some studies such as Pfarr *et al.* (2013), and Mei-Zhav *et al.* (2006), while others were uncertain of its significance as stated on Clinvar (www.ncbi.nhlm.gov). However, the relevant

articles with these outcomes could not be found. Pfarr *et al.* (2013) identified this mutation in a young HHT patient who also presented with idiopathic pulmonary arterial hypertension (IPAH). Based on PolyPhen algorithm scores this mutation was considered damaging (http://genetics.bwh.harvard.edu/pph2/). In another study (Mei-Zhav *et al.,* 2006), this variant was found in a young girl who was diagnosed with HHT at the age of 12 years and also presented with pulmonary and cerebral AVMs. Moreover, her mother was also found with this mutation; even though she did not present with HHT symptoms. It was concluded that the mutation may be disease causing, yet further investigation on this variant is justified (Mei-Zhav *et al.,* 2006). This mutation displays an example of genotypic and phenotypic correlation given the presence of PAVM and IPAH associated with this mutation. Further investigations on this mutation on people who have HHT and IPAH and PAVMs may link this mutation to the described phenotypes, therefore, increasing our understanding in how a mutation causes a HHT symptom.

In the current study, the rs150932144 SNP was only found in one individual (12) that is diagnosed with HHT because of severe nose bleed episodes and family history. Unlike the cases mentioned above, she has not presented with any PAVM or IPAH. Sequence screening results of her son, who was also diagnosed with HHT, revealed that her son does not share the presence of the A allele. The homozygous G allele genotype was present in her son. This could be a case of *ENG* gene mosaicism where leucocyte cell lines reveals the mutation while the other cell lines from tissues (cheek swab) reveal the normal allele. This may be the reason for not finding this mutation in her son. Mosaicism is defined as the presence of various genotypes found in one individual resulting from two or more cell lines (Tørring *et al.,*2017). Her parents were deceased and no biological material was collected from them,

therefore, genotyping could not be performed (Figure 35). The mutation may have been inherited from the parents or be a case of mosaicism but given the lack of their biological samples and breach of this study's ethical agreement (not allowed to collect more biological samples), this possibility cannot be established.

The homozygous presence of the mutation in her and not her son is biologically unlikely thus following explanations may be true. Firstly, contamination could be a factor; however, precautions were taken to avoid contamination. The regular use 70% ethanol, RNase and DNase on the work bench and the autoclaving of tips and micro-centrifuges are implemented as routine laboratory practice. Additionally, a non-template control was added during PCR to confirm that no contamination was present. Furthermore, the source of contamination is also unknown, since it is the rare mutation allele that is not present in any of the other family members and is unlikely present in the investigator.

Secondly, the presence of a sequence artefact is also possible. Artefacts may result because of peaks being too close or when the signal intensity is low as seen in Figure 31. Another possibility is that sample tubes could have been incorrectly labelled or swapped during the sampling process at the clinic in Koffiefontein. This can be corrected by re-sampling but this option unfortunately fell out of the time frame of this study, since ethical approval was only granted for the use of archived samples and additional amendments and permissions from the relevant governing bodies, may prolong the study with several months. However despite the possible discrepancies of this specific mutation, it cannot be regarded as pathogenic since it is present in only one HHT patient.

Exon 11 SNP

rs116330805

The SNP rs116330805 is present in exon 11 at position 9:127818296 as a missense mutation changing from a guanine allele to an adenine allele (c.1510G>A). Sequencing resulted in a heterozygous mutation present in the DNA of study participant 2, who has been diagnosed with HHT (Figure 33). The resulting amino acid is a valine to a methionine (p.Val504Met). This mutation lies on the ZP-C subdomain and on the surface area of the gene (Figure 34) (Ali *et al*., 2011; Mallet *et al*., 2014). This is not a novel mutation and it has been previously found in HHT patients in Funen (Brusgaard *et al.,* 2004), France (Lesca *et al.,* 2004) and United States (McDonald *et al.,* 2009).

There are conflicting interpretations about the pathogenicity of this mutation in HHT. Lesca *et al.,* (2004) and Brusgaard *et al.,* (2004) regarded this mutation as pathogenic, confirming the diagnosis of HHT. However, based on SIFT and PolyPhen scores of 0.13 and 0.286 respectively, the mutation is tolerated or benign (www.ensembl.org). Mallet *et al.* (2014) studied numerous HHT variants, including c.1510G>A (V504M), and found that this mutation is a silent mutation because the cell surface proteins did not differ from what is expressed by the wild type allele. Concluding that that this mutation is not pathogenic, but may be suspected to be associated with a benign HHT phenotype and warrants further functional analysis (Mallet *et al.,* 2014). Similarly, McDonald *et al.,* (2009) regarded this mutation to be benign since in one of his case studies this variant was detected in an unaffected brother, confirming the conclusion of Mallet *et al.* (2014). The clinical significance of this variant is reported as “likely benign” on NCBI, since the variant has been

observed in both cases and controls. Taking above information into account, along with the fact that it is not present in any of the other HHT family members SNP rs116330805 is therefore concluded to be benign.

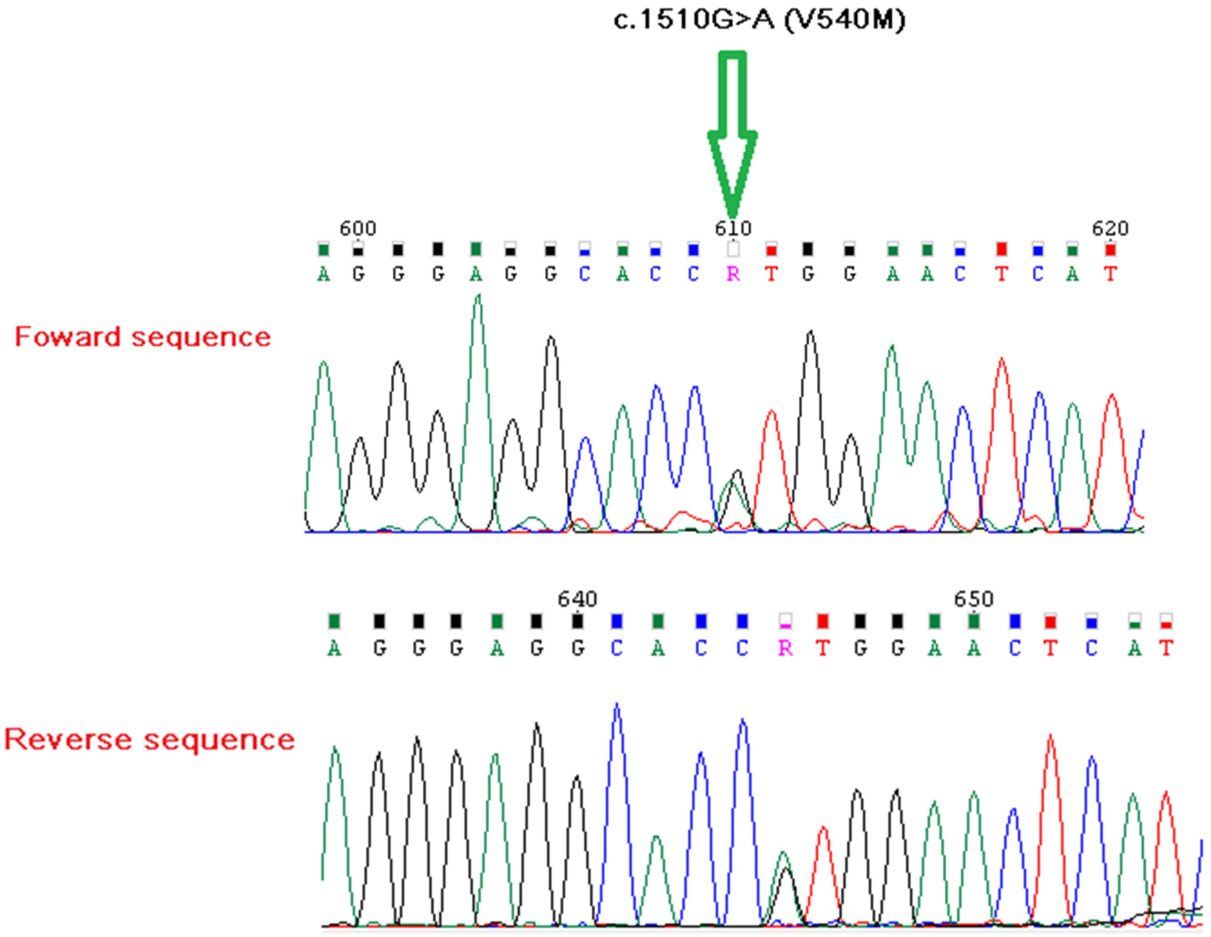


Figure 33: Electropherogram data presenting forward and reverse sequences for the missense mutation of c.1510G>A (rs116330805) (V504M) in HHT patient 2.

This mutation is present in a heterozygous state; the mutant allele (allele A) may have been inherited from either the mother or father. Since neither parents participated in the study, there are no samples for either of the parents, the parent donating the A allele cannot be confirmed (Figure 35).

It is worth mentioning that in terms of population genetics (1000 Genome Project Phase 3) that the C allele is the ancestral allele and is present in 99% of individuals while the T allele is only present in 1% in all population groups. However, the minor allele frequency of the T allele is the highest allele frequency (0.018) in African populations compared to other population groups in Europe, Americas and Asia whose frequencies ranged between 0.001 and 0.002.

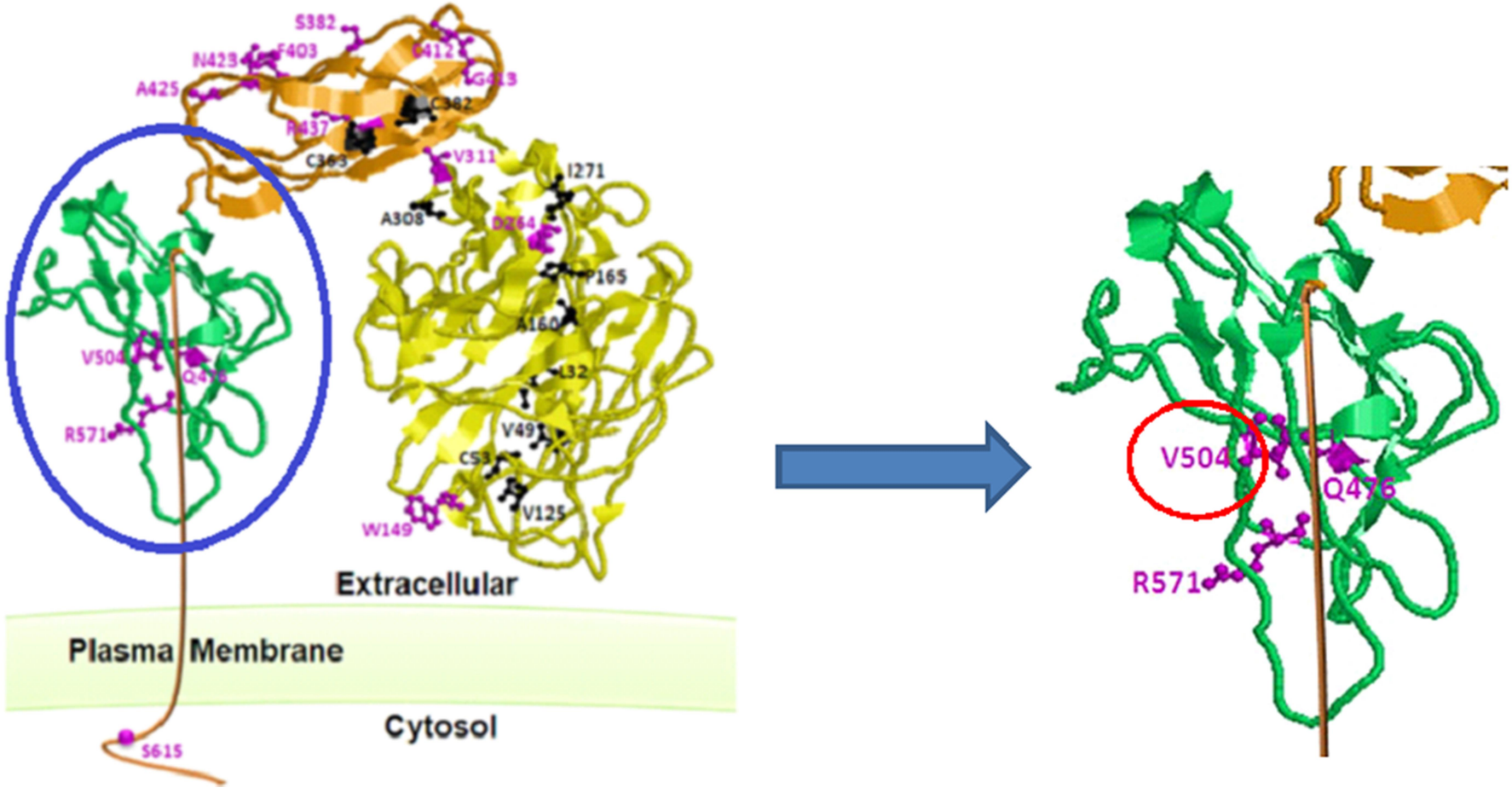


Figure 34: A three-dimensional endoglin monomer indicating 25 missense mutations. Magnification of the blue circle shows the zona pellucida (ZP-C) subdomain (green) where the V504M variant (red circle) is located. The other two domains are ZP-N (orange) and orphan (yellow) (Ali *et al.,* 2011).

TABLE 7: SUMMARY OF MUTATIONS RESULTING FROM THIS STUDY

MUTATION	PROTEIN CHANGE	LOCATION	SNP ID
c.-324A>G	None	Exon 1: 5' UTR 9:127854679	rs7033891
c.-207G>A	None	Exon 1: 5' UTR 9:127854562	rs1002959572
c.640G>A	p.Gly214Ser	Exon 5 9:127825744	rs150932144
c.1510G>A	p.Val504Met	Exon 11 9:127818296	rs116330805

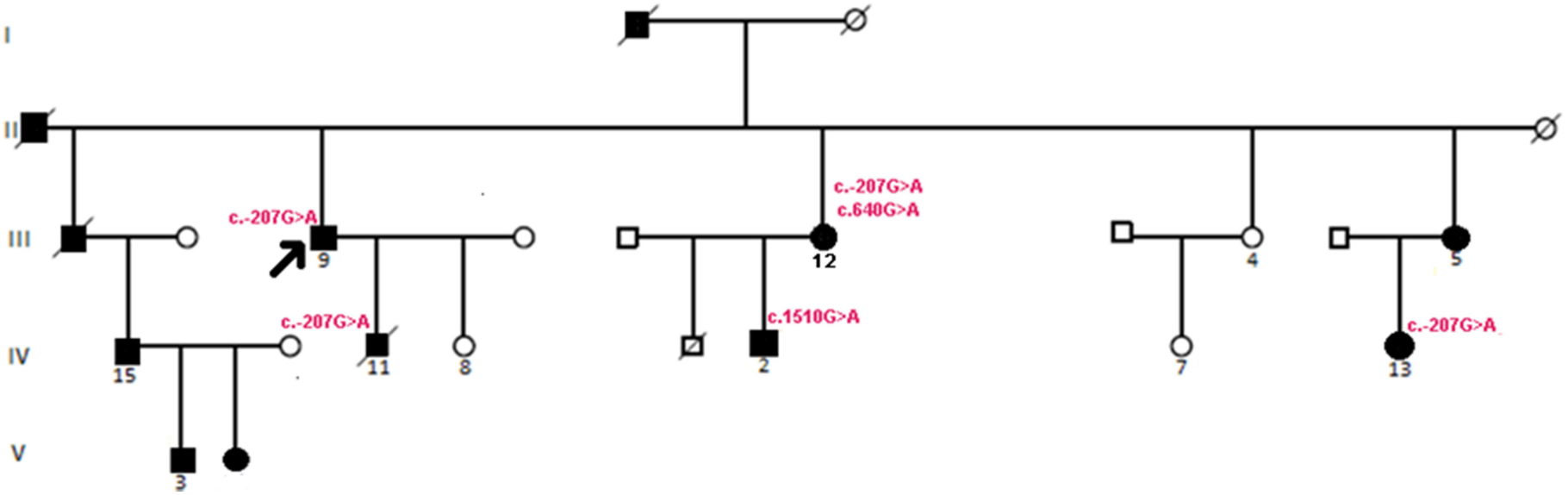


Figure 35: A five generation family pedigree chart displaying the autosomal dominant pattern of inheritance. The proband is participant 9. Polymorphisms resulting from this study is depicted in red with the exception of population variant (c.324A>G; rs7033891) present in all the participants. SNPs are shown for individuals 2 (c.1510G>A; rs116330805), 12 (c.640G>A; rs150932144) and 9, 11, 12 and 13 (c.-207G>A; rs1002959572). The black squares and circles represent affected males and females respectively. The white squares and circles represent the unaffected males and females respectively.

Concluding remarks on results

None of the mutations that have been identified can be concluded to have a causal effect toward HHT. The exon one mutation c.-324A>G (page 102) is an African population variant and since it is located at the UTR region, this mutation is not translated into an amino acid. The c.-207G>A (page 104) was identified in four of the seven individuals who have HHT in this cohort suggesting that this mutation may be linked to HHT. Moreover, given that it is a rare African mutation, this mutation may be an African population variant. However, since the other HHT affected individuals did not have the mutation, this variant is considered benign and further investigation as to its role in HHT is required. The c.640G>A (page 106) has been linked to IPAH and PAVMs in young patients. This woman suffered neither of those symptoms. Given that this mutation was not present in her son, this may be a case of mosaicism or she may have inherited the mutation from her parents. Nevertheless, the mutation was only identified in a single individual and therefore it is likely benign. The c.1510G>A (page 114) missense mutation has been previously suggested to be pathogenic but recent studies have proven contrary. However, this mutation is suggested to be associated with a benign HHT phenotype. Since the mutation was found in only a single individual, it was concluded that it is benign in the study cohort.

The effect of *ENG* mutations on the TGF-β pathway

Even though none of the mutations identified from the study results appear to be pathogenic in this family, some of these mutations rs150932144 (c.640G>A) and rs116330805 (c.1510G>A) have been positively associated to HHT in previous literature (Table 3), however, in this study these mutations were found to be benign even though it has been identified in people affected with HHT. All the SNPs detected in this study (c.-207G>A, c.640G>A and c.1510G>A) were concluded to be

uncertain in terms of pathogenicity; however that doesn't mean that they are not pathogenic. The effect of missense mutations in the TGF-β pathway has not been proven. Mutations in the *ENG* gene may affect numerous pathways downstream, even though the precise mechanism remains unknown (Waite & Eng, 2003).

Two models have been suggested to explain the consequences of mutant *ENG* proteins. The first model predicts that the truncated proteins interfere with TGFβ signalling by being secreted, and thus sequestering extracellular TGFβ, or by binding normal *ENG.* The second model is the dominant haploinsufficiency mechanism that has been supported by independent studies (Bayrak-Toydemir *et al.*, 2004).

To extrapolate the first model, an expression study of a splice site mutation that led to a deletion in exon 3 decreased the expression of *ENG* on the cell surface of both monocytes and endothelial cells by 50%. This suggests that expressed *ENG* mutant types do not form heterodimers with the wild-type, and neither does homodimerization occur resulting in an intercellular degradation. An expression study performed by Lux *et al.* (2000) discovered six missense mutations that were not detected on the cell surface, they hypothesise that the proteins were misfolded. Yet, when these missense mutations were co-expressed with normal *ENG* (Bayrak-Toydemir *et al*., 2004), the normal and the mutant proteins would dimerize and consequently moved to the cell surface. These are examples of dominant negative mutations (Guttmacher *et al*., 1995) and again points out the complexity of HHT disease aetiology.

Of these two models, haploinsufficiency is considered to play the main role in the aetiology of HHT (Bayrak-Toydemir *et al*., 2004; Chen *et al.,* 2013). Haploinsufficiency of surface *ENG* protein may impede TGF-β signalling,

consequently interfering with normal vessel wall repair resulting in a cascade of effects that eventually changes the vascular architecture (Gallione *et al*., 1998). Haploinsufficiency of *ENG* is a result of mutations occurring on the external domain of the gene. Similar vascular defects that occur due to *ENG* haploinsufficiency in humans were also observed in mice with HHT1 (Park *et al.,* 2015).

5'UTR mutations and HHT

Two 5'UTR mutations in exon 1 were identified in this study, namely rs7033891 (c.-324A>G) and rs100295972 (c.-207G>A). The former mutation is a population variant, while the latter mutation may be a regulatory variant. Nucleotides bases on the untranslated region are transcribed but not translated, and therefore do not result as proteins. However, that does not exclude the SNPs from having a pathogenic effect.

Mutations identified in the 5' untranslated region of the *ENG* gene have been proven to be pathogenic in individuals affected with HHT. A study by Damjanovich *et al*., (2011) identified two pathogenic mutations found in the 5' UTR of the *ENG* gene. One such a mutation c.-127C>T, was described in a family in Northern Spain where only a single individual was affected in a family of three. Two individual probands from two other distinct families and some family members that were affected by HHT were positive for this mutation. The c.-127C>T mutation caused a new translation start site resulting in a frame shift product to form a premature truncated protein that differed from the wild type protein (Damjanovich *et al.,* 2011). Results of this mutation were confirmed by another study with a positive HHT family history (Kim *et al.,* 2011).

The other 5' untranslated region mutation c.-9G>A was identified in three HHT families (Damjanovich *et al.,* 2011). This mutation also created a new translation start site but the reading frame was not altered. However, it was suggested that this mutation may be mildly pathogenic as it may result in reduced protein function and production. Both of these mutations may also influence transcriptional regulation of the *ENG* gene (Damjanovich *et al.,* 2011).

Additional genes that are involved in HHT Pathogenesis

Hereditary haemorrhagic telangiectasia is a heterogeneous disease, with at least two additional genes that has been described to cause HHT, besides *ENG*. These include *Activin receptor-like kinase 1* (*ALK1*), SMAD family member 4 (*SMAD4*) genes as well as *bone morphogenetic 9* (*BMP9*) and *RAS p21 protein activator 1* (*RASA1*) (McDonald *et al.,* 2015).

Since none of the participants in this study had a pathogenic mutation caused by the *ENG* gene, the other panel genes such as *ALK1*, *SMAD4, BMP9* and *RASA1* should be screened in order to identify the causative mutation. Figure 36 is a schematic illustration of a proposed molecular testing algorithm based on suspected HHT clinical diagnoses proposed by McDonald *et al.* (2015). The testing includes screening for the presence of classic HHT phenotype, followed by *ENG* and *ALK1* (*ACVRL1*), then the screening of *SMAD4.* If the first three genes test negative, other genes can be tested. Testing of HHT-like phenotype includes the screening of HHT multi-gene panel using next generation sequencing (NGS) and comparative genomic hybridization array (aCGH). If no mutation is identified, research should be done on other additional genes so the mutation can be identified (McDonald *et al.,* 2015). With more research performed on HHT patients from African descent this algorithm

could be adjusted to be more efficient, cost effective and applicable in terms of HHT diagnosis in Africa.

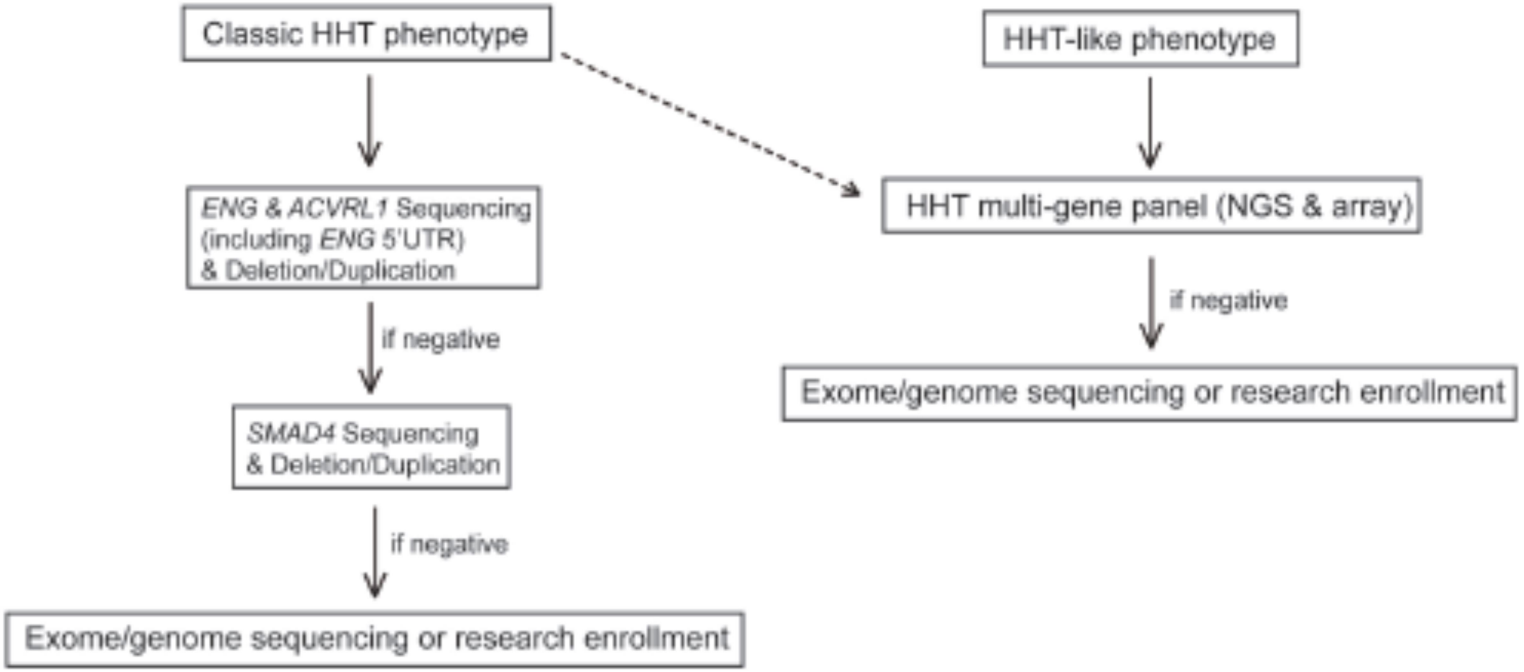


Figure 36: The proposed molecular testing algorithm based on suspected HHT clinical diagnosis. On the left is the classic HHT phenotype, the screening of *ENG*, *ALK1* and *SMAD4*. If negative, research is done on other genes. On the right is HHT-like phenotype, HHT multi-gene panel can be screened using NGS and comparative genomic hybridization array (aCGH). If negative, further research is needed to identify the mutated gene (McDonald *et al.,* 2015).

Epigenetic factors that influence HHT

According to literature, it appears that HHT phenotype is the result not only of genes, but environmental factors as well. This gives rise to the “second hit “hypothesis (Figure 37) (Ruiz-Llorente *et al.,* 2017). Since the presence of mutation is not enough to cause AVMs and telangiectases, there are epigenetic factors that cause the development of these symptoms. Some of these are stimuli resulting from epigenetic factors which include wounds, infection, trauma and hypoxia (Figure 14), all of which have already been mentioned in chapter 2 which may initiate HHT symptoms.

The other factor that has been mentioned is “Fluid shear stress”. Gene expression in endothelial cells exposed to shear stress change due to chromatin modification changes. It was speculated that altered hemodynamic (blood pressure and shear stress) forces at the junction of the vein and artery may lead to the development of AVM (Thomas *et al.,* 2016).

DNA methylation is an epigenetic factor that is associated with HHT. The level of methylation varies in ECs of both vein and arteries. It has been observed that differential methylation of promoter eNOS gene influences vascular function and plays a crucial role in ECs function (Thomas *et al.,* 2016). The roles of epigenetic factors in HHT aid to explain the formation of AVMs as three of the participants in this study have AVMs (Table 4). However, it should be noted that the pathogenesis of AVM development largely remains unknown (Thomas *et al.,* 2016) and therefore it is crucial that epigenetic factors revolving around HHT should be investigated in order to understand its role in this disorder.

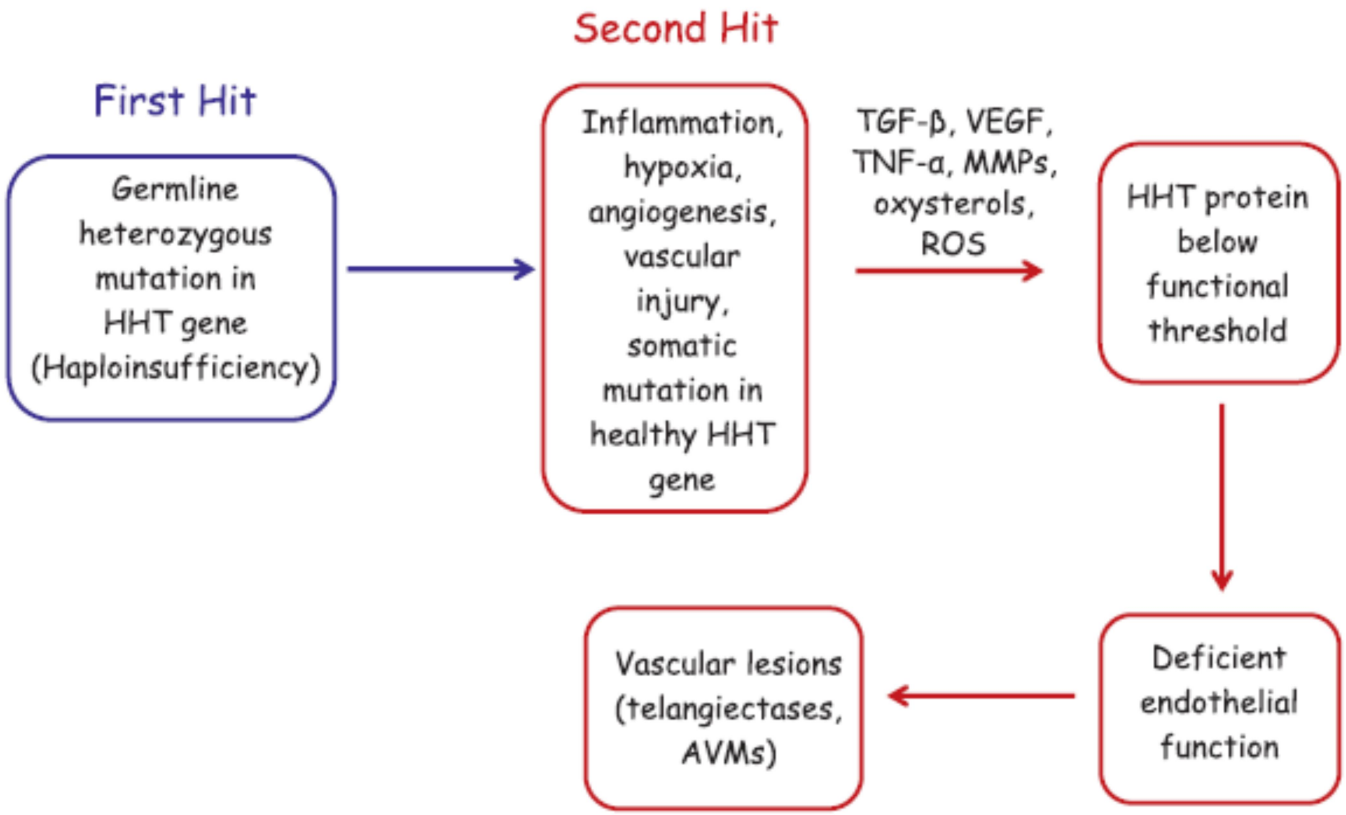


Figure 37: Hypothetical second hit model in HHT. The germline heterozygous mutation in the HHT gene leads to haploinsufficiency (First hit). A subsequent event (Second hit) involves inflammation, hypoxia, neoangiogénesis, vascular injury, or a somatic mutation in the healthy HHT allele. Some of these events induce the expression/activation of mediators, which generate a microenvironment where HHT protein levels are below the needed functional threshold. This drop in the HHT functional protein can also be directly obtained after a second somatic mutation of the healthy HHT allele. The resulting deficiency in endothelial cells function leads to the generation of vascular lesions (Copied from Ruiz-Llorente *et al.,* 2017).

Limitation of Study

The ethical clearance given for this study did not permit the re-collection of blood samples from participants especially from those of whom the genotyping results could not be resolved. The small sample family size is also a limitation. Given more time, and funding a larger group of the family could have been recruited and samples collected. This would have allowed a more broaden view of other undiscovered mutations on the *ENG* gene from this family. Secondly, due to lack of sufficient funds, the screening of other causal genes could not be done. Consequently, we were limited to testing for mutations in one gene.

CHAPTER 5

CONCLUSION

The diagnosis and treatment of HHT requires the participation of multi-disciplinary clinicians. The current Curaçao criteria as applied currently require clinical examinations and family history to determine diagnoses. However, the diagnosis is based on the assumption that the phenotype is already present, thus patients with a mild phenotype or young persons with underlying disease may be misdiagnosed. Therefore genetic testing is vital for a confirmed and early diagnosis.

This study is the first to report on *ENG* gene splice site mutations in a family of which some members have been diagnosed with HHT with African ancestry. Even though, no such mutations were detected, this is a significant result contributing to the paucity of reports on HHT in Africa. According to literature splice site mutations account for 12% of mutations found on the *ENG* gene in patients with HHT, studied in Caucasian and Asian populations. This could not be proven in the studied family from African descent. The small number of participants was a major limitation of this study which may be the reason of discrepancies between this study and other studies.

In this study, four exon region mutations SNPs were identified; c.-324A>G, c.-207G>A, c.640G>A and c.1510G>A (see Figure 36). The last three (c.207G>A, c.640G>A and c.1510G>A) were found to be present in different individuals all diagnosed with HHT. The other mutation (c.-324A>G) is a population variant found to be present in all participants, irrespective of HHT diagnosis. The c.-207G>A mutation was detected in four of seven individuals affected with HHT in this cohort. This suggests that the mutation may be linked to HHT but further investigation is warranted. Additionally, this mutation is also unique to the African population but it is rare. Since none of these mutations were found in other individuals in the family who

are diagnosed with HHT, none of the mutations appear to be pathogenic in the studied family.

One mutation c.-324A>G is a mutation unique to African sub-populations. This is an indication that the people of African ancestry with HHT may have unique *ENG* gene mutations that have not been identified in European, Asian and American population groups. This highlights the need for HHT genetic research on people of African ancestry. Such studies may identify not only mutations unique to the African population but also discover new genes that cause HHT.

The missense mutation c.640G>A has been linked to pulmonary arterial hypertension and pulmonary arteriovenous malformations. This is an example of genotypic and phenotypic correlation. More studies on this mutation on people with the HHT phenotype such as PAVM may increase our understanding on the mechanism from the mutation to the symptom.

Numerous mutations have been identified all through the *ENG* gene, which has a high missense mutation rate. However, few functional studies exist on missense mutations found in the *ENG* gene. The majority of pathogenic missense mutations were mainly based on the scoring algorithms including SIFT and PolyPhen scores. Additionally, these algorithms tend to vary when it comes to classification of a mutation. Therefore, the use of scoring systems however helpful is not an accurate reflection of a mutation's pathogenicity.

Hereditary haemorrhagic telangiectasia is a complex heterogeneous disorder despite its Mendelian autosomal dominant pattern of inheritance. The complexity of this disorder is due to locus heterogeneity, age dependent expression, several phenotypic mechanisms and epigenetic factors. All these aspects affect the

interpretation of a mutation, consequently affecting how a mutation is classified. Classification of mutations whether it is “pathogenic” or “benign” have proved challenging to establish, not only in this study, but also in literature. One of the reasons being, that for mutations to be considered pathogenic, HHT patients should have the mutations exclusively when compared to HHT negative controls. More research should be performed to accurately classify mutations. This may then limit conflicting views on the status of a mutation. Therefore, functional analysis studies performed by using knockout mice have a contributing role to play in understanding the disease pathology and to enable accurate classification of mutations.

Since HHT is a heterogeneous disorder and no familial pathogenic mutation was found in the splice sites and exonic regions of the *ENG gene*, future studies should include the following: firstly to recruit and genotype as many additional family members as possible. With this information a more comprehensive pedigree analysis can be formed on the inheritance and disease causing effect of the identified polymorphisms. Secondly, other associated genes such as *ALK1, SMAD 4, BMP9* and *RASA1* should be comprehensively investigated. This study is the first report of genetic analysis of family with HHT in South Africa and showed by process of elimination that *ENG* gene splice site regions are not the obvious cause of HHT in the studied family.

REFERENCES

Abdalla, A. & Letarte, M. (2006). Hereditary haemorrhagic telangiectais: current views on genetics and mechanisms of the disease. *J Med Genet,* 43:97-110.

Al-Deen, S. & Bachmann-Harildstad, G. (2008). A grading scale for epistaxis in hereditary haemorrhagic teleangectasia. *Rhinology*, 46:281-284.

Ali, B. R., Ben-rebeh, I., John, A., Akawi, N. A., Milhem, R. M., Al-shehhi, N. A., Al-gazali, L. *et al.* (2011). Endoplasmic Reticulum Quality Control Is Involved in the Mechanism of Endoglin-Mediated Hereditary Haemorrhagic Telangiectasia. *PloS ONE*, 6:1-10.

Ammann, K.R., DeCook, K.J., Tran, P.L., Merkle, V.M &Slepian, M.J. (2015). Collective cell migration of smooth muscle and endothelial cell: Impact of injury versus non-injury stimuli. *J Biol Eng,* 9:1-15.

Bayrak-Toydemir, P., Mao, R., Lewin, S. & McDonald, J. (2004). Hereditary hemorrhagic telangiectasia: an overview of diagnosis and management in the molecular era for clinicians. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, 6:175-191.

Bennett, R. L., French, K. S., Resta, R. G. & Doyle, D. L. (2008). Standardized human pedigree nomenclature: update and assessment of the recommendations of the National Society of Genetic Counselors. *Journal of genetic counselling,* 17:424-433.

Bernabéu, C., Conley, B. A. & Vary, C. P. (2007). Novel biochemical pathways of endoglin in vascular cell physiology. *J Cell Biochem,* 102:1375-1388.

Berg, J. N., Guttmacher, A. E., Marchuk, D. A. & Porteous, M. E. M. (1996). Clinical heterogeneity in hereditary haemorrhagic telangiectasia : are pulmonary arteriovenous malformations more common in familied linked to endoglin? *J Med genet,* 33:256-257.

Berg J. N., Gallione C. J., Stenzel, T. T., Johnson, D. W., Allen, W. P., Schwartz, C. E., *et al*. (1997). The activin receptor-like kinase 1 gene: Genomic structure and mutations in hereditary hemorrhagic telangiectasia type 2. *Am J Hum Genet*, 61:60-67.

Blanco, F. J., Grande, M. T., Langa, C., Oujo, B., Velasco, S., Rodriguez-Barbero, A., *et al.*(2008). S-endoglin expression is induced in senescent endothelial cells and contributes to vascular pathology. *Circulation Research*, 103:1383-1392.

Bossler, A.D., Richards, J., George, C., Godmillow, L. & Ganguly, A. (2006) . Novel mutations in ENG and ACVRL1 Identified in a series of 200 individuals undergoing clinical genetic testing for hereditary heamorrhagic telangiectasia (HHT): Correlation of genotype with phenotype. *Human Mutation,* 27:667-675.

Bourdeau, A., Cymerman, U., Paquet, M. E., Meschino, W., McKinnon, W. C., Guttmacher, E., *et al.*(2000). Endoglin expression is reduced in normal vessels but still detectable in arteriovenous malformations of patients with hereditary hemorrhagic telangiectasia type 1. *The American Journal of Pathology*, 156:911-923.

Braverman, I. M., Keh, A., & Jacobson, B. S. (1990). Ultrastructure and three-dimensional organization of the telangiectases of hereditary hemorrhagic telangiectasia. *The Journal of Investigative Dermatology*, 95*:*422-427.

Brusgaard, K., Kjeldsen, A. D, Poulsen, L., Moss, H., Vase, P., Rasmussen, K., Kruse, T. A. & Horder, M. (2004). Mutations in endoglin and in activin receptor-like kinase 1 among Danish patients with hereditary haemorrhagic telangiectasia. *Clin Genet*, 66:556-61.

Canzonieri, C., Centenara, L., Ornati, F., Pagella, F., Matti, E., Alvisi, C., *et al.*(2014a). Endoscopic evaluation of gastrointestinal tract in patients with hereditary hemorrhagic telangiectasia and correlation with their genotypes. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, 16:3-10.

Canzonieri, C., Ornati, F., Matti, E., Chu, F., Manfredi, G., Olivieri, C.,*et al.*. (2014b). Hereditary haemorrhagic telangiectasia in North African and sub-Saharan patients. *South African Medical Journal*, 104:256-257.

Carey J. L., Nader, N. & Chai, P.R. (2017). Drugs and medical devices: adverse events and the impact on women's health*. Clin Ther*, 39:10-22.

Carvalho, R. L. C., Jonker, L., Goumans, M.-J., Larsson, J., Bouwman, P., Karlsson, S. *et al.* (2004). Defective paracrine signalling by TGF beta in yolk sac vasculature of endoglin mutant mice: a paradigm for hereditary haemorrhagic telangiectasia. *Development (Cambridge, England)*, 131:6237-6247.

Castonguay, R., Werner, E. D., Matthews, R. G., Presman, E., Mulivor, A. W., Solban, N. *et al.* (2011). Soluble Endoglin Specifically Binds Bone Morphogenetic Proteins 9 and 10 via Its Orphan Domain , Inhibits Blood Vessel Formation , and Suppresses Tumor Growth. *The Journal of Biological Chemistry,* 286:30034-30046.

Chen, Y. J., Yang, Q. H., Liu, D., Liu, Q. Q., Eyries, M., Wen, L. *et al.* (2013). Clinical and genetic characteristics of Chinese patients with hereditary haemorrhagic telangiectasia-associated pulmonary hypertension. *European Journal of Clinical Investigation*, 43:1016-1024.

Chiu J. & Chien, S. (2011). Effects of disturbed flow on vascular endothelium: Pathophysiological basis and clinical perpectives. *Physiol Rev,* 91:1-106.

Choi, E. J., Chen, W., Jun, K., Arthur, H. M., Young, W. L. & Su, H. (2014). Novel brain arteriovenous malformation mouse models for type 1 hereditary hemorrhagic telangiectasia. *PloS ONE*, 9:1-10.

Cymerman, U., Vera, S., Karabegovic, A., Abdalla, S. & Letarte, M. (2003). Characterization of 17 novel endoglin mutations associated with hereditary hemorrhagic telangiectasia. *Human Mutation*, 21:482-492.

Dallas, N. A., Samuel, S., Xia, L., Fan, F., Gray, M. J., Lim, S. J. & Ellis, L. M. (2008). Endoglin (CD105): A Marker of Tumor Vasculature and Potential Target for Therapy, *Clinical Cancer Research,*14:1931-1937.

Dejana, E. (2004). Endothelial cell-cell junctions: Happy together. *Molecular Cell Biology,* 5:261-270.

Dakeishi, M., Shioya, T., Wada, Y. Shindo, T., Otaka, K. *et al*. (2002). Genetic epidemiology of hereditary haemorrhagic telangiectasia in a local community in the northern part of Japan. *Human Mutation,* 19:140-148.

Damjanovich, K., Langa, C., Blanco, F. J., Mcdonald, J., Botella, L. M., Bernabeu, C. *et al.* (2011). 5'UTR mutations of ENG cause hereditary hemorrhagic telangiectasia. *Orphanet Journal of Rare Diseases*,32:1-11.

Dheyauldeen, S., Geirdal, A., Osnes, T., Vartdal, L. S, & Dollner, R. (2012). Bevacizumab in hereditary hemorrhagic telangiectasia-associated epistaxis: effectiveness of an injection protocol based on the vascular anatomy of the nose. Laryngoscope 122:1210–1214.

Dimopoulos, M. A., Hamilos, G., Zomas, A. Gika, D., Efstathiou, E. *et al*. (2004). Pulsed cyclophosphamide, thalidomide and dexamethasone: an oral regiment for previously treated patients with multiple myeloma. *Hematol J,* 5:112-117.

Fernandez-L, A., Sanz-Rodriguez, F., Zarrabeitia, R., Perez-Molino, A., Morales, C., Restrepo, C. M., *et al*. (2006). Mutation study of Spanish patients with hereditary hemorrhagic telangiectasia and expression analysis of endoglin and ALK1. *Human Mutation*, 27:295-301.

Fulbright, R. K., Chaloupka, J. C., Putman, C. M., Sze, G. K., Merriam, M. M., Lee, G. K. *et al.* (1998). MR of hereditary hemorrhagic telangiectasia: Prevalence and spectrum of cerebrovascular malformations. *American Journal of Neuroradiology*, 19:477-484.

Gaillard, S., Dupuis-Girod, S., Boutitie, F., Riviere, S., Moriniere, S. *et al*. (2014). Tranexamic acid for epistaxis in hereditary hemorrhagic telangiectasia patients: a

European cross-over controlled trial in a rare disease. *J Thromb Haemost*, 12:1494-1502.

Gallione, C. J., Klaus, D. J., Yeh, E. Y., Stenzel, T. T., Xue, Y., Anthony, K. B. *et al.* (1998). Mutation and expression analysis of the endoglin gene in Hereditary Hemorrhagic Telangiectasia reveals null alleles. *Human Mutation*, 11:286-294.

Gallione, C. J., Scheessele, E. A., Reinhardt, D., Duits, A. J., Berg, J. N., Westermann, C. J. J., & Marchuk, D. A. (2000). Two common endoglin mutations in families with hereditary hemorrhagic telangiectasia in the Netherlands Antilles: Evidence for a founder effect. *Human Genetics*, 107:40-44.

Goumans, M. J., Liu, Z. & Ten Dijke, P. (2009). TGF-beta signalling in vascular biology and dysfunction, *Cell Res,* 19:116 -127.

Grand'Maison, A. (2009). Hereditary hemorrhagic telangiectasia. *CMAJ,* 180:833-839.

Guttmacher, A. E, Marchuk, D. A, & White, R. I. (1995). Hereditary Hemorrhagic Telangiectasia. *The New England Journal of Medicine,* 333:647-650.

Hoag, J. B., Terry, P., Mitchell, S., Reh, D, & Merlo, C. A. (2010). An epistaxis severity score for hereditary haemorrhagic telangiectasia. The *Laryngoscope,* 120:838-843.

Huang, X & Miller W. (1991). A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math*, 12:337-357.

Khachatryan, T., Karnwal, S., Hamirani, Y. S., & Budoff, M. J. (2010). Coronary arteriovenous malformation, as imaged with cardiac computed tomography angiography. *J Radiol Case Rep,* 4:1-8.

Kim, M., Kim, S., Lee, H., Lee, K., Seo, J., Lee, J., & Lee, Y. (2011). Clinical and genetic analyses of three Korean families with hereditary haemorrhagic telangiectasia. *BMC Medical Genetics,* 12:1-10.

Kitonyi, G.W., Wambungu, P.M., Oburra, H.O., & Ireri, J.M. (2008). Hereditary haemorrhagic telangiectasia in black adult male: case report. *East Afr Med J*, 85:412-416.

Kjeldsen, A. D., Vase, P., & Green, A. (1999). Hereditary haemorrhagic telangiectasia : a population-based study of prevalence and mortality in Danish patients. *J Intern Med,* 245:31-39.

Lamalice, L., Boeuf, F. Le, & Huot, J. (2007). Endothelial cell migration during angiogenesis. *Circ Res,* 10:782-794.

Lebrin, F., Srun, S., Raymond, K., Martin, S., van den Brink, S., Freitas. *et al.* (2010). Thalidomide stimulates vessel maturation and reduces epistaxis in individuals with hereditary haemorrhagic telangiectasia. *Nature Medicine*,16:420-428.

Lee, H. E., Sagong, C., Yeo, K. Y., Ko, J. Y., Kim, J. S., & Yu, H. J. (2009). A Case of Hereditary Hemorrhagic Telangiectasia. *Ann Dermatol,* 21:206-208.

Lesca, G., Plauchu, H., Coulet, F., Lefebvre, S., Plessis, G., Odent, S. *et al.* (2004). Molecular screening of ALK1/ACVRL1 and ENG genes in hereditary haemorrhagic telangiectasia in France. *Human Mutation,* 23:289-99.

Letterboer, T. G. W., Zewald, R. A. Kamping, E. J., Haas, G, Mager, J. J. *et al.* (2005). Heredidary hemorrhagic telangiectasia: *ENG* and *ALK1* mutations in Dutch patients. *Human Genetics,*116:8-16.

Levine, R. J., Lam, C. L., Qian, C., Yu, K. F., Maynard, S. E. *et al.* (2006). Soluble Endoglin and Other Circulating Antiangiogenic Factors in Preeclampsia. *The new England Journal of Medicine*,10:992-1005.

Li, W.D., Sorensen, L.K., Brooke, B.S., Urness, L.D., Davis, E.C. *et al.* (1999). Defcetive angiogenesis in mice lackinf endoglin. *Science* 284:1534-1537.

Llorca, O., Trujillo, A., Blanco, F.J. & Bernabeu, C. (2007). Structural model of human endoglin, a transmembrane receptor responsible for hereditary haemorrhagic telangiectasia. *J Mol Biol,* 365:694-705.

López-Novoa, J. M., & Bernabeu, C. (2010). The physiological role of endoglin in the cardiovascular system. *American Journal of Physiology. Heart and Circulatory Physiology*, 299:959-974.

Lund V. J & Howard, D. J. (1999). A treatment algorithm for the management of epistaxis in hereditary haemorrhagic telangiectasia. *Am J Rhinol,* 13:319-322.

Lux, A., Gallione, C. J & Marchuk, D.A. (2000). Expression analysis of endoglin missense and truncation mutations: insights into protein structure and disease mechanisms. *Human Molecular Genetics,* 9:745-755.

Mallet, C., Lamribet, K., Giraud, S., Dupuis-Girod, S., Feige, J. J., Bailly, S., & Tillet, E. (2015). Functional analysis of endoglin mutations from hereditary hemorrhagic telangiectasia type 1 patients reveals different mechanisms for endoglin loss of function. *Human Molecular Genetics*, 24:1142-1154.

Mandzia, J. L., Karel, G., Faughnan, M. E., & Hyland, R. H. (1999). Spinal cord arteriovenous malformations in two patients with hereditary hemorrhagic telangiectasia, *Childs Nerv Syst,*15:80-83.

McAllister, K. A., Grogg, K. M., Johnson, D. W., Gallione, C. J., Baldwin, M. A. *et al*. (1994). Endoglin, a TGF-β binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1. *Nat Genet,* 8:345-351.

McDonald J., Gedge F., Burdette A., Carlisle J., Bukjiok C. J., Fox M., *et al*. (2009). Multiple sequence variants in hereditary haemorrhagic telangiectasia cases: illustration of complexity in molecular diagnostic interpretation. *J. Mol. Diag,*11:569-575.

McDonald, J., Bayrak-Toydemir, P., & Pyeritz, R. E. (2011). Hereditary haemorrhagic telangiectasia, an overview of diagnosis, management, and pathogenesis. *Genet. Med*,13:607-616.

McDonald, J., Wooderchak-Donahue, W., VanSant Webb, C., Whitehead, K., Stevenson, D. A., & Bayrak-Toydemir, P. (2015). Hereditary hemorrhagic telangiectasia: Genetics and molecular diagnostics in a new era. *Frontiers in Genetics*, 5:1-8.

Mei-Zahav, M., Letarte, M., Faughnan, M. E., Abdalla, S. A., Cymerman, U. & MacLusky, I. B. (2006). Symptomatic children with hereditary haemorrhagic telangiectasia: a pediatric center experience. *Arch Pediatr Adolesc, Med,* 160:596-601.

Mukwaya, A., Peebo, B., Xeroudaki, M., Ali, Z. *et al*. (2016). Factors regulating capillary remodeling in a reversible model of inflammatory corneal angiogenesis. *Scientific reports,*6:1-15.

Murakami, M. & Simons, M. (2009). Regulation of vascular integrity. *J Mol Med,* 87: 571-582.

NanoDrop technical support bulletin (2007) NanoDrop Technologies Inc 5/07.

Olitsky, S. E. (2010). Hereditary hemorrhagic telangiectasia: diagnosis and management. *American Family Physician*, 82:785-790.

Park, S. O., Wankhede, M., Lee, Y. J., Choi, E., Fliess, N., Choe, S. *et al.* (2009). Real-time imaging of de novo arteriovenous malformation in a mouse model of hereditary hemorrhagic telangiectasia. *J Clin Invest*, 119:3487-3496.

Park, S., Sorenson, C. M., & Sheibani, N. (2015). PECAM-1 isoforms , eNOS and endoglin axis in regulation of angiogenesis, *Clin Sci (Lond),*129:217-234.

Pece, N., Vera, S., Cymerman, U., White, R. I., Wrana, J. L. & Letarte, M. (1997). Mutant endoglin in hereditary hemorrhagic telangiectasia type 1 is transiently expressed intracellularly and is not a dominant negative. *Journal of Clinical Investigation*,100:2568-2579.

Pfarr, N., Fischer, C., Ehlken, N., Becker-grünig, T., López-gonzález, V., Gorenflo, M. *et al.* (2013). Hemodynamic and genetic analysis in children with idiopathic , heritable , and congenital heart disease associated pulmonary arterial hypertension. *Respiratory Research,*14:1-9.

Rebeiz, E. E., Bryan, D. J., Ehrlichman, R. J., & Shapshay, S. M. (1995). Surgical management of life-threatening epistaxis in Osler-Weber- Rendu disease. *Ann Plast Surg*; 35:208-213.

Rubi-Ruiz, M.E., Peredo-Escarcega, A. E., Cano-Martinez, A & Guarner-Lans, V. (2015). An Evolutionary Perspective of Nutrition and Inflammation as Mechanisms of Cardiovascular Disease. *International Journal of Evolutionary Biology,* 15:1-10.

Ruiz-Llorente, L., Gllardo-Vara, E., Rossi, E., Smadja, D. M., Botella, L.M & Bernabeu, C.(2017). Endoglin and alk1 as therapeutic targets for hereditary haemorrhagic telangiectasia. *Experts Opinion on Therapeutic Targets,* 21:933-947.

Sabbá C. (2005). A rare and misdiagnosed bleeding disorder: hereditary hemorrhagic telangiectasia. *J Thromb Haemost*, 3:2201-2210.

Sáez, P. J., Villalobos-labra, R., Westermeier, F., & Sobrevia, L & Farias-Jofre, M.(2014). Modulation of endothelial cell migration by ER stress and insulin resistance : a role during maternal obesity ? *Front Pharmacol,* 189:1-10.

Schwochow, D., Serieys, L.E.K., Wayne, R.K & Thalman, O.(2012). Efficient recovery of whole RNA – a comparison of commercial RNA extraction protocols for high-throughput applications in wildlife species. *BMC Biotechnology,* 12:1-12.

Shin J, H., Park S, J & Sung, K. (2010). Embolotherapy for pulmonary arteriovenous malformations in patients without hereditary haemorrhagic telangiectasia. *Korean J Radiol,* 11:312-19.

Sankelo, M., Halme, M., Laitinen, T & Mattila., P.S. (2008). Hereditary hemorrhagic telangiectasia type 1 and 2 mutations in Finland. *Acta Oto_Laryngologica,* 128: 1238-1241.

Scharpfenecker, M., Floot, B., Russell, N. S., ten Dijke, P., & Stewart, F. A. (2009). Endoglin haploinsufficiency reduces radiation-induced fibrosis and telangiectasia formation in mouse kidneys. *Radiotherapy and Oncology*, 92:484-491.

Sharathkumar A. A & Shapiro, A. (2008). Hereditary haemorrhagic telangiectasia. *Haemophilia,* 14:1269-1280.

Schrader, C., Schelke, A., Ellerbroek, L & Johne, R. (2012). PCR inhibitors – occurrence, properties and removal. *Journal of Applied Microbiology,* 115: 1014-1026.

Schwochow, D., Serieys, L.E.K. & Wayne, R.K. (2012). Efficient ecovery of whole blood RNA – a comparison of commercial RNA extraction protocols for high-throughput applications wildlife species. *BMC Biotechnology,* 12:1-12.

Shovlin, C. L. (2010). Hereditary haemorrhagic telangiectasia: Pathophysiology, diagnosis and treatment. *Blood Reviews*, 24:203-219.

Shovlin, C. L., Hughes, J. M. B., Scott, J., Seidman, C. E., & Seidman, J. G. (1997). Characterization of Endoglin and Identification of Novel Mutations in Hereditary Hemorrhagic Telangiectasia. *Am. J. Hum. Genet.* 61:68-79.

Sreejayan, N. & Ren, J. (2007). Vascular pathology protocols. *Methods of Molecular Medicine,* 139:1-372.

Statistics South Africa.(2012). Census 2011 Census in brief. Statistics South Africa; 2012. Report No. 03-01-41.

Sweet, K., Willis, J., Zhou, X.P., Gallione, C., Sawada, T., Alhopuro, .*et al*. (2005). Molecular classification of patients with unexplained hamartomatous and hyperplastic polyposis. *JAMA,* 294:2465-2473.

Teekakirikul, P., Milewicz, D. M., Miller, D. T., Lacro, R V., Regalado, E.S., *et al*. (2013). Thoracic aortic disease in two patients with juvenile polyposis syndrome and SMAD4 mutations. *American Journal of Medical Genetics,* 161:185-191.

Ten Dijke, P., & Hill, C. S. (2004). New insights into TGF-??-Smad signalling. *Trends in Biochemical Sciences*, 29:265-273.

Thomas, M. J., Surendran, S., Abraham, M., Rajavelu, A & Kartha, C.C. (2016). Genetic and epigenetic mechanisms in the development of arteriovenous malformations in the brain. *Clin Epigenetics*, 8:1-8.

Tørring, P.M., Kjeldsen, A.D., Ousager, L.B & Brusgaard, K. (2017). *ENG* mutational mosaicism in a family with hereditary hemorrhagic telangiectasia. *Molecular Genetics & Genomic Medicine,* 1:1-5.

Tual-Chalot, S., Oh, S. P., & Arthur, H. M. (2015). Mouse models of hereditary hemorrhagic telangiectasia: Recent advances and future challenges. *Frontiers in Genetics*, 5:1-12.

Van Cutsem, E., Rutgeerts, P., Vantrappen, G. (1990). Treatment of bleeding gastrointestinal vascular malformations with oestrogen-progesterone. *Lancet*, 335:953-955.

Waite, K. A & Eng, C. (2003). From developmental disorder to heritable cancer: it's all in the BMP/TGF-beta family. *Nat Rev Genet*, 4:763-773.

Westermann, C. J. J., Rosina, A. F., De Vries, V., & de Coteau, P. (2003). The prevalence and manifestations of hereditary hemorrhagic telangiectasia in the Afro-Caribbean population of the Netherlands Antilles: a family screening. *American Journal of Medical Genetics,*116:324-328.

Williams, W.V., Rosenbaum, H & Weiner, D.B. (1992). Effect of RNA concentration on cDNA synthesis for DNA amplification. *Cold Spring Habour laboratory,* 2:56-88.

Wooderchak-Donahue, W. L., McDonald, J., O'Fallon, B., Upton, P. D., Li, W., Roman, B. L., *et al*. (2013). BMP9 mutations cause a vascular-anomaly syndrome with phenotypic overlap with hereditary haemorrhagic telangiectasia. *Am. J. Hum. Genet*, 93:530-537.

Zhang, W & Dolan, M.E. (2010). Impact of the 1000 Genome Project on the next wave of pharmacogenomics discovery. *Pharmacogenomics,* 11:249-256.

Websites

https://embnet.vital-itch/software/LALIGN_form.html (LALIGN server)

http://genetics.bwh.harvard.edu/pph2/ PolyPhen

http://sift.jvci.org/ (SIFT)

www.arup.utah.edu (HHT mutation database)

www.ensembl.org (Genome browser for analysing results)

www.idtdna.com (IDT OligoAnalyzer Tool version 3.1 to check primer parameters)

www.internationalgenome.org (Project phase 3 population groups)

www.mcdb.ucla.edu (RNA Isolation and RT-PCR Analysis protocol)

www.ncbi.nhlm.gov/clinvar/intro/ (Clinvar, connects the relationship between mutation and human health)

www.thermofisher.com (TRIzol® LS Reagent manual)

www.mcdb.ucla.edu (RNA Isolation and RT-PCR Analysis protocol)

www.researchgate.net (Suggested optimal RNA concentration)

Appendix 1 - Ethics approval letter

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

25 May 2016

MISS KT PETA
DEPT GENETICS
FACULTY OF HEALTH SCIENCES
UFS

Dear Miss KT Peta

HSREC 79/2016
PROJECT TITLE: MUTATION DETECTION IN THE ENDOGLIN GENE IN A FAMILY AFFECTED WITH HEREDITARY HAEMORRHAGIC TELANGIECTASIA

1. You are hereby kindly informed that, at the meeting held on 24 May 2016, the Health Sciences Research Ethics Committee (HSREC) approved the above project after all conditions were met.
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 HSREC for approval.
4. A progress report should be submitted within one year of approval and annually for long term studies.
5. A final report should be submitted at the completion of the study.
6. Kindly use the HSREC NR as reference in correspondence to the HSREC Secretariat.
7. The HSREC 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 46) (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 HSREC of the Faculty of Health Sciences.

Yours faithfully

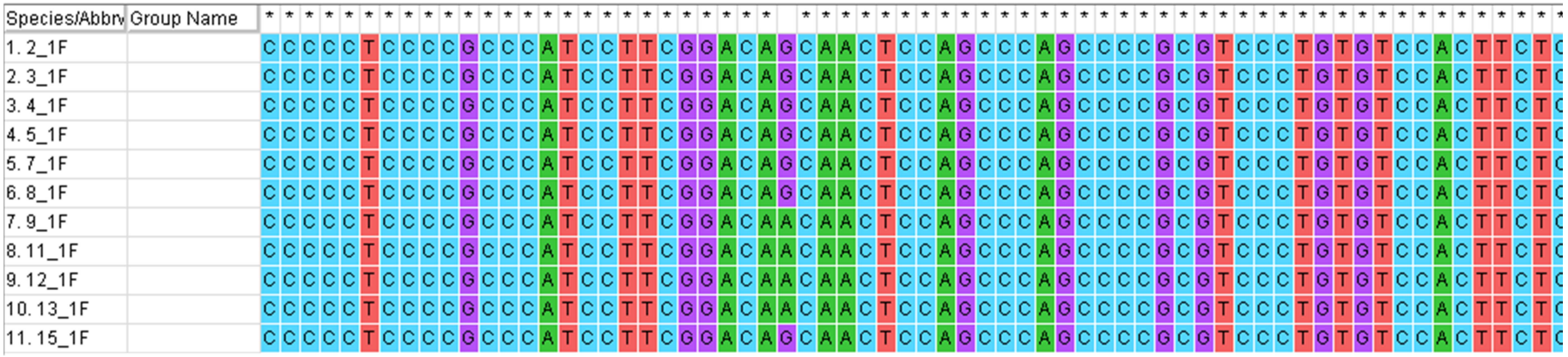
PROF VJ STEINBERG
VICE CHAIR: HEALTH SCIENCES RESEARCH ETHICS COMMITTEE

Cc: Dr G Mare; Prof MJ Coetzee

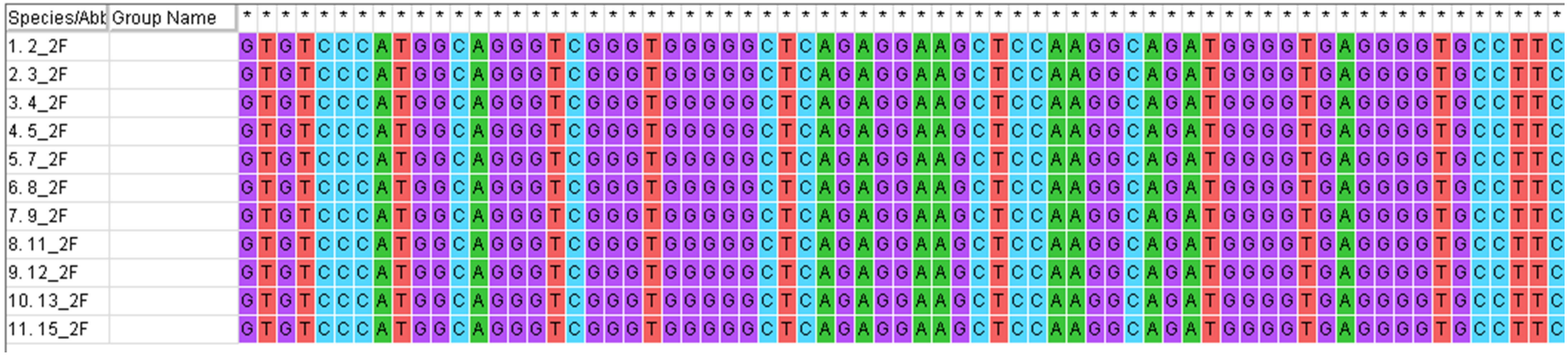
Health Sciences Research Ethics Committee
Office of the Dean: Health Sciences
T: +27 (0)51 401 7795; 7784 | F: +27 (0)51 444 4158 | E: ethicsfhs@ufs.ac.za
Block D, Dean's Division, Room 3104 | P.O. Box/Posbus 339 (Internal Post Box G40) | Bloemfontein 9300 | South Africa
www.ufs.ac.za

Appendix 2

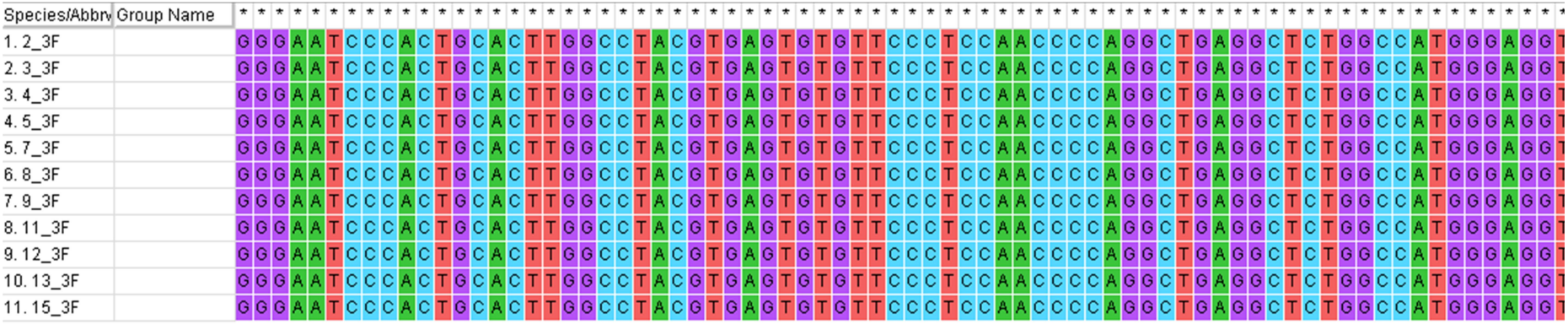
Exon 1



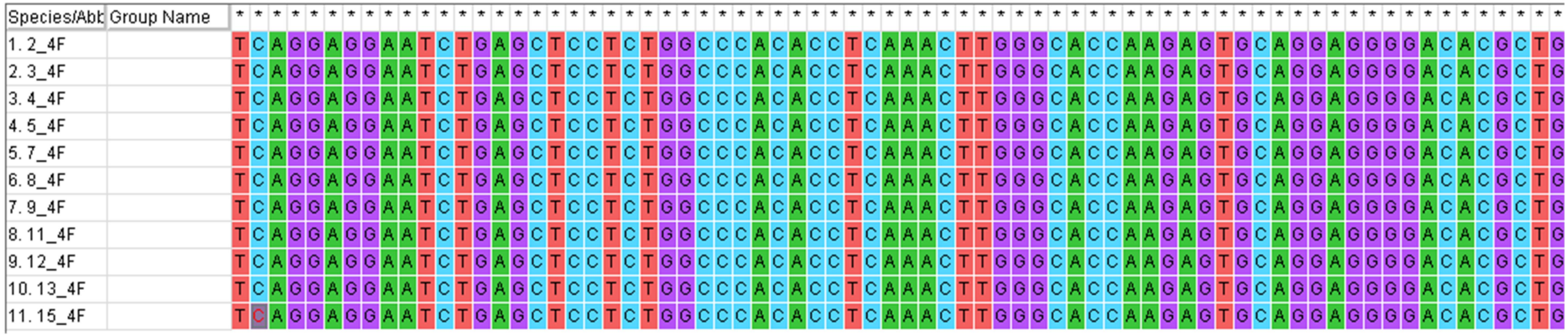
Exon 2



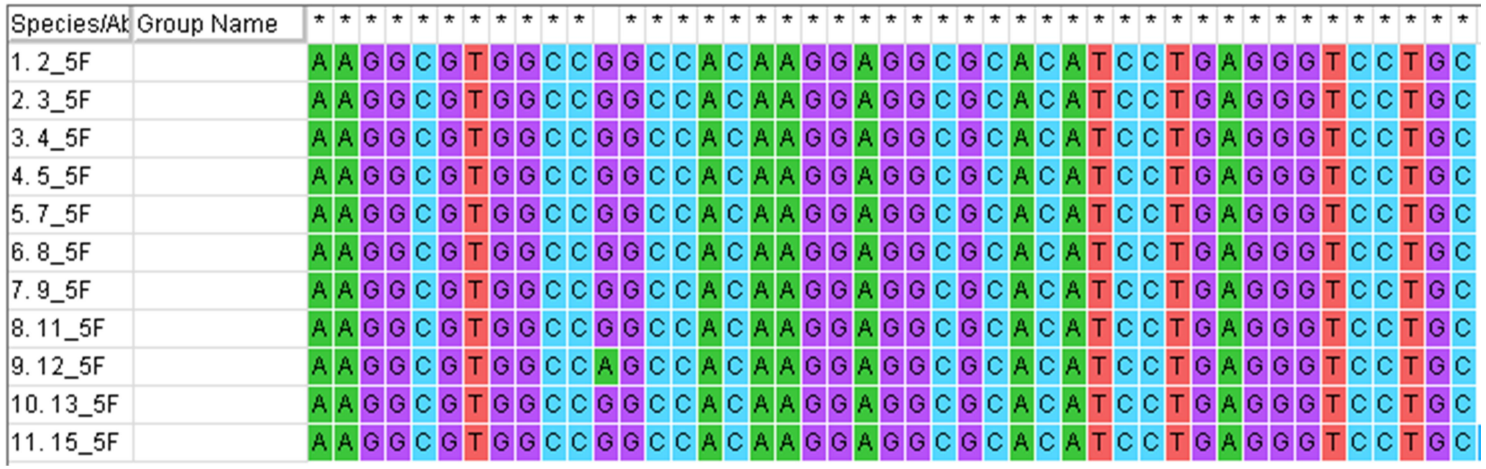
Exon 3


Species/Abbrv
Group Name
1. 2_3F
2. 3_3F
3. 4_3F
4. 5_3F
5. 7_3F
6. 8_3F
7. 9_3F
8. 11_3F
9. 12_3F
10. 13_3F
11. 15_3F

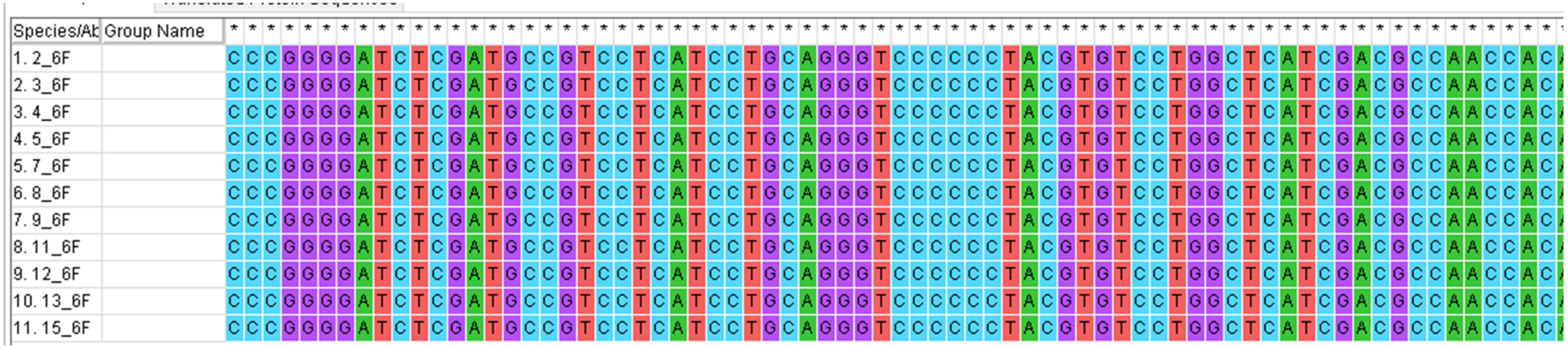
Exon 4


Species/Abb
Group Name
1. 2_4F
2. 3_4F
3. 4_4F
4. 5_4F
5. 7_4F
6. 8_4F
7. 9_4F
8. 11_4F
9. 12_4F
10. 13_4F
11. 15_4F

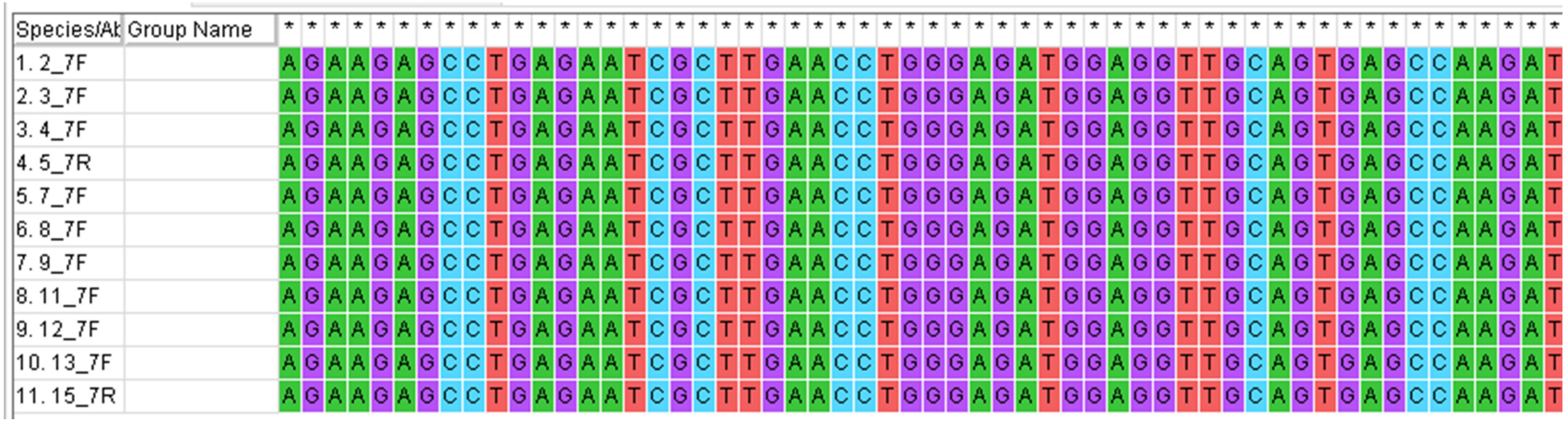
Exon 5


Species/Ab
Group Name
1. 2_5F
2. 3_5F
3. 4_5F
4. 5_5F
5. 7_5F
6. 8_5F
7. 9_5F
8. 11_5F
9. 12_5F
10. 13_5F
11. 15_5F

Exon 6


Species/Ab
Group Name
1. 2_6F
2. 3_6F
3. 4_6F
4. 5_6F
5. 7_6F
6. 8_6F
7. 9_6F
8. 11_6F
9. 12_6F
10. 13_6F
11. 15_6F

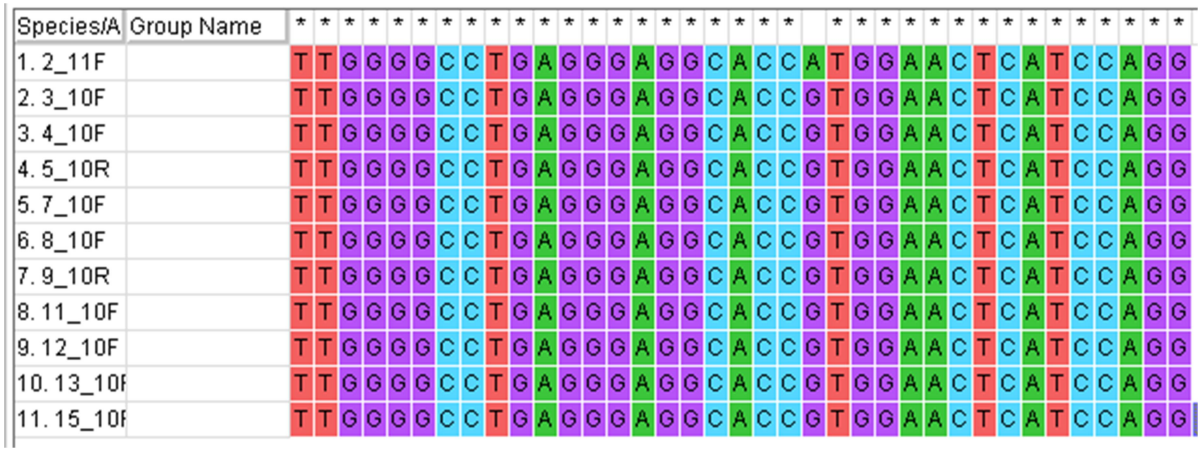
Exon 7-8



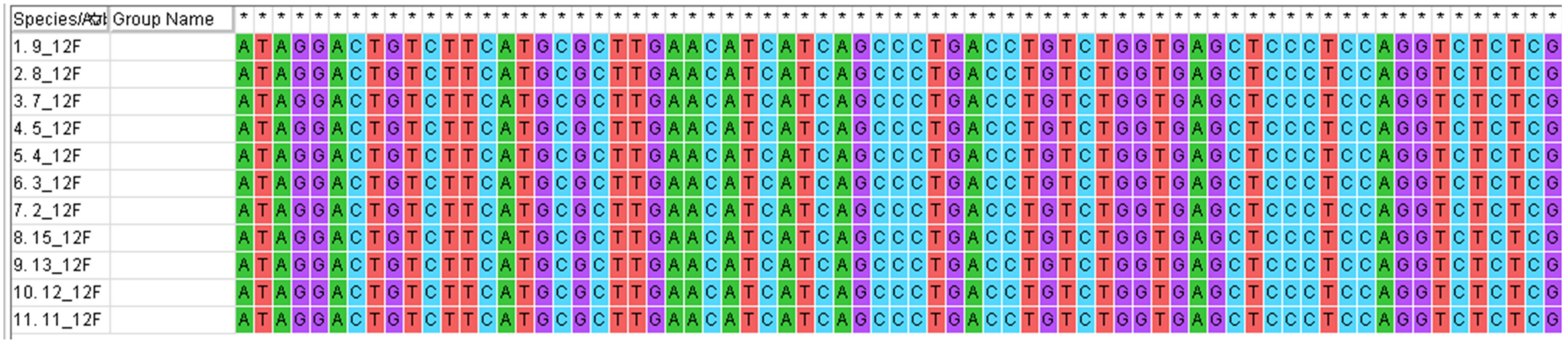
Exon 9

Due to non-specific binding, the sequences could not align and therefore, there is no alignment for this exon.

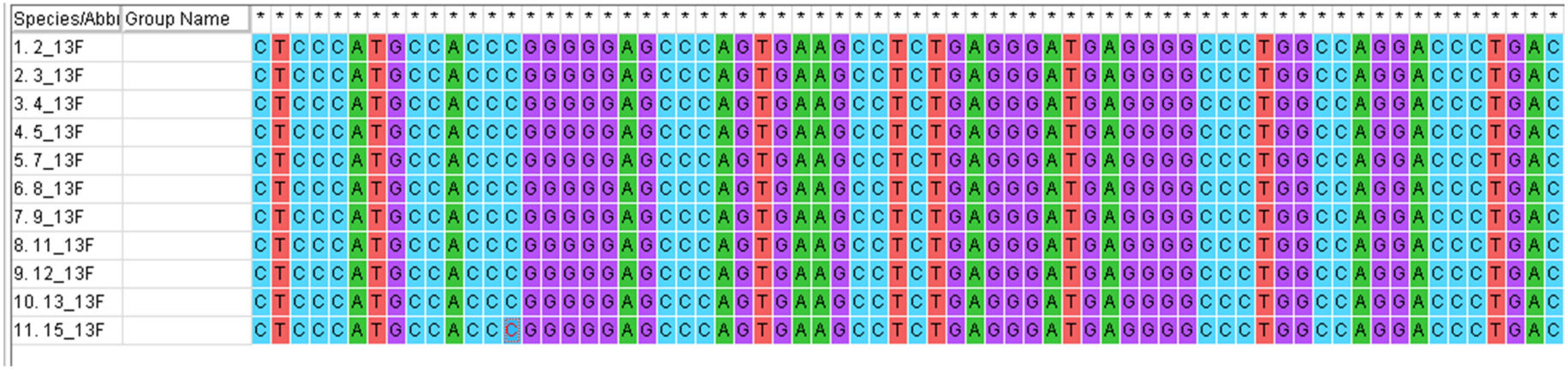
Exon 10-11



Exon 12



Exon 13-14


Species/Abbr
Group Name
1. 2_13F
2. 3_13F
3. 4_13F
4. 5_13F
5. 7_13F
6. 8_13F
7. 9_13F
8. 11_13F
9. 12_13F
10. 13_13F
11. 15_13F

Appendix 3

Guidelines from the SASHG Committee for publication purposes regarding: Nomenclature for South African populations

Guidelines from the SASHG Committee for publication purposes regarding: Nomenclature for South African populations

Prepared by Lisa Roberts, Jacquie Greenberg and Michele Ramsay, with input from Himla Soodyall and the current SASHG committee - June 2013

When publishing genetic or biomedical research studies involving patients, controls or population groups in South Africa, it is often relevant and necessary to identify the groups by their ethnic affiliations. The categories are usually used as proxies for genetic ancestry, but the ways in which individuals "self-identify" are influenced by many factors, including geographic origin, language and culture. As genetic and biomedical researchers, we take cognisance of the fact that these are labels of convenience that do not necessarily accurately reflect ancestry, nor do they define homogenous groups.

We recognise that preferred identifiers for South African ethnic groups vary between individuals, groups, nations and organisations. In citing ethnic labels, we therefore recommend the use of the names or categories that the people call themselves. Today, South Africans broadly consider themselves White, Black (black African), Coloured or Indian. Importantly too, these are also the categories used in the South African population census (also at www.statssa.gov).

Due to the international opinion that "Coloured" is a derogatory term, it is necessary to clarify that in the South African context, populations who self-identify as Coloured have a complex history of ancestrally derived admixture. These groups, also frequently termed "Mixed Ancestry" are ancestrally derived from admixtures of one or more of the indigenous African populations (Khoe- and San-speaking or Bantu-speaking), immigrants from Western Europe, or slave labourers from West Africa, Indonesia, Madagascar, Java, India and Malaysia. The term "Coloured" in South Africa is therefore a name that encompasses a wide range of people who are unique to this country.

As a guide, the SASHG committee therefore recommends, for publication purposes for genetic and biomedical studies, the use of the following nomenclature for South African populations: White, Black African, Coloured and Indian. This terminology should be accompanied by a brief description of the geographic origin and inclusion criteria of the participants of the specific research study.

If necessary, depending on the journal and/or the reviewers' comments, it is acceptable to substitute the term "Mixed Ancestry" for "Coloured", accompanied by a description of the cohort(s) under study.

Black African communities in South Africa also identify themselves by language: the Zulu people or amaZulu (who speak isiZulu), Xhosa or amaXhosa (isiXhosa), Sotho (seSotho), etc. and these affiliations should be used if the data are available and relevant to the study. Similarly, since San hunter-gatherers and Khoe pastoralists identify themselves with their community names (e.g. !Xun, ǂKhomani, //Ganaa, Hai//om, etc.) these affiliations should be used when available and relevant, however if they are grouped together the term Khoe-San is recommended.

We acknowledge that in the future there may no longer be a need for such "labels" but at present, particularly for genetic research purposes, this information is pertinent and relevant.