*School of Medicine*

**Human Genetics Division**

# *Molecular Genetic Studies of Arterial Aneurysms* and *Arterial Dissection Diseases.*

**Doctor of Philosophy**

**Waseem Barakat**

# Abstract:

**Background**

Arterial rupture like subarachnoid haemorrhage (SAH) is a severe, often fatal, neurological consequence of proneness to rupture of the cerebral vasculature, usually arteries. Of cerebrovascular disease, SAH accounts for about 10 %. Five to ten percent of cases are familial and are ascribable to unidentified single major gene defects. Other aneurysms leading to fatal disease like aortic aneurysm are due to mutations in the fibrillin-1 gene (MFS1 stands for Marfan syndrome) and the transforming growth factor beta-receptor II gene (*TGFβRII*). Some reports suggest an association of intracranial aneurysms and other aortic aneurysmal diseases. Molecular studies of familial and sporadic aneurysms have been initiated in order to identify unknown gene (or genes) and pathways involved in this form of arterial rupture.

**Candidate genes**

**Elastin gene *(ELN)***

A genome wide linkage study was performed searching for possible genes that may be involved in the onset of familial SAH. This study was performed on 104 Japanese affected sib pairs and demonstrated that the best linkage evidence was on chromosome 7q11 (with LOD score of 3.22)[1], this region is close to the elastin gene (*ELN*). Moreover, Onda *et al* 2001 shown that homozygous haplotypes in intron-20/intron-23 of the *ELN* are at high risk of developing intracranial aneurysm (ICA)[1]. Furthermore, James *et. al.* 2003 confirmed that chromosome 7q11 locus is a predisposing factor for intracranial aneurysm (ICA)[2].

**Transforming Growth Factor Beta Receptor 2 *(TGFβRII)***

Another locus that is responsible for MFS2 (that shows aortic aneurysms) was evidenced to be associated with *TGFβRII*[3].

**Fibrillin-1 gene (*FBN1)***

In two separate families it was shown that some mutations in the fibrillin-1 gene are responsible for aortic aneurysms, these patients did not show MFS phenotype[4,5].

**Hypotheses:**

I hypothesise that specific major or minor gene effects on occurrence or outcome of cerebral aneurysms and subarachnoid haemorrhage can be identified through a combination of linkage and association studies.

**DNA Samples**

Samples collected for this study are divided in to three categories.

1- Familial SAH: 130 blood DNA samples, of the blood samples 80 samples came from Glasgow and 51 DNA familial samples were collected. The tgfbr2 family was from Southampton/Channel Islands, apparently with some of the Channel Island founders originating from France.

2- Sporadic SAH and control groups: 214 DNA samples (from blood) were collected, 137 are sporadic patients, who suffered from aneurysmal subarachnoid haemorrhage SAH (range, 23 to 75 years; mean age, 50 years; ancestry of white European) and received surgical treatment in the Wessex Neurological Centre, Southampton General Hospital . 10 ml venous Blood samples collected from patients are having a clinical history of SAH with an association of abnormal Computerised Tomography (CT) scan. The second group contains 77 Head injuries control.

3- DNA samples from The British Women's Heart and Health Study (2890 samples).

**Aims and Objectives**

To determine the mutations that could be responsible for the onset of SAH, this will involve screening of 34 exons and part of the introns of the elastin gene using Denaturing

High-Performance Liquid Chromatography (DHPLC), positive results were confirmed by DNA sequencing. In addition, screening of 2.2 kb of 5' and 0.4 kb of 3' of the elastin gene was performed.

Genotyping of three SNPs was performed on both sporadic SAH and controls to investigate any genetic association between elastin and sporadic SAH as proposed by the Japanese paper these SNPs(that form a haplotype) are: intron 20 A/G **C** (rs2856728); exon 20 C/T (rs2071307); intron 23 A/G. (Hideak Onda et al 2001) **do not have rs number.**

Analysis was performed using 2890 BWHHS samples on the Gly422Ser mutation to search for any association of this mutation with stroke. In addition, GT microsatellite analysis was performed to study association of SAH and GT repeat

Linkage analyses of *TGFβRII* using five STR markers (four-tetras {D3S2466, D3S4535, D3S2432, and D3S1768} and one di-nucleotides) were launched to study the possibility of linkage of this gene to SAH in one of the French families, the di-nucleotide was used to perform STR genotyping and association analysis with SAH (D3S3727).

To detect the mutations that are responsible for the lethal aortic rupture (in the fibrillin gene), optimisation of exons 27, 28 and 56 were performed (these exons are involved in fibrillin-elastin interaction (exons 27 and 28) and fibrillin-fibrillin interaction (exon 56), these exons were chosen for the EndoVII method screening of any unknown mutations in sporadic patients with Intracranial aneurysms. Performing *In silico* analysis on any mutation found in any of the above genes.

**Results and Conclusions**

1- No significant association was found between haplotypes in Sporadic SAH and control groups.
2- One exonic mutation was found in exon 20(rs2071307), six mutations were found in the intronic regions of the *ELN* gene.
3- Genotyping results of the exon 20 SNP using 2890 samples of the British Women's Heart and Health Study (BWHHS) evidenced a possible association with stroke(p=0.05), more analysis is required using at least 10 000 sample of a case-control stroke cohort.
4- No significant association was found between *ELN* GT microsatellite and SAH in sporadic and familial *vs*. control in the following five models: additivity model; major expansion model (anticipation); recessive model, loss of heterozygosity model and dominant model.
5- There is no evidence from these studies to support a role for variation in the elastin gene in predisposition to SAH.
6- SAH may not be associated with the elastin gene in European ethnicities.
7- Three novel mutations/SNPs (POSITION -1050 POSITION -1162 C>G POSITION -2253 G>C IVS18+47 G>C) were detected in the 5' region and one known exonic SNP rs2071307 (of the elastin gene).
8- More studies are required to investigate the possibility of 5' mutations involvement in the onset of SAH (which are relevant to promoter function).
9- *ELN* (in SAH) may be associated with the mRNA stability (possibly by exonic mutation) or the amount of mRNA of elastin or unusual alternative splicing.
10- *TGFβRII* may be linked with SAH in the one French family (using D3S3727).
11- Microsatellite studies of a dinucleotide (CA repeat called D3S3727) within the *TGFβRII* gene did not show association with sporadic SAH in any model.
12- Two penta nucleotides and two exonic SNPs were detected in the *FBNI* gene, a study shown a positive association between ATTTT (which is the complementary of TAAAA in my study) and pulse pressure (especially 2-3 genotype of this pentamer p=0.003) in over 50 years[6], this genotype may affect the elastic property of the aorta and other arteries in combination of ageing. Future analysis on a stroke cohort and functional analysis may be needed.

# *To my parents.*

"وقضى ربك ألا تعبدوا إلا إياه وبالوالدين إحسانا إما يبلغن عندك الكبر أحدهما أو كلاهما فلا تقل لهما أف ولا تنهرهما وقل لهما قولاً كريمًا واخفض لهما جناح الذل من الرحمة وقل ربي ارحمهما كما ربياني صغيرًا"

*الإسراء 23-24*

***" And your Lord decreed that you shall not serve except Him, and do good to your parents. When one of them or both of them reaches old age, do not say to them a word of disrespect nor raise your voice at them, but say to them a kind saying. And lower for them the wing of humility through mercy, and Say: "My Lord, have mercy upon them as they have raised me when I was small"".***
***Banî Israel 17:23-24***

# *Acknowledgements*

I would like to take this opportunity to thank my supervisor Professor Ian Day for his encouragement, support, advice and supervision throughout this project. In addition, my thanks go to Dr Beatrice Zhang for her support, encouragement and supervision in the first year of my project. To all members past and present of Human Genetics Division who supported me. All their help and useful contributions are appreciated. My thanks to Dr Mohib Uddin & Dr Ahmad Islam for their help in English corrections.

Thanks to the British Hart foundation for supporting my project.

Thanks to my wife for her support and encouragement throughout my PhD project.

# Table of Contents


**Abstract:** 2
**Index of Figures:** 10
**Index of Tables:** 14
**Abbreviations:** 16
**ACE Angiotensin converting enzyme** 16
**Chapter One** 17
**1.0 Introduction** 17
1.1 Overview 17
1.2 SAH 21
1.3 Incidence of SAH 24
1.4 Risk Factors and Outcomes of SAH 24
1.5 Familial SAH 26
1.6 Anticipation in the Familial SAH 27
1.7 Abdominal Aortic Aneurysm (AAA) 28
1.7.1 Incidence of Abdominal Aortic Aneurysm (AAA) 28
1.7.2 Risk factors for Abdominal Aortic Aneurysm (AAA) 28
1.8 Familial Abdominal Aortic Aneurysm (AAA) 29
1.9 Genetics of Arterial Aneurysms 30
1.10 SNPs, Protein Structure, Function and Disease 33
1.11 Microsatellites 33
1.12 Tropoelastin Gene and Protein 35
1.12.1 Tropoelastin Gene (*ELN*; MIM# *130160) 35
1.12.2 Exons and Introns 35
1.12.2.1 *Alu* Repetitive Sequences 36
1.12.2.2 Intron 18 GT Repetitive Sequence 36
1.12.2.3 Trinucleotides Repetitive Sequences 37
1.12.3 Splicing and Alternative Splicing 37
1.12.4 Elastin Gene Regulation in 5' Region 38
1.12.5 Elastin Gene Regulation at 3' Region 38
1.12.6 Transcription Regulation 39
1.12.6.1 Decreased Production of the Elastin Protein 39
1.12.6.2 Increased Production of Elastin Protein 39
1.12.7 Post-transcriptional Regulation of the Elastin Gene 40
1.12.8 Tropoelastin Protein 41
1.12.9 Elastogenesis 42
1.12.9.1 Modification and degradation of tropoelastin 42
1.12.9.2 Secretion of Tropoelastin 43
1.12.9.3 Post secretion of Tropoelastin 44
1.12.10 Elastin Protein Formation 45
1.12.10.1 Cross-Linking of Tropoelastin 45
1.12.11 Some Functions of Tropoelastin and Elastin 46
1.13 Fibrillin 1(*FBN1*; MIM# 134797) 47
1.13.1 General 47
1.13.2 Structure of *FBN1* Gene 47
1.13.3 *FBN1* Mutations and Diseases 48
1.13.4 Fibrillin-1 protein: 49
1.13.5 Fibrillogenesis and Assembly Matrix incorporation defects and some Interactions 50
1.13.6 Some Functions of Fibrillin-1 53

1.14 Transforming Growth Factor Beta Receptor 2 (*TGFβRII*) and linkage to aortic aneurysm 54
1.14.1 *TGFβRII* General 54
1.14.2 *TGFβRII* Role in Angiogenesis 55
1.14.3 Genetics of *TGFβRII* and Marfan's Syndrome Type2 56
1.14.4 Relationship of *ELN* and *TGFβRII* 58
1.14.5 Relationship of *FBN1* and *TGFβRII* 59
1.14.6 Sporadic SAH and *TGFβRII* STR (D3S3727) 59
1.15 Possible methods for investigating variation in genes 60
1.15.1 Automated DNA Sequencing of PCR product 60
1.15.2 Amplification Refractory Mutation System (ARMS) 61
1.15.3 Restriction fragment length polymorphism 61
1.15.4 Mismatch Cleavage 61
1.15.5 The Endo VII / Double labelled primers 62
1.15.6 Microplate-Array-Diagonal-Gel Electrophoresis (MADGE) 64
1.15.11 GeneScan and Microsatellite Detection 64
1.15.12 Denaturing High-Performance Liquid Chromatography (DHPLC) 65
1.15.13 LightTyper (Odyssey) 66
1.16 *In silico* analysis 68
1.17 Hypotheses and Plan 69
**Chapter Two 70**
**2.0 Material and Methods 70**
2.1 Samples used in this study 70
2.2 Reagents used in this project 73
2.2.1 For PCR and Restriction Enzymes 73
2.2.2 For DNA Sequencing 73
2.2.3 For normal gel preparation and staining and Electrophoresis 73
2.2.4 For DHPLC 75
2.2.5 For GeneScan 75
2.2.6 Reagents for the Endo VII 76
2.2.7 Purification of PCR Probes for the EndoVII 77
2.2.8 Other reagents 77
2.3 Instruments used in this project 77
2.3.1 For PCR 77
2.3.2 Visualisation of PCR product 78
2.3.3 For DHPLC 78
2.3.4 DNA Sequencing and GeneScan 78
2.3.5 MADGE Former 78
2.3.6 Other Instruments and disposables 79
2.4 Primers and Probes Design 79
2.4.1 *ELN* Primers for DHPLC 79
2.4.2 *ELN* Primers for GeneScan 80
2.4.3 *ELN* Probes and Primers for Odyssey (LightTyper) Genotyping 80
2.5 PCR optimisation 80
2.6 Primer Design for *FBN1* . 81
2.7 Primer Design for LightTyper for the *ELN* gene. 81
**Chapter Three 82**
**3.0 Results 82**
3.1 Work Performed in this project 82
3.2 Phase, Arlequine and CONTING. 83
3.3 Results of DHPLC and DNA sequencing 86
3.3.1 Mutations of Exon 20 and Intron 20 86
3.3.2 IVS20+17 T>C 88
3.3.3 AC005056:37759 G>A 90

3.3.3.1 Mutation analysis in relation to splicing process using Exonic Splicing Enhancers finder ESEfinder[303] 93
3.3.3.2 RESCUE-ESE analysis 94
3.3.3.3 PolyPhen analysis 95
3.3.3.4 Genotyping of the *ELN* exon 20 Gly224Ser SNP on the BWHHS 96
3.3.3.5 Association Results of the sporadic SAH and control using 97
3.3.4 Mutation Intron 19 of the *ELN* gene (C/T rs2239691) 99
3.3.5 Exon 33 MIX 2 of the *ELN* gene (ss4943619 C/T rs3757587) 102
3.3.6 Exon 18 (mix 18) mutations of the *ELN* gene 105
3.3.7 IVS18+20DEL2 (rs5884930 –GT) of the *ELN* gene. 108
3.3.7.2 Results of GeneScan Test (*ELN*) 110
3.3.7.3 Results of Models proposed depending on GeneScan (*ELN*) 111
3.3.7.4 Allele Counts /Additivity Model results (*ELN*) 112
3.3.7.5 Major Expansion Model (Anticipation) (*ELN*) 112
3.3.7.6 Loss of Heterozygosity Model (*ELN*) 113
3.3.7.7 Recessive Model (*ELN*) 113
3.3.7.8 Dominant Model (*ELN*) 114
3.3.8 Exon 23 (IVS23+24 T>C) of the *ELN* gene 117
3.3.9 5' Flanking (POSITION -1050 C>T) (in the NCBI is G>A) 120
3.3.10 5' Flanking (POSITION -1162 C>G) of the *ELN* gene 123
3.3.11 5' Flanking (POSITION -1859 G>A rs3757583) of the *ELN* gene 126
3.3.12 5' Flanking (POSITION -2253 G>C) of the *ELN* gene 128
3.4 Results Summary 130
3.5 *In silico* analysis of the 5' region of the elastin gene 131
3.6 Results of linkage analysis performed on the French family on the *TGFRβII* gene 132
3.7 Results Linkage analysis using D3S3727 D3S2432, D3S4535 and D3S1768 markers 135
3.8.1 Allele Counts /Additivity Model results (*TGFβRII*) 143
3.8.2 Major Expansion Model (Anticipation) (*TGFβRII*) 146
3.8.3 Loss of Heterozygosity Model (*TGFβRII*) 148
3.8.4 Recessive Model (*TGFβRII*) 149
3.8.5 Dominant Model (*TGFβRII*) 149
3.9 Results of the Endo VII on exon 56 of the fibrillin-1 gene 155
3.9.1 PolyPhen analysis of the rs363831 Glu [E]> Asp [D] SNP 160
3.9.2 Results ESE finder on exon 56 of the fibrillin-1 gene rs363831 SNP 161
3.10 Results RESCUE-ESE on exon 56 of the fibrillin-1 gene rs363831 SNP 162
3.11 Results of the Endo VII on exon 28 of the fibrillin-1 gene 163
3.11.1 Results of the intronic STR VNTR on splicing[314] 168
3.12 Results summary of DNA sequencing for mutations detected by The Endo VII 169
**Chapter Four 170**
**4.0 Discussion and Conclusion 170**
4.1 Elastin gene SNPs Genotyping 170
4.2 Elastin Gene Scanning and Mutations 171
4.2.1 Hydrophobic domain mutations 172
4.2.2 Cross-Linking Domain Mutations 174
4.2.3 5' and 3' Scanning of the elastin gene 174
4.2.4 Other Mutations 175
4.2.5 Other mutations not detected in our *ELN* samples 177
4.3 GT Repeat Models of *ELN* related to SAH 179
4.3.1 Additivity Model 180
4.3.2 Major Expansion Model 180
4.3.3 Loss of Heterozygosity Model 180
4.3.4 Recessive Model 181

4.3.5 Dominant Model 181
4.4 Linkage *TGFβRII* to aortic dissection 181
4.5 D3S3727 STR in *TGFβRII* related to sporadic SAH 182
4.5.1 Additivity Model 182
4.5.2 Major Expansion Model 182
4.5.3 Loss of Heterozygosity Model 183
4.5.4 Recessive Model 183
4.5.5 Dominant Model 183
4.6 *FBN1* scanning of exons 27 28 and 56 using our sporadic SAH and control samples 184
**Chapter Five** **185**
**5.0 Further Studies** **185**
5.1 Functional analysis of some elastin 5' SNPs/mutations 186
5.2 Further analysis of the exonic SNP 186
5.3 Screening for collagen 1 alpha 2 gene 186
5.4 Studying genes that may be associated with ICAs and arterial architecture 187
5.5 Fibrillin gene studies of two exonic SNPs 187
5.6 Genetic scanning of *TGFβRII* gene 187
**Chapter Six** **188**
**6.0 References** **188**
**Appendix A:** **206**
PCR optimization 1 (For *ELN*) 206
Primers used in PCR optimisation of Elastin gene: 207
PCR condition for each primer is summarised in the following table: 208
DHPLC pre-treatment: 209
Pre-Sequencing Reaction 209
Sequencing Reaction 210
Odyssey Reaction Optimisation 210
Heteroduplex Formation and The Endo VII Cleavage Reaction 210
Heteroduplex Formation For The Endo VII Cleavage Reaction 211
Primers for *TGFBRII* 212
PCR optimization (with Betaine) 212
Primers used in PCR optimisation of The fibrillin 1 gene: 213
**Appendix B: Searching Transcription Factors** **215**
**Appendix C:** **217**
**Appendix D:** **222**
**Appendex E** **231**
**Appendix F:** **260**
**Appendix G:** **261**
**Appendix H:** **268**
**Appendix I :** **270**
**Appendix J :** **271**
**Appendix K:** **273**

# Index of Figures:

Figure 1:( Above) 3D with major structures of an artery adapted from [8]. (Lower) structure of artery cross section adapted from [9] ____ 18
Figure 2: Cerebral arteries are the same as other arteries in the body, but lacking external elastic lamina. There is the lumen (monocellular layer of endothelial cells (EC)); an intima (forming the basement membrane consists of collagen IV; XVIII; laminin; nidogen); internal elastic lamina (IEL) mainly elastin; tunica media containing smooth muscle cells and many ECM proteins like fibrillin-1 and elastin; finally adventitia ( fibroblasts nerve cells and collagen) adapted from[10]. ____ 19
Figure 3: The relationship between these three genes is shown in this figure; *TGFβRII* plays a role in the expression of both *FBN1* and *ELN*. The elastin gene is the functional candidate for the onset of ICA, also linkage analysis were supporting its role, fibrillin-1 gene mutations causing arterial aneurysm (in some cases SAH), both the product of these two genes shows interaction in the ECM; fibrillin-1 binds elastin via exons 27 and 28. The region coded by exon 56 binds fibrillin-1 protein. The general work performed in this study is shown in the pink squares. ____ 21
Figure 4: Common places where saccular intracranial aneurysms take place: ____ 23
Figure 5: Risk factors for SAH ____ 24
Figure 6:Possible effects of MMP in the origin of aneurysms, adapted from [111]. ____ 32
Figure 7: The human elastin gene: ____ 36
Figure 8: Some different splicing products of the elastin gene are present. ____ 37
Figure 9: Some control elements that are present in the 5' region of the elastin gene, it contains several SP1(only one is shown) and AP2 as well as putative glucocorticoid, cAMP, and TPA responsive elements, but no consensus TATA box or functional CAAT box (for more details see Appendex E) ____ 38
Figure 10: Secretion of tropoelastin: ____ 43
Figure 11: Tropoelastin release from 67kDa protein. Elastin binding protein complex is composed of 67-kDa subunit and two other membranal proteins of 61 and 55 kDa. The 67kDa protein binds both tropoelastin and galactosugars via two different binding sites, and then galactosugars binds to the 67kDa tropoelastin-protein complex are no longer attached together (adapted from[188]). ____ 44
Figure 12: *FBN1* gene is present on chromosome 15q15-21, contains 65 exons, and encodes for a large protein (2871 amino acids) this entire gene is present in NT_010194 genomic contig. ____ 48
Figure 13: This figure illustrate a model of elastic fibre formation in the extracellular matrix, (1) Formation of linear and lateral assembly. (2) MAGP-1 bind with the microfibrils at the bead area. (3) Binding of tropoelastin (transglutaminase binding) in another domain of the microfibrils. (4) Possible interaction of tropoelastin to another tropoelastin or to MAGP-1 and (5) Further deposition elastin and cross linking via lysyl oxidase. Adapted from [26] ____ 51
Figure 14: Three possible defects that may lead to MFS occurrence: (1) Mutations responsible for decreased fibrillin-1protein. (2) Mutations that can affect the efficiency of secretion. (3) Mutations that affect fibrillin-1and prevent it from proper interaction with proteins present in the ECM ( adapted from [5] ). ____ 52
Figure 15: Simplified mechanism of target gene to initiate transcription. Adapted from [253] _ 54
Figure 16 "Regulation of ECM behaviour by TGF-beta signalling and corresponding vascular defects observed in mice deficient in TGF-beta components. TGF-beta switches ECM behaviour via two distinct TGF-h type I receptor (ThR-I)/Smad pathways. Upon TGF-beta -induced heteromeric complex formation, activin receptor-like kinase ALK 5 and ALK1 are phosphorylated and activated by ThR-II kinase. Signalling of TGF-h through ALK5 and subsequent Smad2/3 phosphorylation leads to inhibition of ECM proliferation and migration. Signalling of TGF- beta through ALK1 via phosphorylation of Smad1/5 induces ECM proliferation and migration. Moreover, ALK1 signalling indirectly

inhibits ALK5-induced Smad-dependent transcriptional responses. Vascular defects of mice deficient in TGF-h signalling components are listed. Abbreviations: VCAM-1, vascular cell adhesion molecule-1; SMC, smooth muscle cell." Adapted from [262] ____ 56
Figure 17: The exact breakpoint in the *TGFβRII* gene (containing seven exons), breakpoint is in 3p24.1 between exon5 and exon 7. Adapted from[77] ____ 57
Figure 18: *TGFβRII* is shown in red colour, markers used in this study are in oval, one of the markers used is located in the target gene (D3S3727), the location of *TGFβRII* is 30.65 cM, markers selected are located in both sides of the gene, and the completely covered region is around 10 cM. ____ 58
Figure 19: ABI 377 sequencing machine. ____ 60
Figure 20: Principle of the Endo VII digestion: ____ 63
Figure 21: MADGE system can run 96 samples at the same time, a very high throuput method to run high number of samples. "The wells are 2mm square, the angle between the direction of electrophoresis and the 12-well rows of the array is 71.5 degree, and the track length per well is 26.5 mm" [292] ____ 64
Figure 22: Heteroduplex formation will result in four different combinations in heterozygous DNA, it involves denaturation the renaturation of the DNA product. ____ 65
Figure 23: The transgenomic wave DHPLC used in *ELN* scanning for heteroduplexes. ____ 65
Figure 24: Principle is using a single stranded PCR product (that contains a SNP) to be used as a target for a 5' Fluorescence probe, another probe contains a 3' dapcyl is responsible for preventing the FITC from emitting fluorescent light and need higher temperature to be denatured. ____ 67
Figure 25: In the lightTyper an increase in the temperature is performed to allow denaturation of the probes, samples are subjected to a UV light, when the probes are denatured and are away from the dabcyl they will fluoresces light that will be detected by special detectors. Adapted from [297] ____ 67
Figure 26: Our mutation in red box (G>A) yellow box contains the complementary sequence probe (with 5' FITC and 3' phosphate) of the segment that contains the SNP, at the 5' region of this sequence is the Dabcyl probe the dabcyl is in the 5' region. ____ 81
Figure 27: The exact sequence of the probes used in the LightTyper assay. ____ 81
Figure 28: This figure describes the pattern of exon 20 and intron 20 on sample number 5, method used is *ELN* EXON20MIX16@63. ____ 86
Figure 29: This figure is the pattern of exon 20 sample number 2, method used is *ELN* EXON20MIX16@63: ____ 87
Figure 30: Sequencing results of sample 5 (heterozygous) and sample two (normal) ____ 88
Figure 31: Sequence obtained from GeneBank AC005056, exact mutation is shown G ____ 89
Figure 32: Sequencing results of sample 5(normal) ____ 90
Figure 33: Sequencing results of sample 2 (heterozygous) ____ 90
Figure 34: Sequencing results of sample 5(normal) and sample two (heterozygous) ____ 91
Figure 35: Sequence obtained from GeneBank AC005056, exact mutation is shown G ____ 92
Figure 36: Results of the Exonic Splicing Enhancers (ESE) of the wildtype ____ 94
Figure 37: Results of the Exonic Splicing Enhancers (ESE) of the mutant sequence. ____ 94
Figure 38: DHPLC for samples 1,2 and 3. ____ 99
Figure 39: Sequencing results of sample 3(normal); sample 2 (heterozygous) and sample 1 (homozygous mutant). ____ 100
Figure 40: Sequence obtained from GeneBank AC005056 of the *ELN* gene , exact mutation is shown (G) ____ 101
Figure 41: DHPLC for samples 2 and 6. ____ 102
Figure 42: Sequencing results of sample above (normal) and sample down (heterozygous) 103
Figure 43: Sequence obtained from GeneBank AC005056, exact mutation is shown G ____ 104
Figure 44: DHPLC for samples 1 and 2, small difference can be seen. ____ 105
Figure 45: Sequencing results of sample 5(normal) and sample one (heterozygous) ____ 106
Figure 46: Sequence obtained from GeneBank AC005056, exact mutation is shown G ____ 107
Figure 47: Sequencing results of GT repeats are different in sample 1 and sample 8. ____ 108

Figure 48: Scores results of putative donor and acceptor sites for the wild type sequence, actual sites are marked, no changes where found when adding 1-100 GT repeats (complete sequence is preset in Figure 46) ____ 109
Figure 49: DHPLC for samples 1, 2. ____ 117
Figure 50: Sequencing results of sample 4(normal) and sample two (heterozygous) ____ 118
Figure 51 : Sequence obtained from GeneBank AC005056, exact mutation is shown (G )_ 119
Figure 52: Three types of mutations were detected, sample one is homozygous T, sample 2 is heterozygous T and C, sample three is homozygous C. ____ 120
Figure 53: The reverse sequence confirms the mutations, sample one is homozygous T, sample 2 is heterozygous T and C, sample three is homozygous. ____ 121
Figure 54: The exact sequence is shown, the position of T mutation is present. ____ 122
Figure 55: Sample number 2 is normal at C, sample 7 is heterozygous C/G ____ 123
Figure 56: Confirmation of C>G mutation is seen, sample seven has heterozygous mutation. ____ 124
Figure 57: The exact sequence is shown, the position of C>G mutation is present. ____ 125
Figure 58: Sample 5 and normal control sample are heterozygous G/C. Sample 8 is homozygous A ____ 126
Figure 59: G>A mutation is seen, it is at position-1859. ____ 127
Figure 60: Samples 1 and 6 are heterozygous, sample three is normal homozygous ____ 128
Figure 61: G>C mutation is present as G ____ 129
Figure 62: Linkage analysis performed on the French family that was shown to be linked with yellow 132 segment, affected names are with red background, un affected are clear, markers used are on the left side of the chart, also the genetic distance is present. ____ 133
Figure 63: Co-runs of samples 17 and 25; sample 17; sample 25 for D3S2466 are shown, they contain the same genotype blue is the 371 green is the 379 bases. ____ 134
Figure 64: Co-runs of samples 17 and 25, sample 17, sample 25 for the rest of the markers respectively (D3S3727, D3S2432, D3S4535 and D3S1768) alleles 132, 160, 170 and 202 are in common. ____ 135
Figure 65: Co-runs of samples 14 and 25; sample 14; sample 25 for D3S2466 are shown, both have the same alleles. ____ 136
Figure 66: Co-runs of samples 14 and 25; sample 14; sample 25 for the rest of the markers respectively (D3S3727, D3S2432, D3S4535 and D3S1768), alleles 132, 152, 170 206 are in common. Allele 132 always segregates with the affected chromosome. ____ 137
Figure 67: Co-runs of samples 14 and 17; sample 14; sample 17 for the following markers respectively (D3S3727, D3S2432, D3S4535 and D3S1768), alleles 122; 132; 144; 170; 198 are in common. ____ 138
Figure 68: Samples GDO C10, GDO D10 and control 31 all shows the same pattern, control 25 shows different pattern in different filter absorption. ____ 155
Figure 69: Exon 56 is shown in this figure, the exon is highlighted with yellow colour, and red colour represents splice acceptor and donor sites respectively, our mutation is C/G rs363831. ____ 156
Figure 70: Sample 25F shows C/G substitution wile C31F and GDO C10 F are normal, this SNP was reported as rs363831 Glu [E]> Asp [D]. ____ 157
Figure 71: A/G substitution was seemed in the forward and reverse sequencing this SNP is rs363830 Gln [Q]> Gln [Q], next figure shows the normal comparison. ____ 158
Figure 72: The normal forward and reverse sequencing of the same position shown in the previous figure. ____ 159
Figure 73: RESCUE-ESE on the normal sequence of rs363831 ____ 162
Figure 74: RESCUE-ESE on the rs363831 ____ 162
Figure 75: After the Endo VII, three major patterns were seen: G5 with single band; G4 two very close bands and G3 with two bands more separated than G4. ____ 163
Figure 76: This is the sequence of exon 28, exon is highlighted with yellow colour, red colour represents splice acceptor and donor sites respectively, orange and red letters represents two VNTR repeats. ____ 164

Figure 77: Deletion of (ATTT) sequence in the above figure resulted in the appearance of CTTTA twice, the figure below shows one CTTTA, and the arrows describe the direction of sequencing. 165
Figure 78: This is the same sample but with forward and reverse sequencing results, the above sequencing shows one CTTTA because deletion of (GTTAT) occurred on the 5' region of the CTTTA sequence. In the lower diagram, a reverse complement of the same sequence is represented. Because the polymerase amplified two different VNTR alleles (one with deletion of GTTAT sequence) I have seen two sequences with CTTAT, arrows represent the Taq polymerase movement. 166
Figure 79: This figure contains the same reactions as the previous figure but this time with homozygous VNTR. The arrows describe the direction of sequencing. 167
Figure 80: After applying the sequence in the N N splice programme, I have scores of the donor and acceptor sites. No other sites were created in any case. 168
Figure 81: Tropoelastin alignment: 173

## Index of Tables:

Table 1: Some differences between sporadic and familial SAH are summarised in the following table [61 69]: 27
Table 2. Risk conditions for aortic dissection 29
Table 3: Location and heterozygosity of STRs used in linkage analysis, D3S3727 marker was used to search for an association of *TGFβRII* and sporadic SAH [267,268]. 58
Table 4: Number of families collected for this analysis was 16, relevant family medical history is provided. 71
Table 5: Comparison of Arlequine and Phase results on the numbers of haplotypes present in SAH and control samples. 84
Table 6: Comparison of Arlequine and Phase expected results on the numbers of haplotypes present in SAH and control samples. 84
Table 7: P value and Chi-square results of both Arlequine and Phase 85
Table 8: Results of the ESEfinder on the wildtype sequence of the elastin gene 93
Table 9: Results of the ESEfinder on the mutant sequence of the elastin gene 93
Table 10: Association of *Elastin* (rs2071307) with blood pressure, pulse pressure and stroke 97
Table 11: Analysis of sporadic SAH vs. control samples on the exonic SNP using dominant model. 97
Table 12: Analysis of sporadic SAH vs. control samples on the exonic SNP using recessive model 98
Table 13: Analysis of sporadic SAH vs. control samples on the exonic SNP using additivity model 98
Table 14: Analysis of sporadic SAH vs. control samples of BWHHS on the exonic SNP using dominant model 98
Table 15: Summarise GeneScan results depending on the GT repeats chromosome counts 110
Table 16: Summarise GeneScan results depending on the count of genotypes. 111
Table 17: Results of allele count, p value and chi square are present. 112
Table 18: Looking at the expected homozygous numbers of allele 17 in sporadic SAH and control. 112
Table 19: Looking at the expected homozygous numbers of allele 17 in familial SAH and control. 113
Table 20: Sporadic SAH vs. Control 114
Table 21: Familial SAH vs. control 115
Table 22: Shows the results of positive detection of SNPs after performing DHPLC and DNA sequencing, description of the exon name, function (of the exon domain), if any documented mutation is present in literature, then the location of this mutation ( Exon; Intron or 5'UTR ). Any amino acid substitution is described. If the mutations in the intronic region can affect or induce alternative splicing sites will be described, also included the frequency of each SNP (if it is known). Moreover, the expected observation and the observed mutations are included depending on the DHPLC results. 130
Table 23: Using TFSEARCH programme using threshold of 85.0 point (default) 131
Table 24: TRANSFAC Ali Baba results before and after SNP and it s possible creation or abolishing important sites for transcription control. 132
Table 25: Results of GeneScan genotyping of GDO samples sporadic SAH and control samples. 140
Table 26: Results of genotypes of total and sporadic SAH and control samples with there numbers. 141
Table 27: Results of allele count, p value and chi square are present. Taking 132 allele to be important in comparison to all non-132 143
Table 28: Taking 126 allele to be important in comparison to all non-126 144
Table 29: Taking 120 allele to be important in comparison to all non-120 (observed results came from table 25). 145

Table 30: Looking at the expected homozygous numbers of allele 132 (132:132 genotype) in sporadic SAH and control. 146
Table 31: Looking at the expected homozygous numbers of allele 126 (126:126 genotype) in sporadic SAH and control. 147
Table 32: Looking at the expected homozygous numbers of allele 120 (120:120 genotype) in sporadic SAH and control. 148
Table 33: Taking 132:132 genotype with non all 149
Table 34: Presence of allele 126 against all 150
Table 35: Presence of allele 120 against all 151
Table 36: Presence of allele 132 against all 152
Table 37: Presence of alleles group E against H in all 153
Table 38: Presence of alleles group D against G in all 154
Table 39: Results of ESE finder on the normal sequence of exon 56. 161
Table 40: Results of ESE finder on the mutant sequence of exon 56 161
Table 41: Description of the variations detected by DNA sequencing 169
Table 42: Mutations not found in our samples are summarised in this table, they are exonic, intronic and 5' UTR, calculation of the expected frequency was performed. 177
Table 43: Work performed in this project on three different genes, the elastin (*ELN*); the fibrillin-1 (*FBN1*) and the transforming growth factor beta-receptor II gene (*TGFβRII*). 186

# Abbreviations:

| | |
|---|---|
| AAA | Abdominal Aortic Aneurysm |
| ACE | Angiotensin converting enzyme |
| ADPKD | Autosomal Dominant Polycystic Kidney Disease |
| Ala (A) | Alanine |
| Arg (R) | Arginine |
| AP1/CRE | Activator Protein-1-cAMP Response Element |
| ARMS | Amplification Refractory Mutation System |
| Asn (N) | Asparagine |
| Asp (D) | Asparatate |
| bFGF | Basic Fibroblast Growth Factor |
| BWHHS | British Women's Heart and Health Study |
| CAT | Chloramphenicol Acetyl Transferase |
| Cys (C) | Cysteine |
| CFTR | Cystic Fibrosis Transmembrane conductance Regulator gene |
| ddF | Dideoxy Fingerprinting |
| DGGE | Denaturing Gradient Gel Electrophoresis |
| DHPLC | Denaturing High-Performance Liquid Chromatography |
| dNTPs | Dinucleotide Triphosphates |
| EDRF | Endothelium Derived Relaxing Factor |
| EDS | Ehlers-Danlos Syndrome |
| *ELN* | Elastin Gene |
| *ENG* | Endoglin Gene |
| FKBP | FK506-Binding Protein |
| Glu (E) | Glutamate |
| Gln (Q) | Glutamine |
| Gly (G) | Glycine |
| His (H) | Histidine |
| IGF1 | Insulin-Like Growth Factor I |
| ICA | Intracranial Aneurysm |
| IRM | Idiopathic Recurrent Miscarriage |
| Ile (I) | Isoleucine |
| Leu (L) | Leucine |
| Lys (K) | Lysine |
| MAGP | Microfibril-associated glycoprotein |
| MMPs | Matrix metalloproteinases |
| Melt-MADGE | Microplate-Array-Diagonal-Gel Electrophoresis |
| Met (m) | Methionine |
| MFS | Marfan syndrome |
| MFS2 | Marfan syndrome type 2 |
| OsO4 | Osmium Tetroxide |
| Phe (F) | Phenylalanine |
| Pro (P) | Proline |
| PTT | Protein Truncation Test |
| RER | Rough Endoplasmic Reticulum |
| SAH | Subarachnoid Haemorrhage |
| SDS-PAGE | Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis |
| Ser (S) | Serine |
| SRP | Signal Recognition Particle |
| SNP | Single Nucleotide Polymorphism |
| Spl | Transcription Factor 1 |
| SSCP | Single Strand Conformation Polymorphism |
| SSR | Simple Sequence Repeats |
| SVAS | Spravalvar Aortic Stenosis |
| TGF$\beta$ | Tumour Growth factor-Beta |
| *TGF$\beta$RII* | Tumour Growth factor-Beta Receptor II |
| TIMPs | The tissue inhibitors of metalloproteinases |
| Thr (T) | Threonine |
| Trp (W) | Tryptophan |
| Tyr (Y) | Tyrosine |
| Val (v) | Valine |

# Chapter One

## 1.0 Introduction

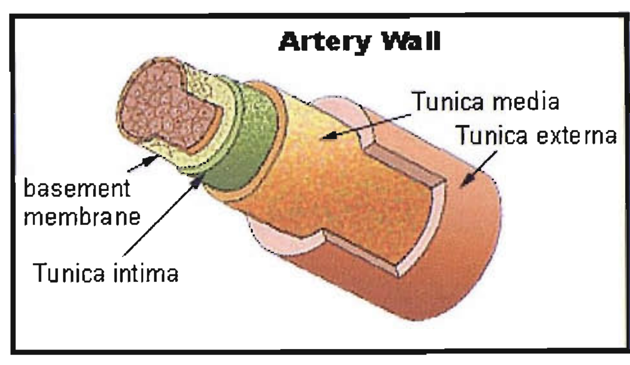
### 1.1 Overview

Arterial aneurysmal diseases (like subarachnoid haemorrhage and aortic aneurysms) may result in fatal or severe neurological consequences, most aneurysms remain asymptomatic until they rupture, understanding the genetic bases of these diseases, will help us to develop a diagnostic tool for detecting these diseases and/or preventing pregnancies that may carry the affected gene(s). Gene therapy is another area to be developed and researched.

Before going to the genetics of the arterial diseases, it is important to look at the anatomy of the arterial system in general.

The general anatomical structure of the arterial blood system is shown in Figure 1, three layers are present: tunica intima; tunica media and tunica externa.

Arterial blood vessels contains many cells (endothelial cells; smooth muscle cells; pericytes and fibroblast cells) attached and embedded in the extra cellular matrix (ECM). Composition and mechanical properties of the ECM are controlled by these cells[7].




A

1- Muscular artery (Thick Walled)
2- Tunica Adventitia (Connective Tissue)
3- Tunica Media (Smooth Muscle)
4- Tunica Intima (Internal Elastic Membrane, Endothelium, Little Connective Tissue)
5- Internal Elastic Membrane
6- External Elastic Membrane

B

**Figure 1:( Above) 3D with major structures of an artery adapted from [8]. (Lower) structure of artery cross section adapted from [9]**

There is a difference in the molecular composition of the cerebral arteries in comparison to the general structure of the arteries, cerebral arteries lack external elastic lamina. Figure 2 shows the molecular and cellular composition of the cerebral arteries.

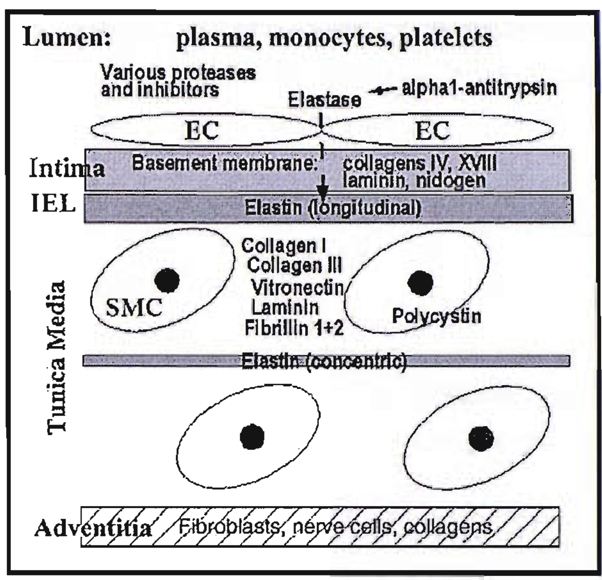



**Figure 2: Cerebral arteries are the same as other arteries in the body, but lacking external elastic lamina. There is the lumen (monocellular layer of endothelial cells (EC)); an intima (forming the basement membrane consists of collagen IV; XVIII; laminin; nidogen); internal elastic lamina (IEL) mainly elastin; tunica media containing smooth muscle cells and many ECM proteins like fibrillin-1 and elastin; finally adventitia ( fibroblasts nerve cells and collagen) adapted from[10].**

Elastin represents at least 20% dry weight of the arterial blood system, giving its elastomeric properties. Mutations in genes coding for structural proteins(ie. The elastin gene; the fibrillin gene and transforming growth factor Beta receptor II gene) in the ECM can lead to many diseases associated with blood vessels: supravalvular aortic stenosis (SVAS); Marfan syndrome (MFS) and Ehlers-Danlos syndrome type IV (EDS IV)[11]. Ruptured abdominal aortic aneurysms (AAA) has been shown to contains a decreased structural proteins like elastin[12]. The disturbance of the ECM in these diseases makes these genes good functional candidates for aneurysmal diseases.

Aneurysms may be associated with other diseases involved in the ECM like: Marfan syndrome (mutations in *FBN1*); Ehlers-Danlos syndrome Type IV (EDS IV) (mutations in collagen type III [13]) and autosomal dominant polycystic kidney disease (ADPKD) (mutations in PKD1 and 2 genes[14-16] that play a role in the interactions (between cell–cell and cell–matrix) in the ECM [17]).

Proteins in ECM when mutated may lead to aortic stiffness, this was supported by a paper that reported an increase risk of aortic aneurysms due to aortic stiffness with fibrillin-1 mutations [18], hence, mutations of genes coding for the ECM may play a role in the onset of arterial aneurysmal diseases.

Complex diseases (like arterial rupture) involve environmental and genetic factors (see Figure 5), genetic factors can be gene pleiotropy (i.e. Marfan syndrome (MFS), which involves three systems: skeletal, ocular and cardiovascular) where some intracranial aneurysms are associated with Marfan syndrome. Another factor is locus heterogeneity (like autosomal dominant polycystic kidney disease (ADPKD)), as a defect of different genes can result in the same disease).

Allelic heterogeneity (as in many mutations in fibrillin-1 gene may result in Marfan syndrome) is also a factor contributing to the complex diseases. On the other hand, different mutations that may affect single gene product may result in different diseases, (i.e. elastin mutations may result in: Supravalvular aortic stenosis[19], cutis laxa[20]. Elastin related like : Hurler disease (impaired elastic fiber assembly) [21] and Costello syndrome [22]).

Penetrance of SAH is not 100%, as in the onset of SAH in one of the identical twins , the second twin was having no ICA[23], another study suggested penetrance to be 70-99%[24], anticipation effect was noticed by Struycken et.al. to be 5-10% of familial SAH cases[25].

Gene-gene interactions can be involved in arterial aneurysms (see Figure 3 which contains the summary of the work performed in this project). Patients with Marfan syndrome type2 (MFS2) were shown to have a mutation in the transforming growth factor beta receptor II gene (*TGFβRII)*, this gene is involved in the expression of both elastin and fibrillin-1 proteins, mutations in *TGFβRII* in these patients showed ruptured arterial aneurysmal disease. Elastin and fibrillin-1 can be functional candidates for the onset of an arterial disease, as they participate in the ECM of the arteries, also evidence of elastin and fibrillin-1 interactions in the ECM may result in stability of the arterial ECM.

Many mutations of fibrillin-1 cause arterial aortic aneurysms. A few papers have suggested (from linkage studies) that the elastin gene may predispose aneurysms[1,2]. Elastin and fibrillin-1 proteins bind to each another and participate in the formation of the ECM[26]. Some mutations may leads to alteration of the arterial property and therefore may give rise to an aneurysmal disorder.

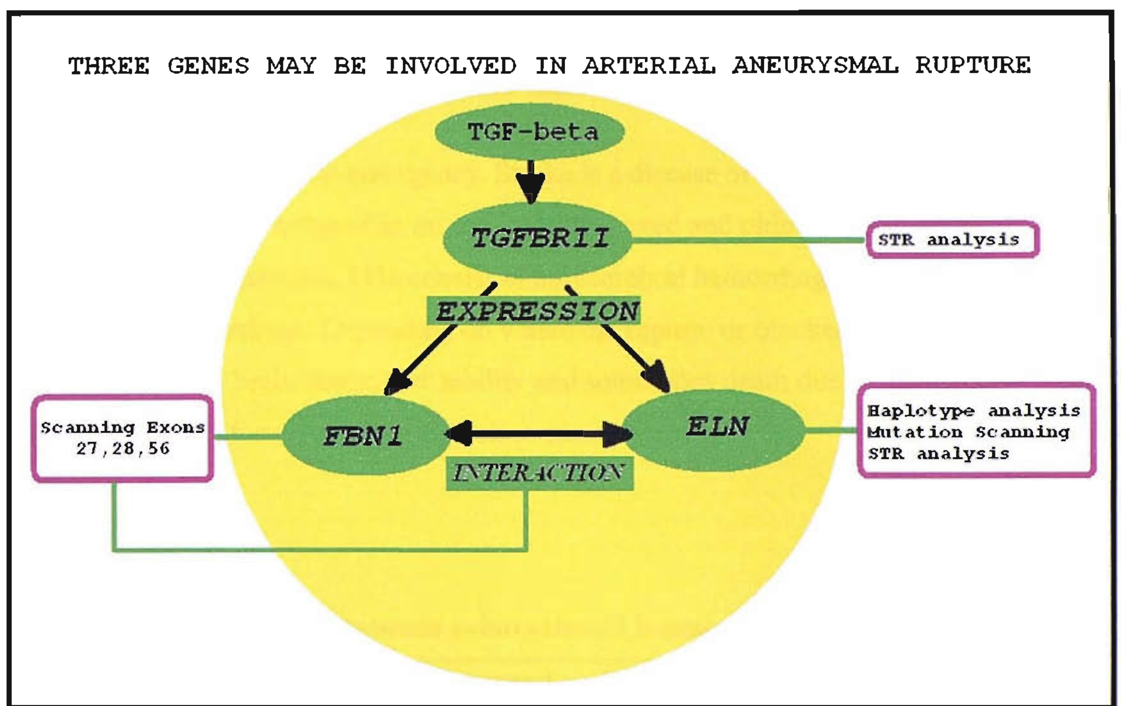



**Figure 3: The relationship between these three genes is shown in this figure; *TGFβRII* plays a role in the expression of both *FBN1* and *ELN*. The elastin gene is the functional candidate for the onset of ICA, also linkage analysis were supporting its role, fibrillin-1 gene mutations causing arterial aneurysm (in some cases SAH), both the product of these two genes shows interaction in the ECM; fibrillin-1 binds elastin via exons 27 and 28. The region coded by exon 56 binds fibrillin-1 protein. The general work performed in this study is shown in the pink squares.**

*TGFβRII* gene is another possible candidate gene for the onset of a ruptured arterial aneurysmal disease, some mutations were involved in MFS type II and SAH [27]. Hence, genetic heterogeneity for the onset of aneurysmal disease. Furthermore, *TGFβRII* gene has been shown to be involved in the gene expression cascade of both of the elastin and the fibrillin-1 genes (see Figure 3).

## 1.2 SAH

Intracranial aneurysms may result in subarachnoid haemorrhage (SAH), this disease is serious disorder with high morbidity and mortality; it occurs when an aneurysmal blood vessel ruptures or leaks blood and accumulates in the subarachnoid space [28]. The aetiology of spontaneous subarachnoid haemorrhage is summarised in **Box 1**. Subarachnoid haemorrhage contributes approximately to 10% of all strokes [29] . There are two main types of stroke: ischemic and hemorrhagic. Ischemic stroke is caused by blockage in an artery that supplies

blood to the brain, resulting in a deficiency in blood flow (ie. ischemia). Hemorrhagic stroke is caused by the bleeding of ruptured blood vessels (haemorrhage) in the brain.

A Stroke is an acute medical emergency. Stroke is a disease of the circulatory system caused by the blockage or rupturing of an artery. In middle aged and older women, about 70% of strokes are thromboembolic, 15% consist of intracerebral hemorrhage, and 10% of subarachnoid hemorrhage. Depending on where the rupture or blocked artery occurs, this can result in permanent brain damage, disability and sometimes death due to oxygen deprivation. SAH also accounts for up to 25% of cerebrovascular deaths [30].

**Box 1: Aetiology of spontaneous subarachnoid haemorrhage [31]**

- Intracranial aneurysms: degenerative 60-70 %
- Peri-mesencephalic haemorrhages 15-20%
- Arteriovenous malformations and associated aneurysms
- Other causes: Dural fistula; venous vascular abnormalities; spinal arteriovenus malformations; cerebral artery dissections; Moyamoya syndrome; vasculopathies; mycotic aneurysms; coagulopathies; neoplasia; pituitary apoplexy; drug abuse (amphetamine and cocaine)

Adapted from [31].

Concerning SAH, eighty five percent of saccular aneurysms occur at the bifurcation of the large anterior arteries mostly on the circle of Willis, see Figure 4 [28].

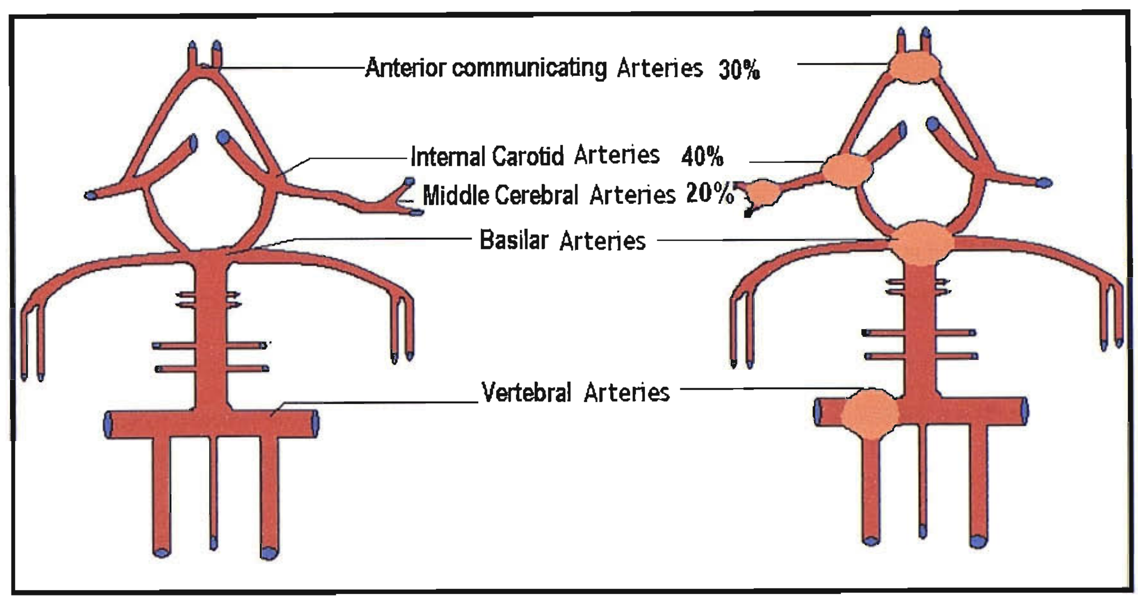



**Figure 4: Common places where saccular intracranial aneurysms take place:**

Most intracranial aneurysms occur in the circle of Willis. The internal carotid arteries display the highest percentage (40%), anterior communicating arteries (30%) and middle cerebral arteries (20%). Other places including basilar and vertebral arteries both displays 5-10 %[10].

Intracranial aneurysms (OMIM # *105800) can be divided into three categories: Giant intracranial aneurysms, which are defined as any aneurysms that have exceeded 25 mm in maximum diameter; large aneurysms with diameters 10-25 mm; and small aneurysms with diameters of <10 mm [32]. The frequency of multiple aneurysms is higher in women than in men and present in all age categories except in the over 80 years old. It is estimated that 12.4% in men and 20.2% in women have multiple aneurysms [33], and frequently occur at the contra-lateral site [28]. Moreover, it is difficult to predict which aneurysms are likely to rupture, but the available data suggests that most ruptured aneurysms are > 7 mm in diameter [28,34]. In another study it was shown that the size factor of the aneurysm can play a major role in rupture, for example, looking to aneurysmal sizes with <5mm; 5-15mm and >15mm they found that annual rupture was 0.4%, 3.3% and 9.9% respectively[35].

Giant aneurysms that occur at the bifurcation sites account for approximately 5% of aneurysmal cases, with a risk of rupture 6-10 % every year [28,34,35].

In some cases defects in internal elastic lamina are noticed in intracranial aneurysm[36]. Furthermore, in IA collagen type IV may be defected [37], other skin biopsies studies shown elastin and collagen disruption is present in IA cases[38].

## 1.3 Incidence of SAH

The incidence of SAH within the general population is 8-15 / 100 000 per annum [28,39,40]. Ruptured intracranial aneurysms (ICA) are the second cause of SAH after head trauma. A study has shown that 3-4% of the US population harboured aneurysms, which accounts for 8-10 million Americans of whom only 25-30 thousand haemorrhage per year [28]. Broadly, the estimated prevalence of ICA in the general population according to autopsy and angiographic studies is 0.2% to 8.9% [34,41-43]. From these studies, it appears that the majority of aneurysms do not rupture.

## 1.4 Risk Factors and Outcomes of SAH

Many factors participate in the onset of SAH like: coffee; drugs abuse such as cocaine and amphetamine[44]; hypertension; alcohol; smoking; pregnancy; low body mass index [45-48] and infections like Gram positive actinomycotic meningitis [49] or fungi like Aspergillus infection [50] See Figure 5

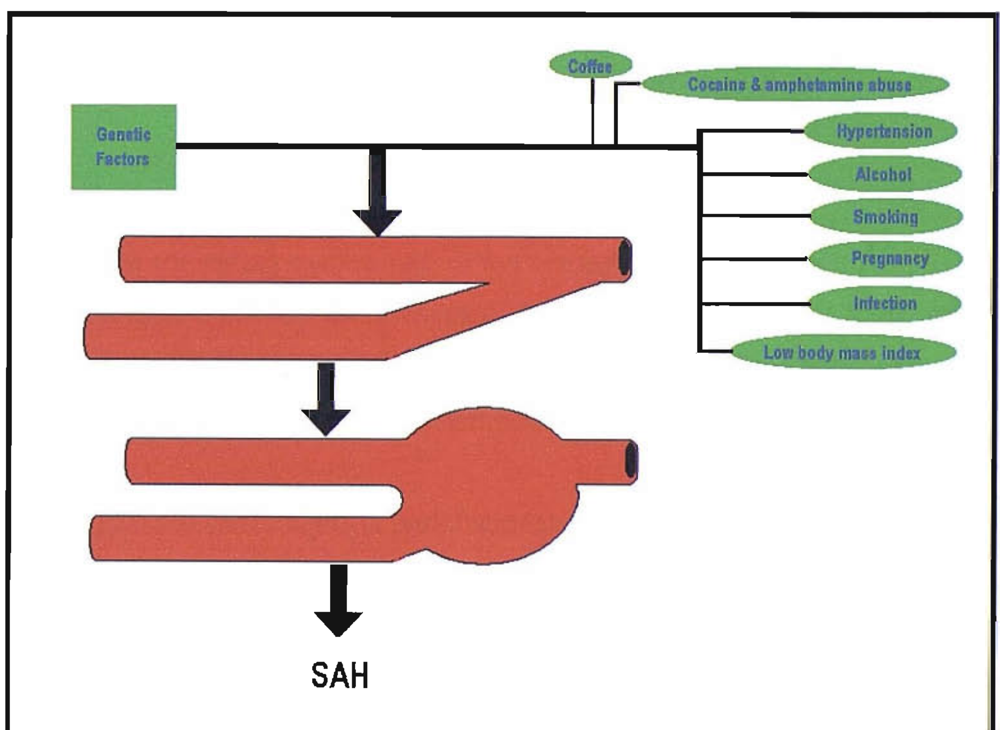



**Figure 5: Risk factors for SAH**

Factors such as hypertension, drug abuse, coffee, alcohol, smoking, infections, pregnancy and low body mass index can affect the integrity of blood vessels, but when they are combined with genetic factors (which can be monogenic, oligogenic or polygenic) there is a high possibility of developing aneurysms in the arterial system, when it is burst can lead to SAH. [46] [45] [41,47].

Other association of SAH with pituitary adenoma was noticed[51], in another case SAH was seen in patients with phaeochromocytoma[52], may be these diseases are associated with high blood pressure leading to ICAs.

Some investigators have observed negative associations between smoking or alcohol consumption and SAH. Furthermore, they have concluded that moderate to extreme physical exertion can triple the risk of SAH[53]. Another Japanese study suggests that hypertension and cigarette smoking seem to be independent risk factors for SAH. They added that high prevalence of hypertension in both sexes and the high prevalence of cigarette smoking in men in the general population might contribute to the high incidence of SAH[54].

In a third study, SAH can be prevented among the young and middle-aged class if behavioural risk factors such as smoking and cocaine use are avoided, and if medication (e.g. factor for hypertension) is improved [55]. An example that shows the importance of controlled medication in a study performed on identical hypertensive twins: one twin (with multiple aneurysms) had SAH; the second (healthy twin) had also a single small aneurysm. The reason behind it seems to be that the healthy twin was in good control of her blood pressure whereas the affected one was in poor medication control [56].

Angiotensin converting enzyme (ACE) was considered as a risk factor for SAH[57], another study suggested a possible involvement of insertions/deletions in the ACE gene in relation to SAH [58]. In the States a study was performed in response to the previous study, they have found no association between SAH and insertion /deletions of the ACE gene[59]

Tumour necrosis factor alpha (a proinflammatory cytokine) and Fas-associated death domain protein (proapoptotic protein) are increased in human aneurysms in general, and may have an effect on cerebral arteries by inducing inflammation and apoptosis in vascular and immune cells, thereby weakening vessel walls and leading to the formation of an aneurysm [60].

About fifty percent of people who have developed a SAH die before reaching hospital. In addition, after one month from the onset of the disease more than 50% of survivors are left with major neurological defects due to initial haemorrhage and cerebral vasospasm [28]. The inflammatory consequences of SAH are summarised in **Box2.**

**Box2: Inflammatory consequences of SAH [31].**

| Inflammatory consequences of SAH |
|---|
| • Red cell lyses and release of catalytic agents (e.g. oxyhaemoglobin) |
| • Free radical generation and lipid peroxidation |
| • Prostaglandin activation |
| • Complement activation |
| • Platelet activation and adhesion |
| • Release of vaso-constrictive agents: calcium ions; growth factors; IgG and complement; 5HT (5 Hydroxytryptamine); bilirubin; neuropeptides |
| • Reduced synthesis of endothelium dependent relaxation factor (Endothelium Derived Relaxing Factor) |

## 1.5 Familial SAH

First evidence was demonstrated through the incidence of SAH in identical twins. Data suggests that if one twin had SAH, then this would increase the risk for the second twin to develop the disease within 10 years [56,61]. The association of SAH in families with other genetic disorders also supports the theory that this disease can be a familial. An example is autosomal dominant polycystic kidney disease (ADPKD): the prevalence of asymptomatic intracranial aneurysms in ADPKD has been estimated to be about 8%, roughly five times higher than that found in the general population [62].

The risk of developing ICA is about 3-5 times higher in familial cases than the general population [63-65]. Furthermore, in first degree relatives the risk can increase to 5-8% [66,67], which support that familial SAH is present. In familial SAH, 77% of the patients are females, and the onset of SAH in familial cases is on average a decade earlier than in sporadic cases. The size of ruptured aneurysm is smaller in familial cases. Furthermore, the occurrence in the middle cerebral artery is relatively higher in familial SAH. Sporadic cases have more ruptured aneurysms in anterior communicating artery [41,68,69]. Some studies suggested that familial asymptomatic aneurysms are more likely to rupture in families having members with aneurysmal subarachnoid haemorrhage than in those

without [70]. Some differences between sporadic and familial SAH are summarised in Table 1:

**Table 1: Some differences between sporadic and familial SAH are summarised in the following table [61 69]:**

| | Identical Twins | Familial SAH | Sporadic SAH |
|---|---|---|---|
| **Average age of onset** | 41.9 years | 42.3 years | 51.4 years |
| **Size of burst aneurysm** | Relatively small | | Relatively large |
| **Female : male incidence** | Higher | | Lower |
| **Occurrence in the anterior communicating artery** | Relatively low | | Relatively high |
| **Occurrence in the middle cerebral artery** | Relatively high | | Relatively low |
| **Risk that ICA will rupture** | Higher | | Lower |
| **Incidence of SAH** | The risk in an affected relative may reach up to 5 fold higher in SAH than in general population. | | |

Variations are present in the prevalence of familial subarachnoid haemorrhage amongst different populations. From a retrospective study performed in Greenland, the rate of a positive family history of presumed SAH and ICA is high among Inuit's [SAH (23.1%) and of ICA (9.6%)], in comparison with Caucasian Danes [SAH (4.3%) and of ICA (1.6%)] [71]. Other studies have shown that the prevalence of familial intracranial aneurysms ranged from 4% to 10.5 % in different populations [42,70,72].

## 1.6 Anticipation in the Familial SAH

A study showed that some familial SAH cases might become manifested at an earlier age in some of the second generations, mean calculations difference between the two generations was about 21 years. Struycken *et al* (2002) suggested that the percentage of anticipation found in familial SAH patients is 5-10% [25]. In Netherlands, a study suggest that anticipation may occur as the age of onset in children (35.4 years) and the parents (55.2 years)[73].

## 1.7 Abdominal Aortic Aneurysm (AAA)

These aneurysms are characterised in their progressive dilatation that may lead to the rupture of the aortic wall [74], AAAs occurs when the aortic diameter exceeds 3.0 cm [75]. Mutations in genes like Type III collagen (responsible for Ehlers–Danlos type IV) [13] and Fibrillin-1 (responsible for Marfan's syndrome) are associated with several forms of heritable aneurysmal like diseases with vascular rupture [76], *TGFβRII* gene was shown to be associated with MFS type 2, in which AAA were seen[77]. Although AAA is rare disorder, it can have fatal consequences. The most common symptom of aortic dissection is pain ( AAA shows none or few symptoms until rupture) [78].

### 1.7.1 Incidence of Abdominal Aortic Aneurysm (AAA)

Incidence of AAA is about 5-30 cases / one million per year [78], in Brazil prevalence of AAA in the population over 50 years is 2-3%, and males over 60 years is 4.3-8.0 % [79].

Only 10-25 % of individuals with AAA cases reach to the hospital alive[80] (in another study 20-40 %), and after they are admitted to the hospital the mortality rate is 30-60 % [81]. In USA, each year, AAA deaths accounts for 9000 cases, the total number of death due to aortic aneurysms is 15000 [80]. AAA accounts for 1% of the total deaths in the western countries, the frequency of it in males is more than four times higher than in females [81].

### 1.7.2 Risk factors for Abdominal Aortic Aneurysm (AAA)

Table 2 describes some risk factors that can lead to aortic dissection[82], some other findings suggested that serum homocysteine elevation[83] and peripheral vascular occlusive disease[84] are other risk factors.

Chronic systemic hypertension can be one of the most common factors (present in 62–78%) of aortic dissection patients [78], many studies suggested that hypertension is a risk factor, others did not have positive association [81].

**Table 2. Risk conditions for aortic dissection**

| Risk Conditions for Aortic Dissection |
|---|
| **Long-standing arterial hypertension** |
| • Smoking, dyslipidemia, cocaine/crack |
| **Connective tissue disorders** |
| • Hereditary fibrillinopathies |
| • MFS |
| • EDS |
| • Hereditary vascular diseases |
| • Bicuspid aortic valve |
| • Coarctation |
| • Vascular inflammation |
| • Giant cell arteritis |
| • Takayasu arteritis |
| • Behcet's disease |
| • Syphilis |
| • Ormond's disease |
| **Deceleration trauma** |
| • Car accident |
| • Fall from height |
| **Iatrogenic factors** |
| • Catheter/instrument intervention |
| • Valvular/aortic surgery |
| • Side- or cross-clamping/aortotomy |
| • Graft anastomosis |
| • Patch aortoplasty |
| • Cannulation site |
| • Aortic wall fragility |

Adapted from [82]

## 1.8 Familial Abdominal Aortic Aneurysm (AAA)

Familial aortic aneurysms have genetic elements, a study was performed on 9 members in two generations (patients did not have symptoms of MFS) showed an autosomal dominant pattern of inheritance of aneurysmal disease in young age [4]. In another family, nine out of 10 were carrying one mutation (Gly1127Ser) in fibrillin-1 gene with no symptoms of MFS but with thoracic aortic aneurysm. That mutation resulted in a reduced fibrillin-1 deposition in cultured fibroblasts[85], and was in the EGF like domain (**Appendix D6**). Identification of a missense mutation D1155N in exon 27 of *FBN1* showed normal amount of fibrillin-1synthesis, but decreased deposition into the ECM, supporting the hypothesis of the origin of AAA is due to a decreased deposition of the fibrillin-1 protein in the ECM in non MFS patients[86].

Smoking family history and age seems to have a positive association with AAA [87], whereas female sex; diabetes; atherosclerotic disease and black race are negatively associated with AAA cases[88].

## 1.9 Genetics of Arterial Aneurysms

Genetic factors are becoming more recognised in the emergence of these multifactorial diseases, 5% of intracranial aneurysms are associated with heritable disorders like Ehlers-Danlos syndrome Type IV; neurofibromatosis type I; Marfan's syndrome and autosomal dominant polycystic kidney disease [89,90]. Some reports suggests an association of intracranial aneurysms and other aortic aneurysmal diseases[91-93].

It is likely that mutations in any gene involved in the structure and function of the artery can leads to AAA or ICA, as well as enzymes encoded by genes responsible for the post-translational alteration of structural proteins in the ECM; construction of the matrix and proteases involved in turnover of matrix components[87]. Penetrance in AAA is low and it inheritance is likely to be autosomal dominant[94]. It was thought that occurrence of both ICA and AAA is purely coincidence, a new paper provides evidence of a familial single gene defect that may lead to the development of either cerebral and aortic aneurysms[95], Some candidate genes that should be considered in the pathogenesis of SAH or rupture of AAA include:

1- **Elastin gene**: A genome wide linkage study and candidate-gene approach of ICA was performed by (Onda *et al* 2001) on 104 Japanese affected sib pairs (with LOD score of 3.22) revealed an association between SAH and chromosome 7q11 (with the best region close to the marker ***D7S2472***). Furthermore, Onda *et al* 2001 stated "The haplotype between the intron-20/intron-23 polymorphism of *ELN* is strongly associated with ICA (P=3.81x10^-6), and homozygous patients are at high risk (P=0.002), with an odds ratio of 4.39. These findings suggest that a genetic locus for ICA lies within or close to the *ELN* locus on chromosome 7" [1].

Another study (James *et. al.* 2003) confirmed that the chromosome 7q11 locus is a predisposing factor for intracranial aneurysm, this study was performed with 85 nuclear families each family with a minimum two affected siblings [2]. Chemical studies concerning

elastin content in ruptured abdominal aortic aneurysms is low when fibrillin-1 and collagen are high (this might be due to a complementary result of decreasing elastin cross-links in the aorta) [12].

2- **Fibrillin-1**: Which is associated with Marfan syndrome and may lead to AAA [87,96]. Also, it was reported that some mutations in this gene that are not associated with MFS cause AAA[85]. About 10 papers showed cases with an association of FBN1 with ICA see review [97].

3- ***TGFβRII*** gene: this gene is located on chromosome 3p25-p24.2 and was shown to cause MFS2 (OMIM 154705)[77], also in some families it is associated with ICA and SAH.

4- **Collagen type I alpha2 (*COL1A2*):** A recent paper suggested that the variations of *COL1A2* could be a genetic risk factor for ICA patients with family history. A 21 SNPs where genotyped from 260 ICA patients and 293 controls, differences in allelic and genotypic frequencies between controls and patients where significant (P= 0.00087) in one of the exonic SNPs, this SNP converts alanine (Ala) to proline at a.a. position 459 [98].

5- **Collagen III** : It is a major structural protein in the matrix, analysis performed on 50 patients showed that pro-collagenin type III has a 2% involvement in aortic aneurysms[99,100]. Ehlers-Danlos syndrome (EDS) type IV shows aneurysmal arterial rupture and it is associated with mutations in Collagen III[13].

6- **Chromosome 17cen; 19q13** and **Xp22**: Genome wide scan in 29 Japanese families gave evidence of linkage to these chromosomes regarding familial Intracranial aneurysm[101]. Another paper linked ICA to chromosome 19q13.3 in a Finnish population [102].

7- **1p34.3-p36.13** chromosome is associated with ICA in autosomal dominant way, this study was performed on a big family with six affected living individuals using Affymetrix 10K Gene Chips then microsatellite analysis of 23 kindred members, LOD score was 4.2, in this paper they suggested that penetrance is 70-99% [24].

8- **Fibrillin-2 gene**: This gene associated with congenital contractural arachnodactyly and aortic diseases that is related to Marfan's syndrome[103].

9- A locus present in chromosome **11q23.2-q24**, due to a linkage study of familial aortic aneurysms on 3 families[104].

10- **Matrix proteinases and tissue inhibitor metalloproteinases**: They are important in tissue repair and tissue remodelling, hyperactivity or over expression may affect the structural matrix and may contribute to the AAA pathology[105-108]. Matrix metalloproteins (MMPs) and how they may cause AAA (and possibly SAH) are shown in Figure 6. The tissue inhibitors of metalloproteinases (TIMPs) act to protect connective tissue from the destructive effects of the metalloproteinases, it was reported that TIMP are decreased in AAA tissue[109].

11- **Alpha-1-antichymotrypsin** (*SERPINA3*) gene present on chromosome 14q32.1 was suggested to be a risk factor for SAH[110].

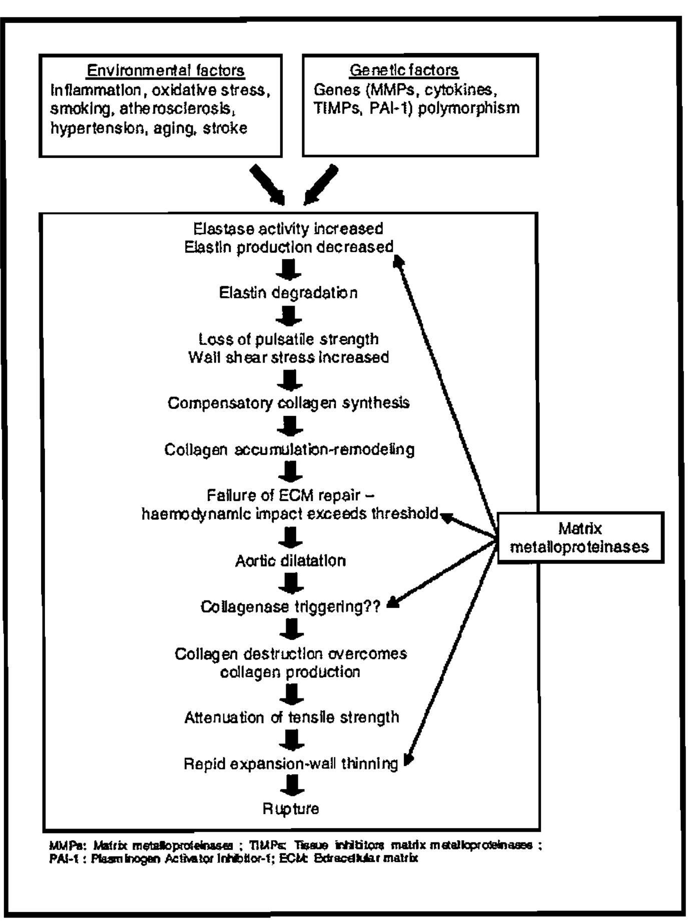



**Figure 6:Possible effects of MMP in the origin of aneurysms, adapted from [111].**

## 1.10 SNPs, Protein Structure, Function and Disease

The most common type of genetic variations in humans (~90%) are single nucleotide polymorphisms (SNPs) [112], and represent about one change per one thousand bases [113]. The most likely SNPs that affect gene function are those occurring both in the coding (called cSNPs) and regulatory regions [114 115 116].

SNPs are very easy to detect, since they have only two alleles. It is hoped that knowing the SNP genotype of individuals may help in identifying their susceptibility to certain diseases and subsequent choice of therapies [117]. Generally, SNPs can affect proteins in many ways, however, some mutations are considered neutral and will not cause any change to the protein structure. Other effects can be due to transcriptional or translational changes and some post-translational modifications [118]. SNPs can affect protein folding, stability, ligand binding, catalysis, post-translational modifications and protein interactions and the amount of protein expressed.

Some SNPs in the coding region that will not create a new a.a. may result in alternative splicing leading for example to drug resistance[119].

## 1.11 Microsatellites

Microsatellites or simple sequence repeats (SSR) are present in the coding and in the non-coding regions [120]. Distribution of SSR seems to be non-random in the genomic DNA, this can be attributed by its functional effect in chromatin organisation; recombination; DNA replication; cell cycle; regulation of gene activity and mismatch repair system [121].

GT microsatellite repeat distribution was significantly associated with recombination frequency in chromosome 22[122].

Advantageous SSR

SSR may provide adaptation advantages; for example, prostate cancer was found to have later time of onset in patients carrying longer CAG trinucleotide repeat. This trinucleotide is present in the amino terminal domain of the androgen receptor gene[123].

Another example of the advantageous effect of SSR expansion occurs in the bacteria, *Escherichia coli* strains lacking both thioredoxin reductase and glutathione reductase, have very poor growth due to this deficiency. A one trinucleotide expansion of (TCT) in the ahpC gene (encoding peroxiredoxin) will add an amino acid that resulted in the convergence of this product (from a peroxidase) to a disulfide reductase, as a result restoration of normal growth was accomplished [124].

SSRs can cause diseases

Most SSR diseases are associated with trinucleotide expansion, resulting predominantly in neurological diseases. Triplet expansion diseases can be due to coding region expansion like poly glutamines (Poly Q diseases). One example of this is Huntington's disease, which is thought to occur due to a gain of function mutation (protein aggregation). Consequently, the severity of this disease progresses with further addition of poly glutamines (the disease will manifest itself when the repeat expansion exceeds the threshold that is 35-40 CAG repeats) [125].

Diseases that occur due to non coding region are like Fragile X syndrome, whereby modifications in the 5'UTR region trigger DNA methylation that will inhibit transcription of the FMRP protein [126]. Other diseases can be caused by intronic expansion, for example in the Epidermal Growth Factor Receptor gene (EGFR) whereby a (CA) repeat in intron one was shown to enhance its transcription and play a critical role in breast carcinogenesis [127].

Some cases with cystic fibrosis were shown to be associated with (TG) and (T) expansion at the 3' end of exon 8. As a result, exon 9 was spliced out (alternative splicing) from the mRNA resulting in non-functional cystic fibrosis transmembrane conductance regulator (CFTR) gene product [128].

Position effect variation was shown to be a result of triplet expansion of in Friedreich's ataxia (GAA) and Myotonic dystrophy (CTG), these repeats trigger the formation of heterochromatin, hence affecting there transcription [129].

Inactivation mutation was shown to be responsible for human colorectal tumours due to an expansion of (A) nucleotide in (>80%) *TGFβRII* gene (Frameshift mutation) [130].

## 1.12 Tropoelastin Gene and Protein

In this study, the elastin gene was the first candidate for the onset of ICA. I will introduce this gene and talk about its product and formation in the coming sections.

### 1.12.1 Tropoelastin Gene (*ELN*; MIM# *130160)

A single copy of elastin gene is present in the haploid human genome, it is carried on the long arm of chromosome seven (7q11.1-21.1) [131-134], human elastin gene (*ELN*) has 34 exons [135] (whilst bovine elastin has 36 exons) [136], in general the exon: intron ratio is 1:20. The total genomic DNA of the elastin gene is about 45 kb [135]. Physical map of the *ELN* gene and its accession numbers are found in reference [137].

### 1.12.2 Exons and Introns

Structural analysis of the elastin gene in many species showed that most of the hydrophobic and cross-linking domains alternate [135,138-140]. Variations exist between different species, for example exons 34 and 35 present in the bovine elastin gene are absent in the human species [135]. In humans, the number of exons coding for hydrophobic domains and cross-linking domains is eighteen and fourteen respectively. As shown in Figure 7, exons coding for hydrophobic domains are [2, 3, 5, 7, 9, 11, 13, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33], and exon coding for cross-linking domains are [4, 6, 8, 10, 12, 15, 17, 19, 21, 23, 25, 27, 29, 31][141]. Most of the exons start with glycine (26 out of 34 exons) three exons starts with alanine; two exons start with valine. Proline; arginine (Arg) and methionine all occurs once in three different exons (Appendix E).

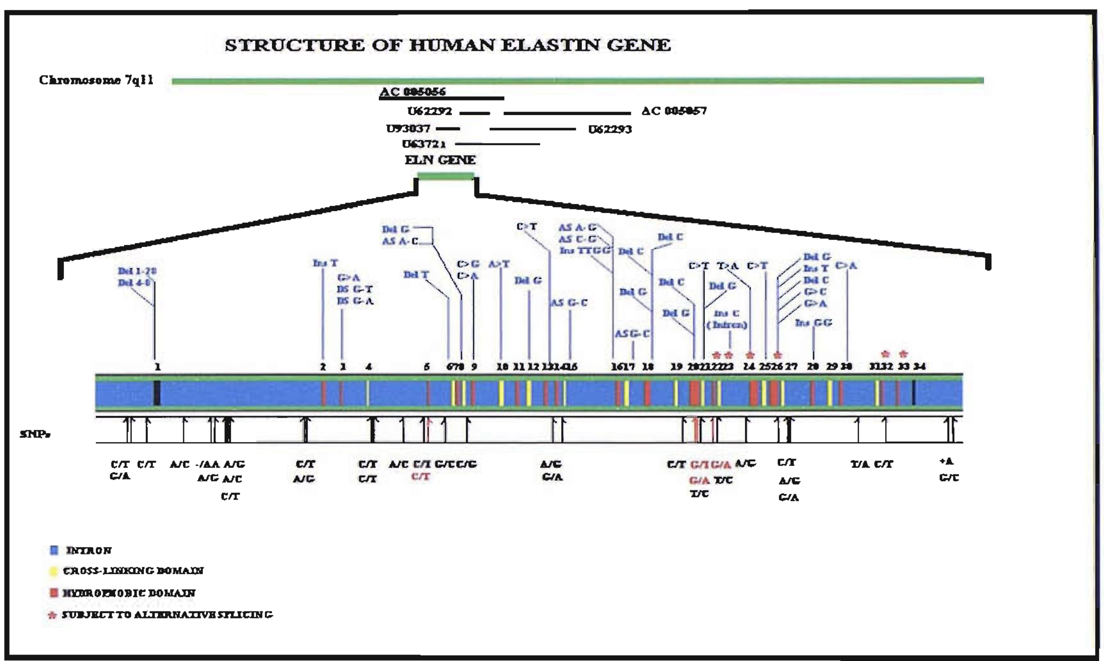



**Figure 7: The human elastin gene:**

General structure of the human elastin gene, some SNPs and other mutations are included, the elastin gene is carried on chromosome 7, cross linking domain is in yellow colour wile hydrophobic domain is in red colour, exons that are subjected to alternative splicing are with (*) .

### 1.12.2.1 *Alu* Repetitive Sequences

Genes that are known to have high *Alu* repeats, are prone to frequent deletions and duplications [126], since *Alu* repeats in the elastin gene is four times higher than other regions of the human genome [141,142], the possibility of instability in this gene through recombination events will be high in the general population [143]. The numbers of *Alu* repeats in elastin gene are 30 or more [1]. Supravalvular aortic stenosis (SVAS) is an example that results from 30 kb deletion within the elastin gene [133,144]. Duplication in exon 18 (1034-1057dup) within elastin gene is another example of SVAS [145].

### 1.12.2.2 Intron 18 GT Repetitive Sequence

This repetitive sequence is present in intron 18, data bases information showed to contain 19 or 20 GT repeats (accession numbers AC005056 and U93037 respectively), a paper described this repeat to have 8 different alleles [146]. The start position of this repeat is

IVS18+13; consequently, this repeat may have an affect on splicing. In addition, it may be involved in anticipation in case of GT expansion.

#### 1.12.2.3 Trinucleotides Repetitive Sequences

Anticipation paper gave me an idea to search for trinucleotides within elastin gene. Actions were taken to perform blasting against the elastin gene with different sequences of trinucleotides that are known to be involved in triplet repeat expansion [126]

### 1.12.3 Splicing and Alternative Splicing

Exon/intron borders in the elastin gene always have phase 1 splicing, which allows alternative splicing to occur without disturbing the reading frame. In humans, some differing splicing products of the elastin gene are shown in Figure 8 [147], See **Appendix D2** for more information about some of the protein products.

Concerning humans, alternative splicing of elastin gene can create 17 different transcripts they are the following: a(2937 bp); b(2850 bp); c(3261 bp); d(2832 bp); e(2794bp); f(3371bp); g(3368bp); h(3074bp); i(3006bp); j(2845bp); l(2026bp); m(1302bp); n(914bp); q(604bp); r(260bp); s(874bp) and t(172bp) [148] [132,149] [141] [133,139,150].

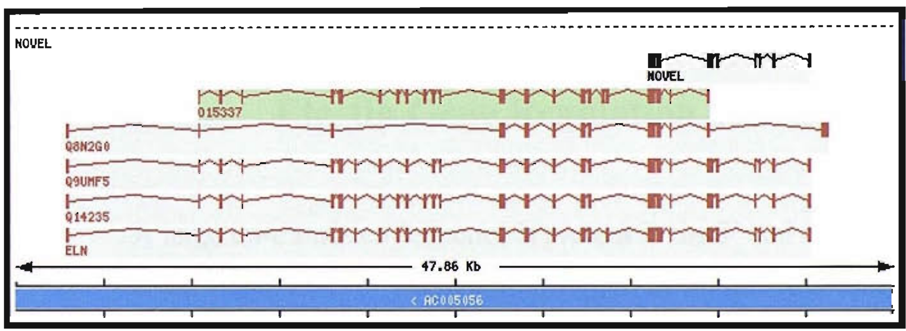



**Figure 8: Some different splicing products of the elastin gene are present.**

### 1.12.4 Elastin Gene Regulation in 5' Region

The basic promoter element is within one to -128 bases of the 5' region of the elastin gene, non-functional CAAT promoter sequence is present (-550), no TATA box is present. The 5' region is rich in CG and multiple binding sites for Sp1 and Sp2 transcription factors. Eight different transcription start regions have been identified in the human elastin gene [132,135] see Figure 9:

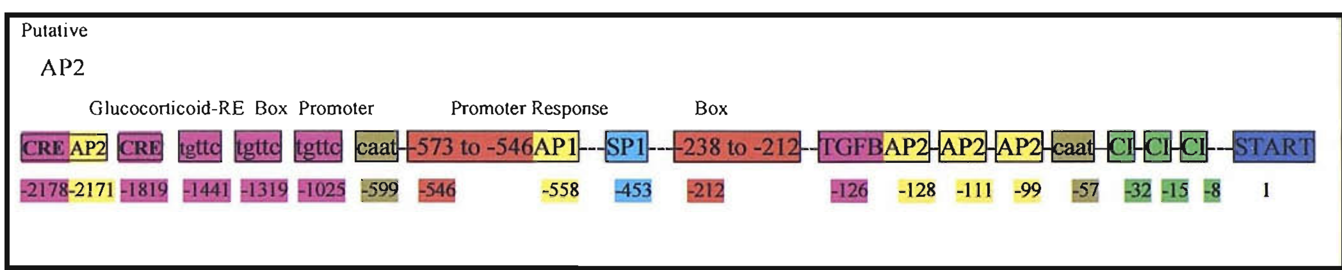



**Figure 9: Some control elements that are present in the 5' region of the elastin gene, it contains several SP1(only one is shown) and AP2 as well as putative glucocorticoid, cAMP, and TPA responsive elements, but no consensus TATA box or functional CAAT box (for more details see Appendex E)**

Primer extension and S1 mapping of the elastin mRNA indicated that transcription is initiated at multiple sites, it seems that transcription of the elastin gene is complex and takes place at several levels[132].

### 1.12.5 Elastin Gene Regulation at 3' Region

A large un-translated sequence is present in the 3' end of the elastin gene. In humans, it is estimated to be 1181 nucleotides long (from the first nucleotide beyond the stop codon to the polyadenelation signal) [142], as a result possible regulation elements may be present in this region [133,139].

### 1.12.6 Transcription Regulation

In early stages of postnatal development, elastin is the major synthetic product of arterial tissues, however, synthesis of elastin generally peaks early during arterial growth, with further development, it decreases rapidly and essentially ceases in adult arterial tissue. Gene expression of tropoelastin is a complex process and it is suggested to be regulated by different pathways (transcriptional and post transcriptional levels) depending on the tissue type [151,152].

#### 1.12.6.1 Decreased Production of the Elastin Protein

It has been reported that Interleukin-1beta (IL1β) reduces the rate of elastin gene transcription by 74% (in neonatal rat lung fibroblasts). This is performed through activation of the nuclear localisation of NF-kB that subsequently interacts with Sp1 to down regulate elastin transcription and expression of the myofibroblast phenotype. IL-1 β does not have an affect on the level of Sp1 but it does induce translocation of the p65 subunit of NF-kappa B. Over expression of (NF-K subunit p65) decrease elastin promoter activity and markedly reduces elastin mRNA [153].

Elastin is also down regulated by basic fibroblast growth factor in aortic smooth muscle cells. This is accomplished by activator protein-1-cAMP response element (AP1/CRE) (-564 to -558-bp) within the elastin gene promoter that mediates the basic fibroblast growth factor (bFGF)-dependent down regulation of elastin gene transcription in SMC [154].

Epidermal growth factor (EGF); angiotensin II and endothelin-1 increase cellular proliferation and decrease elastin expression, SMC proliferation will inhibit elastin expression[155].

#### 1.12.6.2 Increased Production of Elastin Protein

An in-vitro example of the elastin gene regulation at the transcriptional level is insulin-like growth factor I (IGF1). IGF1 disrupts the binding of promoter-selective transcription

factor I (Sp1) (that acts as a negative transcription regulator for tropoelastin), consequently increases tropoelastin mRNA in aortic smooth muscle cells. In lung fibroblast cells, IGF1 has no effect. On the other hand, a cytokine receptor called transforming growth factor-beta receptor type III (*TGFβRIII*) (a product of endoglin gene "*ENG*") was shown to increase tropoelastin steady-state mRNA in lung fibroblasts but not in smooth muscle cells [156-158]. A recent study suggested that *ENG* has no association with ICA. They also found that there is no linkage between the *ENG* locus and ICA in Japanese population, indicating that *ENG* may not be a major susceptibility gene for ICA in that population[159]

Expression of elastin in chick embryo cells (aorta but not in tendon cells) was shown to increase from 2-4 times after addition of TGF-beta[160].

### 1.12.7 Post-transcriptional Regulation of the Elastin Gene

The steady state mRNA levels are determined by rates of transcription and rates of mRNA decay. A study performed on freshly isolated chick aortic tissue and from cultured aortic tissue showed that decreased synthesis of elastin was partially due to the instability of mRNA [151]. This appears to be due to a secondary structure in the mid-region of the 3' UTR of chicken elastin gene that functions as a target for protein binding, decreased protein binding was observed with decreased mRNA stability [161]. Further investigation of this region suggested that this region contains a cis-acting element called the G3A motif. This region is conserved between species, and is a GA-rich sequence. Proteins (trans acting elements) that bind to this motif coexist with mRNA production, similar sequences are present in human acid phosphatase 5 and human pre-procathepsin P[162].

Decreasing tropoelastin expression can be controlled by post-transcriptional regulation, this is performed via rapid decay of tropoelastin mRNA, even though the steady state mRNA production is the same, this was suggested by many studies [151,163-165].

Another involvement in post-transcriptional regulation of elastin mRNA is exon 30. This is due to a 10 nucleotide sequence in mRNA near the 5' end [166], a frame shift mutation that was reported in this exon results in Cutis laxa disease [167]. Also elastin is down regulated (post transcriptional) by exposure to 1,25-dihydroxyvitamin D3[168].

## 1.12.8 Tropoelastin Protein

Tropoelastin is the soluble precursor of elastin; it was first isolated from a copper-deficient porcine aorta [169]. The general structure of tropoelastin showed to contain the following (also see **Appendix D1** for the complete amino acid sequence and sites for cross-linking domains):

1- Hydrophobic domains dominated by the aliphatic residues: alanine (Ala); valine (Val); leucine (Leu); isoleucine (Ile); proline (Pro) and glycine (Gly), altogether forming a repetitive of di-; tri-; tetra-; penta-; hexa- and nona-peptides. However, the general binding blocks of the tropoelastin is Gly; X; Pro; X; Gly; Gly; X and Pro; Gly; X; were X can be one of the following: Val; Ala; Gly; Leu; or Ile. This domain seems to be responsible for the elastic properties of elastin protein.

2- Cross-linking domains: contain lysyl residues within proline or alanine regions, lysine amino acids are oxidised and deaminated by lysyl oxidase to form cross-linking domains. Generally, this domain is rich in lysine (Lys) and alanine (Ala) amino acids. [170,171].

3- Signal sequence (N-terminal leader sequence): it is encoded by exon one and is highly hydrophobic formed by 26 amino acids with the sequence: Met; Ala; Gly; Leu; threonine (Thr); Ala; Ala; Ala; Pro; Arg; Pro; Gly; Val; Leu; Leu; Leu; Leu; Leu; serine (Ser); Ile; Leu; histidine (His); Pro; Ser; Arg; Pro.

4- C-terminal (encoded by exon 34) is hydrophilic and highly basic and it is highly conserved between species (more than 70 %), it contains the only two cysteines amino acids in the tropoelastin protein and terminates with a positive charge sequence. These two cysteine residues form an intra-chain disulfide bond, giving rise to a positively charged pocket [133,172 133,173 142].

5- The human hydrophilic sequence originating from exon 26A is rarely expressed in the human body, this sequence is serine rich (about 57% serine amino acids (a.a.) are present in this exon), also contains numerous charged amino acids like glutamate (Glu); asparatate (Asp) and arginine (Arg) [174].

Exon 26A domain contains the only basic histidine amino acid [175]. It has been reported that tropoelastin lacking exon 26 A is less efficient as a substrate for lysyl oxidase, though

hydrophilicity may not be the main reason behind this [133,176]. Others have suggested that exon 26A may increase in aged or diseased elastic fibres, and as a consequence can be a marker of damage [133,139].

### 1.12.9 Elastogenesis

Various types of cells like smooth muscle cells; microvascular; endothelial cells; fibroblasts and chondroblast do synthesise elastin. In chick embryo, the total synthesis of tropoelastin was varied between studies, the time range was 30 minutes [177] to around 85 minutes [178]. Elastin expression was seen to be expressed minimally at G2/M phase and maximally at the G0 phase of the cell cycle, hence, proliferation state is associated with low elastin production[179]. The first step in tropoelastin protein formation is translation, which takes place on the surface of rough endoplasmic reticulum (RER):

Elastogenesis is complex process, starting in the nucleus (Figure 10); translation of mRNA produces hydrophobic N-terminal peptide that contains the signal sequence (leader sequence). After the start of translation, the leader sequence is recognised by signal recognition particle (SRP), that will bind to both the ribosomes and the signal sequence and then direct them to a SRP receptor found on the RER surface. Tropoelastin peptide enters the lumen of the RER followed by cleavage of the signal peptide by a specialised signal peptidase (Figure 10.2) [126,180].

Tropoelastin then binds to a 67 kDa protein (that was demonstrated to be an inactive alternatively spliced variant of beta-galactosidase) and act as a chaperone [181,182], which will prevent premature intracellular aggregation, also this protein have another binding site for lectin [181].

#### 1.12.9.1 Modification and degradation of tropoelastin

Very little modification can occur to tropoelastin in the cell, the example modification, is the hydroxylation of some prolyl residues of the tropoelastin protein [183]. Inhibitory action against secretion can result in intracellular degradation of tropoelastin [184,185], this is

evidence of a possible way of quality control degradation at the RER level, degradation is performed by cysteine proteases [184].

Tropoelastin in the RER may be subjected to proper folding of proline rich domains by a protein that has prolyl *cis-trans* isomerase activity, called the FK506-binding protein (FKBP) [133,186].

#### 1.12.9.2 Secretion of Tropoelastin

Secretion of tropoelastin is carried out via secretary vesicles [177,187], that are rapidly translocated to the plasma membrane (Figure 10.4) [177]. Tropoelastin may accumulate in the RER and Golgi apparatus without modification (like glycosylation). Then two proteins (55 and 61 kDa) binds to the 67kDa-tropoelastin to form receptor complex protein (Figure 11) [188].

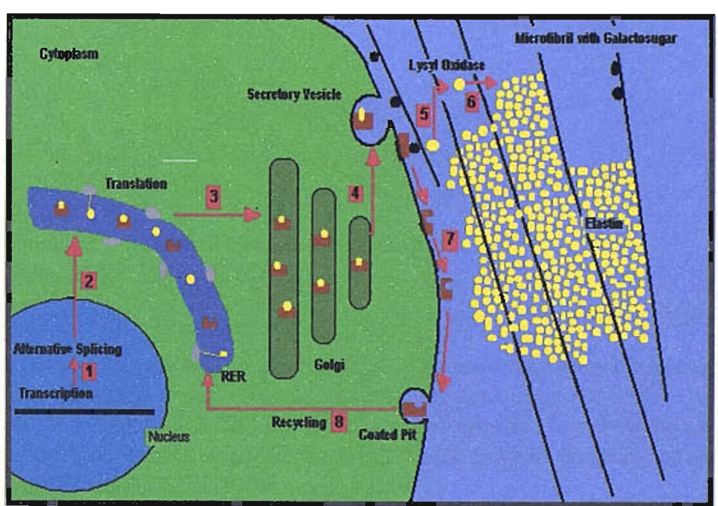



**Figure 10: Secretion of tropoelastin:**

1- Alternative splicing may take place in the nucleus. 2- Translation of mRNA occurs on the surface of rough endoplasmic reticulum (RER), tropoelastin polypeptide is released into the lumen with the cleavage of the signal sequence, tropoelastin is immediately captured by molecular chaperone (this will prevent aggregation of tropoelastin). 3- Elastin goes through the golgi to be packed in secretary vesicle. 4- A complex receptor protein is formed and exposing of the receptor complex to the extra cellular space. 5- Chaperone interact with galactosugar (part of microfibril), tropoelastin-chaperone complex is dissociated. 6- Free tropoelastin is appropriately aligned on the growing elastin, cross-linking reaction is taking place by lysyl oxidase enzyme, and elastin is growing within the microfibrillar scaffold. 7- Endocytosis of chaperone via coated pit. 8- Recycling of chaperone takes place in the RER.

### 1.12.9.3 Post secretion of Tropoelastin

When tropoelastin receptor complex is exposed to the extracellular space, interaction between galactosugars (i.e. lactose or galactose) or glycosaminoglycans (i.e. dermatan sulphate or chondroitin sulphate) occur, see (Figure 11).

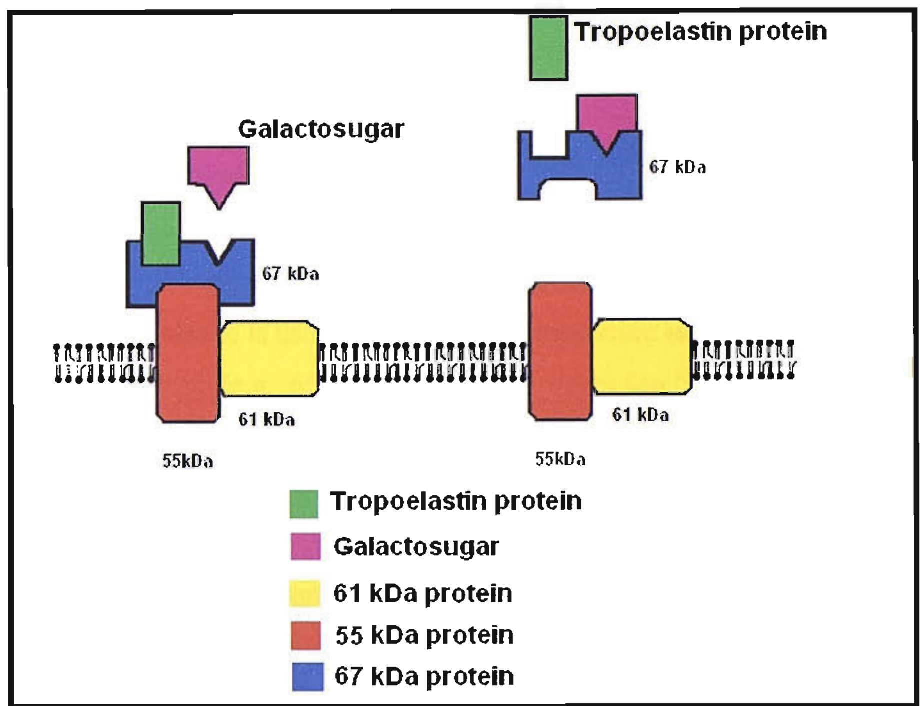



**Figure 11: Tropoelastin release from 67kDa protein. Elastin binding protein complex is composed of 67-kDa subunit and two other membranal proteins of 61 and 55 kDa. The 67kDa protein binds both tropoelastin and galactosugars via two different binding sites, and then galactosugars binds to the 67kDa tropoelastin-protein complex are no longer attached together (adapted from[188]).**

As a result, reduction of the affinity for tropoelastin-67kDa protein complex and anchoring proteins are observed, concurrently the dissociation of tropoelastin-67kDa complex is accomplished, this is due to the proposed interaction between the lectin binding site of the tropoelastin protein complex with highly glycosylated microfibrils [189-191].

Since the production of tropoelastin exceeds the production of the 67kDa protein, it was proposed to be internalised and transferred into endoplasmic reticulum (ER) to bind new formed tropoelastin [133], another possible role for 67kDa protein is to prevent the soluble tropoelastin from extracellular degradation and contributing to the stability of it[182].

Tropoelastin is soluble in cold solution, but at physiological temperature free tropoelastin will deposit on the microfibril via the non covalent interactions between the N-terminal part of the microfibrillar-associated glycoprotein (acidic) with the C-terminal end of tropoelastin that contains two cysteine residues forming an intra-chain disulfide bond, giving rise to a positively charged pocket [133,133,172,173].

## 1.12.10 Elastin Protein Formation

The first step in elastin formation is achieved through self-aggregation of tropoelastin to form elastic fibres in a process called coacervation. Hydrophobic sequences of tropoelastin have been suggested to play an important role in self-assembly, and coacervation of tropoelastin is usually induced by temperature increase, while in many proteins high temperature (e.g. 65° C) can causes denaturation [171,192-194].

Elastin is one of the most hydrophobic proteins known, it is very insoluble protein and contains extensive cross-links at lysine residues, these cross-links are formed by lysyl oxidase enzyme (copper dependent enzyme) that result in allysine residue [169], consequently, condensation reaction is taken place to form cross links [183].

Studies on elastin evidenced that it contains a two-phase model consisting of dynamic and hydrophobic domains in water. The hydrophobic domain of elastin can be described as a compact amorphous structure containing distorted beta-strands, fluctuating turns, buried hydrophobic residues, and main-chain polar atoms that form hydrogen bonds with water. Water plays an important role in determining the conformational behaviour of elastin, making it extremely dynamic in its relaxed state, also providing an important source of elasticity [195].

### 1.12.10.1 Cross-Linking of Tropoelastin

The deamination and oxidation of lysine takes place to give allysine, followed by cross-linking of the allysine molecules of different tropoelastin, which will leads to insolubilisation of tropoelastin [183]. These covalent bonds can be bi-(lysinonorleucine), tri-

(merodesmosine) or tetra-(desmosine and isodesmosine) [170], with three allysines and one lysine contributing to each desmosine or isodesmosine [196].

### 1.12.11 Some Functions of Tropoelastin and Elastin

Elastin is found in many sites of the human body including major vascular vessels and the aorta (28-50%, dry weight), lungs (3-7%) elastic ligaments (50%), tendons (4%) and skin (2-3%) [197,198]; and is responsible for the elastic properties of connective tissue.

Tropoelastin and some soluble forms of elastin can show biological activities like chemotactic activities of monocytes, fibroblast and some tumour cells. Elastin degradation generates products that act on calcium ion channels of several cells. Hydrophobic elastin peptides can work as vaso-relaxant thereby reducing vascular tone and elastin fibres are responsible for the rheological properties of blood vessels [199].

Elastin endows the connective tissues with resilience and permitting deformability and passive recoil without energy input, these properties are important for maintaining artery function, which undergoes repeated cycles of extension and recoil [198].

The elastin peptide sequence VPGVG can enhance SMC proliferation, this may result in the reduction of elastin expression and stabilising the arterial structure[155].

# 1.13 Fibrillin 1(*FBN1*; MIM# 134797)

## 1.13.1 General

Fibrillin-1 protein was discovered in 1986 [200], it is one of the major components of the microfibrils, present as isolated aggregates or closely associated with elastin. It is thought that fibrillin-1 plays an important role in tropoelastin deposition and elastic fibre formation, as well as anchoring function in some tissues.

Many fibrillin-1 mutations are associated with Marfan's syndrome[201] that may result in AAA. Fibrillin-1 is synthesised by many cells like fibroblasts; smooth muscle cells[202]; osteoblasts and osteoblasts like cells[203].

One case report has found an association between AAA and SAH[204], another evidence showing familial aggregation of both aortic aneurysms and cerebral aneurysms, may result in a common genetic background of both cases[95].

## 1.13.2 Structure of *FBN1* Gene

Fibrillin 1(*FBN1*) is a large gene (234912 bases), consists of 65 exons (Figure 12) with a transcript size around 10kb. Of the mRNA 9663 nt., the open reading frame of 8613 nt, the 5' flanking of 134 nt and the 3' flanking of 916 nt[205].

*FBN1* was assigned to chromosome 15q15-21 by in situ hybridisation by using a 1.6 kb of cDNA probe belonging to a PCR product of cDNA of the fibrillin-1 gene [206]. Very high conservation is present in the cDNA of *FBN1* gene between human and pig[207]

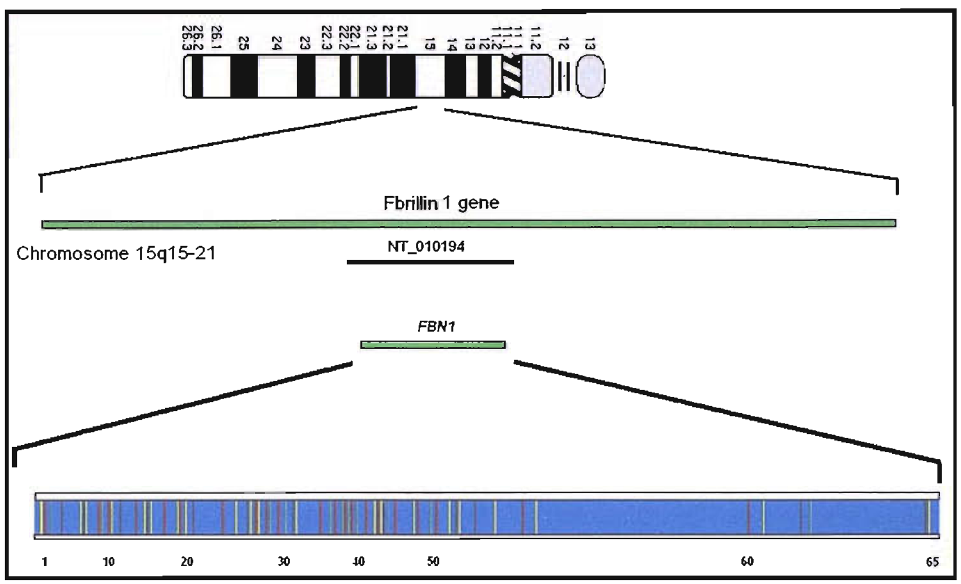



**Figure 12:** ***FBN1*** **gene is present on chromosome 15q15-21, contains 65 exons, and encodes for a large protein (2871 amino acids) this entire gene is present in NT_010194 genomic contig.**

### 1.13.3 *FBN1* Mutations and Diseases

One of the diseases that are associated with mutations in *FBN1* is Marfan's syndrome (MFS); this is a pleiotropic genetic disorder of connective tissue involving skeletal, ocular and cardiovascular abnormalities. Moreover, in vast majority of MFS is considered to be an autosomal dominant with an incidence of 1/(7000-10000) [82,208] and prevalence of 4–30 per 100,000[209-212].

More than 600 different mutations in *FBN1* gene may causes MFS disorder [213,214], most of them are missense mutation (about 75%), twenty percent of MFS mutations accounts for frameshift mutation and 12% were found to affect splice site[215]. Missense mutations occur mostly on the EGF like domains (that are predicted to disrupt Ca++ binding and/or secondary structure of the fibrillin-1 protein) [216], new mutations occurs in 25-30% of cases [211,212].

Most MFS cases are autosomal dominant inheritance[217], germ line mosaicism (which are rare) where confirmed by some molecular studies [218,219].

Severe lethal forms of MFS are associated with skipping exons 24-32 , while mutations in exons 59-65 are associated with mild phenotypes which are characterised by the lack of significant aortic pathology[220].

A phenotype genotype study (on *FBN1* gene) shown that about 34 % of mutations can not be found in MFS patients [221].

Marfanoid-craniosynostosis syndrome (Shprintzen-Goldberg) is another diseases that may be associated with *FBN1* mutations in two described cases[222].

MASS syndrome[223] (Marfan-like syndrome with mild dilatation of the aortic root, MASS letters coming from: mitral valve; aorta; skeleton and skin) is also associated with *FBN1* frameshift mutations leading to premature termination this mutation was described in 1993 and was considered to be MFS mutation [224].

Life expectancy of MFS increased about 25% since 1972, this may due to the improved life expectancy of the whole population or improvements of surgical benefits or discovering of more milder cases due to the improvement of the diagnostics or due to medical therapy like (beta blockers)[225].

## 1.13.4 Fibrillin-1 protein:

Profibrillin-1 (2871 amino acid protein) is a cysteine rich monomer glycoprotein (350kDa.) [226]. Transmission electron microscopy studies indicated a striated tubular appearance with a diameter of 8-12nm[227].

Fibrillin-1 can be divided into five structurally distinct domains namely A to E, and signal peptide sequence for the extracellular recreation [205]. These domains are 1-Epidermal Growth Factor type 2 (EGF-2); 2-Epidermal Growth Factor like (EGF-like), 3-Epidermal Growth Factor like calcium binding EGF-Ca. All of these three domains contain six cysteine residues that are involved in disulphide bonds. Ca dependent domain requires Ca+2 ion for its biological function; 4-Aspartic acid and asparagine hydroxylation site (Asx_hydroxyl_S) which is an EGF like domain that is hydroxylated on aspartic acid and asparagine amino acids; 5- Matrix fibril-associated domains (TB domain) binds to transforming growth factor beta (TGF-beta)[205,228].

Fibrillin-1 protein contains 15 potential sites of glycosylation and one single cell attachment site[229]. Fibrillin-1 protein can be degraded by matrix metalloproteinases (MMP) like MMP2; MMP3; MMP13; MMP 9 and MMP12 proteins [230].

### 1.13.5 Fibrillogenesis and Assembly Matrix incorporation defects and some Interactions

Fibrillin-1 precursor undergoes cleavage at C-terminal, mutation in exon 64 (close to the C terminal) leading to R2726W substitution, and is adjacent immediately to a consensus sequence (R-G-R-K-R-R) involved in cellular protease that will inhibit cleavage, as a result, this mutation will prevent pro-fibrillin-1/fibrillin-1 processing, hence resulting in null fibrillin-1 allele phenotype[231,232].

Molecular investigation of elastic fibre formation on fibrillin-1 showed high affinity calcium-independent binding of elastin (into two overlapping fibrillin-1 fragments), fibrillin-1 exons responsible for this binding are in the centre of fibrillin-1 exons 18-30, also exons 9-17 encoded for a fragment that have a novel transglutaminase cross link with elastin was documented [26]. Figure 13 shows a possible model of fibrillin-1 formation and interactions, starting with head to tail interaction between two fibrillin-1 monomers, when accomplished it will achieve linear extended structure. Microfibril-associated glycoprotein-1 (MAGP1) interact with this structure, then tropoelastin cross-link with fibrillin-1 chain and then further deposition of tropoelastin and cross linking interactions are performed to form a large microfibrillar structure, any mutation in these monomers can result in ECM changes that may leads to a disease [26,226,233]. Exon 56 of fibrillin-1 links by transglutaminase to another fibrillin-1 molecule at residue 2312 [234].

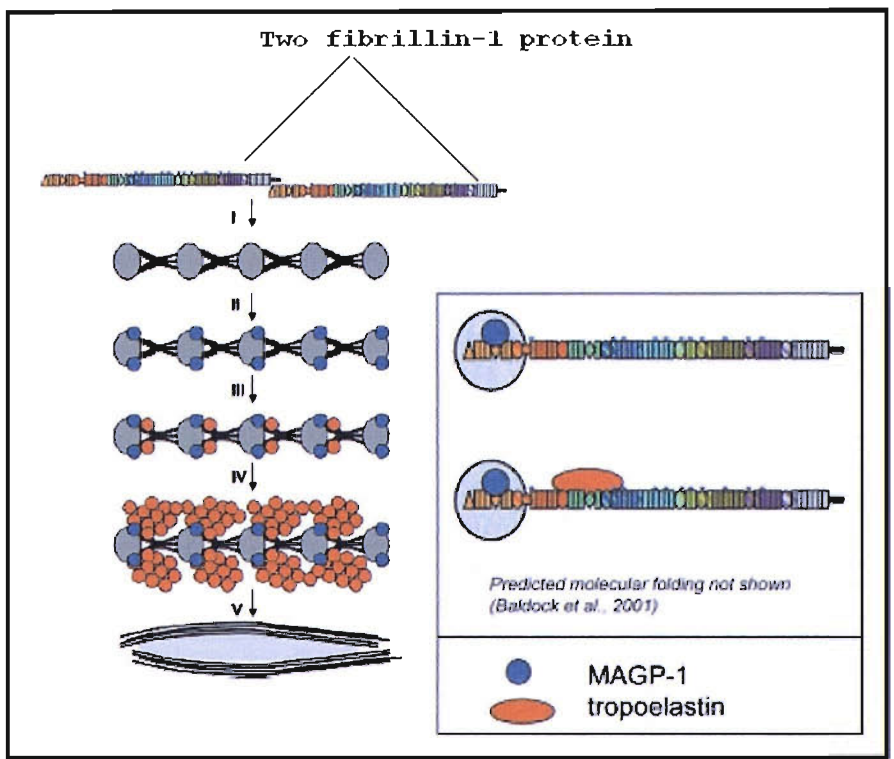



**Figure 13: This figure illustrate a model of elastic fibre formation in the extracellular matrix, (1) Formation of linear and lateral assembly. (2) MAGP-1 bind with the microfibrils at the bead area. (3) Binding of tropoelastin (transglutaminase binding) in another domain of the microfibrils. (4) Possible interaction of tropoelastin to another tropoelastin or to MAGP-1 and (5) Further deposition elastin and cross linking via lysyl oxidase. Adapted from [26]**

A risk factor for AAA disease is associated with Gly1127Ser substitution mutation in *FBN1* gene in exon 28 (see Appendix K for detailed fibrillin-1 gene), this mutation occurs in EGF-like domains, and is not related to MFS, a possible mechanism can be due to a reduced matrix deposition, with weakening of the elastic tissue[85]. Another patient with aortic aneurysms and dissection had a mutation in exon 27 (Asp1155 asparagine (Asn)), which disrupts an amino acid involved in calcium binding. This mutation decreased the amount of fibrillin-1 protein deposition into the pericellular matrix [86]. These two exons were screened in our study using our sporadic SAH samples.

Fibrillin-1 can interact with ECM proteins, one example is interaction with fibulin2 which binds to the N-terminal region of fibrillin-1, another example is binding with laminin B2 to the C terminal of fibrillin-1[235]. Many other proteins that has been reported to co-localise or to associate with fibrillin-1 containing microfibrils.

In 1994, a paper proposed that three possible defects can be responsible for MFS (see Figure 14), decreased fibrillin-1 synthesis, secretion inefficiency of fibrillin-1 and defects in the incorporation of fibrillin-1 with other proteins in the ECM [5].

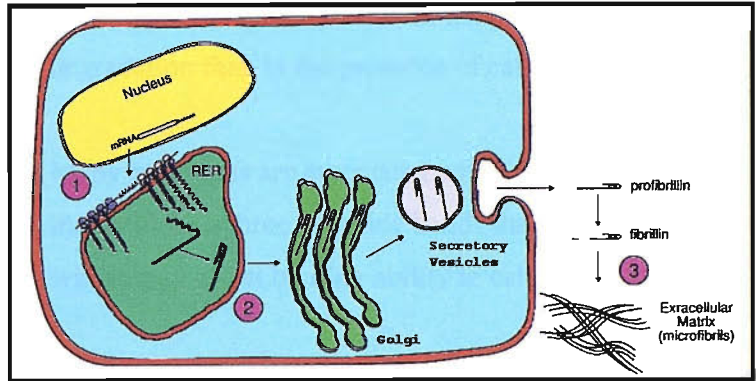



**Figure 14: Three possible defects that may lead to MFS occurrence: (1) Mutations responsible for decreased fibrillin-1protein. (2) Mutations that can affect the efficiency of secretion. (3) Mutations that affect fibrillin-1and prevent it from proper interaction with proteins present in the ECM ( adapted from [5] ).**

Five different pathophysiological models can be responsible for MFS:

**Dominant negative model:**

This model suggests that mutant fibrillin-1 monomer disrupts the assembly of normal fibrillin-1 into microfibrils or is it self miss-incorporated into the microfibril in the ECM [236]. This model is likely to be the strongest, and includes models from below.

**Mutant allele expression model:**

Mutations that involve in the premature termination of mRNA translation can affect the expression level of the mutant allele, which is associated with a range of phenotypic severity. The level of mutant protein modulates the severity of the disease, patients with lowest amount of mutant mRNA have mildest MFS phenotype [237].

**Normal allele expression model (haploinsufficiency):**

Differences in normal *FBN1* expression can be considered to as a potential cause of MFS; this was due to the finding of a *FBN1* deletion case with higher level expression in comparison to a single copy, depending on this case, suggests that variation in normal expression of *FBN1* may cause MFS in some cases [238].

**Stability and interaction model:**
Calcium-binding epidermal growth factor like domain (cbEGF) may be involved the binding with other proteins like the interaction between fibrillin-1 and fibulin in many tissues like elastic intima of blood vessels, and kidney glomerulus [239], also these domains have shown protein interaction like in factor IXa with factor X via EGF domains[240] . It was shown that fibrillin-1 protein in the presence of calcium has significantly slower proteolytic degradation than in the presence of calcium chelating agent (EDTA)[241].

Many mutations are associated with this domain in classical MFS cases, these mutations manly disturb three disulfide bonds that are created by cysteine a.a. [242] in this domain, which may affect binding ability to calcium.

**Dysregulation of transforming growth factor-beta (TGF-beta) model:**
MFS Mice lacking the fibrillin-1 gene have shown to have a TGF-beta dysregulation in signalling and activation, this may result in cell death in the developing lung. From that conclusion, matrix sequestration of cytokines is crucial to their regulated activation and signalling and that small changes of this function may contribute to the pathogenesis of MFS disease [243].

### 1.13.6 Some Functions of Fibrillin-1

The role of microfibrils is not yet well established, but they may have the following functions:
1- Acting as scaffolding for tropoelastin deposition and elastic fibre formation [244].
2- Linking elastic fibres both to each other and to other components in the ECM [245].
3- Extensibility function, they may contribute to the mechanical properties of elastic tissues by means of load redistribution between individual elastic fibres [246].
4- Maintenance of elastic fibres (sustain physiological haemodynamic stress in adventitia)[247].
5- Anchoring epithelial cells to the interstitial matrix[248], structural anchoring to non elastic tissue like ciliary zonules[249].

## 1.14 Transforming Growth Factor Beta Receptor 2 (*TGFβRII*) and linkage to aortic aneurysm

### 1.14.1 *TGFßRII* General

It belongs to the serine-threonine kinase family [250], when binding to TGF-beta family proteins, many functions can be accomplished, examples like: proliferation; differentiation; extracellular matrix production and cell death. In adults, these proteins are involved in tissue repair and immune regulation. Type II receptor have high affinity to TGF-beta than type I receptor [251]. When type II receptor binds to its ligand, a two (type I) and two (type II) complex dimer is formed[252], see Figure 15, the complex will bind; activate and phosphorylate type I homodimer receptor, this may result in the formation of an active tetrameric receptor complex.

Type I receptor in the complex may phosphorylate Smad 1;2;3;5 or 8 protein that may bind to Smad 4 to form a complex that will migrate to the nucleus and activate transcription of specific genes[253,254].

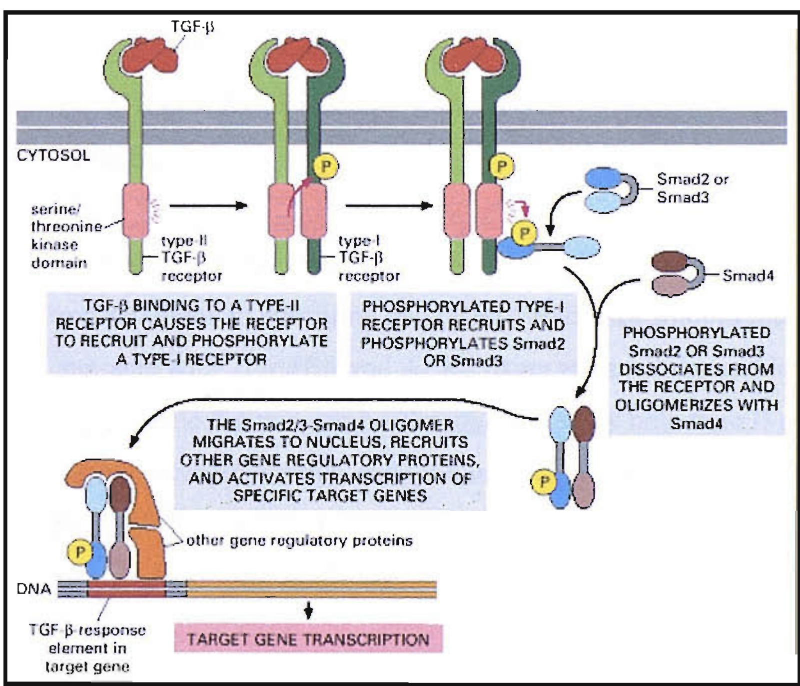



**Figure 15: Simplified mechanism of target gene to initiate transcription. Adapted from [253]**

Many *TGFβRII* mutations are associated with various cancers like colon cancer[255], breast cancer[256], colorectal cancer[257].

### 1.14.2 *TGFßRII* Role in Angiogenesis

Transforming growth factor beta (TGFbeta) cytokines play a role in angiogenesis via type I and II receptors, in knockout mice, homozygous *Tgfbr2* (an inactivation mutation generated via recombination) resulted in an embryonic lethality around 10.5 days of gestation due to a defects in the yolk sac haematopoiesis and vasculogenesis [258].

TGF activates type I receptor, wile type II may play a coordination function[259]. Type I receptor complex (in TGFbeta signalling) can be composed of two proteins (see Figure 16), the first protein is activin receptor-like kinase 5 [ALK5 (when activated by type II receptor it will phosphorylate Smad 1; 5 and 8)]. The second one is activin receptor-like kinase 1 [ALK1 (when activated by type II receptor it will phosphorylate Smad 2 and 3)]. Activation of Smad 2/3 inhibits proliferation migration of endothelial cells (EC), also inhibits Smad 1/5 actions in activating proliferation and migration of EC [260].
Type III receptor is a component of TGFßR system[159,261] , divided into betaglycan and endoglin.

Reports describing that mutations in endoglin are associated with defects in angiogenesis and problems in recruiting and differentiation of smooth muscle cells[262]. Endoglin interacts with *TGFβRII* and ALK5 using extracellular and intracellular domains, its cytoplasmic domain can be phosphorylated by both type II receptor and ALK5[263]. Some studies have shown that endoglin intronic insertion polymorphism is not associated with SAH [264,265]. Moreover, linkage analysis in the Japanese paper shown negative association between endoglin and SAH[159].

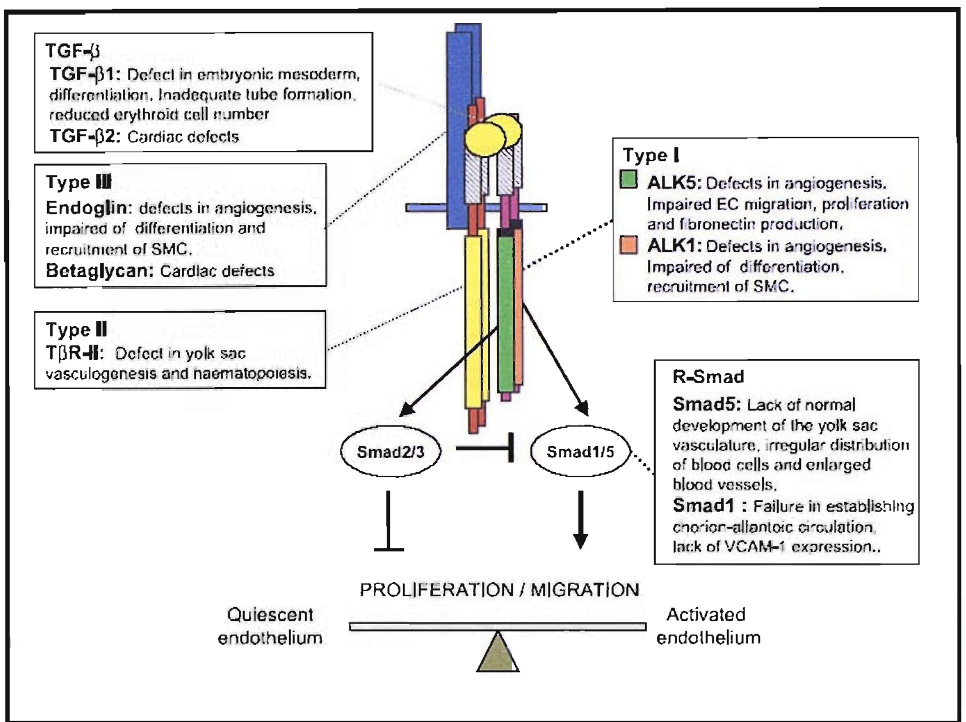



**Figure 16 "Regulation of ECM behaviour by TGF-beta signalling and corresponding vascular defects observed in mice deficient in TGF-beta components. TGF-beta switches ECM behaviour via two distinct TGF-h type I receptor (ThR-I)/Smad pathways. Upon TGF- beta -induced heteromeric complex formation, activin receptor-like kinase ALK 5 and ALK1 are phosphorylated and activated by ThR-II kinase. Signalling of TGF-h through ALK5 and subseqnent Smad2/3 phosphorylation leads to inhibition of ECM proliferation and migration. Signalling of TGF- beta through ALK1 via phosphorylation of Smad1/5 induces ECM proliferation and migration. Moreover, ALK1 signalling indirectly inhibits ALK5-induced Smad-dependent transcriptional responses. Vascular defects of mice deficient in TGF-h signalling components are listed. Abbreviations: VCAM-1, vascular cell adhesion molecule-1; SMC, smooth muscle cell." Adapted from [262]**

### 1.14.3 Genetics of *TGFßRII* and Marfan's Syndrome Type2

*TGFßRII* was shown to be involved in Marfan's syndrome type II in a large French family, this gene is located on chromosome 3p25-p24.2, this was evidenced through linkage analysis that localised MFS2 at D3S2335 with Lod of 4.89 [3]. The same region was mapped for familial aortic aneurysm and dissection (TAAD2) [266].

Also, the identification of a breakpoint in chromosome 3p24.1 in a Japanese person with MFS (this breakpoint was in the coding region of *TGFßRII*) led to consider the involvement of this gene in MFS (see Figure 17). Also mutation 1524G>A (Q508Q) present in the last nucleotide of exon 6, related to abnormal splicing segregated with

MFS2, this mutation resulted in an addition of 23 nucleotides to exon 6 and creating a stop codon at position 525 of the protein product. Moreover, another three mutations (in nine probands of unrelated French families with MF2 syndrome) which are 923T→C (leucine to proline substitution at position 308); 1346C→T (serine to phenylalanine substitution at position 449 and 1690C→T (arginine to cysteine substitution at position 537).

These three missense mutations are in the serine-threonine kinase domain, affecting a highly conserved or chemically similar amino acid in other species like mouse; rat; zebra fish and nematode[77].

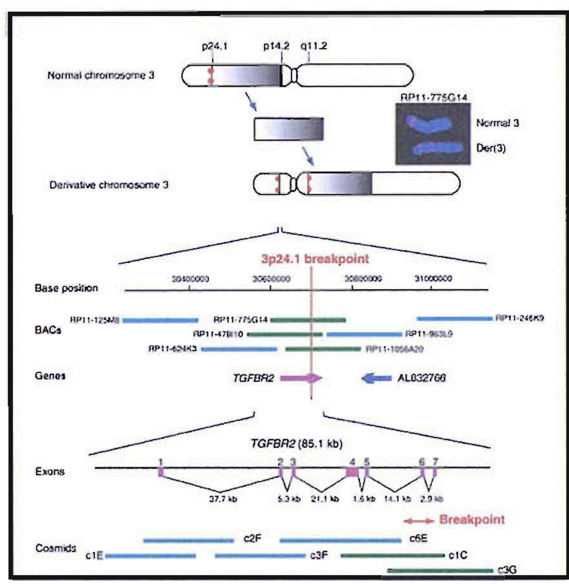



**Figure 17**: **The exact breakpoint in the *TGFβRII* gene (containing seven exons), breakpoint is in 3p24.1 between exon5 and exon 7. Adapted from[77]**

Another French family (in this study) was shown to be affected by MFS2 with SAH disease, investigation started to perform linkage analysis using STR markers used for the linkage analysis see Figure 18 below:

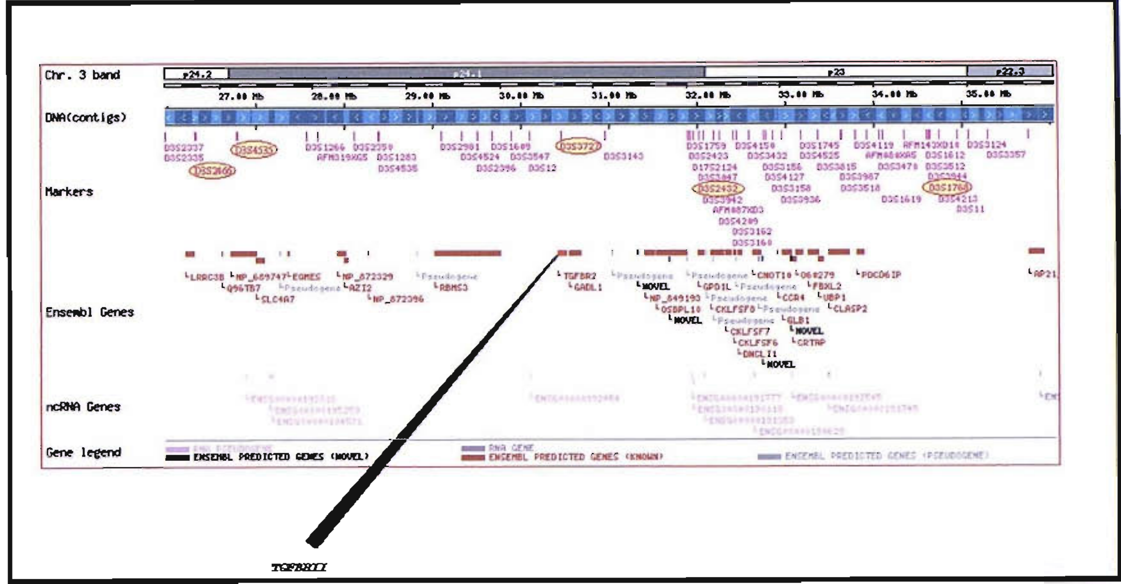


**Figure 18:** ***TGFβRII*** **is shown in red colour, markers used in this study are in oval, one of the markers used is located in the target gene (D3S3727), the location of** ***TGFβRII*** **is 30.65 cM, markers selected are located in both sides of the gene, and the completely covered region is around 10 cM.**

Heterozygosity of the markers is in Table 3:

**Table 3: Location and heterozygosity of STRs used in linkage analysis, D3S3727 marker was used to search for an association of** ***TGFβRII*** **and sporadic SAH** [267,268]**.**

| MARKER NAME | cM | Heterozygosity |
|---|---|---|
| D3S2466 | 26.8 | 0.85 |
| D3S4535 | 27.2 | 0.73 |
| D3S3727(CA) | 30.65 | 0.80 |
| D3S2432 | 32.8 | 0.81 |
| D3S1768 | 35.1 | 0.77 |

After that, I used the best marker to investigate sporadic SAH cases vs. control samples see (Figure 62) .

### 1.14.4 Relationship of *ELN* and *TGFßRII*

TGF-beta 1and2 were shown to be involved at least in post transcriptional regulation of the elastin gene[269], in another study, TGF-beta 1 was shown to increase elastin expression maximum of 3-folds in smooth muscle cells[270,270]. Moreover, TGFbeta was shown to correct the mRNA stability in cultures fibroblast cells of cutis laxa (CL) patients[167]. Smad3 knock out mice were shown to have reduced the amount of elastin mRNA in the

lung [271]. It seems that expression of elastin gene may result via the involvement of *TGFßRII* receptor.

This may suggest that *TGFßRII* receptor is involved in the elastin expression (also see Figure 15 and Figure 16).

### 1.14.5 Relationship of *FBN1* and *TGFßRII*

A paper published in Nature 2003 showed that marked dysregulation of transforming growth factor-beta activation and signalling was associated with fibrillin-1 deficiency in mice. Moreover, it contributes to the pathogenesis of MFS [272]. In another study, heterozygous loss of function mutations in the *TGFßRII* gene resulted in a phenocopy Marfan syndrome [77]. This gene product binds to TGF beta [251], as mentioned previously that changes in the TGFbeta signalling is present in many diseases like arterial aneurysms[27], this suggests that *TGFßRII* mutation may affect pathways leading to abnormal ECM proteins like fibrillin-1 and elastin.

### 1.14.6 Sporadic SAH and *TGFßRII* STR (D3S3727)

The tgfbr2 gene harbouring this microsat, is linked as a major gene to aortic and cerebral aneurysm formation. The microsat was tested for association with sporadic SAH.

One of the alleles that co-segregated in our linkage study was allele 132, which is a D3S3727 Di-nucleotide marker (see Figure 62). Since this allele is linked to blood aortic dissection in our French family, I wanted to test if sporadic SAH is associated with that allele. Hence, I performed genescan run of samples (214 samples 137 of them are sporadic SAH).

## 1.15 Possible methods for investigating variation in genes

### 1.15.1 Automated DNA Sequencing of PCR product

Automated DNA sequencing started at 1986 [276], At the beginning of DNA sequencing, PCR products should be treated with exonuclease I, (which will destroy the single stranded DNA i.e. primers) and the shrimp alkaline phosphatase (which decompose excessive dNTPs). Destruction of dNTPs and single strand DNA will prevent any interference with the chain termination sequencing reaction. The principle of sequencing involves Sanger's dideoxy chemistry with the incorporation of fluorescent dye at the dideoxy terminus, i.e. the elongation reaction is terminated by incorporation of labelled ddNTP that lacks OH groups at the 3' position of the sugar, necessary for polymerisation.

Using high resolving polyacrylamide gel, the sequencing product is subjected to electrophoresis, as result of laser exposure, fluorescent light is emitted and detected by a sensitive photon detector. Since four different incorporated dyes are present, four different fluorescent wavelength lights will be emitted and translated to electropherograme output. Direct DNA sequencing is the golden standard for mutation detection but it remains labour intensive, expensive, ABI 377 was used (Figure 19)

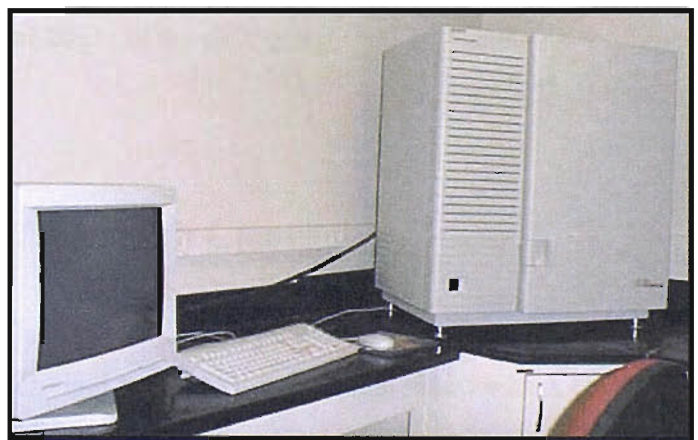


**Figure 19: ABI 377 sequencing machine.**

### 1.15.2 Amplification Refractory Mutation System (ARMS)

This method is used for the detection of specific SNPs or small deletions (invented in 1987), it is fast, sensitive, simple and reliable. ARMS PCR amplification is performed when the 3' end of the primer matches the target DNA, otherwise no successful PCR reaction will occur. To increase the specificity of the reaction, another mismatch is introduced in both forward primers at the –2 position. Since ARMS test for SNPs consists of two complementary reactions, both homozygous (normal and mutant) and heterozygous genotypes can be revealed [277], it can detect only one type of SNP. Therefore, for every known SNP special ARMS primers are prepared. Two mutations were genotyped by this method:

(1) Intron 20 A/G **C** (rs2856728)

(2) Exon 20 C/T (rs2071307)

### 1.15.3 Restriction fragment length polymorphism

This method use amplified DNA for restriction digestion, enzymes used are specific for certain nucleotides, when there is a SNP that is recognised by this enzyme it is possible to use it for genotyping, also we can use it if there is a loss of restriction fragment due to a mutation in the DNA.

TspRI enzyme was used for the genotyping of one SNP (Intron 23 A/G. **(HIDEAKI ONDA et al 2001**) of the *ELN* gene).

### 1.15.4 Mismatch Cleavage

Chemical cleavage of heteroduplex DNA using: osmium tetroxide $OsO_4$ (which reacts with a T mismatch) and hydroxylamine (which reacts with a C mismatch), after the addition of pipridine cleavage all modified DNA will be accomplished. The product is then analysed by electrophoresis on a denaturing polyacrylamide gel. The efficiency of this method is almost 100% and can work on 1.7kb in size. This method can give information about the

position of the mutation. However, the disadvantages are the use of toxic chemicals and it is very labour demanding [278] [286]. This method invented in 1989 [287].

The second type of cleavage is an enzymatic cleavage, one of the enzymes that can be used is called T4 Endonuclease VII, which is the product of gene 49 of the T4 bacteriophage [288] and involved in the DNA repair [289]. This enzyme will cut within six bases on the 3' side to the site of DNA distortion.

The concentration of the enzyme should be optimised, since excess enzyme leads to over digestion (hence loosing the PCR products) and low concentrations of the enzyme results in under digestion (see **Appendix** A and Table A 9).

The cleavage efficiency of this enzyme varies between different mismatched structures in three different levels, low efficiency with all G mismatches ( 10 % cleavage), intermediate efficiency in A/A A/C, C/A, T/C and T/T mismatches (30 % cleavage), high efficiency in C/C and C/T mismatches (50 % cleavage) [289].

Maximum DNA size that this method can work on is 1.5 kb, but unlike the chemical method, it cannot define what the exact mutation is. Detection limit is less than the chemical method [278]. this method was used to resolve Holliday structures in 1982 [287,288].

### 1.15.5 The Endo VII / Double labelled primers

This method can be used in establishing an assay for the screening of fibrillin-1gene, with a combination of double universal primers see Figure 20:

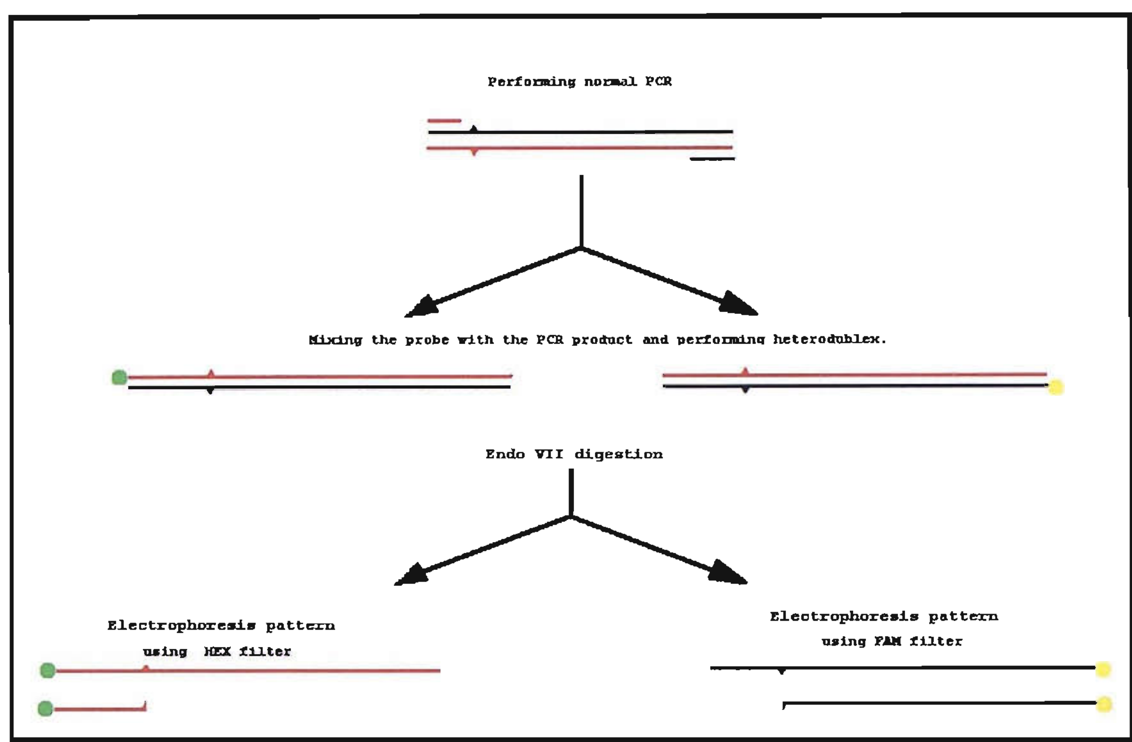



**Figure 20: Principle of the Endo VII digestion:**

To perform a double labelling reaction, I have added both labelled forward and reverse primers and performed PCR on normal sample. Now probe formation is accomplished. After I performed the PCR on unknown samples, both probe and PCR product are incubated to form heteroduplex structure. The product is incubated with The Endo VII (this enzyme will cut at one strand with heteroduplex structure), I stop the reaction then performing denaturant electrophoresis, an example of a mutation and expected finding using two different filters to detect mismatches are shown here.

## 1.15.6 Microplate-Array-Diagonal-Gel Electrophoresis (MADGE)

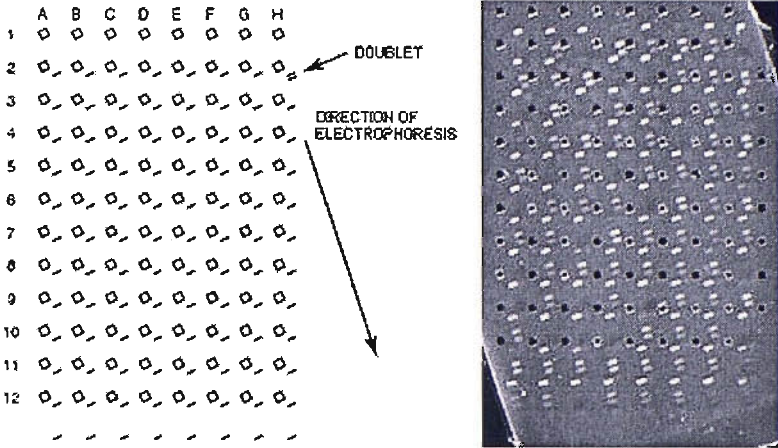



**Figure 21: MADGE system can run 96 samples at the same time, a very high throughput method to run high number of samples. "The wells are 2mm square, the angle between the direction of electrophoresis and the 12-well rows of the array is 71.5 degree, and the track length per well is 26.5 mm" [292]**

This method was invented in 1994, it is high throughput compared to conventional electrophoresis, we use acrylamide gel to run DNA samples we can run up to 96 samples at the same time [292].

## 1.15.11 GeneScan and Microsatellite Detection

This method is used to calculate microsatellite length, like di; tri; tetra and penta nucleotide repeats, this method depends on the migration of single strand DNA segment in a denaturant gel, a marker is run with the sample to estimate the length of the denatured PCR segment. This method was used in 1995 [293].

6-Carboxyfluorescein (6-FAM) and 4,7,2',4',5',7' -Hexachloro-6-carboxyfluorescein (HEX) are examples that can be used to label the amplified segment, a marker is mixed with the PCR product, a computer programme is used to calculate the fragment length. GeneScan is used in conjunction with the ABI sequencer.

## 1.15.12 Denaturing High-Performance Liquid Chromatography (DHPLC)

DHPLC is a rapid automated scanning method for mutation detection used in DNA analysis in 1999 [294]. Heterodublex formation is an important step before the DNA samples enter the DHPLC (Figure 22).

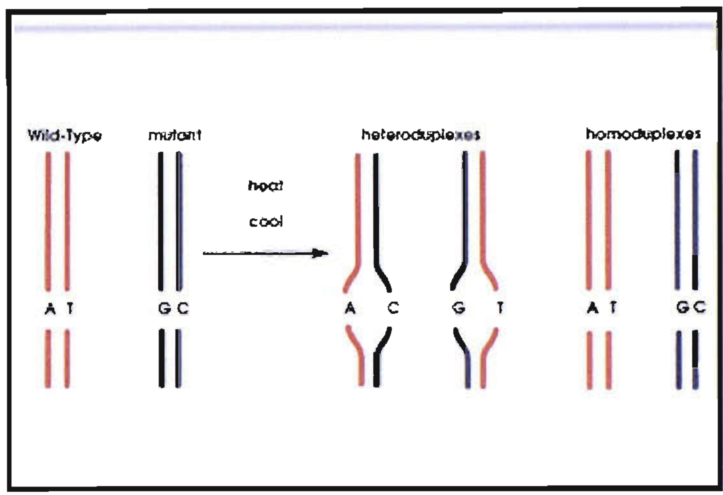



**Figure 22: Heteroduplex formation will result in four different combinations in heterozygous DNA, it involves denaturation the renaturation of the DNA product.**

However, it is not important to know the nature and location of the mutation. Visualisation of mutations is performed through a characteristic pattern of peaks in comparison of the wild type (Figure 23).

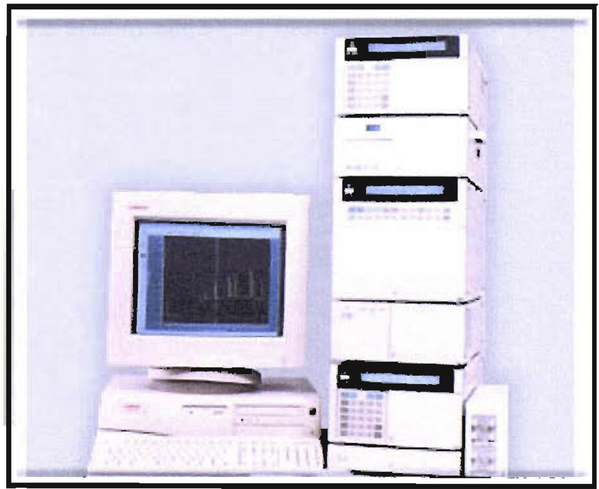


**Figure 23: The Transgenomic wave DHPLC used in *ELN* scanning for heteroduplexes.**

This method can compare two different chromosomes after the formation of the heteroduplex product, which is achieved by denaturation and renaturation of the PCR product (see **Appendix A**, Table A 5). The PCR product will run on a reverse phase chromatography column and is subjected to a gradient increase in the denaturation chemical (acetonitrile), the ion paring chemical triethylammonium acetate (TEAA) ensures that the DNA interacts with the column (electrostatic interaction between TEAA and DNA and hydrophobic interaction between the column and TEAA). Heteroduplex products will have lower retention times, so that they will emerge before the homoduplexes. DHPLC can detect more than one mutation in the amplicon strand. Some limitations are that some homozygous mutants cannot be detected unless they are mixed with the normal PCR amplicons. Furthermore, the ideal PCR product length is from 200-500 bases, otherwise sensitivity will be compromised, broadly, the sensitivity of the DHPLC is more than 97% [295].

## 1.15.13 LightTyper (Odyssey)

LightTyper is very high throughput method for genotyping of known SNPs, it will take about 10 minutes to analyze 384 samples in (384 well plate system). The principle of this assay is to perform an asymmetric PCR, this will create a single stranded amplicon that will be used as a base for probe hybridisation. When the probe hybridise on a wild type product and form a homoduplex then it will need higher denaturation temperature in comparison to the heteroduplex hybridisation as if a mutation is present. First publication using this machine was in 2003[296]. Figure 24 shows the principle of this technique:

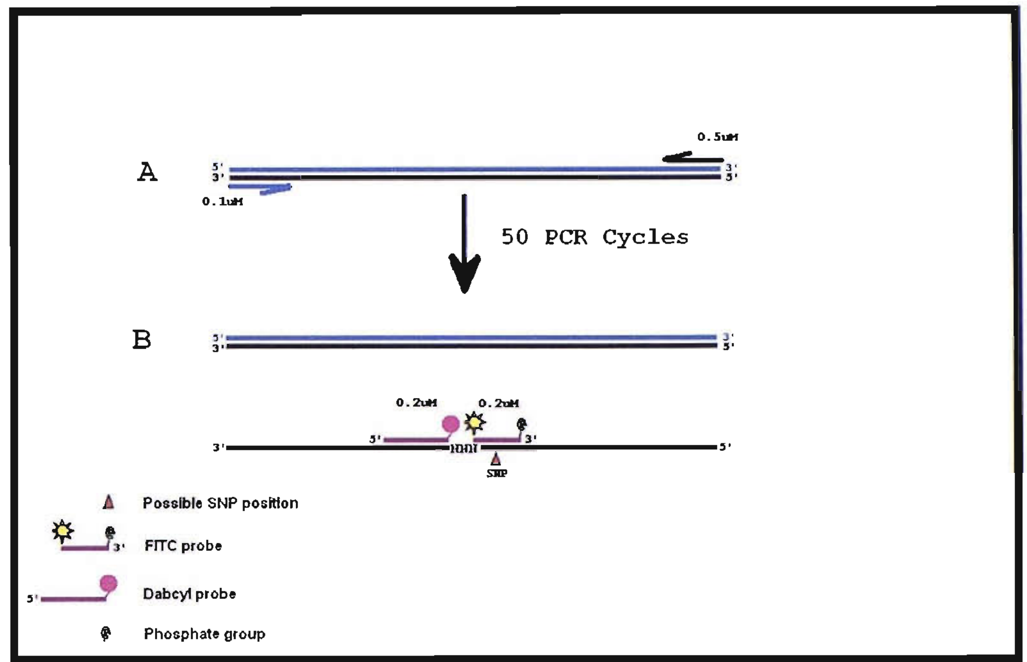



**Figure 24: Principle is using a single stranded PCR product (that contains a SNP) to be used as a target for a 5' Fluorescence probe, another probe contains a 3' dapcyl is responsible for preventing the FITC from emitting fluorescent light and need higher temperature to be denatured.**

In the Odyssey the PCR and probes are subjected to an increase in temperature (Figure 25), this will allow the denaturation of the probes. Firstly, the heteroduplexes probes are denatured then the homoduplexes ones, it is important that the dabcyl probe should have about 10 degrees higher denaturation temperature than the FITC homoduplexes probe.

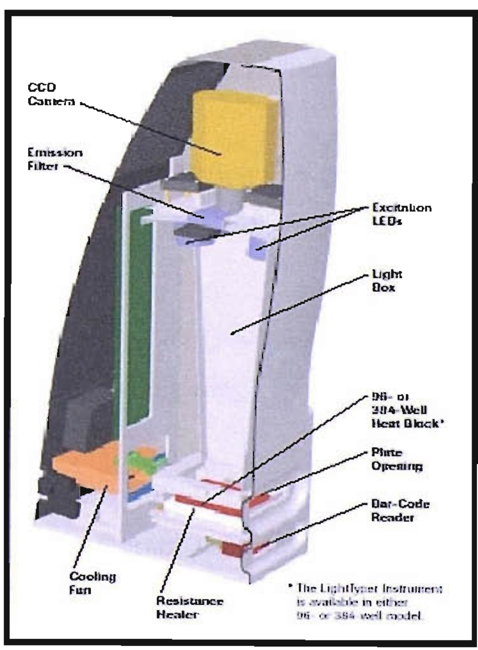



**Figure 25: In the LightTyper an increase in the temperature is performed to allow denaturation of the probes, samples are subjected to a UV light, when the probes are denatured and are away from the dabcyl they will fluoresces light that will be detected by special detectors. Adapted from [297]**

When the denatured FITC probes are exposed to the UV light, the FITC probe will fluoresce a light with higher wavelength that can be detected by the machine. I have used this machine in genotyping the exonic SNP (Exon 20 C/T (rs2071307)) using the BWHHS samples.

## 1.16 *In silico* analysis

Frequency estimation of haplotypes in the population using EM (Expectation Maximisation) can be performed through **Arlequine** software. **Phase** software applies information about the haplotype for each tested individual, and provides estimation of the haplotype frequency in the population. It has been suggested that using Phase can increase the accuracy of the haplotype frequency by up to 50 % compared to Arlequine[298].

It is important to see the effect of any cSNP detected in our study; this can be accomplished by *In silico* studies. Many programmes are available like **ESE finder**; **RESCUE-ESE** ( these two programmes are designed to look for exonic splicing sites effect created by a coding SNP (cSNP), other program that I have used is **N,N splice**, this one I can input sequences that are intronic to investigate any possible effect on splicing.

**PolyPhen** (*Poly*morphism *Phen*otyping) program can help in prediction of the possible effect of the SNP/mutation on the structure and function of the protein.

**Ali Baba 2.1** is a program for predicting transcription factor binding sites in an unknown DNA sequence using the binding sites collected in TRANSFAC.

**TRANSFAC**: is a database on eukaryotic cis-acting regulatory DNA elements and trans-acting factors. It covers the whole range from yeast to human, when I used it under the TFBLAST.

**TFSEARCH**: is another program used to see the possible effect of gene expression, looking for transcription factors sites and the possible creation or destruction of a 5' transcription sites by SNP.

## 1.17 Hypotheses and Plan

I have launched a molecular study trying to find the reasons behind these aneurysmal diseases and concentrating more on ICA and SAH disease (Figure 3).

I hypothesise that specific major or minor gene effects on occurrence or outcome of cerebral aneurysms and subarachnoid haemorrhage can be identified through a combination of linkage and association studies.

# Chapter Two

## 2.0 Material and Methods

## 2.1 Samples used in this study

During 1998-2000, Day and colleagues in Human Genetics Division undertook a pilot study working in collaboration with Fausto Iannotti, late Professor of Neurosurgery and a research nurse (Lesley Foulkes) funded by a local grant from the Wessex Medical Trust.

In this study, Day and his group attempted to review all families known from the previous five years to the Wessex Neuro Unit, in which two or more members were affected by SAH:

8 families were known to the neurosurgical unit and were contacted directly
200 idiopathic SAH (potential probands) were identified
191 questionnaires were sent, for nine there were no contact details

they received 133 replies:

22 reported at least one other SAH in the family
16 families were identified in total, the detail from pilot data collection from closely related kindred is in Table 4:

**Table 4: Number of families collected for this analysis was 16, relevant family medical history is provided.**

| HIGH RISK SUBARACHNOID HAEMORRHAGE FAMILIES | | |
|---|---|---|
| **FAMILY No.** | **RELEVANT FAMILY MEDICAL HISTORY** | **CONSENT Y/N** |
| 500 | 3 SAH'S, 1 AAA, 4 MIGRAINES | Y |
| 501 | 2 SAH'S, 2 MIGRAINES | Y |
| 502 | 3 SAH'S | Y |
| 503 | 2 SAH'S, 2 MIGRAINES | Y |
| 504 | 2 SAH'S | N |
| 505 | 2 SAH'S, 2 ANEURYSM | Y |
| 506 | 2 SAH'S, 1 SDH + AAA | Y |
| 507 | 5 SAH'S, 1 ANEURYSM, 1 MIGRAINE | Y |
| 508 | 2 SAH'S. KNOWN COLLAGEN DEFICIENCY | Y |
| 509 | 3 SAH'S, 1 AVM | Y |
| 510 | 3 SAH'S, 1 ANEURYSM | Y |
| 511 | 3 SAH'S, 4 AAA, 15 MIGRAINES - MARFANS | ON HOLD |
| 512 | 3 SAH'S, 1 AAA, 1 CVA, 2 MIGRAINES | Y |
| 513 | 2 SAH'S, 2 MIGRAINES | Y |
| 514 | 4 SAH'S | Y |
| 43 | 2 SAH'S, 2 MIGRAINES, 2 CVA'S, 3 DIED IN SLEEP | Y |

While two SAHs could be a (rare but positively ascertained) event in a family, the occurrence of three or more can be taken as strong evidence of a major gene[299], particularly when, as is the case for most of these families, vertical transmission over two or more generations is observed. Intergenerational transmission, coupled with affected cousins living in separate households from birth, argues against a household environmental factor of major effect.

Because of the high mortality rate at diagnosis, it is rare to find more than 2-3 living affected individuals known in one family (i.e. DNA available). This presents some specific challenges, which probably explain why SAH genes have not yet been identified by linkage studies.

It is also evident that some families seem to display both SAH and other aneurysms, notably aortic. Such heterogeneities and those noted for known aortic aneurysm loci, render these loci quite plausible in SAH also.

Samples collected for this study are divided in to three categories and they are from three different sources;

1- Familial SAH: 130 blood DNA samples, of the blood samples 80 samples came from Glasgow and 51 DNA familial samples were collected. Five samples from familial were used in the DHPLC Scan, they are 506.01; 505.01;509.01;510.01;506.11. Three samples from the Glasco samples from the affected SAH with AAA were used. French family used for GenScan analysis: The tgfbr2 family was from Southampton/Channel Islands, apparently with some of the Channel Island founders originating from France, they are part of the familial SAH samples.

2- Sporadic SAH and control groups: 214 DNA samples (from blood) were collected, 137 are sporadic patients, who suffered from aneurysmal subarachnoid haemorrhage SAH (range, 23 to 75 years; mean age, 50 years; ancestry of white European) and received surgical treatment in the Wessex Neurological Centre, Southampton General Hospital . 10 ml venous blood samples collected from patients having a clinical history of SAH with an association of abnormal computerised tomography (CT) scan. The second group contains 77 Head injuries control. For more information see ref [300]:

DNA extraction was performed by salting out procedure[301].

The study was approved by the South and West Local Research Ethics Committee (submission No. 170/98), and written consent was obtained from the participants.

3- British women heart and health study (BWHHS cohort):

About 3000 women where recruited from 23 different towns (around 175 from each town) in England Wales and Scotland. All data regarding like phenotypes are present in the department of social medicine, University of Bristol, ages 60-70 years (see reference [302]) of those 2890 samples were genotyped on an exonic mutation on the elastin gene using blood pressure as a phenotype.

## 2.2 Reagents used in this project

### 2.2.1 For PCR and Restriction Enzymes

| | |
|---|---|
| Taq DNA polymerase: | Recombinant Invitrogen 500U Cat No 10342-020 |
| | From Promega cat No: M1665, 500U |
| | 8mM dNTPs Promega. Cat No: U1330 100mM each |
| TspRI | Cat No: R0582S, New England BIOLABS (for *ELN* SNP) |
| Tsp45 I | Cat No: R05836, New England BIOLABS (for *TGFβRII*) |

### 2.2.2 For DNA Sequencing

DNA sequencing Kit BigDye Terminator cycle sequencing ready reaction. Part No: 4303152
Shrimp Alkaline Phosphatase. Amersham biosciences, cat No: E70092Y, 500U
Exonuclease I. Amersham biosciences, cat No: E70073Z , 2500U
Methanol BDH Prod No: 291926g

### 2.2.3 For normal gel preparation and staining and Electrophoresis

**Preparation of 10X TBE for 1 L:**

108 g Tris (hydroxymethyl)-methylamine. Fisher Chemicals. Code T/3710/60
55g Boric Acid. Sigma. Cat No B-0394
9.3 EDTA (Ethylenediamietetra-acetic acid sodium salt), dihydrate Sigma. Cat No: E-5134
Up to 1L Deionised water

**Agarose gel for electrophoresis 2%:**

2.0 g Agarose Ultra Pure electrophoresis gradient GIBCO BRL. Cat No 15510-027
100 ml 1xTBE buffer

**Acrylamide gel for MADGE electrophoresis 5%:**

8.3 mL 30% w/v Acrylamide ratio 19:1 Bis Acrylamide Severn Biotech Ltd. Cat No 20-2300-10

5 mL 10X TBE

35.7 mL Deionised water

150 µL (20%) Ammonium Persulfate (Free radical donor) (APS). Promega. Part No. V313a

150 µL TEMED ( N,N,N',N'-Tetramethylethylendiamine) (catalyses free radical formation) Sigma cat No T-7024

NB: "In the presence of free radicals (provided by ammonium persulfate, catalyzed by TEMED), a reaction occurs such that the acrylamide monomers polymerise (form chains) and the bis molecules, where incorporated, provide cross-links between the chains. This forms a regular matrix with "holes" that serve as pores in the polyacrylamide gel." From: http://biotech.nhctc.edu/BT210/tutorial10.html

**Gel staining and Ladder Marker:**

Ethidium Bromide Solution Fisher Scientific Code E/P800/03

50 bp DNA Ladder Invitrogen Cat No 10416014

100 bp DNA Ladder, Promega Cat No G2101

**Acrylamide gel for DNA sequencing and GeneScan:**

30 mL Gene-Page Plus 5.0% 6M UREA AMRESCO UN No 2810

300 µL (10%) Ammonium Persulfate (APS) Promega. Part No. V313a

30 µL TEMED ( N,N,N',N'-Tetramethylethylendiamine). Sigma, cat No T-7024

**Preparation of Formamide Dye Mix for 10 ml:**

9800 µL Formamide 98% deionised. Sigma. Cat No F-9037

200 µL 10mM EDTA (pH 8.0, 0,5M)

1.5 mg Xylene Cyanol FF. Sigma. Cat No X4126

**Sticky saline composed of the following for each (100 ml):**

500 µL Glacial Acetic Acid BDH. Prod 270134x

500 µL Trimethoxysilylpropyl-methacrylate. Fisher Scientific, code T/3350/48

99 mL Methanol BDH Prod No: 291926g

### 2.2.4 For DHPLC

**Buffer A: 0.1M TEAA from stock solution 2M**

| | |
|---|---|
| 50mL | Triethyl Ammonium Acetate (TEAA) WAVE Ion Paring Agent TransGenomic. Part no. 553303 |
| Up to 1L | Deionised water( DHPLC grade) |

**Buffer B: 0.1M TEAA from stock solution 2M and 250 Acetonitrile (25%)**

| | |
|---|---|
| 50mL | Triethyl Ammonium Acetate (TEAA) WAVE Ion Paring Agent TransGenomic. Part No 553303 |
| 250mL | Acetonitrile HPLC grad ALDRICH. Cat No 27,071-1 |
| Up to 1L | Deionised water (DHPLC grade) |

**Buffer C: 75% Acetonitrile Cleaning solution:**

| | |
|---|---|
| 750mL | Acetonitrile HPLC grad ALDRICH. Cat No 27,071-1 |
| Up to 1L | Deionised water (DHPLC grade) |

**Buffer D: Syringe wash solution 8% Acetonitrile**

| | |
|---|---|
| 80mL | Acetonitrile HPLC grad ALDRICH. Cat No 27,071-1 |
| Up to 1L | Deionised water (DHPLC grade) |

### 2.2.5 For GeneScan

**For *ELN* CA repeats**

| | |
|---|---|
| 1 µL | GeneScan-500 ROX Size Standard Part No. 401734 Applied Biosystems. |
| 5 µL | Formamide 98% deionised Sigma. Cat No F-9037. |
| 1 µL | Blue Dye. |
| 1 µL | Optimised PCR concentration (1:10). |

**For *TGFβRII* Tetra STRs (4 reactions)**

| | |
|---|---|
| 1μL | GeneScan-500 Tamra Size Standard Cat No. K3215, Genetix. |
| 5μL | Formamide 98% deionised Sigma Cat No F-9037. |
| 1μL | Blue Dye. |
| 1μL | Optimised PCR concentration (1:5 each reaction). |

## 2.2.6 Reagents for the Endo VII

**10X Denaturation Buffer for <u>two Litres</u> stoke (add hot water)**

| | | | |
|---|---|---|---|
| Tris base | (900mM) | 216 | g |
| Boric Acid | (900mM) | 110 | g |
| EDTA | (20 mM) | 18.8 | g |
| Urea | (about 7.5 M) | 1.0 | Kg |

**Diluent (<u>1 Litre</u>) (add hot water)**

| | | | |
|---|---|---|---|
| Urea | (7M) | 410 | g |

**Stocks for the Endo VII Reaction Buffer to prepare for 20 T4EVII reactions.**

| | | | |
|---|---|---|---|
| 1M K2PO4 | 10 | μL | (pH control) |
| 1M MgCl2 | 1 | μL | (enzyme activator, add last thing to prevent protein precipitation) |
| 100mM DDT | 2 | μL | (Unstable at room temperature. Prevent S-S bonding) |
| 0.1mg/ml EndoVII | 5 | μL | |
| DW | 22 | μL | |

**5X T4 Endo VII Reaction buffer:**

| | | |
|---|---|---|
| 250mM | K2PO4 | (pH 6.5) |
| 25mM | MgCl2 | |
| 5mM | DTT | |
| 0.1mg/ml | Endo VII | |

**Stop solution (20 mL)**

| | | | |
|---|---|---|---|
| 10 mM | NaOH | 40 | µL [5.0 M concentration stock] |
| 50 mM | EDTA | 400 | µL [0.5 M concentration stock] |
| 80% | Formamide | 16 | mL |
| 0.25% | Bromophenol Blue | 0.05 | g |
| 0.25% | Xylene Cyanol FF | 0.05 | g |

### 2.2.7 Purification of PCR Probes for the EndoVII

Wizard PCR Preps DNA purification system kit was used to purify the probes used in the Fibrillin-1gene, this kit is from Promega cat number A7170 www.promega.com .

### 2.2.8 Other reagents

Alconox detergent ALDRICH Cat No 24,298-5 (prepare 1%) used in cleaning up the prism after DNA sequencing.

Betaine anhydrous: Cat no: EC 203-490-6 sigma, add 5.8575 g and up to 10 mL for 5M betaine.

## 2.3 Instruments used in this project

### 2.3.1 For PCR

DNA Engine Tetrad MJResearch Gradient cycler Model PTC-225.
Pur1TE serial No 19420.
Whirl Mixer serial number 28201.
200 V/A 50-60 Hz output Power supply, BioRad laboratories INC,CA USA.

### 2.3.2 Visualisation of PCR product

FluorImager595 (Molecular Dinamix) Amersham Pharmacia Biotech, serial No. 86297.
UVP San Gabriel CA 91778 USA. Model TM-20. 95-0173-05.
Typhoon Trio+ 9400/9410, serial No. 98017 Amersham Biosciences USA.

### 2.3.3 For DHPLC

D7000 HPLC System Manager Hitachi. Model D7000 virsion3.0-2.2 ( Transgenomic).
Millipore Cat No SIMS 5VOCS. Serial No FODN53771H.
DNA Engine Tetrad MJResearch Gradient cycler. Model PTC-225.
Pur1TE. Serial No 19420.

### 2.3.4 DNA Sequencing and GeneScan

ABI Prism 377 DNA Sequencer version 2.1 Perkin Elmer Corp PE Applied Biosystems.
Heraeus centrifuge. Serial No MC0230.
Whirl Mixer. Serial number 28201.
Techne Dri-Block Heater DB.2D. Model FDB0200 Serial No 70469-15.
DNA Engine Tetrad MJResearch. Gradient cycler. Model PTC-225.
Pur1TE Serial No 19420.

### 2.3.5 MADGE Former

96 (8x12) well, industry standard plated supplied by Thermo-Fast, USA Cat No. MSA-5001.
Glass Plates 110mmx170mm.
MADGE Former. Supplied by MADGE Bio Ltd, Nottingham. UK.
96-Pin Passive Replicator , provided by MADGE Bio Ltd, Nottingham. UK.

### 2.3.6 Other Instruments and disposables

Sartorius GMBH Type 1207, No: 3002081, made in Germany.
Pipettes: (2μL, 10μL, 20μL, 200μL) Gilson SA Villiers-le-Bel, FranceDisposables lab tips: 5-200μL Finntip 200, 100-1000μL Finntip 1000 from Finntip® - Standard Pipette Tips - Offered By Thermo_Labsystems0.5-10μL Tips from Fisherbrand. Cat No: FB56196.

**Microtubes:**
Microtubes 1.5 ml cat No: LW2375. Microtubes 0.5 ml, cat No LW 2372 from Alpha laboratories.

**96 well PCR Tubes:**
From Abgene thermo-Fast 96 skirted. Cat No: AB-0800 www.abgen.com.

**Gloves:**
Powder free Latex Exam gloves from Kimberly-Clark Corporation. Ref No. E330.

## 2.4 Primers and Probes Design

### 2.4.1 *ELN* Primers for DHPLC

Primers used are from Sigma-Genosys (http://www.sigma-genosys.co.uk), Primer 3 software (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) used to design the primers. These primers are designed to cover the 34 exons of elastin gene, also they are designed to cover some of the 5' and 3' ends of each amplified exon. In addition, primers were designed (for ARMS and Restriction digest of PCR) assays for Genotyping of 3 SNPs present in Exon 20 Intron 20 and Intron 23. See **Appendix A** Table A 3.

### 2.4.2 *ELN* Primers for GeneScan

Primers designed for GeneScan analysis were modified by adding at the 5' end 6-FAM. Bought from MWG.

### 2.4.3 *ELN* Probes and Primers for Odyssey (LightTyper) Genotyping

Normal primers are designed using primer 3, probes are of two types the 5' fluorescence probe is 16-20 nt long, because it is 5' probe then phosphate group is attached to the 3, end to prevent extension during PCR. SNP should be near the centre of the fluorescence probe. The second one is 3'dabcyl probe 18-28 nt long, also it should have 10 degrees higher TM than the fluorescence probe. Energy from fluorescein molecule is absorbed by dabcyl, resulting in reduced fluorescence when stimulated by LED light from the Odyssey. After melting the fluorescein molecule is separated from the quencher(s) and therefore fluorescence increases and is detected by the camera in the Odyssey.

## 2.5 PCR optimisation

Table A 1; Table A 2 and Table A 11 shows the preparation of PCR mix that was used in the optimisation reaction using Gradient thermocycler. After PCR reaction is performed, the cleanest band with the right bp length is chosen for further work (i.e. DHPLC and DNA sequencing). Thirty-one assays were designed to amplify thirty-four exons of elastin gene, PCR optimisation was carried out for the coding region of the elastin gene. For the genotyping of the three SNPs used in my analysis I have used please see **Appendix A**

Optimisation for the Odyssey [for the exonic sequence(rs2071307)] is described in Table A 8, this optimisation was performed in 10 µl PCR reaction mix and used in 384 with 5 µl PCR reaction mix.

## 2.6 Primer Design for *FBN1* .

Primers designed to cover splice donor and aceptor elements that are involved in RNA splicing (the GT and AG sequences) . All exons are covered. I did not use all of the optimised primers in our study, I have used primers covering exon 27;28 and 56. Primer3 was used to design the primers. Primers from Sigma Genosys.

## 2.7 Primer Design for LightTyper for the *ELN* gene.

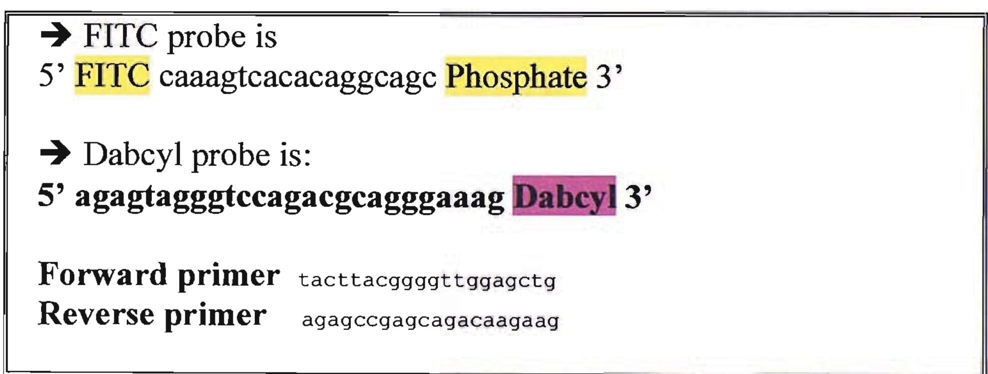
This is the sequence used to design probes for the LightTyper see Figure 26, the probe should have a phosphate group in the 3' region, in the 5' region FITC modification is added, the second probe is designed in the upstream of the FITC probe and should have Dabcyl group in the 3' end. Temperatures 53.6 and 63.7 are the calculated denaturation temperature. Figure 26 contains the exonic sequence(rs2071307) that was analysed in this study primers used are in **Appendix A** also this SNP was genotyped by ARMS.

```
     67561 ggcaacctac aggagtacct gacgcggcat gtcatcagct gggaggacct gcgcaagctg
      67621 ggcagctccc tcgcccgggg gattgctcac ctccacagtg atcacactcc atgtgggagg
     67681 cccaagatgc ccatcgtgca cagggacctc aagagctcca atatcctcgt gaagaacgac
                                                                cttcttgctg

     67741 ctaacctgct gcctgtgtga ctttgggctt tccctgcgtc tggaccctac tctgtctgtg
3'           gattggacga cggacacact gaaacccgaa agggacgcag acctgggatg agacagacac   5'
                  P       53.6C       FITC     Dabcyl      63.7C
     67801 gatgacctgg ctaacagtgg gcaggtaagt tagagctagt gctagatccc ctttaccttg
     67861 agcctggcct caccctacct cttgatccat atctcctggc tcttatctca aacagccctg
     67921 tactctggac actggtctag ggaatctagc caaagtatgg agtctgcctt gagcatactc
     67981 tgctctgtcc tgcctgagca tttttgctaa tggacagcat ttctcctcct atcttcaaat
```

**Figure 26: Our mutation in red box (G>A) yellow box contains the complementary sequence probe (with 5' FITC and 3' phosphate) of the segment that contains the SNP, at the 5' region of this sequence is the Dabcyl probe the dabcyl is in the 5' region.**

Figure 27 shows the exact sequence of the probes used in the LightTyper assay.


➔ FITC probe is
5' FITC caaagtcacacaggcagc Phosphate 3'

➔ Dabcyl probe is:
**5' agagtagggtccagacgcagggaaag Dabcyl 3'**

**Forward primer** tacttacggggttggagctg
**Reverse primer** agagccgagcagacaagaag

**Figure 27: The exact sequence of the probes used in the LightTyper assay.**

# Chapter Three

## 3.0 Results

## 3.1 Work Performed in this project

1- Detailed map of the elastin gene sequence with SNPs, exons and introns from the NCBI (Appendex E).
2- Complete amino acid sequence of the elastin gene with exons and cross-linking regions (Appendix D:).
3- PCR optimisation for 34 exons of the elastin gene (Table A 3 and Table A 4).
4- Scanning of the 34 exons using DHPLC of the *ELN* gene. Genotyping of three SNPs on sporadic SAH using ARMS and RFLP of the *ELN* gene and they are: Intron 20 A/G **C** (rs2856728); Exon 20 C/T (rs2071307) and Intron 23 A/G. **(HIDEAKI ONDA et al 2001)**
5- (Appendix C:)
6- Statistical association analysis by the use of Phase and Arlequine on sporadic SAH and control using the results of the three SNPs in the previous point (Table 7).
7- DNA sequencing for the positive DHPLC results of the *ELN* gene.
8- Fluorescence PCR optimisation 6-carboxyfluorescein (6-FAM) to amplify the GT microsatellite region of the elastin gene.
9- GeneScan analysis of GT microsatellite in intron 18 of the elastin gene and describing five models.
10- Examining the mutations that were seen in *ELN* 5' flanking region to see any possible functions (Table 9 and Table 23).
11- Genotyping of exon 20 SNP on sporadic SAH and control of the *ELN* gene.
12- Optimizing and genotyping 3000 subjects using Odyssey (LightTyper) for exon 20 SNP of the *ELN* gene see Table A 8.
13- Analysis the these results (Table 10)
14- Applying dominant, recessive and additivity models
15- Detailed map of the *FBN1* gene sequence with cSNPs, exons and introns form the NCBI (Appendix K:).
16- Complete amino acid sequence of the fibrillin-1gene with protein domains regions (Appendix D:).

17- PCR optimisation for 65 exons of the *FBN1* gene Table A 12.

18- Scanning 56, 27 and 28 exons of the *FNB1* gene using The EndoVII MADGE for mutations.

19- DNA sequencing on new mutations found by The EndoVII technique of the *FNB1*.

20- Linkage analysis using a family that may have aortic dissection and linked to *TGFβRII* gene using 5 STRs (Figure 62).

21- Map of the markers used in the linkage analysis used in the *TGFβRII* gene (Figure 18).

**22-** Association studies using one of the STRs (of the *TGFβRII*) in sporadic *vs.* control markers used are: (D3S2466, D3S4535, D3S2432, and D3S1768) and one CA repeat marker (D3S3727).

## 3.2 Phase, Arlequine and CONTING.

To examine whether the haplotypes present in introns 20 and 23 have any association with ICA and hence SAH, another three assays were developed to genotype three SNPs on sporadic SAH using ARMS and RFLP (**Appendix C**) these SNPs are:

(1) Intron 20 A/G **C** (rs2856728)

(2) Exon 20 C/T (rs2071307)

(3) Intron 23 A/G. **(HIDEAKI ONDA et al 2001) does not have rs number**

The first two were performed by ARMS method, the third one was performed by using the restriction enzyme **TspRI.** SNP results were shown to be in the HW equilibrium. **(Appendix C)**

Haplotype frequencies were estimated for two different groups (the first one composed of 137 individuals associated with sporadic SAH, the second group was composed of 77 individuals not associated with SAH). These groups were compared by measuring of Chi-square contingency table using the program CONTING.

Table 5 shows the results of expected numbers haplotypes using Arlequine and Phase.

**Table 5**: Comparison of Arlequine and Phase results on the numbers of haplotypes present in SAH and control samples.

| Hap | Arlequine | | Phase | |
|---|---|---|---|---|
| | SAH | Control | SAH | Control |
| 111 | 96 | 48 | 98 | 52 |
| 112 | 10 | 14 | 8 | 10 |
| 121 | 26 | 13 | 22 | 10 |
| 122 | 87 | 45 | 91 | 48 |
| 211 | 27 | 18 | 28 | 17 |
| 212 | 28 | 16 | 27 | 17 |
| Total Count | 274 | 154 | 274 | 154 |

The expected results are in Table 6:

**Table 6**: Comparison of Arlequine and Phase **expected results** on the numbers of haplotypes present in SAH and control samples.

| Haplotype | Arlequine | | Phase | |
|---|---|---|---|---|
| | SAH | Control | SAH | Control |
| 111 | 92.19 | 51.81 | 96.03 | 53.97 |
| 112 | 15.36 | 8.64 | 11.52 | 6.48 |
| 121 | 24.97 | 14.03 | 20.49 | 11.51 |
| 122 | 84.50 | 47.50 | 88.99 | 50.01 |
| 211 | 28.81 | 16.19 | 28.81 | 16.19 |
| 212 | 28.71 | 15.83 | 28.17 | 15.83 |

Table 7 bellow describes the chi-square and P value calculated from table 5 (degrees of freedom=5)

**Table 7**: P value and Chi-square results of both Arlequine and Phase

| | **Arlequine** | **Phase** |
|---|---|---|
| **Chi-square** | **6.29** | **3.74** |
| **p value** | **0.279** | **0.587** |

Analysis result showed no significant association of SNPs haplotype with SAH. This was due to non-significant change in frequencies of haplotypes between these groups

## 3.3 Results of DHPLC and DNA sequencing

**Scanning of the elastin gene (34 exons) was performed, for any mutation found sequencing was performed.**

### 3.3.1 Mutations of Exon 20 and Intron 20

Multiple mutations were detected (in **exon and intron 20**) after the performance of DHPLC on samples 2 and 5 (Figure 28 and Figure 29):

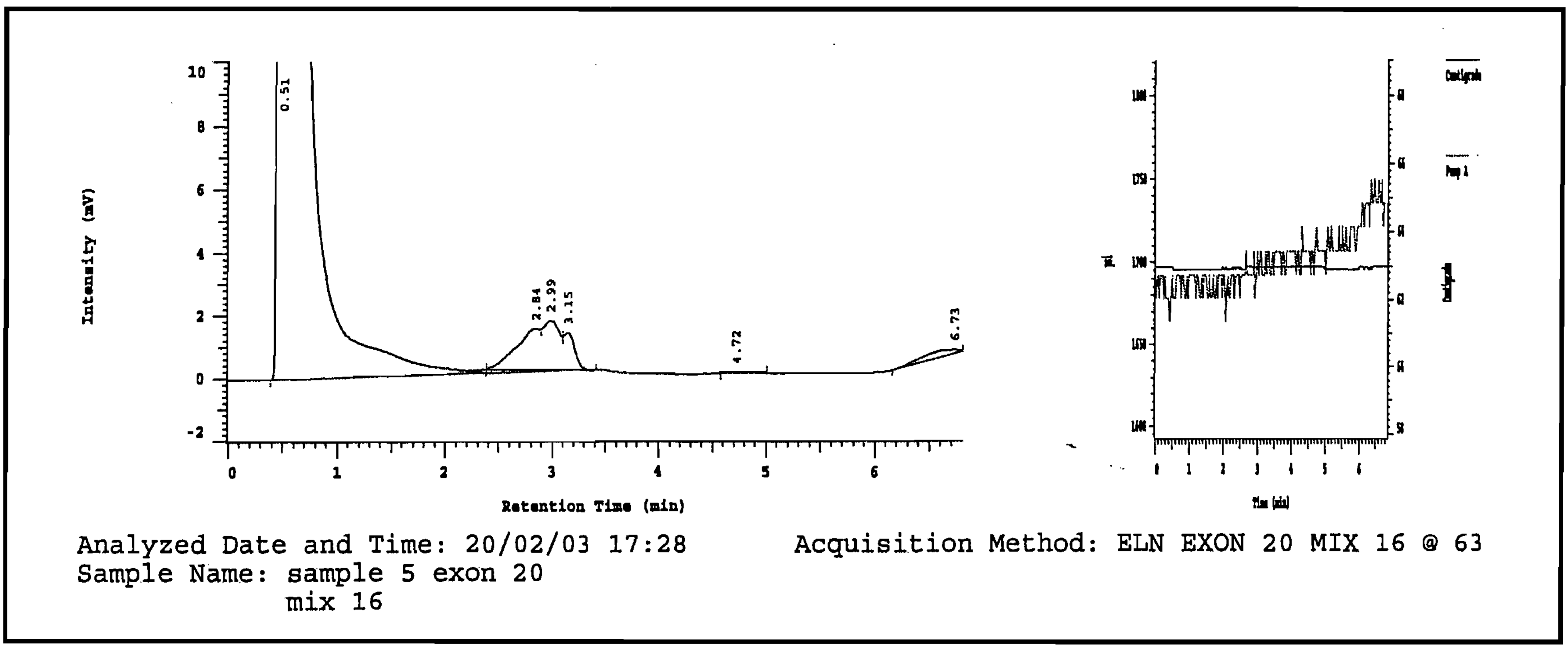



**Figure 28: This figure describes the pattern of exon 20 and intron 20 on sample number 5, method used is *ELN* EXON20MIX16@63.**

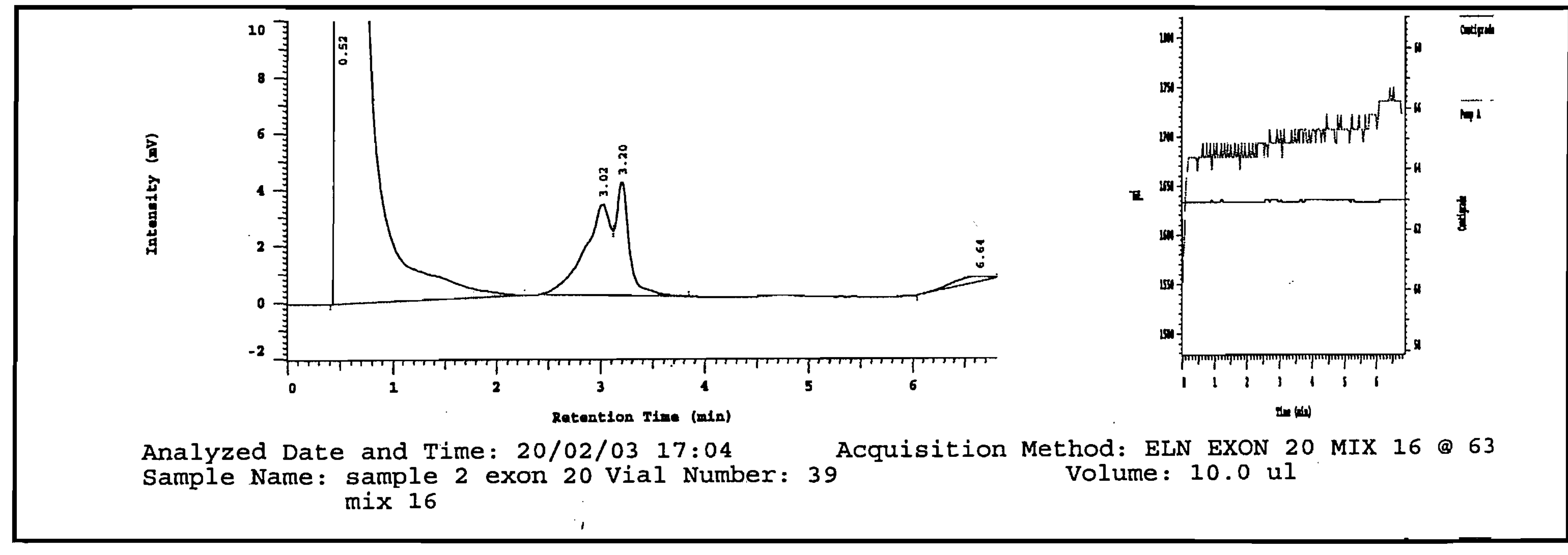



**Figure 29: This figure is the pattern of exon 20 sample number 2, method used is *ELN* EXON20MIX16@63:**

### 3.3.2 IVS20+17 T>C

Sequencing reaction was performed on these two samples 5 and 2 respectively, sample 5 is C/T heterozygous while sample 2 is C/C homozygous.

Position of this SNP is IVS20+17 T>C (ss4044368 T/C rs2856728) Figure 30:

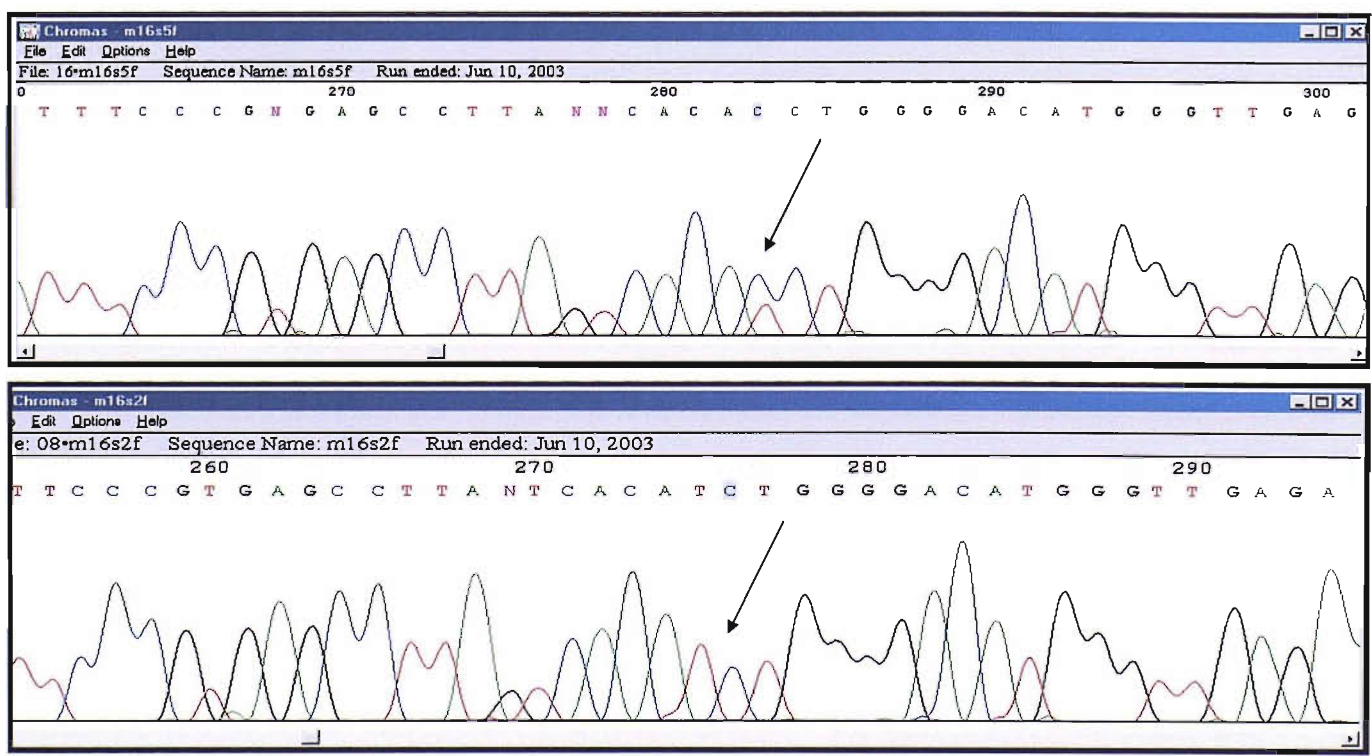



**Figure 30: Sequencing results of sample 5 (heterozygous) and sample two (normal)**

The exact position of this intronic SNP is : C see details in Figure 31

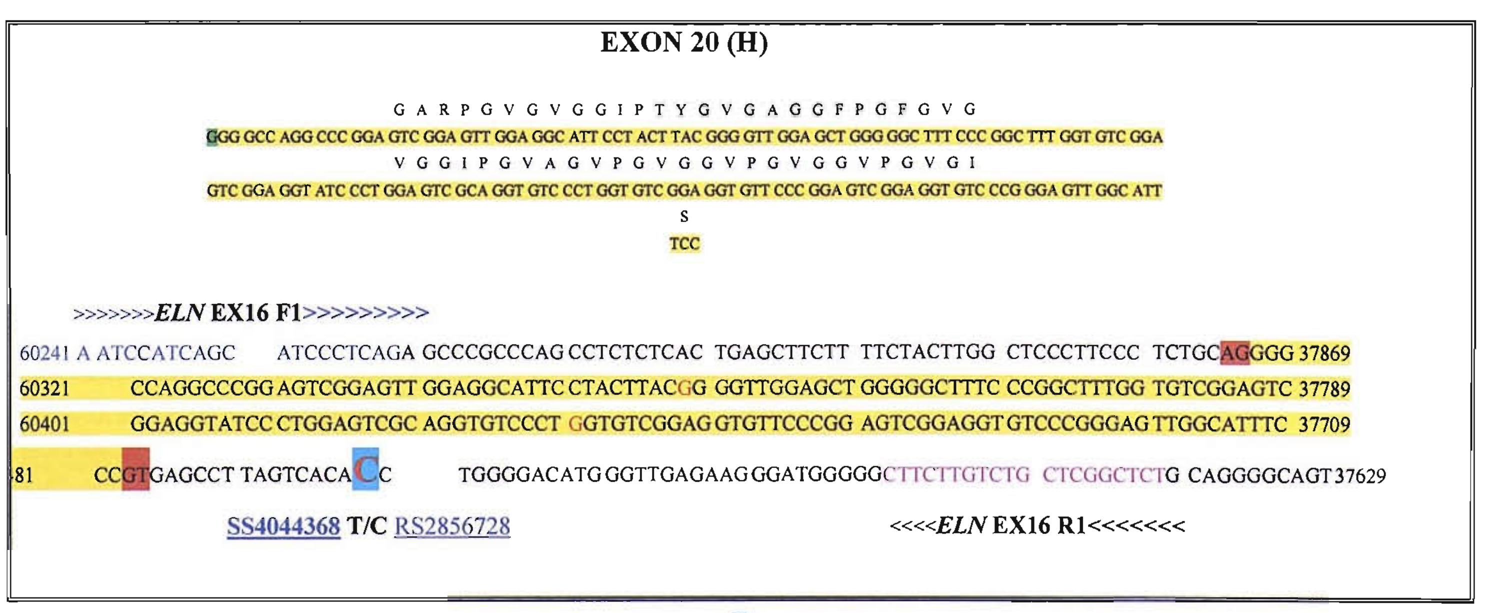



**Figure 31: Sequence obtained from GeneBank AC005056, exact mutation is shown C**

### 3.3.3 AC005056:37759 G>A

The second mutation found in this domain is in sample 2, A/G heterozygous was detected, this SNP is exonic and it is non-synonymous mutation leading to an amino acids substitution of glycine to serine. Reference rs2071307 **G/A** Figure 32 and Figure 33

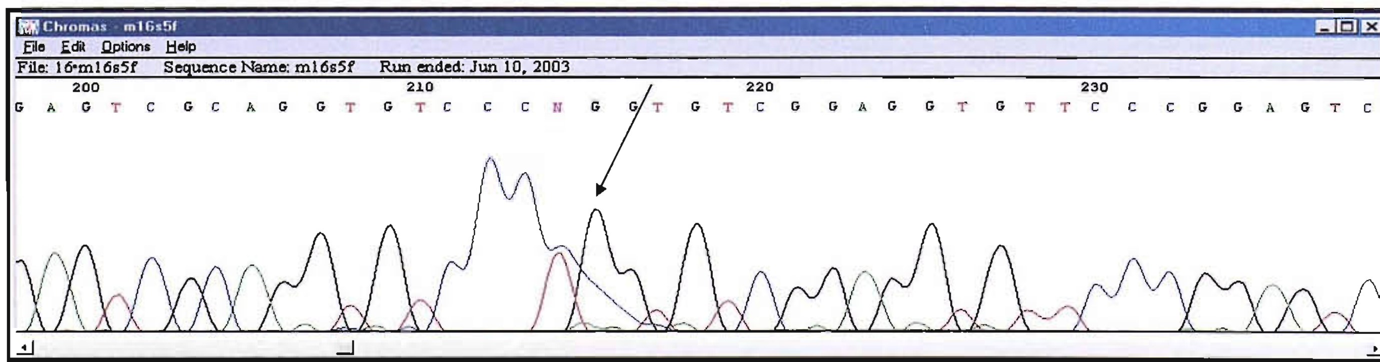



**Figure 32: Sequencing results of sample 5(normal)**

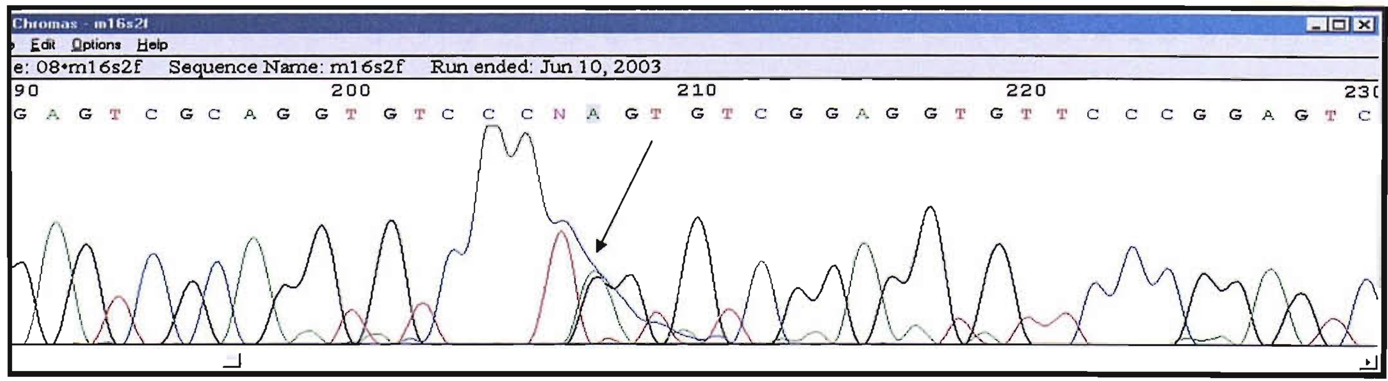



**Figure 33: Sequencing results of sample 2 (heterozygous)**

The reverse sequencing of this mutation is in Figure 34

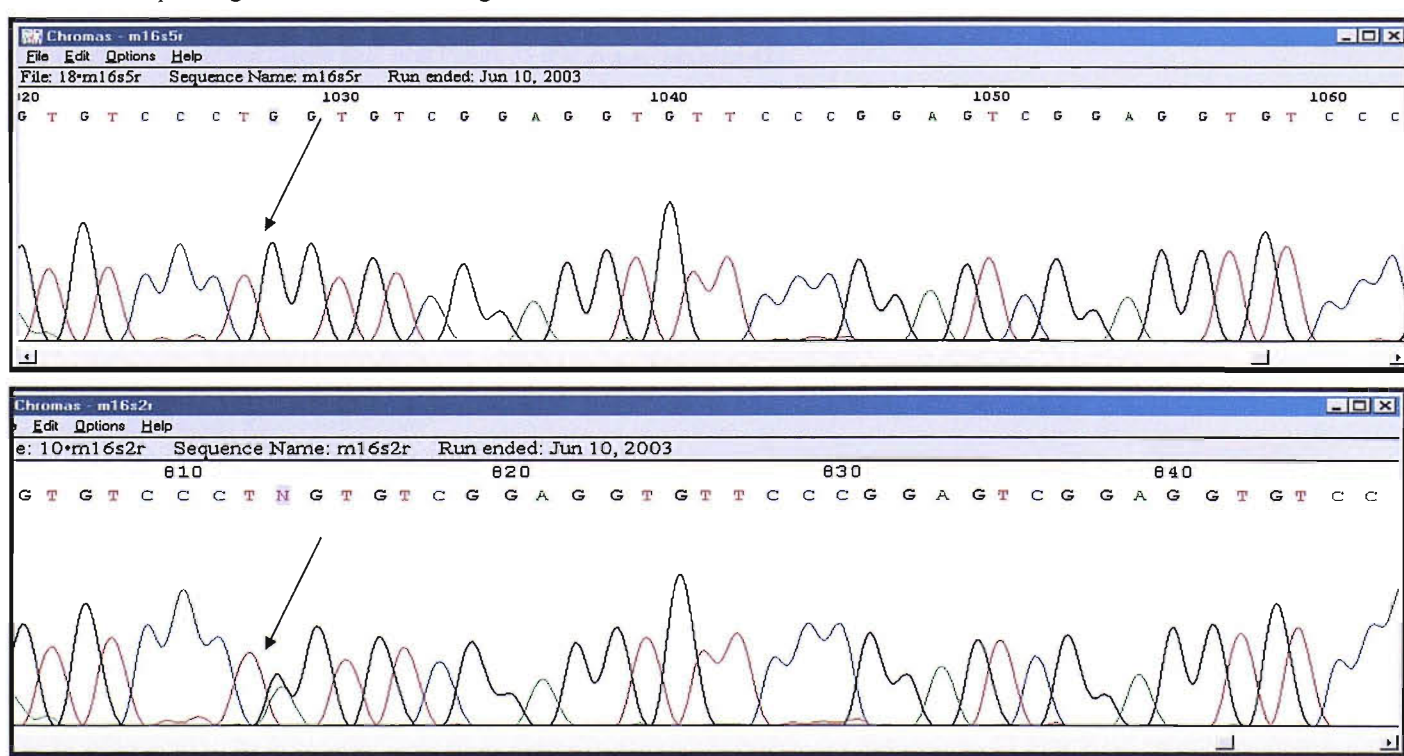



**Figure 34: Sequencing results of sample 5(normal) and sample two (heterozygous)**

The exact location of this exonic SNP is g Figure 35 **it is in a.a. position 422:**

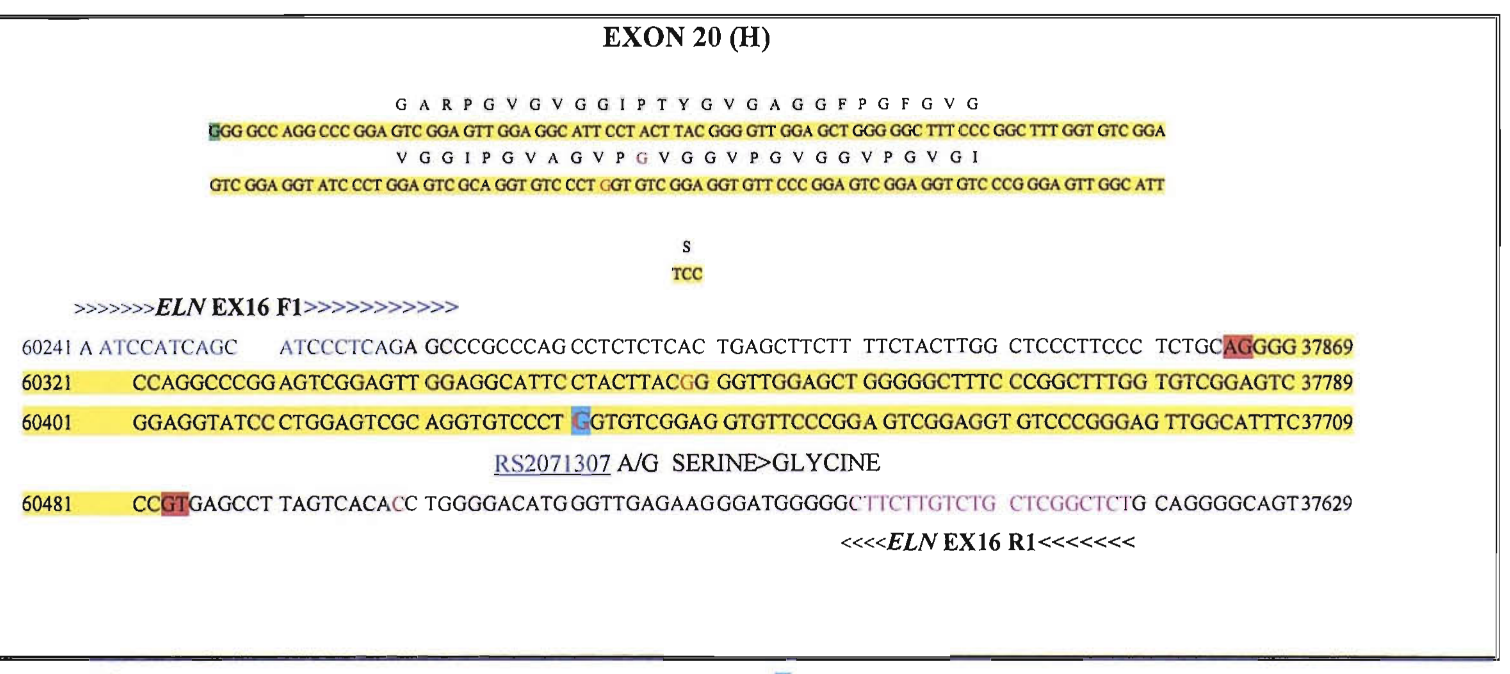



**Figure 35: Sequence obtained from GeneBank AC005056, exact mutation is shown G**

### 3.3.3.1 Mutation analysis in relation to splicing process using Exonic Splicing Enhancers finder ESEfinder[303]

**To see if this mutation has an effect on splicing, *In silico* analysis as performed using ESE finder:**

**AC005056:37759 G>A (rs2071307) amino acids substitution of glycine to serine of the elastin gene.**

**Sequence ID: AC005056 wild type**
**Sequence:**
AGGTGTCCCTGGTGTCGGAG
Length=20
Results of the ESE finder is found in Table 8

**Table 8: Results of the ESE finder on the wild type sequence of the elastin gene**

| | SF2/ASF Thr=1.956 | | | SC35 Thr=2.383 | | | SRp40 Thr=2.67 | | | SRp55 Thr=2.676 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Position** | **Motif** | **Score** | **Position** | **Motif** | **Score** | **Position** | **Motif** | **Score** | **Position** | **Motif** | **Score** |
| | | | 4 | TGTCCCTG | 3.833084 | 6 | TCCCTGG | 3.683126 | | | |

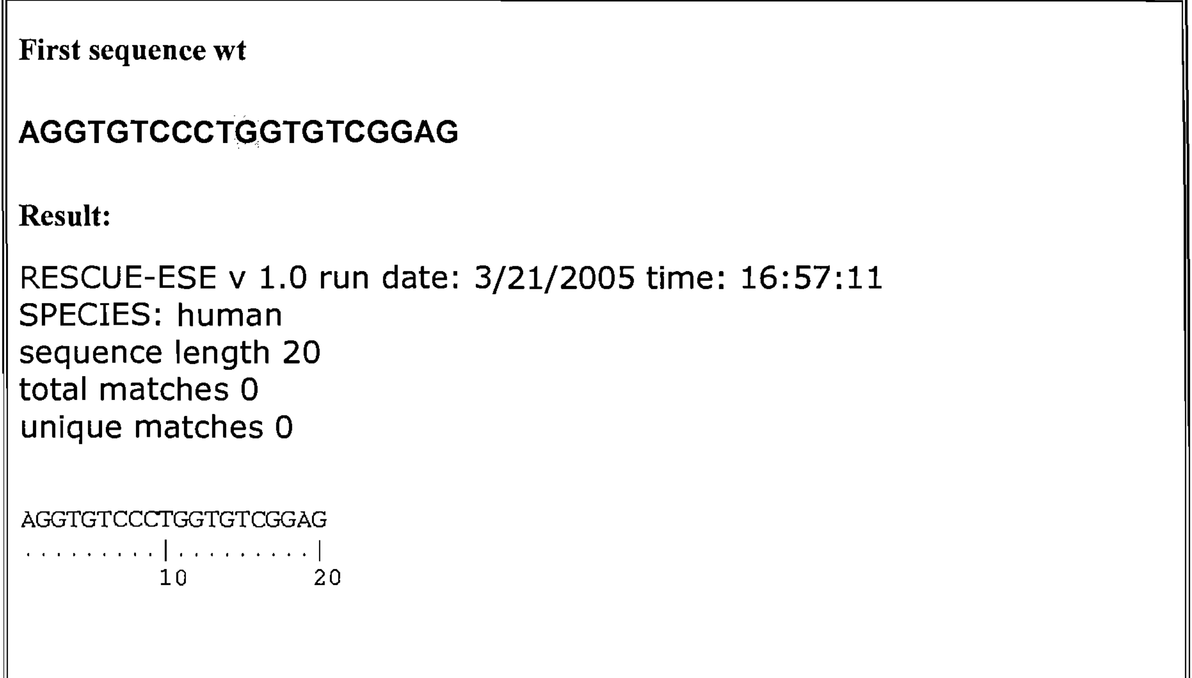
I will insert the mutant sequence in the ESE finder:
**Sequence ID: Mutant AC005056**
**Sequence:**
AGGTGTCCCTAGTGTCGGAG
Length=20
Results of the ESE finder is found in Table 9

**Table 9: Results of the ESE finder on the mutant sequence of the elastin gene**

| | SF2/ASF Thr=1.956 | | | SC35 Thr=2.383 | | | SRp40 Thr=2.67 | | | SRp55 Thr=2.676 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Position** | **Motif** | **Score** | **Position** | **Motif** | **Score** | **Position** | **Motif** | **Score** | **Position** | **Motif** | **Score** |
| | | | 4<br>5 | TGTCCCTA<br>GTCCCTAG | 3.383188<br>3.051854 | 6 | TCCCTAG | 3.115297 | 11 | AGTGTC | 2.697407 |

### 3.3.3.2 RESCUE-ESE analysis

Using another program to predict exonic splicing enhancers in a programme called: RESCUE-ESE Web Server http://genes.mit.edu/burgelab/rescue-ese/ , which is an online tool to annotate exons with Exonic Splicing Enhancers (ESE), using the wild type sequence results shown in Figure 36



**First sequence wt**

**AGGTGTCCCTGGTGTCGGAG**

**Result:**

RESCUE-ESE v 1.0 run date: 3/21/2005 time: 16:57:11
SPECIES: human
sequence length 20
total matches 0
unique matches 0

```
AGGTGTCCCTGGTGTCGGAG
.........|.........|
        10        20
```

**Figure 36: Results of the Exonic Splicing Enhancers (ESE) of the wild type**

When adding the second sequence (mutant) sequence results are the same as the wild type see Figure 37

**AGGTGTCCCTAGTGTCGGAG**

RESCUE-ESE v 1.0 run date: 3/21/2005 time: 16:58:28
SPECIES: human
sequence length 20
total matches 0
unique matches 0

```
AGGTGTCCCTTGTGTCGGAG
.........|.........|
        10        20
```

**Figure 37: Results of the Exonic Splicing Enhancers (ESE) of the mutant sequence.**

### 3.3.3.3 PolyPhen analysis

**Using PolyPhen[304] programme to assess the possible effect on amino acid substitution**

Glycine amino acid substitution with Serine in this domain gave the following results:

## Query

| Acc number | Position | $AA_1$ | $AA_2$ | Description |
|---|---|---|---|---|
| P15502 | 422 | S | G | Elastin precursor (Tropoelastin). LENGTH: 730 AA |

## Prediction

This variant is predicted to be unknown (no data for prediction)

## Details

PSIC PROFILE SCORES FOR TWO AMINO ACID VARIANTS

| Score1 | Score2 | \|Score1-Score2\| | Observations | Diagnostics | Multiple alignm |
|---|---|---|---|---|---|
| N/A | N/A | N/A | 0 | all sequences filtered out | N/A |

MAPPING OF THE SUBSTITUTION SITE TO KNOWN PROTEIN 3D STRUCTURES

| Database | Initial number of structures | Number of structures |
|---|---|---|
| PQS | 147 | 0 |

#### 3.3.3.4 Genotyping of the *ELN* exon 20 Gly224Ser SNP on the BWHHS cohort (results)

Results of genotyping was going with HW equilibrium (see **Appendix J**)
In my study the frequency of my mutant allele is A is 39 %.

Three models proposed on the cohort are suggested, each proposed model is supported by references below:

**Dominant negative model:** references for Intracranial aneurysms are [24,69,305,306] and for abdominal aortic aneurysms and/or thoracic aortic aneurysms are [3,86,201,238].

**Additivity model**: good for complex diseases, for intracranial aneurysms papers supporting this model are [307,308].

**Recessive model** references in Intracranial aneurysms are [2,102,309,310] and for abdominal aortic aneurysms and/or thoracic aortic aneurysms are [311,312].

MFS (intron 28) is associated with high pulse pressure, some papers suggests that some fibrillin-1 mutations may leads to high blood pressure via increased arterial stiffness, hence may in aneurysmal diseases[18,313]. I am looking for the effect of this SNP (in the elastin gene) on blood pressure phenotype, especially the pulse; systolic and diastolic pressure and stroke patients. Results of the exonic elastin SNP on the BWHHS are shown in Table10.

**Table 10: Association of *Elastin* (rs2071307) with blood pressure, pulse pressure and stroke**

| | Mean (SD) or N (%) by genotype | | | P anova[a] | P trend[b] |
|---|---|---|---|---|---|
| | 11<br>N = 1070 | 12<br>N = 1315 | 22<br>N = 445 | | |
| Systolic (mmHg) | 147.6 (25.5) | 146.9 (25.3) | 145.5 (25.7) | 0.3 | 0.2 |
| Diastolic (mmHg) | 79.9 (11.7) | 79.1 (11.6) | 79.1 (12.0) | 0.2 | 0.1 |
| Pulse pressure (mmHg) | 67.7 (19.4) | 67.9 (19.1) | 66.4 (19.0) | 0.4 | 0.3 |
| Stroke N (%) | 61 (5.7) | 77 (5.9) | 39 (8.8) | 0.07 | 0.05 |

[a] Testing null hypothesis of no difference between any of three categories
[b] Testing null hypothesis of linear trend across the three categories

Notes:
Stroke includes prevalent and incident cases combined. Strangely, blood pressure seems to reduce with each addition of minor allele but stroke shows an increase.
For information. per allele, odds ratio for stroke = 1.24 (1.00, 1.53)
HWE exact test p = 0.2
*"Regret data not currently available* for how does the elastin SNP correlate with different types of stroke"

### 3.3.3.5 Association Results of the sporadic SAH and control using *ELN* exonic SNP

Three models were tested in this association analysis, 138-sporadic samples and 77-control sample were used. Table 11 shows the dominant model

**Table 11: Analysis of sporadic SAH vs. control samples on the exonic SNP using dominant model.**

| HOMOZYGOUS DOMINANT | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observed | | | | | EXPECTED | | | |
| | 11 | NON 11 | Total | | | 11 | NON 11 | Total |
| Sporadic | 48 | 90 | 138 | | sporadic | 49.42 | 88.58 | 138.0( |
| Control | 29 | 48 | 77 | | control | 27.58 | 49.42 | 77.00 |
| Total | 77 | 138 | 215 | | Total | 77.00 | 138.00 | 215.0( |
| X^2 | 0.177741 | | | | | | | |
| P- VALUE | 0.673322 | | | | | | | |

Table 12 shows the recessive model.

**Table 12: Analysis of sporadic SAH vs. control samples on the exonic SNP using recessive model**

| HOMOZYGOUS RECESSIVE | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observed | | | | | EXPECTED | | | |
| | 22 | NON 22 | Total | | | 22 | NON 22 | Total |
| Sporadic | 24 | 114 | 138 | | sporadic | 21.82 | 116.18 | 138.00 |
| Control | 10 | 67 | 77 | | control | 12.18 | 64.82 | 77.00 |
| Total | 34 | 181 | 215 | | Total | 34.00 | 181.00 | 215.00 |
| X^2 | 0.738202 | | | | | | | |
| P- VALUE | 0.390237 | | | | | | | |

Table 13 shows the additivity model.

**Table 13: Analysis of sporadic SAH vs. control samples on the exonic SNP using additivity model**

| ADDITIVITY | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observed | | | | | EXPECTED | | | |
| | ALLELE 1 | NON ALLELE 1 | Total | | | ALLELE 1 | NON ALLELE 1 | Total |
| Sporadic | 162.00 | 114 | 276 | | sporadic | 165.60 | 110.40 | 276.00 |
| Control | 96 | 58 | 154 | | control | 92.40 | 61.60 | 154.00 |
| Total | 258 | 172 | 430 | | Total | 258.00 | 172.00 | 430.00 |
| | | | | | | | | |
| X^2 | 0.548055 | | | | | | | |
| P- VALUE | 0.459113 | | | | | | | |

Obtaining the control results from BWHHS (since the control number is 77 samples, the BWHHS contains 2645 non-stroke samples which can be used as a normal control) and comparing it with our sporadic SAH results shown in Table 14.

**Table 14: Analysis of sporadic SAH vs. control samples of BWHHS on the exonic SNP using dominant model**

| SPORADIC *vs.* CONTROL WOMEN OF THE BWHHS | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observed | | | | | EXPECTED | | | |
| | ALLELE 22 | NON ALLELE 22 | Total | | | ALLELE 22 | NON ALLELE 22 | Total |
| Sporadic | 24 | 114 | 138 | | sporadic | 21.25 | 116.75 | 138.00 |
| BWHHS | 406 | 2248 | 2654 | | control | 408.75 | 2245.25 | 2654.00 |
| Total | 430 | 2362 | 2792 | | Total | 430.00 | 2362.00 | 2792.00 |
| | | | | | | | | |

### 3.3.4 Mutation Intron 19 of the *ELN* gene (C/T rs2239691)

It appears that a heterozygous mutation is present in exon 19, the pattern of DHPLC for samples 1,2 and 3 are in Figure 38:

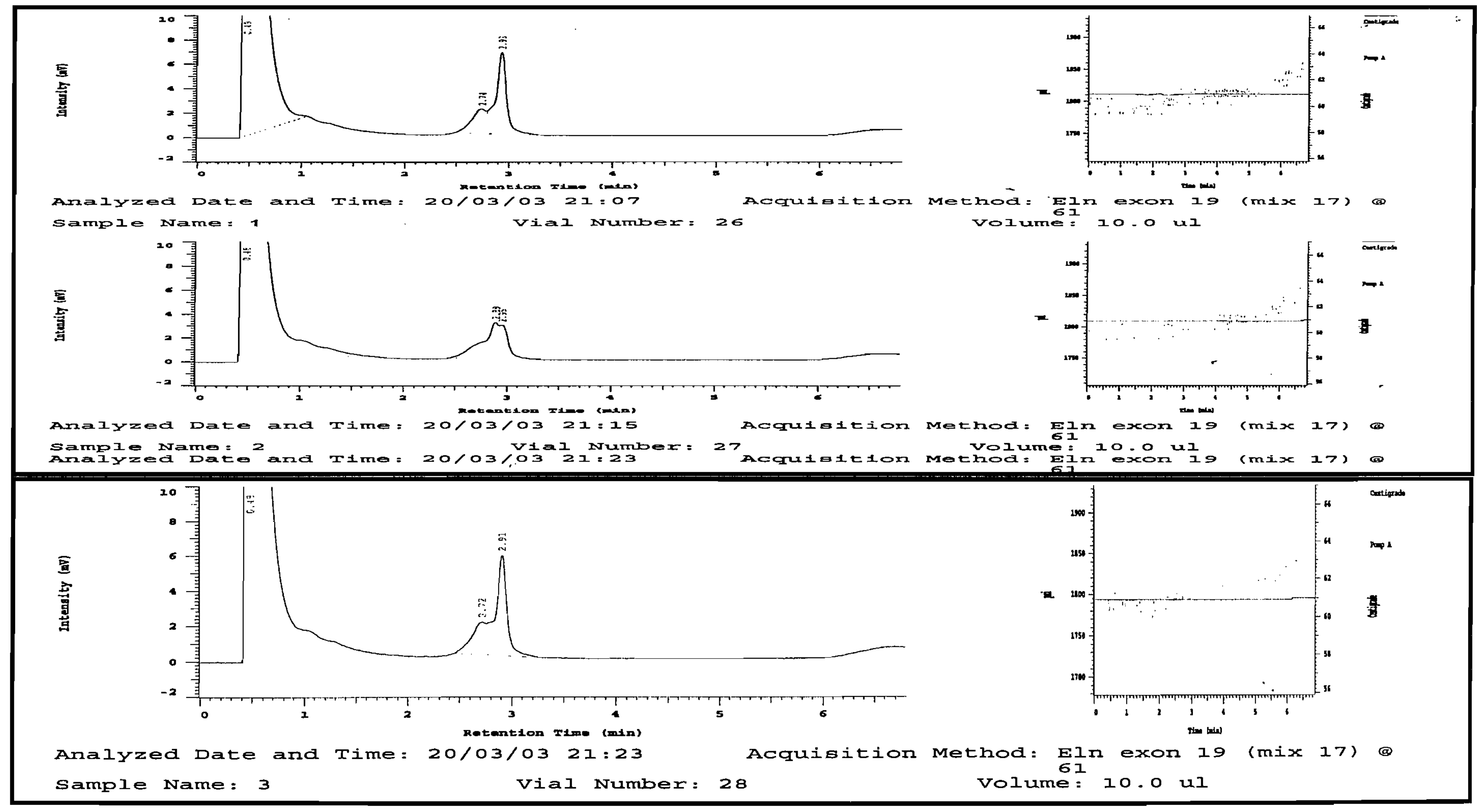



**Figure 38: DHPLC for samples 1,2 and 3.**

After DNA sequencing of these samples it appears that: Sample 1 (homozygous T/T Genotype), sample 2 (C/T genotype), Sample 3(homozygous C/C Genotype). Reference ss3195004 C/T and ss4044366 **T/C** Figure 39

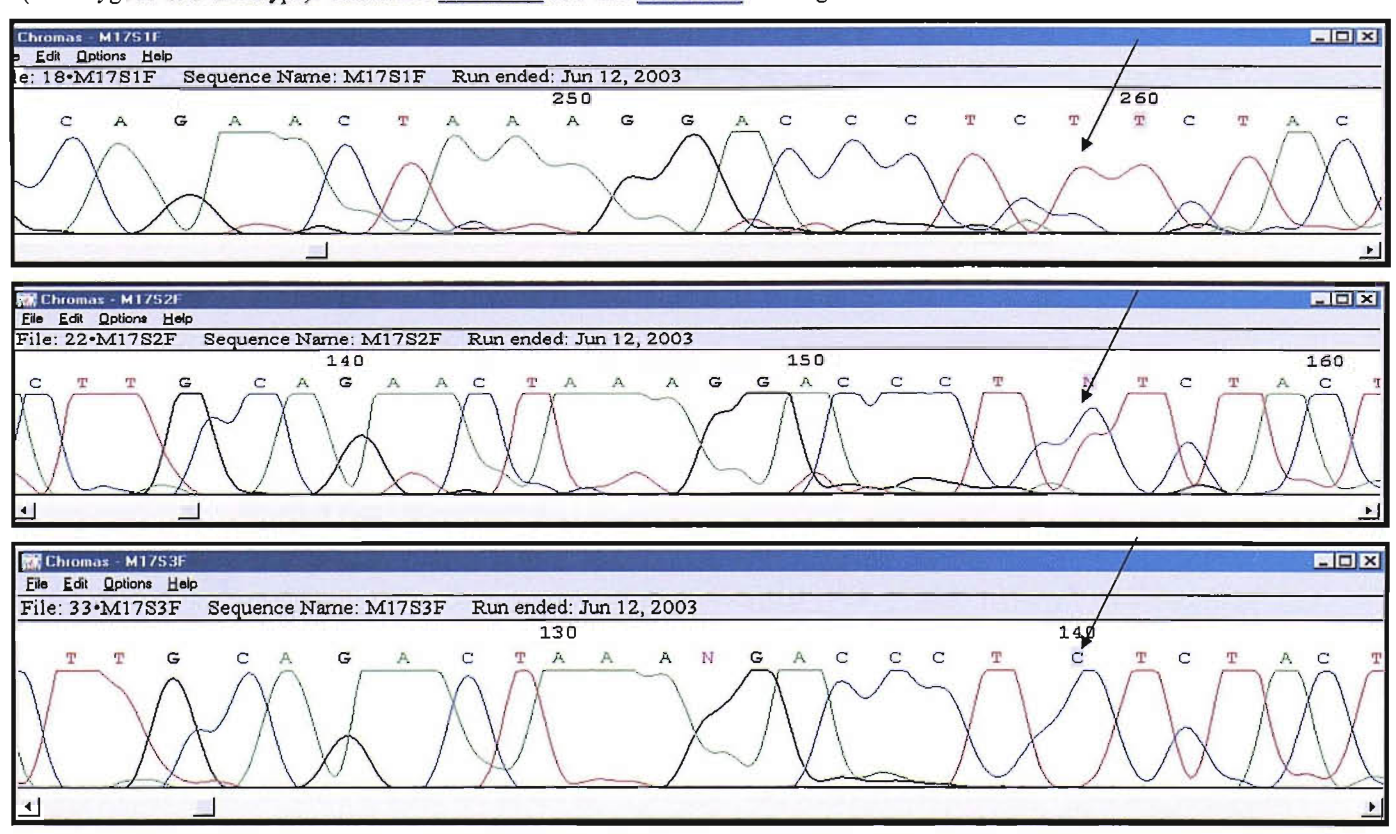



**Figure 39: Sequencing results of sample 3(normal); sample 2 (heterozygous) and sample 1 (homozygous mutant).**

The exact location of this intronic SNP is below: c see Figure 40

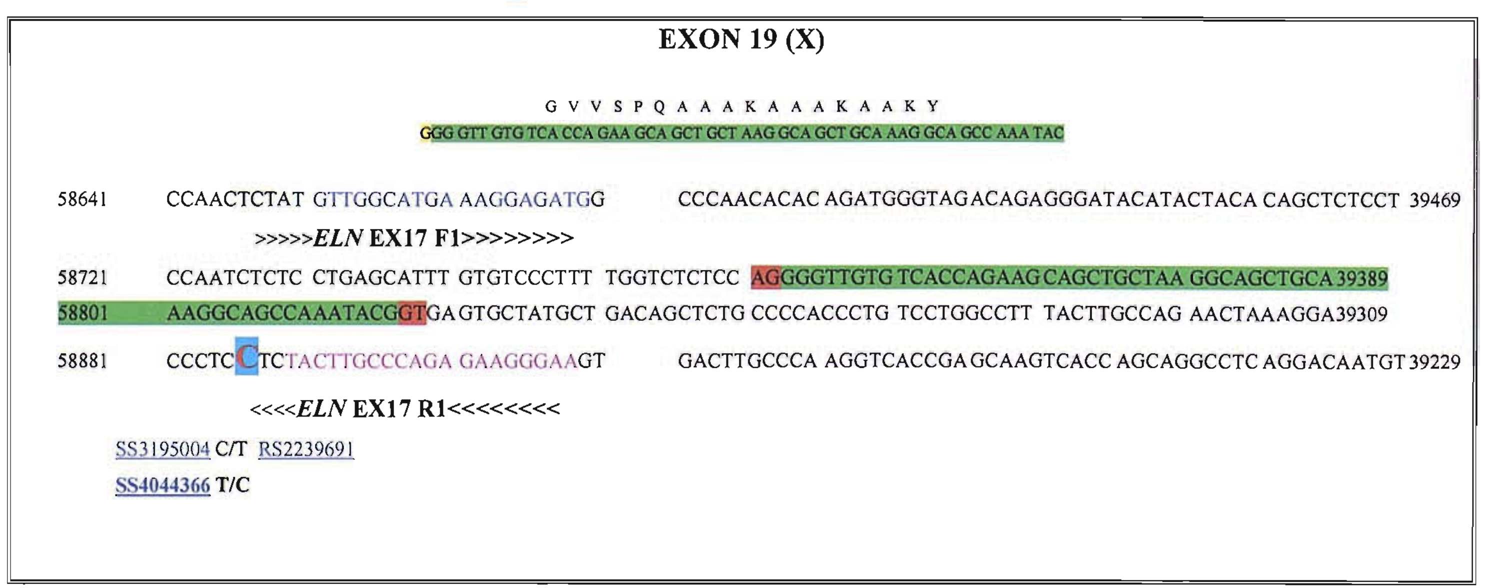



**Figure 40: Sequence obtained from GeneBank AC005056 of the *ELN* gene , exact mutation is shown (C)**

### 3.3.5 Exon 33 MIX 2 of the *ELN* gene (ss4943619 C/T rs3757587)

On familial samples with positive DHPLC on **exon 33 mix 2,** the following results are shown Figure 41:

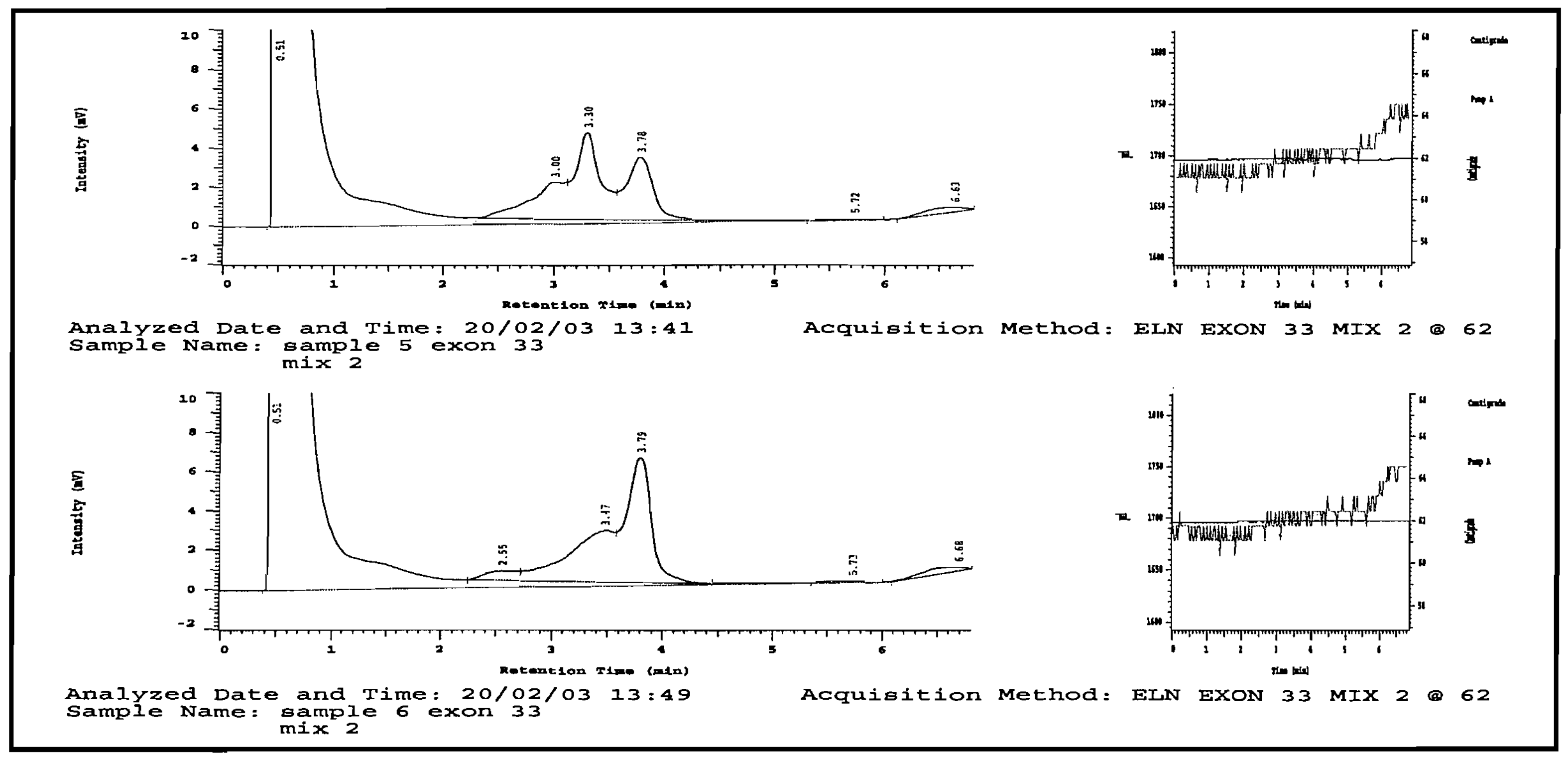



**Figure 41: DHPLC for samples 2 and 6.**

Sequencing was performed, results: ss4943619 C/T rs3757587 Figure 42:

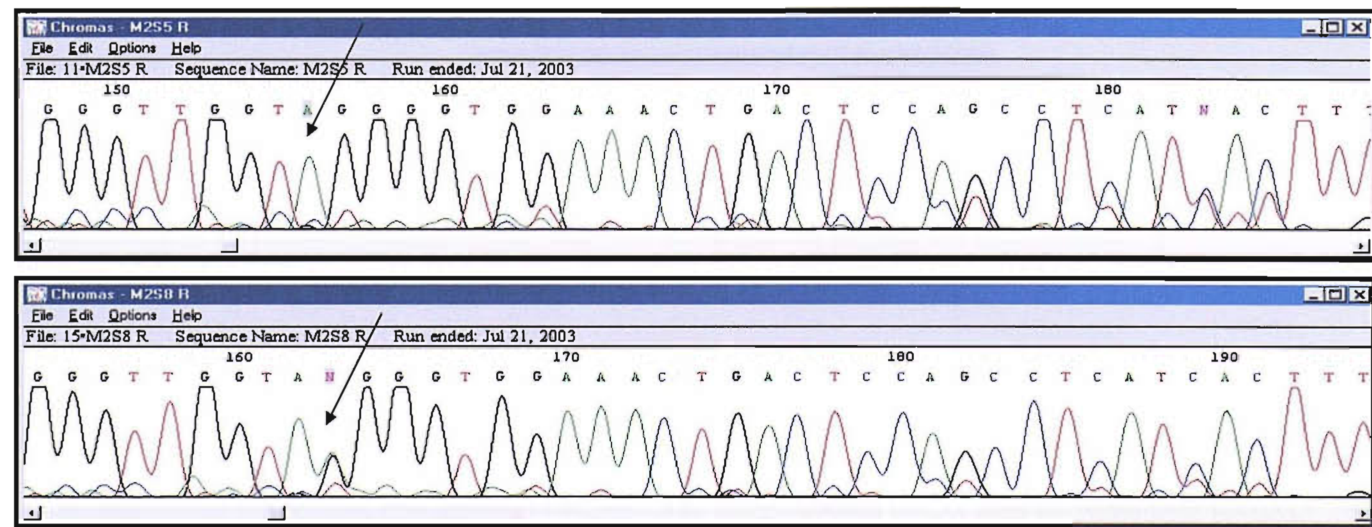



**Figure 42: Sequencing results of sample above (normal) and sample down (heterozygous)**

The exact position of this mutation is in Figure 43:

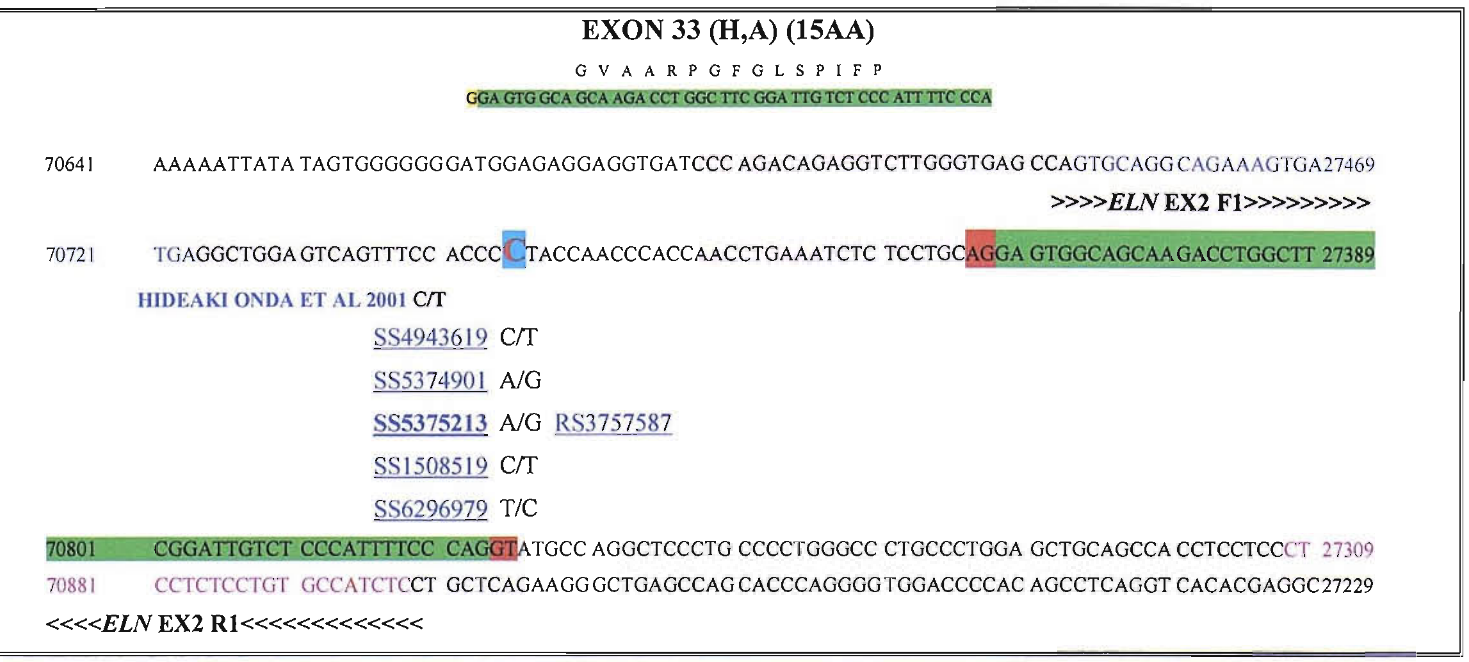



**Figure 43: Sequence obtained from GeneBank AC005056, exact mutation is shown C**

## 3.3.6 Exon 18 (mix 18) mutations of the *ELN* gene

Small variation was detected Figure 44:

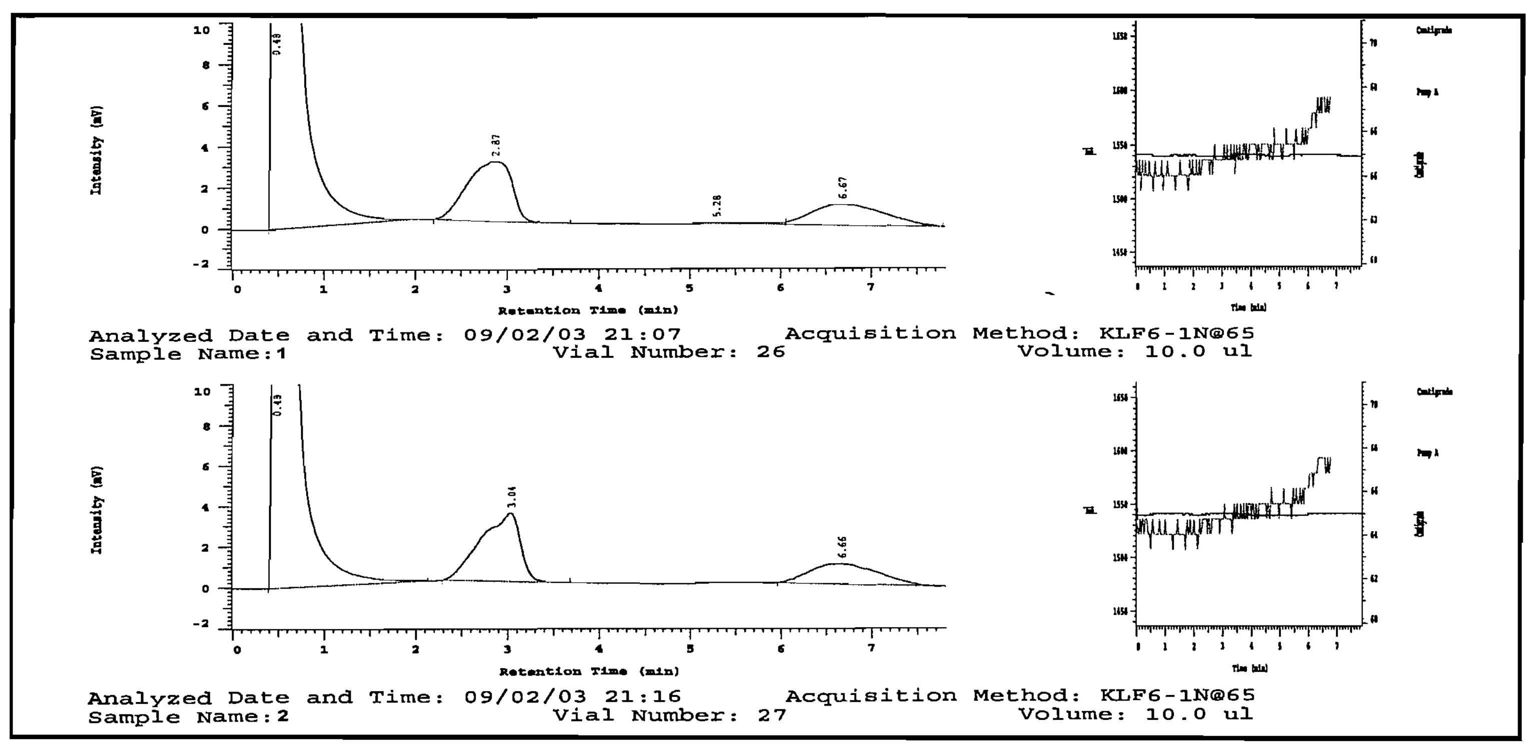



**Figure 44: DHPLC for samples 1 and 2, small difference can be seen.**

Sequencing results showed the presence of the following mutation:

**IVS18+47 G>C:**

Sequencing of sample 1 and sample 5 showed in Figure 45:

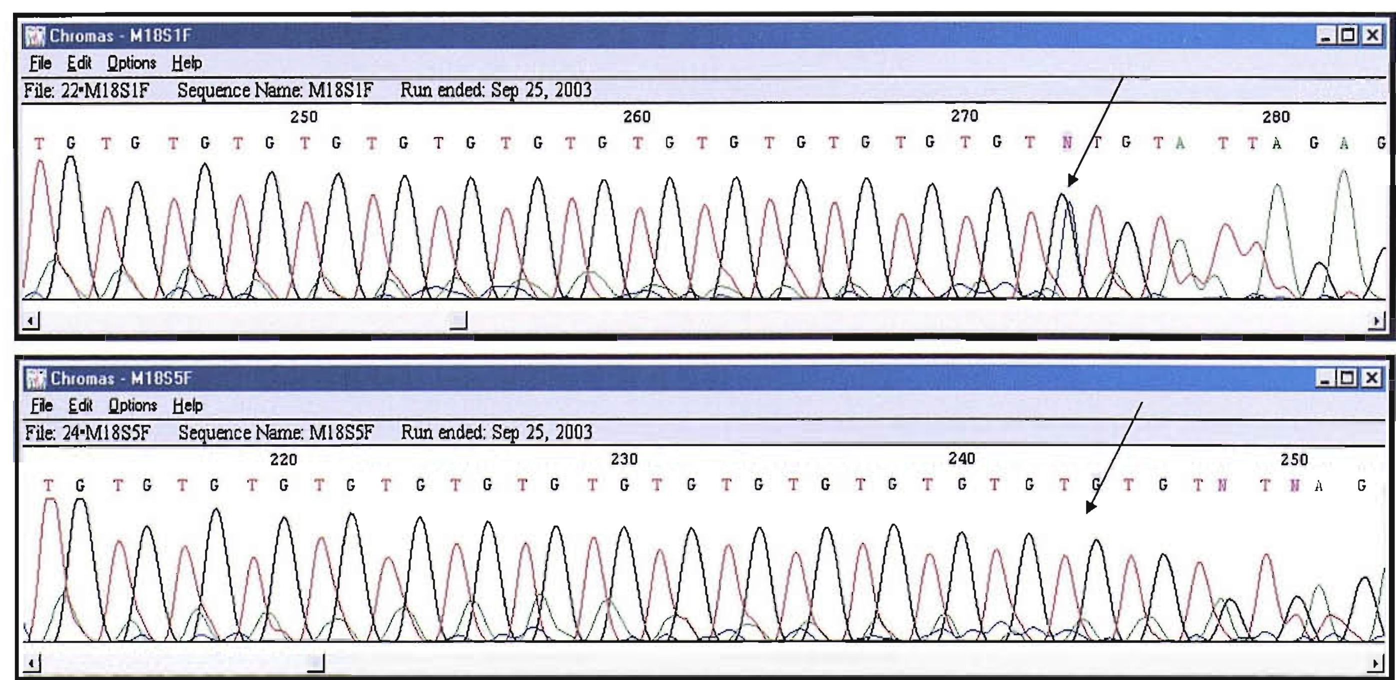



**Figure 45: Sequencing results of sample 5(normal) and sample one (heterozygous)**

The exact position of this mutation is Figure 46:

**EXON 18(H)(49AA)**

G A A A G L V P G G P G F G P G V V G V P G A G V P G
GGA GCT GCT GCA GGC TTA GTG CCT GGT GGG CCA GGC TTT GGC CCG GGA GTA GTT GGT GTC CCA GGA GCT GGC GTT CCA GGT

V G V P G A G I P V V P G A G I P G A A V P
GTT GGT GTC CCA GGA GCT GGG ATT CCA GTT GTC CCA GGT GCT GGG ATC CCA GGT GCT GCG GTT CCA

57041 TCTTAGTCTC TCCACATCTC TCTGATGAGT AGGATCCATG CAGAGGAAATGTCAACCCAC CTGCAATCCT GCATTCAGGA 41069

**>>>*ELN* EX18 F1>>>**

57121 CCAACTGTCA CTTCCATACT CTACTAACCA CCCTTCTAGC CCCTCTGAGG TTCCCATAGG TTAGGGGAAC AATGCTTTTT 40989

57201 CTTCCACAGG AGCTGCTGCA GGCTTAGTGC CTGGTGGGCC AGGCTTTGGC CCGGGAGTAGTTGGTGTCCC AGGAGCTGGC40909

57281 GTTCCAGGTG TTGGTGTCCC AGGAGCTGGGATTCCAGTTG TCCCAGGTGC TGGGATCCCA GGTGCTGCGG TTCCAGGTGA 40829

57361 GCTGGGCTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTATTA GAGAGAAATA TTGAGACTATTGCCAAAATT 40749

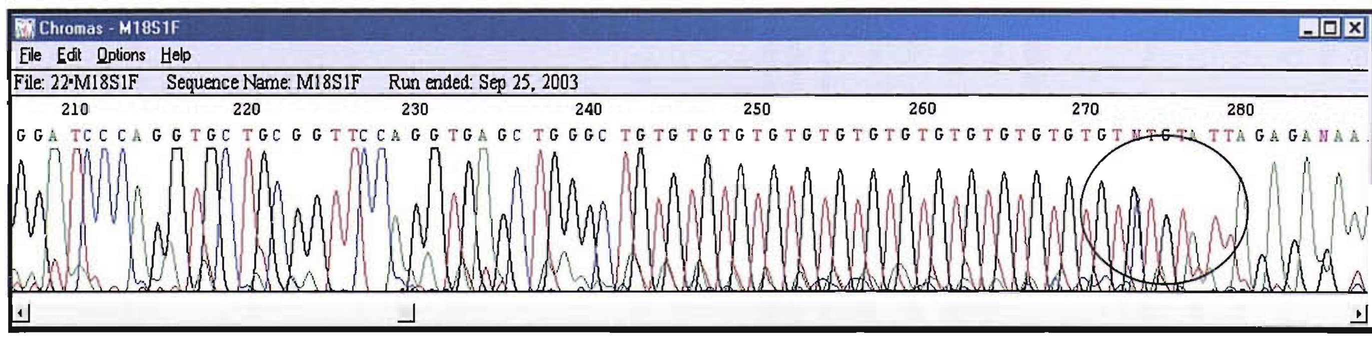
SS8153909 -/GT RS5884930 G/C WASEEM

57441 TTTGCATTCT CCCTAACACC ATAACCATCT GCCCATACCC TTGACCACGT CTCATCCCCT CATCTTCTCT TCCTTGGGCT

**Figure 46: Sequence obtained from GeneBank AC005056, exact mutation is shown G**

### 3.3.7 IVS18+20DEL2 (rs5884930 –GT) of the *ELN* gene.

Sequencing results concerning the GT repeats are: Genotype 17:17 GT repeats and Genotype 18:19 GT repeats respectively Figure 47:



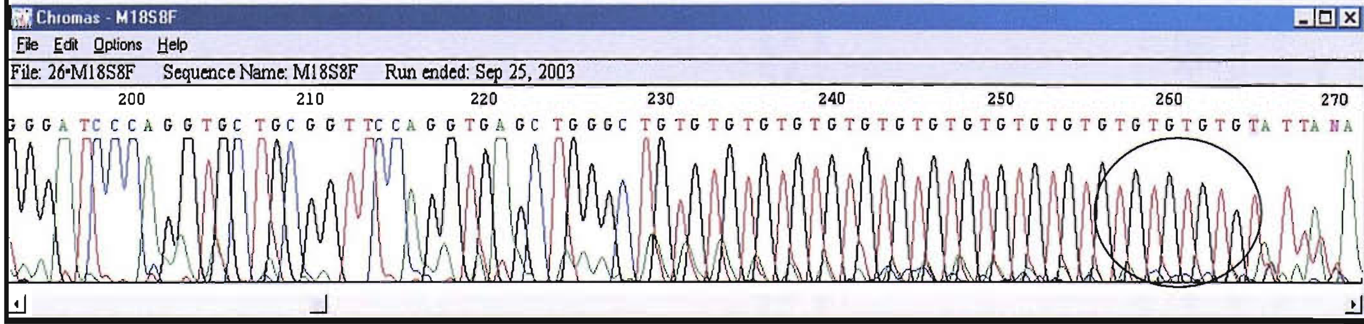




**Figure 47: Sequencing results of GT repeats are different in sample 1 and sample 8.**

### 3.3.7.1 Results of NNSPLICE 0.9 [314] on the IVS18+20DEL2 (rs5884930 – GT) of the *ELN* gene.

Splice site predictions for GT wild type sequence with donor score cut-off 0.40, acceptor score cut-off 0.40 (exon/intron boundary shown in larger font), see Figure 48, when adding 1 to 100 repeat of GT to the sequence I did not have any changes in the scores, the true donor and acceptor sites are marked.

**Acceptor site predictions for gi:**

| Start | End | Score | Intron Exon |
|---|---|---|---|
| 99 | 139 | 0.55 | ctctactaaccacccttctagcccctctgaggttcccatag |
| 119 | 159 | 0.85 | gcccctctgaggttcccatagg ttaggggaacaatgctttt |
| 149 | 189 | 0.97 | acaatgctttttcttccac**ag**gagctgctgcaggcttagtg |
| 266 | 306 | 0.67 | ctgggattccagttgtcccaggtgctgggatcccaggtgct |

**Donor site predictions for gi:**

| Start | End | Score | Exon Intron |
|---|---|---|---|
| 133 | 147 | 0.83 | cccataggttagggg |
| 295 | 309 | 0.51 | atcccaggtgctgcg |
| 310 | 324 | 0.98 | gttccag**gt**gagctg |

**Figure 48: Scores results of putative donor and acceptor sites for the wild type sequence, actual sites are marked, no changes where found when adding 1-100 GT repeats (complete sequence is preset in Figure 46)**

### 3.3.7.2 Results of GeneScan Test (*ELN*)

GeneScan results were summarised in two different tables, Table 15 describes chromosomes with allele and Table 16, which describes genotypes).

**Table 15: Summarise GeneScan results depending on the GT repeats chromosome counts**

| | All | SAH FAM | SAH Sporadic | Mixed samples | Control HI |
|---|---|---|---|---|---|
| ChromosomeGT 17 | 219 | 11 | 120 | 18 | 70 |
| ChromosomeGT 18 | 9 | 2 | 6 | 0 | 1 |
| ChromosomeGT 19 | 76 | 2 | 44 | 9 | 21 |
| ChromosomeGT 20 | 60 | 0 | 37 | 3 | 20 |
| ChromosomeGT 21 | 9 | 0 | 6 | 0 | 3 |
| ChromosomeGT 22 | 9 | 1 | 6 | 1 | 1 |
| ChromosomeGT 23 | 4 | 0 | 1 | 1 | 2 |
| TOTAL | 386 | 16 | 220 | 32 | 118 |
| No. of non GT 17 | 167 | 5 | 100 | 14 | 48 |

**Table 16: Summarise GeneScan results depending on the count of genotypes.**

| | All | SAH FAM | SAH Sporadic | Different | control HI |
|---|---|---|---|---|---|
| Genotype 17:17 | 64 | 4 | 36 | 4 | 20 |
| Genotype 19:19 | 10 | 0 | 8 | 0 | 2 |
| Genotype 17:18 | 4 | 1 | 3 | 0 | 0 |
| Genotype 17:19 | 36 | 1 | 16 | 7 | 12 |
| Genotype 17:20 | 39 | 0 | 22 | 2 | 15 |
| Genotype 17:21 | 4 | 0 | 3 | 0 | 1 |
| Genotype 17:22 | 5 | 1 | 3 | 1 | 0 |
| Genotype 17:23 | 3 | 0 | 1 | 0 | 2 |
| Genotype 18:19 | 4 | 1 | 2 | 0 | 1 |
| Genotype 18:20 | 1 | 0 | 1 | 0 | 0 |
| Genotype 19:20 | 10 | 0 | 6 | 1 | 3 |
| Genotype 19:21 | 3 | 0 | 2 | 0 | 1 |
| Genotype 19:22 | 2 | 0 | 2 | 0 | 0 |
| Genotype 19:23 | 1 | 0 | 0 | 1 | 0 |
| Genotype 20:20 | 4 | 0 | 3 | 0 | 1 |
| Genotype 20:21 | 1 | 0 | 1 | 0 | 0 |
| Genotype 20:22 | 1 | 0 | 1 | 0 | 0 |
| Genotype 21:22 | 1 | 0 | 0 | 0 | 1 |
| TOTAL GENOTYPES | 193 | 8 | 110 | 16 | 59 |
| TOTAL non 17:17 | 129 | 4 | 74 | 12 | 39 |
| Total 17 allele genotypes | 155 | 7 | 84 | 14 | 50 |
| Total non17 allele genotypes | 38 | 1 | 26 | 2 | 9 |
| Total Chromosomes | 386 | 16 | 220 | 32 | 118 |

### 3.3.7.3 Results of Models proposed depending on GeneScan (*ELN*)

Five models were proposed to be tested using familial SAH, sporadic SAH, and control samples.

### 3.3.7.4 Allele Counts /Additivity Model results (*ELN*)

Looking for allele 17 (17 GT repeat) vs. non-17. Does the presence of allele 17 contribute to SAH? Results are present in Table 17:

**Table 17: Results of allele count, p value and chi square are present.**

| | 17 GT | NON 17GT | P, X^2, Yates |
|---|---|---|---|
| SAH Sporadic | 120 | 100 | $\chi2$ = 0.714<br>P = 0.398<br>Yates<br>$\chi2$ =0.53<br>P = 0.466 |
| Control | 70 | 48 | |
| SAH Familial | 11 | 5 | $\chi2$ = 0.538<br>P = 0.463<br>Yates<br>$\chi2$ =0.20<br>P=0.651 |
| Control | 70 | 48 | |

### 3.3.7.5 Major Expansion Model (Anticipation) (*ELN*)

Looking for the expected value of sporadic SAH *vs*. control (Table 18), are the real results the same as the expected?

**Table 18: Looking at the expected homozygous numbers of allele 17 in sporadic SAH and control.**

| OBSERVED | 17:17 GT | Non 17:17 |
|---|---|---|
| Sporadic SAH | 36 | 74 |
| Control | 20 | 39 |
| | | |
| EXPECTED | 17:17 GT | Non 17:17 |
| Sporadic SAH | 36.45 | 73.55 |
| Control | 19.55 | 39.45 |

Looking for the expected value of familial SAH vs. control see Table 19

**Table 19: Looking at the expected homozygous numbers of allele 17 in familial SAH and control.**

| OBSERVED | 17:17 GT | Non 17:17 |
|---|---|---|
| Familial SAH | 4 | 4 |
| Control | 20 | 39 |
| | | |
| EXPECTED | 17:17 GT | Non 17:17 |
| Familial SAH | 2.87 | 5.13 |
| Control | 21.13 | 37.87 |

### 3.3.7.6 Loss of Heterozygosity Model (*ELN*)

Testing wither I have genotype 17:17 as expected, using the results present in Table 18 and Table 19.

The expected results are almost the same as the real ones.

### 3.3.7.7 Recessive Model (*ELN*)

The result of this model goes with the expansion model as described in Table 18 and Table 19. Furthermore, the results of chi-square and p-value are the following:

**Sporadic SAH vs. control**

Chi-square = **0.0237** p value = **0.877**

When using Yates formula:

Chi-square =0.00 p value =1.00

**Familial SAH vs. control**

Chi-square = 0.766 p value = 0.381

When using Yates formula:

Chi-square =0.25 p value =0.618

### 3.3.7.8 Dominant Model (*ELN*)

Investigation of the presence of allele 17 in genotypes against none 17 genotypes Table 20:

**Table 20: Sporadic SAH vs. Control**

| Observed | Genotypes with 17 | Genotypes without 17 | | Expected | Genotypes with 17 | Genotypes without 17 |
|---|---|---|---|---|---|---|
| Sporadic SAH | 84 | 26 | | Sporadic SAH | 87.22 | 22.78 |
| Control | 50 | 9 | | Control | 46.78 | 12.22 |

**Chi-square** = 1.706 **p value** = 0.192

When using Yates formula:

**Chi-square = 1.17 p value = 0.278**

Investigation of the presence of allele 17 in genotypes against none 17 genotypes in familial vs. control Table 21:

**Table 21: Familial SAH vs. control**

| Observed | Genotypes with 17 | Genotypes without 17 | | Expected | Genotypes with 17 | Genotypes without 17 |
|---|---|---|---|---|---|---|
| Familial SAH | 7 | 1 | | Familial SAH | 6.81 | 1.19 |
| Control | 50 | 9 | | Control | 50.19 | 8.81 |

**Chi-square** = 0.044 **p value** = 0.834

When using Yates formula:

**Chi-square = 0.00 p value = 1.00**

Performing search on tri nucleotide in the Elastin gene, searching results showed the following:

1- (AAT)/(ATT)*5 repeats:

This repeat was found in -12395 position from the start codon, one of the triplets does not agree.

2- (AGG)/ (CCT)*nine repeats and *six repeats:

IVS23+385 and IVS23+373 both are interrupted by other sequences, no more than four consecutive repeats without disruption.

3- (AAG)/ (CTT)*19 and *five repeats:

IVS23+523; IVS23+569 and IVS23+664 some interruptions by other sequences no more that five consecutive repeats without disruption.

4- (AAC)/ (GTT)*five repeats :

IVS5+459 one of the triplets does not agree.

5- (AGC)/ (CTG)*six and three repeats:

Six*Exon 15 many a.a. are present the fifth triplet is not the same. In addition, three repeats in exon 17.

6- (CCG)/ (CGG)*four repeats: Exon 29

7- (ATC)/ (GAT) : Negative blast results were seen.

8- (ACC)/ (GGT) : Negative blast results were seen.

9- (ACG)/ (CGT) : Negative blast results were seen.

10-(ACT)/ (AGT) : Negative blast results were seen.

### 3.3.8 Exon 23 (IVS23+24 T>C) of the *ELN* gene

DHPLC result shows the following pattern see Figure 49:

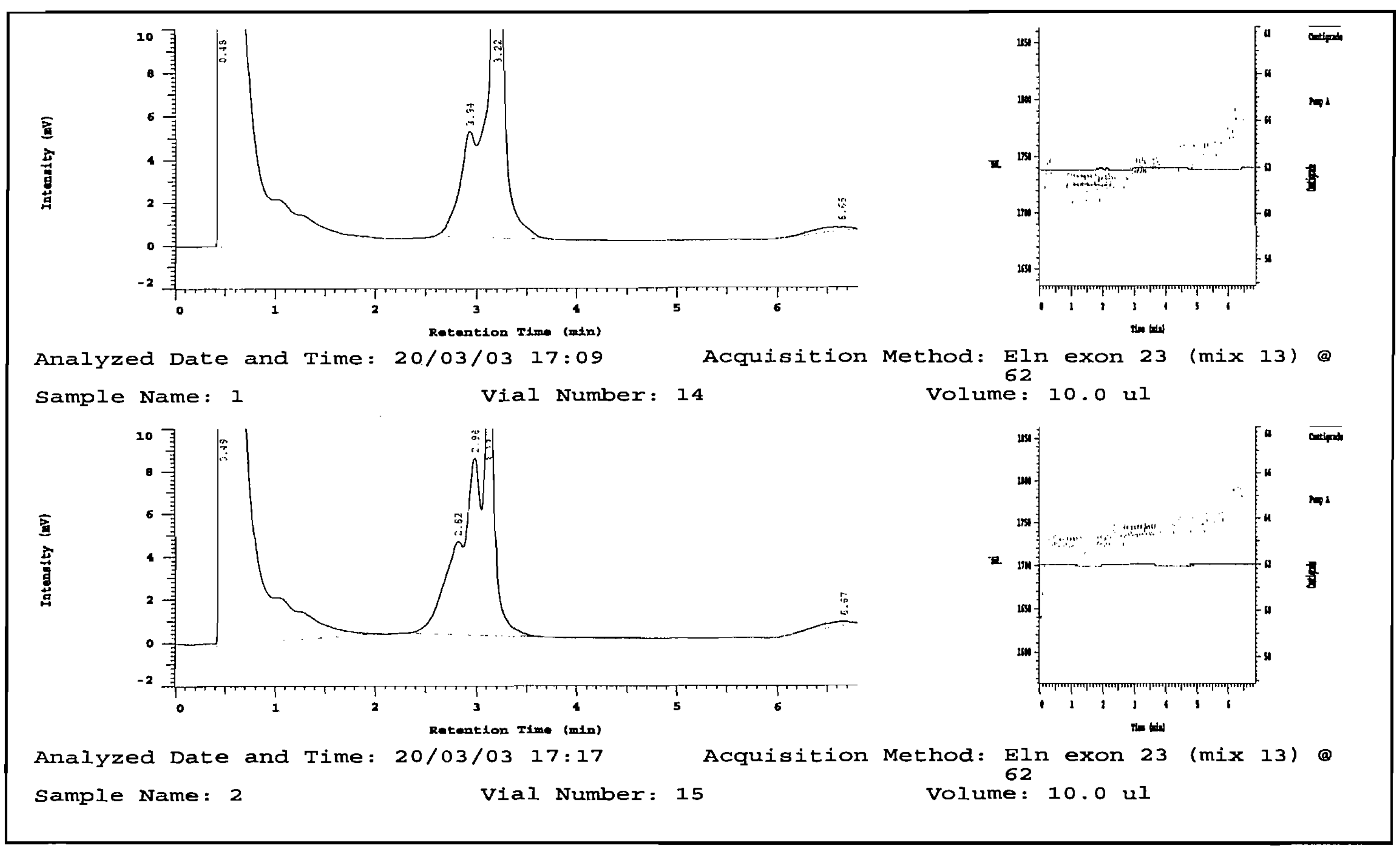
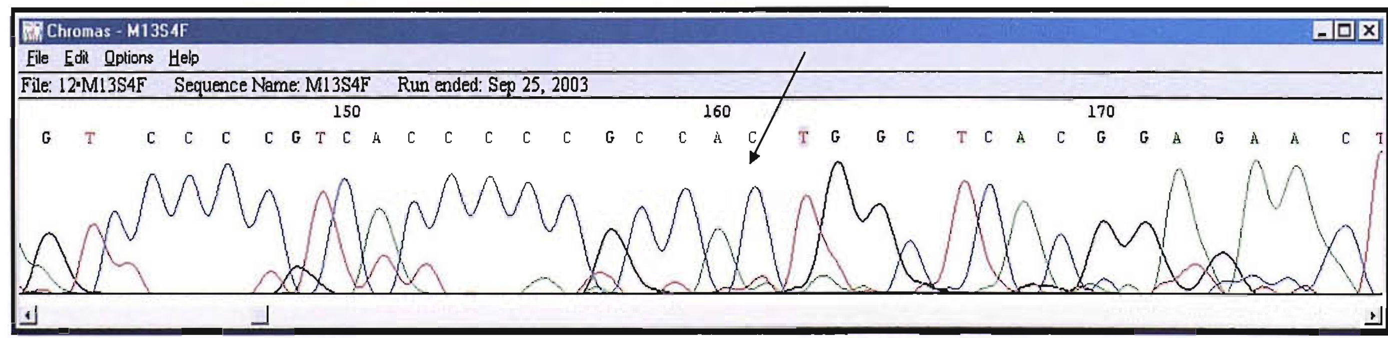



**Figure 49: DHPLC for samples 1, 2.**

DNA sequencing heterozygous **C/T:** (T/C **HIDEAKI ONDA et al 2001**) reverse sequence could not confirm this mutation Figure 50:



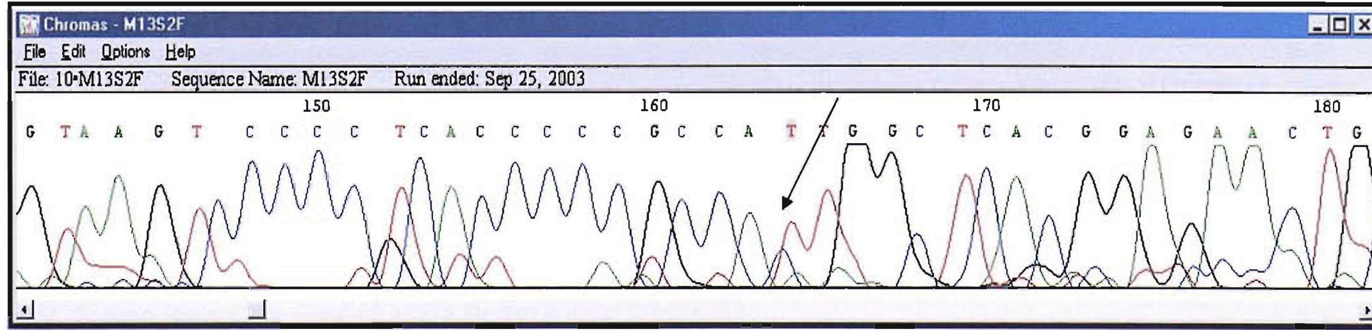




**Figure 50: Sequencing results of sample 4(normal)** and **sample two (heterozygous)**

The exact mutation sequence is shown in Figure 51:

**Figure 51 : Sequence obtained from GeneBank AC005056, exact mutation is shown (G)**

### 3.3.9 5' Flanking (POSITION -1050 C>T) (in the NCBI is G>A) of the *ELN* gene

This mutation occurs at the 5' flanking region position -1050 (Figure 52)

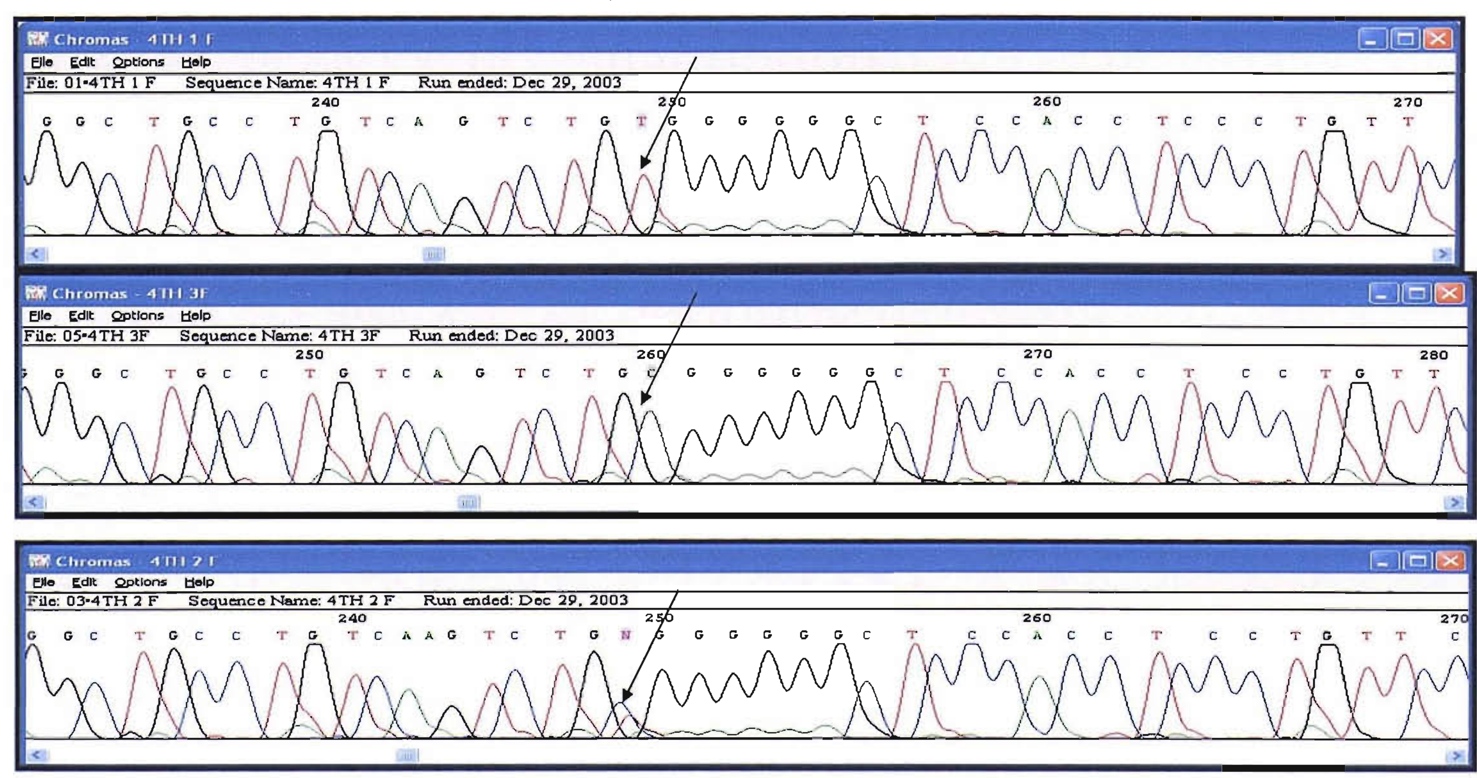



**Figure 52: Three types of mutations were detected, sample one is homozygous T, sample 2 is heterozygous T and C, sample three is homozygous C.**

The reverse sequence confirms the same mutations (Figure 53)

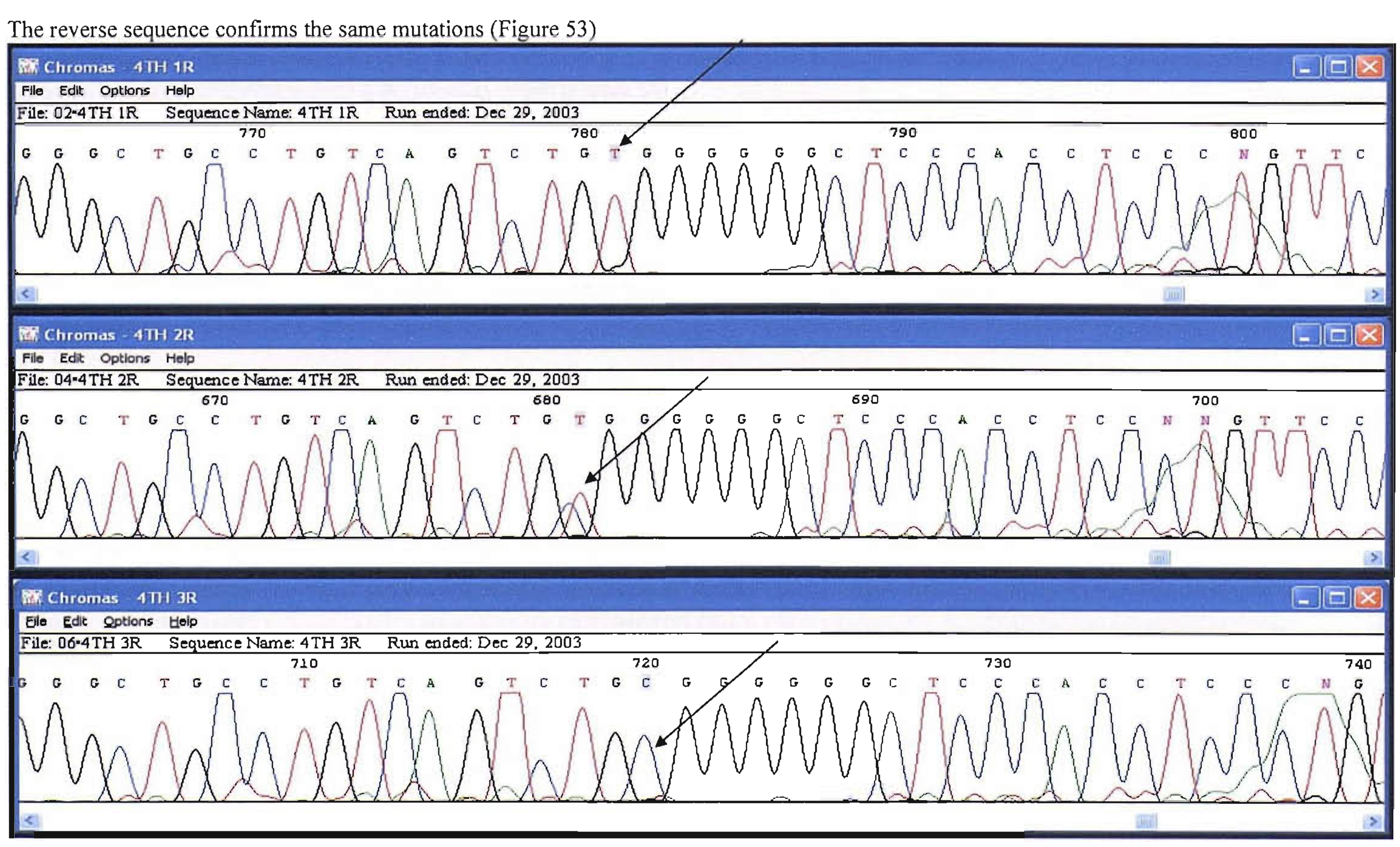



**Figure 53: The reverse sequence confirms the mutations, sample one is homozygous T, sample 2 is heterozygous T and C, sample three is homozygous.**

Mutation is known ONDA et al 2001 C/T (position -1050) (Figure 54)

**GLUCOCORTICOID-RESPONSIVE ELEMENT-1319**

```
30881   GAAGTGAGGGGTGGGGGACACGACCCTCCC CTGTTCCATC CCACACTCCA ACCCCCAAAA TCCCCCAGGG TCCCCGTCCA 67229
                                        >>>>>>> 4TH 5'UTR>>>>>
30961   GCTCAGTCCT GGGGGCAGAAATGCAGAGTT CTCCAGGAAC GTGGTCCCAG CTGTTTCAGT GCAGGCCGCC CCCTCCTGGC 67149
          <<< 5TH 5'UTR REV<<<<<
31041   CACCAGCGGA ATGTCAGCCT TCCCAGAGGG GCCGGGAGAACAGCAGTCGA GAAGCTCCCA GACTGGTGTG GGCGCTAGCT 67069
                                                              HIDEAKI ONDA ET AL 2001 C/T -1050
31121   GTGCTCAGCG TGGGGATGGGAGGTGACCCA GTGATAATGG GAAGCTGGGCTGCCTGTCAG TCTGCGGGGGGCTCCCACCT 66989
```

**GLUCOCORTICOID-RESPONSIVE ELEMENT-1025**

```
31201   CCCTGTTCCC CCACAGGGCA CCTGGGGATC CAGCCTGATT TTTACCAGAC CCTGCGGCCT GCATGGGGCT GGGTATAGGG 66909
31281   CTGTGACCTT GACCCATGCA GAATAGAACC CTGTGTGTCG GGATCCTCCA TGTGCTCCAG ATGCCCCTGG GGACAGCACC 66829
             >>>3RD 5 UTR FOR >>>>>               <<<<4TH 5UTR REV<<<<<<<
```

**Figure 54: The exact sequence is shown, the position of T mutation is present.**

### 3.3.10 5' Flanking (POSITION -1162 C>G) of the *ELN* gene

Another mutation was seen in the same domain, position -1162 C>G (Figure 55)

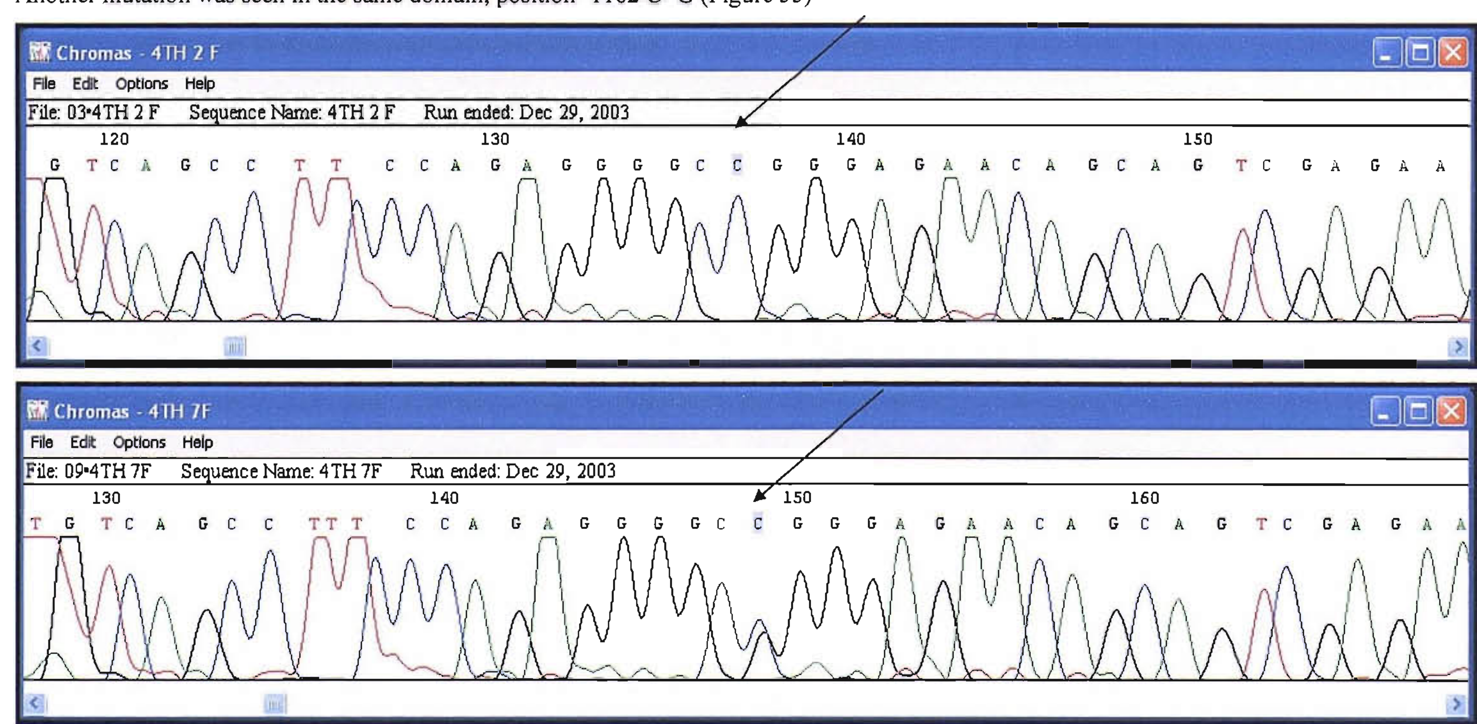



**Figure 55: Sample number 2 is normal at C, sample 7 is heterozygous C/G**

The reveres primer sequence confirms this mutation (C>G) Figure 56

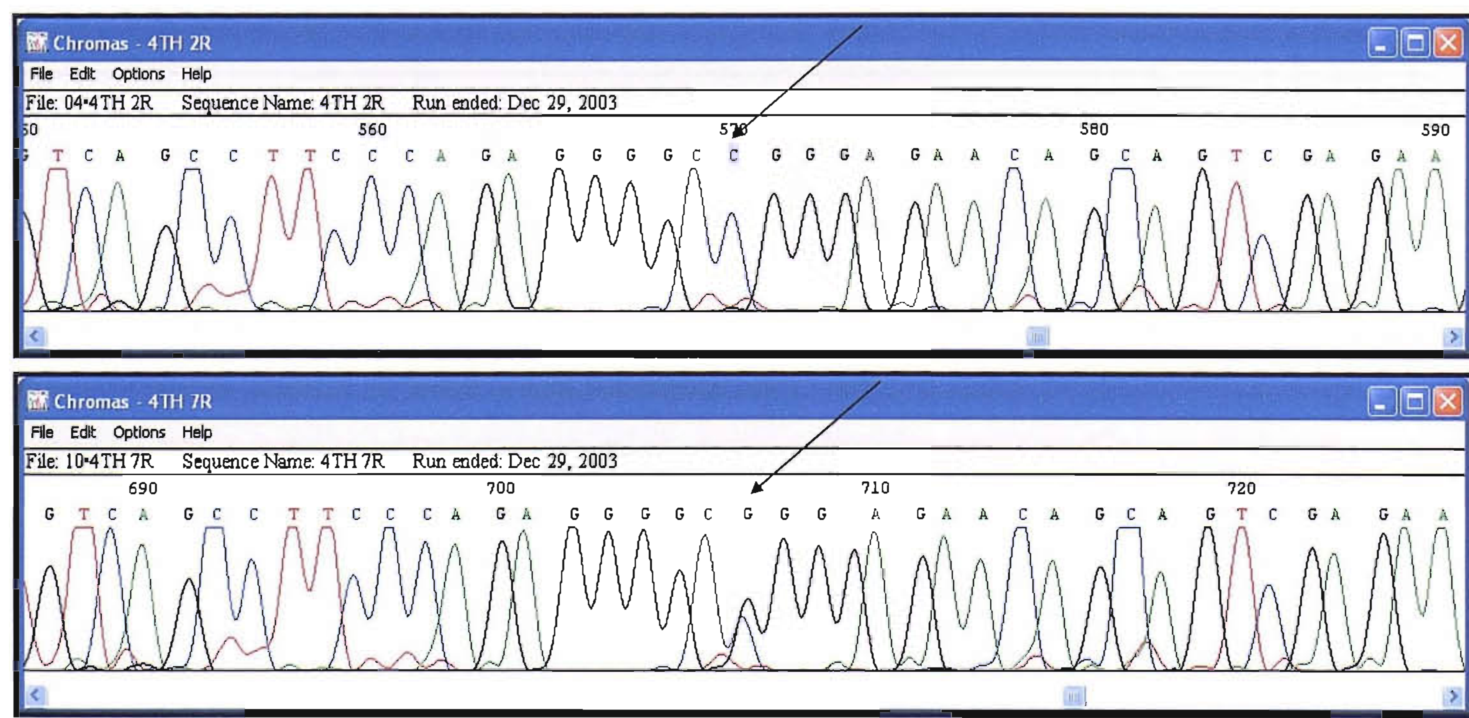



**Figure 56: Confirmation of C>G mutation is seen, sample seven has heterozygous mutation.**

Mutation is known ONDA et al 2001 C/G the exact sequence is shown in (position -1162) (Figure 57):

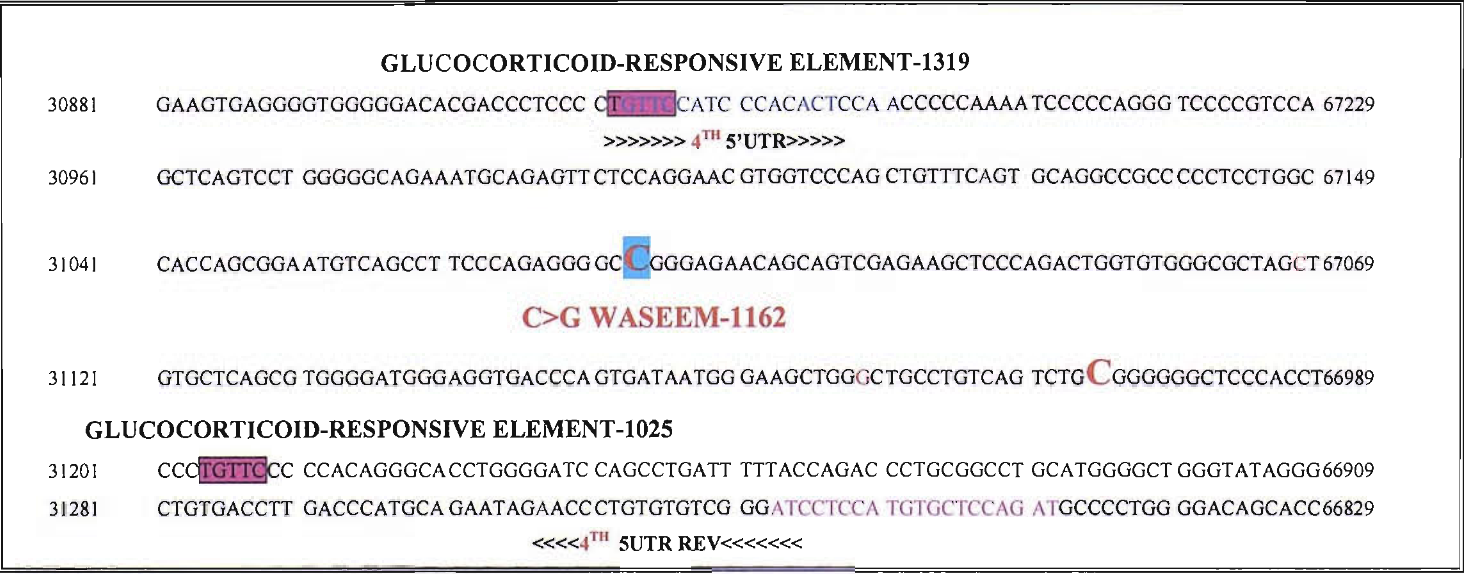



**Figure 57: The exact sequence is shown, the position of C>G mutation is present.**

### 3.3.11 5' Flanking (POSITION -1859 G>A rs3757583) of the *ELN* gene

A heterozygous mutation was found in sample 5 and normal control sample Figure 58

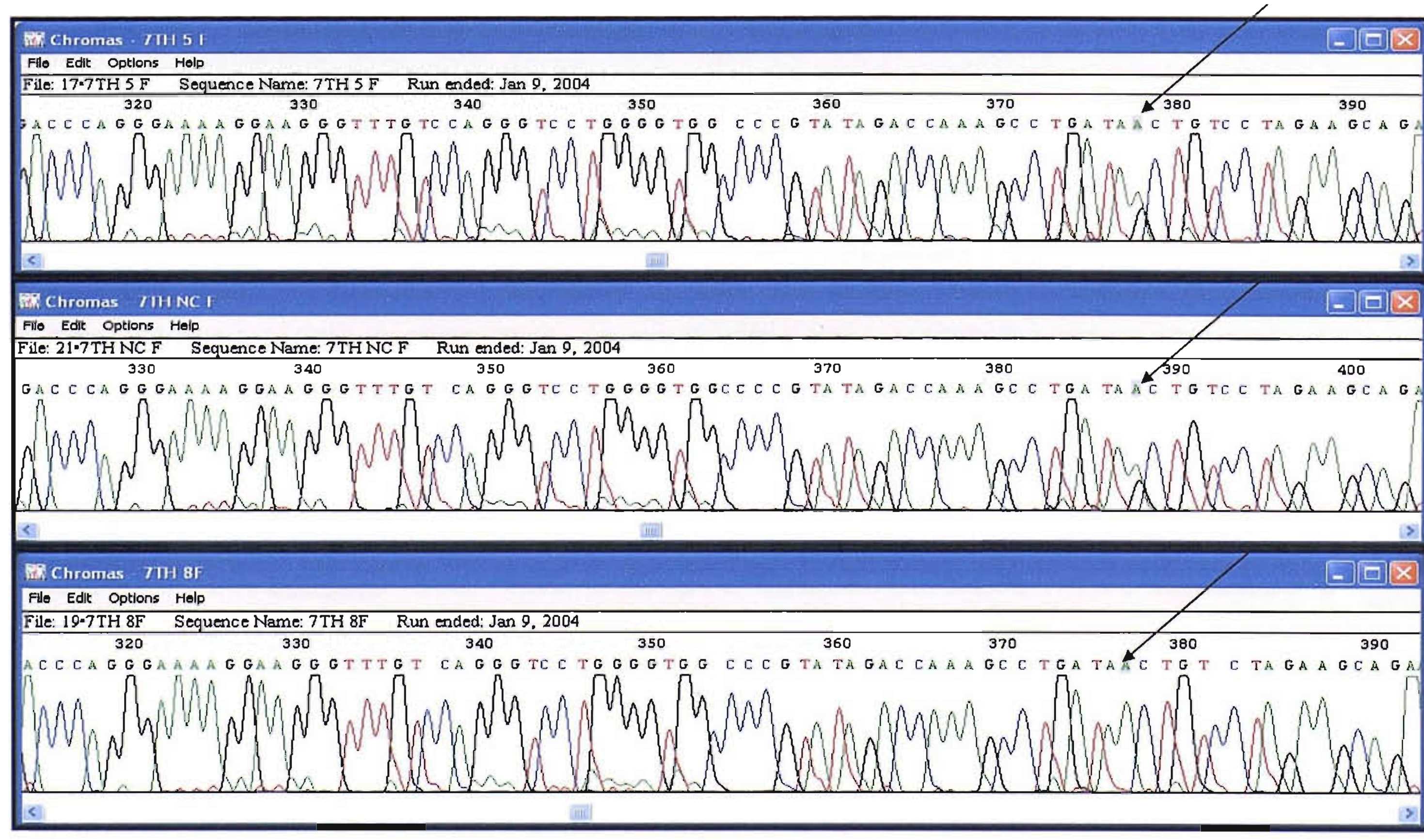



**Figure 58: Sample 5 and normal control sample are heterozygous G/C. Sample 8 is homozygous A**

This SNP A>G is known as ss4943615, Figure 59 shows the exact mutation position.

```
29921   ACTGCGGTTT GCAAGTCTGA AAGCTGGTTC CTGCCCGTGT CACTGCCTCG AGAAGAGAGGGTCCAGCTCCCCACAGTAG 68189
                                                    >>> 7TH 5 UTR FOR>>>>>>>
                                                    CRE-2178  PUTATIVE AP2-2171
30001   CCCCTGCCTT CCTCCTTCCC AGGCAGGCGGAGGCACGCAGATATACCATT GACTTCCCCT CCCCTGCAGC AGGCACATCC 68109
  <<<<8TH 5UTR REV<<<<<<
30081   TGGGCATCGA GCTTCAGACC CTGCCCCTGA GCAGCCCCTA ACCCCACCAA CAAAGGGTGGCTTGGGGGGG CTTTCACCCC 68029
30161   AGCATAATCT CCATCAGCTA CCCTCAAAGC ACCCCCAAAT AAACACACAC CGTAAGTAAG AGCTGTACAC TGGCTGTGTG 67949
30241   CGTACATCTT CAAGACAATT CTCCCAGCAT GCCCCTACCT TCCAAAATTC CAGAGCTGCT CCCTCCAAAG ACCCAGGGAA 67869
30321   AAGGAAGGGTTTGTCCAGGG TCCTGGGGTG GCCCCGTATA GACCAAAGCC TGATAGCTGTCCTAGAAGCAGAGTACTTGC 67789
>>>6TH 5'UTR FOR>>>>>>>>                          SS4943615 A/G
                CRE-1819                       SS6500846 G/A RS3757583
30401   AGAGCGAGTGA GGCAACTG TGGTATTGAC ACCAGTCCTA GCACCAGCTG AACACAGAGCATTTTTGATC TAGCAGAAAT 67709
```

**Figure 59: G>A mutation is seen, it is at position-1859.**

### 3.3.12 5' Flanking (POSITION -2253 G>C) of the *ELN* gene

Homozygous and heterozygous mutations were seen Figure 60

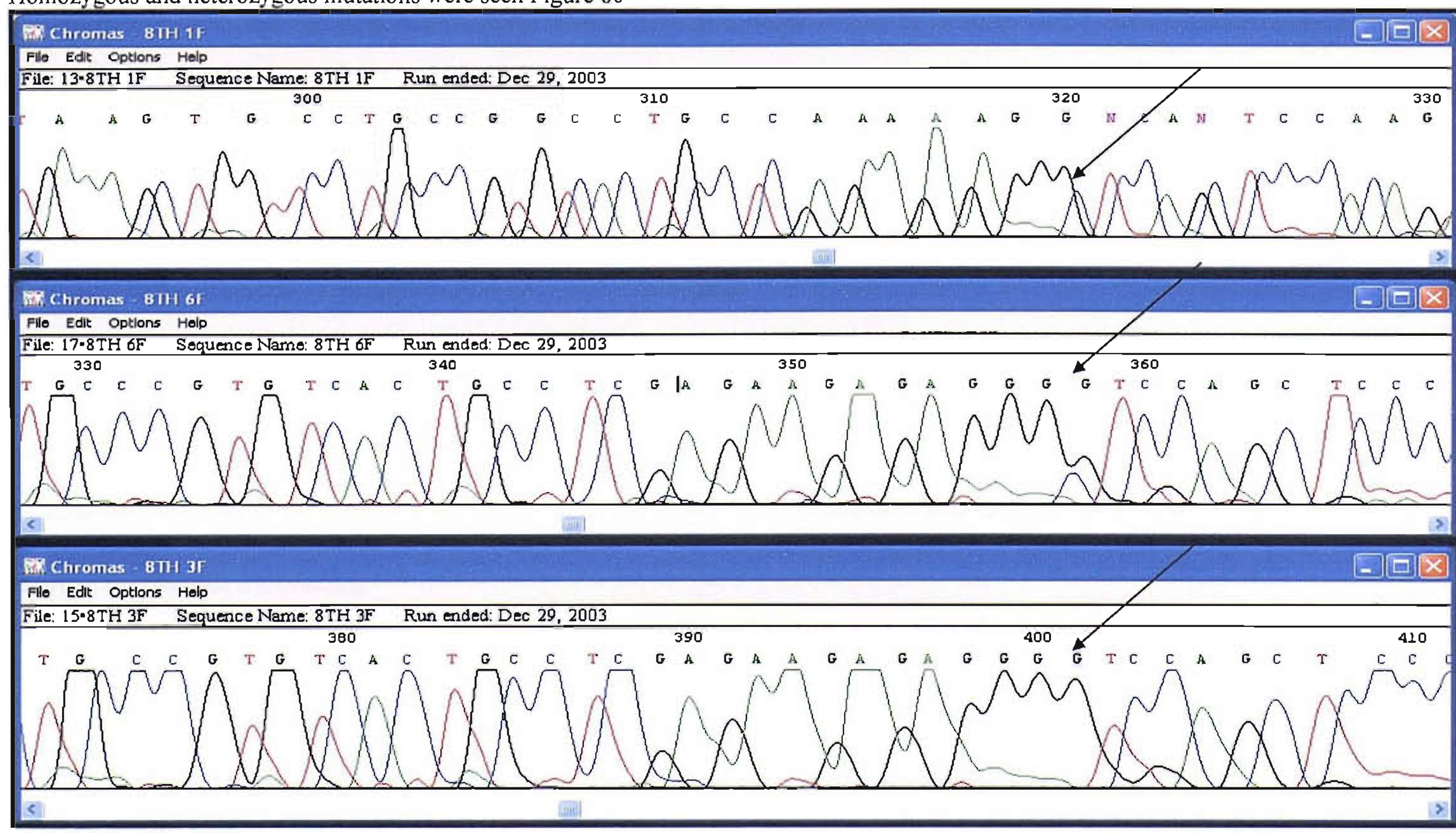



**Figure 60: Samples 1 and 6 are heterozygous, sample three is normal homozygous**

The exact mutation G is shown in Figure 61

```
29601   GTATCTTATG  CATCAGACTA CACGGCATGG ATATTCAGTG CAATGGATTT GATTGTGAAT TGAGGCCCCC TTTTGCAATC  68509
                    >>> 8TH 5 UTR FOR>>>>>
29681   TTGTTGCATG  TTATGTTCGT  GGAGCTCCTG CAATGAGCAG AGGCTGCTAA GCAGAAGCAGCGCACAGGCT GGGGGCAGGA68429
29761   CACAGGGCCC CCGGCCTCAG GGGAGAGGTCTGTTCCAGCC  ACCTCAAGTC TATTCCTGTG  AGTGCCCCTC CACACCCCTC 68349
29841   CTGGATCCCC CCACTGCAAA CTCTCTCCCC  CTGGCTCCGG ACTCTCTCCC  CGTCCTGGAA TTAATAAAGA CTCTGTGCAG 68269
29921   ACTGCGGTTT  GCAAGTCTGA AAGCTGGTTC CTGCCCGTGT  CACTGCCTCG AGAAGAGAGGGGTCCAGCTCCCCACAGTAG68189
                                                             >>> 7TH 5 UTR FOR>>>>>>>
                                                                  G>C WASEEM
```

**Figure 61: G>C mutation is present as G**

## 3.4 Results Summary

Summary of ready results are described in the following Table 22:

**Table 22: Shows the results of positive detection of SNPs after performing DHPLC and DNA sequencing, description of the exon name, function (of the exon domain), if any documented mutation is present in literature, then the location of this mutation ( Exon; Intron or 5'UTR ). Any amino acid substitution is described. If the mutations in the intronic region can affect or induce alternative splicing sites will be described, also included the frequency of each SNP (if it is known). Moreover, the expected observation and the observed mutations are included depending on the DHPLC results.**

| Exon Name | Domain Function | Method Name of the Positive DHPLC. Mutations not detected are *undetected* | Any Known mutations | Sequencing Result | Mutation Type | Amino Acid Substitution | Known Possible Effect and Comments | Average allele frequency and ethnicity | frequency of rare allele | the expected number of detected alleles = 16 :: frequency of rare allele | the actual number of detected alleles depending on |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | Hydrophobic Domain | *ELN* EXON20MIX16 @ 63 | Known | IVS20+17 T>C (ss4044368 T/C rs2856728) | Intronic | None | ? | T=0.867, C=0.133, HapMap-CEPH-30-trios(Northern and Western Europe) | 0.133 | 2.128 | 3 |
| 20 | Hydrophobic Domain | *ELN* EXON20MIX16 @ 63 | Known | AC005056 :37758 G>A (rs2071307) | Exonic | Glycine > Serine | Glycine and serine are very similar amino acids, no major structural change is expected. | (A=0.189, G=0.811 ONDA et al 2001) ,Also see Japanese :PMID: 12436197 | 0.189 | 3.024 | 3 |
| 19 | CROSS LINKING | *ELN* exon 19 (mix 17) @61 | Known | IVS19+70 T>C (rs2239691) | Intronic | None | ? | Overlap_SNPs_by_SsahaSNP | 0 | 0 | 4 |
| 33 | CROSS LINKING | *ELN* exon 33 mix 2 @62 | Known | IVS33-34 C>T (rs3757587) | Intronic | None | ? | T=0.052, C=0.948 Japanese (ONDA et al 2001) | 0.052 | 0.832 | 2 |
| 18 | Hydrophobic Domain | WAS E18M18@60 | Known | IVS18+20DEL2 rs5884930 -GT | Intronic | None | ? | Not available! | 0 | 0 | 2 |
| 18 | Hydrophobic Domain | *ELN* E18M18@64+1 | New | IVS18+47 G>C | Intronic | None | ? | Not available | 0 | 0 | 1 |
| 23 | CROSS LINKING | *ELN* exon23 ( mix 13) @ 62 | Known | IVS23+24 T>C | Intronic | None | ? | C=0.294, T=0.706Japanese (ONDA et al 2001) | 0.294 | 4.704 | 5 |
| 7TH | PROMOTER | 7THSUTR @59-1 | Known | POSITION -1859 G>A rs3757583 | 5'UTR | None | ? | A=0.725, G=0.275, SSAHA and WIBR fosmid using SsahaSNP | 0.275 | 4.4 | 2 |
| 4TH | PROMOTER | 4THSUTR @62-1 also 64 and 65 | New | POSITION -1050 | 5'UTR | None | ? | Not available | 0 | 0 | 2 |
| 4TH | PROMOTER | 4THSUTR @62-1 also 64 and 65 | New | POSITION -1162 C>G | 5'UTR | None | ? | Not available | 0 | 0 | 2 |
| 8TH | PROMOTER | 8THSUTR @ 59-1 *also 63 and 64 | New | POSITION -2253 G>C | 5'UTR | None | ? | Not available | 0 | 0 | 1 |

## 3.5 *In silico* analysis of the 5' region of the elastin gene

DNA mutations on promoter region can change the percentile of expressed product, studies on mice vascular smooth muscle cells lacking the elastin gene, have shown that elastin can induce actin stress fibre organisation; inhibit proliferation; regulates migration and signals. This alteration in gene expression of *ELN* may play an important role in the onset of SAH.

Three novel mutations were detected in the 5' flanking region and one previously described mutation (summary of mutations detected are in Table 22). To investigate the putative function of these SNPs I have used a prediction programme to see the possible functions for these sequences (programme name is TFSEARCH: Searching Transcription Factor Binding Sites (version 1.3) the website is in [315]) TRANSFAC Ali Baba programme [316] is a database on eukaryotic cis-acting regulatory DNA elements and trans-acting factors. It covers the whole range from yeast to human, when I used it under the TFBLAST [317] I could not have any positive hits regarding the normal and the mutant sequences. Results of the TFSEARCH programme are shown in Table 23:

**Table 23: Using TFSEARCH programme using threshold of 85.0 point (default)**

| MUTATION | MUTATION NOTE | TF FUNCTION WT | SCORE | TF FUNCTION AFTER SNP/MUTATION | SCORE |
|---|---|---|---|---|---|
| G>C-2253 | New 5' UTR | M00083 MZF1 | 95.7 | M00083 MZF1<br>M00085 ZID | 95.7<br>87.0 |
| C>G -1162 | New 5' UTR | M00008 Sp1 | 87.7 | M00008 Sp1<br>M00083 MZF1 | 95.9<br>93.0 |
| C>T-1050 | New 5' UTR | NONE | < 85.0 | NONE | < 85.0 |
| **G>A -1859** | **rs3757583 5'UTR** | M00075 **GATA-1**<br>M00076 **GATA-2**<br>M00127**GATA-1** | 95.5<br>94.1<br>85.5 | M00075 **GATA-1**<br>M00076 **GATA-2**<br>M00127**GATA-1**<br>M00128**GATA-1** | 88.2<br>87.8<br>91.6<br>87.8 |

Table 24 shows the results of TRANSFAC Ali Baba before and after the SNP.

**Table 24: TRANSFAC Ali Baba results before and after SNP and it s possible creation or abolishing important sites for transcription control.**

| MUTATION | MUTATION NOTE | TF FUNCTION WT | TF FUNCTION AFTER SNP/MUTATION |
|---|---|---|---|
| G>C-2253 | New 5' UTR | Sp1 | NONE |
| C>G -1162 | New 5' UTR | Sp1 GR | SP1 GR ETF MIG1 |
| C>T-1050 | New 5' UTR | Sp1 | Sp1 |
| **G>A -1859** | **rs3757583 5'UTR** | GATA-1 | NONE |

## 3.6 Results of linkage analysis performed on the French family on the *TGFRβII* gene

Linkage analysis was performed using five STRs. The markers used are defined on the left box of the chart (Figure 62) and their linkage distances in cM are shown. In five markers used in this analysis, four markers are a tetra nucleotide repeats and they are (D3S2466, D3S4535, D3S2432, and D3S1768) and one CA repeat marker (D3S3727) that is present in intron one of the *TGFRβII* gene. Coloured markers are the ones that were typed, the uncoloured are the ones inferred. The yellow chromosome seems the one that carries the diseased gene, and it manifests itself in all affected persons. Sample 1 family branch represented by the yellow chromosome is easily recognised while in the Sample 2 branch, it seems to be involved in two recombination, the first one can be started with sample 2 or her dissent, and the second recombination occurring in the sample 25 family. All the affected patients had the yellow chromosome segment, see Figure 62.

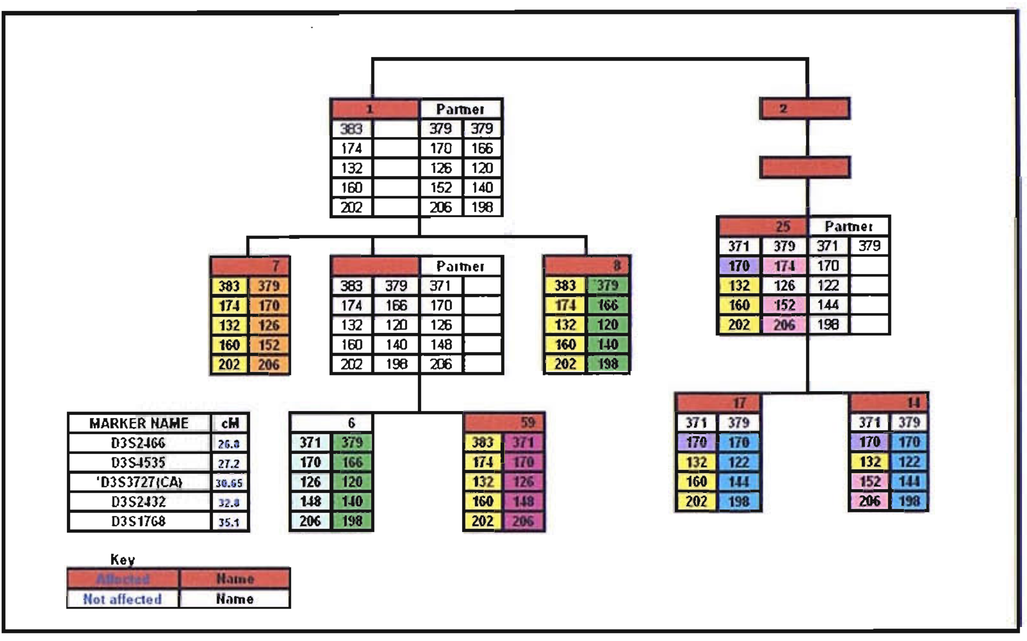



**Figure 62:** Linkage analysis performed on the French family that was shown to be linked with yellow 132 segment, affected names are with red background, un affected are clear, markers used are on the left side of the chart, also the genetic distance is present.

I could not assign the phase for the 371/379s of D3S2466 in sample-25 branch. In sample-1 branch: sample-7; sample-8 and sample-59 were used to identify the putative affected haplotype, our results shows that 383-174-132-160-202 are the markers on the affected chromosome. In sample-25 case, it seems to have one recombination leading to (317 or 379)-170-132-160-202 haplotype, I do not know were the exact recombination happened, hence four meioses occurred between *TGFβRII* and D3S4535, a 3.45cM interval (from sample-1 to sample-25). Concerning sample-17 and sample-14 siblings; they seem to be inherited the same chromosome from their father, the maternal chromosome (for sample-17 case ) is the same affected one from her mother, while sample-14 mother chromosome seems to be involved in a recombination between *TGFβRII* and D3S2432 in the 2.15 cM interval. The only marker that did not have any recombination in all of the affected persons is the 132 of the D3S3727 (this one is present in all of the affected patients and is in the intronic region of the *TGFβRII* gene)

I have 10cM (10% to have recombination per meioses) and I do have 11 meioses in total (concerning only the affected chromosome), so to have no recombination / meioses then 100-10=90%/meioses i.e. 0.9, now if I have 11 meioses then the possibility of having no recombination is 0.9^11=about 31% (approximately 30%). Then the recombination

possibility is about 70%.

Slippage of tetras at mutation rates of 10^-3 or less would be a much less frequent source of transmission 'inconsistencies.

Runs and co runs were performed to each of the samples, here three samples results are shown to show the possibility of two recombination discussed before.

Co-runs of samples 17 and 25, sample 17, sample 25 for D3S2466 are shown in Figure 63.

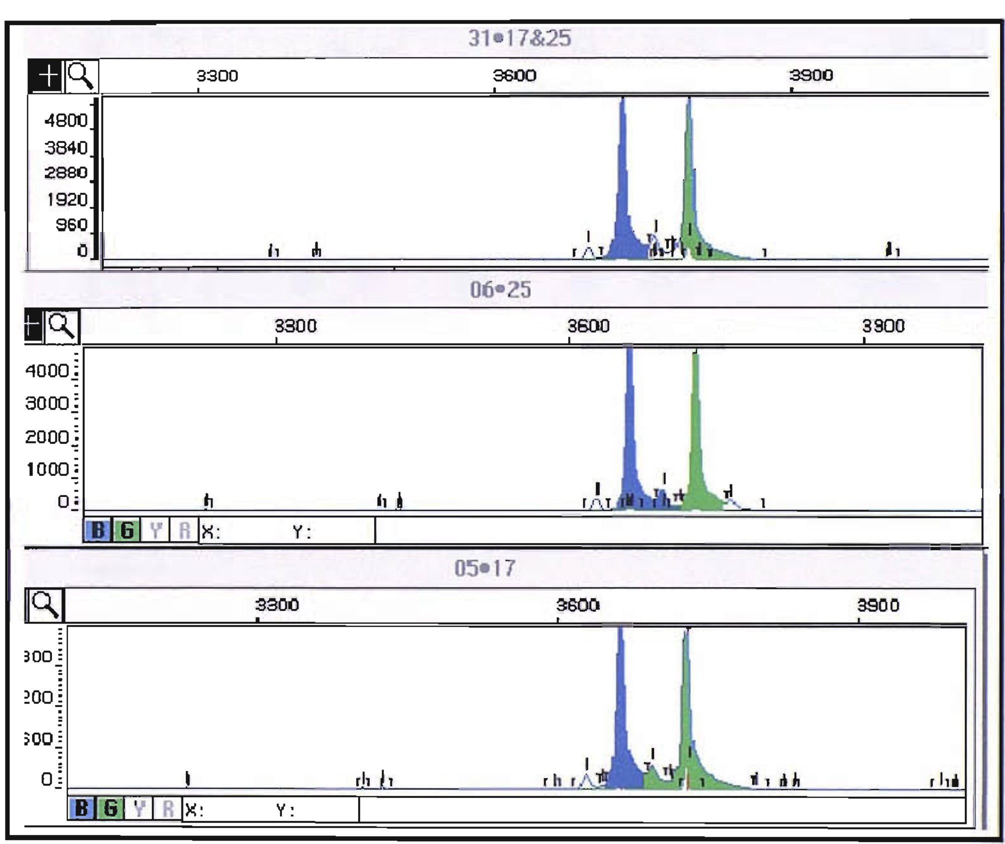



**Figure 63: Co-runs of samples 17 and 25; sample 17; sample 25 for D3S2466 are shown, they contain the same genotype blue is the 371 green is the 379 bases.**

## 3.7 Results Linkage analysis using D3S3727 D3S2432, D3S4535 and D3S1768 markers

Co-runs of samples 17 and 25; sample 17 and sample 25 for the rest of the markers respectively (D3S3727, D3S2432, D3S4535 and D3S1768) and are shown in Figure 64

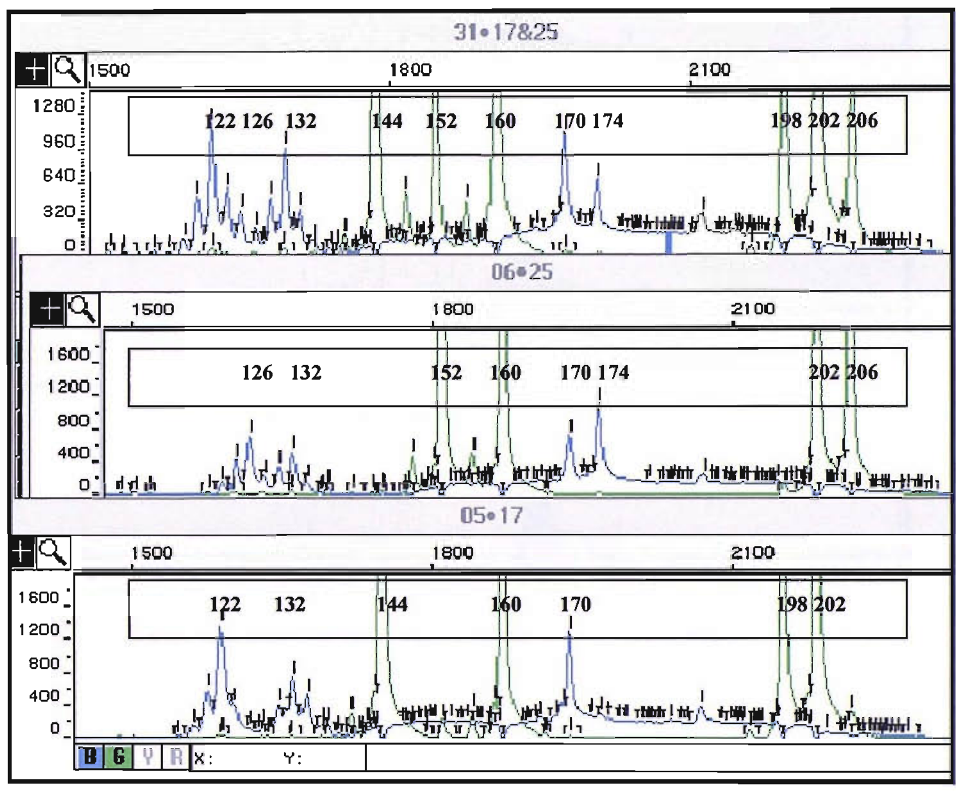



**Figure 64: Co-runs of samples 17 and 25, sample 17, sample 25 for the rest of the markers respectively (D3S3727, D3S2432, D3S4535 and D3S1768) alleles 132, 160, 170 and 202 are in common.**

Co-runs of samples 14 and 25; sample 14; sample 25 for D3S2466 are shown in Figure 65

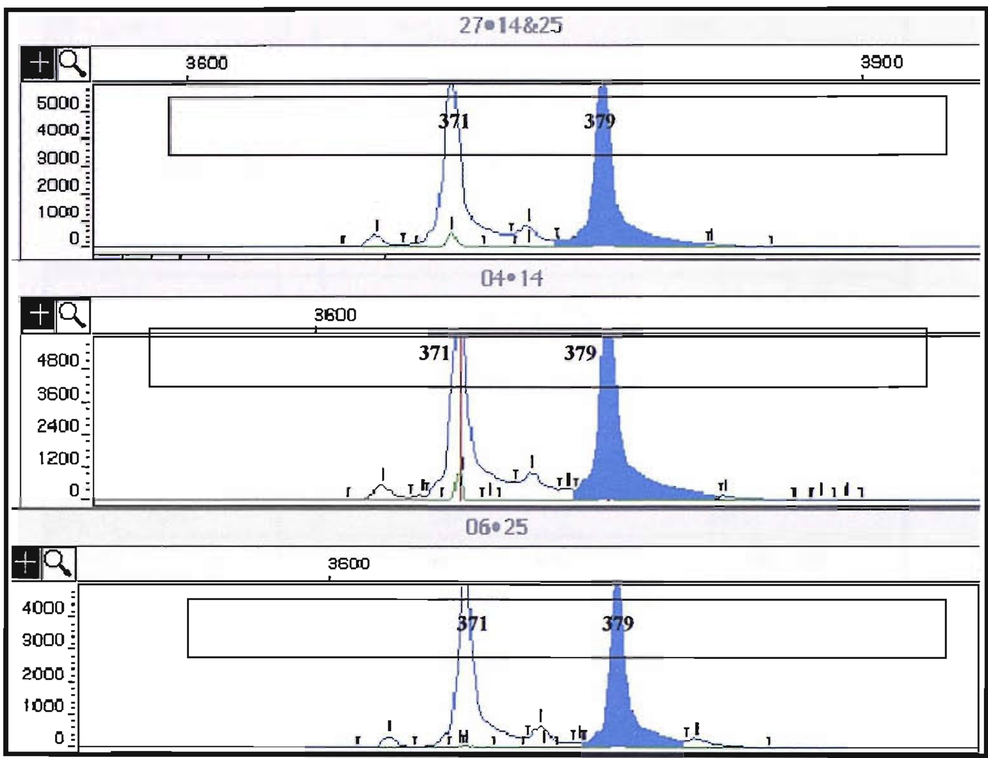



**Figure 65: Co-runs of samples 14 and 25; sample 14; sample 25 for D3S2466 are shown, both have the same alleles.**

Co-runs of samples 14 and 25; sample 14; sample 25 for the rest of the markers respectively (D3S3727, D3S2432, D3S4535 and D3S1768) are shown in Figure 66:

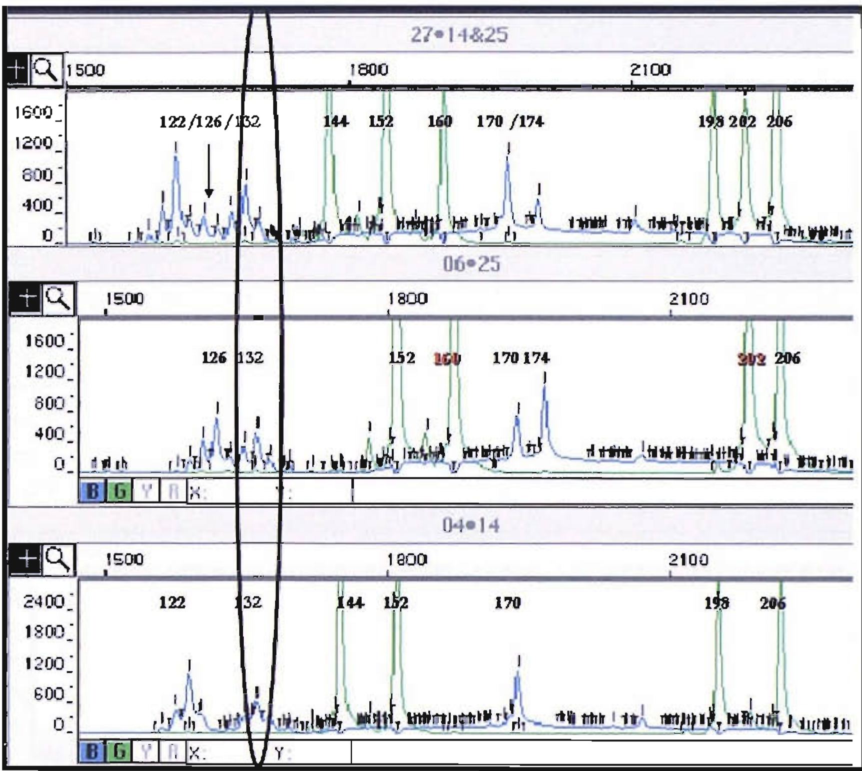



**Figure 66: Co-runs of samples 14 and 25; sample 14; sample 25 for the rest of the markers respectively (D3S3727, D3S2432, D3S4535 and D3S1768), alleles 132, 152, 170 206 are in common. Allele 132 always segregates with the affected chromosome.**

Co-runs of samples 14 and 17, sample 14, sample 17 for the following markers respectively (D3S3727, D3S2432, D3S4535 and D3S1768) are shown in Figure 67.

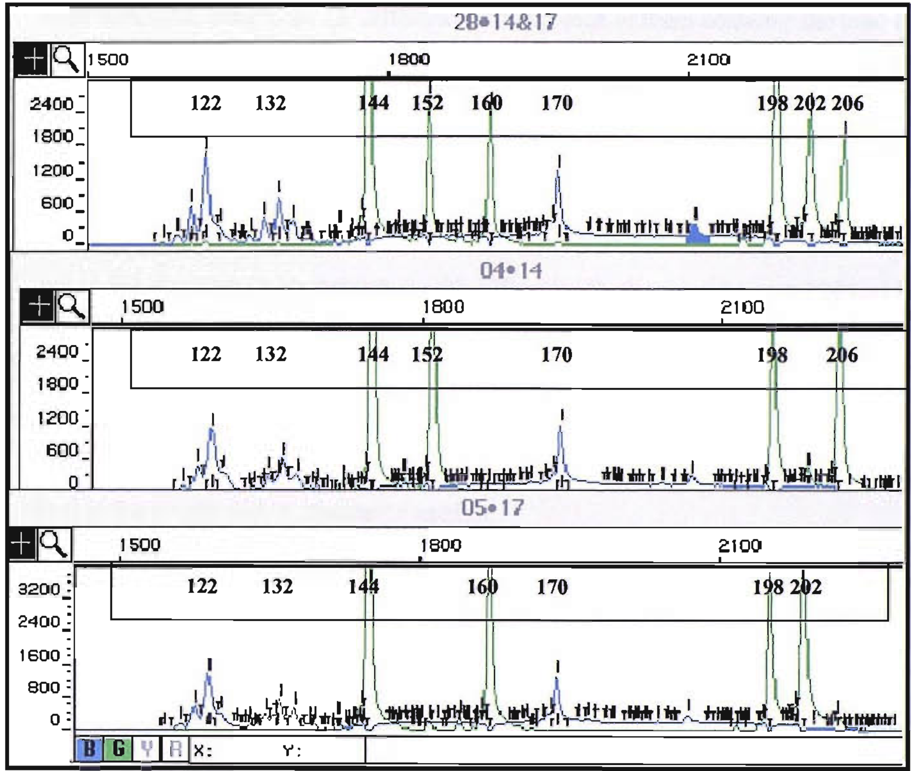



**Figure 67: Co-runs of samples 14 and 17; sample 14; sample 17 for the following markers respectively (D3S3727, D3S2432, D3S4535 and D3S1768), alleles 122; 132; 144; 170; 198 are in common.**

## 3.8 Results of the association analysis using D3S3727 on *TGFβRII*

Using sporadic SAH vs. control, this marker (in particular allele 132) was associated with aortic aneurysm in the French family as in Figure 62, it is an opportunity to perform STR analysis and test them for five different models

Table 25 shows the results of the genotyping of D3S3727 microsatellite on the *TGFβRII* gene, alleles present in this study were : 116;118;120;122;124;126;128;130;132;134;136 and 138. This table is divided into three horizontal subgroups (in different columns) each of them contains: the total counts of the alleles; total count without major allele; group name; the total count of the group and the total count without the major allele in the group. Vertically, on the left are alleles present 116 to 138.

Allele 132 ( the one present in the French family and is associated with aortic aneurysm and is forming a small cluster group with 130 and134), the second one is 120 that is forming a cluster group with 118, the third one is allele 126 that is close to the cluster family 124)

Nine group families are shown (A to I), and there is much emphasis on family groups (D to I) in the results and in discussion section.

Table 25: Results of GeneScan genotyping of GDO samples sporadic SAH and control samples.

| ALLELE | TOTAL | TOTAL-COUNT without major allele | Group Name | Total Group Count | Without Major Allele | Sporadic SAH | TOTAL-COUNT without major allele | Total Group Count | Group Count | Without Major Allele | HI | TOTAL-COUNT without major allele | Group Name | Total Group Count | Without Major Allele |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 116 | 1 | 427 | | | | 1 | 273 | | | | 0 | 154 | | | |
| 118 | 14 | 414 | | | | 9 | 265 | | | | 5 | 149 | | | |
| 120 | 92 | 336 | A | 106 | 14 | 58 | 216 | D | 67 | 9 | 34 | 120 | G | 39 | 5 |
| 122 | 11 | 417 | | | | 8 | 266 | | | | 3 | 151 | | | |
| 124 | 58 | 370 | | | | 35 | 239 | | | | 23 | 131 | | | |
| 126 | 108 | 320 | B | 166 | 58 | 73 | 201 | E | 108 | 35 | 35 | 119 | H | 58 | 23 |
| 128 | 21 | 407 | | | | 14 | 260 | | | | 7 | 147 | | | |
| 130 | 44 | 384 | | | | 25 | 249 | | | | 19 | 135 | | | |
| 132 | 62 | 366 | | | | 41 | 233 | | | | 21 | 133 | | | |
| 134 | 11 | 417 | C | 117 | 55 | 6 | 268 | F | 72 | 31 | 5 | 149 | I | 45 | 24 |
| 136 | 3 | 425 | | | | 1 | 273 | | | | 2 | 152 | | | |
| 138 | 3 | 425 | | | | 3 | 271 | | | | 0 | 154 | | | |
| TOTAL | 428 | | | 389 | | 274 | | | 247 | | 154 | | | 142 | |
| Samples | 214 | | | | | 137 | | | | | 77 | | | | |

Genotype result of GeneScan on the CA dinucleotide (D3S3727) in intron 1 of *TGFβRII* gene, 78 genotypes were detected, see Table 26.

**Table 26: Results of genotypes of total and sporadic SAH and control samples with there numbers.**

| GENOTYPE | TOTAL | Sporadic SAH | Control |
|---|---|---|---|
| 116:116 | 0 | 0 | 0 |
| 116:118 | 0 | 0 | 0 |
| 116:120 | 1 | 1 | 0 |
| 116:122 | 0 | 0 | 0 |
| 116:124 | 0 | 0 | 0 |
| 116:126 | 0 | 0 | 0 |
| 116:128 | 0 | 0 | 0 |
| 116:130 | 0 | 0 | 0 |
| 116:132 | 0 | 0 | 0 |
| 116:134 | 0 | 0 | 0 |
| 116:136 | 0 | 0 | 0 |
| 116:138 | 0 | 0 | 0 |
| 118:118 | 1 | 1 | 0 |
| 118:120 | 6 | 3 | 3 |
| 118:122 | 0 | 0 | 0 |
| 118:124 | 1 | 1 | 0 |
| 118:126 | 3 | 2 | 1 |
| 118:128 | 0 | 0 | 0 |
| 118:130 | 2 | 1 | 1 |
| 118:132 | 0 | 0 | 0 |
| 118:134 | 0 | 0 | 0 |
| 118:136 | 0 | 0 | 0 |
| 118:138 | 0 | 0 | 0 |
| 120:120 | 14 | 11 | 3 |
| 120:122 | 0 | 0 | 0 |
| 120:124 | 10 | 5 | 5 |
| 120:126 | 26 | 18 | 8 |
| 120:128 | 0 | 0 | 0 |
| 120:130 | 8 | 4 | 4 |
| 120:132 | 10 | 5 | 5 |
| 120:134 | 3 | 0 | 3 |
| 120:136 | 0 | 0 | 0 |

| | | | |
|---|---|---|---|
| 120:138 | 0 | 0 | 0 |
| 122:122 | 1 | 1 | 0 |
| 122:124 | 2 | 1 | 1 |
| 122:126 | 4 | 4 | 0 |
| 122:128 | 1 | 0 | 1 |
| 122:130 | 1 | 1 | 0 |
| 122:132 | 1 | 0 | 1 |
| 122:134 | 0 | 0 | 0 |
| 122:136 | 0 | 0 | 0 |
| 122:138 | 0 | 0 | 0 |
| 124:124 | 2 | 1 | 1 |
| 124:126 | 15 | 11 | 4 |
| 124:128 | 4 | 2 | 2 |
| 124:130 | 8 | 5 | 3 |
| 124:132 | 12 | 7 | 5 |
| 124:134 | 0 | 0 | 0 |
| 124:136 | 1 | 0 | 1 |
| 124:138 | 1 | 1 | 0 |
| 126:126 | 11 | 7 | 4 |
| 126:128 | 2 | 1 | 1 |
| 126:130 | 14 | 8 | 6 |
| 126:132 | 18 | 12 | 6 |
| 126:134 | 4 | 3 | 1 |
| 126:136 | 0 | 0 | 0 |
| 126:138 | 0 | 0 | 0 |
| 128:128 | 2 | 2 | 0 |
| 128:130 | 4 | 2 | 2 |
| 128:132 | 3 | 2 | 1 |
| 128:134 | 2 | 2 | 0 |
| 128:136 | 1 | 1 | 0 |
| 128:138 | 0 | 0 | 0 |
| 130:130 | 1 | 0 | 1 |
| 130:132 | 3 | 3 | 0 |
| 130:134 | 2 | 1 | 1 |
| 130:136 | 0 | 0 | 0 |
| 130:138 | 0 | 0 | 0 |
| 132:132 | 6 | 5 | 1 |
| 132:134 | 0 | 0 | 0 |
| 132:136 | 1 | 0 | 1 |
| 132:138 | 2 | 2 | 0 |
| 134:134 | 0 | 0 | 0 |
| 134:136 | 0 | 0 | 0 |
| 134:138 | 0 | 0 | 0 |
| 136:136 | 0 | 0 | 0 |
| 136:138 | 0 | 0 | 0 |
| 138:138 | 0 | 0 | 0 |

Five models were proposed to be tested using sporadic SAH, and control samples against the following:

### 3.8.1 Allele Counts /Additivity Model results (*TGFßRII*)

Looking for allele 132 vs. non-132, does the presence of allele 132 contribute to SAH? Results are present in Table 27:

**Table 27: Results of allele count, p value and chi square are present. Taking 132 allele to be important in comparison to all non-132**

| Taking 132 allele to be important in comparison to all | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Observed | | | | | Expected | | | | |
| | 132 | NON 132 | Total | | | 132 | NON 132 | Total | |
| Sporadic | 41 | 233 | 274 | | sporadic | 39.69 | 234.31 | 274.00 | |
| Control | 21 | 133 | 154 | | control | 22.31 | 131.69 | 154.00 | |
| Total | 62 | 366 | 428 | | Total | 62.00 | 366.00 | 428.00 | |
| X^2 | 0.141227 | | | | | | | | |
| P- VALUE | 0.707064 | | | | | | | | |

Looking for allele 126 vs. non-126, does the presence of allele 126 contribute to SAH? Results are present in Table 28:

**Table 28: Taking 126 allele to be important in comparison to all non-126**

| Taking 126 allele to be important in comparison to all | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |
| Observed | | | | | Expected | | | |
| | 126 | NON 126 | Total | | | 126 | NON 126 | Total |
| Sporadic | 73 | 201 | 274 | | Sporadic | 69.14 | 204.86 | 274.00 |
| Control | 35 | 119 | 154 | | Control | 38.86 | 115.14 | 154.00 |
| Total | 108 | 320 | 428 | | Total | 108.00 | 320.00 | 428.00 |
| X^2 | 0.809476 | | | | | | | |
| P- VALUE | 0.368275 | | | | | | | |

Looking for allele 120 vs. non-120 , does the presence of allele 120 contribute to SAH? Results are present in Table 29:

**Table 29: Taking 120 allele to be important in comparison to all non-120 (observed results came from table 25).**

| Taking 120 allele to be important in comparison to all | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observed | | | | | Expected | | | |
| | 120 | NON 120 | Total | | | 120 | NON 120 | Total |
| Sporadic | 58 | 216 | 274 | | sporadic | 58.90 | 215.10 | 274.00 |
| Control | 34 | 120 | 154 | | control | 33.10 | 120.90 | 154.00 |
| Total | 92 | 336 | 428 | | Total | 92.00 | 336.00 | 428.00 |
| | | | | | | | | |
| X^2 | 0.048249 | | | | | | | |
| P- VALUE | 0.826138 | | | | | | | |

## 3.8.2 Major Expansion Model (Anticipation) (*TGFßRII*)

Looking for the expected value of sporadic SAH vs. control (Table 30) for allele 132

**Table 30: Looking at the expected homozygous numbers of allele 132 (132:132 genotype) in sporadic SAH and control.**

| Taking 132:132 genotype with non all | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observed | | | Observed | Expected | | | | |
| | 132:132 | NON | | 132:132 | NON | Total | 132:132 | |
| Sporadic | 5 | 132 | Sporadic | 3.84 | 133.16 | 137.00 | 3.84 | |
| Control | 1 | 76 | Control | 2.16 | 74.84 | 77.00 | 2.16 | |
| Total | 6 | 208 | Total | 6.00 | 208.00 | 214.00 | 6.00 | |
| | | | | | | | | |
| X^2 | 1.125525 | | | | | | | |
| P- VALUE | 0.288732 | | | | | | | |

Looking for the expected value of sporadic SAH vs. control (Table 31) for allele 126

**Table 31: Looking at the expected homozygous numbers of allele 126 (126:126 genotype) in sporadic SAH and control.**

| Taking 126:126 genotype with non all | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observed | | | | | Expected | | | |
| | 126:126 | NON | Total | | | 126:126 | NON | Total |
| Sporadic | 7 | 130 | 137 | | sporadic | 7.04 | 129.96 | 137.00 |
| Control | 4 | 73 | 77 | | control | 3.96 | 73.04 | 77.00 |
| Total | 11 | 203 | 214 | | Total | 11.00 | 203.00 | 214.00 |
| X^2 | 0.000735 | | | | | | | |
| P- VALUE | 0.978374 | | | | | | | |

Looking for the expected value of Sporadic SAH vs. control (Table 32) for allele 120

**Table 32: Looking at the expected homozygous numbers of allele 120 (120:120 genotype) in sporadic SAH and control.**

| Taking 120:120 genotype with non all | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observed | | | | | Expected | | | |
| | 120:120 | NON | Total | | | 120:120 | NON | Total |
| Sporadic | 11 | 126 | 137 | | sporadic | 8.96 | 128.04 | 137.00 |
| Control | 3 | 74 | 77 | | control | 5.04 | 71.96 | 77.00 |
| Total | 14.00 | 200 | 214 | | Total | 14.00 | 200.00 | 214.00 |
| X^2 | 1.486418 | | | | | | | |
| P- VALUE | 0.222773 | | | | | | | |

## 3.8.3 Loss of Heterozygosity Model (*TGFßRII*)

Testing wither I have genotypes (132:132); (126:126) AND (120:120) as expected, if they are within the expected range then heterozygosity does not likely to be the cause, using the results present in Table 30; Table 31 and Table 32.

### 3.8.4 Recessive Model (*TGFßRII*)

The result of this model goes with the expansion model as described in Table 30; Table 31 and Table 32. The results of chi-square and p-value are shown in these tables.

### 3.8.5 Dominant Model (*TGFßRII*)

Investigation of the presence of allele 132 in genotypes against none-132 genotypes, see Table 33.

**Table 33: Taking 132:132 genotype with non all**

| Taking 132 against all | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Observed** | | | | | **Expected** | | | |
| | **132** | **NON** | **Total** | | | **132** | **NON** | **Total** |
| **Sporadic** | **41** | **233** | **274** | | **sporadic** | **39.69** | **234.31** | **274.00** |
| **Control** | **21** | **133** | **154** | | **control** | **22.31** | **131.69** | **154.00** |
| **Total** | **62** | **366** | **428** | | **Total** | **62.00** | **366.00** | **428.00** |
| **X^2** | **0.141227** | | | | | | | |
| **P- VALUE** | **0.707064** | | | | | | | |

Investigation of the presence of allele 126 in genotypes against none-126 genotypes, see Table 34.

**Table 34: Presence of allele 126 against all**

| presence of allele 126 against all | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Observed** | | | | **Expected** | | | | |
| | **126** | **NON** | **Total** | | | **126** | **NON** | **Total** |
| **Sporadic** | **73** | **201** | **274** | | **Sporadic** | **69.14** | **204.86** | **274.00** |
| **Control** | **35** | **119** | **154** | | **Control** | **38.86** | **115.14** | **154.00** |
| **Total** | **108** | **320** | **428** | | **Total** | **108.00** | **320.00** | **428.00** |
| **X^2** | **0.809476** | | | | | | | |
| **P- VALUE** | **0.368275** | | | | | | | |

Investigation of the presence of allele 120 in genotypes against none 120 genotypes, see Table 35.

**Table 35: Presence of allele 120 against all**

| presence of allele 120 against all | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observed | | | | | Expected | | | |
| | 120 | NON | Total | | | 120 | NON | Total |
| Sporadic | 58 | 216 | 274 | | Sporadic | 58.90 | 215.10 | 274.00 |
| Control | 34 | 120 | 154 | | Control | 33.10 | 120.90 | 154.00 |
| Total | 92 | 336 | 428 | | Total | 92.00 | 336.00 | 428.00 |
| X^2 | 0.048249 | | | | | | | |
| P- VALUE | 0.826138 | | | | | | | |

In this test, the purpose is to look at related groups i.e. allele 132 have two derivatives: 132 and 134) the total number of the alleles is 41+25+6=72 (group F), I will compare this group to the control group I and perform X^2 test.

Comparisons between two groups (Table 25) these groups are group F and group I in Table 36.

**Table 36: Presence of allele 132 against all**

| Presence of alleles group F against I in all | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observed | | | | | Expected | | | |
| | F | NON | Total | | | I | NON | Total |
| Sporadic | 72.00 | 175.00 | 247 | | sporadic | 74.29 | 172.71 | 247.00 |
| Control | 45 | 97 | 142 | | control | 42.71 | 99.29 | 142.00 |
| Total | 117 | 272 | 389 | | Total | 117.00 | 272.00 | 389.00 |
| X^2 | 0.275548 | | | | | | | |
| P- VALUE | 0.599634 | | | | | | | |

Comparisons between two groups (see Table 37) these groups are: group E and group H in Table 25

**Table 37: Presence of alleles group E against H in all**

| Presence of alleles group E against H in all | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observed | | | | | Expected | | | |
| | E | NON | Total | | | H | NON | Total |
| Sporadic | 108.00 | 139.00 | 247 | | sporadic | 105.40 | 141.60 | 247.00 |
| Control | 58 | 84 | 142 | | control | 60.60 | 81.40 | 142.00 |
| Total | 166 | 223 | 389 | | Total | 166.00 | 223.00 | 389.00 |
| X^2 | 0.306161 | | | | | | | |
| P- VALUE | 0.580046 | | | | | | | |

Comparisons between two groups (see Table 38) these groups are: group D and group G in Table 25.

**Table 38: Presence of alleles group D against G in all**

| Presence of alleles group D against G in all | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observed | | | | Expected | | | | |
| | D | NON | Total | | G | NON | Total |
| Sporadic | 67.00 | 180.00 | 247 | sporadic | 67.31 | 179.69 | 247.00 |
| Control | 39 | 103 | 142 | control | 38.69 | 103.31 | 142.00 |
| Total | 106 | 283 | 389 | Total | 106.00 | 283.00 | 389.00 |
| X^2 | 0.005232 | | | | | | |
| P- VALUE | 0.942337 | | | | | | |

## 3.9 Results of the Endo VII on exon 56 of the fibrillin-1 gene

Two mutations were detected using FAM and HEX, no novel mutation is expected. See Figure 68.

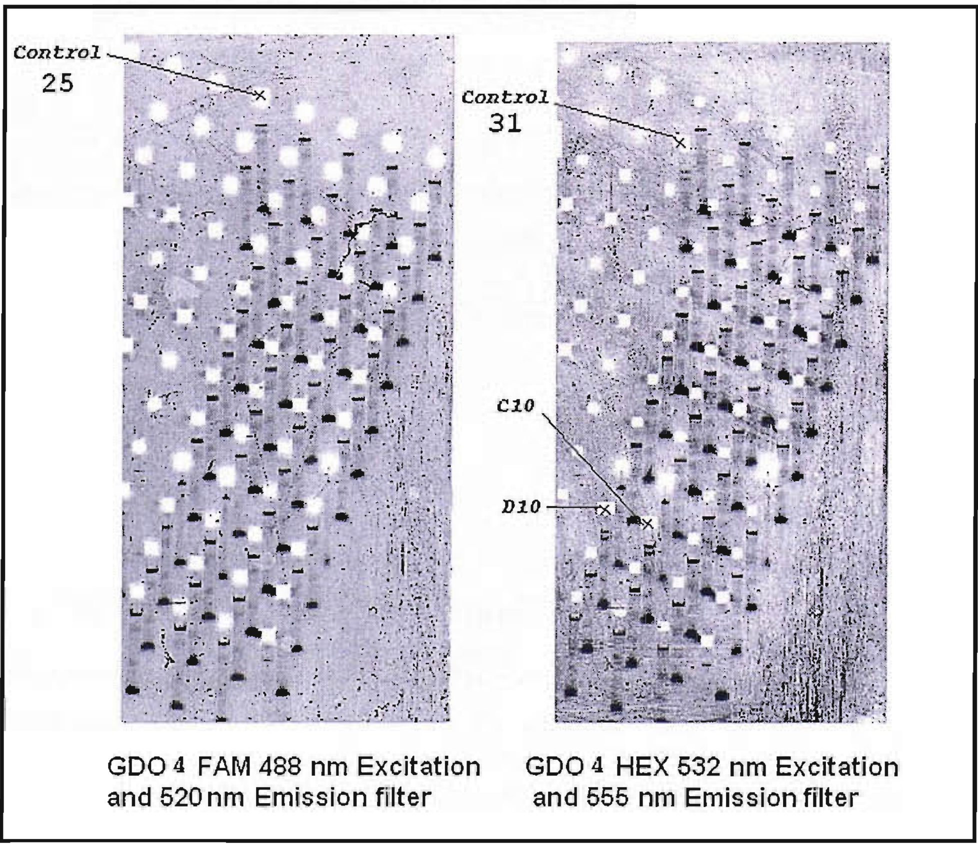



**Figure 68: Samples GDO C10, GDO D10 and control 31 all shows the same pattern, control 25 shows different pattern in different filter absorption.**

Exon 56 sequence DNA and aa. sequence forward and reveres primers are shown in blue and pink colour respectively see Figure 69

**EXON 56 (176 bp)**

```
preceding intron phase: 1
aataaaatcaaacag<-flank
frame 1 (1):   D  E  N  E  C  Q  T  K  P  G  I  C  E  N  G  R  C  L  N  T  R  G  S  Y  T  C  E  C
N  D
19511225
ATGAGAATGAATGTCAGACGAAGCCAGGGATCTGTGAGAATGGGCGCTGCCTCAACACCCGTGGGAGCTACACCTGTGAGTGTAATGATG
7005
.........................................................................................
               D  E  N  E  C  Q  T  K  P  G  I  C  E  N  G  R  C  L  N  T  R  G  S  Y  T  C  E  C
N  D
7005
.........................................................................................
               D  E  N  E  C  Q  T  K  P  G  I  C  E  N  G  R  C  L  N  T  R  G  S  Y  T  C  E  C
N  D

                                           intron phase: 1
                                           flank->gtgagtacagttggc
frame 1 (1):   G  F  T  A  S  P  N  Q  D  E  C  L
19511135       GGTTTACCGCCAGCCCCAACCAGGACGAGTGCCTTG
7095           ....................................
               G  F  T  A  S  P  N  Q  D  E  C  L
7095           ....................................
               G  F  T  A  S  P  N  Q  D  E  C  L
```

```
   216361 aaggagctcc atcctctata aaatggtcag atgactcttc ttgtttttgg
tccttcaata
   216421 aaatcaaaca gatgagaatg aatgtcagac gaagccaggg atctgtgaga
atgggcgctg
                  ss20009974 A/G rs363830 Gln [Q]> Gln [Q]
   216481 cctcaacacc cgtgggagct acacctgtga gtgtaatgat gggtttaccg
ccagccccaa
   216541 ccaggacgag tgccttggtg agtacagttg gcaccgcact ttcctaacct
cagcctccac
          ss461299 C/G rs363831 Glu [E]> Asp [D]
   216601 actgggatgc tggaaaccca gacttcttat ttaaaataca agaaaatgtc
aaaatctgag
```

Sequencing of 25F showed C/G heterozygous substitution, samples 31F and GDO C10F showed wild type C, **rs363831** Glu [E]> Asp [D] see Figure 70.

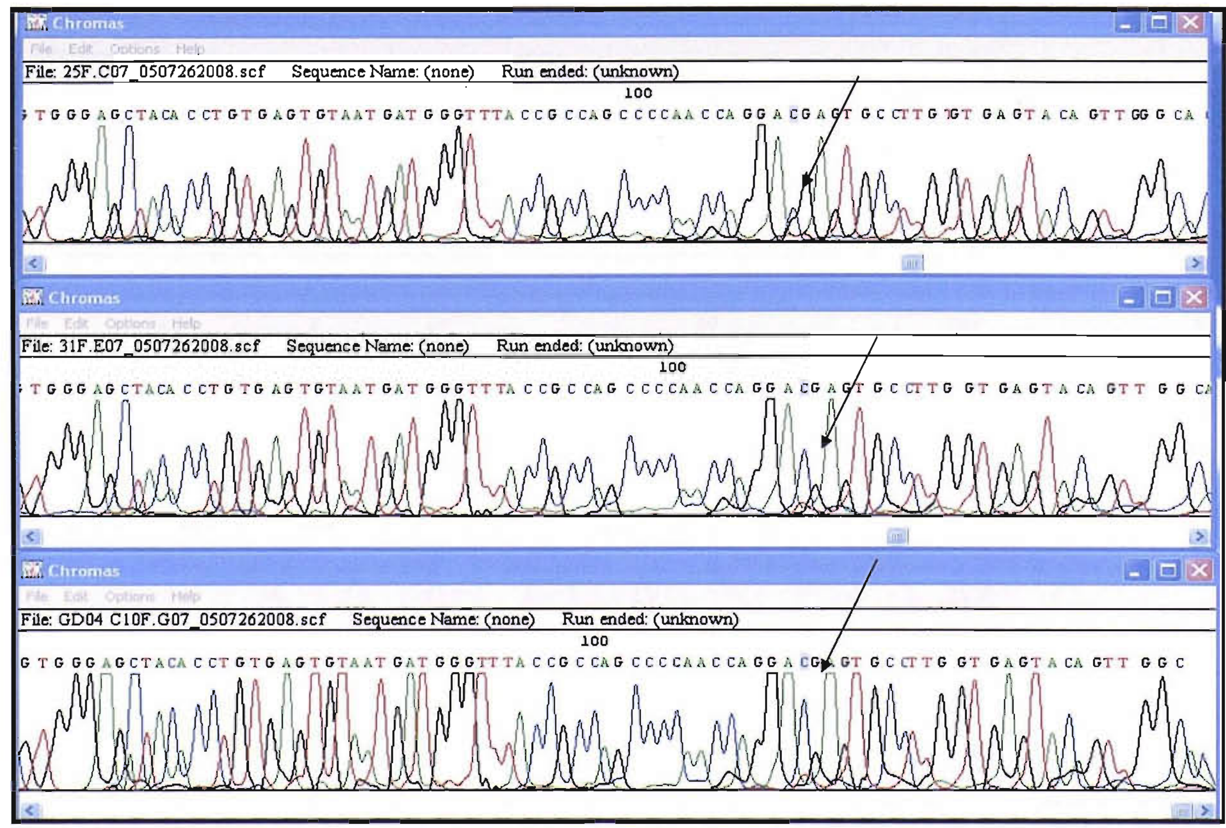



**Figure 70: Sample 25F shows C/G substitution wile C31F and GDO C10 F are normal, this SNP was reported as rs363831 Glu [E]> Asp [D].**

This sequencing shows a SNP that can result in the Endo VII pattern shown in samples GDO C10 which is the same as control 31( rs363831 Glu [E]> Asp [D]), here I shows the forward and reverse complement of the reverse sequencing (Figure 71), control 25 dose not show this variation ( Figure 72 )

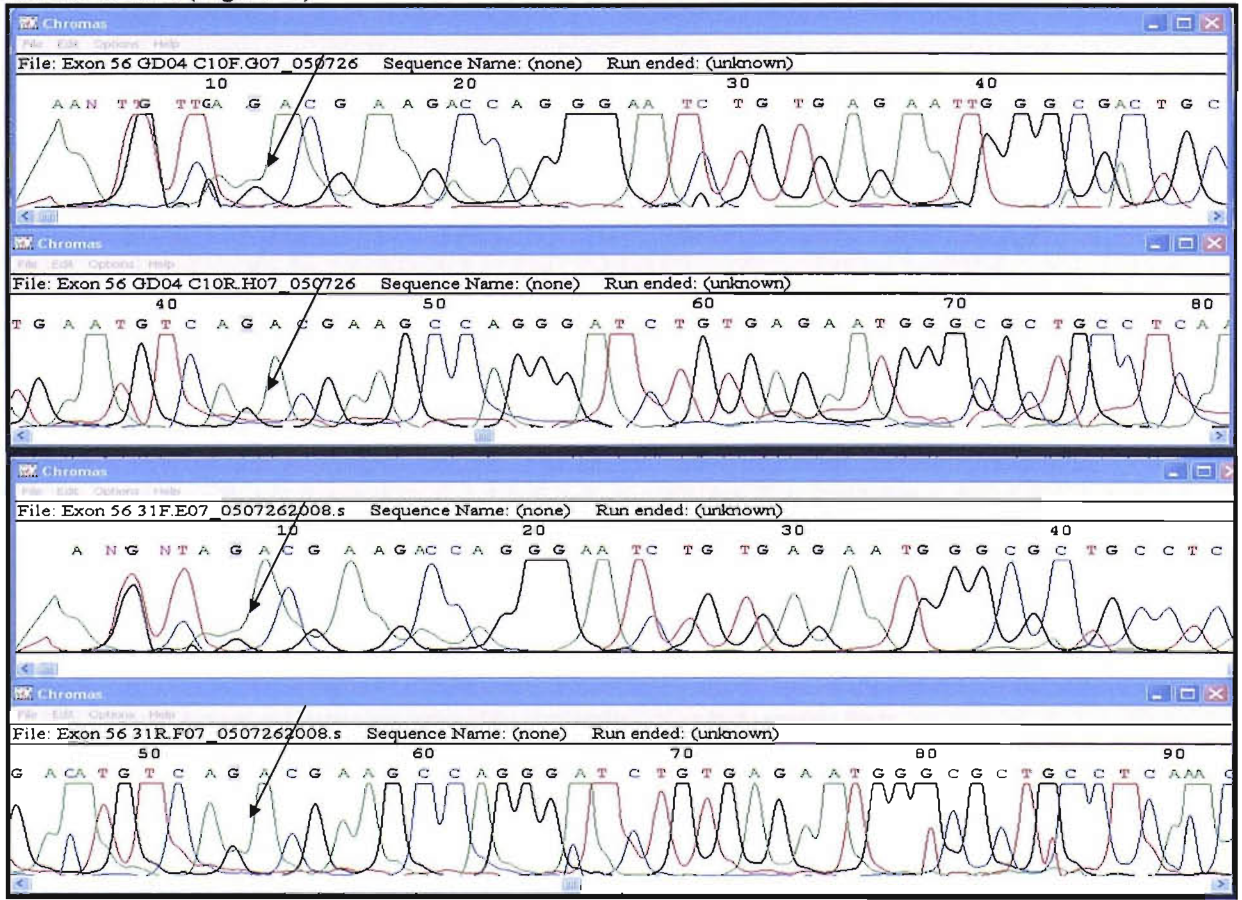



**Figure 71: A/G substitution was seemed in the forward and reverse sequencing this SNP is rs363830 Gln [Q]> Gln [Q], next figure shows the normal comparison.**

In Figure 72, sample 25 **does not** show G/A SNP.

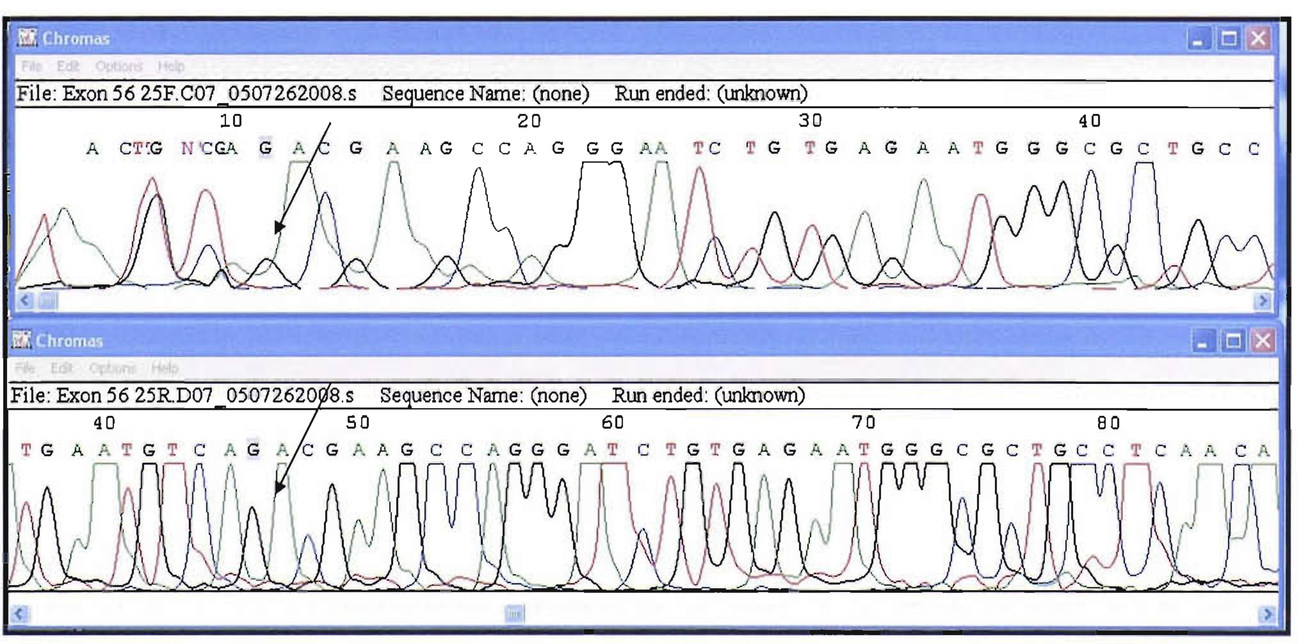



**Figure 72: The normal forward and reverse sequencing of the same position shown in the previous figure.**

### 3.9.1 PolyPhen analysis of the rs363831 Glu [E]> Asp [D] SNP

## Query

| Acc number | Position | $AA_1$ | $AA_2$ | Description |
|---|---|---|---|---|
| P35555 | 2329 | D | E | Fibrillin 1 precursor. LENGTH: 2871 AA |

## Prediction

**This variant is predicted to be unknown (no data for prediction)**

## Details

**PSIC PROFILE SCORES FOR TWO AMINO ACID VARIANTS**

| Score1 | Score2 | \|Score1-Score2\| | Observations | Diagnostics | Multiple alignment around substitution position |
|---|---|---|---|---|---|
| N/A | N/A | N/A | 0 | huge blast | N/A |

**MAPPING OF THE SUBSTITUTION SITE TO KNOWN PROTEIN 3D STRUCTURES**

| Database | Initial number of structures | Number of structures |
|---|---|---|
| PQS | 1038 | 0 |

### 3.9.2 Results ESE finder on exon 56 of the fibrillin-1 gene rs363831 SNP

**Sequence ID: Seq#1 wild type**
**Sequence:**
AGGACGAGTG

Length=10

Table 39 below shows the results of ESE finder on the normal sequence of exon 56

**Table 39: Results of ESE finder on the normal sequence of exon 56.**

| | SF2/ASF Thr=1.956 | | | SC35 Thr=2.383 | | | SRp40 Thr=2.67 | | | SRp55 Thr=2.676 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Position** | **Motif** | **Score** | **Position** | **Motif** | **Score** | **Position** | **Motif** | **Score** | **Position** | **Motif** | **Score** |
| 1 | AGGACGA | 3.390451 | | | | | | | | | |

**Sequence ID: Seq#1 rs363831 SNP**
**Sequence:**
AGGAGGAGTG

Length=10

Table 40 below shows the results of ESE finder on the mutant sequence of exon 56

**Table 40: Results of ESE finder on the mutant sequence of exon 56**

| | SF2/ASF Thr=1.956 | | | SC35 Thr=2.383 | | | SRp40 Thr=2.67 | | | SRp55 Thr=2.676 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Position** | **Motif** | **Score** | **Position** | **Motif** | **Score** | **Position** | **Motif** | **Score** | **Position** | **Motif** | **Score** |
| 1 | AGGAGGA | 2.786336 | | | | | | | | | |

## 3.10 Results RESCUE-ESE on exon 56 of the fibrillin-1 gene rs363831 SNP

RESCUE-ESE on the normal sequence of rs363831 is in Figure 73.

```
RESCUE-ESE output for 39913:993:16:16:12
RESCUE-ESE v 1.0 run date: 7/28/2005 time: 16:16:12  wild type
SPECIES: human
sequence length 10
total matches 0
unique matches 0

AGGACGAGTG
.........|
        10
```

**Figure 73: RESCUE-ESE on the normal sequence of rs363831**

RESCUE-ESE on the rs363831 is in Figure 74.

```
RESCUE-ESE output for 39914:7145:16:17:26
RESCUE-ESE v 1.0 run date: 7/28/2005 time: 16:17:26  rs363831 SNP
SPECIES: human
sequence length 10
total matches 2
unique matches 2

 GAGGAG
GGAGGA
AGGAGGAGTG
.........|
        10
```

**Figure 74: RESCUE-ESE on the rs363831**

## 3.11 Results of the Endo VII on exon 28 of the fibrillin-1 gene

Three patterns of bands were seen in Figure 75 three of these sequences were analysed by sequencing:

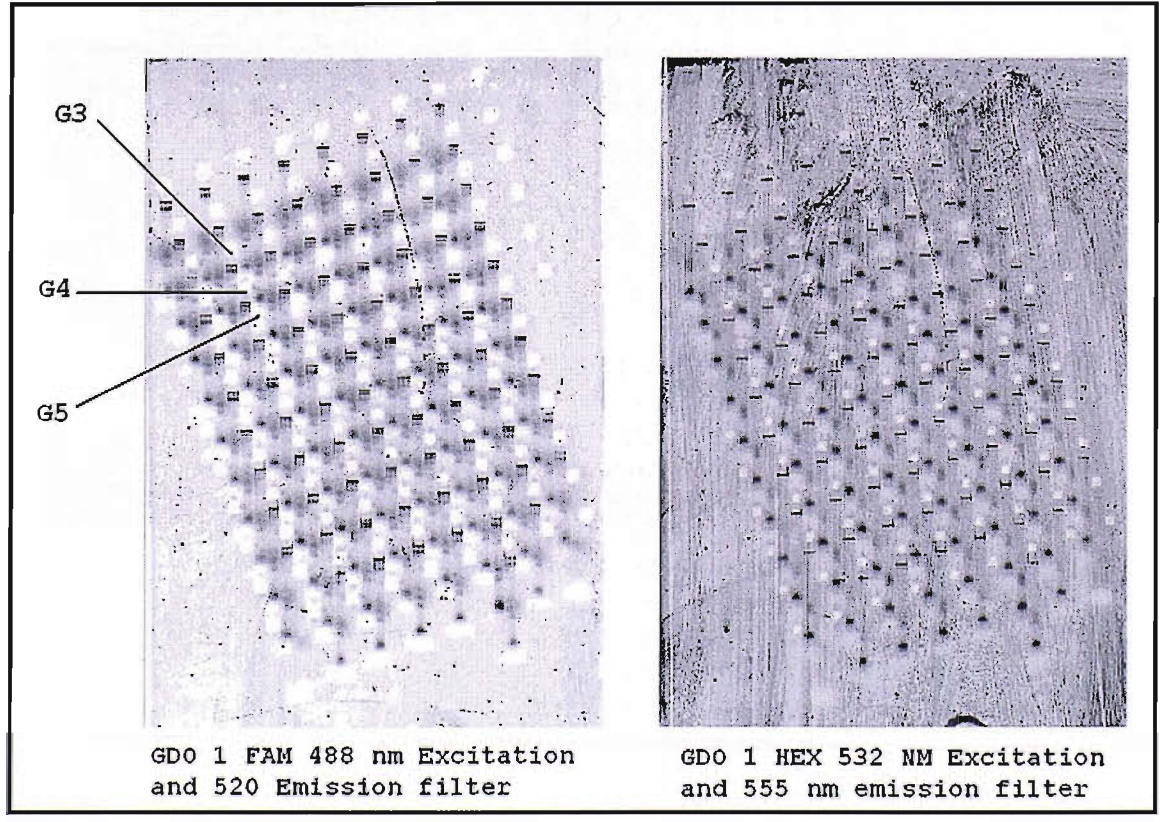



**Figure 75: After the Endo VII, three major patterns were seen: G5 with single band; G4 two very close bands and G3 with two bands more separated than G4.**

The sequence of exon 28 with its translation is in Figure 76, penta repeats are shown in red and orange colours.

```
                                   EXON 28 ( 298 bp)
preceding intron phase: 1
ttattttcccgacag<-flank
frame 1 (1):   D  I  N  E  C  E  L  S  A  H  L  C  P  N  G  R  C  V  N  L  I  G  K  Y  Q  C  A  C  N  P
19569954
ACATCAATGAATGTGAGCTGAGTGCACACCTGTGCCCCAATGGCCGTTGCGTGAACCTCATAGGGAAGTATCAGTGTGCCTGCAACCCTG
3597
..........................................................................................
               D  I  N  E  C  E  L  S  A  H  L  C  P  N  G  R  C  V  N  L  I  G  K  Y  Q  C  A  C  N  P
1076
..........................................................................................
               D  I  N  E  C  E  L  S  A  H  L  C  P  N  G  R  C  V  N  L  I  G  K  Y  Q  C  A  C  N  P
3597
..........................................................................................
               D  I  N  E  C  E  L  S  A  H  L  C  P  N  G  R  C  V  N  L  I  G  K  Y  Q  C  A  C  N  P

                                intron phase: 1
                                flank->gtaagttctttttta
frame 1 (1):   G  Y  H  S  T  P  D  R  L  F  C  V
19569864       GCTACCATTCAACTCCCGATAGGCTATTTTGTGTTG
3687           ....................................
               G  Y  H  S  T  P  D  R  L  F  C  V
1166           ....................................
               G  Y  H  S  T  P  D  R  L  F  C  V
3687           ....................................
               G  Y  H  S  T  P  D  R  L  F  C  V

   157621 ttgtggtcag ttgtttgaat gacatcattg ccaaagttgg aagcttatgt ttgggtgttt
   157681 gttcttctta ttttcccgac agacatcaat gaatgtgagc tgagtgcaca cctgtgcccc
   157741 aatggccgtt gcgtgaacct catagggaag tatcagtgtg cctgcaaccc tggctaccat
   157801 tcaactcccg ataggctatt ttgtgttggt aagttctttt ttattttatt ttattttatt
   157861 ttattttact ttattttatg ttatgttatg ttatgttatt ttattttatt ttatttcatc
   157921 tctccctaaa acactctatg tactctgagt tttgttttca tcgctacttt tacttatctt
```

Figure 77 shows sample GDO G3 forward with one less penta ATTTT, sample GDO G5 forward shows homozygous repeats. The approximate position is 150 (the electrophoresis of G3 shows two bands that are well separated).

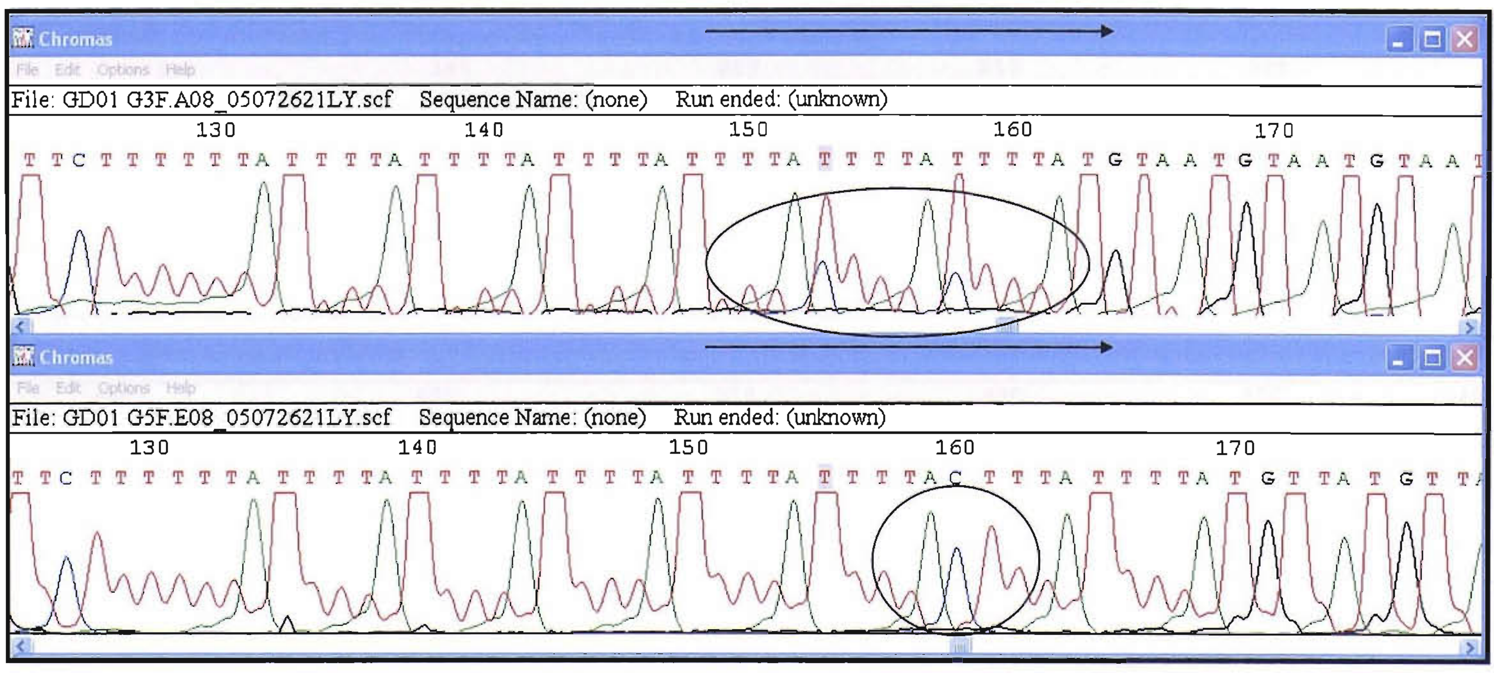



**Figure 77: Deletion of (ATTT) sequence in the above figure resulted in the appearance of CTTTA twice, the figure below shows one CTTTA, and the arrows describe the direction of sequencing.**

In this case the forward and reverse of sample GDO1G4 (the second and third sequences) shows two different sequencing results, this is due to a deletion of GTTAT in Figure 78. Note that G is half peak this made the reverse sequence with C move early. Please see Figure 79 to compare the forward and reverse on the same sequence, the approximate position of this insertion is 200 which goes with more close two bands in comparison to the G3 insertion as in the previous figure.

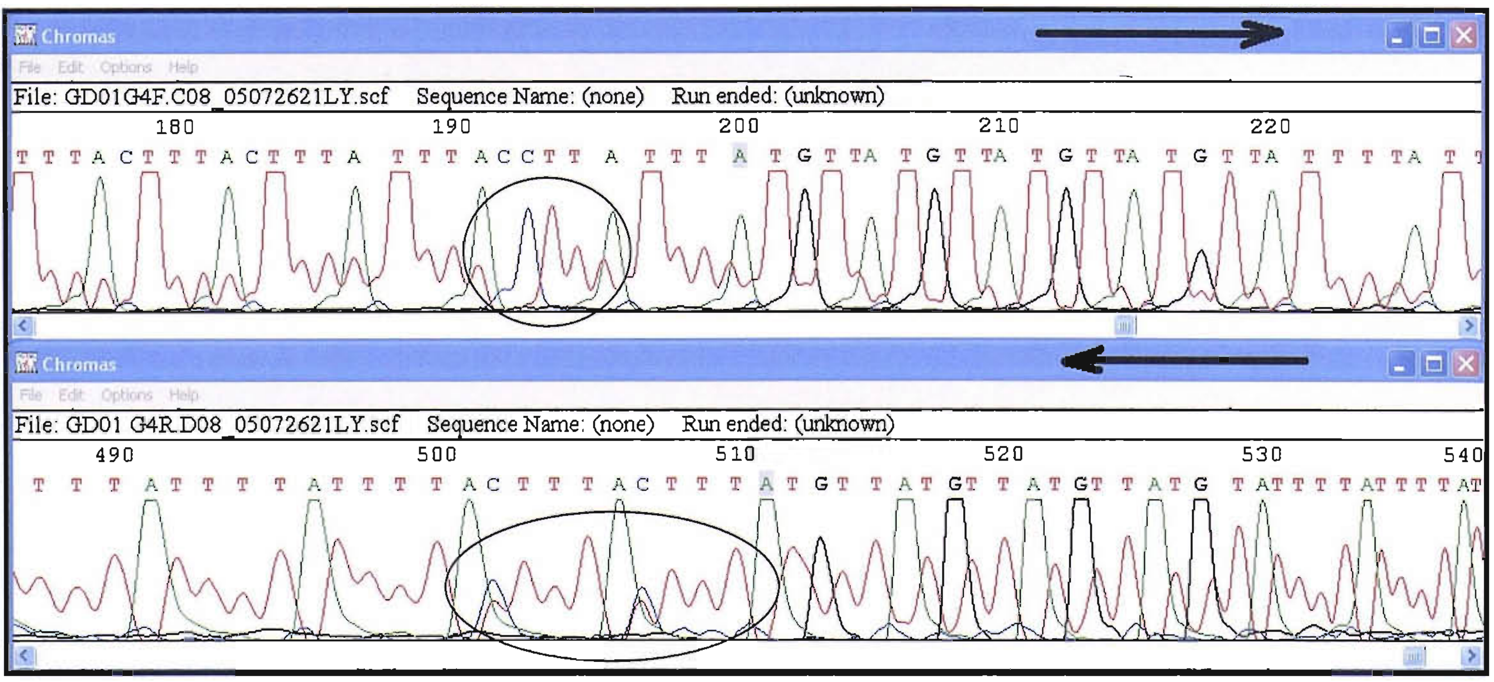



**Figure 78: This is the same sample but with forward and reverse sequencing results, the above sequencing shows one CTTTA because deletion of (GTTAT) occurred on the 5' region of the CTTTA sequence. In the lower diagram, a reverse complement of the same sequence is represented. Because the polymerase amplified two different VNTR alleles (one with deletion of GTTAT sequence) I have seen two sequences with CTTAT, arrows represent the Taq polymerase movement.**

Figure 79 shows sample GDO G5 forward and reverse I did not have two peaks of C, no insertion was seen after the A nucleotide.

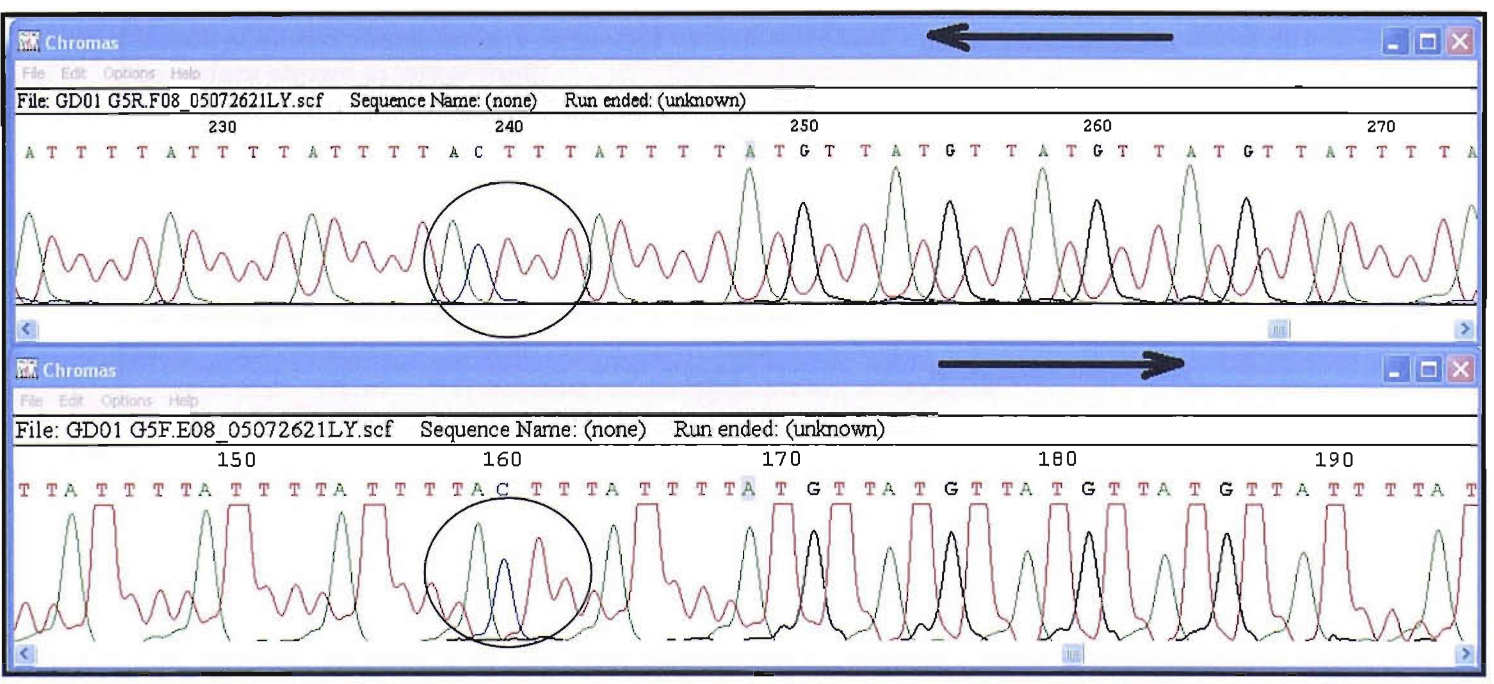



**Figure 79: This figure contains the same reactions as the previous figure but this time with homozygous VNTR. The arrows describe the direction of sequencing.**

### 3.11.1 Results of the intronic STR VNTR on splicing[314]

For the NCBI sequence the following splice donor and acceptor sites were predicted which shows no change of what is found in mRNA as in the following (Figure 80):

**Splice site predictions for one sequence with donor score cut-off 0.40, acceptor score cut-off 0.40 (exon/intron boundary shown in larger font):**

**Donor site predictions for 152.78.6.222.14222.0 :**

| Start | End | Score | Exon Intron |
|---|---|---|---|
| 202 | 216 | 0.99 | tgtgttggtaagttc |

**Acceptor site predictions for 152.78.6.222.14222.0 :**

| Start | End | Score | Intron Exon |
|---|---|---|---|
| 62 | 102 | 0.89 | ttcttcttattttcccgacagacatcaatgaatgtgagctg |

**Figure 80: After applying the sequence in the N N splice programme, I have scores of the donor and acceptor sites. No other sites were created in any case.**

After addition and deletion of 1-8 (atttt) penta to the 3' end of the intronic sequence, no change was seen of the prediction of the donor or acceptor splice sites.

For the addition and deletion of the 1-8 (gttat) penta to the 3' of the intronic sequence, no change was seen of the prediction of the donor or acceptor splice sites.

Addition and deletion of both 1-8 of the VNTRs showed no difference in the scores of the splicing sites.

## 3.12 Results summary of DNA sequencing for mutations detected by The Endo VII

Table 41 Shows the summary of sequencing results

**Table 41: Description of the variations detected by DNA sequencing**

| Exon Name | Domain Function | Any Known mutations | Sequencing Result | Mutation Type | Amino Acid Substitution | Known Possible Effect and Comments | Average allele frequency and ethnicity |
|---|---|---|---|---|---|---|---|
| 56 | **Fibrillin-Fibrillin interaction** | **Known** | G>A rs363830 | Exonic | Gln [Q]> Gln [Q] | **Synonymous, this SNP was seen in control sample ESE finder and rescue ESE were negative.** | **G=0.923, A=0.077 mixed** |
| 56 | **Fibrillin-Fibrillin interaction** | **Known** | G>C rs363831 | Exonic | Asp [D]> Glu [E] | **PolyPhen results shows unknown effect** | **G=0.5, C=0.5 mixed** |
| 28 | **Fibrillin-elastin interaction** | **Known** | ATTTT penta polymorphism | Intronic | None | **Association with arterial pulse pressure** | ? |
| 28 | **Fibrillin-elastin interaction** | **Unknown** | GTTAT penta polymorphism | Intronic | None | ? | ? |

# Chapter Four

## 4.0 Discussion and Conclusion

### 4.1 Elastin gene SNPs Genotyping

Statistical association analyses on sporadic SAH were performed using Phase and Arlequine (see Table 5). There was no significant haplotypic association with sporadic SAH and SAH in comparison to normal population.

In conclusion, this association shows that SAH may be associated with the elastin gene in different ethnicities or probably this may due to an increased genetic heterogeneity of intracranial aneurysm in Europeans compared with Japanese.

I have performed our SNPs haplotype analysis to investigate the association between sporadic SAH and haplotypes present in the elastin gene, a previous analysis was performed on Japanese cases and a positive association was seen. Also another study (James *et. al.* 2003) confirmed that the chromosome 7q11 locus is a predisposing factor for intracranial aneurysm (page 30) [1,2]. A third study was performed on 167 sporadic cases (Dutch patients) and found positive association between the elastin locus and SAH[318]. It is not necessary that the elastin gene is involved in the ICA and SAH, there is a possibility that this locus may contains a gene other than elastin that is involved in this type of arterial disease.

Our results gave negative association, this may be due to the reason that the original haplotype association were identified in Japanese subjects. For the Dutch study, it may suggest that there is more than one locus responsible for the onset of SAH. Consistent with our results, a study performed using two genotyped SNPs of the elastin gene in association study of SAH, in subjects from central Europe, showed negative association [319]. Another linkage study of the *ELN* locus performed on 14 families with 64 members concluded that the majority of aggregated intracranial aneurysms in the Japanese families may have a negative linkage to chromosome 7q11 [320]. In addition, another study found no association between haplotypes in the *ELN* gene in Caucasian populations and the presence of IA. This study was published late 2004 and used 120 case-control study and 170 controls to associate 8 different SNPs with ICA [321].

One of the studies suggested that chromosome 1p34.3-p36.13 is associated with ICA in autosomal dominant way, this study was performed on a big family with six affected living individuals using Affymetrix 10K Gene Chips then microsatellite analysis of 23 kindred members, LOD score was 4.2 [24]. In Finnish families, a confirmation of linkage to SAH was seen in chromosome 19 q13.3 using 139 affected sib pairs [102].

A genome-wide scan of 29 Japanese families suggested linkage region on chromosome 17cen and two studies speculated chromosomes 19q13 and Xp22, this study was performed on 29 ICA families each with three or more affected persons with SAH [322]. Finally the same group who suggested that *ELN* was linked to ICA (see the abstract) published another paper showing that *COL1A2* (and not the elastin gene) is the candidate gene for ICA[98].

## 4.2 Elastin Gene Scanning and Mutations

Mutations/ SNPs detected using these assays (concerning the coding region) were:
1-AC005056; 37759 G>A (rs2071307) glycine>serine (the only exonic SNP)
2- IVS20+17 T>C (ss4044368 T/C rs2856728)
3-IVS19+70 T>C (rs2239691)
4- IVS33-34 C>T (rs3757587)
5-HIDEAKI ONDA et al 2001 IVS23+24 T>C
6- IVS18+47G>C
7- IVS18+20DEL2 rs5884930 -GT.

One mutation was detected in the coding region of the elastin gene. A similar study performed by Nicole Berthelemy-Okazaki *et. al.* on 14 different patients affected with ICA. They have used DHPLC for mutation detection, for positive DHPLC results, sequencing was performed. The conclusion was that this analysis does not support *ELN* as the gene responsible for familial IA in the linked Utah IA pedigrees[323].

I have screened 34 exons and about 2 kb of 5' and 0.4 kb of the 3' regions of the elastin gene. In addition, more investigation should be performed to discover any possible association with ICA. Otherwise, there is a possibility that other regions in the elastin gene may cause ICA.

This gene was chosen because of two reasons, its potential role as a functional candidate gene and the genetic linkage evidence as described previously.
A detailed map of the elastin gene sequence with SNPs, exons and introns (**Appendix** E) was constructed, using gene bank accession number AC005056. Primers were designed to cover the exons and the splice junction sequences of the 5' flanking region and 3' flanking region of each exon.

### 4.2.1 Hydrophobic domain mutations

One mutation was seen in exon 20 (Table 22). This exon encodes for one of the hydrophobic domains, which may play an important role in the alignment of the tropoelastin protein, leading to the growth of the elastin chain. This specific alignment will permit proper cross-linking reactions (Figure 81).

The transition mutation (G to A), leading to glycine>serine substitution was seen in samples other than SAH patients. Moreover, glycine and serine are very similar amino acids (**Appendix D3** for glycine and **Appendix D4** for serine), so no major structural changes are expected. Further studies were performed to find if this mutation has a role in the onset of this disease.

*In silico* analyses were performed. The first programme called exonic splicing enhancers (ESE finder), I noticed that I have gained two sites: SC35 and SRp55 sequences ( Table 8, for the wild type and Table 9 for the SNP). In another programme called RESCUE-ESE, I did not find any sequence that can affect alternative splicing (Figure 36 and Figure 37).

This SNP was associated with Carotid Artery Distensibility Disorder in cases over 50 years old [324]. I have performed genotyping analysis using the LightTyper to analyse about 3000 samples from the BWHHS. Our results gave marginal evidence of an involvement of stroke (Table 10). This interesting result needs more investigation using stroke case study samples. I did not see any association with systolic; diastolic or pulse pressure, also I have performed association analysis between this SNP and sporadic SAH *vs*. control samples using additivity; dominant and recessive models no positive associations were seen. If this SNP is associated with stroke then a possible explanation can be due to a change in the arterial structure that facilitates occlusion that may lead to a stroke.

Hydrophilic substitution within hydrophobic domain can result in a defective alignment of the tropoelastin-cross linking domains, leading to the absence or reduced cross-links. These mutations may include transitions or transversion. Representation of the postulated alignment of the tropoelastin domains to allow proper interaction of the cross domains [180] is shown in Figure 81:

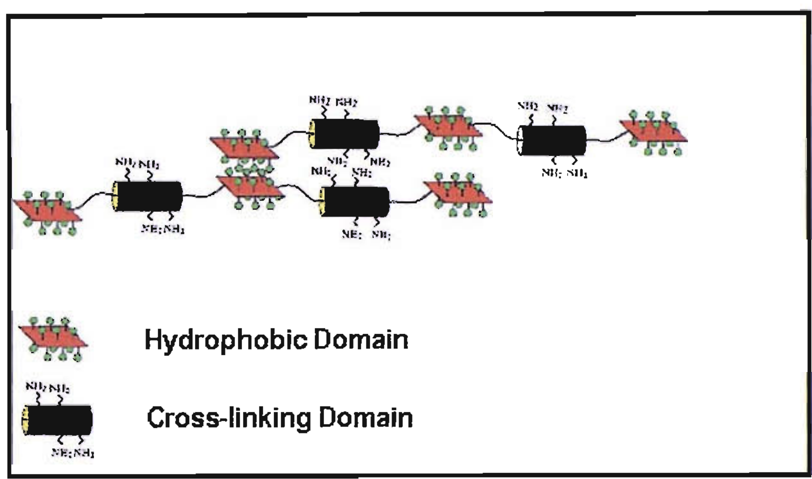



**Figure 81: Tropoelastin alignment:**

**Proper alignments of hydrophobic domain give a chance for the cross-linking domain to form cross-links by lysyl oxidase enzyme, depending on this assumption, point mutations on the hydrophobic domain (i.e. substitution with hydrophilic amino acid or micro deletions) can affect that proper alignment, while change on specific sites of the cross linking domain, results in changing in the cross-linking bonds.**

Mutation like glycine>serine (rs2071307) that may cause weakening of the elastin structure near the gap region of the bifurcation region. Aneurysms may start in early lifetime in an infant artery. A little turbulence may occur of blood flow past the V-shaped carina of the artery bifurcation, this may form a haemodynamic stress damages to the endothelium (this damage may occur faster in case of the glycine>serine mutation), pads of fibrous scar tissue may occur and change the conformation of the vessel carina, with the change in the direction of the haemodynamic stress, there is out-pouching of an aneurysm at the carina that in the future may leads to SAH.

### 4.2.2 Cross-Linking Domain Mutations

I did not detect any transition or transversion point mutations in the cross linking domain, a synonymous mutation (rs6979788 A>G ss10419219) present in exon 17 was not detected in our samples.

Concerning the cross-linking domain, the key amino acid is lysine, since it is responsible for the formation of desmosine / or isodesmosine cross links, between different tropoelastin domains. Other important mutations that were expected are the presence of aromatic amino acids like tyrosine or phenylalanine on the C-terminal side of lysine, this may prevent oxidation performed by lysyl oxidase [133,325] and favours the formation of lysinonorleucine [183].

Anyway, a study was performed to investigate the genetic variants in the lysyl oxidase gene. This study showed that the lysyl oxidase gene might not be involved in the aetiology of intracranial aneurysms. This analysis was performed on central Europe resident patients with ICA[326]. In this analysis lysyl oxidase gene was sequenced (seven exons; exon/intron splice sites and of the putative promoter region) in 25 patients.

### 4.2.3 5' and 3' Scanning of the elastin gene

No mutations were found in the 3' region of the elastin gene, this region was documented to have a function in the steady state stability of the mRNA of elastin protein[133,139] . In addition, 3'UTR distortion may contribute to low mRNA stability as it was proposed by Hew Y *et. al.*[162].

5' scanning also have functional domains[151,152], scanning of this region showed four variations, three of them are novel, see Table 23 and Table 24:

Creation of three domains (ZID; MZF1 and GATA-1) in the 5' region may be involved in the regulation of the elastin gene (as in Table 23). Creation of MTF and MIG1 with the abolishing of an SP1 domains (as in Table 24), suggests that more investigation is needed. Literature investigation shows no information in any of these variations until now. It will be an opportunity to search for these variations in relation to gene expression.

### 4.2.4 Other Mutations

All other mutations mentioned are intronic and they do not appear to cause any significant changes to the splicing morphology. Depending on our results, it seems that *ELN* (coding region) of our samples may not be associated with the intracranial aneurysm, furthermore, SAH may be associated with *ELN* in different ethnicities.

Our primers were designed to cover the splice donor and acceptor sites of each exon. Mutations at these sites can result in a truncated protein due to exon skipping or can result in the retention of a whole intron. In some instances, partial exclusion of normal exons can result, also a new exonic sequence can be seen in the case of introduction of a cryptic splice site. The result may also be a truncated protein due to a frameshift mutation. Unstable RNA transcripts or non-functional products due to loss of a crucial domain or part of a domain can result from exon deletions.

An example of a splice site mutation in *ELN* is C to G transversion within the acceptor splice site of exon 16 that was responsible for supravalvar aortic stenosis (SVAS). This mutation results in two abnormal elastin mRNA species, the first mRNA is generated by the activation of a cryptic splice site that lies within intron 15, consequently addition of 44 bp of intronic sequence to exon 16 was seen. This insertion creates a frame shift that results in an abnormal protein sequence resulting in a termination codon in exon 17. The second (smaller) abnormal mRNA arises as a consequence of the skipping of exon 16 [327].

Synonymous (silent) substitutions can result in a new codon coding for the same amino acid but can be a pathogenic; such substitutions can cause activation of a cryptic splice donor sequence that may lead to a loss or gain of translated amino acid sequences and possibly a frameshift mutation. After screening *ELN* gene, no mutations were observed at splice donor / accepter sites.

Single nucleotide deletions in *ELN* were reported to cause Cutis laxa. This deletion was present in exon 30, and caused translational frameshift mutation. Two such deletions were associated with the same disease [167]. No micro deletions were found in the coding region of *ELN*. Alternative polyadenelation sites may play a role in the stability of

mRNA and can influence the amount of expressed protein depending on the poly A signals, this could be a future functional study.

## 4.2.5 Other mutations not detected in our *ELN* samples

Some mutations were not detected in our samples and are summarised in Table 42. The reasons for not detecting them are also discussed.

Table 42: Mutations not found in our samples are summarised in this table, they are exonic, intronic and 5' UTR, calculation of the expected frequency was performed.

| Exon Name | Domain Function | Method Name of the Positive DHPLC Mutations not detected are "undetected" | Any Known mutations | Sequencing Result | Mutation Type | Amino Acid Substitution | Known Possible Effect and Comments | Average allele frequency and ethnicity | frequency of rare allele | the expected number of detected alleles = 16 x frequency of rare allele | the actual number of detected alleles depending on DHPLC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 26 | ophobic Domain | *undetected* | | DEAKI ONDA et al 2001 AC005056 : 33061 C/T | Intronic | | | .016, C=0.984 Japanese (ONDA et al 2001) | 0.016 | 0.256 | 0 |
| 22 | ophobic Domain | *undetected* | | s5996510 G/C rs4464848 | Exonic | anine > Proline | | ailable, TSC using SsahaSNP and SSAHA | 0 | 0 | 0 |
| 22 | ophobic Domain | *undetected* | | DEAKI ONDA et al 2001 | Exonic | synonymous | | A=0.011, G=0.989 Japanese | 0.011 | 0.176 | 0 |
| 17 | ROSS LINKING | *undetected* | | 79788 A>G ss10419219 | Exonic | synonymous | | Not available | 0 | 0 | 0 |
| 14 | ROSS LINKING | *undetected* | | DEAKI ONDA et al 2001 :005056: 45667 G>A+23 | Intronic | ? | | .016,G=0.984 Japanese (ONDA et al 2001) | 0.016 | 0.256 | 0 |
| 5 | ophobic Domain | *undetected* | | KI ONDA et al 2001 C/T | Exonic | lanine> Valine | | .021,C=0.979 Japanese (ONDA et al 2001) | 0.021 | 0.336 | 0 |
| 1st | PROMOTER | *undetected* | | ON-38 F95C/T HIDEAKI ONDA et al 2001 | 5'UTR | ? | | .021,C=0.979 Japanese (ONDA et al 2001) | 0.021 | 0.336 | 0 |
| 4TH | PROMOTER | *undetected* | | N -1042 C>T HIDEAKI ONDA et al 2001 | 5'UTR | ? | C>T Sequencing in six s(s1,s2,s3,s6,s7,s8) did not show this mutation | 202, C=0.798 Japanese (ONDA et al 2001) | 0.202 | 3.232 | 0 |
| 4th | PROMOTER | *undetected* | | ON -972 G>A HIDEAKI ONDA et al 2001 | 5'UTR | ? | G>A Sequencing in six s(s1,s2,s3,s6,s7,s8) did not show this mutation | .178, A=0.822 Japanese (ONDA et al 2001) | 0.178 | 2.848 | 0 |
| 6th | PROMOTER | *undetected* | | 4943616 C/A rs3757584 | 5'UTR | ? | equencing in two samples(s1,s3) did not show this mutation | .803, A=0.197 Japanese :PMID: 12436197 | 0.197 | 3.152 | 0 |
| 6th | PROMOTER | *undetected* | | 6500846 G/A rs3757583 | 5'UTR | ? | equencing in two samples(s1,s3) did not show this mutation | 0.725, G=0.275, SSAHA and WIBR fosmid SsahaSNP and Japanese :PMID: 12436197 | 0.275 | 4.4 | 0 |
| 20 | ophobic Domain | *undetected* | Known | AC005056 :37831 G>T (rs2229427 | Exonic | ne> Tryptophan | G>T Sequencing in six s(s1,s2,s3,s4,s5,s7) did not show this mutation | G=0.964, T=0.036 Multinational | 0.036 | 0.576 | 0 |

Six mutations were not detected in our samples (Table 42) and are having very low frequency in the mutant allele and they are:

1- AC005056 : 33061 C/T intronic (frequency of rare allele is 0.016 and the expected number to observe is 0.256)
2- G/A ONDA et al 2001exonic (frequency of rare allele is 0.011 and the expected number to observed is 0.176)
3- G/A ONDA et al 2001 AC005056: 45667 G>A+23 intronic (frequency of rare allele is 0.016 and the expected number to observed is 0.256)
4- ONDA et al 2001 C/T exonic (frequency of rare allele is 0.021 and the expected number to observe is 0.336)
5- Position-38 F95C/T ONDA et al 2001 5'UTR. (Frequency of rare allele is 0.021 and the expected number to observed is 0.336)
6- AC005056 :37831 G>T (rs2229427 exonic (frequency of rare allele is 0.036 and the expected number to observed is 0.576) concerning this mutation 12 out of 16 allele were sequenced with no evidence of this mutation.

Two mutations are not known to have any information regarding the frequency and they are (Table 42):
1- ss5996510 G/C rs4464848 exonic
2- rs6979788 A>G ss10419219 exonic

The last four mutations were having relatively high frequency mostly in Japanese and they are in (Table 42):

1- Position -1042 C>T ONDA et al 2001 5'UTR (the expected alleles number is 3.232)
2- Position -972 G>A ONDA et al 2001 5'UTR ( the expected alleles number is 2.848 )
3- ss4943616 C/A rs3757584 5'UTR (the expected alleles number is 3.152)
4- ss6500846 G/A rs3757583 5'UTR (the expected alleles number is 4.4)

To confirm the accuracy of our results, I have sequenced 12 out of 16 alleles (in each of: 1-position -1042 C>T ONDA et al 2001 2-position -972 G>A ONDA et al 2001), no mutations were detected. Moreover, four alleles of each of the third and fourth SNP were sequenced, no mutations were found.
DHPLC can detect up to 95-97% of mutations, a possibility of 3-5% missing mutations may explain these findings.

## 4.3 GT Repeat Models of *ELN* related to SAH

The *ELN* GT microsatellite (in any of the five models) did not show positive association in relation to sporadic or familial SAH.

A paper suggesting genetic anticipation in SAH emerged in 2003 (Struycken *et. al.*), I have observed a GT microsatellite mutation very close to the 3' end of exon 18 (13 nucleotides from the end of exon 18) called IVS18+20DEL2 rs5884930 –GT. Another study suggested a probable anticipation in familial SAH[73]. Most genetic diseases showing anticipation are due to expansions of triplet repeats, some GT repeats were involved in diseases, in particular haem oxygenase1 (HO-1) gene was shown to be associated with idiopathic recurrent miscarriage (IRM) via GT repeat variation [328].

GT repeat may be involved in splicing. Therefore, the sequence was entered to a software that predict possible splicing sites (a programme called NNSPLICE 9.0 version was used) [329]. Using this programme, many additions and deletions of the GT repeats in intron 18 of the *ELN* gene were performed, but the results of these modifications showed no differences in the predicted efficiency of splicing, nor creation of new splicing sites (see Figure 48). However, it is possible that the programme will not predict well the structural effect on the genome created by addition or deletion of the GTs.

GeneScan method was established to examine GT genotypes in familial and sporadic *vs.* controls. Primers labelled with 6-Carboxyfluorescein (6-FAM) were used to amplify the GT microsatellite region of intron 18 of the elastin gene.

The most prominent allele found in the GT study was allele 17 (i.e. 17 GT repeats), other alleles were detected are: allele 18; allele 19; allele 20; allele 21; allele 22 and allele 23 (Table 15 and Table 16). Five models were tested (in familial and sporadic SAH *vs.* control group).

### 4.3.1 Additivity Model

The first model was to test if the additivity of allele 17 is more associated with SAH than in control group, as the complex genetic disorders may involve more that one mutation that can contribute to the onset of a genetic disease[307,308], I tested this mutation if it can be a contribution factor for the onset of our disease. The result was negative see Table 17.

### 4.3.2 Major Expansion Model

The second model was to test whether the 17:17 genotype (homozygous of 17 GT repeat) is within the expected count. Because if the number of expected 17:17 genotype is significantly less that the real count, then the possibility of technical PCR failure or GeneScan detection failure is expected ( i.e. If the real genotype is 17 and 1000, then the large repeat fragment may not be amplified by PCR or it may not be detected by GeneScan method). The results of expected repeats were approximately the same as the real ones, so the major expansion model cannot explain the case. (See Table 18 and Table 19). For dinucleotides, very big expansions may not be the optimal expectation for the onset of this disease. However, anticipation was suggested in few papers and may be present in up to 10% of familial cases[25,73]. To rollout the possibility of major expansion I have tested this model, the weakness of this model is that I are trying to find significant results in uncommon cases (10% of the families).

### 4.3.3 Loss of Heterozygosity Model

In the third model I wanted to test whether SAH is associated with loss of heterozygosity, in this case I will have only one allele in SAH patients, results showed that the expected numbers of 17:17 are approximately the same as observed). This result was expected, since many cases of SVAS are associated with *ELN* deletion (dosage effect)[141,144,145]. However, a micro deletion involved in this region may contribute to the disease. See Table 18 and Table 19.

### 4.3.4 Recessive Model

The fourth model was to test if the 17:17 genotype may contribute to the SAH (recessivity for allele 17), the same calculations (of the major expansion model) are applied, and p-value was not significant. See Table 18 and Table 19. Some studies have shown a possible recessive type of inheritance, this type of inheritance may be applied to some cases of ICA leading to SAH [2,102,309,310].

### 4.3.5 Dominant Model

Finally, the dominant model, it was proposed to investigate of the presence of allele 17 in genotypes against none 17 genotypes in SAH vs. control group, also the result of this test was not significant. See Table 20 and Table 21. many papers suggested an autosomal dominant pattern of inheritance, this maybe the best model to be used in this study[24,69,305,306].

## 4.4 Linkage *TGFβRII* to aortic dissection

One of our families was having six affected persons with Marfan syndrome Type-2 (MFS2), also they showed ruptured intracranial aneurysms. Linkage investigation for markers spanning the *TGFβRII* gene was performed. The only marker that segregated with all disease cases was the 132 allele of D3S3727 marker (see Figure 62). This suggests that *TGFβRII* may be used to scan for SAH in some families, which is another factor of heterogeneity.

Because this STR was linked to this disease, microsatellite investigation was launched on sporadic SAH cases, I have used five different models as in the next section.

A new paper documented that mutation at position 460 in this gene is responsible for about 5% of familial thoracic aortic aneurysm [330]. This suggests that genetic screening of large number of cases may show few families that are associated with SAH, hence, I cannot detect statistical variation on 5% of cases.

## 4.5 D3S3727 STR in *TGFβRII* related to sporadic SAH

The *TGFβRII* microsatellite (in any of the five models) did not show positive association in relation to intracranial aneurysms.

The repeat D3S3727 (CA repeats) was found in intron one of the *TGFβRII*, there is no sufficient evidence to support functional effect of type II receptor gene. Since this marker was linked to MFS2 that shows ICA in our French family, it was worthwhile to scan our sporadic cases for possible association. Looking at Table 25 I can see that three major allele families are found and they are: The first is (118; 120); the second is (124, 126); the third family is (130,132,134).

From each of the three families, the most common member is 120 allele; the second family is 126; the third family is 132. These members were used to test our five models bellow, also I have tested these models on each family as a whole i.e. the counts in the family 118+120 (called group D) in sporadic SAH is 67 (see Table 25).

### 4.5.1 Additivity Model

The first model was to test if the additivity model on patients vs. control using alleles 120;126;132 and in (118+120 "groups D *vs*. G"); (124+126 "groups E *vs*. H"); (130;132;134 "groups F *vs*. I" ) to examine if they are more associated with sporadic SAH than in control group, the result was negative. (See Table 27Table 28Table 29, pages 143; 144 and 145). I did not find any association.

### 4.5.2 Major Expansion Model

The second model was to test whether the genotype [homozygous of 120; 126; 132 and (118+120 "groups D *vs*. G"); (124+126 "groups E *vs*. H"); (130; 132; 134 "groups F *vs*. I")] are within the expected count. Because if the number of expected genotypes are significantly less that the real count, then again it is possible that I have missed the expanded genotypes due to a problems mentioned before in the major expansion model in the *ELN* gene. The results of expected repeats were approximately the same as the

real ones, so the major expansion model cannot explain the case. (See Table 30; Table 31; Table 32 and pages 146; 147 and 148).

### 4.5.3 Loss of Heterozygosity Model

In this model I wanted to test whether intracranial aneurysms are associated with loss of heterozygosity, in this case I will have only one allele in SAH patients, results showed that the expected numbers of 120;126;132 and (118+120); (124+126); (130;132;134) genotypes are approximately the same as observed). (See Table 30; Table 31; Table 32 and pages 146; 147 and 148).

### 4.5.4 Recessive Model

The fourth model was to test if the recessivity of 120; 126; 132 and (118+120); (124+126); (130; 132; 134) genotypes may contribute to the intracranial aneurysms, the same calculations (of the major expansion model) are applied, p-value was not significant. (Table 30; Table 31; Table 32 and pages 146; 147 and 148).

### 4.5.5 Dominant Model

Finally, the dominant model, it was proposed to investigate of the homozygosity of the alleles 120; 126; 132. (Table 33 and Table 34 and Table 35). The group families (118+120); (124+126); (130;132;134) genotypes against none homozygous genotypes in SAH vs. control group see Table 36 and Table 37 and Table 38. In addition, the result of this test was not significant in all cases.

## 4.6 *FBN1* scanning of exons 27 28 and 56 using our sporadic SAH and control samples

Table 41 summarises DNA variations detected by the EndoVII method. Exons 27 and 28 may play a role in abdominal aneurysmal diseases, so these exons were screened in our sporadic SAH cases and exon 27 did not show any mutations.

For exon 56 two SNPs were detected both of which had been previously documented in NCBI databases. *In silico* analysis were performed using the PolyPhen programme, no prediction for Glu>Asp mutation was shown further analysis concerning potential splice enhancer effects of the SNP were performed using ESE-finder (see Table 39 and Table 40) and this showed no real difference.

However, using the RESCU-ESE program, I found that two new domains were created (see Figure 73 and Figure 74). These results suggest that functional analysis for possible splice effects would be worthwhile in the future.

For exon 28 three main pentanucleotides were seen (Figure 75 samples G3; G4 and G5). After sequencing these samples, it was shown that the band patterns in the electrophoresis were due to two types of penta nucleotide repeats (Figure 77 and Figure 78). In order to explore the possible effect on splicing, I used the NNSPLICE program, but no differences in splicing scores were detected (see Figure 80). A study showed a positive association between ATTTT repeat length polymorphism (which is the reverse complement of TAAAA and represents the same marker described in my study) and arterial pulse pressure in subjects over 50 years[6].

This marker is present in the intronic region of the fibrillin gene, NNSPLICE 9.0 program did not show any splicing effect, hence no functional property of this STR is expected, and it seems that this marker is acting as a linkage disequilibrium marker for some other functional sites close to it.

The detection of this marker by my new mutation scanning approach illustrates the potential of this approach to readily scan for microsatellite polymorphisms. Further studies of this locus would have been worthwhile, for example, association studies in other categories of stroke, but I did not have time for this during my thesis work

Page 185 missing

# Chapter Five

## 5.0 Further Studies

The out come of this research suggests the need for more investigation on the regulatory elements; SNPs and gene scanning for other genes. For review of the work performed in this project, see Table 43:

**Table 43: Work performed in this project on three different genes, the elastin (*ELN*); the fibrillin-1 (*FBN1*) and the transforming growth factor beta-receptor II gene (*TGFβRII*).**

| *ELN* | *FBN1* | *TGFβRII* |
| --- | --- | --- |
| 1- Detailed map of the elastin gene sequence with SNPs, exons and introns (**Appendix E**).<br>2- Complete amino acid sequence of the elastin gene with exons and cross-linking regions (**Appendix D**).<br>3- PCR optimisation for 34 exons of the elastin gene.<br>4- Scanning of the 34 exons using DHPLC.<br>5- Genotyping of three SNPs on sporadic SAH using ARMS and RFLP (**Appendix C**)<br>6- Statistical association analysis by the use of Phase and Arlequine on sporadic SAH and control.<br>7- DNA sequencing for the positive DHPLC results.<br>8- Fluorescence PCR optimization 6-carboxyfluorescein (6-FAM) to amplify the GT microsatellite region of the elastin gene.<br>9- GeneScan analysis of GT microsatellite in intron 18 of the elastin gene and describing five models.<br>10- Examining the mutations that were seen in *Eln* 5' flanking region to see any possible functions.<br>11- Optimising and genotyping 3000 subjects using odyssey **LightTyper** for exon 20 mutation.<br>12- Analysis the 3000 results.<br>13- Hypothesis for the SNPs on BWHHS samples.<br>14- Analysis of exon 20 SNP of sporadic SAH vs. control in GDO samples. | 15- Detailed map of the *FBN1* gene sequence with cSNPs, exons and introns (**Appendix K**).<br>16- Complete amino acid sequence of the fibrillin-1 gene with protein domains regions (**Appendix D**).<br>17- PCR optimisation for 65 exons of the *FBN1* gene.<br>18- Scanning three exons using the Endo VII MADGE for mutation detection on sporadic SAH.<br>19- DNA sequencing on new mutations found in the Endo VII technique and /or light scanner.<br>20- Analysis of the mutations. | 21- Linkage analysis using a family that may have aortic dissection an linked to *TGFβRII* gen using 5 STRs<br>22- Map of the markers used in the linkage analysis. ( Appendix I)<br>23- GeneScan for CA o control samples to see th frequency of the large allele A.<br>24- Perform about 214 HI and sporadic SAH samples looking for the allele frequency.<br>25- Association study with sporadic SAH.<br>26- Map of the CA in *TGFβRII* microsatellite |

Since I did not have enough time in my PhD to perform more analysis, I thought to put them as a possible future work.

### 5.1 Functional analysis of some elastin 5' SNPs/mutations

It is not yet known what are the functional impacts of the 5' polymorphisms (found in Table 22), especially the effect on gene expression. One of the SNPs with C>G at position -1162 was analysed by Ali-Baba and TFSEARCH programmes. Analysis results showed creation of functional domains. Alteration of gene expression due to these SNPs may be a causing factor for ICA. To investigate that, I need to perform real time PCR to compare the level of RNA expression of the elastin gene in normal and affected patients.

Mobility shift assay may also be applied to analyse the difference of protein binding(i.e. transcription factors/repression factors) on normal and mutant sequences, if there is any difference in the protein binding to the normal 5' DNA sequence, this may shows a different functionality and hence a possible effect on the onset of ICA.

### 5.2 Further analysis of the exonic SNP

The genotyping results of the [G>A (rs2071307) glycine>serine] performed on the BWHHS(Table 10) gave us an interesting results regarding the association with stroke. The p value of this analysis was 0.05, the number of stroke-affected patients is less than 200.

To gain more reliable results I need to perform this test again on a big stroke case-control study (about 10 000 subjects) to see if it is associated with stroke. Unfortunately, I do not have this big case-control stroke study samples to perform this analysis.

### 5.3 Screening for collagen 1 alpha 2 gene

Performing genetic screening for the collagen 1 alpha 2 gene (COL1A2 ) for the evidence of linkage analysis of the Japanese paper [98]. COL1A2 is considered as functional candidate gene[155,331]. Performing the same strategies as in this thesis to study this gene is

worthwhile, I may find few families that carry a defective gene that may be involved in ICA and SAH.

### 5.4 Studying genes that may be associated with ICAs and arterial architecture

SAH is associated with some familial genetic disorders like fibrillin-1, which is associated with lethal aortic rupture. The MADGE / T4 endonucleases digestion method may be performed to screen sporadic and familial SAH for mutations.

### 5.5 Fibrillin gene studies of two exonic SNPs

Using BWHHS cohort to perform genotyping of these SNPs (two exonic and two penta nucleotide repeats), RESCUE-ESE programme result on the rs363831 showed creation of two domains wile rs363830 did not show any difference. Summary of these variations are in Table 41. Looking for any association with pulse pressure; stroke; heart diseases are reasonable.

### 5.6 Genetic scanning of *TGFβRII* gene

The presence of a mutation at position 460 of this gene[330] suggests screening our sporadic and familial cases. There is a possibility to find this mutation in some familial or sporadic cases.

# Chapter Six

## 6.0 References

Reference List


1. Onda, H. *et al.* Genomewide-linkage and haplotype-association studies map intracranial aneurysm to chromosome 7q11. *Am.J Hum.Genet.* **69**, 804-819 (2001).

2. Farnham, J. M. *et al.* Confirmation of chromosome 7q11 locus for predisposition to intracranial aneurysm. *Hum.Genet.* **114**, 250-255 (2004).

3. Collod, G. *et al.* A second locus for Marfan syndrome maps to chromosome 3p24.2-p25. *Nat.Genet.* **8**, 264-268 (1994).

4. Nicod, P. *et al.* Familial aortic dissecting aneurysm. *J.Am.Coll.Cardiol.* **13**, 811-819 (1989).

5. Milewicz, D. M. Identification of defects in the fibrillin gene and protein in individuals with the Marfan syndrome and related disorders. *Tex.Heart Inst.J.* **21**, 22-29 (1994).

6. Powell, J. T., Turner, R. J., Henney, A. M., Miller, G. J., & Humphries, S. E. An association between arterial pulse pressure and variation in the fibrillin-1 gene. *Heart* **78**, 396-398 (1997).

7. Bou-Gharios, G., Ponticos, M., Rajkumar, V., & Abraham, D. Extra-cellular matrix in vascular networks. *Cell Prolif.* **37**, 207-220 (2004).

8. Internet. **Classification & Structure of Blood Vessels**. SEER . 2005.
Ref Type: Electronic Citation

9. Cardiovascular System. NEW YORK MEDICAL COLLEGE . 2005.
Ref Type: Electronic Citation

10. Zhang, B. *et al.* Molecular pathogenesis of subarachnoid haemorrhage. *Int.J Biochem Cell Biol.* **35**, 1341-1360 (2003).

11. rteaga-Solis, E., Gayraud, B., & Ramirez, F. Elastic and collagenous networks in vascular diseases. *Cell Struct.Funct.* **25**, 69-72 (2000).

12. Osakabe, T. *et al.* [Quantitative changes of elastin, fibrillin and collagen in abdominal aortic aneurysms]. *Nippon Ronen Igakkai Zasshi* **37**, 979-983 (2000).

13. Schwarze, U. *et al.* Haploinsufficiency for one COL3A1 allele of type III procollagen results in a phenotype similar to the vascular form of Ehlers-Danlos syndrome, Ehlers-Danlos syndrome type IV. *Am.J.Hum.Genet.* **69**, 989-1001 (2001).

14. Veldhuisen, B. *et al.* A spectrum of mutations in the second gene for autosomal dominant polycystic kidney disease (*PKD2*). *Am.J.Hum.Genet.* **61**, 547-555 (1997).

15. Mochizuki, T. *et al.* *PKD2*, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* **272**, 1339-1342 (1996).

16. Peral, B. *et al.* Identification of mutations in the duplicated region of the polycystic kidney disease 1 gene (*PKD1*) by a novel approach. *Am.J.Hum.Genet.* **60**, 1399-1410 (1997).

17. Hughes, J. *et al.* The polycystic kidney disease 1 (*PKD1*) gene encodes a novel protein with multiple cell recognition domains. *Nat.Genet.* **10**, 151-160 (1995).

18. Medley, T. L. *et al.* Fibrillin-1 genotype is associated with aortic stiffness and disease severity in patients with coronary artery disease. *Circulation* **105**, 810-815 (2002).

19. Curran, M. E. *et al.* The elastin gene is disrupted by a translocation associated with supravalvular aortic stenosis. *Cell* **73**, 159-168 (1993).

20. Olsen, D. R., Fazio, M. J., Shamban, A. T., Rosenbloom, J., & Uitto, J. Cutis laxa: reduced elastin gene expression in skin fibroblast cultures as determined by hybridizations with a homologous cDNA and an exon 1-specific oligonucleotide. *J.Biol.Chem.* **263**, 6465-6467 (1988).

21. Hinek, A. & Wilson, S. E. Impaired elastogenesis in Hurler disease: dermatan sulfate accumulation linked to deficiency in elastin-binding protein and elastic fiber assembly. *Am.J.Pathol.* **156**, 925-938 (2000).

22. Hinek, A. *et al.* Decreased elastin deposition and high proliferation of fibroblasts from Costello syndrome are related to functional deficiency in the 67-kD elastin-binding protein. *Am.J.Hum.Genet.* **66**, 859-872 (2000).

23. Puchner, M. J., Lohmann, F., Valdueza, J. M., Siepmann, G., & Freckmann, N. Monozygotic twins not identical with respect to the existence of intracranial aneurysms: a case report. *Surg.Neurol.* **41**, 284-289 (1994).

24. Nahed, B. V. *et al.* Mapping a Mendelian form of intracranial aneurysm to 1p34.3-p36.13. *Am.J.Hum.Genet.* **76**, 172-179 (2005).

25. Struycken, P. M. *et al.* Anticipation in familial intracranial aneurysms in consecutive generations. *Eur.J Hum.Genet.* **11**, 737-743 (2003).

26. Rock, M. J. *et al.* Molecular basis of elastic fiber formation. Critical interactions and a tropoelastin-fibrillin-1 cross-link. *J.Biol.Chem.* **279**, 23748-23758 (2004).

27. Loeys, B. L. *et al.* A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. *Nat.Genet.* **37**, 275-281 (2005).

28. Harrisons. *Principle of internal medicineHarrisons*, **vol.2**, 2388-2391 (2001).

29. Drazen *et al. Cecile, Text Book of Medicne*, **2**, 2109-2112 (2004).

30. Fogelholm, R., Hernesniemi, J., & Vapalahti, M. Impact of early surgery on outcome after aneurysmal subarachnoid hemorrhage. A population-based study. *Stroke* **24**, 1649-1654 (1993).

31. Kirkpatrick, P. J. Subarachnoid haemorrhage and intracranial aneurysms: what neurologists need to know. *J.Neurol.Neurosurg.Psychiatry* **73 Suppl 1**, i28-i33 (2002).

32. Choi, I. S. & David, C. Giant intracranial aneurysms: development, clinical presentation and treatment. *Eur.J Radiol.* **46**, 178-194 (2003).

33. Kaminogo, M., Yonekura, M., & Shibata, S. Incidence and outcome of multiple intracranial aneurysms in a defined population. *Stroke* **34**, 16-21 (2003).

34. White, P. M. & Wardlaw, J. M. Unruptured intracranial aneurysms. *J.Neuroradiol.* **30**, 336-350 (2003).

35. Matsumoto, K. M. *et al.* [Natural history of unruptured cerebral aneurysms of the unoperated and observed cases]. *No Shinkei Geka* **33**, 35-41 (2005).

36. Cajander, S. & Hassler, O. Enzymatic destruction of the elastic lamella at the mouth of cerebral berry aneurysm? An ultrastructural study with special regard to the elastic tissue. *Acta Neurol.Scand.* **53**, 171-181 (1976).

37. Skirgaudas, M., Awad, I. A., Kim, J., Rothbart, D., & Criscuolo, G. Expression of angiogenesis factors and selected vascular wall matrix proteins in intracranial saccular aneurysms. *Neurosurgery* **39**, 537-545 (1996).

38. Grond-Ginsbach, C. *et al.* Ultrastructural connective tissue aberrations in patients with intracranial aneurysms. *Stroke* **33**, 2192-2196 (2002).

39. Longstreth, W. T., Jr., Nelson, L. M., Koepsell, T. D., & van, B. G. Clinical course of spontaneous subarachnoid hemorrhage: a population-based study in King County, Washington. *Neurology* **43**, 712-718 (1993).

40. Linn, F. H., Rinkel, G. J., Algra, A., & van, G. J. Incidence of subarachnoid hemorrhage: role of region, year, and rate of computed tomography: a meta-analysis. *Stroke* **27**, 625-629 (1996).

41. Tomasello, F., D'Avella, D., Salpietro, F. M., & Longo, M. Asymptomatic aneurysms. Literature meta-analysis and indications for treatment. *J Neurosurg.Sci.* **42**, 47-51 (1998).

42. Rinkel, G. J., Djibuti, M., Algra, A., & Van Gijn, J. Prevalence and risk of rupture of intracranial aneurysms: a systematic review. *Stroke* **29**, 251-256 (1998).

43. King, J. T., Jr. Epidemiology of aneurysmal subarachnoid hemorrhage. *Neuroimaging Clin.N.Am.* 7, 659-668 (1997).

44. Auer, J., Berent, R., Weber, T., Lassnig, E., & Eber, B. Subarachnoid haemorrhage with "Ecstasy" abuse in a young adult. *Neurol.Sci.* **23**, 199-201 (2002).

45. Kissela, B. M. *et al.* Subarachnoid hemorrhage: a preventable disease with a heritable component. *Stroke* **33**, 1321-1326 (2002).

46. Isaksen, J., Egge, A., Waterloo, K., Romner, B., & Ingebrigtsen, T. Risk factors for aneurysmal subarachnoid haemorrhage: the Tromso study. *J Neurol.Neurosurg.Psychiatry* **73**, 185-187 (2002).

47. Knekt, P. *et al.* Risk factors for subarachnoid hemorrhage in a longitudinal population study. *J Clin.Epidemiol.* **44**, 933-939 (1991).

48. Kurth, T. *et al.* Smoking and risk of hemorrhagic stroke in women. *Stroke* **34**, 2792-2795 (2003).

49. Koda, Y., Seto, Y., Takeichi, S., & Kimura, H. Fatal subarachnoid hemorrhage complicating actinomycotic meningitis. *Forensic Sci.Int.* **134**, 169-171 (2003).

50. Horten, B. C., Abbott, G. F., & Porro, R. S. Fungal aneurysms of intracranial vessels. *Arch.Neurol.* **33**, 577-579 (1976).

51. Cohn-Zurita, F., Sandoval-Rivera, J. M., & Hernandez-Hernandez, F. [Association of aneurysm with pituitary adenoma]. *Cir.Cir.* **72**, 495-498 (2004).

52. Gelis, A., Pelissier, J., Blard, J. M., & Pages, M. [Intracranial haemorrhage associated with phaeochromocytoma]. *Rev.Neurol.(Paris)* **160**, 945-948 (2004).

53. Anderson, C. *et al.* Triggers of subarachnoid hemorrhage: role of physical exertion, smoking, and alcohol in the Australasian Cooperative Research on Subarachnoid Hemorrhage Study (ACROSS). *Stroke* **34**, 1771-1776 (2003).

54. Inagawa, T. Risk factors for aneurysmal subarachnoid hemorrhage in patients in Izumo City, Japan. *J.Neurosurg.* **102**, 60-67 (2005).

55. Broderick, J. P. *et al.* Major risk factors for aneurysmal subarachnoid hemorrhage in the young are modifiable. *Stroke* **34**, 1375-1381 (2003).

56. Sharma, P. & Brown, M. J. Neurovascular lessons from a pair of identical twins with cerebral aneurysms. *Postgrad.Med.J* 77, 197-198 (2001).

57. Slowik, A. *et al.* II genotype of the angiotensin-converting enzyme gene increases the risk for subarachnoid hemorrhage from ruptured aneurysm. *Stroke* **35**, 1594-1597 (2004).

58. Takenaka, K. *et al.* Angiotensin I-converting enzyme gene polymorphism in intracranial saccular aneurysm individuals. *Neurol.Res.* **20**, 607-611 (1998).

59. Pannu, H. *et al.* Lack of an association between the angiotensin-converting enzyme insertion/deletion polymorphism and intracranial aneurysms in a Caucasian population in the United States. *J.Neurosurg.* **103**, 92-96 (2005).

60. Jayaraman, T. *et al.* Tumor necrosis factor alpha is a key modulator of inflammation in cerebral aneurysms. *Neurosurgery* **57**, 558-564 (2005).

61. Hagen, T., Neidl, K., & Piepgras, U. Multiple cerebral aneurysms in identical twins. *AJNR Am.J Neuroradiol.* **18**, 973-976 (1997).

62. Rossetti, S. *et al.* Association of mutation position in polycystic kidney disease 1 (PKD1) gene and development of a vascular phenotype. *Lancet* **361**, 2196-2201 (2003).

63. Ronkainen, A. *et al.* Familial intracranial aneurysms. *Lancet* **349**, 380-384 (1997).

64. De Braekeleer, M., Perusse, L., Cantin, L., Bouchard, J. M., & Mathieu, J. A study of inbreeding and kinship in intracranial aneurysms in the Saguenay Lac-Saint-Jean region (Quebec, Canada). *Ann.Hum.Genet.* **60** ( **Pt 2**), 99-104 (1996).

65. Raaymakers, T. W., Rinkel, G. J., & Ramos, L. M. Initial and follow-up screening for aneurysms in families with familial subarachnoid hemorrhage. *Neurology* **51**, 1125-1130 (1998).

66. Gaist, D. *et al.* Risk of subarachnoid haemorrhage in first degree relatives of patients with subarachnoid haemorrhage: follow up study based on national registries in Denmark. *BMJ* **320**, 141-145 (2000).

67. Rossetti, S. *et al.* Association of mutation position in polycystic kidney disease 1 (PKD1) gene and development of a vascular phenotype. *Lancet* **361**, 2196-2201 (2003).

68. Leblanc, R., Melanson, D., Tampieri, D., & Guttmann, R. D. Familial cerebral aneurysms: a study of 13 families. *Neurosurgery* **37**, 633-638 (1995).

69. Bromberg, J. E. *et al.* Familial subarachnoid hemorrhage: distinctive features and patterns of inheritance. *Ann.Neurol.* **38**, 929-934 (1995).

70. Kojima, M. *et al.* Asymptomatic familial cerebral aneurysms. *Neurosurgery* **43**, 776-781 (1998).

71. Lindgaard, L., Eskesen, V., Gjerris, F., & Olsen, N. V. Familial aggregation of intracranial aneurysms in an Inuit patient population in Kalaallit Nunaat (Greenland). *Neurosurgery* **52**, 357-362 (2003).

72. Raaymakers, T. W. Aneurysms in relatives of patients with subarachnoid hemorrhage: frequency and risk factors. MARS Study Group. Magnetic Resonance Angiography in Relatives of patients with Subarachnoid hemorrhage. *Neurology* **53**, 982-988 (1999).

73. Ruigrok, Y. M., Rinkel, G. J., Wijmenga, C., & Van Gijn, J. Anticipation and phenotype in familial intracranial aneurysms. *J.Neurol.Neurosurg.Psychiatry* **75**, 1436-1442 (2004).

74. Michel, J. B. Acquired abdominal aortic aneurysm. *Nephrol.Dial.Transplant.* **13 Suppl 4**, 20-24 (1998).

75. Johnston, K. W. *et al.* Suggested standards for reporting on arterial aneurysms. Subcommittee on Reporting Standards for Arterial Aneurysms, Ad Hoc Committee on Reporting Standards, Society for Vascular Surgery and North American Chapter, International Society for Cardiovascular Surgery. *J.Vasc.Surg.* **13**, 452-458 (1991).

76. Milewicz, D. M., Urban, Z., & Boyd, C. Genetic disorders of the elastic fiber system. *Matrix Biol.* **19**, 471-480 (2000).

77. Mizuguchi, T. *et al.* Heterozygous *TGFBR2* mutations in Marfan syndrome. *Nat.Genet.* **36**, 855-860 (2004).

78. Khan, I. A. & Nair, C. K. Clinical, diagnostic, and management perspectives of aortic dissection. *Chest* **122**, 311-328 (2002).

79. Puech-Leao, P., Molnar, L. J., Oliveira, I. R., & Cerri, G. G. Prevalence of abdominal aortic aneurysms--a screening program in Sao Paulo, Brazil. *Sao Paulo Med.J.* **122**, 158-160 (2004).

80. Fleming, C., Whitlock, E. P., Beil, T. L., & Lederle, F. A. Screening for abdominal aortic aneurysm: a best-evidence systematic review for the U.S. Preventive Services Task Force. *Ann.Intern.Med.* **142**, 203-211 (2005).

81. Singh, K., Bonaa, K. H., Jacobsen, B. K., Bjork, L., & Solberg, S. Prevalence of and risk factors for abdominal aortic aneurysms in a population-based study : The Tromso Study. *Am.J.Epidemiol.* **154**, 236-244 (2001).

82. Nienaber, C. A. & Eagle, K. A. Aortic dissection: new frontiers in diagnosis and management: Part I: from etiology to diagnostic strategies. *Circulation* **108**, 628-635 (2003).

83. Takagi, H. & Umemoto, T. Homocysteinemia is a risk factor for aortic dissection. *Med.Hypotheses* **64**, 1007-1010 (2005).

84. Hoshina, K., Miyata, T., Hatakeyama, T., Shigematsu, H., & Nagawa, H. Hemodynamic influence of peripheral vascular occlusive disease on abdominal aortic aneurysms. *Int.Angiol.* **23**, 373-378 (2004).

85. Francke, U. *et al.* A Gly1127Ser mutation in an EGF-like domain of the fibrillin-1 gene is a risk factor for ascending aortic aneurysm and dissection. *Am.J.Hum.Genet.* 56, 1287-1296 (1995).

86. Milewicz, D. M. *et al.* Fibrillin-1 (*FBN1*) mutations in patients with thoracic aortic aneurysms. *Circulation* **94**, 2708-2711 (1996).

87. van Vlijmen-van Keulen CJ, Pals, G., & Rauwerda, J. A. Familial abdominal aortic aneurysm: a systematic review of a genetic background. *Eur.J.Vasc.Endovasc.Surg.* **24**, 105-116 (2002).

88. Lederle, F. A. *et al.* The aneurysm detection and management study screening program: validation cohort and final results. Aneurysm Detection and Management Veterans Affairs Cooperative Study Investigators. *Arch.Intern.Med.* **160**, 1425-1430 (2000).

89. Schievink, W. I. Genetics of intracranial aneurysms. *Neurosurgery* **40**, 651-662 (1997).

90. Rossetti, S. *et al.* Association of mutation position in polycystic kidney disease 1 (*PKD1*) gene and development of a vascular phenotype. *Lancet* **361**, 2196-2201 (2003).

91. Harikrishnan, S., Stigimon, J., & Tharakan, J. M. Intracranial aneurysms, coronary aneurysms and descending aortic coarctation--unreported association. *Int.J.Cardiol.* **99**, 329-330 (2005).

92. Triantafillidi, H. *et al.* Coronary artery ectasia, aneurysm of the basilar artery and varicose veins: common presentation or generalized defect of the vessel wall? A case report. *Angiology* **52**, 287-291 (2001).

93. Ruttimann, M., Perez, J. P., Richard, R., Brinquin, L., & Bonsignour, J. P. [Intracranial aneurysm and coronary ectasia]. *Presse Med.* **26**, 1141-1143 (1997).

94. Verloes, A., Sakalihasan, N., Koulischer, L., & Limet, R. Aneurysms of the abdominal aorta: familial and genetic aspects in three hundred thirteen pedigrees. *J.Vasc.Surg.* **21**, 646-655 (1995).

95. Kim, D. H., Van, G. G., & Milewicz, D. M. Familial aggregation of both aortic and cerebral aneurysms: evidence for a common genetic basis in a subset of families. *Neurosurgery* **56**, 655-661 (2005).

96. Milewicz, D. M. *et al.* Reduced penetrance and variable expressivity of familial thoracic aortic aneurysms/dissections. *Am.J.Cardiol.* **82**, 474-479 (1998).

97. van den Berg, J. S., Limburg, M., & Hennekam, R. C. Is Marfan syndrome associated with symptomatic intracranial aneurysms? *Stroke* **27**, 10-12 (1996).

98. Yoneyama, T. *et al.* Collagen type I alpha2 (COL1A2) is the susceptible gene for intracranial aneurysms. *Stroke* **35**, 443-448 (2004).

99. Kontusaari, S., Tromp, G., Kuivaniemi, H., Ladda, R. L., & Prockop, D. J. Inheritance of an RNA splicing mutation (G+ 1 IVS20) in the type III procollagen gene (COL3A1) in a family having aortic aneurysms and easy bruisability: phenotypic overlap between familial arterial aneurysms and Ehlers-Danlos syndrome type IV. *Am.J.Hum.Genet.* **47**, 112-120 (1990).

100. Menashi, S., Campa, J. S., Greenhalgh, R. M., & Powell, J. T. Collagen in abdominal aortic aneurysm: typing, content, and degradation. *J. Vasc.Surg.* **6**, 578-582 (1987).

101. Yamada, S. *et al.* Genome-wide scan for Japanese familial intracranial aneurysms: linkage to several chromosomal regions. *Circulation* **110**, 3727-3733 (2004).

102. van, d., V *et al.* Intracranial aneurysms in Finnish families: confirmation of linkage and refinement of the interval to chromosome 19q13.3. *Am.J.Hum.Genet.* **74**, 564-571 (2004).

103. Tsipouras, P. *et al.* Genetic linkage of the Marfan syndrome, ectopia lentis, and congenital contractural arachnodactyly to the fibrillin genes on chromosomes 15 and 5. The International Marfan Syndrome Collaborative Study. *N.Engl.J.Med.* **326**, 905-909 (1992).

104. Vaughan, C. J. *et al.* Identification of a chromosome 11q23.2-q24 locus for familial aortic aneurysm disease, a genetically heterogeneous disorder. *Circulation* **103**, 2469-2475 (2001).

105. Crowther, M., Goodall, S., Jones, J. L., Bell, P. R., & Thompson, M. M. Increased matrix metalloproteinase 2 expression in vascular smooth muscle cells cultured from abdominal aortic aneurysms. *J. Vasc.Surg.* **32**, 575-583 (2000).

106. Lindholt, J. S., Vammen, S., Fasting, H., Henneberg, E. W., & Heickendorff, L. The plasma level of matrix metalloproteinase 9 may predict the natural history of small abdominal aortic aneurysms. A preliminary study. *Eur.J. Vasc.Endovasc.Surg.* **20**, 281-285 (2000).

107. Busuttil, R. W., Rinderbriecht, H., Flesher, A., & Carmack, C. Elastase activity: the role of elastase in aortic aneurysm formation. *J.Surg.Res.* **32**, 214-217 (1982).

108. Mao, D., VanVickle, S. J., Curci, J. A., & Thompson, R. W. Expression of matrix metalloproteinases and TIMPs in human abdominal aortic aneurysms. *Ann. Vasc.Surg.* **13**, 236-237 (1999).

109. Brophy, C. M., Marks, W. H., Reilly, J. M., & Tilson, M. D. Decreased tissue inhibitor of metalloproteinases (TIMP) in abdominal aortic aneurysm tissue: a preliminary report. *J.Surg.Res.* **50**, 653-657 (1991).

110. Slowik, A. *et al.* Alpha1-antichymotrypsin gene (SERPINA3) A/T polymorphism as a risk factor for aneurysmal subarachnoid hemorrhage. *Stroke* **36**, 737-740 (2005).

111. Kadoglou, N. P. & Liapis, C. D. Matrix metalloproteinases: contribution to pathogenesis, diagnosis, surveillance and treatment of abdominal aortic aneurysms. *Curr.Med.Res.Opin.* **20**, 419-432 (2004).

112. Collins, F. S., Brooks, L. D., & Chakravarti, A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**, 1229-1231 (1998).

113. Taillon-Miller, P., Gu, Z., Li, Q., Hillier, L., & Kwok, P. Y. Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. *Genome Res.* **8**, 748-754 (1998).

114. Collins, F. S., Guyer, M. S., & Charkravarti, A. Variations on a theme: cataloging human DNA sequence variation. *Science* **278**, 1580-1581 (1997).

115. Chakravarti, A. It's raining SNPs, hallelujah? *Nat.Genet.* **19**, 216-217 (1998).

116. Syvanen, A. C., Landegren, U., Isaksson, A., Gyllensten, U., & Brookes, A. First International SNP Meeting at Skokloster, Sweden, August 1998. Enthusiasm mixed with scepticism about

single-nucleotide polymorphism markers for dissecting complex disorders. *Eur.J Hum.Genet.* **7**, 98-101 (1999).

117. Masood, E. As consortium plans free SNP map of human genome. *Nature* **398**, 545-546 (1999).

118. Wang, Z. & Moult, J. SNPs, protein structure, and disease. *Hum.Mutat.* **17**, 263-270 (2001).

119. Bracco, L. & Kearsey, J. The relevance of alternative RNA splicing to pharmacogenomics. *Trends Biotechnol.* **21**, 346-353 (2003).

120. Li, Y. C., Korol, A. B., Fahima, T., & Nevo, E. Microsatellites within genes: structure, function, and evolution. *Mol.Biol.Evol.* **21**, 991-1007 (2004).

121. Li, Y. C., Korol, A. B., Fahima, T., Beiles, A., & Nevo, E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol.Ecol.* **11**, 2453-2465 (2002).

122. Majewski, J. & Ott, J. GT repeats are associated with recombination on human chromosome 22. *Genome Res.* **10**, 1108-1114 (2000).

123. Hardy, D. O. *et al.* Androgen receptor CAG repeat lengths in prostate cancer: correlation with age of onset. *J.Clin.Endocrinol.Metab* **81**, 4400-4405 (1996).

124. Ritz, D., Lim, J., Reynolds, C. M., Poole, L. B., & Beckwith, J. Conversion of a peroxiredoxin into a disulfide reductase by a triplet repeat expansion. *Science* **294**, 158-160 (2001).

125. Masino, L. & Pastore, A. A structural approach to trinucleotide expansion diseases. *Brain Res.Bull.* **56**, 183-189 (2001).

126. Tom Strachan and Andrew P.Read. *Human Molecular Genetics* Oxford (2004).

127. Tidow, N. *et al.* Distinct amplification of an untranslated regulatory sequence in the egfr gene contributes to early steps in breast cancer development. *Cancer Res.* **63**, 1172-1178 (2003).

128. Pagani, F. *et al.* Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. *J.Biol.Chem.* **275**, 21041-21047 (2000).

129. Saveliev, A., Everett, C., Sharpe, T., Webster, Z., & Festenstein, R. DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. *Nature* **422**, 909-913 (2003).

130. Duval, A. & Hamelin, R. Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer Res.* **62**, 2447-2454 (2002).

131. Fazio, M. J. *et al.* Human elastin gene: new evidence for localization to the long arm of chromosome 7. *Am.J.Hum.Genet.* **48**, 696-703 (1991).

132. Rosenbloom, J. *et al.* Regulation of elastin gene expression. *Ann.N.Y.Acad.Sci.* **624**, 116-136 (1991).

133. Hinek, A., Keeley, F. W., & Callahan, J. Recycling of the 67-kDa elastin binding protein in arterial myocytes is imperative for secretion of tropoelastin. *Exp.Cell Res.* **220**, 312-324 (1995).

134. Olliver, L. *et al.* The gene coding for tropoelastin is represented as a single copy sequence in the haploid sheep genome. *Coll.Relat Res.* 7, 77-89 (1987).

135. Bashir, M. M. *et al.* Characterization of the complete human elastin gene. Delineation of unusual features in the 5'-flanking region. *J.Biol.Chem.* **264**, 8887-8891 (1989).

136. Yeh, H. *et al.* Structure of the bovine elastin gene and S1 nuclease analysis of alternative splicing of elastin mRNA in the bovine nuchal ligament. *Biochemistry* **28**, 2365-2370 (1989).

137. Internet. GenBank Overview. Internet . 2004.
Ref Type: Electronic Citation

138. Cicila, G. *et al.* Structure of the 3' portion of the bovine elastin gene. *Biochemistry* **24**, 3075-3080 (1985).

139. Indik, Z. *et al.* Structure of the elastin gene and alternative splicing of elastin mRNA: implications for human disease. *Am.J Med.Genet.* **34**, 81-90 (1989).

140. Pierce, R. A., Deak, S. B., Stolle, C. A., & Boyd, C. D. Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning. *Biochemistry* **29**, 9677-9683 (1990).

141. Tassabehji, M. *et al.* Elastin: genomic structure and point mutations in patients with supravalvular aortic stenosis. *Hum.Mol.Genet.* **6**, 1029-1036 (1997).

142. Indik, Z. *et al.* Structure of the 3' region of the human elastin gene: great abundance of Alu repetitive sequences and few coding sequences. *Connect.Tissue Res.* **16**, 197-211 (1987).

143. Rosenbloom, J., Abrams, W. R., & Mecham, R. Extracellular matrix 4: the elastic fiber. *FASEB J.* **7**, 1208-1218 (1993).

144. Olson, T. M. *et al.* A 30 kb deletion within the elastin gene results in familial supravalvular aortic stenosis. *Hum.Mol.Genet.* **4**, 1677-1679 (1995).

145. Urban, Z. *et al.* Supravalvular aortic stenosis: genetic and molecular dissection of a complex mutation in the elastin gene. *Hum.Genet.* **109**, 512-520 (2001).

146. Foster, K. *et al.* Description of a dinucleotide repeat polymorphism in the human elastin gene and its use to confirm assignment of the gene to chromosome 7. *Ann.Hum.Genet.* **57 ( Pt 2)**, 87-96 (1993).

147. Internet. *ELN* gene. 2004.
Ref Type: Internet Communication

148. Internet. *ELN* GENE NCBI. 2004.
Ref Type: Internet Communication

149. Pena, J. D., Roy, S., & Hernandez, M. R. Tropoelastin gene expression in optic nerve heads of normal and glaucomatous subjects. *Matrix Biol.* **15**, 323-330 (1996).

150. Parks, W. C. & Deak, S. B. Tropoelastin heterogeneity: implications for protein function and disease. *Am.J Respir.Cell Mol.Biol.* **2**, 399-406 (1990).

151. Johnson, D. J., Robson, P., Hew, Y., & Keeley, F. W. Decreased elastin synthesis in normal development and in long-term aortic organ and cell cultures is related to rapid and selective destabilization of mRNA for elastin. *Circ.Res.* **77**, 1107-1113 (1995).

152. Keeley, F. W. & Johnson, D. J. Measurement of the absolute synthesis of soluble and insoluble elastin by chick aortic tissue. *Can.J Biochem Cell Biol.* **61**, 1079-1084 (1983).

153. Kuang, P. P. *et al.* NF-kappaB induced by IL-1beta inhibits elastin transcription and myofibroblast phenotype. *Am.J Physiol Cell Physiol* **283**, C58-C65 (2002).

154. Carreras, I., Rich, C. B., Panchenko, M. P., & Foster, J. A. Basic fibroblast growth factor decreases elastin gene transcription in aortic smooth muscle cells. *J Cell Biochem* **85**, 592-600 (2002).

155. Seyama, Y. & Wachi, H. Atherosclerosis and matrix dystrophy. *J.Atheroscler.Thromb.* **11**, 236-245 (2004).

156. Jensen, D. E., Rich, C. B., Terpstra, A. J., Farmer, S. R., & Foster, J. A. Transcriptional regulation of the elastin gene by insulin-like growth factor-I involves disruption of Sp1 binding. Evidence for the role of Rb in mediating Sp1 binding in aortic smooth muscle cells. *J Biol.Chem.* **270**, 6555-6563 (1995).

157. McGowan, S. E. Influences of endogenous and exogenous TGF-beta on elastin in rat lung fibroblasts and aortic smooth muscle cells. *Am.J Physiol* **263**, L257-L263 (1992).

158. Kucich, U., Rosenbloom, J. C., Abrams, W. R., Bashir, M. M., & Rosenbloom, J. Stabilization of elastin mRNA by TGF-beta: initial characterization of signaling pathway. *Am.J Respir.Cell Mol.Biol.* **17**, 10-16 (1997).

159. Onda, H. *et al.* Endoglin is not a major susceptibility gene for intracranial aneurysm among Japanese. *Stroke* **34**, 1640-1644 (2003).

160. Marigo, V., Volpin, D., & Bressan, G. M. Regulation of the human elastin promoter in chick embryo cells. Tissue-specific effect of TGF-beta. *Biochim.Biophys.Acta* **1172**, 31-36 (1993).

161. Hew, Y., Grzelczak, Z., Lau, C., & Keeley, F. W. Identification of a large region of secondary structure in the 3'-untranslated region of chicken elastin mRNA with implications for the regulation of mRNA stability. *J Biol.Chem.* **274**, 14415-14421 (1999).

162. Hew, Y., Lau, C., Grzelczak, Z., & Keeley, F. W. Identification of a GA-rich sequence as a protein-binding site in the 3'-untranslated region of chicken elastin mRNA with a potential role in the developmental regulation of elastin mRNA stability. *J Biol.Chem.* **275**, 24857-24864 (2000).

163. Swee, M. H., Parks, W. C., & Pierce, R. A. Developmental regulation of elastin production. Expression of tropoelastin pre-mRNA persists after down-regulation of steady-state mRNA levels. *J Biol.Chem.* **270**, 14899-14906 (1995).

164. Kahari, V. M. Dexamethasone suppresses elastin gene expression in human skin fibroblasts in culture. *Biochem Biophys.Res.Commun.* **201**, 1189-1196 (1994).

165. McGowan, S. E., Liu, R., Harvey, C. S., & Jaeckel, E. C. Serine proteinase inhibitors influence the stability of tropoelastin mRNA in neonatal rat lung fibroblast cultures. *Am.J Physiol* **270**, L376-L385 (1996).

166. Zhang, M., Pierce, R. A., Wachi, H., Mecham, R. P., & Parks, W. C. An open reading frame element mediates posttranscriptional regulation of tropoelastin and responsiveness to transforming growth factor beta1. *Mol.Cell Biol.* **19**, 7314-7326 (1999).

167. Zhang, M. C. *et al.* Cutis laxa arising from frameshift mutations in exon 30 of the elastin gene *(ELN)*. *J Biol.Chem.* **274**, 981-986 (1999).

168. Pierce, R. A., Kolodziej, M. E., & Parks, W. C. 1,25-Dihydroxyvitamin D3 represses tropoelastin expression by a posttranscriptional mechanism. *J.Biol.Chem.* **267**, 11593-11599 (1992).

169. Sandberg, L. B., Weissman, N., & Smith, D. W. The purification and partial characterization of a soluble elastin-like protein from copper-deficient porcine aorta. *Biochemistry* **8**, 2940-2945 (1969).

170. Partridge, S. M. *et al.* Biosynthesis of the desmosine and isodesmosine cross-bridges in elastin. *Biochem.J.* **93**, 30C-33C (1964).

171. Keeley, F. W., Bellingham, C. M., & Woodhouse, K. A. Elastin as a self-organizing biomaterial: use of recombinantly expressed human elastin polypeptides as a model for investigations of structure and self-assembly of elastin. *Philos.Trans.R.Soc.Lond B Biol.Sci.* **357**, 185-189 (2002).

172. Brown, P. L., Mecham, L., Tisdale, C., & Mecham, R. P. The cysteine residues in the carboxy terminal domain of tropoelastin form an intrachain disulfide bond that stabilizes a loop structure and positively charged pocket. *Biochem.Biophys.Res.Commun.* **186**, 549-555 (1992).

173. Brown-Augsburger, P. *et al.* Microfibril-associated glycoprotein binds to the carboxyl-terminal domain of tropoelastin and is a substrate for transglutaminase. *J.Biol.Chem.* **269**, 28443-28449 (1994).

174. Debelle, L. & Tamburro, A. M. Elastin: molecular description and function. *Int.J.Biochem.Cell Biol.* **31**, 261-272 (1999).

175. Indik, Z. *et al.* Alternative splicing of human elastin mRNA indicated by sequence analysis of cloned genomic and complementary DNA. *Proc.Natl.Acad.Sci.U.S.A* **84**, 5680-5684 (1987).

176. Bedell-Hogan, D., Trackman, P., Abrams, W., Rosenbloom, J., & Kagan, H. Oxidation, cross-linking, and insolubilization of recombinant tropoelastin by purified lysyl oxidase. *J Biol.Chem.* **268**, 10345-10350 (1993).

177. Saunders, N. A. & Grant, M. E. The secretion of tropoelastin by chick-embryo artery cells. *Biochem.J.* **230**, 217-225 (1985).

178. Kao, W. W., Bressan, G. M., & Prockop, D. J. Kinetics of the incorporation of tropoelastin into elastic fibers in embryonic chick aorta. *Connect.Tissue Res.* **10**, 263-274 (1982).

179. Wachi, H., Seyama, Y., Yamashita, S., & Tajima, S. Cell cycle-dependent regulation of elastin gene in cultured chick vascular smooth-muscle cells. *Biochem.J.* **309** ( **Pt 2**), 575-579 (1995).

180. Saunders, N. A. & Grant, M. E. Elastin biosynthesis in chick-embryo arteries. Studies on the intracellular site of synthesis of tropoelastin. *Biochem.J.* **221**, 393-400 (1984).

181. Privitera, S., Prody, C. A., Callahan, J. W., & Hinek, A. The 67-kDa enzymatically inactive alternatively spliced variant of beta-galactosidase is identical to the elastin/laminin-binding protein. *J.Biol.Chem.* **273**, 6319-6326 (1998).

182. Hinek, A. & Rabinovitch, M. 67-kD elastin-binding protein is a protective "companion" of extracellular insoluble elastin and intracellular tropoelastin. *J.Cell Biol.* **126**, 563-574 (1994).

183. Vrhovski, B. & Weiss, A. S. Biochemistry of tropoelastin. *Eur.J Biochem* **258**, 1-18 (1998).

184. Davis, E. C. & Mecham, R. P. Selective degradation of accumulated secretory proteins in the endoplasmic reticulum. A possible clearance pathway for abnormal tropoelastin. *J.Biol.Chem.* **271**, 3787-3794 (1996).

185. Ikeda, K., Wachi, H., Seyama, Y., & Tajima, S. Effects of monensin on tropoelastin metabolism in vascular smooth muscle cells: monensin causes intracellular degradation of accumulated tropoelastin. *J.Biochem.(Tokyo)* **121**, 5-7 (1997).

186. Davis, E. C., Broekelmann, T. J., Ozawa, Y., & Mecham, R. P. Identification of tropoelastin as a ligand for the 65-kD FK506-binding protein, FKBP65, in the secretory pathway. *J.Cell Biol.* **140**, 295-303 (1998).

187. Damiano, V., Tsang, A., Kucich, U., Weinbaum, G., & Rosenbloom, J. Immuno electron microscopic studies on cells synthesizing elastin. *Connect.Tissue Res.* **8**, 185-188 (1981).

188. Mecham, R. P. *et al.* Elastin binds to a multifunctional 67-kilodalton peripheral membrane protein. *Biochemistry* **28**, 3716-3722 (1989).

189. Hinek, A., Mecham, R. P., Keeley, F., & Rabinovitch, M. Impaired elastin fiber assembly related to reduced 67-kD elastin-binding protein in fetal lamb ductus arteriosus and in cultured aortic smooth muscle cells treated with chondroitin sulfate. *J.Clin.Invest* **88**, 2083-2094 (1991).

190. Mecham, R. P., Whitehouse, L., Hay, M., Hinek, A., & Sheetz, M. P. Ligand affinity of the 67-kD elastin/laminin binding protein is modulated by the protein's lectin domain: visualization of elastin/laminin-receptor complexes with gold-tagged ligands. *J.Cell Biol.* **113**, 187-194 (1991).

191. Hinek, A., Wrenn, D. S., Mecham, R. P., & Barondes, S. H. The elastin receptor: a galactoside-binding protein. *Science* **239**, 1539-1541 (1988).

192. Bressan, G. M. *et al.* Banded fibers in tropoelastin coacervates at physiological temperatures. *J Ultrastruct.Res.* **82**, 335-340 (1983).

193. Cox, B. A., Starcher, B. C., & Urry, D. W. Coacervation of alpha-elastin results in fiber formation. *Biochim.Biophys.Acta* **317**, 209-213 (1973).

194. Tamburro, A. M., De Stradis, A., & D'Alessio, L. Fractal aspects of elastin supramolecular organization. *J Biomol.Struct.Dyn.* **12**, 1161-1172 (1995).

195. Li, B. & Daggett, V. Molecular basis for the extensibility of elastin. *J Muscle Res.Cell Motil.* **23**, 561-573 (2002).

196. Foster, J. A. *et al.* Isolation and characterization of cross-linked peptides from elastin. *J Biol.Chem.* **249**, 6191-6196 (1974).

197. Uitto, J. Biochemistry of the elastic fibers in normal connective tissues and its alterations in diseases. *J.Invest Dermatol.* **72**, 1-10 (1979).

198. Robert, L., Jacob, M. P., & Fulop, T. Elastin in blood vessels. *Ciba Found.Symp.* **192**, 286-299 (1995).

199. Ostuni, A., Lograno, M. D., Gasbarro, A. R., Bisaccia, F., & Tamburro, A. M. Novel properties of peptides derived from the sequence coded by exon 26A of human elastin. *Int.J.Biochem.Cell Biol.* **34**, 130-135 (2002).

200. Sakai, L. Y., Keene, D. R., & Engvall, E. Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils. *J.Cell Biol.* **103**, 2499-2509 (1986).

201. Dietz, H. C. *et al.* Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* **352**, 337-339 (1991).

202. Kielty, C. M. & Shuttleworth, C. A. Synthesis and assembly of fibrillin by fibroblasts and smooth muscle cells. *J.Cell Sci.* **106 ( Pt 1)**, 167-173 (1993).

203. Kitahama, S. *et al.* Expression of fibrillins and other microfibril-associated proteins in human bone and osteoblast-like cells. *Bone* **27**, 61-67 (2000).

204. Kanai, H., Umezu, M., Koide, K., & Hato, M. Ruptured intracranial aneurysm associated with unruptured abdominal aortic aneurysm--case report. *Neurol.Med.Chir (Tokyo)* **41**, 260-263 (2001).

205. Pereira, L. *et al.* Genomic organization of the sequence coding for fibrillin, the defective gene product in Marfan syndrome. *Hum.Mol.Genet.* **2**, 961-968 (1993).

206. Magenis, R. E., Maslen, C. L., Smith, L., Allen, L., & Sakai, L. Y. Localization of the fibrillin (*FBN*) gene to chromosome 15, band q21.1. *Genomics* **11**, 346-351 (1991).

207. Biery, N. J. *et al.* Revised genomic organization of *FBN1* and significance for regulated gene expression. *Genomics* **56**, 70-77 (1999).

208. Godfrey, M. From fluorescence to the gene: the skin in the Marfan syndrome. *J.Invest Dermatol.* **103**, 58S-62S (1994).

209. von, K. Y., Raghunath, M., & Nienaber, C. A. [Marfan syndrome: prevalence and natural course of cardiovascular manifestations]. *Z.Kardiol.* **87**, 150-160 (1998).

210. Lalchandani, S. & Wingfield, M. Pregnancy in women with Marfan's Syndrome. *Eur.J.Obstet.Gynecol.Reprod.Biol.* **110**, 125-130 (2003).

211. Pyeritz, R. E. The Marfan syndrome. *Annu.Rev.Med.* **51**, 481-510 (2000).

212. Gray, J. R. *et al.* Ascertainment and severity of Marfan syndrome in a Scottish population. *J.Med.Genet.* **31**, 51-54 (1994).

213. Collod-Beroud, G. *et al.* Update of the UMD-FBN1 mutation database and creation of an *FBN1* polymorphism database. *Hum.Mutat.* **22**, 199-208 (2003).

214. **The FBN1 mutations database.** Internet . 2005.

Ref Type: Electronic Citation

215. Robinson, P. N. *et al.* Mutations of *FBN1* and genotype-phenotype correlations in Marfan syndrome and related fibrillinopathies. *Hum.Mutat.* **20**, 153-161 (2002).

216. Lack, J. *et al.* Solution structure of the third TB domain from LTBP1 provides insight into assembly of the large latent complex that sequesters latent TGF-beta. *J.Mol.Biol.* **334**, 281-291 (2003).

217. Montgomery, R. A. *et al.* Multiple molecular mechanisms underlying subdiagnostic variants of Marfan syndrome. *Am.J.Hum.Genet.* **63**, 1703-1711 (1998).

218. Rantamaki, T., Kaitila, I., Syvanen, A. C., Lukka, M., & Peltonen, L. Recurrence of Marfan syndrome as a result of parental germ-line mosaicism for an *FBN1* mutation. *Am.J.Hum.Genet.* **64**, 993-1001 (1999).

219. Collod-Beroud, G. *et al.* Demonstration of the recurrence of Marfan-like skeletal and cardiovascular manifestations due to germline mosaicism for an FBN1 mutation. *Am.J.Hum.Genet.* **65**, 917-921 (1999).

220. Palz, M. *et al.* Clustering of mutations associated with mild Marfan-like phenotypes in the 3' region of *FBN1* suggests a potential genotype-phenotype correlation. *Am.J.Med.Genet.* **91**, 212-221 (2000).

221. Loeys, B., Nuytinck, L., Delvaux, I., De, B. S., & De, P. A. Genotype and phenotype analysis of 171 patients referred for molecular study of the fibrillin-1 gene *FBN1* because of suspected Marfan syndrome. *Arch.Intern.Med.* **161**, 2447-2454 (2001).

222. Sood, S., Eldadah, Z. A., Krause, W. L., McIntosh, I., & Dietz, H. C. Mutation in fibrillin-1 and the Marfanoid-craniosynostosis (Shprintzen-Goldberg) syndrome. *Nat.Genet.* **12**, 209-211 (1996).

223. Glesby, M. J. & Pyeritz, R. E. Association of mitral valve prolapse and systemic abnormalities of connective tissue. A phenotypic continuum. *JAMA* **262**, 523-528 (1989).

224. Dietz, H. C. *et al.* Four novel FBN1 mutations: significance for mutant transcript level and EGF-like domain calcium binding in the pathogenesis of Marfan syndrome. *Genomics* **17**, 468-475 (1993).

225. Silverman, D. I. *et al.* Life expectancy in the Marfan syndrome. *Am.J.Cardiol.* **75**, 157-160 (1995).

226. Sakai, L. Y., Keene, D. R., Glanville, R. W., & Bachinger, H. P. Purification and partial characterization of fibrillin, a cysteine-rich structural component of connective tissue microfibrils. *J Biol.Chem.* **266**, 14763-14770 (1991).

227. Sherratt, M. J. *et al.* Fibrillin-rich microfibrils of the extracellular matrix: ultrastructure and assembly. *Micron.* **32**, 185-200 (2001).

228. *FBN1*. GENE CARDS . 2005.
Ref Type: Electronic Citation

229. Whiteman, P. & Handford, P. A. Defective secretion of recombinant fragments of fibrillin-1: implications of protein misfolding for the pathogenesis of Marfan syndrome and related disorders. *Hum.Mol.Genet.* **12**, 727-737 (2003).

230. Ashworth, J. L. *et al.* Fibrillin degradation by matrix metalloproteinases: implications for connective tissue remodelling. *Biochem.J.* **340 ( Pt 1)**, 171-181 (1999).

231. Milewicz, D. M. *et al.* A mutation in *FBN1* disrupts profibrillin processing and results in isolated skeletal features of the Marfan syndrome. *J.Clin.Invest* **95**, 2373-2378 (1995).

232. Raghunath, M. *et al.* Carboxy-terminal conversion of profibrillin to fibrillin at a basic site by PACE/furin-like activity required for incorporation in the matrix. *J.Cell Sci.* **112 ( Pt 7)**, 1093-1100 (1999).

233. Baldock, C. *et al.* The supramolecular organization of fibrillin-rich microfibrils. *J.Cell Biol.* **152**, 1045-1056 (2001).

234. Qian, R. Q. & Glanville, R. W. Alignment of fibrillin molecules in elastic microfibrils is defined by transglutaminase-derived cross-links. *Biochemistry* **36**, 15841-15847 (1997).

235. Reinhardt, D. P. *et al.* Fibrillin-1 and fibulin-2 interact and are colocalized in some tissues. *J Biol.Chem.* **271**, 19489-19496 (1996).

236. Ramirez, F., Gayraud, B., & Pereira, L. Marfan syndrome: new clues to genotype-phenotype correlations. *Ann.Med.* **31**, 202-207 (1999).

237. Dietz, H. C. *et al.* Four novel *FBN1* mutations: significance for mutant transcript level and EGF-like domain calcium binding in the pathogenesis of Marfan syndrome. *Genomics* **17**, 468-475 (1993).

238. Hutchinson, S. *et al.* Allelic variation in normal human *FBN1* expression in a family with Marfan syndrome: a potential modifier of phenotype? *Hum.Mol.Genet.* **12**, 2269-2276 (2003).

239. Reinhardt, D. P. *et al.* Fibrillin-1 and fibulin-2 interact and are colocalized in some tissues. *J.Biol.Chem.* **271**, 19489-19496 (1996).

240. Rees, D. J. *et al.* The role of beta-hydroxyaspartate and adjacent carboxylate residues in the first EGF domain of human factor IX. *EMBO J.* **7**, 2053-2061 (1988).

241. Reinhardt, D. P., Ono, R. N., & Sakai, L. Y. Calcium stabilizes fibrillin-1 against proteolytic degradation. *J.Biol.Chem.* **272**, 1231-1236 (1997).

242. Schrijver, I., Liu, W., Brenn, T., Furthmayr, H., & Francke, U. Cysteine substitutions in epidermal growth factor-like domains of fibrillin-1: distinct effects on biochemical and clinical phenotypes. *Am.J.Hum.Genet.* **65**, 1007-1020 (1999).

243. Neptune, E. R. *et al.* Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. *Nat.Genet.* **33**, 407-411 (2003).

244. Raghunath, M. *et al.* Fibrillin and elastin expression in skin regenerating from cultured keratinocyte autografts: morphogenesis of microfibrils begins at the dermo-epidermal junction and precedes elastic fiber formation. *J Invest Dermatol.* **106**, 1090-1095 (1996).

245. Kielty, C. M. & Shuttleworth, C. A. Fibrillin-containing microfibrils: structure and function in health and disease. *Int.J.Biochem.Cell Biol.* **27**, 747-760 (1995).

246. Keene, D. R., Maddox, B. K., Kuo, H. J., Sakai, L. Y., & Glanville, R. W. Extraction of extendable beaded structures and their identification as fibrillin-containing extracellular matrix microfibrils. *J Histochem.Cytochem.* **39**, 441-449 (1991).

247. Pereira, L. *et al.* Targetting of the gene encoding fibrillin-1 recapitulates the vascular aspect of Marfan syndrome. *Nat.Genet.* **17**, 218-222 (1997).

248. Sakamoto, H. *et al.* Cell-type specific recognition of RGD- and non-RGD-containing cell binding domains in fibrillin-1. *J.Biol.Chem.* **271**, 4916-4922 (1996).

249. Wheatley, H. M. *et al.* Immunohistochemical localization of fibrillin in human ocular tissues. Relevance to the Marfan syndrome. *Arch.Ophthalmol.* **113**, 103-109 (1995).

250. Lin, H. Y., Wang, X. F., Ng-Eaton, E., Weinberg, R. A., & Lodish, H. F. Expression cloning of the TGF-beta type II receptor, a functional transmembrane serine/threonine kinase. *Cell* **68**, 775-785 (1992).

251. Wrana, J. L., Attisano, L., Wieser, R., Ventura, F., & Massague, J. Mechanism of activation of the TGF-beta receptor. *Nature* **370**, 341-347 (1994).

252. Yamashita, H., Ten, D. P., Franzen, P., Miyazono, K., & Heldin, C. H. Formation of hetero-oligomeric complexes of type I and type II receptors for transforming growth factor-beta. *J.Biol.Chem.* **269**, 20172-20178 (1994).

253. Bruce Alberts *et al. Molecular Biology of the Cell*(2002).

254. Wrana, J. L. TGF-beta receptors and signalling mechanisms. *Miner.Electrolyte Metab* **24**, 120-130 (1998).

255. Grady, W. M. *et al.* Mutational inactivation of transforming growth factor beta receptor type II in microsatellite stable colon cancers. *Cancer Res.* **59**, 320-324 (1999).

256. Lucke, C. D. *et al.* Inhibiting mutations in the transforming growth factor beta type 2 receptor in recurrent human breast cancer. *Cancer Res.* **61**, 482-485 (2001).

257. Parsons, R. *et al.* Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. *Cancer Res.* **55**, 5548-5550 (1995).

258. Oshima, M., Oshima, H., & Taketo, M. M. TGF-beta receptor type II deficiency results in defects of yolk sac hematopoiesis and vasculogenesis. *Dev.Biol.* **179**, 297-302 (1996).

259. Goumans, M. J., Lebrin, F., & Valdimarsdottir, G. Controlling the angiogenic switch: a balance between two distinct TGF-b receptor signaling pathways. *Trends Cardiovasc.Med.* **13**, 301-307 (2003).

260. Goumans, M. J. *et al.* Balancing the activation state of the endothelium via two distinct TGF-beta type I receptors. *EMBO J.* **21**, 1743-1753 (2002).

261. Cheifetz, S. *et al.* Endoglin is a component of the transforming growth factor-beta receptor system in human endothelial cells. *J.Biol.Chem.* **267**, 19027-19030 (1992).

262. Lebrin, F., Deckers, M., Bertolino, P., & Ten, D. P. TGF-beta receptor function in the endothelium. *Cardiovasc.Res.* **65**, 599-608 (2005).

263. Guerrero-Esteo, M., Sanchez-Elsner, T., Letamendia, A., & Bemabeu, C. Extracellular and cytoplasmic domains of endoglin interact with the transforming growth factor-beta receptors I and II. *J.Biol.Chem.* **277**, 29197-29209 (2002).

264. Pera, J. *et al.* Endoglin gene insertion polymorphism not associated with aneurysmal subarachnoid hemorrhage. *J.Neurosurg.* **102**, 879-881 (2005).

265. Peters, D. G., Kassam, A. B., & Chang, Y. F. A DNA sequence polymorphism in the endoglin gene is not associated with intracranial aneurysm or aneurysmal subarachnoid hemorrhage. *Cerebrovasc.Dis.* **20**, 96-100 (2005).

266. Hasham, S. N. *et al.* Mapping a locus for familial thoracic aortic aneurysms and dissections (TAAD2) to 3p24-25. *Circulation* **107**, 3184-3190 (2003).

267. H.H.Stassen, A.Schneeberger, & C.Scharfetter. **CHLC Physical Map (Tri- and Tetranucleotides)**. 2005.
Ref Type: Internet Communication

268. The GDB Human Genome Database. 2005.
Ref Type: Internet Communication

269. Kahari, V. M. *et al.* Transforming growth factor-beta up-regulates elastin gene expression in human skin fibroblasts. Evidence for post-transcriptional modulation. *Lab Invest* **66**, 580-588 (1992).

270. Zhu, C. *et al.* [Overexpression of TGFbeta1 increases ECM expression and adhesion of endothelial cells on matrix]. *Sheng Wu Yi.Xue.Gong.Cheng Xue.Za Zhi.* **22**, 271-274 (2005).

271. Chen, H. *et al.* Abnormal mouse lung alveolarization caused by Smad3 deficiency is a developmental antecedent of centrilobular emphysema. *Am.J.Physiol Lung Cell Mol.Physiol* **288**, L683-L691 (2005).

272. Neptune, E. R. *et al.* Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. *Nat.Genet.* **33**, 407-411 (2003).

273. Southern, E. M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol.Biol.* **98**, 503-517 (1975).

274. Gold, B., Radu, D., Balanko, A., & Chiang, C. S. Diagnosis of Fragile X syndrome by Southern blot hybridization using a chemiluminescent probe: a laboratory protocol. *Mol.Diagn.* **5**, 169-178 (2000).

275. Nogueira, M. L. *et al.* Diagnosis of mucocutaneous herpetic infections by PCR without DNA extraction. *Mem.Inst.Oswaldo Cruz* **93**, 213-214 (1998).

276. Ansorge, W., Sproat, B. S., Stegemann, J., & Schwager, C. A non-radioactive automated method for DNA sequence determination. *J.Biochem.Biophys.Methods* **13**, 315-323 (1986).

277. Newton, C. R. *et al.* Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17**, 2503-2516 (1989).

278. Graham R.Taylor. *Laboratory methods for the detection of mutations and polymorphisms in DNA*(2004).

279. Gross, E., Arnold, N., Goette, J., Schwarz-Boeger, U., & Kiechle, M. A comparison of BRCA1 mutation analysis by direct sequencing, SSCP and DHPLC. *Hum.Genet.* **105**, 72-78 (1999).

280. Orita, M., Suzuki, Y., Sekiya, T., & Hayashi, K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874-879 (1989).

281. Thilly, W. G. *et al.* Direct measurement of mutational spectra in humans. *Genome* **31**, 590-593 (1989).

282. Fischer, S. G. & Lerman, L. S. DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proc.Natl.Acad.Sci.U.S.A* **80**, 1579-1583 (1983).

283. Myers, R. M., Lumelsky, N., Lerman, L. S., & Maniatis, T. Detection of single base substitutions in total genomic DNA. *Nature* **313**, 495-498 (1985).

284. Myers, R. M., Fischer, S. G., Lerman, L. S., & Maniatis, T. Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Res.* **13**, 3131-3145 (1985).

285. Sheffield, V. C., Cox, D. R., Lerman, L. S., & Myers, R. M. Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc.Natl.Acad.Sci.U.S.A* **86**, 232-236 (1989).

286. Cotton, R. G., Rodrigues, N. R., & Campbell, R. D. Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc.Natl.Acad.Sci.U.S.A* **85**, 4397-4401 (1988).

287. Bhattacharyya, A. & Lilley, D. M. The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles). *Nucleic Acids Res.* **17**, 6821-6840 (1989).

288. Mizuuchi, K., Kemper, B., Hays, J., & Weisberg, R. A. T4 endonuclease VII cleaves holliday structures. *Cell* **29**, 357-365 (1982).

289. Solaro, P. C., Birkenkamp, K., Pfeiffer, P., & Kemper, B. Endonuclease VII of phage T4 triggers mismatch correction in vitro. *J Mol.Biol.* **230**, 868-877 (1993).

290. Sarkar, G., Yoon, H. S., & Sommer, S. S. Dideoxy fingerprinting (ddE): a rapid and efficient screen for the presence of mutations. *Genomics* **13**, 441-443 (1992).

291. Sarkar, G., Yoon, H. S., & Sommer, S. S. Dideoxy fingerprinting (ddE): a rapid and efficient screen for the presence of mutations. *Genomics* **13**, 441-443 (1992).

292. Day, I. N. & Humphries, S. E. Electrophoresis for genotyping: microtiter array diagonal gel electrophoresis on horizontal polyacrylamide gels, hydrolink, or agarose. *Anal.Biochem.* **222**, 389-395 (1994).

293. Song, S. Application of fluorescent dUTP on PCR and automated PCR product detection by using internal lane size standard. *Nucleic Acids Symp.Ser.* 211-213 (1995).

294. Giordano, M. *et al.* Identification by denaturing high-performance liquid chromatography of numerous polymorphisms in a candidate region for multiple sclerosis susceptibility. *Genomics* **56**, 247-253 (1999).

295. Xiao, W. & Oefner, P. J. Denaturing high-performance liquid chromatography: A review. *Hum.Mutat.* **17**, 439-474 (2001).

296. Bennett, C. D. *et al.* The LightTyper: high-throughput genotyping using fluorescent melting curve analysis. *Biotechniques* **34**, 1288-5 (2003).

297. **LightTyper System -- Introduction.** 2005.
Ref Type: Internet Communication

298. Stephens, M., Smith, N. J., & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am.J Hum.Genet.* **68**, 978-989 (2001).

299. De Braekeleer, M., Perusse, L., Cantin, L., Bouchard, J. M., & Mathieu, J. A study of inbreeding and kinship in intracranial aneurysms in the Saguenay Lac-Saint-Jean region (Quebec, Canada). *Ann.Hum.Genet.* **60 ( Pt 2)**, 99-104 (1996).

300. Zhang, B. *et al.* Polymorphisms in matrix metalloproteinase-1, -3, -9, and -12 genes in relation to subarachnoid hemorrhage. *Stroke* **32**, 2198-2202 (2001).

301. Miller, S. A., Dykes, D. D., & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).

302. Ebrahim, S., Montaner, D., & Lawlor, D. A. Clustering of risk factors and social class in childhood and adulthood in British women's heart and health study: cross sectional analysis. *BMJ* **328**, 861 (2004).

303. Luca Cartegni, Jinhua Wang, Zhengwei Zhu, Michael Q.Zhang, & Adrian R.Krainer. ESEfinder. [2]. 2004.
Ref Type: Computer Program

304. Vasily Ramensky, Peer Bork, & Shamil Sunyaev. PolyPhen prediction of functional effect of human non-synonymous SNPs. 2004.
Ref Type: Computer Program

305. Kanne, J. P. & Talner, L. B. Autosomal dominant polycystic kidney disease presenting as subarachnoid hemorrhage. *Emerg.Radiol.* **11**, 110-112 (2004).

306. Kim, D. H., Van, G. G., & Milewicz, D. M. Incidence of familial intracranial aneurysms in 200 patients: comparison among Caucasian, African-American, and Hispanic populations. *Neurosurgery* **53**, 302-308 (2003).

307. Ruigrok, Y. M., Rinkel, G. J., & Wijmenga, C. Genetics of intracranial aneurysms. *Lancet Neurol.* **4**, 179-189 (2005).

308. Wills, S. *et al.* Familial intracranial aneurysms: an analysis of 346 multiplex Finnish families. *Stroke* **34**, 1370-1374 (2003).

309. Lilova, M. I. & Petkov, D. L. Intracranial aneurysms in a child with autosomal recessive polycystic kidney disease. *Pediatr.Nephrol.* **16**, 1030-1032 (2001).

310. Neumann, H. P., Krumme, B., van, V., V, Orszagh, M., & Zerres, K. Multiple intracranial aneurysms in a patient with autosomal recessive polycystic kidney disease. *Nephrol.Dial.Transplant.* **14**, 936-939 (1999).

311. Coady, M. A. *et al.* Familial patterns of thoracic aortic aneurysms. *Arch.Surg.* **134**, 361-367 (1999).

312. Gardella, R. *et al.* Characterization of mutations leading to recessive dystrophic epidermolysis bullosa and Marfan syndrome in a single patient. *Clin.Exp.Dermatol.* **26**, 710-713 (2001).

313. Dijk, J. M., van der, G. Y., Grobbee, D. E., Banga, J. D., & Bots, M. L. Increased arterial stiffness is independently related to cerebrovascular disease and aneurysms of the abdominal aorta: the Second Manifestations of Arterial Disease (SMART) Study. *Stroke* **35**, 1642-1646 (2004).

314. Internet. NNSPLICE 0.9 **Berkeley Drosophila Genome Project**. 2004.
Ref Type: Internet Communication

315. Internet. TFSEARCH. Internet . 2005.
Ref Type: Electronic Citation

316. TRANSFAC 4.0 Alibaba2.1. Internet . 2005.
Ref Type: Electronic Citation

317. Transfac 6.0 public. Internet . 2005.
Ref Type: Electronic Citation

318. Ruigrok, Y. M. *et al.* Association of polymorphisms and haplotypes in the elastin gene in Dutch patients with sporadic aneurysmal subarachnoid hemorrhage. *Stroke* **35**, 2064-2068 (2004).

319. Hofer, A. *et al.* Elastin polymorphism haplotype and intracranial aneurysms are not associated in Central Europe. *Stroke* **34**, 1207-1211 (2003).

320. Yamada, S. *et al.* Absence of linkage of familial intracranial aneurysms to 7q11 in highly aggregated Japanese families. *Stroke* **34**, 892-900 (2003).

321. Krex, D., Konig, I. R., Ziegler, A., Schackert, H. K., & Schackert, G. Extended single nucleotide polymorphism and haplotype analysis of the elastin gene in Caucasians with intracranial aneurysms provides evidence for racially/ethnically based differences. *Cerebrovasc.Dis.* **18**, 104-110 (2004).

322. Yamada, S. *et al.* Genome-wide scan for Japanese familial intracranial aneurysms: linkage to several chromosomal regions. *Circulation* **110**, 3727-3733 (2004).

323. Berthelemy-Okazaki, N. *et al.* Examination of ELN as a candidate gene in the Utah intracranial aneurysm pedigrees. *Stroke* **36**, 1283-1284 (2005).

324. Hanon, O. *et al.* Aging, carotid artery distensibility, and the Ser422Gly elastin gene polymorphism in humans. *Hypertension* **38**, 1185-1189 (2001).

325. Foster, J. A. *et al.* Isolation and characterization of cross-linked peptides from elastin. *J Biol.Chem.* **249**, 6191-6196 (1974).

326. Hofer, A. *et al.* Mutations in the lysyl oxidase gene not associated with intracranial aneurysm in central European families. *Cerebrovasc.Dis.* **18**, 189-193 (2004).

327. Urban, Z. *et al.* Supravalvular aortic stenosis: a splice site mutation within the elastin gene results in reduced expression of two aberrantly spliced transcripts. *Hum.Genet.* **104**, 135-142 (1999).

328. Denschlag, D. *et al.* The size of a microsatellite polymorphism of the haem oxygenase 1 gene is associated with idiopathic recurrent miscarriage. *Mol.Hum.Reprod.* **10**, 211-214 (2004).

329. Internet. Splicing. 2004.
Ref Type: Internet Communication

330. Pannu, H. *et al.* Mutations in transforming growth factor-beta receptor type II cause familial thoracic aortic aneurysms and dissections. *Circulation* **112**, 513-520 (2005).

331. Hedin, U. Extracellular matrix components and integrins in the control of arterial smooth muscle cell structure and function. *J.Atheroscler.Thromb.* **1 Suppl 1**, S39-S46 (1994).

## Appendix A:

## PCR optimization 1 (For *ELN*)

Mix preparation for PCR optimisation using gradient thermocycler, in this table 1%w1 is used.
Oligo concentration is 1ug/1µL

**Table A 1: PCR optimization with 1%W1**

| DNA µL | 10X Buffer µL | 8 mM dNTP µL | 1% w1 µL | $MgCl_2$ 25 mM µL | Oligo 1 µL | Oligo 2 µL | Taq µL | H2O µL |
|---|---|---|---|---|---|---|---|---|
| 1 | 2.5 | 0.625 | 1.25 | 0.5 | 0.1 | 0.1 | 0.2 | 18.73 |
| 1 | 2.5 | 0.625 | 1.25 | 1.0 | 0.1 | 0.1 | 0.2 | 18.23 |
| 1 | 2.5 | 0.625 | 1.25 | 1.5 | 0.1 | 0.1 | 0.2 | 17.73 |
| 1 | 2.5 | 0.625 | 1.25 | 2.0 | 0.1 | 0.1 | 0.2 | 17.23 |
| 1 | 2.5 | 0.625 | 1.25 | 2.5 | 0.1 | 0.1 | 0.2 | 16.73 |
| 1 | 2.5 | 0.625 | 1.25 | 3.0 | 0.1 | 0.1 | 0.2 | 16.23 |

## PCR optimization 2 (For *ELN*)

Mix preparation for PCR optimisation using gradient thermocycler, in this table 1%w1 is **not** used.
Oligo concentration is 1ug/1µL

**Table A 2 : PCR optimization without 1%W1**

| DNA in µL | 10X Buffer µL | 8 mM dNTP µL | $MgCl_2$ 25 mM µL | Oligo 1 µL | Oligo 2 µL | Taq µL | H2O µL |
|---|---|---|---|---|---|---|---|
| 1 | 2.5 | 0.625 | 0.5 | 0.1 | 0.1 | 0.2 | 19.98 |
| 1 | 2.5 | 0.625 | 1.0 | 0.1 | 0.1 | 0.2 | 19.48 |
| 1 | 2.5 | 0.625 | 1.5 | 0.1 | 0.1 | 0.2 | 18.98 |
| 1 | 2.5 | 0.625 | 2.0 | 0.1 | 0.1 | 0.2 | 18.48 |
| 1 | 2.5 | 0.625 | 2.5 | 0.1 | 0.1 | 0.2 | 17.98 |
| 1 | 2.5 | 0.625 | 3.0 | 0.1 | 0.1 | 0.2 | 17.48 |

## Primers used in PCR optimisation of Elastin gene:

**Table A 3: primers used in PCR optimization for the elastin gene.**

| Exon | forward ID | Primer sequence | Reversed ID | Primer sequence |
|---|---|---|---|---|
| 1 | Elastin-Ex34F1 | gtctagtcacctggcccaaa | Elastin-Ex34R1 | ctctctccctccctctttcc |
| 2 | Elastin-Ex33F1 | ggcgtgtcaatgttcctacc | Elastin-Ex33R1 | tgggttttgccattgaaagt |
| 3 | Elastin-Ex32F1 | caggtctgaggatgcatgtg | Elastin-Ex32R1 | cctggcagaagtaccgatga |
| 4 | Elastin-Ex31F1 | cctcgctgtctctcaatgct | Elastin-Ex31R1 | ggttgggggttggataagtag |
| 5 | Elastin-ExK F1 | ctgatcacagcactgcccta | Elastin-ExKR1 | ggcagttggtatcagcatca |
| 6 | Elastin-ExA F1 | gccagagcgtaggagtcttc | Elastin-ExAR1 | gttgagggaaggttctttgc |
| 8 and 7 | Elastin-ExG+H F1 | cccactgttccttacgcaat | Elastin-Ex28R1 | tggagagggctctgttcct |
| 9 | Elastin-Ex (I)F1 NEW | acagaggctgtgggtttgag | Elastin-Ex (I) R1 NEW | agtccctgtttccctccttg |
| 10 | Elastin-ExCF1 | agtcagtcccaagggaggtc | Elastin-ExCR1 | ccagaaggtggttggagagt |
| 11 | Elastin-ExLF1 | agaactggccattccttgg | Elastin-Ex24R1 | caggcttggatgggatctt |
| 12 | Elastin-ExDF1 NEW | tgggacctgaacttgctctc | Elastin-ExDR1OLD | aagtgatctgcccgcctta |
| 13 | Elastin-Ex(J)F1 | cctgtgggggtagatctgt | Elastin-Ex(J)R1 | cacagcacccttgctaggac |
| 14 | Elastin-Ex(F)F1 | ggcagcagtggtgatgtct | Elastin-Ex(F)R1 | gctgctttcaggagggaac |
| 15 | ElastinEx XXF1 | gggtatgtaggggccacttt | ElastinEx XXF1 | ccatcagcctctgcctactc |
| 17 and 16 | Elastin-Ex19,20F1 | gaacaaaggccaagtccatc | Elastin-Ex19R1 | ggagggtccttgggaaacta |
| 18 | Elastin-Ex18F1 | gcattcaggaccaactgtca | Elastin-Ex18R1 | tggtgttagggagaatgcaa |
| 19 | Elastin-Ex17F1 | tgttggcatgaaaggagatg | Elastin-Ex17R1 | ttcccttctctgggcaagta |
| 20 | Elastin-Ex16F1 | aatccatcagcatccctcag | Elastin-Ex16R1 | agagccgagcagacaagaag |
| 21 | Elastin-Ex15F1 | aggagttgggggagaagaag | Elastin-Ex15R1 | agtttgccctgaggttggac |
| 22 | Elastin-Ex14F1 | agaattgaaggtgccaggaa | Elastin-Ex14R1 | aaaatggtgcagtcgacctt |
| 23 | Elastin-Ex13F1 | aaggtcgactgcaccatttt | Elastin-Ex13R1 | gaaagcagttctccgtgagc |
| 24 | Elastin-Ex12F1 | cccctcaggctcattgact | Elastin-Ex12R1 | atccagggtcacacagcaa |
| 25 | Elastin-Ex11F1 | ggtggagttgcaggtgagtt | Elastin-Ex11R1 | accagctctgagatcgttgg |
| 26 | Elastin-Ex10F1 | tctgggactaggctcagctc | Elastin-Ex10R1 | ttacccccagatgcttagga |
| 27 | Elastin-Ex9F1 | gaagcaatagaggccaagga | Elastin-Ex9R1 | tattgtgaccaccccagtcc |
| 28 and 29 | Elastin-Ex6,7,8F1 | ctgtctgcttgccttgtgtc | Elastin-Ex6,7,8R1 | gtcagaagctcctcccacac |
| 30 | Elastin-Ex5R1 | agacctcaggctccacctgt | Elastin-Ex5R1 | tgtctcgcatacacacacca |
| 31 | Elastin-Ex4F1 | ggcgaaggagtgagactctg | Elastin-Ex4R1 | caagatcttcaggggtaggg |
| 32 | Elastin-Ex3F1 | agggatatcagggcctcttc | Elastin-Ex3R1 | ggagtccccactgctagatg |
| 33 | Elastin-Ex2F1 | gtgcaggcagaaagtgatga | Elastin-Ex2R1 | gagatggcacaggagaggag |
| 34 | Elastin Ex1F1* | ttctccaccaagcagtagca | Elastin Ex1R1* | Gggattagagccgaaactga |
| 1st 5` | Forward | ctctttctggcgggaaca | Reverse | gaggggtggaggatggac |
| 2nd 5` | Forward | taccttccaggccattcaac | Reverse | actttcccccatctctttcc |
| 3rd 5` | Forward | tgacccatgcagaatagaacc | Reverse | gaggcattgggcaggtct |
| 4th 5` | Forward | gttccatcccacactccaac | Reverse | atctggagcacatggaggat |
| 5th 5` | Forward | tcc acc aat acc tgc ctt tg | Reverse | att tct gcc ccc agg act |
| 6th 5' | Forward | gggaaaaggaagggtttgtc | Reverse | tcagcgtggaaaggtcaaat |
| 7th 5` | Forward | cgagaagagaggggtccag | Reverse | gttgccgtcactcgctct |
| 8th 5` | Forward | catcagactacacggcatgg | Reverse | gaggaaggcaggggctact |
| 3`UTR | Forward | aaccctttgtaaccccatcc | Reverse | caggtgcttgggtaccaact |
| Intron20F mt | Forward | caa ccc atg tcc ccc gA | | |
| Intron20 wt | Forward | caa ccc atg tcc ccc gG | Reverse | aat cca tca gca tcc ctc ag |
| | | gggaacacctccgacccT | | |
| Exon 20F mt | Forward | | | |
| Exon 20F wt | Forward | gggaacacctccgacccC | Reverse | aatccatcagcatccctcag |
| Intron23 | Forward | ccaccccagaatgtgacag | Reverse | cagcccagagatgggtttg |

## PCR condition for each primer is summarised in the following table:

Table A 4: showes the PCr conditions for each PCR amplicon.

| Exon Name | Mg CONCENTRATION | Annealing temp | Exon Name | Mg CONCENTRATION | Annealing temp |
|---|---|---|---|---|---|
| | | | 18 | 2.0 | 60 |
| 1 | 1.0 | 61 | 19 | 1.5 | 60 |
| 2 | 2.0 | 58 | 20 | 1.5 | 63 |
| 3 | 2.0 | 61 | 21 | 1.5 | 60 |
| 4 | 2.0 | 63 | 22 | 2.0 | 63 |
| 5 | 1.5 | 61 | 23 | 1.5 | 59 |
| 6 | 1.0 | 58 | 24 | 1.0 | 56 |
| 8 and 7 | 1.5 | 63 | 25 | 2.5 | 68 |
| 9 | 2.0 | 63 | 26 | 1.0 | 63 |
| 10 | 2.5 | 65 | 27 | 1.0 | 62 |
| 11 | 1.5 | 60 | 28 and 29 | 1.5 | 63 |
| 12 | 1.0 | 58 | 30 | 2.0 | 59 |
| 13 | 1.5 | 60 | 31 | 1.5 | 63 |
| 14 | 1.5 | 59 | 32 | 1.5 | 58 |
| 15 | 1.5 | 59 | 33 | 1.5 | 63 |
| 17 and 16 | 1.0 | 61 | 34 | 1.0 | 56 |

## DHPLC pre-treatment:

Exon where subjected to PCR reaction on 8 familial samples, pre-treatment of PCR product before DHPLC run was performed table below for details:

**Table A 5: using this program for the formation of PCR Heteroduplex**

| PCR treatment for DHPLC | |
|---|---|
| Temperature | Time in minutes |
| 94.0 | 1.0 |
| 93.0 | 1.0 |
| 92.0 | 1.0 |
| 91.0 | 1.0 |
| 90.0 | 1.0 |
| 89.0 | 1.0 |
| 88.0 | 1.0 |
| 87.0 | 1.0 |
| 86.0 | 1.0 |
| 85.0 | 1.0 |
| 84.0 | 1.0 |
| 83.0 | 1.0 |
| 82.0 | 1.0 |
| 81.0 | 1.0 |
| 80.0 | 1.0 |
| 79.0 | 1.0 |
| 78.0 | 1.0 |
| 77.0 | 1.0 |
| 76.0 | 1.0 |
| 75.0 | 1.0 |
| 74.0 | 1.0 |
| 10.0 | Forever |

# Pre-Sequencing Reaction

**Table A 6: Pre sequencing preparation of PCR product**

| | |
|---|---|
| **PCR product** | **5.0 µL** |
| **Exonuclease** | **0.1 µL** |
| **Shrimp Alkaline Phosphatase (SAP)** | **1.0 µL** |
| **SAP Dilution Buffer** | **0.9 µL** |

## Sequencing Reaction

**Table A 7: Sequencing reaction preparation:**

| | |
|---|---|
| **½ sequencing buffer** | **1 µL** |
| **Big Dye** | **2 µL** |
| **Primer (1:50)** | **5 µL ( =1.75 µL of 10 nmol)** |
| **Treated PCR Product** | **3 µL** |
| **Total volume** | **11 µL** |

## Odyssey Reaction Optimisation

**Table A 8: For 100 ul mix preparation and primers I have used and with the following concentrations:**

| | | |
|---|---|---|
| **10 X BUFFER** | **10 µL** | |
| **8 mM dNTPs** | **2.5 µL** | **200 µM** |
| **10 pmol / µL (1:10) primer F** | **5 µL** | **500 nM** |
| **10 pmol / µL (1:10) primer R** | **1 µL** | **100 nM** |
| **10 pmol / µL (1:10) FITC probe** | **1 µL** | **100 nM** |
| **10 pmol / µL (1:10) DABCYL probe** | **1 µL** | **100 nM** |
| **25 mM MGCl2** | **6 µL** | **1.5 mM** |
| **Tag (5U/µL)** | **0.4 µL** | |
| **Water** | **71.1 µL** | |
| **Forward primer**<br>**Reverse primer** | tacttacggggttggagctg<br>agagccgagcagacaagaag | |

## Heteroduplex Formation and The Endo VII Cleavage Reaction

1- Mix fluorescently labelled probe and amplified PCR sample to form 8 µL sample mix, usually 2.5ul of probe to 5.5 of the amplified sample( these values can be altered)
2- Denaturation as in Table A 9.
3- Add 2ul of EndoVII reaction mix
4- Incubate at 37 C for 20 minutes
5- Add 3µL of stopping solution
6- Denature samples then place it on ice
7- Load sample into gel
8- Electrophoresis for 30-35 minutes tank temperature is 60 degrees Celsius.

## Heteroduplex Formation For The Endo VII Cleavage Reaction

**Table A 9: Heteroduplex formation before using the Endo VII is summarised bellow:**

| PCR treatment for the Endo VII | |
|---|---|
| Temperature | Time in minutes |
| 94.0 | 2.0 |
| 93.0 | 1.0 |
| 92.0 | 1.0 |
| 91.0 | 1.0 |
| 90.0 | 1.0 |
| 89.0 | 1.0 |
| 88.0 | 1.0 |
| 87.0 | 1.0 |
| 86.0 | 1.0 |
| 85.0 | 1.0 |
| 84.0 | 1.0 |
| 83.0 | 1.0 |
| 82.0 | 1.0 |
| 81.0 | 1.0 |
| 80.0 | 1.0 |
| 79.0 | 1.0 |
| 78.0 | 1.0 |
| 77.0 | 1.0 |
| 76.0 | 1.0 |
| 75.0 | 1.0 |
| 74.0 | 1.0 |
| 73.0 | 1.0 |
| 72.0 | 1.0 |
| 71.0 | 1.0 |
| 70.0 | 1.0 |
| 69.0 | 1.0 |
| 68.0 | 1.0 |
| 67.0 | 1.0 |
| 65.0 | 1.0 |
| 60.0 | 30 |
| 37.0 | 15 |

## Primers for *TGFBRII*

**Table A 10: shows primers used in linkage study, for GeneScan: labelled primers are on the 5' site**

| NAME FORWARD | SEQUENCE | REVERSE NAME | SEQUENCE | BP |
|---|---|---|---|---|
| D3S2466 FAM For | GCAGAACTTCAGATAAAAGATGC | D3S2466 Rev | TGGTGGGATTTCACTGAAGT | 353-393 |
| D3S4535 FAM For | TTCCTCAGAATGTATCCCCA | D3S4535 Rev | AGGACGCTGAATGAAATGAG | 161-177 |
| D3S2432 HEX For | GGCAGGCAGGTAGATAGACA | D3S2432 Rev | ACACTAAACAAGCATAGTCAGGC | 118-170 |
| D3S1768 HEX For | GGTTGCTGCCAAAGATTAGA | D3S1768 Rev | CACTGTGATTTGCTGTTGGA | 186-206 |
| D3S3727 FAM For | CTAGTACGGGCCGGGT | D3S3727 Rev | GGTAGGTAGTTCCAGTGTGAAA | 115-135 |

Last STR was optimised manually:
($MgCl_2$=2.5)

93 for 3 minutes
93 for 50 sec
53 for 50 sec
72 for 50 sec
goto step 2 for 30 cycles
72 for 2 hours

Other PCR mixes and PCR conditions were from NCBI: UniSTS:47331; UniSTS:78496; UniSTS:52729; UniSTS:55754.

## PCR optimization (with Betaine)

Table A 11: pcr optimization when using Betaine

| DNA µL | 10X Buffer µL | 8 mM dNTP µL | 1% w1 µL | MgCl2 25 mM µL | Oligo 1 µL | Oligo 2 µL | Taq µL | H2O µL |
|---|---|---|---|---|---|---|---|---|
| 1 | 2.5 | 0.625 | 6.5 | 0.5 | 0.2 | 0.2 | 0.2 | 13.0 |
| 1 | 2.5 | 0.625 | 6.5 | 1.0 | 0.2 | 0.2 | 0.2 | 12.5 |
| 1 | 2.5 | 0.625 | 6.5 | 1.5 | 0.2 | 0.2 | 0.2 | 11.5 |
| 1 | 2.5 | 0.625 | 6.5 | 2.0 | 0.2 | 0.2 | 0.2 | 11.0 |
| 1 | 2.5 | 0.625 | 6.5 | 2.5 | 0.2 | 0.2 | 0.2 | 10.5 |
| 1 | 2.5 | 0.625 | 6.5 | 3.0 | 0.2 | 0.2 | 0.2 | 10.0 |

**Oligo concentration is 100 uM/1 µL**

# Primers used in PCR optimisation of The fibrillin 1 gene:

**Table A 12: primers used in FBN1 optimization, PCR conditions are described, "B" stands for Betaine**

| Exon | forward ID | Primer sequence | Reversed ID | Primer sequence | PCR TEM | MgCl2 | NOTE | bp |
|---|---|---|---|---|---|---|---|---|
| 1 | FBNF1 | gcaagaggcggcgggag | FBNR1 | ttgaaacttgggagacccac | 61(57-65) | 5 | B | 246 |
| 2 | FBNF2 | ttggccatctcttcctcttc | FBNR2 | tgcagaatgacaagttttct | 60(57-62) | 5 | B | 175 |
| 3 | FBNF3 | caaatcgtgttccaaatcca | FBNR3 | caggaaagaggaaagccaaa | 58(57-58) | 1.5 | B | 239 |
| 4 | FBNF4 | cctgtgagctgttgcaatct | FBNR4 | cgaagaaaatccatcagcactt | 55(48-61) | 1.5 | B | 250 |
| 5 | FBNF5 | aaagcgtctcagctctctcc | FBNR5 | agtagccatgcagacccaat | 60(50-63) | 1.5 | B | 215 |
| 6 | FBNF6 | cctgcttttctggattttca | FBNR6 | tggctctccagagcaaataag | 55(52-60) | 5 | B | 300 |
| 7 | FBNF7 | tttttctctctgtcttctg | FBNR7 | cccccaactgcaaagcataa | 55(50-57) | 5 | B | 246 |
| 8 | FBNF8 | gctgtttccagggacatgat | FBNR8 | aaccatgcatgctgtttgtc | 57(51-61) | 5 | B | 267 |
| 9 | FBNF9 | ctcagcgatgtgtgtgtgtg | FBNR9 | agggctgggatgggatatt | 58(53-60) | 1.5 | B | 245 |
| 10 | FBNF10 | tgtgttttgttttgttgtgtttttcta | FBNR10 | aacaatgcaagaaaaataactagatg | 55(48-58) | 1.5 | B | 250 |
| 11 | FBNF11 | cctttgcccaaagagtatcc | FBNR11 | agacccttggtgccaaccta | 55(46-60) | 1.5 | B | 299 |
| 12 | FBNF12 | aaggaacccagaaagtcttagaa | FBNR12 | tatgtcccacattccacgtc | 55(51-63) | 5 | B | 231 |
| 13 | FBNF13 | ggagggaggggggaaataaa | FBNR13 | actgcaatggaaggagagga | 50(48-55) | 1.5 | B | 244 |
| 14 | FBNF14 | tcataagaaaatgtatgttt | FBNR14 | gaacatgatctagggtttta | 51(51-55) | 5 | B | 191 |
| 15 | FBNF15 | ttccccattttcaagggtta | FBNR15 | aaggttagccatgatgttttctt | 55(53-60) | 5 | B | 243 |
| 16 | FBNF16 | gggggttctcatctgtttga | FBNR16 | cagtacgagggcatctccat | 55(52-63) | 5 | B | 243 |
| 17 | FBNF17 | ttggaggaaatgatgtgtgc | FBNR17 | acccacaagaaagcctgatg | 60(52-63) | 5 | B | 211 |
| 18 | FBNF18 | cctcctgtagctcctaaggtca | FBNR18 | cagcaatgaaagaaggaatgc | 60(57-63) | 5 | B | 299 |
| 19 | FBNF19 | caaagtttgggcccttttta | FBNR19 | tggcattccaaaagatagca | 62(60-63) | 5 | B | 227 |
| 20 | FBNF20 | ggcccaagactagattttagca | FBNR20 | tttgcaggaaaagctgacatt | 60-(53-62) | 3 | B | 242 |
| 21 | FBNF21 | ttccaaggtgtatgtttgaattttt | FBNR21 | aaccacagcatgggtttctc | 63(61-63) | 5 | B | 224 |
| 22 | FBNF22 | tgtcagaactgcaaagtctgg | FBNR22 | gctgcatatttctccctgtga | 63(55-67) | 3.5 | | 233 |
| 23 | FBNF23 | acttaccaggttcaaaatg | FBNR23 | gtgtgtctgtacctgaagc | 50(48-53) | 5 | B | 320 |
| 24 | FBNF24 | acctcccttgattccctctg | FBNR24 | attggccatggaaaacgtaa | 55(52-60) | 3 | | 280 |
| 25 | FBNF25 | gggcattgagacctcctgac | FBNR25 | aaacagcaagtggcagcaaa | 60(55-65) | 5 | B | 311 |
| 26 | FBNF26 | tcatttgctgccacttgctg | FBNR26 | tctgtgttgatcaaatgatc | 57(50-60) | 5 | B | 190 |
| 27 | FBNF27 | gtctggtggaggagatgagg | FBNR27 | ccaactttggcaatgatgtc | 55(48-58) | 1.5 | B | 250 |
| 28 | FBNF28 | tggaagcttatgtttgggtgt | FBNR28 | agagtgttttagggagagatgaaa | 60(55-65) | 1.5 | B | 298 |
| 29 | FBNF29 | gggacagacatccaaaccat | FBNR29 | aaagcctgggccctaaacta | 63(51-67) | 1.5 | | 248 |
| 30 | FBNF30 | aacctgtggttgttggtttt | FBNR30 | tgaaaaattctgtcttctttgctt | 60(58-61) | 5 | B | 183 |
| 31 | MSF31 | gtactcaatgatatcaaatagc | MSR31 | accaatctcttaactacttaat | 57(52-57) | 5 | B | 239 |
| 32 | FBNF32 | catttgtgctgagcctttttc | FBNR32 | tgcagtccttgataagcaacc | 50(45-55) | 1.5 | B | 212 |
| 33 | FBNF33 | ggttttaaataccacccttctgtt | FBNR33 | gcctgagaaatgtggaatgc | 55(52-58) | 1.5 | B | 244 |
| 34 | FBNF34 | gagtaacgtgtgtttctttc | FBNR34 | ggctcccagtggcttccccatc | 55(50-60) | 5 | B | 183 |
| 35 | FBNF35 | gtttttgctttttctccc | FBNR35 | gctgattttgatgccagtgg | 51(51-63) | 1.5 | | 188 |
| 36 | FBNF36 | ccactactcactgttcggtt | FBNR36 | ttctctgaaaagtttttaag | 51(51-57) | 3 | W1% | 176 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 37 | FBNF37 | ttttgtgtttgtatatggta | FBNR37 | taaataggaggatgtccact | 50(50-53) | 5 | B | 213 |
| 38 | FBNF38 | gattcaaaacaactcaattt | FBNR38 | gctttaagacaaaggaaacac | 51(51-57) | 1.5 | | 140 |
| 39 | FBNF39 | ggaatgcctttgttttgatttt | FBNR39 | ttctggttttgcaggtcagtt | 55(51-63) | 1.5 | | 217 |
| 40 | FBNF40 | tattcacataccactttctc | FBNR40 | catgcattactgagaaaagc | 50(50-57) | 5 | B | 185 |
| 41 | FBNF41 | tcctttttttacctccctt | FBNR41 | gataatggagaaactaaaact | 53(51-55) | 1.5 | | 207 |
| 42 | FBNF42 | ctcccatcccacctttgtt | FBNR42 | gaaagttctgacaatgccgt | 55(51-63) | 1.5 | | 133 |
| 43 | FBNF43 | tgtcactcatgaatgactac | FBNR43 | tggatatgataaagtcatga | 55(53-55) | 1.5 | | 187 |
| 44 | FBNF44 | tccttcaaattcagttctct | FBNR44 | aggcatgtccagcctgtggg | 52(51-55) | 1.5 | | 185 |
| 45 | FBNF45 | agttctcacttaagatgctt | FBNR45 | aataataataattgcatact | 53(53-55) | 5 | | 180 |
| 46 | FBNF46 | tatgtttctttatggccttt | FBNR46 | ctttgctgatgcacaatttt | 55(51-57) | 1.5 | | 178 |
| 47 | FBNF47 | tattaaaggattgttgggga | FBNR47 | tttccaggtctttctaagtc | 53(51-57) | 1.5 | | 190 |
| 48 | FBNF48 | cctcttccttatttttccct | FBNR48 | ctcatttgctacaactgata | 51(51-57) | 1.5 | | 181 |
| 49 | FBNF49 | tgactttgtttgactcatgt | FBNR49 | tgaaagcccaaagccttcaa | 55(51-61) | 1.5 | | 187 |
| 50 | FBNF50 | tttgctatggtgcaatacgg | FBNR50 | gcccagagagaaatgcagat | 51(51-65) | 1.5 | | 247 |
| 51 | FBNF51 | ttctatctattaatgagtgt | FBNR51 | gaaatgctgagaatccagcac | 53(51-55) | 1.5 | | 129 |
| 52 | FBNF52 | cttgttattcactatttttt | FBNR52 | atggaagaaaacttattact | 51 | 5 | B | 188 |
| 53 | FBNF53 | atgttttggacacattcctg | FBNR53 | actgttctctgtttaagaga | 53(51-55) | 5 | B | 178 |
| 54 | FBNF54 | tcccttatttacttactctc | FBNR54 | ggcttagatgaccttgaacac | 53(50-57) | 5 | B | 185 |
| 55 | FBNF55 | tttgtgattgtacatttttt | FBNR55 | tcaaagctccttccacaggg | 55 | 5 | B | 186 |
| 56 | FBNF56 | tggtccttcaataaaatcaa | FBNR56 | gaaagtgcggtgccaactgta | 55(50-57) | 5 | B | 176 |
| 57 | FBNF57 | ctgacatccccttttgccata | FBNR57 | tccctgcaagtatttttggac | 55(46-58) | 1.5 | B | 277 |
| 58 | FBNF58 | cactgaagtgacccccctaca | FBNR58 | tccacttgaggataagccatc | 62(60-63) | 5 | B | 259 |
| 59 | FBNF59 | agaccctgtggaaattgagc | FBNR59 | cagccatgtgtcaggagcta | 55(47-61) | 1.5 | B | 246 |
| 60 | FBNF60 | gactcaaatgcctctcttgca | FBNR60 | tcgctacaatccatgtagga | 51(51-60) | 1.5 | | 189 |
| 61 | FBNF61 | aattttaacccctctttgcc | FBNR61 | gatcgcagctgaagtctccac | 52(51-55) | 5 | B | 191 |
| 62 | FBNF62 | ggcatcatggtggctctgcttc | FBNR62 | tgagagtgaggaaaagttac | 55(54-60) | 5 | B | 174 |
| 63 | FBNF63 | gctgccacacatgccgcttc | FBNR63 | tccaaccatgaccaggaaga | 60(57-63) | 5 | B | 298 |
| 64 | FBNF64 | catctatgctccccttctgc | FBNR64 | ttccaccacaggagacatca | 55(46-58) | 1.5 | B | 243 |
| 65A | FBNF65A | catattgccatgtgtctttcc | FBNR65A | aatgaataggttccagccact | 51(51-61) | 1.5 | | 283 |
| 65B | FBNF65B | ccagtggctggaacctattc | FBNR65B | tgattctgattgggggaaaa | 51(51-57) | 1.5 | | 232 |

# Appendix B: Searching Transcription Factors

## B1 :Waseem G>C position -2253:

### TFSEARCH Search Result

```
** TFSEARCH  ver.1.3 **        (c)1995 Yutaka Akiyama (Kyoto Univ.)

 This simple routine searches highly correlated sequence fragments
 versus TFMATRIX transcription factor binding site profile database
 by E.Wingender, R.Knueppel, P.Dietze, H.Karas (GBF-Braunschweig).

  <Warning> Scoring scheme is so straightforward in this version.
            score = 100.0 * ('weighted sum' - min) / (max - min)
            The score does not properly reflect statistical significance!

 Database:  TRANSFAC MATRIX TABLE, Rel.3.3 06-01-1998
 Query:     untitled (50 bases)
 Taxonomy:  Vertebrate
 Threshold: 85.0 point


 TFMATRIX entries with High-scoring:

   1 AAGCTGGTTC CTGCCCGTGT CACTGCCTCG AGAAGAGAGG GGTCCAGCTC entry
score


 Total 0 high-scoring sites found.

** No TFMATRIX entry hit for your sequence. **
```

## B2: Waseem C>G -1162

### TFSEARCH Search Result

```
** TFSEARCH  ver.1.3 **        (c)1995 Yutaka Akiyama (Kyoto Univ.)

 This simple routine searches highly correlated sequence fragments
 versus TFMATRIX transcription factor binding site profile database
 by E.Wingender, R.Knueppel, P.Dietze, H.Karas (GBF-Braunschweig).

  <Warning> Scoring scheme is so straightforward in this version.
            score = 100.0 * ('weighted sum' - min) / (max - min)
            The score does not properly reflect statistical significance!

 Database:  TRANSFAC MATRIX TABLE, Rel.3.3 06-01-1998
 Query:     untitled (50 bases)
 Taxonomy:  Vertebrate
 Threshold: 85.0 point


 TFMATRIX entries with High-scoring:

   1 CACCAGCGGA ATGTCAGCCT TCCCAGAGGG GCCGGGAGAA CAGCAGTCGA entry
score
                  <-------------                          M00087 Ik-2
90.4
                    <----------                           M00141 Lyf-1
88.3
                              ---------->                 M00008 Sp1
87.7


 Total 3 high-scoring sites found.
 Max score: 90.4 point, Min score: 87.7 point
```

## B3: Waseem C>T -1050

### TFSEARCH Search Result

```
** TFSEARCH  ver.1.3 **        (c)1995 Yutaka Akiyama (Kyoto Univ.)

 This simple routine searches highly correlated sequence fragments
 versus TFMATRIX transcription factor binding site profile database
 by E.Wingender, R.Knueppel, P.Dietze, H.Karas (GBF-Braunschweig).

  <Warning> Scoring scheme is so straightforward in this version.
            score = 100.0 * ('weighted sum' - min) / (max - min)
            The score does not properly reflect statistical significance!

 Database:  TRANSFAC MATRIX TABLE, Rel.3.3 06-01-1998
 Query:     untitled (50 bases)
 Taxonomy:  Vertebrate
 Threshold: 85.0 point


 TFMATRIX entries with High-scoring:

   1 GTGATAATGG GAAGCTGGGC TGCCTGTCAG TCTGCGGGGG GCTCCCACCT entry
score
     -------->                                                 M00077 GATA-3
88.4
     ------------->                                            M00137 Oct-1
86.5
       ------>                                                 M00101 CdxA
85.7
```

## B4: Waseem G>A rs3757583

### TFSEARCH Search Result

```
** TFSEARCH  ver.1.3 **        (c)1995 Yutaka Akiyama (Kyoto Univ.)

 This simple routine searches highly correlated sequence fragments
 versus TFMATRIX transcription factor binding site profile database
 by E.Wingender, R.Knueppel, P.Dietze, H.Karas (GBF-Braunschweig).

  <Warning> Scoring scheme is so straightforward in this version.
            score = 100.0 * ('weighted sum' - min) / (max - min)
            The score does not properly reflect statistical significance!

 Database:  TRANSFAC MATRIX TABLE, Rel.3.3 06-01-1998
 Query:     untitled (50 bases)
 Taxonomy:  Vertebrate
 Threshold: 85.0 point

 TFMATRIX entries with High-scoring:

   1 TCCTGGGGTG GCCCCGTATA GACCAAAGCC TGATAGCTGT CCTAGAAGCA entry
score
                                   ---------->                 M00075 GATA-1
95.5
                                   ---------->                 M00076 GATA-2
94.1
        ----->                                                 M00271 AML-1a
87.4
                                 -------------->               M00127 GATA-1
85.5


 Total 4 high-scoring sites found.
 Max score: 95.5 point, Min score: 85.5 point
```

# Appendix C:

## SAH ONLY

SAH ONLY

### Intron 20 A(1)G(2)

| | Observed | | | | |
|---|---|---|---|---|---|
| | 12 | 12 | 12 | | |
| Genotypes | **11** | **12** | **22** | Total | (Dropouts) |
| Frequency | 86 | 47 | 4 | 137 | 0 |
| %Frequency | 62.77 | 34.31 | 2.92 | 100.00 | |
| | p | q | p+q | | |
| | 0.80 | 0.20 | 1 | | |

| | Expected | | | |
|---|---|---|---|---|
| | 12 | 12 | 12 | |
| Genotypes | **11** | **12** | **22** | Total |
| Frequency | 87.5 | 44.0 | 5.5 | 137 |
| % Frequency | 63.9 | 32.1 | 4.0 | 100.0 |

| | Chi^2 test and level of significance | | | | |
|---|---|---|---|---|---|
| | Obs. | Exp. | (o-e)^2/e | Chi^2 | Sig level |
| 11 | 86 | 87.5 | 0.0 | | |
| 12 | 47 | 44.0 | 0.2 | | |
| 22 | 4 | 5.5 | 0.4 | | |
| | | | | 0.7 | 4.18E-01 |
| | | | | Not sig at 1 df | |

SAH ONLY

## EXON 20 C(1)/T(2)

| | Observed | | | | |
|---|---|---|---|---|---|
| | 12 | 12 | 12 | | |
| **Genotypes** | **11** | **12** | **22** | **Total** | (Dropouts) |
| **Frequency** | 48 | 66 | 24 | 138 | 0 |
| **%Frequency** | 34.78 | 47.83 | 17.39 | 100.00 | |
| | p | q | p+q | | |
| | 0.59 | 0.41 | 1 | | |

| | Expected | | | |
|---|---|---|---|---|
| | 12 | 12 | 12 | |
| **Genotypes** | **11** | **12** | **22** | **Total** |
| **Frequency** | 47.5 | 66.9 | 23.5 | 138 |
| **% Frequency** | 34.5 | 48.5 | 17.1 | 100.0 |

| | Chi^2 test and level of significance | | | | |
|---|---|---|---|---|---|
| | **Obs.** | **Exp.** | (o-e)^2/e | **Chi^2** | **Sig level** |
| **11** | 48 | 47.5 | 0.0 | | |
| **12** | 66 | 66.9 | 0.0 | | |
| **22** | 24 | 23.5 | 0.0 | | |
| | | | | 0.0 | 8.73E-01 |
| | | | | **Not sig at 1 df** | |

SAH ONLY

## Intron 23 A=1/G=2

| | Observed | | | | |
|---|---|---|---|---|---|
| | 12 | 12 | 12 | | |
| **Genotypes** | **11** | **12** | **22** | **Total** | (Dropouts) |
| **Frequency** | 40 | 66 | 24 | 130 | 0 |
| **%Frequency** | 30.77 | 50.77 | 18.46 | 100.00 | |
| | p | q | p+q | | |
| | 0.56 | 0.44 | 1 | | |

| | Expected | | | |
|---|---|---|---|---|
| | 12 | 12 | 12 | |
| **Genotypes** | **11** | **12** | **22** | **Total** |
| **Frequency** | 41.0 | 64.0 | 25.0 | 130 |
| **% Frequency** | 31.5 | 49.2 | 19.2 | 100.0 |

| | Chi^2 test and level of significance | | | | |
|---|---|---|---|---|---|
| | Obs. | Exp. | (o-e)^2/e | Chi^2 | Sig level |
| **11** | 40 | 41.0 | 0.0 | | |
| **12** | 66 | 64.0 | 0.1 | | |
| **22** | 24 | 25.0 | 0.0 | | |
| | | | | 0.1 | 7.24E-01 |
| | | | | **Not sig at 1 df** | |

# HI ONLY

HI ONLY

## Intron 20 A(1)G(2)

| | Observed | | | | |
|---|---|---|---|---|---|
| | 12 | 12 | 12 | | |
| **Genotypes** | **11** | **12** | **22** | **Total** | (Dropouts) |
| **Frequency** | 45 | 30 | 2 | 77 | 0 |
| **%Frequency** | 58.44 | 38.96 | 2.60 | 100.00 | |
| | **p** | **q** | **p+q** | | |
| | 0.78 | 0.22 | 1 | | |

| | Expected | | | |
|---|---|---|---|---|
| | 12 | 12 | 12 | |
| **Genotypes** | **11** | **12** | **22** | **Total** |
| **Frequency** | 46.8 | 26.5 | 3.8 | 77 |
| **% Frequency** | 60.7 | 34.4 | 4.9 | 100.0 |

| | Chi^2 test and level of significance | | | | |
|---|---|---|---|---|---|
| | Obs. | Exp. | (o-e)^2/e | Chi^2 | Sig level |
| **11** | 45 | 46.8 | 0.1 | | |
| **12** | 30 | 26.5 | 0.5 | | |
| **22** | 2 | 3.8 | 0.8 | | |
| | | | | 1.3 | 2.45E-01 |
| | | | | **Not sig at 1 df** | |

HI ONLY

## EXON 20 C(1)/T(2)

| | Observed | | | | |
|---|---|---|---|---|---|
| | 12 | 12 | 12 | | |
| Genotypes | **11** | **12** | **22** | Total | (Dropouts) |
| Frequency | 29 | 38 | 10 | 77 | 0 |
| %Frequency | 37.66 | 49.35 | 12.99 | 100.00 | |
| | p | q | p+q | | |
| | 0.62 | 0.38 | 1 | | |

| | Expected | | | |
|---|---|---|---|---|
| | 12 | 12 | 12 | |
| Genotypes | **11** | **12** | **22** | Total |
| Frequency | 29.9 | 36.2 | 10.9 | 77 |
| % Frequency | 38.9 | 47.0 | 14.2 | 100.0 |

| | Chi^2 test and level of significance | | | | |
|---|---|---|---|---|---|
| | Obs. | Exp. | (o-e)^2/e | Chi^2 | Sig level |
| **11** | 29 | 29.9 | 0.0 | | |
| **12** | 38 | 36.2 | 0.1 | | |
| **22** | 10 | 10.9 | 0.1 | | |
| | | | | 0.2 | 6.54E-01 |
| | | | | Not sig at 1 df | |

HI
ONLY

## ıntron 23 A=1/G=2

| | Observed | | | | |
|---|---|---|---|---|---|
| | 12 | 12 | 12 | | |
| Genotypes | **11** | **12** | **22** | Total | (Dropouts) |
| Frequency | 18 | 43 | 16 | 77 | 0 |
| %Frequency | 23.38 | 55.84 | 20.78 | 100.00 | |
| | p | q | p+q | | |
| | 0.51 | 0.49 | 1 | | |

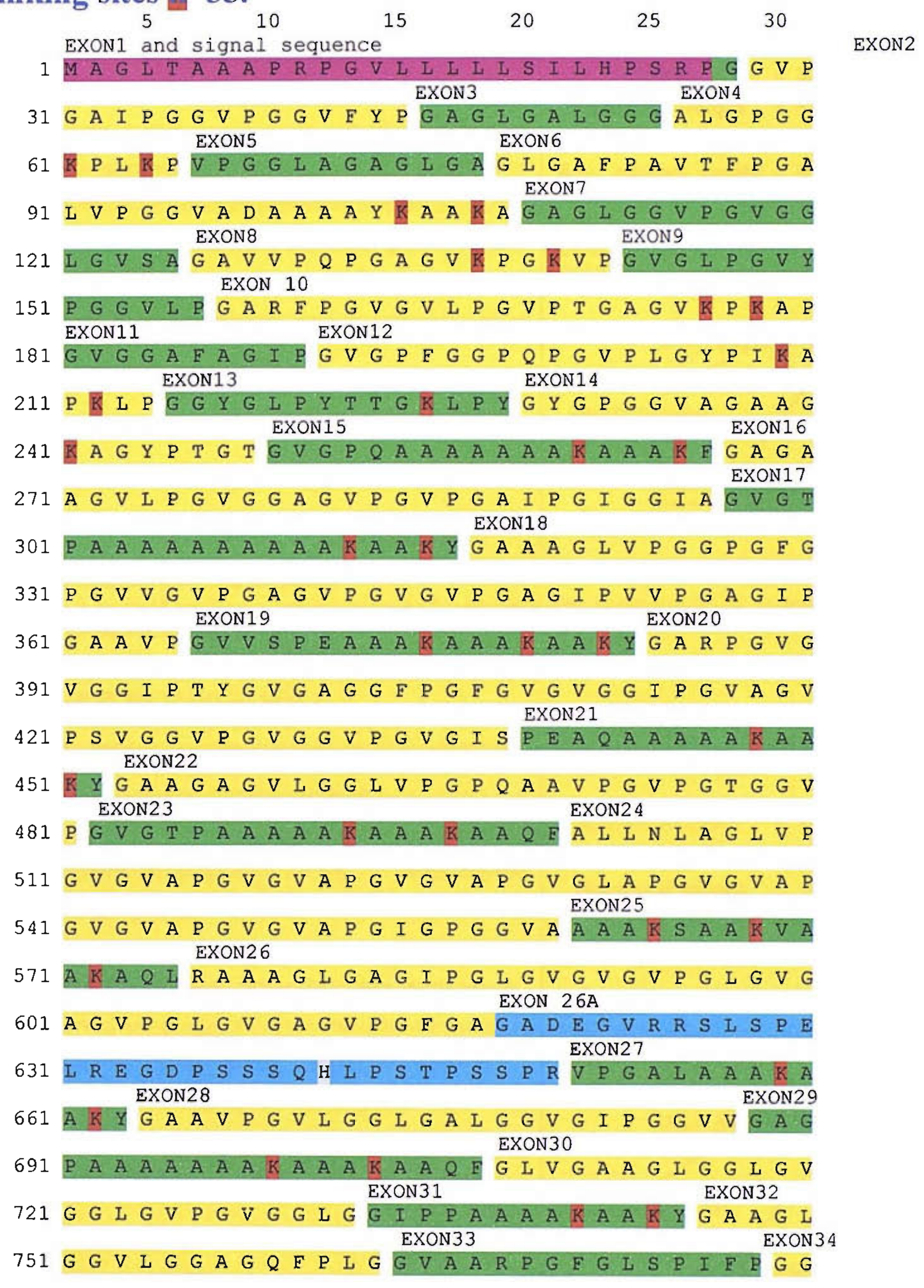
| | Expected | | | |
|---|---|---|---|---|
| | 12 | 12 | 12 | |
| Genotypes | **11** | **12** | **22** | Total |
| Frequency | 20.3 | 38.5 | 18.3 | 77 |
| % Frequency | 26.3 | 50.0 | 23.7 | 100.0 |

| | Chi^2 test and level of significance | | | | |
|---|---|---|---|---|---|
| | Obs. | Exp. | (o-e)^2/e | Chi^2 | Sig level |
| **11** | 18 | 20.3 | 0.3 | | |
| **12** | 43 | 38.5 | 0.5 | | |
| **22** | 16 | 18.3 | 0.3 | | |

| | | |
|---|---|---|
| | 1.1 | 3.02E-01 |
| | **Not sig at 1 df** | |

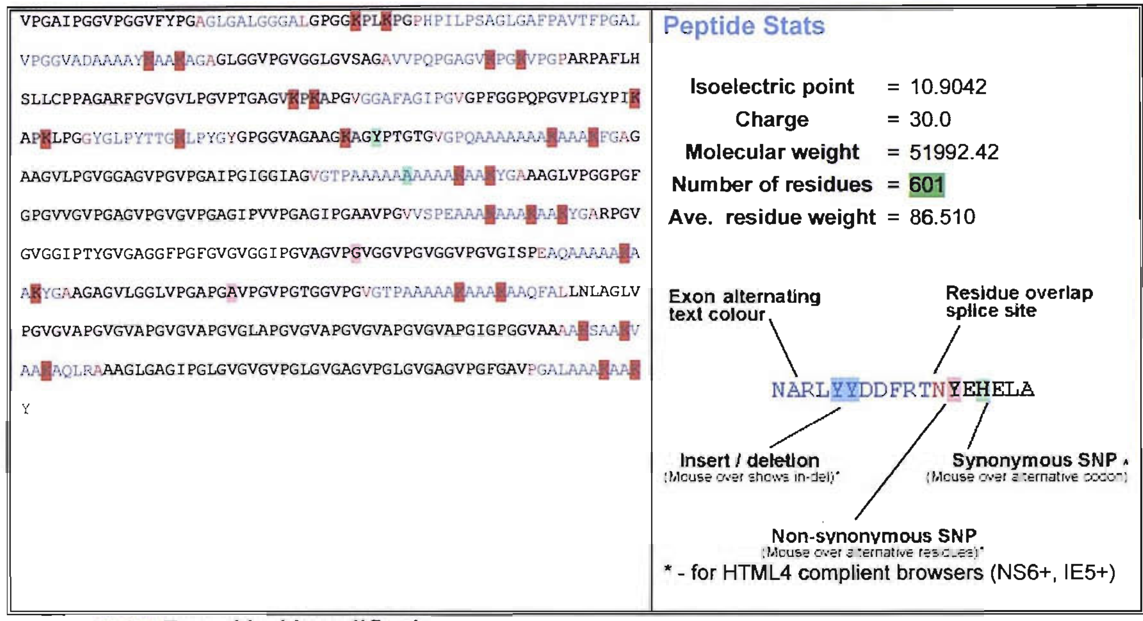
## Appendix D:

### D.1: Human elastin protein sequence with exon position and cross linking sites K=35.



## D2 : Some *ELN* transcripts

**DIFFERENT PEPTIDES ENCODED BY DIFFERENT TRANSCRIPT OF ELASTIN GENE:**

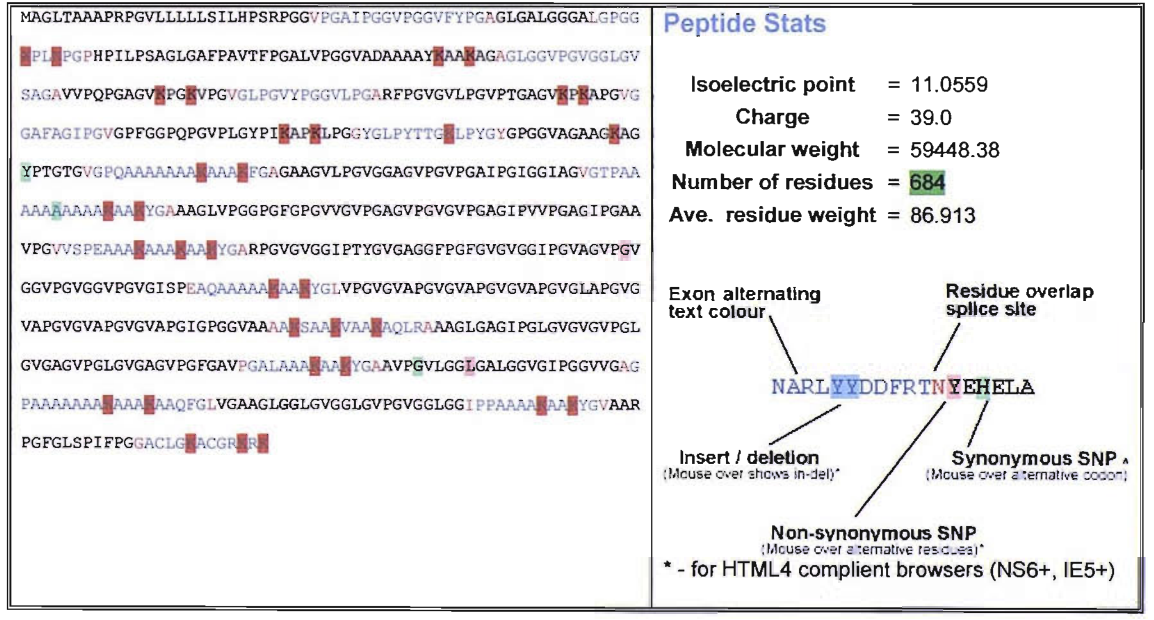



**FROM** Ensembl with modification

**http://www.ensembl.org/Homo_sapiens/geneview?gene=ENSG00000049540#ENST00000252034**

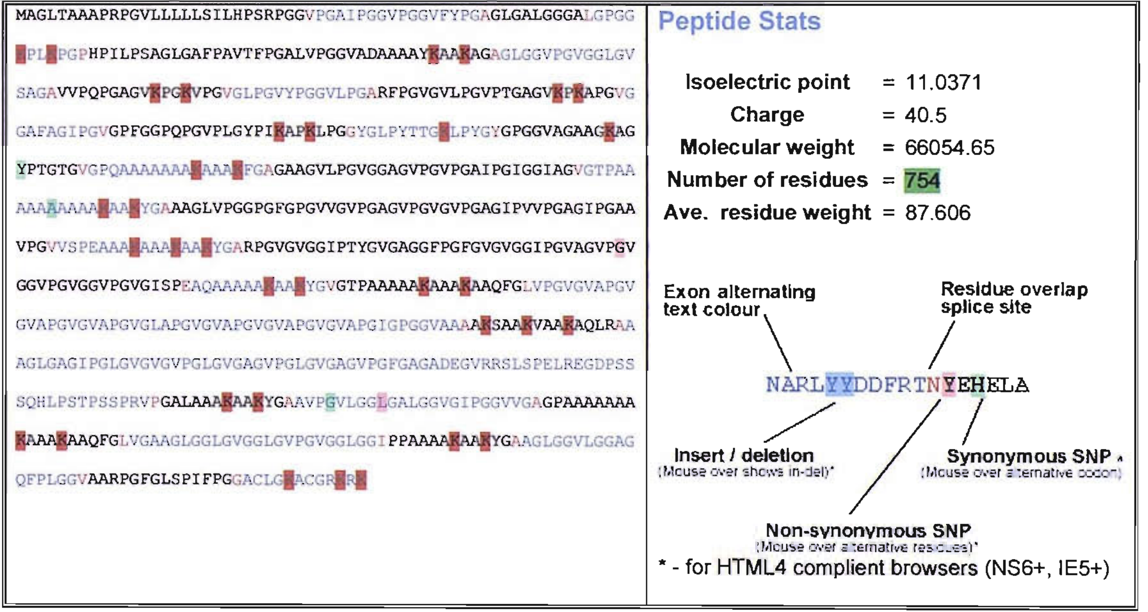
**O15337** (SPTREMBL ID) NUMBER OF K= 28

**Q14235** (SPTREMBL ID) NUMBER OF K= 33

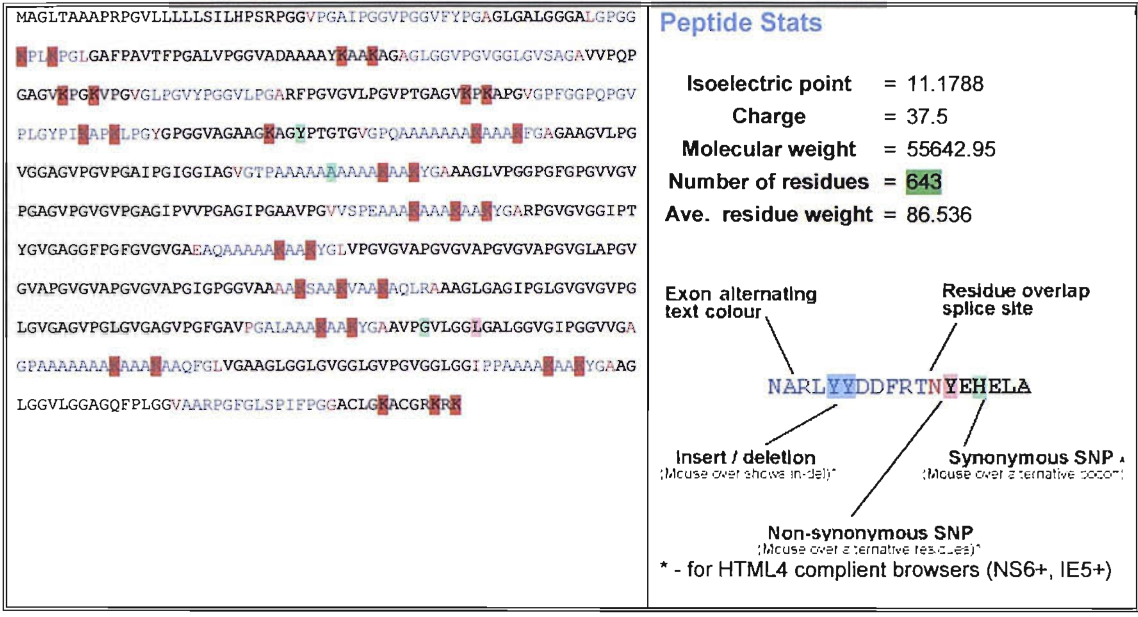


**Q8N2G0** (SPTREMBL ID) NUMBER OF K= 14

***ELN*** (HUGO ID) NUMBER OF K= 35




**Q8NBI4** (SPTREMBL ID) NUMBER OF K= 32



MAGLTAAAPRPGVLLLLLSILHPSRPGGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGG
KPLKPGLGAFPAVTFPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQP
GAGVKPGKVPGVGLPGVYPGGVLPGARFPGVGVLPGVPTGAGVKPKAPGVGPFGGPQPGV
PLGYPIKAPKLPGYGPGGVAGAAGKAGYPTGTGVGPQAAAAAAAKAAAKFGAGAAGVLPG
VGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVVGV
PGAGVPGVGVPGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGARPGVGVGGIPT
YGVGAGGFPGFGVGVGAEAQAAAAAKAAKYGLVPGVGVAPGVGVAPGVGVAPGVGLAPGV
GVAPGVGVAPGVGVAPGIGPGGVAAAAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGVPG
LGVGAGVPGLGVGAGVPGFGAVPGALAAAKAAKYGAAVPGVLGGLGALGGVGIPGGVVGA
GPAAAAAAAKAAAKAAQFGLVGAAGLGGLGVGGLGVPGVGGLGGIPPAAAAKAAKYGAAG
LGGVLGGAGQFPLGGVAARPGFGLSPIFPGGACLGKACGRKRK
Peptide Stats
Isoelectric point = 11.1788
Charge = 37.5
Molecular weight = 55642.95
Number of residues = 643
Ave. residue weight = 86.536
Exon alternating text colour
Residue overlap splice site
NARLYYDDFRTNYEHELA
Insert / deletion
Synonymous SNP
Non-synonymous SNP
* - for HTML4 complient browsers (NS6+, IE5+)

# D3: Aliphatic a.a.

| | Amino Acid | $pK_a$'s[2] | Pro Structure[3] | Chemical Structure[4] |
|---|---|---|---|---|
| A<br>l<br>i | **Glycine, Gly, G**<br>smallest amino acid<br>no charge<br>not hydrophilic (0.67)[1]<br><br>Molec. Wt. = 57<br>Mole % = 7.5 | N=9.60<br><br>C=2.35<br><br>pI=5.97 | a =0.43<br><br>ß =0.58<br><br>t =1.77 | H<br>$H_3N^+$—C—$CO_2^-$<br>H |
| p<br>h<br>a | **Alanine, Ala, A**<br>like glycine<br>no charge<br>hydrophobic (1.0)<br><br>Molec. Wt. = 71<br>Mole % = 9.0 | N=9.69<br><br>C=2.34<br><br>pI=6.01 | a =1.41<br><br>ß =0.72<br><br>t =0.82 | $CH_3$<br>$H_3N^+$—C—$CO_2^-$<br>H |
| t<br>i<br>c | **Valine, Val, V**<br>no charge<br>hydrophobic (2.3)<br><br>Molec. Wt. = 99<br>Mole % = 6.9 | N=9.62<br><br>C=2.32<br><br>pI=5.97 | a =0.90<br><br>ß =1.87<br><br>t =0.41 | $CH_3$ $CH_3$<br>CH<br>$H_3N^+$—C—$CO_2^-$<br>H |
| | **Leucine, Leu, L**<br>no charge<br>isomer of isoleucine<br>hydrophobic (2.2)<br><br>Molec. Wt. = 113<br>Mole % = 7.5 | N=9.60<br><br>C=2.36<br><br>pI=5.98 | a =1.34<br><br>ß =1.22<br><br>t =0.57 | $CH_3$ $CH_3$<br>CH<br>$CH_2$<br>$H_3N^+$—C—$CO_2^-$<br>H |
| | **Isoleucine, Ile, I**<br>no charge<br>isomer of leucine<br>hydrophobic (3.1)<br><br>Molec. Wt. = 113<br>Mole % = 4.6 | N=9.68<br><br>C=2.36<br><br>pI=6.02 | a =1.09<br><br>ß =1.67<br><br>t =0.47 | $CH_3$<br>$CH_2$<br>$CH_3$—CH<br>$H_3N^+$—C—$CO_2^-$<br>H |
| | **Proline, Pro, P**<br>no charge<br>promotes turns<br>not hydrophobic (-0.29)<br><br>Molec. Wt. = 97<br>Mole % = 4.6 | N=10.96<br><br>C=1.99<br><br>pI=6.48 | a =0.34<br><br>ß =0.31<br><br>t =1.32 | $H_2$<br>C<br>$H_2C$ $CH_2$<br>$H_2N^+$—C—$CO_2^-$<br>H |

From: http://www.mcb.ucdavis.edu/courses/bis102/Aliphatic.html#Glycine

## D4: polar / uncharged a.a.

| | Amino Acid | pKa's[2] | Pro Structure[3] | Chemical Structure[4] |
|---|---|---|---|---|
| Polar/Uncharged | **Serine, Ser, S**<br>no charge<br>hydrogen bonding<br>hydrophilic (-1.1)<br><br>Molec. Wt. = 87<br>Mole % = 7.1 | N=9.15<br><br>C=2.21<br><br><br>pI=5.68 | a =0.57<br><br>ß =0.96<br><br>t =1.22 | OH / CH2 / $H_3N^+$—C—$CO_2^-$ / H |
| | **Threonine, Thr, T**<br>no charge<br>hydrogen bonding<br>hydrophilic (-0.75)<br><br>Molec. Wt. =101<br>Mole % = 6.0 | N=9.62<br><br>C=2.11<br><br><br>pI=5.87 | a =0.76<br><br>ß =1.17<br><br>t =0.90 | OH / CH—CH3 / $H_3N^+$—C—$CO_2^-$ / H |
| | **Asparagine, Asn, N**<br>amide of Asp<br>hydrogen bonding<br>no charge<br>hydrophilic (-2.7)<br><br>Molec. Wt. =114<br>Mole % = 4.4 | N=8.08<br><br>C=2.02<br><br><br>pI=5.41 | a =0.76<br><br>ß =0.48<br><br>t =1.34 | O, NH2 / C / CH2 / $H_3N^+$—C—$CO_2^-$ / H |
| | **Glutamine, Gln, Q**<br>amide of Glu<br>hydrogen bonding<br>no charge<br>hydrophilic (-2.9)<br><br>Molec. Wt. = 128<br>Mole % = 3.9 | N=9.13<br><br>C=2.17<br><br><br>pI=5.65 | a =1.27<br><br>ß =0.98<br><br>t =0.84 | O, NH2 / C / CH2 / CH2 / $H_3N^+$—C—$CO_2^-$ / H |

From: http://www.mcb.ucdavis.edu/courses/bis102/Polar.html#Serine

# D5: Aromatic a.a.

| | Amino Acid | pK$_a$'s[2] | Pro Structure[3] | Chemical Structure[4] |
|---|---|---|---|---|
| Aromatic | **Phenylalanine, Phe, F**<br>no charge<br>absorbs UV<br>hydrophobic (2.5)<br><br>Molec. Wt. = 147<br>Mole % = 3.5 | N=9.13<br><br>C=1.83<br><br>pI=5.48 | a =1.16<br><br>ß =1.33<br><br>t =0.59 | $CH_2$<br>$H_3N^+$—C—$CO_2^-$<br>H |
| | **Tyrosine, Tyr, Y**<br>weak charge<br>absorbs UV<br>hydrogen bonding<br>not hydrophilic (0.08)<br><br>Molec. Wt. = 163<br>Mole % = 3.5 | N=9.11<br><br>C=2.20<br><br>R=10.07<br><br>pI=5.66 | a =0.74<br><br>ß =1.45<br><br>t =0.76 | H<br>O<br>$CH_2$<br>$H_3N^+$—C—$CO_2^-$<br>H |
| | **Tryptophan, Trp, W**<br>largest amino acid<br>rarest amino acid<br>no charge<br>absorbs UV<br>hydrogen bonding<br>hydrophobic (1.5)<br><br>Molec. Wt. = 186<br>Mole % = 1.1 | N=9.39<br><br>C=2.38<br><br>pI=5.89 | a =1.02<br><br>ß =1.35<br><br>t =0.65 | H<br>N<br>$CH_2$<br>$H_3N^+$—C—$CO_2^-$<br>H |

From: http://www.mcb.ucdavis.edu/courses/bis102/Aromatic.html#Tryptophan

# D6: FIBRILLIN-1 PROTEIN WITH ITS DOMAINS

FROM:
http://us.expasy.org/cgi-bin/niceprot.pl?P35555

```
        1          11         21         31          41          51
                                                          <<<<EXON1
   1 MRRGRLLEIA LGFTVLLASY TSHGADANLE AGNVKETRAS RAKRRGGGGH DALKG|PNVCG    60
                       <<<<EXON2                          <<<<EXON3
                                          DOMAIN   81   112   EGF-like 1.
  61 SRYNAYCCPG WKTLPGGNQC IV|PICRHSCG DGFCSRPNMC TCPSGQIAPS CGSRSI|QHCN   120
                            <<<<EXON4                          <<<<EXON5
          DOMAIN   115  146   EGF-like 2.               DOMAIN   147   178   EGF-like 3.
 121 IRCMNGGSCS DDHCLCQKGY IGTHCGQ|PVC ESGCLNGGRC VAPNRCACTY GFTGPQCER|D   180
 181 YRTGPCFTVI SNQMCQGQLS GIVCTKQLCC ATVGRAWGHP CEMCPAQPHP CRRGFIPNIR   240
  <<<<EXON6                                    <<<<EXON7
                    DOMAIN   246   287   EGF-like 4, calcium-binding
 241 TGACQ|DVDEC QAIPGLCQGG NCINTVGSFE CKCPAGHKLN EVSQKCE|DID ECSTIPGICE   300
                                    <<<<EXON8
DOMAIN 288-329 EGF-like 5,calcium-binding.
 301 GGECTNTVSS YFCKCPPGFY TSPDGTRCI|D VRPGYCYTAL TNGRCSNQLP QSITKMQCCC   360
                               <<<<EXON9
        DOMAIN   330   401   TGFBP 1.
 361 DAGRCWSPGV TVAPEMCPIR AT|EDFNKLCS VPMVIPGRPE YPPPPLGPIP PVLPVPPGFP   420
                      <<<<EXON10
DOMAIN   402   446   Pro-rich.                   DOMAIN   449   489   EGF-like 6.
 421 PGPQIPVPRP PVEYLYPSRE PP|RVLPVNVT DYCQLVRYLC QNGRCIPTPG SYRCECNKGF   480
  <<<<EXON11                                     <<<<EXON12
                         DOMAIN   490   529   EGF-like 7, calcium-binding.
 481 QLDLRGECI|D VDECEKNPCA GGECINNQGS YTCQCRAGYQ STLTRTECR|D IDECLQNGRI   540
                                   <<<<EXON13
DOMAIN   530   571   EGF-like 8, calcium-binding          DOMAIN   572   612   EGF-like 9, calcium-binding.
 541 CNNGRCINTD GSFHCVCNAG FHVTRDGKNC E|DMDECSIRN MCLNGMCINE DGSFKCICKP   600
               <<<<EXON14                                     <<<<EXON15
                              DOMAIN   613   653   EGF-like 10, calcium-binding
 601 GFQLASDGRY CK|DINECETP GICMNGRCVN TDGSYRCECF PGLAVGLDGR VCV|DTHMRST   660
                                          <<<<EXON16
                              DOMAIN   656   721   TGFBP 2.
 661 CYGGYKRGQC IKPLFGAVTK SECCCASTEY AFGEPCQPCP AQNS|AEYQAL CSSGPGMTSA   720
<<<<EXON17                                      <<<<EXON18
                     DOMAIN   723   764   EGF-like 11, calcium-binding.
 721 GS|DINECALD PDICPNGICE NLRGTYKCIC NSGYEVDSTG KNCV|DINECV LNSLLCDNGQ   780
                        <<<<EXON19
DOMAIN 765-806 EGF-like 12,calcium-binding.           DOMAIN 807-846 EGF-like 13, calcium-binding
 781 CRNTPGSFVC TCPKGFIYKP DLKTCE|DIDE CESSPCINGV CKNSPGSFIC ECSSESTLDP   840
  <<<<EXON20                                         <<<<EXON21
 841 TKTICI|ETIK GTCWQTVIDG RCEININGAT LKSQCCSSLG AAWGSPCTLC QV|DPICGKGY   900
  <<<<EXON22                                           <<<<EXON23
                            DOMAIN   910   951   EGF-like 14, calcium-binding
 901 SRIKGTQCE|D IDECEVFPGV CKNGLCVNTR GSFKCQCPSG MTLDATGRIC L|DIRLETCFL   960
                   DOMAIN   952   1018   TGFBP 3.
 961 RYEDEECTLP IAGRHRMDAC CCSVGAAWGT EECEECPMRN TPEYEELCPR GPGFATKEIT   1020
  <<<<EXON24                                        <<<<EXON25
                            DOMAIN   1028   1069   EGF-like 15, calcium-binding
1021 NGKPFFK|DIN ECKMIPSLCT HGKCRNTIGS FKCRCDSGFA LDSEERNCT|D IDECRISPDL  1080
                               <<<<EXON26
      DOMAIN 1070-1112 EGF-like 16, calcium-binding        DOMAIN 1113-1154 EGF-like 17, calcium-binding
1081 CGRGQCVNTP GDFECKCDEG YESGFMMMKN CM|DIDECQRD PLLCRGGVCH NTEGSYRCEC  1140
                <<<<EXON27                                   <<<<EXON28
                                   DOMAIN   1155   1196   EGF-like 18, calcium-binding
1141 PPGHQLSPNI SACI|DINECE LSAHLCPNGR CVNLIGKYQC ACNPGYHSTP DRLFCV|DIDE  1200
                                     <<<<EXON29
     DOMAIN   1197   1237   EGF-like 19, calcium-binding.        DOMAIN 1238-1279 EGF-like 20,calcium-binding
1201 CSIMNGGCET FCTNSEGSYE CSCQPGFALM PDQRSCT|DID ECEDNPNICD GGQCTNIPGE  1260
                 <<<<EXON30
                                           DOMAIN   1280   1321   EGF-like 21, calcium-binding
1261 YRCLCYDGFM ASEDMKTCV|D VNECDLNPNI CLSGTCENTK GSFICHCDMG YSGKKGKTGC  1320
<<EXON31                                  <<<<EXON32
                DOMAIN   1322   1362   EGF-like 22, calcium-binding
1321 T|DINECEIGA HNCGKHAVCT NTAGSFKCSC SPGWIGDGIK CT|DLDECSNG THMCSQHADC  1380
                <<<<EXON33
DOMAIN 1363-1403 EGF-like 23,calcium-binding          DOMAIN   1404   1445   EGF-like 24, calcium-binding
```

```
1381 KNTMGSYRCL CKEGYTGDGF TCT|DLDECSE NLNLCGNGQC LNAPGGYRCE CDMGFVPSAD  1440
  <<<<EXON34                                       <<<<EXON35
                         DOMAIN    1446   1486   EGF-like 25, calcium-binding
1441 GKACE|DIDEC SLPNICVFGT CHNLPGLFRC ECEIGYELDR SGGNCT|DVNE CLDPTTCISG  1500
                            <<<<EXON36
     DOMAIN 1487-1527 EGF-like 26,calcium-binding
1501 NCVNTPGSYI CDCPPDFELN PTRVGCV|DTR SGNCYLDIRP RGDNGDTACS NEIGVGVSKA  1560
                       <<<<EXON37                   <<<<EXON38
   DOMAIN    1528   1599   TGFBP 4
1561 SCCCSLGKAW GTPCEMCPAV NT|SEYKILCP GGEGFRPNPI TVILE|DIDEC QELPGLCQGG  1620
                            <<<<EXON39
     DOMAIN 1606-1647 EGF-like 27,calcium-binding         DOMAIN    1648   1688   EGF-like 28, calcium-binding
1621 KCINTFGSFQ CRCPTGYYLN EDTRVCD|DVN ECETPGICGP GTCYNTVGNY TCICPPDYMQ  1680
     <<<<EXON40
                                                  DOMAIN    1689   1758   TGFBP 5
1681 VNGGNNCM|DM RRSLCYRNYY ADNQTCDGEL LFNMTKKMCC CSYNIGRAWN KPCEQCPIPS  1740
<<<<EXON41                <<<<EXON42
                                                   DOMAIN 1766-1807 EGF-like 29, calcium-binding
1741 T|DEFATLCGS QRPGFVIDIY TGLPV|DIDEC REIPGVCENG VCINMVGSFR CECPVGFFYN  1800
    <<<<EXON43                                    <<<<EXON44
                        DOMAIN    1808   1848   EGF-like 30, calcium-binding
1801 DKLLVCE|DID ECQNGPVCQR NAECINTAGS YRCDCKPGYR FTSTGQCN|DR NECQEIPNIC  1860
                              <<<<EXON45
     DOMAIN 1849-1890 EGF-like 31, calcium-binding          DOMAIN 1891-1929 EGF-like 32,calcium-binding
1861 SHGQCIDTVG SFYCLCHTGF KTNDDQTMCL|  DINECERDAC GNGTCRNTIG SFNCRCNHGF  1920
     <<<<EXON46                                        <<<<EXON47
                           DOMAIN    1930   1972   EGF-like 33, calcium-binding
1921 ILSHNNDCI|D VDECASGNGN LCRNGQCINT VGSFQCQCNE GYEVAPDGRT CV|DINECLLE  1980
                                 <<<<EXON48
     DOMAIN    1973-2012 EGF-like 34, calcium-binding       DOMAIN 2013-2054 EGF-like 35, calcium-binding
1981 PRKCAPGTCQ NLDGSYRCIC PPGYSLQNEK CE|DIDECVEE PEICALGTCS NTEGSFKCLC  2040
           <<<<EXON49
                                                    DOMAIN    2055   2121   TGFBP 6
2041 PEGFSLSSSG RRCQ|DLRMSY CYAKFEGGKC SSPKSRNHSK QECCCALKGE GWGDPCELCP  2100
<<<<EXON50                <<<<EXON51
                                               DOMAIN    2127   2165   EGF-like 36, calcium-binding
2101 TEPD|EAFRQI CPYGSGIIVG PDDSAV|DMDE CKEPDVCKHG QCINTDGSYR CECPFGYTLA  2160
  <<<<EXON52                                       <<<<EXON53
                         DOMAIN    2166   2205   EGF-like 37, calcium-binding
2161 GNECV|DTDEC SVGNPCGNGT CKNVIGGFEC TCEEGFEPGP MMTCE|DINEC AQNPLLCAFR  2220
                           <<<<EXON54
     DOMAIN 2206-2246 EGF-like 38,calcium-binding          DOMAIN 2247-2290 EGF-like 39, calcium-binding
2221 CVNTYGSYEC KCPVGYVLRE DRRMCK|DEDE CEEGKHDCTE KQMECKNLIG TYMCICGPGY  2280
        <<<<EXON55                                        <<<<EXON56
                               DOMAIN    2291   2332   EGF-like 40, calcium-binding
2281 QRRPDGEGCV|  DENECQTKPG ICENGRCLNT RGSYTCECND GFTASPNQDE CL|DNREGYCF  2340
                               DOMAIN    2333   2400   TGFBP 7
2341 TEVLQNMCQI GSSNRNPVTK SECCCDGGRG WGPHCEICPF QGTVAFKKLC PHGRGFMTNG  2400
<<EXON57                                          <<<<EXON58
                    DOMAIN    2402   2443   EGF-like 41, calcium-binding
2401 A|DIDECKVIH DVCRNGECVN DRGSYHCICK TGYTPDITGT SCV|DLNECNQ APKPCNFICK  2460
                           <<<<EXON59
   DOMAIN 2444-2484 EGF-like 42,calcium-binding            DOMAIN 2485-2523 EGF-like 43,calcium-binding
2461 NTEGSYQCSC PKGYILQEDG RSCK|DLDECA TKQHNCQFLC VNTIGGFTCK CPPGFTQHHT  2520
<<<EXON60                                          <<<<EXON61
                         DOMAIN    2524   2566   EGF-like 44, calcium-binding
2521 SCI|DNNECTS DINLCGSKGI CQNTPGSFTC ECQRGFSLDQ TGSSCE|DVDE CEGNHRCQHG  2580
                              <<<<EXON62
     DOMAIN 2567-2606EGF-like 45,calcium-binding           DOMAIN 2607-2647   EGF-like 46, calcium-binding
2581 CQNIIGGYRC SCPQGYLQHY QWNQCV|DENE CLSAHICGGA SCHNTLGSYK CMCPAGFQYE  2640
                                               <<<<EXON63
                    DOMAIN    2648   2687   EGF-like 47, calcium-binding
2641 QFSGGCQDIN ECGSAQAPCS YGCSNTEGGY LCGCPPGYFR IGQG|HCVSGM GMGRGNPEPP  2700
                                          <<<<EXON64

2701 VSGEMDDNSL SPEACYECKI NGYPKRGRKR RSTNETDASN IE|DQSETEAN VSLASWDVEK  2760

2761 TAIFAFNISH VSNKVRILEL LPALTTLTNH NRYLIESGNE DGFFKINQKE GISYLHFTKK  2820
                                                     <<<<EXON65

2821 KPVAGTYSLQ ISSTPLYKKK ELNQLEDKYD KDYLSGELGD NLKMKIQVLL H|
```

Page 231 missing

# Appendex E

The sequence of the **Elastin gene** taken from the NCBI reference number AC005056

Reverse flip of **AC005056**

Key:

```
CI              cluster of initiation
nnnnnnnnnnnn    forward primer
nnnnnnnnnnnn    reverse primer
nnnnnnnnnnnn    Exons
                Splicing site
NNNNNNNNNNNN    3'or 5' DOMAINs

All of the following are from MAYADA TASSABEHJI 1997
(H)             hydrophobic domain
(X)             cross linking domain
(A)             subject to alternative splicing
(C)             C-terminus: conserved cysteines and four terminal basic
                residues.
```

**SAH Reference** **AC005056 Reference**

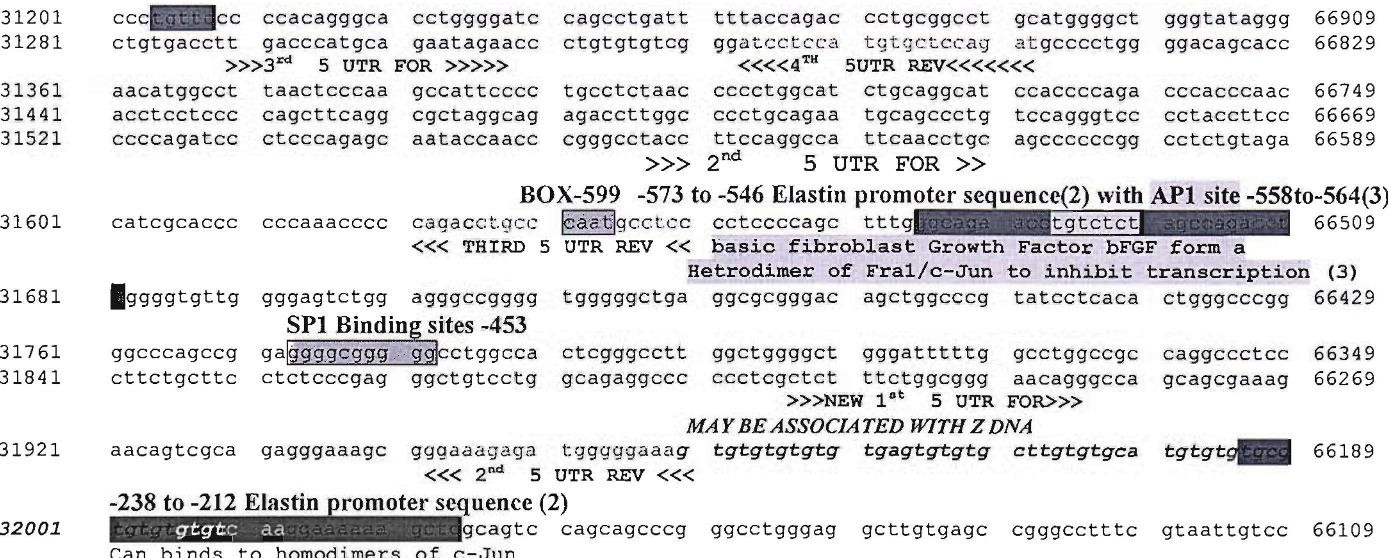
```
1     aagcttgcac  tacagaggag  tcaagggtca  agtgcataat  gagagtggca  gggcctggcc  gaggtgggca  gggagaggtc  98109
81    agtgggttcc  agggaatttg  aggcccatga  ctaagctgcc  tcagggggc   tgatgggcag  gtattataag  caggttgcag  98029
161   aggagaaaag  aggctctgga  aaagccagga  ggtgtagaca  gccaggtggg  ctcaccaggt  atgacgatgt  gggcacagga  97949
241   agtgaaggtt  aaagccaaga  gtggcacagt  ctcctgggga  tgctgagccc  atgtcccaag  gtcggggtcc  tggttccagt  97869
321   gtccagctct  aaccccagcc  actgtgtgtg  tctgggcaag  ctgccaaccc  tctctgggcc  tcagattctc  tttgtcaaag  97789
401   gacaggcacc  tctctgtgcc  atgatgctgt  gaacctcttc  cctgctgatg  gggtatccct  acaccccсaa  cccactcttc  97709
481   ctcttccaag  acagcctgag  gctagtttct  ctgccctgtt  ttcccaccag  tgacccaata  tcctccccat  cttttttttt  97629
561   tttttttttg  agacagagtc  ttgctctgtt  gcccagtctg  gagtgcaggg  gtgtgatctt  ggctcactgc  agcctccgcc  97549
```

.X
.X
.X
.X
.X
.X

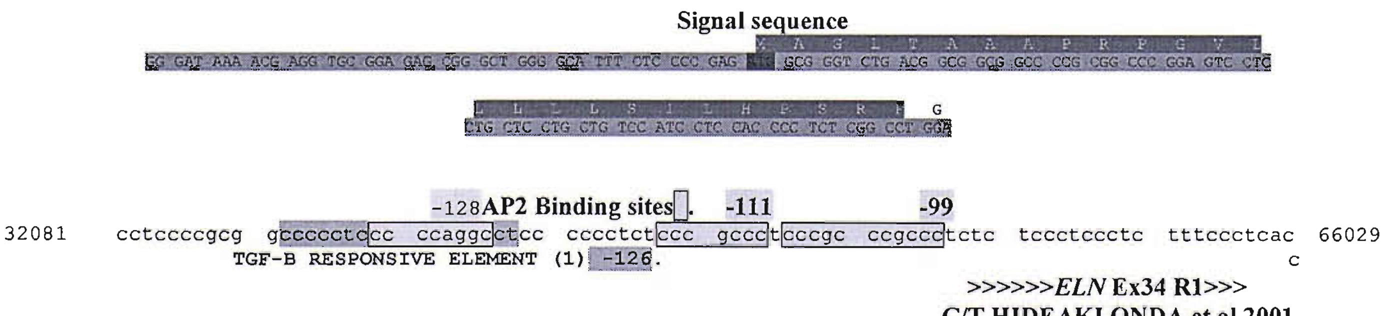
.X
.X
.X
.X

```
26321   cagctcaggg  gagcaggcag  ctcggccctg  ggtcccacaa  catgtgcagc  ccagggaaga  gaaggctctg  ctttccaggt  71789
26401   gtggagccga  acagagatca  tcgctccacc  ctaaagcagt  gtggctggag  gagtgaaata  ctctagaggc  ttgagctcca  71709
26481   ggtcacagtg  cctttgggtg  accttccgct  ggtcatgtgt  ttcctcaagg  atccgtttcc  tcacctgtgt  ggcaggatag  71629
26561   tttattgagt  catgaatatc  tgttcttatc  caccctataa  ccctcccctc  ttgccacttt  ccaagcggct  ctccacgttg  71549
26641   caagcaaaaa  gccgccgaaa  gtcatctttg  agatgagagt  gagtatctta  gactggaagc  atctcctgcc  agcccggaga  71469
26721   tgacagaatg  tcctgcaggc  cactcccttg  gaagcacctc  tggaggtggt  ccagccccga  gcttcccagc  ctcagattcc  71389
26801   agcacagtct  tggggtgcac  acgagcacac  aggctccaac  aagggagtca  gggagtcagg  gcccctacta  tctccaggct  71309
26881   ccatctaaga  tgcctggatt  ttggtggggg  gtggaagtgg  aggaagacag  acccagtaac  tgggatggaa  ttcactgcca  71229
26961   cctacccgtg  tgacattggc  ttaaattaaa  ttaaacagat  ctaaggcaga  caccaagatg  ctctttctgg  tggggccgag  71149
27041   tttatagaca  ggactcctaa  cccttctgct  tgttaaccgg  cataataata  tccccattgg  aaaatggaag  tgtgctttgc  71069
27121   aaacggtcaa  gtgtcaaaca  aacaagaggc  gtgctctgga  ctcctccatc  ccgggtccca  tgggaagggg  gtgcacagag  70989
27201   gagggccggc  tcacgatgtt  cccctctaga  tggcaacaga  cagcccgtga  ggcactgtag  gaacagcaca  gggcagaaca  70909
27281   gcccgtccgg  gacacccacc  tccccacccc  cggctcacac  tggccaaaga  tccagactga  caaatgactc  agacacacaa  70829
27361   acgcccgggg  ctaagaggaa  acacggcatg  gcccggaggt  tgcagggggc  agcactcaca  gggatggggc  tttctgctat  70749
27441   gcctacccct  ccacccagcc  atgatcagtg  gcctgaaatg  gtgtcatgga  cccccctccca  tgtttgagca  cttgctgtat  70669
27521   accaaattct  tacagatatg  tttgtactct  taaatggtta  tgaagatcct  gtcttcatag  gacgaaacca  gggattctag  70589
27601   gggtcaatta  ctgttgatgt  caccatcctt  cattattatt  agataatctg  gccaaactca  tacactaatc  aacatcagtc  70509
27681   cccatcacaa  ccacaacaac  gactgctgac  ccttccgaac  ccactcccag  cctgggggt  tgcagaagca  ccgagttcca  70429
27761   gaatttcagc  ctccctccca  caacgtccaa  cgcacccct  cgaagcacag  cctgtccttc  ttttgagctg  gctctgtgtg  70349
27841   atgaagggac  acccagcaaa  tggcccagac  ccatgggcat  tcgatctatt  ttcggttgca  gtccctttac  ctctgacacc  70269
27921   tgtgcctgcg  ccagacccct  tcctttcctg  gccggccatc  cgaggggtcc  ctggtggcct  ctgggaagag  aggaactgca  70189
28001   gaagtccctt  tcatttcca  gctgaggtgc  tggctctggg  ccgatgctgc  cgggcacccc  acaggaggta  gccgccctct  70109
28081   ctggctcagt  caccagcacg  agggcagggc  cactcctttg  gggctgcttg  tcacgggtgc  cctttgccag  cctccaagga  70029
28161   catccttcct  tctctgtccc  tctcccctgt  cctcaagccc  cgtcccattt  tgaacctctg  catagcagga  ttcagattcc  69949
28241   cttccaagcg  catcagaact  aaggctagct  ctggctgggc  atggtggttc  acgcctgtaa  tcccagcact  tagggaagtc  69869
28321   gaggcgggag  gatcacctga  ggtcaggagt  tcaagatcag  cctggccaaa  acggagaaac  cccgtctcta  ctaaaaatac  69789
28401   aaaaattagc  caggcgtggt  ggcgcgtgcc  tgtaatccca  gctactcagg  aggctgaggc  aggagaattg  cttgaacccg  69709
28481   ggaggcggag  gttgcagtga  gccaagatcg  tgccactgca  ctccagcctg  ggtgacagag  aaagactccc  tctcaaaaaa  69629
28561   gaaaacaaac  caccaaaaaa  gaactaaggc  tagctcatct  caccagtcct  ttatgatcta  aaccagtgct  gtccccggtg  69549
28641   gcagctgctg  gccacatgtg  gctaatggag  cacttgaaat  gtgacaagtc  cagattgaga  cgtaccctcg  atgcaaaata  69469
28721   cacaccagtt  ttaagacctg  gtacaaaaaa  gaatgcaaga  tatctcaata  acttttattt  tgattggatg  ttgaaaggat  69389
28801   tatactttag  atatgttgtg  ttaaatatat  tactaagaat  agttttactt  atgttttctt  ttatttcata  tggctactag  69309
28881   taaatttaag  atcataggag  gggctggatg  cggtagctca  cacctgtaat  cccagcattt  tgggaggccg  aggcggcgg  69229
28961   atcacctgag  ctcaggagtt  tgagaccagc  ctggccaaca  tggtgaaatc  cggtctttac  taaaaataca  aaaattagct  69149
29041   ggcagtggtg  gcggatgcct  gtaatcccag  ctactcgggg  gctgaggcag  gagaattgct  tgaacctggg  aggcagaaat  69069
29121   tttagtgagc  ggagatcacg  ccactgcact  ccagcctggg  tgacagaacg  agactccgtc  tcaaaaaaaa  aatctgaact  68989
29201   ttgtctggat  agcgacggag  ggtgtctgag  caaagcacga  tgttctcaga  gccatccaga  gactgggaac  tctgtcctac  68909
29281   caaaatgggt  gcacccagag  gagctggaag  atagagatgc  cctgtgggtg  actcatctgg  ttgcccttg  gtggcactga  68829
29361   gccaggcaac  aagcgggaaa  ggcttgtcaa  tgataacggc  aagaccaacg  atcaccaggc  agccaggtgc  acccggtagc  68749
```

```
29441   cctctgtggg  caattctgca  tgtatgacct  tatttagtca  tacagctagg  caaggaggtg  acactaagca  tttgagtact  68669
29521   gatggggtct  tgaatgaaca  agtaaatgaa  tgaatgaatg  gatggtggcc  gatgactgtc  atctctagtt  cttacattat  68589
29601   gtatcttatg  catcagacta  cacggcatgg  atattcagtg  caatggattt  gattgtgaat  tgaggccccc  ttttgcaatc  68509
                    >>> 8th 5 UTR FOR>>>>>
29681   ttgttgcatg  ttatgttcgt  ggagctcctg  caatgagcag  aggctgctaa  gcagaagcag  cgcacaggct  gggggcagga  68429
29761   cacagggccc  ccggcctcag  gggagaggtc  tgttccagcc  acctcaagtc  tattcctgtg  agtgcccctc  cacacccctc  68349
29841   ctggatcccc  ccactgcaaa  ctctctcccc  ctggctccgg  actctctccc  cgtcctggaa  ttaataaaga  ctctgtgcag  68269
29921   actgcggttt  gcaagtctga  aagctggttc  ctgcccgtgt  cactgcctcg  agaagagagg  gGtccagctc  cccacagtag  68189
                                                            >>> 7TH 5 UTR FOR>>>>>>>
                                                                              G>C WASEEM
                                                         CRE-2178 PUTATIVE AP2-2171
30001   cccctgcctt  cctccttccc  aggcaggcgg  aggcacgcag  atataccatt  gacttccct  cccctgcagc  aggcacatcc  68109
  <<<<8TH 5UTR REV<<<<<<
30081   tgggcatcga  gcttcagacc  ctgcccctga  gcagccccta  accccaccaa  caaagggtgg  cttggggggg  ctttcacccc  68029
30161   agcataatct  ccatcagcta  ccctcaaagc  acccccaaat  aaacacacac  cgtaagtaag  agctgtacac  tggctgtgtg  67949
30241   cgtacatctt  caagacaatt  ctcccagcat  gcccctacct  tccaaaattc  cagagctgct  ccctccaaag  acccagggaa  67869
30321   aaggaagggt  ttgtccaggg  tcctggggtg  gccccgtata  gaccaaagcc  tgatagctgt  cctagaagca  gagtacttgc  67789
>>>6TH 5'UTR FOR>>>>>>>>                                        ss4943615 A/G
                CRE-1819                                      ss6500846 G/A rs3757583
30401   agagcgagtg  acggcactg  tggtattgac  accagtccta  gcaccagctg  aacacagagc  atttttgatc  tagcagaaat  67709
           ss4943616 C/A rs3757584
        <<<7TH 5 UTR R<<<<<<<
30481   acaagaccac  gttgtatttg  tctttgcaat  aatctcttag  ctaggaatac  tgatcacctg  tagacagata  aggaaactga  67629
30561   tgctctgtgg  agaggttttc  ctaccagaaa  ggctagagcc  agaaatttac  ttctaggtcc  accaatacct  gcctttgacc  67549
                                                                        >>> 5TH 5'UTR FOR>>>>>>
30641   aatgcctgca  tttgaccttt  ccacgctgag  ccacccctgc  tggcactcca  gactgccaca  gtgctcctgc  ctccacaagg  67469
                   <<< 6TH 5 UTR REV <<<<<<<                    Glucocorticoid-responsive element-1441
30721   ggtctttaac  tcatccctcg  gagccatcgt  ggtgcaggga  aaagcccaca  gggcgtgtgg  cttccatgct  gttccctgac  67389
30801   tggctgtgac  ctaggacaag  gaacaagttt  ccctctccta  ttctctaggt  ctcacatttc  ttctcctcta  gcagtagtgg  67309
                               Glucocorticoid-responsive element-1319
30881   gaagtgaggg  gtgggggaca  cgaccctccc  ctgttcatc  ccacactcca  accccaaaa  tcccccaggg  tccccgtcca  67229
                                               >>>>>>> 4TH 5'UTR>>>>>
30961   gctcagtcct  ggggcagaa  atgcagagtt  ctccaggaac  gtggtcccag  ctgtttcagt  gcaggccgcc  ccctcctggc  67149
          <<< 5TH 5'UTR REV<<<<<
31041   caccagcgga  atgtcagcct  tcccagaggg  gcCgggagaa  cagcagtcga  gaagctccca  gactggtgtg  ggcgctagct  67069
                                  C>G WASEEM                          HIDEAKI ONDA et al 2001 C/T-1042
31121   gtgctcagcg  tggggatggg  aggtgaccca  gtgataatgg  gaagctgggc  tgcctgtcag  tctgCggggg  gctcccacct  66989
   Glucocorticoid-responsive element-1025     HIDEAKI ONDA et al 2001 G/A -972       C>T WASEEM
```




## Exon 1 Signal Sequence (27aa)




BOX-57 Cluster of Initiation (CI) -32-33 CI-15-16 CI-7-8

32161 agccgacgag gcaacaatta ggctttgggg ataaaacgag gtgcggagag cgggctgggg catttctccc cgagatggcg 65949
start

32241 ggtctgacgg cggcggcccc gcggcccgga gtcctcctgc tcctgctgtc catcctccac ccctctcggc ctggaggtaa 65869
<<<NEW 1st 5UTR REV<<<<
32321 ggacccctcg cccctgtccc cagcgctgcc cacagctgcg ggccctttgg gccaggtgac tagacgctca aggggagac 65789
<<<<< *ELN* **Ex34 F1**<<<<<<
32401 cttcgtccct gggaactgca aggggtccca ggtgggagcc cctcagccac tgggtcctcg gaactgcccc tcccaggagc 65709
32481 cagccccgtg cccaggagga acctgtcaaa gagacaccct cgactcagaa tagtcccagc ccagcatcca ggctttgctg 65629
32561 ccctgcgagg gctgaggggg cccctgctca ggtgcccggg tgcccgccct gtagggggat cagagcatcc tgggggaggg 65549
32641 cccagcggga acccctgggt ctggagacac cacagtctgc acacctgtgg acacagctgg ggtgcacacc agccctgggg 65469
32721 acagggaacc tgggcttgat accagctcaa ccgcagtgag ctgtgtgacc tcagcaggtt ccttgacccc tgggaagccg 65389
32801 ggatcccccc gtcaggaagg tcagggctcc gctgaggatg ggtttgcagg ttgggaagct gctgctgtgt tctccctgct 65309
32881 gaggcaggca gggactcctg gtatgcattt gtgtgtgtgt gtgtgtgtgc atgcacacag gaatgagcgt gctgtgtgct 65229
32961 cgtgtgtgta tgtgtgcatg cacaggtgtg catgtgttca gatgcccggc tgtgtcactg aacagggaaa gggatttctg 65149
**ss5010350** G/A rs3823878 **ss2733494 A/C** rs1859763
33041 gaggggtaca tatgtctgct cacaggttca tccttaggca ggtgtatgcg tatgtttgga ggaggtatat gaagctgctc 65069
33121 cttagtgtgt gcaaaagtct gccacctgac agacatgtag acatgtacac atgtagacat gggcgatctg gcatccaggt 64989
33201 gtgacactgg tgtggactgt cactgtgtgt gttatggagt gtgtacatct gtgtgtgtga atgtgtgtgc gatgtggcac 64909
33281 gcaagcatgg aaaggcgggc tggagaaggt gtctgtctcc tggcagaggc agtgtggctg agacagaggc tggagtggat 64829
33361 cctggggcac tgctgggcca tctgctatgt gccccaaccc aggcacagtt ggttatgggg tttgagggtg tgccctgtgg 64749
33441 ggctgctatg ccgcggccaa cgggggagga cagaagaggc tcccgggcat gggaaaaagg agggcatttg gtaggggggt 64669
33521 ctgggtggag tctggtttcc tcgggagtcc tccttgccat gagaaccaaa cgctgagtgg ggggcgggag ggccacgggg 64589
33601 tggagtcaag ggcctgggca ggagctgcag ctgtttccgt gagattgaac agcttcctgt gcaggtccg ctgctggacc 64509
33681 tggatgaggc caagggacat gggtacccac aggcgggccc cagcaggggc aacgagccag ggcctgaaag cagcagcccc 64429
33761 gggtccgtct cagcctcctc cccaggtgta gcccccaaac acccagccac ccttcttccc agatggggcc accacctata 64349
33841 gggtggcagt tcatcatcaa ggaaccttga gaattagaaa ttattatcat cacttgacaa acaaggaagc tgaggctccg 64269
33921 gcaggctaag gcagtccccg ggggttagta tcagccagct ctctgcccca ggtcaagggg caaggggggtt aggagtgagc 64189
34001 tagcaggcaa catttgggga ggggaaggct tggggctgag tccttaaagc aggcatgccc taggcagaat cagctctctc 64109
34081 cttgtgtcct cagtgacacc gctaaacgcc cagctggcct cacactttgc agagatcaat ctcagctctg tcccaggtgg 64029
34161 ctggacttgc taggtccaaa taaggggtgt gtgaactgat gagctcgtgt ggggttttgc acagaagtgc tactgaagtc 63949
34241 tgcagaggtc ggggagccaa cacagggagt ggagcgctca gtcacatgct cactgccact cacctggctg tcctggggga 63869
34321 gcctggcttc tccctcaagg aggagggtgt cacttatgca agttcagcgg ctccaaggaa ggaaccctgt tgcgggggtg 63789
34401 ggggagcagg aggaggggac aagccgcaca gctcccaggc tgtttcaaaa gtgagcatga gaaattaacc cagggataat 63709
34481 tcttccccat cccccagaaa tagtatcagg gaNttttttt tttttttttc ctgatttcag aactgaagat ggtttggaac 63629
**ss4314057 -/AA** rs3082600
34561 agaggtttcc ttttctttta ttttcctttt ttattttttat ttttttacct tttttatttt ttttaaattg agacagggtc 63549
34641 tcgctgtgtt gtccaggctg gtctgaattc ctgggctcaa gtgatcctcc cacctccgcc tcccaaagtg cttggattac 63469
34721 aggcatgagc cactgcctct ttttaaatat ttttttaatt attgttttcc gtccttaagt aggaacaggg gttttcaacc 63389
34801 tggctatcta ttagacttcc cacccaagaa tccctgcttg ctttggttgg agagggacct ggacacagac ttttcaggct 63309
**ss1289977 A/G** rs868005
34881 tcccaggtga cggtggggtc gccagggcca agaagcactg gtttagaggt cgatcccgct ctgtgtgtct gccgaggccc 63229
34961 cggcatgagg gcaaatgtgc cctgtgtggg aggggacaca gtcacccagt ggaggacagc ttgtttttga aaacccttgg 63149

```
                                                                        ss8153908 -/T rs5884929
35041   cagccttctg  catgcaaaca  acaaacagct  tcgtgtggtc  ttatctgctc  cccaagccag  ccccctggac  tttggtttgg  63069
35121   caaagccccc  tgccaagatc  ggctggcccc  aggctcgctc  ccaaacaggc  tgcttccagc  tctttgaacc  tggagcccag  62989
35201   ggcccaggca  aatggcccta  catgtgactt  ggggtcagct  gggggtccca  gaccaggctc  ttcagctcac  tggttctcca  62909
35281   tttccctggc  tgcaaaatgg  ggagaaacag  ccctgcctga  agtacctcct  agggccagtg  agagcttcta  gcaatgtaat  62829
35361   atctgggggg  aaaaaaagca  tcaaataaaa  ggaccgtggg  aatgggcatt  tttaaaataa  tctattctct  tattcagagt  62749
35441   ctgctaattt  gggctcatgg  atccttgtaa  ttagtgcaca  ttgctaaggt  ttaatactcc  ctctcactac  ttacaccatg  62669
                                                                        ss1312277 A/G - rs884843
35521   ttttgtctag  tggaggggca  gctaatgtgt  aaatagatgt  gcaatgatga  aagaagcaga  caggactcgg  ttgggggtgg  62589
35601   ccagaaggag  ccttgagaaa  tccaggaagg  cttcccagca  gaagtgacat  ttgagttaaa  cctaaagaga  aaagtaaatg  62509
    ss2733493 A/C  rs1859762                                  ss3312460 C/T rs2356532
35681   gggccaggtg  cagtgactca  tgcctgtaat  cctagcactt  tgggaggccc  acatgggcag  attgcttgag  cccaggagtt  62429
35761   tgagaccagc  ctgggcaaca  taggaaggcc  cccgtctcta  tttaaaaaat  aaatgaatat  ttaaaaagaa  agaaaaagaa  62349
35841   agaaagagag  agagagagag  agaaaggaga  gagagagaga  cagaaagaaa  gaaaggaagg  aaggaaagaa  ggaaggaagg  62269
                                                                              ss5375216 -/TTCC rs4027708
35921   aaggaaggaa  ggaaggaagg  aaggaaggaa  ggaaggaagg  aaaatgttgg  aggacctgtc  tgtggatgag  gggtaagtcc  62189
36001   ccagactaca  ggccctggca  ggcagaactg  aagagacaga  agagcagatt  ctcgattgga  ggaagccctg  tctgcagtta  62109
36081   gtgctgacag  catttggagg  tggtgatttt  cccatcatgg  gaggtatgca  agcagaacac  gaaggatgtt  tgtggaagga  62029
36161   ggtccagccc  aggccagaag  actattatga  tccctgagag  tctaggatcg  tgattctaag  atgcaagaga  caagttccac  61949
36241   cctgggctag  tcggtgaaac  atgccaaaac  cacataaaac  aagacaagtg  ttctccaggc  caactgctta  gggccagctg  61869
36321   agcagcacag  tcagggagag  ccgtgggctc  gggaaaggca  ggcgggagag  tctgagctca  gccctggtgc  gggggtagga  61789
36401   cttggacaga  ggcaaggaca  ttcctggcag  gagaaacggc  ttgtgccaag  gagtggaggg  agaagtgagg  atcgccacgg  61709
36481   ggcctgggga  agcagctgtt  accttttcaa  agtcctttcc  aatttccagc  tgctgtcatt  ttattctcac  ctctctcact  61629
36561   tgcagagagg  caggaatcct  tcgtagcccc  taaaccttct  ccatttcctc  attgcaaagt  ggggattaaa  acctcgggct  61549
36641   ccccaggccg  tcaggttcag  gcatgaagtt  tggcaaatgc  caccatcatc  agggtctgaa  cggtgggggt  ggagagtggg  61469
36721   gacagaggag  atctggggga  ggggaagcta  agtcttcccc  ccagaaagtg  cctggggtgc  aggagggcag  aaccccagga  61389
36801   gaaaggaagt  cagggttcag  ccaggcccag  gggctcaaca  agcccaaggc  agggggactc  ccagggcaga  tggggatgaa  61309
36881   acaggttaga  gctggattca  cagacaacca  ccctgccacc  cctctactta  attcacccat  catgtattga  gtaactactg  61229
36961   catacctggc  cctgggccac  ggactgtgca  gtgaaatggg  aaaaaatgaa  acgtaagtgc  cactgtcttt  ttatttttga  61149
37041   gcaattactg  tctgccaggt  ccagaagagg  acatagccta  ttacttgaac  ccttttacaa  cttgcagcaa  ccttgcaacc  61069
37121   ttgcagcaac  cttgcaacct  tgcaacaacc  aagtgattat  ctgccccatg  tcacaaaaaa  ggaaaccgat  gttcagagag  60989
37201   gttgagtgac  tttcccaagg  tcgcacagat  ggtaagaagc  agcgaaggca  gaattagaag  ccagaggggt  ctgttctgag  60909
                      ss1297500 C/T rs873647
37281   cactttgctc  ttccccccac  ctccaccaca  ctgttccctg  cagtctgctc  tggagccaga  gagtagagag  gaagggagag  60829
37361   aggcacgtga  agaatgaggg  agagggggag  aaagaggagg  aaggagagag  gccgaggcag  agaaagacgc  ataagagact  60749
              ss1293096 A/G rs870424
37441   gttgagagct  ggaggccgga  gccagggaag  gcatcagagg  cccgaggcgg  gccatgtgtg  tgcttggcag  cggatgacag  60669
37521   ggcccaagtg  ctccttctca  gcaggcccag  ccccccctggg  gccaccactc  cgagccttga  ggttctgacc  agattacttg  60589
37601   cacacttcag  gcctggactg  cagtgcctga  ggagccacaa  gccctgggag  ggggtggcgc  cggggtcttg  gctccggtgg  60509
37681   ctgtgacagc  tccacagccc  tgaccctgct  gggctgcagc  tgaggctgac  agggccctgg  ggccagcaag  cccagctgtg  60429
37761   cagcgtcgcc  aggcctcaga  ggccagacgt  ctggccgcga  ggcctccctg  gacacattgc  agagacaccc  gttcccccag  60349
37841   ccccgcccat  gccccacggg  gctcccggcc  cctttcagca  aaagcagtaa  gggagggctg  ggctggggcg  ttggcaaggt  60269
37921   ggcttgtcaa  gctcagaggg  gggccttctg  aggtggaggc  agggcccccg  gtttgggacc  ctgtctgttg  ccatggcgat  60189
```

38001 ggctcaggca gggtctggat ccagcatcac agcgttcccc agtgagagct tcactggtcc atggggactg aactcgtgat 60109

ss4943617 C/T rs3757585

38081 cctgtctggc cggtgaattc gcttttcctt aggaaactat tcactgtgct cattcatggc gtccccctgc ggccccggcg 60029
38161 cctcctaaga cagcccgtgt cagacggcca gagccccacc caagctcggc tgcccagtgg ctgaggacgc tgacctcacg 59949
38241 ccgtcctatc agtgtgggaa tggctgcatt tgtgctggca gctgcacctc cttaatgctc gctgtagcag cccagcctgc 59869
38321 ccttcccaaa gctctggcct ctacccggca gctcagctcg ggaagggagt gggctgcctg gtgagaggtg ctcagaccca 59789
38401 gtgtcttcca ccccaggccc tctccctgac tccagggaga accaggtaac acagcaagaa caatcatact ctaaggactg 59709
38481 gcagcagggc gcaggggcct gtcatcctag cactttagga ggccaaggca ggaggatcgc ttccaggagt tccagaccag 59629
38561 cctggggaac atactgacac cccatcttta caaaaaaatc aattagccag atgtggaggt gcacgcctgt agtcccagct 59549
38641 acttgggaga ctgaggcagg agaatcgctt gaacctagga ggcagaagtt gcagtgagct gagatcttgc tactgcactc 59469
38721 cagcctgggt gacagagcaa gactctgtct caaaaataaa gcactggcct gagcttcaat ccaggccccg tctctttcca 59389
38801 gctgtgtggc cccaggcatg tcccttaacc tctctgagcc tgatttcatg tcttaagatg atagtctctg ggagcggtgg 59309
38881 ctcacacctg taatcccagc actttgggag gccgaggcag gcagatcact tgaggtcagg ggttcaagac cagcctgccc 59229
38961 aacatagtga aaccctgtct ccactaaaaa tacagaaatt agccgtgctt ggtggcatgc acctgtaatc ccagctactc 59149
39041 aggaggctga ggcagcagaa tcgcttgacc caggcagacg gaggttgcag tgagctgaga tcccaccact gcactccagc 59069
39121 ctgggtgaaa ggccaagact ccgtctcagt ctcaaacaaa acaaaatgat gatagtaagc cctgcccttc caacttcaca 58989
39201 ggccgatttc aaagagcaag aatgagaaca cactttgtaa acagaaactg ctctggagac aagcaatgtc attattgcca 58909

## Exon 2(H)(17aa)

G V P G A I P G G V P G G V F Y P
GGG GTC CCT GGG GCC ATT CCT GGT GGA GTT CCT GGA GGA GTC TTT TAT CCA

39281 tcagcattat tcttgtttcc atgtaattgt gggttttgcc attgaaagta atggcaaaaa tcgcagttac ttttgcacca 58829

>>>>>*ELN* Ex33 R1>>>>>>>

39361 gcctaatagt tctggctcct ggaggactga ctctacctgt ttcctttc[illegible] gggtccctgg ggccattcct ggtggagttc 58749
39441 ctggaggagt ctttatcca g[illegible]aacgtac atgaaacttc cacacaccca ggtcatgcgg atgatgctga tgtccataat 58669
39521 agatgcacat tttgacacta cagaaggtag gaacattgac acgcctacca gaagtcacac ccacttaaaa atgcaattaa 58589

<<<<<<*ELN* Ex33 F1<<<<<

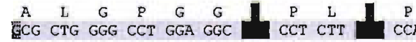
39601 caagacattg atttacagct attaagagca tacaggctgg acatggtggc ttatgcctgt aatctcaaca ctctgggagg 58509
39681 cagagaaaca aggattgctt gaggccagga gtttcagacc agcctggaca acatagtgag atcccccctct ctacacacac 58429
39761 acacacacac aaatttaaaa ttagccatgt tctgcattga gaaaaatgaa aatgaagtat agaaattaaa aaaaaattag 58349
39841 ctgggcatag tggtgcacgc ctgtagtcct agctactggg gaggctgagg caggaggatc acttgagctt aggagttcaa 58269
39921 ggctgcagtg agctataata gcaccactgc actgaagtct gggtgacaga gtaaaaccct gtctctaaaa agaaaagggg 58189
40001 aaaaaaagaa aagaaaaatc aaaagtaata ctttcaatac ttaaaaagta tttgaggcca ggcacggtgg ctcaagcctg 58109
40081 taatcccagc actttgggag gctgaggcag gcggatcaca aggtcaggag atcaagacca tcctggctaa cacggtgaaa 58029
40161 ccctgtctct actaaaaata caaaaaatta accaggtgtg gtggcgagag cctctagtcc cagctactcg ggaggctaag 57949
40241 gcaggagaat gatgtgattg aacccaggag gcggagctta cagtgagccg agatcgcgcc actgcactcc agcctgggcg 57869
40321 acagagggag actccgtata aaaaaaaatt aaaaaattaa aaaaagcatt tgagtgtaca gaagtgacct ggaagtgggg 57789
40401 agagacgcca gcactccccc agatacatac tgacactgga ctactggcca agcagaagag aaaaccgagg cttgcagaga 57709

## Exon 3(H)(10aa)

```
 G   A   G   L   G   A   L   G   G   G
GGG GCT GGT CTC GGA GCC CTT GGA GGA GGA
```

```
40481    gcaggtcttg  cccaaggtca  cgtagttagg  cagaggtgga  ttcagccata  gctgggggtg  gcagcgggct  tgcctggcag  57629
                                                                                                   >>>>>>>
40561    aagtaccgat  gatctctctt  tctctttctc  tccccccac   gggctggtc   tcggagccct  tggaggagga  g gagctca  57549
ELN Ex32 R1>>>>>>>>>>>
40641    gaaaccacac  ttgttcatca  ctgaaagggc  ctgggtttca  cccgagccac  atgcatcctc  agacctgaga  accctgggag  57469
                                                    <<<<<<<<< ELN Ex32 F1<<<<<<
40721    acccgagcat  caaggactcc  ctcatttcac  agacagcatc  caggcggtgg  gaggaggctt  ctttgagaac  cagaacccat  57389
40801    tggccttccc  agtccctccc  ctagacttta  tgtctggtgc  cctctgcctt  cctatttctt  cttttttttg  gtgggggagg  57309
40881    acgggggac   agagtctcgt  tctgtcgccc  aggctggagt  gcagtggcgt  gatctcagct  cactgcagca  tctgcctccc  57229
40961    aggttcaagt  gatgctcctc  cctcagcctc  ccaattagcg  aggactacag  gcgcgtgcca  ccacgcccag  cttatttttg  57149
41041    tatttttttct agagacaggg  tttcatcatg  ttggccaggc  tggtctcgaa  ctcccaactt  caagtgatct  gcctgactcg  57069
41121    gcctctcaaa  gtgctgggat  tacaggcgtg  agccaccttg  cccggcccct  tgcccggccc  cttcctattt  ctttttatcc  56989
41201    tcaaaaacct  ccctgccaag  ccggctgagc  aggcaccatc  ttctttttttt tttttttgag  acaaagtctt  gctcttgtca  56909
41281    cccaggctgg  agtgcagtgg  cgcgatctca  gctcactgta  acctccacct  cccaggttca  ggtgattctc  ctgcctcagc  56829
41361    ctcctgagta  gctgggatta  caggcacatg  ccaccacgcc  tggctaattt  ttgtattttt  agtagagaca  gggtttcacc  56749
41441    atgtttgcca  ggctggtctc  aaactcctga  cctcagatga  tccgcctgcc  tcagcttccc  aaagtgctag  gattacaggc  56669
41521    gtgagccacc  acccccagcc  catcatcccc  ttttgaggct  gagactctgg  gactcagaga  gtataaacca  cgagttagaa  56589
```

## Exon 4 (X)(11aa)




```
41601    gtagagctag  gttcccaggg  gctccaagtc  tgagcgggag  gacctggggt  gtgtgattcc  acactgccca  cactttgccc  56509
41681    gggttggggg  ttggataagt  agtagatgga  taagctgggc  caccccattc  actatcttct  cttccctctg  c cgctggg  56429
```

```
          >>>>>>>ELN Ex31 R1>>>>>>
41761   gcctggaggc [illegible]cctctt[illegible]ccag[illegible]aa gacccaaggc ctcggagcat tgagagacag cgagggagct ggggagggag 56349
                                                            <<<<<ELN Ex31 F1<<<<<<<<
41841   gagcctaccc agctgggaat gggacaagga aactaggaac aggacaagga agccaacggg caggaggaag gagggaggtg 56269
    ss3248153 rs2301995 C/T          ss6501831 G/C rs4717864
41921   tggacaccga ttagcctccc aaggatgagt aggccggggc caggtcccag ggcttccagg aacaagaggc tggaagcagc 56189
                                                            rs2301994 ss3248152 C/T
42001   tccatgtcct ccctgtgtga gggcgtctag catctaccct acatgtgcat gtgtgttcac ccagctgtcc agagactcct 56109
42081   ctctgcaaag gagaggctgt tagaggcaac atcagggatt gctccggttc cagtggcctc caagccttac atgaccctcg 56029
42161   ggagctcaga cacagatcct cagccccaga cttcccctga gtggtcccct agcccttgac cccagaccaa taccccagat 55949
42241   caccctgacc ttgatgccat tctgagccct tctcctgact ccagatggag cactggtctc aagccaaagc ttaaccccag 55869
42321   cccaagccct gagcctaatc ccaggctgaa ccctgaccct gctgaccctg gtcccacagg acccccaatc ctggccttag 55789
42401   gccaagccct gaccccaagc ccaagctgag caccgaccat ggccccggat aagctcccct cctaatctca ggccctttcc 55709
42481   agctgtgtcc agaccacgga gttttccagt aagaaggtgt ggccggatct atcctggcct cgcccaaggg gctgggatct 55629
42561   gccacgaggg tccactctca gctctgggga atgcaccccc aggaaggaag gtggggcagg tccctccggg aaatactggc 55549
                                                                 ss4987843 G/A rs3801458
42641   ggataaggag caggtggagg aaacagcgtc tctgtccaga ttgatggtcc ctcggcatga gacgctccac atgtgacttt 55469
42721   ggccgcccca caccctgagc tgtgttgccc tggctcagaa gcccaagtct ccaggacaca gtttagatgt ccaagagctc 55389
42801   ttatgaaatc ctgatagttg ataatccaag tgcaaatgaa tcttcaagac ccagattccc ctgcgctaaa agtccatctg 55309
42881   ggttgtgaaa caccagggct ggagccagga gaggcgggag aacggggtgc ctggggtcta cagggctctg caaggccagg 55229
42961   agaccctggg agcttgccac ttatccccgc cagcacctat accctgtctt cctggagcag gtgagggttc ttagcagggt 55149
43041   ctttgacggt gaagagagga caggggagat gatagtggtg cagaggtgtg gaggtaggtc tgtgcttggc ctagaagacc 55069
43121   caaattgagg aggttgtgcc tgcaaagctg gagggtacat tcattgatcc cactaggagc aagagacaca tagcaggttg 54989
43201   tggttaaaca ccagaggaaa gacagacaac tccctacgaa gttccaggat gtgtagttgc aaaagtgaag cttttgttaa 54909
43281   tgaaagattt aggttagact tctgaaagaa cttccagctc agggtgaagg agtcaatcca aaggaaagtc agagcatctc 54829
43361   cttctgataa aaagggcagt gtttccaggc aaggggctgg attccatgac ctccccgaag actcagggct acagaagggt 54749
43441   cggcagagcc cccatggcac ccccagggag gcccccacat ccgggctgcc aggacgtctg ggcctgccca ggcaggatgg 54669
43521   aatggagcac aggcagcgcg tgtcatgctt gggaacaatt aatcccttgg tgacagaggg tgggtggcct ggcagctgct 54589
43601   tttccacgtg ggcgccacca caggtccaag ctggccctgc cctgcctcag ctctcagcgg gggaccaggc agctggaagc 54509
43681   ctgcacaccc aacaccctga gcactagcca gaggccagcc gcccactgag aggggcacgg gcatggacaa gaaggaactg 54429
43761   ggagcaccca ctgtgtgccc tgggtaggtt gcacaactag gtgctttctg cacacatgct cactggtccc agcaaggcag 54349
43841   gtgatataat ggccattttg tagatggcaa aaactaagcc tcatcggaga ggcaaattga cctgcttgtg ctcatggagg 54269
                                                                        ss6501832 G/A rs4717865
43921   ctgggtagca gcagcagggt ctgaatctag aactctctga ctctagaaag aaggaggatg gagctggcca ggtccctgga 54189
44001   ctcccagctg tggtcagccc ctccccgctt tctcctcctt tccagatgct agaatccttg gggcatttgg ggcaggtgag 54109
44081   tgcaaaggaa ggagggcagg cagagacaga agtggtcagg ccacatgaca ggatgctcag gctgggatca tgggaaaaag 54029
                                                ss4987844 T/G rs3801459
44161   gaagatggcc ccagagaagc tgcctgcccc ccactgcagc cccagcctct tctcacttcc tgtgcagtcc tcagcttccc 53949
                          ss4987845 C/T rs3801460
44241   ctcccagccc ctctgtggac aggaaaggtc tgagcatgtc ccttgccctc tcagagctca ggaaacatgc ctggcttgag 53869
44321   agaagcgact gccagcgcag gtgggtctgg cccgggtgct gggaagtgac tcctgcctga tcgccaggca ctcccgcccg 53789
                   rs1859761 ss2733492 A/C
```

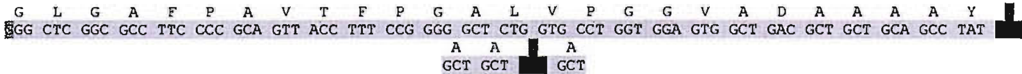
```
44401    gaggcaaaga  tgcttcttac  atgaggctgg  ctgagtttca  gaggggtcca  gctcccacta  gtaggaaatg  agggacgagg  53709
44481    agtggttagt  aacggagacc  acacccaggg  aaagccaggg  gggacctctc  cagggtacag  caggctccat  gtttgggatc  53629
                           ss5010351 G/A rs3823879
44561    catccctgga  gaggactcac  ccctgaggct  caggactagc  acttgtagga  gagatcgtca  cataattaga  aacaaacctg  53549
44641    gcatggttgt  gaactctgac  ctgatggccc  attggtcctg  gccaaggcag  tggaatgtct  cccgcagtgg  gagacgaagc  53469
44721    gattgggccc  agatgtgggg  aggagagagc  cagagggacc  catttggcgt  ctcataaaca  tcttagtagg  aggctcctgg  53389
44801    gctgcagggc  aggctggatg  gaaggacaga  tgggtagtgg  ggacacagga  gttccccgat  gcaggtgaag  gggaggggac  53309
44881    tgagtcaaga  gatatctgca  aggaagcaga  agagagacct  cactggcctg  gggtgaggtc  tcgctcacgg  actctgctct  53229
44961    gtcccagccc  ttgctgaaag  ccctgctgaa  tcttgttagc  cagcagggct  tctaggagaa  gcacaggcct  ggcccaggct  53149
                              ss5001663 A/G rs3815251
45041    tgttagggat  cggtgcaatg  acacctgcac  tgcacatagt  agtcgctcac  taaactactg  gtggatgcta  ttttatcagg  53069
                                                                 rs2286257 ss3226639 C/T
```

## Exon 5 (H) (12aa)

```
45121    atcgaccctg  agcatcacag  gcttagggac  cagcctaggg  acctgagtgg  ctgatcacag  cactgcccta  actccaggag  52989
                                                                 >>>>>>ELN Ex K F1>>>>>
45201    acatttccca  ctctgggcct  aggaacactg  cctacactcc  tgtctctgtt  tcttatccac  agttcccgga  gggcttgcgg  52909
                                                                        HIDEAKI ONDA et al 2001  C/T
45281    gtgctggcct  tggggcagg  gagtgctgac  acccaagaaa  gatatcccct  gtggggacca  gcccctgagc  tcaacccagg  52829
45361    gctggtatgc  agcacggtca  tggaacaagg  gtgcaggcca  ggttccgtcc  tgggcactga  cggggactga  tgctgatacc  52749
                                                                             <<<<ELN Ex K R1<<<<<<
45441    aactgcccca  agcagctgcc  caggacatga  ggaaatcaca  gctgagtacc  tgggttcctt  tataacatag  tagtgaaggc  52669
45521    tgcatggagg  aagccccatg  gaccaaggaa  gcccaaggga  gtcgcttaac  tcagcagggg  ttcaggaagg  cattctagag  52589
45601    gaggtgggtg  gggaaatttg  accccaaaag  atccatgcag  ttttcaggaa  ggctgaagtg  agaggatccc  ttgagcccag  52509
45681    gaggctgagg  cttcagagag  ctatgatcat  gtcactgcac  tccagcctgg  gcgacagaga  aagacctcat  ctcttttgtt  52429
45761    gttcttgttg  ttgttgagac  agagtcccac  tctgtcaccc  aggctggagt  gcagtggctc  agtctcagct  ccccgcacct  52349
45841    ggtgcagttg  atctcagctc  accacaacct  ccacctcccc  ggttcaagtg  attcttgtgt  ctcagcctcc  caaatagctg  52269
45921    gaattaccag  tgtgcaccac  cacgctcggc  taatttttgt  attcttagta  gagacggggt  ttcaccatgt  tggccaggct  52189
46001    ggtctcgaac  tcctggcctc  aagtgatcca  ctcacctcgg  cctcccaaag  tgctgggatg  acaggctcaa  gccaccgtgc  52109
46081    ccagccacaa  gaccccatct  ctaaagggaa  aaaaaaaaaa  agagccatgc  agttttgtca  gagctgtcta  acacattaac  52029
46161    ctcggcacac  aatctagttc  tggtacttag  gcgcggtggc  tcacgcctgt  aatcctagca  ctttgggagg  ctgagacggg  51949
46241    cagatggctt  gagcccagga  gttcgagacc  agcctgggca  acatggtaaa  atcttgtcac  tacaaaaaat  acaaaaaaaa  51869
46321    aaaaatttgc  cgggcacagt  ggcatgcacc  tgtagttcca  cctacttgga  aggctgaggc  gggaggattg  cttgagcctg  51789
                                          rs2856729 ss4044369 G/C
```

46401 ggtggttgag gctgcagtga gccaagatcg caccactgca ctccagcctg ggtgacagag tgagaccttg tgtcaaaaaa 51709
46481 aaaaaaatcc attaccggtg agcagtgctc acatggatgt cctgcaggca ttgatgtctg gcagagagcg gaagagcctc 51629

## Exon 6 (X) (31aa)




46561 caatgtgctt cctgagtggg gcacagccag gcagggccag agcgtaggag tcttcatagg tgtggtagct caggacctca 51549

>>>>>>*ELN* Ex A F1>>>>>>>>

46641 ccccatcctc ccctccgc ggctcggcgc cttccccgca gttacctttc cgggggctct ggtgcctggt ggagtggctg 51469
46721 acgctgctgc agcctat gctgct g ctg gagtg gtgctctttg ggacatgcca caagccctct ggcttccgtg 51389
46801 gggccctccg ctttgcaaag aaccttccct caactcaggc ttgggagccg ggtgggtggg attgtcagtt gcaatctact 51309

<<<<<*ELN* Ex A R1<<<<<<<

## Exon 7 (H) (17aa)

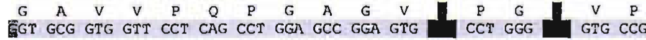
G A G L G G V P G V G G L G V S A
GGC GCT GGG CTT GGT GGT GTC CCA GGA GTT GGT GGC TTA GGA GTG TCT GCA

46881 ggggctcagg gaggcagcta gctgtgccca gtctgaaggt gagttagtaa cggggccaga cctcaatgct gggccctcag 51229
46961 catgcagccc cgggcccttc tgcctcccca ctgttcctta cgcaatgcct cacctgtcct ggctctgc gcgctgggct 51149

>>>*ELN* Ex GandH F1>>>>>>

47041 tggtggtgtc ccaggagttg gtggcttagg agtgtctgca g acgatgg ctatccccga actccctggg tcaaagttgc 51069

## Exon 8 (X) (17aa)



```
47121   aggcctgggt  ggagccaact  ctgatgcagc  cccttctgtg  cc[illegible]gtgcgg  tggttcctca  gcctggagcc  ggagtg[illegible]c  50989
47201   ctggg[illegible]gt  gccgg[illegible]cag  tgcggaatcc  ctggggctgg  aggacagagg  gcagggaggg  gcagagggca  gggaggaaca  50909
47281   gagccctctc  catgccactg  cacttgggga  ggaagggcag  ggcctggctt  cagtcgggca  gagaaactag  ccaggctggt  50829
```

<<<<*ELN* **Ex G+H R1**<<<<< <u>rs2528796</u> <u>ss3544769</u> C/G

```
47361   gcctgcgtct  gtgaaatggg  gaggaggggc  ctggcccact  tcccgggccc  ttctcccggg  atcttggggt  agaaagagac  50749
47441   gggctctgtg  gcaggctgtc  gggcgacaga  tggggaaact  gaggctcaaa  gaggcgagtc  agtgtggaag  ggcctgcagc  50669
47521   gagtcccggc  agagctgggg  gagccccgtg  tcctctgact  ccccatctgg  tactcgttct  gccccaccaa  gccctttagg  50589
47601   aagtgacttc  aggttaaaca  aaagactgcc  tgagtctacg  cagcagggag  ggacatctgg  aagctgttag  ccagaggtgg  50509
47681   gctggcccag  ccatgtaaac  aggctccagg  agacggagat  aaatccacac  accgaagagt  ttgcagatga  cttccgaaac  50429
```

## Exon 9 (H) (14 aa)

```
G   V   G   L   P   G   V   Y   P   G   G   V   L   P
GGT GTG GGG CTG CCA GGT GTA TAC CCA GGT GGC GTG CTC CCG
```

```
47761   tcgtgggagg  agggttgtgg  ccaccccacg  tctgtccctg  gctgcccctg  tcgggcacag  aggctgtggg  tttgagggcc  50349
```

>>>>>>*ELN* **Ex I F1**>>>>>>>>

```
47841   ttggagctgc  ctgggtggga  agggctgggg  aggggtccct  ggaggctgag  ctgctgctag  taactttgct  ttcttttggc  50269
47921   cacaggtgtg  gggctgccag  gtgtataccc  aggtggcgtg  ctcccagatg  agagcaagga  gggaaacagg  gactctatag  50189
```

<<<<<*ELN* **Ex I R1**<<<<<<<<<

```
48001   gaagaaagca  gccaggacgc  agtggctcat  gcctataatc  ccattgcttt  gggagactga  ggcaggagga  tggcttgagg  50109
48081   ccaggagttt  gagaccagcc  tgggcatcat  agtaaggccc  tcgtctctac  aaaaaaattt  agccaggcat  ggtggtgtgc  50029
48161   acctgtagtc  ccagctactc  gggaggctga  ggtgggagga  ttgcttgagc  ccagaaggtc  caggctgcag  tgagctacga  49949
48241   tggtgccact  gcacagcagc  ctgggtgaca  gagcaagacc  ctgtctcaaa  acaaagaaaa  agaaaaagga  aaagaaaagc  49869
48321   agccccctca  gcctccctag  tacccccctc  gcctccccga  aaagcagagg  ccaccagaag  ccctgggtcc  tgacctgagc  49789
48401   catttcccca  aactccaaag  ctcaggctaa  cagacataag  gtccctggca  gtccgtccat  cccctgaagg  tcaaccagcc  49709
48481   ttcctgcacc  ccacccaagc  cctgatccca  ggcacagacc  atcatcacag  cccgaggcgg  gcccccgagg  acagatccca  49629
48561   gtgtgaccag  ctatttccca  aaagctgtga  gtcaccaagg  ggccacccaa  tgagtctgcc  ctgcaggaga  ccctaaagag  49549
48641   ctgcccctca  ccctccttc  ccctgccagg  gcctgaccac  cccacccaac  atgtgacctc  ctgaattccc  ccaaaagccc  49469
48721   agatttaggc  tgttatgtgg  gatttgctga  tggagtggcc  cttcctccct  ctctccccca  ctacgctgaa  agccaaggct  49389
48801   gcccagcccc  aggggagaca  ggagaggtgg  aagacgctgg  catggtccca  gctgtagctg  gatggctgct  ttgggggtct  49309
48881   gggttgtgag  cgagagcccc  agagagcatg  ggcctccctg  gtctacgggc  gagtggcacc  gcgcagcagc  tgtccctctc  49229
48961   tttacctttt  ctgcggaggg  tggtcgagac  ggcttttgtg  attaactcca  gcatagagat  gaggctgact  gaggcaccag  49149
49041   ctcacatggc  tgggaaatag  gattaaaact  cagatcttct  gaccttgagc  cagacaggga  aggaaagtca  gcactaaaca  49069
```

## Exon 10 (X) (24aa)

```
 G   A   R   F   P   G   V   G   V   L   P   G   V   P   T   G   A   G   V       P       A   P
GGA GCT CGG TTC CCC GGT GTG GGG GTG CTC CCT GGA GTT CCC ACT GGA GCA GGA GTT     CCC     GCT CCA
```

```
49121   ggtcctttct  ccacccaccc  cgtgagccgt  gcctgtcact  gacagtcagt  cccaagggag  gtcagctggg  ggacccgagg  48989
                                                    >>>>ELN Ex C F1>>>>>>>>
49201   aggggagggg  ttcccagcag  ggcctgcaag  gcctgccttc  ctacactcac  tgctttgtcc  cccggc  ga  gctcggttcc  48909
49281   ccggtgtggg  ggtgctccct  ggagttccca  ctggagcagg  agtt    ccc     gctccag    atgcagct  gtctggacag  48829
49361   agggctgatg  gcagggactc  tccaaccacc  ttctggcccc  gggtgtgaaa  tggggtggga  tcctggactg  gctcagggcc  48749
             T/C Waseem          <<<<<<ELN Ex C R1<<<<<<
49441   ctctgggtga  cacttttttat  gttccagccc  tgggcagcct  ggtgctgaaa  aacctcagag  tgtggctggg  cacactggct  48669
49521   cacacctgta  atcccagtat  tttgggaggg  tgagctggga  ggatcgcttg  agcccaggag  ttaaagacca  gcctgggcaa  48589
49601   catagcaaga  ccttatctct  acaaaaacat  gattaaaaaa  ttagccggac  atggtcacac  acacctatga  tcccaactat  48509
49681   tcaggaggct  aagacaggag  gatctcttga  gcccaggaga  tcaaggttgt  cgtgagctat  gatcacacca  ctgcactcca  48429
49761   gcctgggcaa  cagagacccc  gttcccccgc  aaaataagaa  agaaagaaaa  gagggccaga  tgcagtggct  catgcctgta  48349
49841   atcccagcac  tttgggaggc  cgaggtgggc  ggatcacctg  aggtcagcct  gaccaacacg  gtgaagccct  gtctctacta  48269
49921   ataatacaaa  aattagctgg  gcgtggtggc  ctgcacctgt  aatcccagct  acttgggagg  ctgaggcggg  agaatcgcca  48189
50001   gaacctggga  ggcagaggtt  gcggtgagtt  gagatcaccc  cattccaccc  cagcctgggc  gacaagagcg  aaactccatc  48109
```

## Exon 11 (H) (10aa)

```
 G   V   G   G   A   F   A   G   I   P
GGT GTA GGT GGA GCT TTT GCT GGA ATC CCA
```

```
50081   tccgaaaaaa  gaaaccccag  agttcatgtg  agcgcagcat  gcgatgactg  gtctgggaag  aactggccat  tccttgggcc  48029
                                                              >>>>>>ELN Ex L F1>>>>>>
50161   tcctggcccc  ttggtgctgt  ctggcccagt  gtccacagtt  ccagggctgt  agtgacagct  ttttatcatt  ac  gtgtag  47949
```

```
50241   gtggagcttt  tgctggaatc  ccaggagagg  caaggctggt  gggagaagca  gggtggccag  ccaggcagag  gctctggcgt  47869
50321   tgggaggggt  tgggcaccca  agatcccatc  caagcctgcc  caatttctcc  cacgcctgca  gagccaggtc  ttgcagtggc  47789
                          <<<<<<ELN Ex L R1<<<<<<
50401   tgtagcacaa  agtggcaccc  atcccaggcc  tgctcccccc  actggcatag  cagcccaaaa  ttcaaaggag  tccagaaaac  47709
50481   cctccagcct  gtgccaaggg  ctgcaggccc  aaactggggc  ccaggcccca  ggagggagag  aatcccgtag  cttcctggag  47629
50561   gaagtggcct  gtcactggga  atttctcaac  tgacccatga  tggagattca  gggagtccct  cgaagcagga  tggtttctgg  47549
```

## Exon 12 (X) (24 aa)

```
50641   atgctgtagc  aagctccttc  tgtctgagca  gaaagtggag  tgggtggcgg  agggtttgga  agggggtgct  gggacctgaa  47469
                                                                                      >>ELN ExD F1 NEW>>
50721   cttgctctct  ttattcccac  gagttgga  ccctttgggg  gaccgcaacc  tggagtccca  ctggggtatc  ccatc  gc  47389
50801   cccc  ctg  cctg  aagt  cagagggacg  gttcaagatg  caccactcgg  ccgggtgtgg  tggttcacac  ctgtaatccc  47309
50881   agcactttgg  gaggctaagg  cgggcagatc  acttgaggtc  aggagttcaa  gactagcctg  gccaacatgg  caaaaccccg  47229
                    <<<<<ELN Ex D R1<<<<<<
50961   tccctactaa  agatacaaaa  attagccagg  tgtggtggca  taggcctata  atcccagcta  ctggggaggc  tgaggcagga  47149
51041   gaatctcttg  aacccagaag  gcagaggttg  cagtgaactg  agatcgcgct  actgcactcc  agcctgggta  acagagtgag  47069
51121   actctcctcc  aaaaaaagaa  aaagaagaaa  aaaaaaatgt  accactcact  gcacctccct  gcacagaagt  cagcctgggt  46989
51201   acaggtgtct  ctggtggcag  ggaaggggtg  ttgaagcccc  tgtatggtca  ccagccaagg  agagcatggg  aaagtcatct  46909
51281   gcaggtattg  aactcacaca  cacacgctca  tgcacagaga  cccatagtcc  cgatctgaag  ctattaggct  ggtggaaagg  46829
51361   aacacggttc  attggaaaga  tccctctaac  atccacccac  tcgtgctccc  gcccctacct  ctgcaatcag  cttagacaat  46749
51441   aagatgcttt  ccactgagga  gggggtgtaa  ggaaaatact  cagactccag  ggccatgatg  gggcttgaat  ttgtaggggg  46669
51521   atgggtgttc  catgggcctc  gggggcagag  gtgtcctctc  cctccttcac  ccagcgcctt  tgctctcctg  ggagcagctc  46589
```

## Exon 13 (H) (14aa)

```
51601   aggaccctga  tgtgggaggt  ctcaagcttt  agcacctgtg  gggtagatc  tgtccaccca  ggttggtggg  agcccagcaa  46509
```

>>>>>*ELN* **Ex J F1**>>>>>>>

```
51681   ggcatggggc  agcccctgag  tttgctctgt  cctctctcca  [illegible]gtggctatg  gactgcccta  caccacaggg  [illegible]ctgccct  46429
51761   atg[illegible]gagtg  agacccttct  agactgtggg  cttccagctc  tttccctctc  cagggtccta  gcaagggtgc  tgtgcttcca  46349
```

T>A WASEEM???

<<<<<<*ELN* **Ex J R1**<<<<<<<

```
51841   gccctgggcc  aggagagcac  ctcgctgggg  cagggttggg  gtcttggagt  gggaatctca  gaaggaaagg  gcatggaatt  46269
51921   tggactcagc  atgggtctcc  atccctgccc  tcccacctat  caataatgag  accttaagtg  agctgtgttg  cttctttgag  46189
52001   cctctgtgtt  ctcatctgta  aaatggacat  aacaactcca  tacatgtgtt  tgtatttgtt  ttcactctgg  ctgggggtga  46109
```

rs1009879 ss1472470 A/G

## Exon 14 (H) (20aa)

```
 G   Y   G   P   G   G   V   A   G   A   A   G       A   G   Y   P   T   G   T
GGC TAT GGG CCC GGA GGA GTG GCT GGT GCA GCG GGC     GCT GGT TAC CCA ACA GGG ACA
```

```
52081   caggtgcaga  ctcaggacag  cttgggcccc  gagggcagag  caggggggagg  gggagggcag  cagtggtgat  gtctgcacag  46029
```

>>>>>*ELN* **Ex F F1**>>>>>>>

```
52161   atgaccatca  agcctctctg  ttttgc[illegible]gc  tatgggcccg  gaggagtggc  tggtgcagcg  ggc[illegible]gctg  gttacccaac  45949
52241   agggacag[illegible]  aaggaaagcc  tcacgtcact  tccagccaag  ggagcactga  tcttccaggc  tccagagccc  tggggtgggt  45869
52321   gaggttccct  cctgaaagca  gcagcccacc  ctgcatccag  accctggtcc  aaacctggag  caggatcctg  ggggaggagt  45789
```

<<<<<<*ELN* **Ex F R1**<<<<<<<<

## Exon 15 (X) (18 aa)

```
 G   V   G   P   Q   A   A   A   A   A   A   A       A   A   A       F
GGG GTT GGC CCC CAG GCA GCA GCA GCA GCG GCA GCT     GCA GCA GCA     TTC
```

```
52401   ggggcagctc  catcagcctc  tgcctactct  gaagctccca  tgtataccca  catgtcagtg  gattggctct  cttggggctg  45709
```

>>>>>*ELN* **Ex XX R1**>>>>>>

```
52481   ggaacaagtg  ggctctggag  atacaggagc  actgtttcaa  ggtctctccc  ctctgcttcc  ttccccc[illegible]g  ggttggcccc  45629
```

**G/A HIDEAKI ONDA et al 2001**

```
52561   caggcagcag  cagcagcggc  agct[illegible]gca  gcagca[illegible]t  tcg[illegible]gagtg  cccctggagt  ccccacctgg  tggcctccag  45549
52641   gcccctagcc  tctccattcc  cattactatt  gacagcctgc  ctccaaagtg  gcccctacat  acccccattt  actcaaaatt  45469
```

<<<<<<*ELN* **Ex XX F1**<<<<<<

```
52721   ttcaacatca  cccatctatc  tatccctccc  tccaaccatt  cttccatgca  tccatgtatc  cattcatcct  tccatcaatc  45389
```

```
52801   tattactttc  tccatccctc  cctccatcca  ttcccattcc  tccatgcatc  catctatcca  tccatccatt  catgtactca  45309
52881   tctgcccatc  catgcatcca  tccactcatc  cgtccattca  ttcatccatc  catctctccc  tccatccatt  catccatcca  45229
52961   tccatccatt  cctccatgca  tccatctatc  catccatcca  cttatctatc  cattcatcta  cttatttaca  catccatgca  45149
53041   tccatccatc  catccactca  tccatccact  cattcatcca  tccatctctc  cctccatcca  ttcatccatc  catctatcca  45069
53121   tccatccatc  catccatcca  tccatccatt  tctccatgga  tccatctatc  catccatctg  tccacacatc  catccatcca  44989
53201   tttatccatc  cacccactca  tccatctatc  cattcatcca  cccatccatc  actacctccc  tccattcatc  catcctccat  44909
53281   tcctccatgc  agccatctat  ccatccatcc  atccacttat  ccacccattc  atccatccat  caatccctcc  ctccatccat  44829
53361   tcttccatcc  atcaatccct  ccctccatcc  attcttccat  ccatccatcc  acgcatccat  ccattcctcc  atgcatccat  44749
53441   ccatctatgc  acccatccat  ccatcccctc  atccatccat  ccatttatcc  atccatcact  ccctcgctcc  attcatccat  44669
53521   ccaccatcta  tccatctatt  cccccatata  cccattcatc  catccaacca  ttcctccatg  catccatcta  tccctccatc  44589
53601   cattcctcta  ggcacccatc  catccatttc  tccatgcatg  catccatcct  tccatccatc  cttccatcca  tccatccacc  44509
53681   atctattcct  ccatgcatcc  attcatccat  ccatccattc  ctccatgcat  tcatccttcc  atccatccat  ccacccatca  44429
53761   attcttccct  ccatccattc  ttccctctat  ccatccactc  atccatccat  tcttccctct  atccgtccac  tcatccatcc  44349
53841   attcctccat  gcatccattc  ctccatgcat  ccatctatcc  atccatccat  tcatccatcc  actcatccat  ccatccattt  44269
53921   acccatccat  ccactcaatc  catccatcca  tccatccata  catcaatcca  tacatcaatc  cacccatcca  tcactccctc  44189
54001   catcgattct  tccatccatc  catctgccca  ttcttccatg  catccctcca  tccattcctc  catgcaccca  cccatccatt  44109
54081   tctccatgca  tgcatccatc  cttccatcca  ttcatccatc  catccatcca  tccatctatt  cctccatgca  tccatccacc  44029
54161   catccatcca  tttctccatg  catccttcca  tccatccatc  catccatcca  tccatccatc  cactcaccca  tctatccatc  43949
54241   catccatcca  tctacccatc  aattcttcct  tccatccatt  cttccctcta  tccatccact  cattcatcca  ttcttccctc  43869
54321   tatccatcca  ctcatccatc  cattcctcca  tgcatccatt  cctccatgca  tccatctatc  catccatcca  ttcatccatc  43789
54401   cactcatcca  tccatccatt  tacccatcca  tccactcgtc  catccatcca  tccatccatc  catccatcaa  tccatctatc  43709
54481   catcactccc  tccattgatt  cttccatcca  tccatccatc  catctgccca  ttcttccgtg  catccctcca  tccattcctc  43629
54561   catgcaccca  tccatccatt  tctccatgca  tgcatccatc  catccatcca  ttcctccatg  catccatgca  cccatccatc  43549
54641   cattcaccca  tgcattcatt  cattcattca  ttcattcatt  cattcattct  ttcctccatg  cattcatcca  tgcctccttc  43469
```

**ss5023641** -/ATTC rs3837128

```
54721   cttccctccc  tcactgcctc  ctcccaccca  tccatatcag  agatcaggga  caaacaagag  ttctcttaga  gaatatagcc  43389
54801   actgttgtgg  gagtcaagaa  acgggaggca  tttttttcta  cctgaggaaa  gttagaaagg  cttcccagag  aaagtgacat  43309
54881   gtgatttagg  ccttaaaaat  aaagatggag  cttaccatgt  ggttaaaaca  ggaaagagga  ggccaagcag  agggaatggg  43229
54961   taccacggtg  gttagaaaga  taattttagg  tggtgcatat  ataagcattt  aaaatattta  atagttatgt  atcaatttaa  43149
55041   tgtgtactat  aaaaatgtta  actcactgct  caggtcagtg  agttcttggc  tagcatggct  gcagatggaa  cgtggtattc  43069
55121   ctgtgcccgt  gcagtgggtg  gagtgggcat  ctgtcccctg  ccccaggtag  tgcccacctc  agggtcagac  cactagggct  42989
55201   gaagtcctcc  ttcatcctca  gccccaggga  agccctttcc  ttatctcccc  accagcccga  gagagcgaga  atgtggggag  42909
55281   aagcctgaag  ctgggcctcc  cagtggaggc  cccgcaggcc  cccctcccag  cacccgaggc  tccttggccc  cagcggctgg  42829
55361   tggggacggc  tgcaatgtgg  gagcgggaga  gcagggctgt  gaggggctgc  cagagccaag  cagccaggcg  cttggattac  42749
55441   aaacttggct  gcatcttcgg  aacacaggga  gaggaagtct  tgaacattcc  tgcaggggac  cctctggccc  agggagcggc  42669
55521   cacttgtggt  ttctcagtat  gtggcagtga  ttagaatggg  atttgtctga  aaacatacaa  gtcccttaat  gagtgtgttg  42589
55601   aaatggacac  tttgggggag  agtcaaggaa  cagtggagtg  gggtgggggc  ctccccagac  aggcccatct  ggagacaccc  42509
```

## Exon 16(H)(30aa)

```
 G   A   G   A   A   G   V   L   P   G   V   G   G   A   G   V   P   G   V   P   G   A   I   P   G   I   G
GGT GCT GGA GCA GCC GGA GTC CTC CCT GGT GTT GGA GGG GCT GGT GTT CCT GGC GTG CCT GGG GCA ATT CCT GGA ATT GGA
```

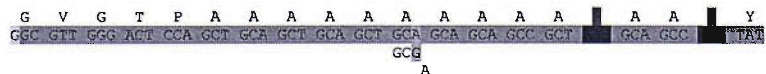
G I A
GGC ATC GCA

```
55681     gggccccatt  cctggataag  atcacactgg  tgaaaacgcc  ggcgtctaag  tggccatcct  gcctgtcctc  aggagggtcc  42429
```

**>>*ELN* Ex19and20 R1>**

```
55761     ttgggaaact  acattgcact  gtccccatct  caac[illegible]gtgc  tggagcagcc  ggagtcctcc  ctggtgttgg  aggggctggt  42349
55841     gttcctggcg  tgcctggggc  aattcctgga  attggaggca  tcgcag[illegible]aa  catctgtccc  agcaggggc  gggtgtgtcc  42269
```

## Exon 17(5)(X)(20aa)



```
55921     ttgagatggc  cacagggcaa  ggacctcacc  ctctgtggct  gtgttttc[illegible]  gcgttgggac  tccagctgca  gctgcagctg  42189
56001     cagcagcagc  cgct[illegible]gca  gcc[illegible]tatg  [illegible]gagtgcct  cccggggtgg  caagtccacg  gctcgggccc  ctgcatagac  42109
```

**rs6979788** A>G **ss10419219**

```
56081     ctcggagacc  ctagccgcaa  agccagatgg  acttggcctt  tgttccttcc  caaatatgca  ttgttcatgc  ctccttacct  42029
```

<<<<<<*ELN* Ex19and20 F1<<<

**A>C WASEEM**

```
56161     ttgcccctte  tgatcactct  acctgagatg  ccatctctat  tgttttgccc  tgattaactc  agcagaggga  ggggaccctg  41949
56241     cagaggggac  atggctcctc  ccactccatc  ccctccaggg  ccagcccaca  ggtgtctgct  gcatcaacta  aatgggtgcc  41869
56321     cagtggtgag  aattctgact  gtgctttagg  aatagatcac  attctagcta  tgcacagtgg  ctcacgcctg  taatcccaac  41789
56401     aatttgggag  gctaaaacca  ggagtttgag  accagcctga  gcaacatact  gagaccccat  ctttccaaaa  agtatttaaa  41709
56481     aattatctag  gcatggtggc  acatgcctgt  ggtcccagcc  acctgcaagg  ctgaggtgag  aggattgctt  gagcctagta  41629
```

ss6500850 G/A rs4717127

```
56561     gttcaaggct  gcagtgagct  atgatcatgc  cactgcactc  cagcctgggt  aagtgagaat  ttggttcaaa  aaaaaaggaa  41549
56641     agagagacag  aaagagagag  acaggaagga  aggaaggaag  gaaggaaaga  aggaaggaag  gaaggaaaga  cggaaggaag  41469
56721     gaaggaagga  aggaagggag  ggagggaggg  aaggaaggaa  atgaaggaag  ggagggaggg  agagagagag  gcaggaagga  41389
56801     aagaaaggaa  gaaacaaaag  agagaaagag  aaagaaagaa  agggaaagga  aggaaggaag  gaaaaagaaa  agagggaggg  41309
56881     agggagagag  agagagggag  ggagagagaa  agaaagaaag  agagagagag  agagagaaag  aaagaaagaa  agagaaagga  41229
56961     aggaaagaaa  agaaaagaaa  aagaaagaga  tcacattcct  ccagctcact  gattcaaatc  ctagagctct  ttagggaccc  41149
```

## Exon 18(H)(49aa)

```
 G   A   A   A   G   L   V   P   G   G   P   G   F   G   P   G   V   V   G   V   P   G   A   G   V   P   G
GGA GCT GCT GCA GGC TTA GTG CCT GGT GGG CCA GGC TTT GGC CCG GGA GTA GTT GGT GTC CCA GGA GCT GGC GTT CCA GGT
```

```
V   G   V   P   G   A   G   I   P   V   V   P   G   A   G   I   P   G   A   A   V   P
GTT GGT GTC CCA GGA GCT GGG ATT CCA GTT GTC CCA GGT GCT GGG ATC CCA GGT GCT GCG GTT CCA
```

```
57041   tcttagtctc  tccacatctc  tctgatgagt  aggatccatg  cagaggaaat  gtcaacccac  ctgcaatcct  gcattcagga  41069
                                                                                            >>>ELN Ex18 F1>>>
57121   ccaactgtca  cttccatact  ctactaacca  cccttctagc  ccctctgagg  ttcccatagg  ttaggggaac  aatgcttttt  40989
57201   cttccac g   agctgctgca  ggcttagtgc  ctggtgggcc  aggctttggc  ccgggagtag  ttggtgtccc  aggagctggc  40909
57281   gttccaggtg  ttggtgtccc  aggagctggg  attccagttg  tcccaggtgc  tgggatccca  ggtgctgcgg  ttccag ga   40829
57361   gctgggctgt  gtgtgtgtgt  gtgtgtgtgt  gtgtgtgtgt  gtgtgtatta  gagagaaata  ttgagactat  tgccaaaatt  40749
                    ss8153909 -/GT rs5884930            G/C WASEEM                 A/C WASEEM
57441   tttgcattct  ccctaacacc  ataaccatct  gcccataccc  ttgaccacgt  ctcatcccct  catcttctct  tccttgggct  40669
         <<<<<<ELN Ex18 R1<<<<<<
57521   ataccaatct  ccttattagc  ttctaatcag  tatcatattt  tccaattgac  cttctggcta  tcagtgttct  ctgggggca   40589
57601   ggaaccgtgt  ctttttcaac  tccatgttcc  tagcccttag  ctgagtaggt  gttcagttta  tggtggataa  aacggtaagt  40509
57681   gggtggatag  atggataagt  ggatgaatgg  gtgggtggat  gaatgaatgg  atggatagat  gggtgggtgg  atggatggat  40429
57761   gggtggatgg  gtgggtggat  ggatggatgg  gtggatgggt  gggtggatgg  atggatggat  ggatgaataa  gtggatggat  40349
57841   gaatgggtgg  atagatgggt  aggtgagtgg  atgtgtgggt  ggatgggtgt  gggatggata  ggtgagtgga  tggatggggg  40269
57921   catggatgga  tgggaggatg  gatggaagaa  tggatgaatg  ggtagatgga  tgagggatgg  atggatggga  ggctgaatga  40189
58001   aagaacgggt  ggatgaatgg  gtagatgggt  ggatgaatga  gcacatggtt  ggaggggcag  atgaatagac  gggtgggtgg  40109
58081   aatggtgggc  aagcaaatgc  aaaatggatg  gttggctaga  tggttagatg  aatggataga  taggcagatg  ggataagttg  40029
58161   gcagatagat  gagtggacag  agagttaggt  ggttgggtgg  gtggattgat  atccaaggat  agacaaatgg  acagccggga  39949
58241   gcagtggctc  acgcctataa  tctcagcact  ttgggaggcc  aaggtgggcg  gatcacctga  ggtgaggagt  tcgagaccag  39869
58321   cctggccaat  atggtgaaat  cccatctcca  ctaaaaatag  aaaaaaatta  gccaggcatg  gtggtgggtg  cctgtaatcc  39789
58401   tagctacttg  ggaggctgag  gcaggagaag  tgcttgaacc  ggggaggcgg  agattgcagt  gagctgagat  cgcgctattg  39709
58481   cactccagcc  tgggtgacaa  gattgagact  ccatctcaaa  aaaaaaaaaa  aagacaaatg  gacaggtata  gaggtgggtc  39629
58561   attaggtaga  tggatgatgg  gggtggctgg  gtatacagat  gggcaggtgg  gtggacatca  gtgcataaat  ggatgtgtag  39549
```

## Exon 19 (X) (18aa)

```
58641   ccaactctat  gttggcatga  aaggagatgg  cccaacacac  agatgggtag  acagagggat  acatactaca  cagctctcct  39469
                    >>>>>ELN Ex17 F1>>>>>>>>>
58721   ccaatctctc  ctgagcattt  gtgtcccttt  tggtctctcc   gggttgtg  tcaccagaag  cagctgct    gcagctgca  39389
58801      gcagcc   tacg  ga    gtgctatgct  gacagctctg  ccccaccctg  tcctggcctt  tacttgccag  aactaaagga  39309
58881   ccctcctcta  cttgcccaga  gaagggaagt  gacttgccca  aggtcaccga  gcaagtcacc  agcaggcctc  aggacaatgt  39229
                  <<<<ELN Ex17 R1<<<<<<<<<
```

ss3195004 C/T rs2239691
ss4044366 T/C

```
58961   ctcccccatt tgtctcccac cacagggcca tggggctgag tggcgggaaa gtcccaggat ggcattccca gtggggacag 39149
59041   tgaccctgag cttccctgct ctggccaagg ccctctcaga ggagcccaaa actgcctggg gatgtggatt ctgccaatag 39069
59121   tctctctgca tccaacaaag ggggtcttcc cgaagtgctc agagaggaga ggggccagag gaggactgaa gagtgtcagt 38989
59201   aaagggctgg gtgcagtggc ttacacctgt aatcccagca atttgggaga ccaaggtagg aggattgctt gaggccagga 38909
59281   attagagacc agcatgggca aaatagcaag accctgtctc tacataaaat gcaaaaatta gctgggcata gtggcatgtg 38829
59361   cttgtttgtt ccagctactt gggaggtcaa ggcaggagga ttgcttgagg ccaggagttt gagaccagcc tgggcaacat 38749
59441   agcaagaccc cataaataaa ataaaaataa aaataaaaaa taaaataaat aaaaaataaa tgaaaaattt taattttttt 38669
59521   ttctttttct ttttcttttt ttctgagata gagtctcact ctgtcgccca ggctggagtg cagtggcatg attttggctc 38589
59601   actgcaatct ctgcatccca ggttcaagca attctcatgc atcagcctcc caagaagctg ggattacaga catgcaccac 38509
59681   catgcctggc taatttttgt attttcagta gagacagggt tttgccttgt tggccaggct ggtctcgaac tcctggactc 38429
59761   aaatgatcca cctgcctcgt ggatcccaaa gtgctgggat tatgggcatg agccactgca cccggccaaa aaaaagaaat 38349
59841   tttttttttt tgagacggag tcttgctctg tcacccaggc tggagtgcag tggcgtgatc tccgctcact gcaagctcca 38269
59921   ccttctgggt tcaagtgatt ctcctacctc agcctcccat ataactggga ttacaggtgc ccgccaccac gcccggctaa 38189
60001   tttttgtatt tttagtagag atggggtttc accatattgg ccaggctggt ctcgaactcc tgacctcagg tgatccacct 38109
60081   gcctcagcct cccaaagtgc taggattaca ggcatgagcc actgcacccg gccccacatt ttttttaaag gcttcatggt 38029
```

## Exon 20 (H) (54aa)

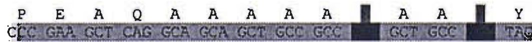
```
G   A   R   P   G   V   G   V   G   G   I   P   T   Y   G   V   G   A   G   G   F   P   G   F   G   V   G
GGG GCC AGG CCC GGA GTC GGA GTT GGA GGC ATT CCT ACT TAC GGG GTT GGA GCT GGG GGC TTT CCC GGC TTT GGT GTC GGA
                                                        Try
                                                        TGG
V   G   G   I   P   G   V   A   G   V   P   G   V   G   G   V   P   G   V   G   G   V   P   G   V   G   I
GTC GGA GGT ATC CCT GGA GTC GCA GGT GTC CCT GGT GTC GGA GGT GTT CCC GGA GTC GGA GGT GTC CCG GGA GTT GGC ATT
                                            S
                                            ATG  MAYADA TASSABEHJI 1997 and NCBI
                                                     S
                                                     TCC
```

```
60161   ggaggtgctg gggacccagg catcccagtt ttctgtcttt tatggacaag gcctggggga aatttacatc ctctttccca 37949
                                                                              >>>>>>>ELN Ex16 F1
60241   atccatcagc atccctcaga gcccgcccag cctctctcac tgagcttctt ttctacttgg ctcccttccc tctgc  ggg 37869
60321   ccaggcccgg agtcggagtt ggaggcattc ctacttacgg ggttggagct gggggctttc ccggctttgg tgtcggagtc 37789
                                   ss3176453 G/T rs2229427
60401   ggaggtatcc ctggagtcgc aggtgtccct ggtgtcggag gtgttcccgg agtcggaggt gtcccgggag ttggcatttc 37709
                        ss2984889 G/A
                        ss3186370 G/A  rs2071307
                        ss4044367 A/G
```

```
60481     cc gagcct  tagtcacacc  tggggacatg  ggttgagaag  ggatgggggc  ttcttgtctg  ctcggctctg  caggggcagt  37629
```

**ss4044368** **T/C** rs2856728 <<<<*ELN* **Ex16 R1**<<<<<<<

## Exon = 21 (X) (14aa)



```
60561     ggggactgta  gatcgggctt  gaatgtgctc  agggaggagt  tgggggagaa  gaagggaggt  cgtatccatg  ccttacaggg  37549
```

>>>>>>*ELN* **Ex15 F1**>>>>>>>

```
60641     cagaagagct  ttaaacacgg  ctcggaggag  acccaggcac  ggcttctgag  ggtctctttc  tttctcgttt  ccttgt cc  37469
60721     gaagctcagg  cagcagctgc  cgcc  gct  gcc  tacg   aagtgccc  ctgccctgcc  tgtccccaag  tcctgctctc  37389
60801     ccgcggggct  cagggtccaa  cctcagggca  aactggctcc  caggcctcca  ggactgaaat  ggcctggacc  aaggtcacga  37309
```

<<<<<*ELN* **Ex15 R1**<<<<<<<

```
60881     ggcccctgcc  ccatcttcca  aagccacact  ccactcttac  tgtccttttc  agatgccccc  tgttaggact  gttggtcact  37229
60961     ggctcatcac  cccaccccccc  acccccagaa  ttgaaggtgt  ctggcagggt  tgggtggctg  atgcctgtaa  tcccagcact  37149
61041     ttgggaggcc  aaggcaggtg  gatcccttga  ggccaggagt  tcgagaccag  cctggccaat  atggtgaaat  gctgtctgta  37069
61121     ctaaaaatac  aaaaattagc  tgggcatggt  ggcgggcacc  tatagtccca  gctacttggg  aggctgaggc  aggagaatgg  36989
61201     cttgaacccg  ggaggtagag  gctgcagtga  gcagagatta  taccattgca  ctccagcctg  ggtgacagag  caagactcca  36909
```

## Exon 22 (29aa) (H,A)

```
 G   A   A   G   A   G   V   L   G   G   L   V   P   G   P   Q   A   A   V   P   G
GGT GCT GCA GGA GCA GGA GTG CTG GGT GGG CTA GTG CCA GGT CCC CAG GCG GCA GTC CCA GGT
                                                                    P
                                                                   CCA

 V   P   G   T   G   G   V   P
GTG CCG GGC ACG GGA GGA GTG CCA
```

```
61281     actctaaaaa  taaaaaaaaa  agaaaaagaa  aaagaattga  aggtgccagg  aagccattct  cttcttcctc  cgattctccc  36829
```

>>>>>>*ELN* **Ex14 F1**>>>>>>>

61361 acccaccctt gctcccccaa aaagtgagta ctgggagggg caaggctgaa agttctccac tccccg gt gctgcaggag 36749
61441 caggagtgct gggtgggcta gtgccaggtg ccccaggcgc agtcccaggt gtgccgggca cgggaggagt gccag gag 36669
G/A **HIDEAKI ONDA et al 2001** **ss5996510** G/C rs4464848

61521 ctgtgtctcc agcccagaga tgggtttggt ttgtctcatg gaagggtccc tggagttgac caaggtcgac tgcaccattt 36589
t
<<<<<***ELN*** **Ex14 R1**<<<<

# Exon 23 (X,A) (19aa)

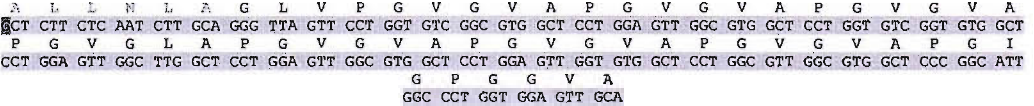
G V G T P A A A A A A A A A Q F
GGA GTG GGG ACC CCA GCA GCT GCA GCT GCT GCA GCC GCC GCC GCC CAG TTT

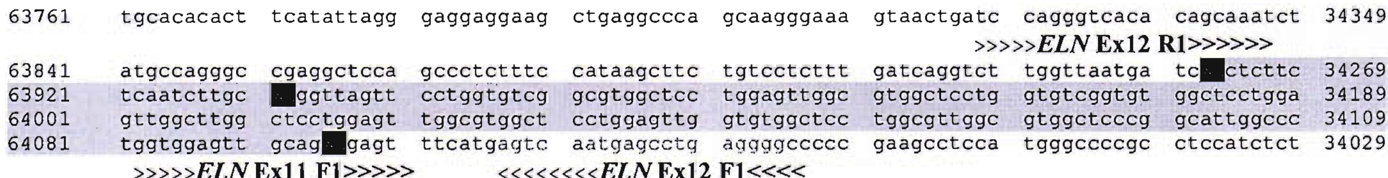
aaggtcgac tgcaccattt
>>>>***ELN*** **Ex13F 1**>>>>

61601 tacaaatggg aagactgagc ctagagatgg gaagcaggga ggggtgtgag agattactct ctcacccctt ctcttcacac 36509
61681 ctcc gagt ggggacccca gcagctgcag ctgct gc agccgcc gccgcccagt ttg aagtc cccctcaccc 36429
61761 ccgccactgg ctcacggaga actgctttct cctgtgccct gctctggggt ctgaccgccc agcttcctgt tcctttccac 36349
<<<<<***ELN*** **Ex13 R1**<<<<<<<
T/C **HIDEAKI ONDA et al 2001**
61841 cccacttaag ctgtcacatt ctggggtggg ccctccttag accttttggc ccactgatga atgacctcta ggagtgtggg 36269
61921 tgatgtttct gattagggga gcagggtgag cagtgtgagc ctccctgttc ctaaagcccc tggtgcctcc caggctattg 36189
62001 gggacctgac ctcatgctga gctccagctc cccttgaggg actccagttc tccccctctt ccttctcttt ctccttctcc 36109
62081 ttcttcttct tgtgctcctc ttccttcttc ttcttcttct ttcttctcct cctcctcctt ctccttctcc ctcctcctcc 36029
62161 ttctcattct tcttcttctc cttcctcttc ttctcctcct gcttcttttc cttctcctcc ctcttcctcc tcctcctgct 35949
62241 tctcctttct tctccttctt cttctttctt cttcttcttc ctctttttct ccttctttct tcttctttct ccttcttctt 35869
62321 cttcttcctc tgcttctttc tcctccttct tctcttctct tccgtctttc ttctcctcct tttcctcctt cttctttctt 35789
62401 ctccttctcc ttttttcttg agatagggtc tagctctgtc acccaggatg gggtacagtg ccacaatcat agctcactac 35709
62481 agcctcaacc tcccagtctc aagcagtctg cctgcttccg ccccccaaga gctgagacca caggtgccca ccaccatgcc 35629
62561 tggctaattt tttaattttt ttgtagcgac agcggtctca ctatgttgct caggctggtc tcaaactcct aggctaaagc 35549
62641 gatcctcctg cctctgcctc ccaaagtcct gggattacag gcgtgagcca ccacacccgg cctgcagtac ttcttgttcc 35469
62721 ccatctcttg ctacatttga gggccacccct ggcagcccca ggtgcccaca cttttctgaa catggcaaat cgtggcagca 35389
62801 ccaattgtag agctcaactg tatgtcaggc cctgggcatg gggtctgtag gccttggccc tagggacctg tgggctgaaa 35309
62881 ggttcagatc agaatctcta ggactgaata ggccagagag catttcgact gcaggtctgc tgagccccat attctcacac 35229
62961 acagcaatct ttattattta tttatttgag acggggtctc acattgtcgc ccaggttgga gtgcagtgat gccgtcagct 35149

```
63041   cgccgcaggc  ctcaaactcc  tagcttaagc  cattattccc  ccttagtatc  cctagtagct  ggggctacag  tcacatgcca  35069
63121   ccatgcccag  tttaaaaaaa  aaaaaaattg  tatgcgatcc  tcccacattg  gcctctcaaa  gtgctgggat  gacaggcatg  34989
63201   agccaacgtg  tctggcctac  aaaaatctta  cagagttgat  tttatttttc  ccattttaca  gatgtggaaa  ctgaggttcc  34909
          ss3544768 A/G rs2528795
63281   cagagcttaa  gtaacttgcc  tacagttgca  cagctaaatg  gtggctgagc  tgagatttga  acccaaagcc  tttctgtctt  34829
63361   acaaagtccc  ttatataatg  taaatctgcc  tccatcagcc  tcaaatctcc  aaggggtcct  tgtcactgaa  aaggttaaga  34749
63441   actcctggcc  aaatgcagca  gctcacaact  ataatcccag  aactttggga  ggccaagtcg  ggtggatcac  ccaaggtcag  34669
63521   gagtttaaga  ccagcctggc  caacatggtg  aaaccctgtc  tctactaaaa  atacaaaaaa  attagccggg  catggtggtg  34589
63601   cgcacctgta  gtcccagcta  ctcaggaggc  tgaggcagga  gactcacttg  aactcgggag  gtggtggttg  cagtgagtcg  34509
63681   agatcacgcc  attgcactcc  agcctgggcg  atagagtgag  actctgtctc  caaaaaaaca  aagttatgaa  ctcctgagcc  34429
```

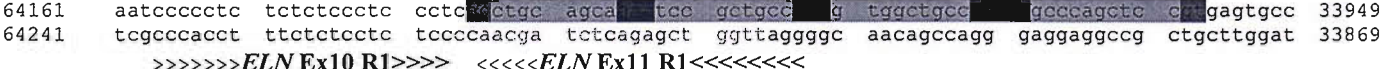
## Exon 24 (H,A) (54aa/60aa)



```
63761   tgcacacact  tcatattagg  gaggaggaag  ctgaggccca  gcaagggaaa  gtaactgatc  cagggtcaca  cagcaaatct  34349
                                                                               >>>>>ELN Ex12 R1>>>>>>
63841   atgccagggc  cgaggctcca  gccctctttc  cataagcttc  tgtcctcttt  gatcaggtct  tggttaatga  tc ctcttc  34269
63921   tcaatcttgc   ggttagtt  cctggtgtcg  gcgtggctcc  tggagttggc  gtggctcctg  gtgtcggtgt  ggctcctgga  34189
64001   gttggcttgg  ctcctggagt  tggcgtggct  cctggagttg  gtgtggctcc  tggcgttggc  gtggctcccg  gcattggccc  34109
64081   tggtggagtt  gcag gagt  ttcatgagtc  aatgagcctg  aggggccccc  gaagcctcca  tgggccccgc  ctccatctct  34029
         >>>>>ELN Ex11 F1>>>>>      <<<<<<<<ELN Ex12 F1<<<<
```

## Exon 25 X (15aa)




## Exon 26 (A,H) (75aa/ )

```
 R   A   A   A   G   L   G   A   G   I   P   G   L   G   V   G   V   G   V   P   G   L   G   V   G   A   G
GGA GCT GCA GCT GGG CTT GGT GCT GGC ATC CCT GGA CTT GGA GTT GGT GTC GGC GTC CCT GGA CTT GGA GTT GGT GCT GGT
 V   P   G   L   G   V   G   A   G   V   P   G   F   G   A   G   A   D   E   G   V   R   R   S   L   S   P
GTT CCT GGA CTT GGA GTT GGT GCT GGT GTT CCT GGC TTC GGG GCA GGT GCA GAT GAG GGA GTT AGG CGG AGC CTG TCC CCT
             E   L   R   E   G   D   P   S   S   S   Q   H   L   P   S   T   P   S   S   P   R
            GAG CTC AGG GAA GGA GAT CCC TCC TCC TCT CAG CAC CTC CCC AGC ACC CCC TCA TCA CCC AGG
```

```
64321   ctgggcccct  tcctctggga  ctaggctcag  ctccctgggc  aggacacctc  cttaggggca  tgctccctgc  ctgctgtcgc  33789
64401   caccactgcc  ctctgtctgc    gagctgca  gctgggcttg  gtgctggcat  ccctggactt  ggagttggtg  tcggcgtccc  33709
64481   tggacttgga  gttggtgctg  gtgttcctgg  acttggagtt  ggtgctggtg  ttcctggctt  cggggcag    gcagatgagg  33629
64561   gagttaggcg  gagcctgtcc  cctgagctca  gggaaggaga  tccctcctcc  tctcagcacc  tccccagcac  cccctcatca  33549
64641   cccaggg  g  cata   aaa  tccttgttag  gttctcctaa  gcatctgggg  gtaacagata  gcgggaggag  ggcagaccag  33469
64721   gccaagggac  cccaggctgc  ccaattgcag  agtatctgcc  tcctcagtag  aggggtggca  gggctccaga  ctgagaaaaa  33389
                                                                                 <<<<ELN Ex10 F1<<<<<
64801   gcctgccctc  caaccaggga  aggagcaggt  agatcagcct  ggctcccctt  cagcagtgct  ctgtggccag  gggaccacta  33309
```

## Exon 27 (X) (13aa)

```
64881    ggaactccag  ttcttcacca  gctggtgact  gagccccact  ctgtgcccag  cctcaggcag  tccatgctag  gcctggggag  33229
64961    agaaggaagg  ggcagacgga  aggtggaggc  actgttcagc  cctaaagctc  tgtgcctgta  cagcttcagg  gctttgagga  33149
>>>>>>ELN Ex 9 R1>>>>>>>>>>>
65041    agcaatagag  gccaaggaag  tcagggaggg  ctctctagag  gaggcggcag  aactcccagg  cacagagctc  ggctcctgac  33069
65121    cactccccaa  cttttctttc  tccccagtac  ctggagccct  ggctgccgct  [illegible]gcagcc[illegible]  tatg[illegible]ga  gtgcacccca  32989
            C/T HIDEAKI ONDA et al 2001
65201    caaccacttg  tggctccctt  gccaccacac  catccctgac  agcacaggac  tggggtggtc  acaatagaaa  aaggcagttt  32909
                                                                 <<<<ELN Ex9 F1<<<<<<<
65281    cggccgggcg  tggtggctca  cacctgtaat  cccagcactt  tgggaggccg  aggcaggcag  atcacctgag  gtcaggagtt  32829
65361    ccagaccagc  ctgaacaata  tggttaaacc  ccatctctac  taaaaattca  aaaattagcc  aggcgtggtg  gcgggcacct  32749
65441    gtaatcccag  cactttggga  ggccgaggtg  ggcagatcaa  ctgaggtcgg  gggttccaga  ccagcctgac  caatatggtg  32669
65521    aaaccccatc  tctactaaaa  acacaaaaat  tagccaggtg  tggtggcggg  cacctgtaat  cccagctact  cgggaggctg  32589
65601    aggcaggaga  atcgcttgaa  cccaggagat  ggaggttgca  gtgagccgag  atcacgccac  tgcactccgg  cctggacgac  32509
65681    agagcgagac  tgtatctcaa  aaaaagaaaa  gaaaagaaaa  agaaaaaggc  agtttctagg  acacgtttat  gacagtttaa  32429
65761    aaacctggcc  cctgcccact  aaatgcttat  ggtgccttca  acccctgtga  ccaccaaaaa  caccccttgaa  atcccagttg  32349
65841    ccccccagga  ggcaattcca  ccatccctaa  gcttgccctg  accctgacag  ttacatggtc  cctgtgtcca  ggaagggact  32269
65921    gggcctgctg  tgggtatgag  gagtctgggc  agtctctgcc  tccaccccag  ctctctggcc  cgagcaaagg  agctggtatt  32189
                             ss1055286 A/G              ss1055285 G/A
                             ss2068364 A/G rs810556 ss2068363 G/Ars810555
                             ss4042979 C/T              ss3544767 G/A

66001    cctcatctgg  gcaccccggg  cccccgcccg  ccttcctgtc  ctgccagcag  gaccgctccc  accacgagcc  agcccagagg  32109
66081    acgtggcagt  cccacagcct  ctgcacttgg  cgctctggca  gccctgcctg  gccagaccca  cgctgtctgt  agcccctgaa  32029
66161    ttcctggcct  cctgtgagca  gcgttggtgg  gaggtggaag  actcccccag  ggatccacca  accagccggg  caggtccact  31949
66241    tggggccagt  ttctacctcg  gaaaatcctg  gcttcctcag  cgccgacccc  cagcagccct  ggcctcctcc  cctgctcact  31869
66321    gctgactgct  tggcaggcct  gagtcaacag  acattgccag  ggaaacagag  tgacgtcatg  tggtggagcc  gcgaagggcg  31789
66401    caggctcccc  agccacttcc  cgtccctgaa  gccctgagca  cactgcttaa  cctctgctct  gcctcagcgt  tctcatctgc  31709
66481    aaaatgggaa  tgacaatagt  gcctaccttt  cctgctgcaa  tgcataaaaa  atgggctgct  aagtgtgaag  cactcaaaat  31629
66561    gttaagtgcc  tgctgtagat  actattaccg  ttgtttgttt  gtttgtttgt  ttgttttgac  agagtcttgc  tctgtcgccc  31549
66641    aggctggaat  gcagtggtac  gatcttggct  ccctgcagcc  tccatctccc  gggttcaagt  gattctcctg  cctcagcctt  31469
66721    ccaagtagct  gggattacag  gcgcccgcca  ccacgcctgg  ctaattttg  tatttttcgt  agaaacaggg  tttcaccgtg  31389
66801    ttggccaggc  tggtctcaaa  ctcctgacct  caggtgatcc  actagcctca  gcctcccaaa  gtgctgggat  tacaggtgtg  31309
66881    agccaccgca  cccggcttac  aaaagaactt  ttaaggccag  gcacagtggt  tcacaccttc  gggaggctaa  ggcaggagga  31229
66961    tcgcatgagc  ccaggagttg  gaggctgcag  tgaactatga  ttgtaccact  gcactccagc  cggggtgaca  gagcaaaacc  31149
```

## Exon 28 (H) (24aa)

```
 G   A   A   V   P   G   V   L   G   G   L   G   A   L   G   G   V   G   I   P   G   G   V   V
GGA GCA GCA GTG CCT GGG GTC CTT GGA GGG CTC GGG GCT CTC GGT GGA GTA GGC ATC CCA GGC GGT GTG GTG
```

```
67041    ccatctcaaa  atgaaacaaa  atatggactg  gacttcctgt  ccactgctcc  tccacagtgt  cacatggccc  ctgccacctg  31069
                                                                                              >>ELN Ex6-8 F1>>>>
67121    tctgcttgcc  ttgtgtccct  ggggcaggga  gacccatcgt  tcagaaatgg  aacactcatt  ttccctcctc  tccccgc  g  30989
67201    agcagcagtg  cctggggtcc  ttggagggct  cggggctctc  ggtggagtag  gcatcccagg  cggtgtggtg  g  gagttga  30909
```

## Exon 29 (X) (20aa)

```
 G   A   G   P   A   A   A   A   A   A   A       A   A   A       A   A   Q   F
GGA GCC GGA CCC GCC GCC GCC GCT GCC GCA GCC     GCT GCT GCC     GCC GCC CAG TTT
```

```
67281    aaccccagga  ggggcagggt  ggggagggaa  tctaaccagt  acagagtgcc  tccctgaact  cggtctgtgt  tccc  gagc  30829
67361    cggacccgcc  gccgccgctg  ccgcagcc    gctgctgcc     gccgccc  agtttg  ga  gcactgggtg  gaggtgggag  30749
67441    ctgccgccag  gcccccaggc  ccccaggtg  tgggaggagc  ttctgaccag  gcactgtaga  ctcagagtcc  ctgccccaga  30669
                                   <<<<<<<ELN Ex6-8 R1<<<<<
```

## Exon 30 (H) (25aa)

```
 G   L   V   G   A   A   G   L   G   G   L   G   V   G   G   L   G   V   P   G   V   G   G   L   G
GGC CTA GTG GGA GCC GCT GGG CTC GGA GGA CTC GGA GTC GGA GGG CTT GGA GTT CCA GGT GTT GGG GGC CTT GGA
```

```
67521    cacctcctgg  ctccactgtg  ccatcgaagg  ccaggggaga  cctcaggctc  cacctgtgtc  cccagaggac  acctccgccc  30589
                                                   >>>>ELN Ex5  F1>>>>>>>>>
67601    tccacaggcc  gaggcttcag  tcccaccttt  ctgaccagcg  gagtctaatg  ctcagctgtc  tccac  gcc  tagtgggagc  30509
67681    cgctgggctc  ggaggactcg  gagtcggagg  gcttggagtt  ccaggtgttg  ggggccttgg  ag  gagagt  tgttctgaaa  30429
                                                                       CC WASEEM
67761    tcagtgagtg  tgtgtggtgt  gtgtatgcga  gacagagatg  gagacagaga  cagagacaga  gactttcgtt  cccacccctg  30349
                   <<<<ELN Ex5 R1<<<<<<<<<<<
67841    gcacgtcttt  gctcaagatg  tcctcttggc  caggtgcggt  ggctcacgcc  tgcaatccca  gcactctagt  aggccaaggt  30269
67921    gggcggatca  ctagaggtga  ggagtttgag  accagcctgg  gtgacatggc  gaaacctcat  ctctaccaaa  aatacaaaaa  30189
68001    taagccgggc  gtggtggtgg  gcacctgtat  ttccagctac  ttgagaggct  gaggccagag  gatcgcttga  gcccaggagg  30109
68081    cagaggctgc  agtgagctga  gatggtacca  ctgcattcca  gcttgggcag  cagagtgaga  ccctgtcatc  taaaaaaaaa  30029
68161    aaagaaagaa  agaaaagaaa  agaggccagg  catggtggct  cacgcctgtg  atcccagcac  tttgggaggc  tgaggtgggc  29949
68241    agatcacgag  gtcagatcaa  gaccatcctg  gctaacacag  tgaaaccccg  tctctactaa  aaatacaaaa  aattagctgg  29869
68321    gcgtggtggc  agacgcctgt  agtcccagct  actcaggagg  ctgaggcaga  agaatggcgt  gaacctggga  ggcggagctt  29789
68401    gcagtgagcc  gagatcgcgc  cactgcactc  cagcctgggc  gacagagcga  gactccgtct  caaaaaaaaa  atatatatat  29709
                                                                                   ss1055284 T/A  rs810554
```

```
                                                                                  ss2068362 T/A
68481   atatatatat  atgtatgtat  gtatgtatat  aaattagctg  ggcatggtga  cacacgcctg  taatcccagc  tactcgggag  29629
68561   cctgagacag  gagaatcact  tgaacccagg  aggcagaagt  tgcagtgagc  caagatcacg  ccgctgcact  ccagcctgag  29549
68641   caacagagca  agacctcatc  ccaaaacaaa  ataaaaacaa  gaaatgtggc  aaacgggagc  tggggctatc  ccttgccgcc  29469
68721   agggtcccag  tcctccccct  ctcctctctc  tttcaaccca  cctgaccact  gcggtgggag  taaattagga  gaatgatggt  29389
68801   tatgccagtt  ctccaagaaa  gaaatcaaag  cttggagaga  aattgctttc  cctaagtctc  caccatggtg  gccgaccttg  29309
68881   tgtggccatt  tccccctttc  ttgggctgtc  attgtgtctc  cccctgtaaa  ctagcctgca  cagctcttcc  tgtgagtgcc  29229
68961   acaagaccag  acacacagta  ggagcttaac  aattaagact  tttggggccg  ggcatggtgg  ctcatgcctg  taatcccaac  29149
69041   actttgggag  gcttaggtag  gtggattgct  tgaggccagg  agttcaagac  cagcctggac  aatgtagcaa  gacccccccc  29069
69121   caccaccacc  tctacgaaat  atttaaaaat  tagctgggcg  tggtgtgtgc  ctgtagtttc  agctacttgg  gaggctgagg  28989
69201   caggaagatc  acttgaaccc  gggagttcag  ggctgcagtg  agctgtgatc  aagccactgc  attccagcct  gggcaacaga  28909
69281   gcaagaagct  ggatagatga  aggagagatg  gggggataga  ctgggaggat  gagtccatgg  gcagacggta  aagggtgagt  28829
69361   aggtcagaag  acagggccta  gccaggtgca  gtggctcaca  cctgtaatgt  cagcactttg  ggaggccgag  gcaggtggat  28749
69441   cacctgaggt  caggagttcg  cgaccagcct  ggccaacatg  gtgaaacctg  tctctactaa  aaatacaaaa  attagccagg  28669
69521   cgtggtggtg  ggtgcctgta  atcccagctg  ctcaggaggc  tgaggcagga  gaaccgcttg  aacccaggag  atggaggttg  28589
```

## Exon 31 (X)(3)(13aa)

```
                          G   I   P   P   A   A   A   A       A   A       Y
                          GGT ATA CCT CCA GCT GCA GCC GCT ... GCA GCT ... TAC

69601   cagtgagccg  ggatcgcgcc  actgcactcc  agcccaggcg  aaggagtgag  actctgcctc  aagaaaaaaa  aaaaaaaaaa  28509
                                              >>>>>ELN Ex4 F1>>>>>>>>>
69681   agacagggcc  tgacaggtgg  cattggcatt  cctgagccgt  catgtgcctc  atctcccc    gtatacctcc  agctgcagcc  28429
69761   gct…gcag    ct…tacg…    gagttcccc   tctgatgcct  tcctgccagt  ggcctgcacc  ccctgccatg  cccatcgcca  28349
69841   ccctccccca  gcccagctca  ggcctccctc  tggctccccc  taccccctgaa gatcttgtct  gggacattcc  ttaacccaga  28269
                                              <<<<<ELN Ex4 R1<<<<<<<<<
```

## Exon 32 (H,A)(18aa)

```
                    G   A   A   G   L   G   G   V   L   G   G   A   G   Q   F   P   L   G
                    GGT GCT GCT GGC CTT GGA GGT GTC CTA GGG GGT GCC GGG CAG TTC CCA CTT GGA

69921   acccagcagg  gatatcaggg  cctcttcccg  atgggggtgt  cttatcctga  ccccacctgc  ctcttctc    gtgctgctgg  28189
                  >>>>>>ELN Ex3 F1>>>>>>>>>>
70001   ccttggaggt  gtcctagggg  gtgccgggca  gttcccactt  ggag…aggg   gtggccagct  ctgctacgta  gtcctcagct  28109
70081   ctgtcccgat  ctagaggggg  cctgtccatc  tagcagtggg  gactcccaga  gcccatgtcc  acacaaggac  aggagactgg  28029
```

```
                                <<<<<ELN Ex3 R1<<<<<<<<<
70161   ggctggtgag  ggcactttag  gattgcagat  gtactgggca  aacgggcaag  ctataaggtt  ggggacagtg  ccaggctgag  27949
70241   ctcagggttg  aggccatgag  gtgccacacc  tggttgctag  gtggcggcat  gttgtgttga  gaaagccact  ctggctgtag  27869
70321   tgaggggat   tggctgggcg  tggtggctca  cgcctgtaat  cccagcactt  tgggaggcct  aggtgggtgg  atcacttgag  27789
70401   gtcaggagtt  cgagaccagt  ctggtcaaca  tggtgaaacc  ctgtctctac  taaaaaaaat  gcaaaaatta  gccaaacgtg  27709
70481   gtggacgcct  gtaatcccag  ctactcggga  ggctgaggcg  ggagaatcac  tggagcctgg  gaagcggagg  ttgcagtgag  27629
                                          ss3544766 C/T rs2528794
70561   ccaagatcgc  accactgcac  tccagcctgg  gtgacagagc  aagaccccat  ctcaaaaaaa  taataataaa  ataaaatata  27549
```

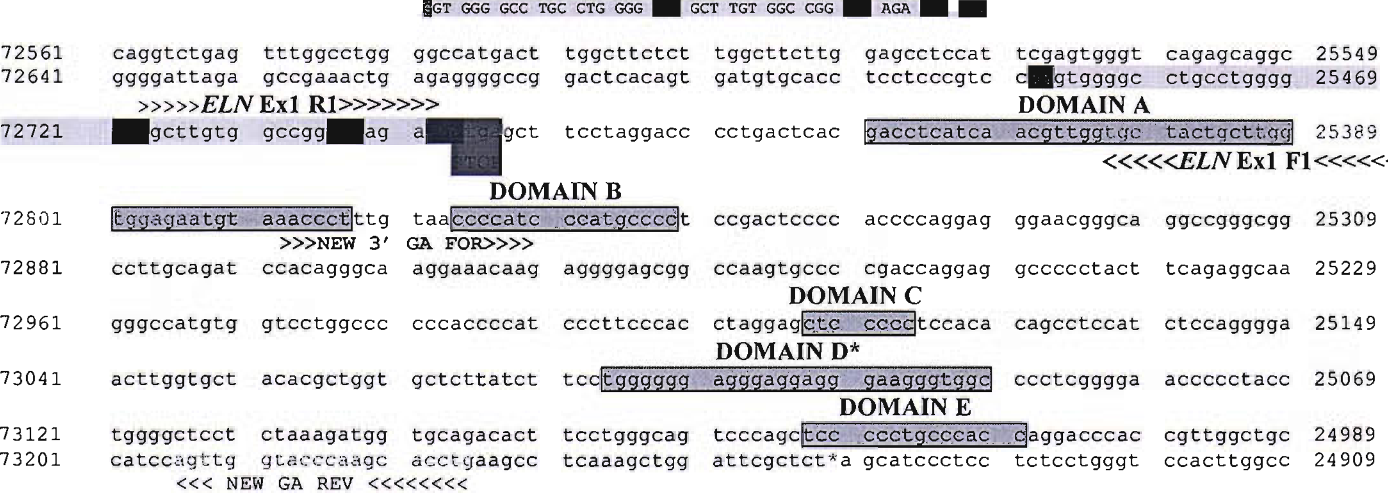
## Exon 33 (H,A) (15aa)

```
                              G   V   A   A   R   P   G   F   G   L   S   P   I   F   P
                              GGA GTG GCA GCA AGA CCT GGC TTC GGA TTG TCT CCC ATT TTC CCA

70641   aaaaattata  tagtgggggg  gatggagagg  aggtgatccc  agacagaggt  cttgggtgag  ccagtgcagg  cagaaagtga  27469
                                                                            >>>>ELN Ex2 F1>>>>>>>>>
70721   tgaggctgga  gtcagtttcc  acccctacca  acccaccaac  ctgaaatctc  tcctgc[illegible]ga  gtggcagcaa  gacctggctt  27389
                   HIDEAKI ONDA et al 2001 C/T
                     ss4943619 C/T
                     ss5374901 A/G
                     ss5375213 A/G rs3757587
                     ss1508519 C/T
                     ss6296979 T/C
70801   cggattgtct  cccattttcc  cag[illegible]atgcc  aggctccctg  cccctgggcc  ctgccctgga  gctgcagcca  cctcctccct  27309
70881   cctctcctgt  gccatctcct  gctcagaagg  gctgagccag  cacccagggg  tggaccccac  agcctcaggt  cacacgaggc  27229
<<<<ELN Ex2 R1<<<<<<<<<<<<<<
                  ss8153910 A/- rs5884931
70961   tggaccccga  gctgaatgta  gagcctcccc  tccttcttgc  tgacctctta  taaacacagg  gaacatttgc  tttttaaaaa  27149
71041   cacagttcat  ggccaggtgt  gatggctcac  acctgtaatc  ccagcacttt  gggaggctga  ggagggcaga  tcacctgagg  27069
71121   tcaggagttc  aagaccagcc  tggccaacat  ggcgaaaccc  tgtctctatt  aaaaatacta  aaattagccg  ggtgtggtgg  26989
71201   ctcacgcctg  taatcccagc  cctttgggag  gccaaggcag  gtggatcacg  aggtcaggag  atggaaacca  tcctggctga  26909
71281   aacggtgaaa  ccccgtctct  actaaaaata  ctatagtctt  cttttttttt  ttgagatgga  gtcttgctct  gttgcccagg  26829
```

71361 ctggagtgca gtggcacgat cttggctcac tgcaacctct gcctcctagg ttcaagcgat tcttctgcct cagcctccca 26749
71441 agtagctggg actacaggca tacaccacta cgcccagcta atttttgtat ttttagtaga gacggggttt caccatattg 26669
71521 gccaggctgg tctcgaactc ctgacctcgt gatccgcccg cctcggcctc ccaaagtgct gggattacag gagtgagcca 26589
71601 ccacgtccgg ccgtctctac taaaaataca aaaaattagc cgggcatggt ggcgggcacc tgtagtctca gctactcggg 26509
71681 aggctcacgc aggagaatgg tgtgaacccg ggaggcggag catggtgtga acccgggagg cggagcttgc agtgagccga 26429
71761 gattgcgcca ctgcactcca gcctgggcgg cacagcaaga ctccgtctca aaaaaaaaaa aaaaattagc caggtgtggt 26349
71841 ggtgcacgcc tgtaatccca gctactctgg aggctgagac aggagaattg cttgaaccca ggaggcggag gttgcagtga 26269
71921 gctgagacta caccactgcc ctctagcctg ggtgatgaca gggcaagacg attgcgtctc aaaaaaaaaa aaaaaaaaaa 26189
72001 aaaaaaacct tcagaggcag atgtctgttc ctgccccact ctctgctcac tgccaccagg tggcggtaag gagccatatc 26109
72081 tggctcaagc agggttagga actgctccgg gccgcgctac tagacaatgg tgccttacct aatacagact gatgccttga 26029
72161 aacagggttc acccatcaga ctattcccag gagggggtag gaatcctggg tttgagtccc agctcagaca ctatgtggat 25949
72241 aagaccctct ctccctctac tgggccttga tttacttatc tgatgccctc tcagatgtat gggtccaccc cactccctga 25869
72321 gagggccgtc tcctgcagac actgactctc ccttcatccc atgttccttg atggagctgc tcttagcctt tccccgaaag 25789
72401 gcttcttgga gcagtggagc caggagccag tgccaggccc cgtccttccc cagccaggcc ccatgacctg ccccctttc 25709
72481 tgcgagcgtt ggtttcatcc aagtaatatc ttgggggctt ctcccgcccc atctgtccag tggaagttga tggcatggat 25629

## Exon 34 (C) (14aa)

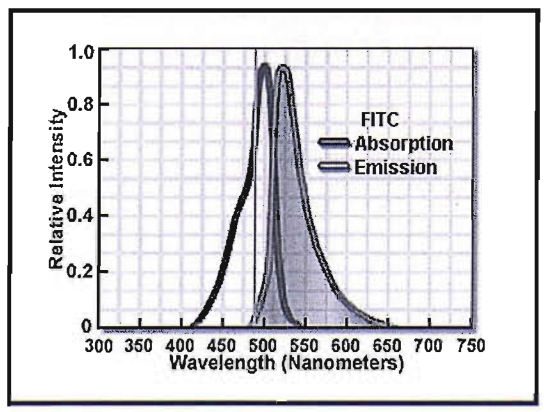


```
73281    gtctcctccc caccgatcgc tgttccccac atctggggcg cttttgggtt ggaaaaccac cccacactgg gaatagccac 24829
73361    cttgcccttg tagaatccat ccgcccatcc gtccattcat ccatcggtcc gtccatccat gtccccagtt gaccgcccgg 24749
                                                      ss10300 G/C
                                                      ss1513096 G/C
                                                      ss4391394 G/C rs8326 - and HIDEAKI ONDA et al 2001
                                                      ss4406392 G/C
73441    caccactagc tggctgggtg cacccaccat caacctggtt gacctgtcat ggccgcctgt gccctgcctc cacccccatc 24669
73521    ctacactccc ccagggcgtg cggggctgtg cagactgggg tgccaggcat ctcctcccca cccggggtgt ccccacatgc 24589
73601    agtactgtat accccccatc cctccctcgg tccactgaac ttcagagcag ttcccattcc tgccccgccc atctttttgt 24509
73681    gtctcgctgt gatagatcaa taaatatttt attttttgtc ctggatattt ggggattatt tttgattgtt gatattctct 24429
73761    tttggtttta ttgttgtggt tcattgaaaa aaaaagataa tttttttttc tgatccgggg agctgtatcc ccagtagaaa 24349
73841    aaacatttta atcactctaa tataactctg gatgaaacac accttttttt ttaataagaa aagagaatta actgcttcag 24269
73921    aaatgactaa taaatgaaaa acctttaaag gaaactgtgt cttagcttcc ttcgtatgat ttaatctgcc ttcaactcgc 24189
74001    tggcctggat ggggataagg gctctgcttc agggaacctc caccacccac agtgtatttg agaggttgcc caaccaaaag 24109
74081    cccctgctgc tggcttctgg ggcatcgatc tttctccaag tttggcttgt actttaacaa tgaggccagg agtttgagac 24029
74161    ccacctggac aacatagctg tccagtggag aatcactttg taccatatag ctacaaaaaa aaaaaaaatt acccaagatt 23949
74241    ggtggtggtg catgcctgtg gtcccagcta ctcaggaggc tgaggtggga ggatcgcttg agcccaggat ttcgaggctg 23869
74321    cagtgagcta tgattgcgcc actgcactcc agcctgggtg agacagcaag atcctgtctc gagagagaga gagagagaaa 23789
```

## Appendix F:

Adopted from

http://www.olympusfluoview.com/java/excitationefficiency/



# Appendix G:

Ali baba file for the 5' end:
Prediction of transcription factor binding sites by constructing matrices on the fly from TRANSFAC 4.0 sites.

# C>T -1050 New 5' UTR

```
==========================================================================
====
seq(    0..   59)     tgcctgtcagtctgCggggggctcccacct
Segments:
2.3.1.0    11   25                =======Sp1=====

1 segments in this sequence identified as potential binding sites
```

1 segments in complete file identified as potential binding sites

## AliBaba2.1 predicts the following sites in your sequence

### Sequence seq_259

```
==========================================================================
====
seq(    0..   59)     tgcctgtcagtctgcggggggctcccacct
Class      lbp  rbp
2.3.1.0    11   20                skGsGksrGG
2.3.1.0.2  -R03382                gggcgggggg
2.3.1.0.2  +R03612                ctgggtgggg
2.3.1.0.2  -R01680                gggcgggggg
2.3.1.0.2  -R03317                gggcggcggg
2.3.1.0.2  -R00378                gggcgggagg
                                  ||||||||||
                                  ||||||||||
                                  ||||||||||
                                  ||||||||||
                                    | |  ||

2.3.1.0    12   21                 GGsssGGGGC
2.3.1.0.2  -R00287                 ggggcggggc
2.3.1.0.2  -R01762                 ggtgggggc
2.3.1.0.2  +R03281                 ggcgggctgg
2.3.1.0.2  -R03348                 ggggcggggc
2.3.1.0.2  +R00821                 ggcgggggc
2.3.1.0.2  +R01991                 ggcgggtgac
2.3.1.0.2  +R00315                 ggggcggggc
2.3.1.0.2  +R02085                 gggcggggc
2.3.1.0.2  -R02891                 ggggcggggc
2.3.1.0.2  -R00447                 tgggcgggac
2.3.1.0.2  +R00816                 ggggcggggc
2.3.1.0.2  -R00817                 ggggcggggc
2.3.1.0.2  +R00823                 ggggcggggc
```

```
2.3.1.0.2  +R01039          ggggcggggc
2.3.1.0.2  -R01309          gggcgggggc
2.3.1.0.2  -R02446          ggggaggggc
2.3.1.0.2  +R03287          ggggcggggc
2.3.1.0.2  +R00281          ggggcggggc
2.3.1.0.2  -R02572          ggggcggggc
2.3.1.0.2  +R04034          ggggcggggc
2.3.1.0.2  -R00224          ggggcggggc
2.3.1.0.2  -R00381          tgcggggggc
2.3.1.0.2  +R00818          tgggcggggc
2.3.1.0.2  +R01307          ggggcggggc
2.3.1.0.2  -R01756          ggggaggggc
2.3.1.0.2  +R01909          gaggcggggc
2.3.1.0.2  +R01927          ggggcggggc
2.3.1.0.2  +R04007          gggggggggc
2.3.1.0.2  -R01680          gggcgggggg
2.3.1.0.2  +R03868          gaggcggggc
2.3.1.0.2  +R00149          gggagggggc
2.3.1.0.2  +R03051          ggcgggggcc
2.3.1.0.2  +R01308          ggcgcggggc
                            ||||||||||
                            ||||||||||
                            |||| |||||
                            || | |||||
                             |   ||| |

2.3.1.0    16   25              GGGGGCkGsn
2.3.1.0.2  +R03448              gggggcggcc
2.3.1.0.2  -R03348              gggggcgggg
2.3.1.0.2  -R01728              gggggcggcg
2.3.1.0.2  +R04005              gggggcggtg
2.3.1.0.2  -R02891              gggggcgggg
2.3.1.0.2  +R00826              gggggcgggc
2.3.1.0.2  +R00281              gggggcgggg
2.3.1.0.2  +R01770              gggggcggca
2.3.1.0.2  -R04124              gggggcggcc
2.3.1.0.2  -R00570              gggggctgca
2.3.1.0.2  +R03486              tggggctcct
                                ||||||||||
                                |||||||||
                                ||||||||
                                ||||||||
                                 |||||

Segments:
2.3.1.0     11   25          =======Sp1=====

1 segments in this sequence identified as potential binding sites
```

---

1 segments in complete file identified as potential binding sites

# C>G -1162 NEW 5' UTR

## AliBaba2.1 predicts the following sites in your sequence

---

**Sequence seq_260**

```
================================================================================
====
seq(    0..   59)     tcccagaggggccgggagaacagcagtcga
Segments:
2.3.1.0    6    16          =====Sp1===
2.1.1.1    12   21                =====GR===

2 segments in this sequence identified as potential binding sites
```

---

```
2 segments in complete file identified as potential binding sites
```

---

## AliBaba2.1 predicts the following sites in your sequence

---

**Sequence seq_260**

```
================================================================================
====
seq(    0..   59)     tcccagaggggccgggagaacagcagtcga
Class      lbp  rbp
2.3.1.0    6    15          rsGGGssGGG
2.3.1.0.2  +R01769          gggggcgggg
2.3.1.0.2  -R02572          gggggcgggg
2.3.1.0.2  +R01991          acggggcggg
2.3.1.0.2  -R01680          acgggcgggg
                            ||||||||||
                            ||||||||||
                            ||||||||||
                              |||  |||
                              |||  |||

2.3.1.0    7    16           ssGGsCGGsn
2.3.1.0.2  +R01704           cggggcggga
2.3.1.0.2  -R02891           gggggcgggg
2.3.1.0.2  +R00149           gggggcggga
2.3.1.0.2  -R00817           gggggcgggg
2.3.1.0.2  +R00821           gggggcgggg
2.3.1.0.2  +R00823           gggggcgggg
2.3.1.0.2  +R00826           gggggcgggc
2.3.1.0.2  -R00955           gggggcgggc
2.3.1.0.2  +R01754           gggggagggg
2.3.1.0.2  -R02435           ggggacgggc
2.3.1.0.2  +R03051           ggggccgggc
2.3.1.0.2  -R03382           cggggcgggg
2.3.1.0.2  +R04282           cggaggcgga
2.3.1.0.2  +R00281           gggggcgggg
```

```
2.3.1.0.2  +R01769          gggggcgggg
2.3.1.0.2  +R02085          gggggcgggg
2.3.1.0.2  -R02572          gggggcgggg
2.3.1.0.2  -R01309          gggggcgggg
2.3.1.0.2  +R01770          gggggcggca
2.3.1.0.2  +R01927          gggggcgggg
2.3.1.0.2  +R02440          gcgggcgggg
2.3.1.0.2  -R00337          gcgggcggga
2.3.1.0.2  -R00570          gggggctgca
2.3.1.0.2  -R02432          gggggcggga
2.3.1.0.2  +R01908          ggggccggcc
2.3.1.0.2  -R00224          ggggccgggg
2.3.1.0.2  +R02857          gcggccgggg
                            ||||||||||
                            ||||||||||
                            |||||||||
                            |||| ||||
                              ||   |

2.3.1.0    7   16           ksGGssGsGm
2.3.1.0.2  +R04034          ggggcggggc
2.3.1.0.2  +R00818          tgggcggggc
2.3.1.0.2  +R00825          gcgggcggga
2.3.1.0.2  -R01444          ggggcggcgc
2.3.1.0.2  -R00287          ggggcggggc
                            ||||||||||
                            ||||||||||
                            ||||||||||
                            ||||||||||
                              ||  | |

2.3.1.0    7   16           GnGGssGGrm
2.3.1.0.2  +R01307          ggggcggggc
2.3.1.0.2  +R01308          ggggcggggc
2.3.1.0.2  -R00385          gaggcgggga
2.3.1.0.2  +R02660          ggggcgggaa
2.3.1.0.2  +R04407          gtgggcggga
                            ||||||||||
                            ||||||||||
                            | ||||||||
                            | |||||||
                            | ||   ||

Segments:
2.3.1.0    6   16          =====Sp1===
2.1.1.1    12  21                =====GR===

2 segments in this sequence identified as potential binding sites
```

---

```
2 segments in complete file identified as potential binding sites
```

---

AND

## AliBaba2.1 predicts the following sites in your sequence

---

**Sequence seq_260**

```
===========================================================================
====
seq(     0..    59)     tcccagaggggccgggagaacagcagtcga
Class     lbp  rbp
2.1.1.1    12   21                   snrGAGrACA
2.1.1.1.1  +R00973                   ccagagaaca
2.1.1.1.1  -R01030                   cgggaggaca
2.1.1.1.1  -R01547                   ctggagaaca
2.1.1.1.1  -R01549                   ggggagaaca
2.1.1.1.1  +R03543                   ggagagaaca
                                     ||||||||||
                                     ||||||||||
                                     | ||||||||
                                        |||||||
                                        ||| |||

Segments:
2.3.1.0    6    16             =====Sp1===
2.1.1.1    12   21                   =====GR===

2 segments in this sequence identified as potential binding sites
```

```
2 segments in complete file identified as potential binding sites
```

# G>A -1859 rs3757583

## AliBaba2.1 predicts the following sites in your sequence

**Sequence seq_261**

```
===========================================================================
====
seq(     0..    59)     gaccaaagcctgatagctgtcctagaagca
Segments:

0 segments in this sequence identified as potential binding sites
```

```
0 segments in complete file identified as potential binding sites
```

**G>C -2253**

**AliBaba2.1 predicts the following sites in your sequence**

---

## Sequence seq_262

```
==============================================================================
====
seq(    0..   59)     cgagaagagagggggtccagctccccacagt
Segments:
2.3.1.0    9    22            ======Sp1=====
2.3.1.0    15   24                  ====Sp1===

2 segments in this sequence identified as potential binding sites
```

---

2 segments in complete file identified as potential binding sites

## Sequence seq_262

```
============================================================================
====
seq(     0..    59)      cgagaagagaggggtccagctccccacagt
Class       lbp   rbp
2.3.1.0      9     18              AGGGGkssrG
2.3.1.0.2   +R00149                agggggcggg
2.3.1.0.2   -R02891                agggggcggg
2.3.1.0.2   -R00955                aggggggcgg
2.3.1.0.2   -R03118                aggggtggag
                                   ||||||||||
                                   ||||||||||
                                   ||||||||||
                                   |||||    |
                                   |||||    |

2.3.1.0     13     22                  synmAGCyCC
2.3.1.0.2   +R01756                    ccccagctcc
2.3.1.0.2   -R02428                    gtccagcccc
2.3.1.0.2   +R02593                    gcacagctcc
2.3.1.0.2   +R04160                    cccaagctcc
2.3.1.0.5   -R03321                    gcgcagctcc
                                       ||||||||||
                                       ||||||||||
                                       || |||||||
                                        | |||||||
                                           |||  ||

Segments:
2.3.1.0      9     22              ======Sp1=====
2.3.1.0     15     24                    ====Sp1===

2 segments in this sequence identified as potential binding sites
```

```
2 segments in complete file identified as potential binding sites
```

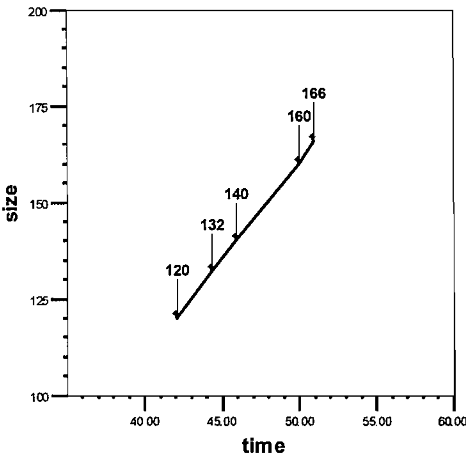
# Appendix H:

Extrapolation calculations depensing on three different programs:

To calculate the allelic size (extrapolation) depending on three known sizes (markers) that are close to the unknown size.

SPSS results
r^2=1 (interpretation as the proportion of the variation of one variabll 'explained' by the other)

## Interactive Graph




## Regression

**Variables Entered/Removed(b)**

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | TIME(a) | . | Enter |

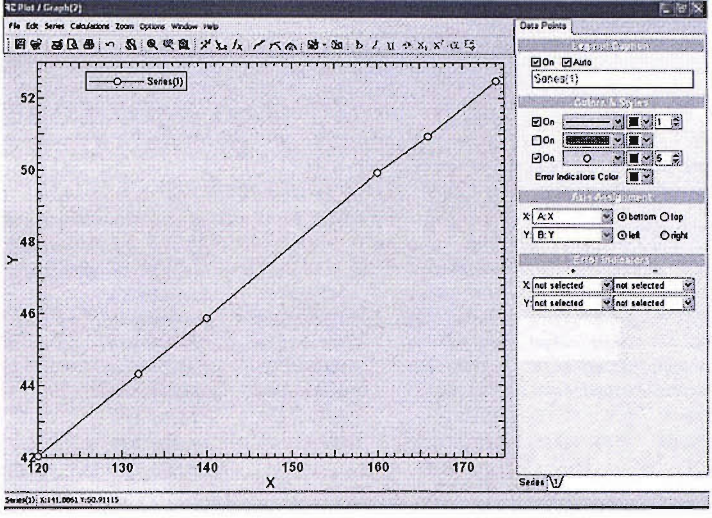
a All requested variables entered.
b Dependent Variable: SIZE

**Model Summary**

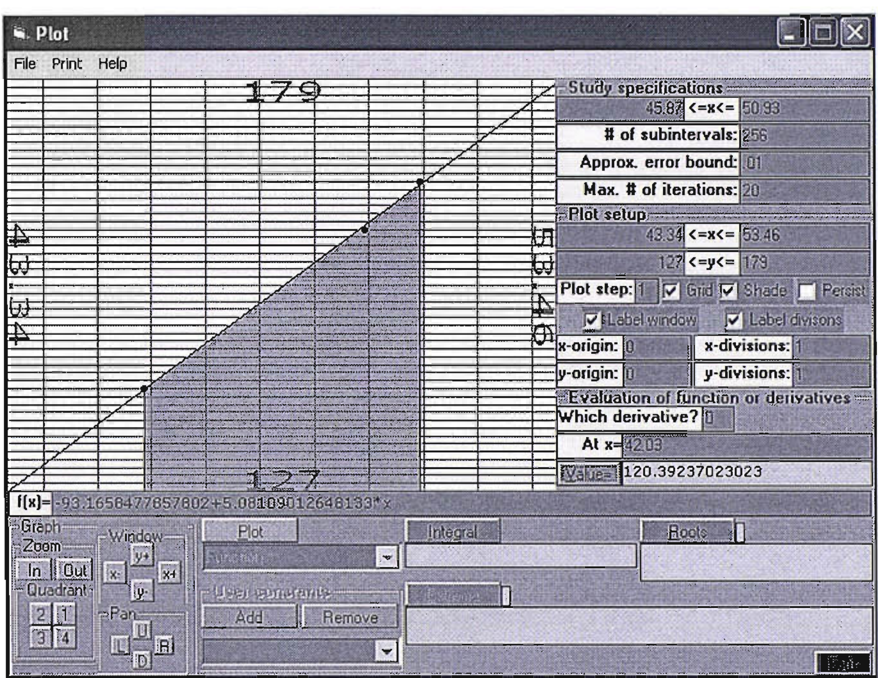
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | 1.000(a) | 1.000 | 1.000 | .38666 |

a Predictors: (Constant), TIME

Another programe was used:
RT Plot science shows results of both alleles in linear correlation.




a third programe called linear regression:



## Appendix I :

## *TGFBRII* EXONS 1and2

The location of D3S3727 (CA) SSR is present in about 29000 base from 5' of intron 1 (all intron one is 37769 bases) other STRs are about 5Mb upstream and down stream of the CA marker (not in this gene)

```
1       ggagagggag   aaggctctcg   ggcggagaga   ggtcctgccc   agctgttggc   gaggagtttc
61      ctgtttcccc   cgcagcgctg   agttgaagtt   gagtgagtca   ctcgcgcgca   cggagcgacg
121     acacccccgc   gcgtgcaccc   gctcgggaca   ggagccggac   tcctgtgcag   cttccctcgg
181     ccgccggggg   cctccccgcg   cctcgccggc   ctccaggccc   cctcctggct   ggcgagcggg
241     cgccacatct   ggcccgcaca   tctgcgctgc   cggcccggcg   cggggtccgg   agagggcgcg
301     gcgcggaggc   gcagccaggg   gtccgggaag   gcgccgtccg   ctgcgctggg   ggctcggtct
361     atgacgagca   gcggggtctg   ccatgggtcg   ggggctgctc   aggggcctgt   ggccgctgca
```

## EXON1

```
421     CATCGTCCTG   TGGACGCGTA   TCGCCAGCAC   GATCCCACCG   CACGTTCAGA   AGTCGG
477     gtgagtggtc   cccagcccgg   gctcggcggg   gcgccggggg   tcttcctggg   gtccccgcct
                                       X
                                       X
                                       X
                                       X
                                       X
29457   tttaaaggca   agtaactgat   tcacatgagg   ttgccgttgt   taatgttgcc   tctcaaacta
29517   gaatataagc   cataggaggg   tagggaccaa   cctggtgttt   gttttagatg   tcttgttgtg
29577   tggagagtat   cccagacatt   tgatatcctt   aattgttggc   cagatggtag   gtagtttcag
```

## D3S3727

```
29637   tgtgaaaata   tatatgtatg   tgtgtgtgtg   tgtgtgtgtg   tgtgtgtgtg   tgtgtgtgtg
29697   tgtatagtac   tgggattaca   ggcgtgagcc   accgcacccg   gcccgtacta   ggtattttac
29757   tagaattatt   tcctgctcta   agagatttta   gagtattgat   tatgtcattc   ttgacaggtt
X
X
X
X
X
38097   taatctgatg   tgaaggaatt   attttgcctt   tcttcagatt   cattctcatg   acatcaagtt
38157   catttgaaat   tgcataacat   cttcaggaat   tcattggcag   gctgcctggc   agttggataa
38217   tcatttaata   tatctttctc   tctcctcag
```

## EXON 2

```
38247   TTAATAACGA   CATGATAGTC   ACTGACAACA   ACGGTGCAGT   CAAGTTTCCA   CAACTGTGTA
38307   AATTTTGTGA   TGTGAGATTT   TCCACCTGTG   ACAACCAGAA   ATCCTGCATG   AGCAACTGCA
38367   GCATCACCTC   CATCTGTGAG   AAGCCACAGG   AAGTCTGTGT   GGCTGTATG
38417   gtaagcaagc   cttttaagaa   gttattcttt   cttttcccct   ttttacataa   tgtattctca
38477   tagtacacac   agtcagtgta   tctctgtctc   ctaaatgtaa   acacctgttc   catttccctt
```

## Appendix J :

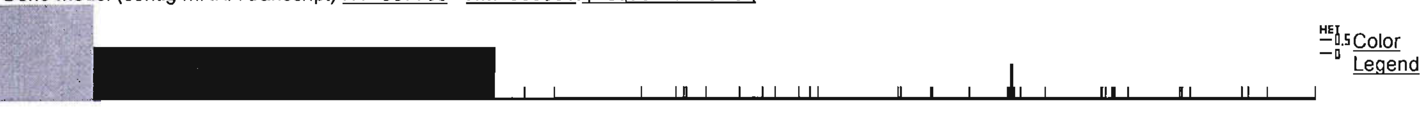
1- Results of the genotyping using the Light Typer on the British women heart cohort, results are with HW equilbrium

| Observed | | | | | | |
|---|---|---|---|---|---|---|
| | **11** | **12** | **22** | **0** | **unk** | **Total** |
| **Full set** | 1091 | 1341 | 458 | 788 | 546 | 4224 |
| | | **p** | **q** | **Total** | | |
| | Full set | 0.610 | 0.390 | | | |
| Expected | | | | | | |
| | **11** | **12** | **22** | **0** | **Other** | **Total** |
| **Full set** | 1073.66 | 1375.68 | 440.66 | 788.00 | 546.00 | 4224 |
| Chi-squared | | | | | | |
| | **11** | **12** | **22** | **Total X^2** | **P-value** | **Significant** |
| **Full set** | 0.28 | 0.87 | 0.68 | 1.84 | 0.1754 | **no** |

2- Results of SNP investigation using the NCBI site on **rs2071307 (Gly > Ser):**

**GeneView via analysis of contig annotation:** *ELN* elastin (supravalvular aortic stenosis, Williams-Beuren syndrome)
Click to see [all] [cSNP] [has frequency] [double hit] [haplotye tagged] variations associated with this gene.

Gene Model (contig mRNA transcript) NT_007758->NM_000501: [Sequence Viewer]




| Contig accession | Contig position | mRNA accession | mRNA orientation | Protein accession | Function | dbSNP allele | Protein residue | Codon position | Amino acid position |
|---|---|---|---|---|---|---|---|---|---|
| NT_007758 | 11504058 | NM_000501 | forward | NP_000492 | nonsynonymous | A | Ser [S] | 1 | 422 |
| | | | | | contig reference | G | Gly [G] | 1 | 422 |

Assay sample size (number of chromosomes): 110
Population data sample size (number of chromosomes):
Total number of populations with frequency data: 0
Total number of individuals with genotype data: 32 Genotype Detail NEW
Hardy-weinberg Probability: Pr(*chiSg= 0.168,df=1*) = 0.752
Average estimated heterozygosity: 0.404
Average Allele Frequency:

| | |
|---|---|
| **G** | **0.719** |
| **A** | **0.281** |

**Comment [U66]: Assumptions of Hardy-Weinberg Principle:**
As with many mathematical models, constraints are placed on them to simplify the model and test each component of it. The basic assumptions of the HWeqm are given below...
1.**Diploid Organism**
2.**Sexual Reproduction**
3.**Non-overlapping Generations**
4.**Equal allele frequencies in males and females**
5.**Random Mating**
6.**Large Population Size**
7.**No Migration**
8.**No Mutation**
9.**No Selection**
**Deviations from Hardy-Weinberg:**
Deviations from Hardy-Weiberg equilibrium may be caused by violations of any of the assumptions that are given above. For example, consider a locus with the alleles described above, except in this instance the locus is for an important gene required for aerobic resperation, and the a allele is a recessive mutation. The mutation is so severe that mutant homozygotes (individuals with the genotype aa) can not survive, and the fetus us spontaneously aborted before pregnancy. As a consequence no one is ever seen with the genotype aa, and any tests performed at this locus would result in a deviation from Hardy-Weinberg equilibrium.
This is of course just an example (however a large number of zygotes are spontaneously aborted, and it is thought that one of the contributing factors is grose genetic abnormalities), but violation of the other assumptions would also result in deviations from Hardy-Weinberg equilibrium. Migration would distort the allele frequencies, differnces in allele frequencies between males and females would result in non-random mating.

# Appendix K:

Fibrillin gene sequence

10194. Homo sapiens chro...[gi:37540936]

```
LOCUS       NT_010194              234913 bp    DNA     linear   CON 19-FEB-2004
DEFINITION  Homo sapiens chromosome 15 genomic contig.
ACCESSION   NT_010194 REGION: complement(19492744..19727656)
VERSION     NT_010194.16  GI:37540936
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 234913)
  AUTHORS   International Human Genome Sequencing Consortium.
  TITLE     The DNA sequence of Homo sapiens
  JOURNAL   Unpublished (2003)
COMMENT     GENOME ANNOTATION REFSEQ:  Features on this sequence have been
            produced for build 34 version 3 of the NCBI's genome annotation
            [see documentation].
            On Oct 7, 2003 this sequence version replaced gi:29801767.
            The DNA sequence is part of the second release of the finished
            human reference genome. It was assembled from individual clone
            sequences by the Human Genome Sequencing Consortium in consultation
            with NCBI staff.
```

# CODING SNPS FROM THIS WEBSITE:

http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=2200andview+rs+=view+rs+andchooseRs=codingand.cgifields=chooseRs

X
X
X
X
X
X
X

```
156841 ctttgatggc tcatagccta tgttccaggt tcttcctgct gccccagtgc aaccagtgaa
156901 gttaggaatc agtctgcaac agagggagac tgaggtagaa agaatgaggg aattcatgct
156961 gtgggtttgt gggtgatggc ggagaccagt tgtgggccct tgagaagtga ttttaacacc
157021 tgaaatggat acccaggcgg aagagaggac aattgagaca gaagctgtag tagaggtccc
157081 tttacaagat gtctgtccaa cacttctaga aaaggtgatt ggtaaacttg ttgctttagg
157141 tctggtttcc tagaaacaaa gtttgaaatg gagattcttg caaaagtggt atttcttact
157201 gtgagcaaat tctctcagaa ggtacttgtc aggaagtaag gagagcaggg tagtgcagga
```

## EXON 27(250bp)*

```
preceding intron phase: 1
atttccattttgcag<-flank
frame 1 (1):   D  I  D  E  C  Q  R  D  P  L  L  C  R  G  G  V  C  H  N  T  E  G  S  Y  R  C  E  C  P  P
19570191       ATATTGATGAGTGTCAGAGAGATCCTCTCCTATGCCGAGGTGGTGTTTGCCATAACACAGAGGGAAGTTACCGCTGTGAATGCCCGCCTG
3471           ..........................................................................................
               D  I  D  E  C  Q  R  D  P  L  L  C  R  G  G  V  C  H  N  T  E  G  S  Y  R  C  E  C  P  P
950            ..........................................................................................
               D  I  D  E  C  Q  R  D  P  L  L  C  R  G  G  V  C  H  N  T  E  G  S  Y  R  C  E  C  P  P
3471           ..........................................................................................
```

D I D E C Q R D P L L C R G G V C H N T E G S Y R C E C P P

intron phase: 1
flank->gtaaggagaaagact

frame 1 (1): G H Q L S P N I S A C I
19570101 GCCATCAGCTGTCCCCCAACATCTCCGCGTGTATCG
3561 ....................................
G H Q L S P N I S A C I
1040 ....................................
G H Q L S P N I S A C I
3561 ....................................
G H Q L S P N I S A C I

157261 gaagaaaata aaccaagagt ttggctgcat tttggtttta gtctgatcct gcagggagat
157321 ctggtgtatg aatagcacca gagagttgtc ccaccttgag acaaggaggt caagatggac
157381 acccagcaat gggtggggga ggagtgcttg gtctggtgga ggagatgagg cccccacctt
>>>>>>FBN 27 new F>>>
157441 taacatggtc atttccattt tgc█tatt gatgagtgtc agagagatcc tctcctatgc
**ss149094** A/G rs140597 Asp [D]> Gly [G]
157501 cgaggtggtg tttgccataa cacagaggga agttaccgct gtgaatgccc gcctggccat
157561 cagctgtcc█ ccaacatctc cgcgt█tatc g█aaggaga aagactttca caccatttac
**ss10788289** C/Grs7175654 Pro [P]> Ala [A] **ss149096** G/A rs140599 Cys [C]> Tyr [Y]
157621 ttgtggtcag ttgtttgaat gacatcattg ccaaagttgg aagcttatgt ttgggtgttt
<<<<<<<<FNB27R new <<<<<

## EXON 28 ( 298 bp)

preceding intron phase: 1
ttattttcccgacag<-flank
frame 1 (1): D I N E C E L S A H L C P N G R C V N L I G K Y Q C A C N P
19569954 ACATCAATGAATGTGAGCTGAGTGCACACCTGTGCCCCAATGGCCGTTGCGTGAACCTCATAGGGAAGTATCAGTGTGCCTGCAACCCTG
3597 ..........................................................................................
D I N E C E L S A H L C P N G R C V N L I G K Y Q C A C N P
1076 ..........................................................................................
D I N E C E L S A H L C P N G R C V N L I G K Y Q C A C N P
3597 ..........................................................................................
D I N E C E L S A H L C P N G R C V N L I G K Y Q C A C N P

intron phase: 1
flank->gtaagttctttttta

frame 1 (1): G Y H S T P D R L F C V
19569864 GCTACCATTCAACTCCCGATAGGCTATTTTGTGTTG
3687 ....................................
G Y H S T P D R L F C V
1166 ....................................
G Y H S T P D R L F C V
3687 ....................................
G Y H S T P D R L F C V

157621 ttgtggtcag ttgtttgaat gacatcattg ccaaagttgg aagcttatgt ttgggtgttt
157681 gttcttctta ttttcccgac █acatcaat gaatgtgagc tgagtgcaca cctgtgcccc
157741 aatggccgtt gcgtgaacct catagggaag tatcagtgtg cctgcaaccc tggctaccat
157801 tcaactcccg ataggctatt ttgtgttg█ aagttctttt ttattttatt ttattttatt
157861 ttattttact ttattttat█ ttatgttatg ttatgttatt ttattttatt ttatttcatc
157921 tctccctaaa acactctatg tactctgagt tttgttttca tcgctacttt tacttatctt
157981 tcttgtttcc ctgcaatatt ctttttatag tggctcacga tctttggatc tcatctccat
158041 tatctgttta aagagccagt taagtgaagg acataggatg ttttagctct aataagcata
158101 atttgaccca tgccgtttgg atagggaaca gtagcagcat gtgctttgtc tacacacatg
158161 cataccaatg cacattaaca aatggtaact aaccacctga aataaactaa acacggtgag
158221 tgctgtggaa aagagcagaa agctgagttg aaatcactgc ccttatgtgg tttagatcag
158281 ctaaaagatt tcaaaaaaat aactattgga aagaagacca tggaatgttc atatcttaca
158341 ctgcaattga ggctatgcct ggctctcctt aaagatgggt ttttgtggga caagaatttt

X
X
X
X
X
X
X
X
X
X

```
215221 tctcaacctc tgctgcactt tggaatcatt tggggagctc aagtccaacc tcagatattc
215281 tgattgaatt gttctggggg tggctggggc attgcattgg ggttttgaac ctgttttaac
215341 actgttctgg attcaccttc caactttctc tctttcctag tgagagctcc atctgaggga
215401 ccctgaacta ttctcaaagg aagggaagca cctcagcaaa ttatttttga tctttctaaa
215461 taatcatcag ataattctaa gaatctgtag atattttaat gttttttttaa acctgggaaa
215521 ctgagttgta gtttatgatt tcttaataat attgttgaaa tattacaata aatgcaacct
215581 atattggaga tataatctag gttaatctga agattaatct agacaaactc ccccagatca
215641 atattgacca ttacagatat tattctaaca tgtattaaaa tatttctttc atatgctgat
215701 attcacattt ttccctcctt cacttgaaat aacactttga gagtcctttt tatgtcctaa
215761 tatgaaaaga tttaaatttt actaagtgaa aaatgtaagt gtgcataaaa tgctaccatc
215821 tatgtttaag tatataaaat gctatcagtt aggcttttaa aagatctgaa gtgcacacac
215881 acatacacat attatgtatg tatttgcttt tatatgcata atccagctct ggaaggatac
215941 attagaaatc ggccccagaa ttggttgctt ttcaggaggg gagtttggtg tctgggaaat
216001 aggaaatgaa gggggtactt ttgctgaatc tcctttttgta cctttttaat tttgaaccat
216061 gtgaagagat aagctattta caaactatat aggaaagaag aaaatctatt cttctataaa
216121 agattcagat ggctctttct gttttcagtc tttcaatgaa accaaacagt taagaatgaa
216181 ttgaagtctc ttttatacct tttaaatttt gagccatgtg aacagattag tgattcaaaa
216241 gctaagttaa gaaggaaaga tgtgagagag ggaaggaagg tgagagggag ggaagggagg
216301 aaggaaagga gaaaggaaca aagggaggga aggagggagg gaggaaggaa ggaacgaagg
```

## EXON 56 (176 bp)

```
preceding intron phase: 1
aataaaatcaaacag<-flank
frame 1 (1):   D  E  N  E  C  Q  T  K  P  G  I  C  E  N  G  R  C  L  N  T  R  G  S  Y  T  C  E  C  N  D
19511225       ATGAGAATGAATGTCAGACGAAGCCAGGGATCTGTGAGAATGGGCGCTGCCTCAACACCCGTGGGAGCTACACCTGTGAGTGTAATGATG
7005           ..........................................................................................
               D  E  N  E  C  Q  T  K  P  G  I  C  E  N  G  R  C  L  N  T  R  G  S  Y  T  C  E  C  N  D
7005           ..........................................................................................
               D  E  N  E  C  Q  T  K  P  G  I  C  E  N  G  R  C  L  N  T  R  G  S  Y  T  C  E  C  N  D

                                                  intron phase: 1
                                                  flank->gtgagtacagttggc
frame 1 (1):   G  F  T  A  S  P  N  Q  D  E  C  L
19511135       GGTTTACCGCCAGCCCCAACCAGGACGAGTGCCTTG
7095           ....................................
               G  F  T  A  S  P  N  Q  D  E  C  L
7095           ....................................
               G  F  T  A  S  P  N  Q  D  E  C  L
```

```
216361 aaggagctcc atcctctata aaatggtcag atgactcttc ttgtttttgg tccttcaata
216421 aaatcaaac██atgagaatg aatgtca█ac gaagccaggg atctgtgaga atgggcgctg
                                ss20009974 A/G rs363830 Gln [Q]> Gln [Q]
216481 cctcaacacc cgtgggagct acacctgtga gtgtaatgat gggtttaccg ccagccccaa
216541 ccagga█gag tgccttg█g agtacagttg gcaccgcact ttcctaacct cagcctccac
                  ss461299 C/G rs363831 Glu [E]> Asp [D]
216601 actgggatgc tggaaaccca gacttcttat ttaaaataca agaaaatgtc aaaatctgag
216661 gaaggataaa aaatgttcat attttggaga tgccgtaatg actgtgattg tccattgggc
216721 tcagcaccac cctgcagcta aattcttcct ttgctaattg gatcctgaat cacttgtttg
216781 gaatttcttg gctgcctttg aagcccttgg tgatcaggat ccacttccgt atgtttctct
216841 gcccctctgt ctgtaagcat ggctattccc ctgtatttct gggagcagag agagttataa
216901 gcttgtctag gccaatgtct ttctgcctac ttactgaatg tcttatttgt ttccctcaga
216961 cgtcatcctt tctcttttac tgctgtctcc agctttcccc tcttgcttct tctcacccag
```

## EXON 57 (277 bp)*

```
preceding intron phase: 1
taatgtcccttccag<-flank
frame 1 (1):   D  N  R  E  G  Y  C  F  T  E  V  L  Q  N  M  C  Q  I  G  S  S  N  R  N  P  V  T  K  S  E
19510527       ACAATCGGGAAGGGTACTGCTTCACAGAGGTGCTACAAAACATGTGTCAGATCGGCTCCAGCAACAGGAACCCCGTCACCAAATCGGAAT
7131           ..........................................................................................
               D  N  R  E  G  Y  C  F  T  E  V  L  Q  N  M  C  Q  I  G  S  S  N  R  N  P  V  T  K  S  E
7131           ..........................................................................................
               D  N  R  E  G  Y  C  F  T  E  V  L  Q  N  M  C  Q  I  G  S  S  N  R  N  P  V  T  K  S  E

frame 1 (1):   C  C  C  D  G  G  R  G  W  G  P  H  C  E  I  C  P  F  Q  G  T  V  A  F  K  K  L  C  P  H
19510437       GCTGCTGTGACGGAGGGAGAGGCTGGGGTCCCCACTGTGAGATCTGCCCTTTCCAGGGGACTGTGGCTTTCAAGAAACTCTGTCCCCATG
7221           ..........................................................................................
```

```
             C   C   C   D   G   G   R   G   W   G   P   H   C   E   I   C   P   F   Q   G   T   V   A   F   K   K   L   C   P   H
7221         ..........................................................................................
             C   C   C   D   G   G   R   G   W   G   P   H   C   E   I   C   P   F   Q   G   T   V   A   F   K   K   L   C   P   H

                                 intron phase: 1
                                 flank->gtacttcatttatag
frame 1 (1):  G  R  G  F  M  T  N  G  A
19510347      GCCGAGGATTCATGACCAATGGAGCAG
7311          ...........................
              G  R  G  F  M  T  N  G  A
7311          ...........................
              G  R  G  F  M  T  N  G  A

   217021 ggtaaagtgt tacatccttt tttggttttt atatctgacc aaatttttaa tattttgttt
   217081 gctcttaaaa tttcctgaca tcccctttgc catataatgt cccttcc█a caatcgggaa
                              >>>>>>>MS 57 f>>>>>>>>>>
   217141 gggtactgct tcacagaggt gctacaaaac atgtgtcaga tcggctccag caacaggaac
   217201 cccgtcacca aatcggaatg ctgctgtga█ ggagggagag gctggggtcc ccactgtgag
                                 ss1466682 C/T rs1005074 Asp [D]> Asp [D]
   217261 atctgccctt tccaggggac tgtggctttc aagaaactct gtccccatgg ccgaggattc
   217321 atgaccaatg gagcag█ac ttcatttata gtccaaaaat acttgcaggg aatctattta
                                         <<<<<<<<<< MS 57 r<<<<
   217381 tttgtttttt gtgtgaaaca cagatgaaaa tatggagttt gcaatatgtg cctaggttga
   217441 attgcacagc tgaggccaac aaaaattctt cattatggag ttttagacat ttgaggtcat
   217501 gctgccaaag tggccttggt ccacttgatt aagctgtgtg tgcttaggac cctccctctg
   217561 cattccagcg agtcttcaac tttttaaacc gtttactata agtcaattgg ctactcagaa
   217621 tataattttt aaatttttta gataaacttt gaaataaatg attattaaat gtgcaaacca
   217681 gtccaccaag atgtggaact cctaaaactg gccaaactgt attgcataaa agaagtataa
   217741 atcatgctga tatcctggca attagacaca attgcatctg gaaaagcaca agctccttaa
   217801 ctagagcagg atgaggaagt gtcagactgt ctaaggcaac tctagagata tctaggcctg
   217861 tcaccaatat gattcctgtg tttatcaact gtaactacaa ttggttttta gttcataaaa
   217921 ggactgagaa aaggtgagag ctttttttgag cttcttggga gggaattaga agtgaaaggg
   217981 gaagaaggag gttaagaagg atttgtgttt gtattggaaa gatgtggctg aaggcagaat
   218041 aatggaactg gatgtgattt aaaaaaaaag aaaaaacaga cctagaaagg taagatggag
   218101 tgaccaaaag caatttatga ttctttggag gatgtagcaa taagatagcc gaaaaaactc
   218161 acttctggag accaataaga gtgtgctaaa ttgcacttca gagtagactt gtaaaggtca
   218221 gatacaccca gctgttggta taaaatggtt atttggcagt caagggctga gaacccatgt
   218281 agaaagtgct tgtatcggca cttcgccaag cagtaaggag gtaaaatggg tcaagatgtg
   218341 gaagacagtg gttttaatgc tcaacagtgc atattgaccc agtctttctc tctgtggtac
   218401 agccatgagt gagggtctcc cttccttagc aggctgcaaa aacccagcac tgtcatccac
   218461 ctcagagccc tggttagttt ccccccctcac cagaggacag aagttcatct catttcccca
   218521 atattatgta atttgaactt tttcaggatt aaccacaaag acaacaaaaa taatgtaaaa
   218581 acaaaaaaca aaaaacctta cttatttga gcatgcaata ctaggggggtg gaaatggggt
   218641 ttcctcatca aaaggatctg cttctgtgat tcatttatta gatgtcagca ggaggcaagc
   218701 ttaatggcat gagatgacac aaaaagccaa atatgtctag atggggtggc tgttctatgg
   218761 atccagaaga gaaaaatata catttagaag accagttgtt tttaaagttt ttctgtggct
   218821 gaattatttt atcaattgta gtctaatgaa gttatttgtt taccttttga tatagctact
   218881 gttacatatt aaatttatgc tgtcatctta ctggtttaat ttcttaggcc caaaatatag
```

## Exon 58 (259 bp)

```
preceding intron phase: 1
tttgttaaattacag<-flank
frame 1 (1):  D  I  D  E  C  K  V  I  H  D  V  C  R  N  G  E  C  V  N  D  R  G  S  Y  H  C  I  C  K  T
19508618      ATATCGATGAATGCAAGGTTATTCACGATGTTTGCCGAAATGGGGAATGTGTCAATGACAGAGGATCATATCATTGCATTTGTAAAACTG
7338          ..........................................................................................
              D  I  D  E  C  K  V  I  H  D  V  C  R  N  G  E  C  V  N  D  R  G  S  Y  H  C  I  C  K  T
7338          ..........................................................................................
              D  I  D  E  C  K  V  I  H  D  V  C  R  N  G  E  C  V  N  D  R  G  S  Y  H  C  I  C  K  T

                                   intron phase: 1
                                   flank->gtaagtgtctatttc
frame 1 (1):  G  Y  T  P  D  I  T  G  T  S  C  V
19508528      GGTACACTCCAGATATAACTGGGACTTCCTGTGTAG
7428          ....................................
              G  Y  T  P  D  I  T  G  T  S  C  V
7428          ....................................
              G  Y  T  P  D  I  T  G  T  S  C  V

   218941 taacacaatt tattagtatt tacactgaag tgacccccta catattaatg ttgtcaattt
```

```
                                      >>>>>>>>>MS 58 F >>>>>>
   219001 tatgatatat ttcttaattt atatttgtta aattacat atcgatgaat gcaaggttat
   219061 tcacgatgtt tgccgaaatg gggaatgtgt caatgacaga ggatcatatc attgcatttg
   219121 taaaactggg tacactccag atataactgg gacttcctgt gtagaagt gtctatttct
   219181 gatggcttat cctcaagtgg aaattttaga ttatggaaaa aaaaaaaccc aaagctaaaa
    <<<<<<<<<<MS 58 R <<<<<
   219241 atctaaaaag tctgtgcagt ttcataggaa agcacaggac aatcatcaaa gtctacacag
```

## EXON 59 (246bp)*

```
preceding intron phase: 1
tttctttgatcatag<-flank
frame 1 (1):   D  L  N  E  C  N  Q  A  P  K  P  C  N  F  I  C  K  N  T  E  G  S  Y  Q  C  S  C  P  K  G
19508245       ATCTGAACGAGTGCAACCAGGCTCCCAAACCCTGCAATTTTATCTGCAAAAACACAGAAGGGAGTTACCAGTGTTCATGCCCGAAAGGCT
7464           ..........................................................................................
               D  L  N  E  C  N  Q  A  P  K  P  C  N  F  I  C  K  N  T  E  G  S  Y  Q  C  S  C  P  K  G
7464           ..........................................................................................
               D  L  N  E  C  N  Q  A  P  K  P  C  N  F  I  C  K  N  T  E  G  S  Y  Q  C  S  C  P  K  G

                                    intron phase: 1
                                    flank->gtaaagtagaattga
frame 1 (1):   Y  I  L  Q  E  D  G  R  S  C  K
19508155       ACATTCTGCAAGAGGATGGAAGGAGCTGCAAAG
7554           .................................
               Y  I  L  Q  E  D  G  R  S  C  K
7554           .................................
               Y  I  L  Q  E  D  G  R  S  C  K

   219301 agtttccttt ttctttcccc gaaactaaaa ttcttcgtta gaccctgtgg aaattgagcg
                                                           >>>>>>> MS 59 F >>>>>
   219361 tgtacacatc atttttagat gcacagtcac gctgtatttc tttgatcat atctgaacg
   219421 agtgcaacca ggctcccaaa ccctgcaatt ttatctgcaa aaacacagaa gggagttacc
   219481 agtgttcatg cccgaaaggc tacattctgc aagaggatgg aaggagctgc aaagaaag
   219541 tagaattgac cattgcccct cacctagctc ctgacacatg gctgcattcc tttgcctgta
                                       <<<<<<<< MS 59 R <<<<<<
   219601 agaactccac aggcaagccg aaagacctgc tctacaggca ggaggtgctt ccctggggtt
   219661 aatgcctcca tgggacctac ctcctgtgct ggagggaagg aggctaaagc cagggagatt
                                       X
                                       X
                                       X
                                       X
                                       X
                                       X
```

## EXON 60 (189 bp)

```
preceding intron phase: 1
tgcattttcttgtag<-flank
frame 1 (1):   D  L  D  E  C  A  T  K  Q  H  N  C  Q  F  L  C  V  N  T  I  G  G  F  T  C  K  C  P  P  G
19504822       ATCTTGATGAGTGTGCAACCAAGCAACACAACTGCCAGTTCCTATGTGTTAACACCATTGGCGGCTTCACATGCAAATGTCCTCCCGGAT
7587           ..........................................................................................
               D  L  D  E  C  A  T  K  Q  H  N  C  Q  F  L  C  V  N  T  I  G  G  F  T  C  K  C  P  P  G
7587           ..........................................................................................
               D  L  D  E  C  A  T  K  Q  H  N  C  Q  F  L  C  V  N  T  I  G  G  F  T  C  K  C  P  P  G

                              intron phase: 1
                              flank->gtgagtaggagagga
frame 1 (1):   F  T  Q  H  H  T  S  C  I
19504732       TTACCCAACACCATACGTCCTGCATTG
7677           ...........................
               F  T  Q  H  H  T  S  C  I
7677           ...........................
               F  T  Q  H  H  T  S  C  I

   222721 tagtcagggt catttgagac ctccaaatca aacgtggagc tgcttcatag ggtcagcttc
   222781 cctgatcctg ttttgttggc ttgactcaaa tgcctctctt gcattttctt gtatcttg
   222841 atgagtgtgc aaccaagcaa cacaactgcc agttcctag tgttaacacc attggcggct
                                   ss461278 A/T rs363810 SER>CYS
   222901 tcacatgcaa atgtcctccc gatttaccc aacaccatac gtcctgcatt ggagtagg
                            ss461279 A/G rs363811 GLY>ARG
   222961 agaggaaaaa atcctacatg gattgtagcg attcttttaa gggattattt tctatttcct
```

# EXON 61 (191 bp)

```
preceding intron phase: 1
cactgcttctcatag<-flank
frame 1 (1):  D  N  N  E  C  T  S  D  I  N  L  C  G  S  K  G  I  C  Q  N  T  P  G  S  F  T  C  E  C  Q
19504440      ATAACAATGAATGCACCTCTGACATCAATCTGTGCGGGTCTAAGGGCATTTGCCAGAACACTCCTGGAAGCTTCACCTGTGAATGCCAGC
7704          ..........................................................................................
              D  N  N  E  C  T  S  D  I  N  L  C  G  S  K  G  I  C  Q  N  T  P  G  S  F  T  C  E  C  Q
7704          ..........................................................................................
              D  N  N  E  C  T  S  D  I  N  L  C  G  S  K  G  I  C  Q  N  T  P  G  S  F  T  C  E  C  Q

                                          intron phase: 1
                                          flank->gtgggtggagacttc
frame 1 (1):  R  G  F  S  L  D  Q  T  G  S  S  C  E
19504350      GGGGATTCTCACTTGATCAGACCGGCTCCAGCTGTGAAG
7794          .......................................
              R  G  F  S  L  D  Q  T  G  S  S  C  E
7794          .......................................
              R  G  F  S  L  D  Q  T  G  S  S  C  E
```

```
223021 ctgctgttgg gataagaaaa taaaagctca aagaaatata tgagtgcatg tatgtgtgag
223081 cacacctgta catgtatgtg aagcgttgtt ggccttattt ggccttttcc gagttatcct
                                   rs1820488 a/c
223141 tctaattttc ttttaaatga tacaaagaga gctttgggga attttaaccc ctctttgccc
223201 ccactgcttc tcat■ataa caatgaatgc acctctgaca tcaatctgtg cgggtctaag
223261 ggcatttg■c agaacactcc tggaagcttc acctgtgaat gccagcgggg attctcactt
          ss8073671 -/G rs5812451
223321 gatcagaccg gctccagctg tgaag■ggg tggagacttc agctgcgatc cagctggtga
223381 atccttgtgg aggtggcctg tgtggctatt ggcaccttca tcatcagcct ctatgagata
223441 gcagatctga gcccaggggg gcactcagct aaaatagtgt gcacaggtcc tgtatcttag
223501 agagtcgtgt tttgtttcat gcttgattgt gtctaacaca catgcctttc ctaagtatat
223561 caagactttt tctgctaagt ttttaattta acctaatagg gaataactca gtcaagaaaa
223621 atatacttga ttttgattca ttaattttct ggtcactggg tccagagcca taaagggatt
223681 tcccccatgt ggactaaaca tatgaggaat agctcttctg atgacattta attagaagtt
223741 taaatgatat ttttagaagt tgactgtggc tttgacacgt tttggtttct aattgtggca
223801 ccaaaaataa aaaaaaaaaa gagcaaacac aaataactta taacttacag agctgtccca
223861 gagagtgctt tgggctttgc actaatttcc tgacaatttt tatttgtaga ctttgccagg
223921 gctctctgaa tgattttctc cttggactta gcagcagttc cagaagagag attcttgaag
```

# EXON 62 (174 bp)

```
preceding intron phase: 1
ctgcttcttttcag<-flank
frame 1 (1):  D  V  D  E  C  E  G  N  H  R  C  Q  H  G  C  Q  N  I  I  G  G  Y  R  C  S  C  P  Q  G  Y
19503560      ACGTGGACGAGTGTGAGGGTAACCACCGCTGCCAGCATGGCTGCCAGAACATCATTGGGGGCTACAGGTGCAGCTGCCCCCAGGGCTACC
7833          ..........................................................................................
              D  V  D  E  C  E  G  N  H  R  C  Q  H  G  C  Q  N  I  I  G  G  Y  R  C  S  C  P  Q  G  Y
7833          ..........................................................................................
              D  V  D  E  C  E  G  N  H  R  C  Q  H  G  C  Q  N  I  I  G  G  Y  R  C  S  C  P  Q  G  Y

                                   intron phase: 1
                                   flank->gcaagtaacttttcc
frame 1 (1):  L  Q  H  Y  Q  W  N  Q  C  V
19503470      TCCAGCACTACCAGTGGAACCAGTGTGTTG
7923          ..............................
              L  Q  H  Y  Q  W  N  Q  C  V
7923          ..............................
              L  Q  H  Y  Q  W  N  Q  C  V
```

```
223981 tttttggtgg tagaataatg tgtaggatgt gtaggggcca gatttcttat tagaatccat
224041 ctggcttcag agagagatgt tgagttggca tcatggtggc tctgcttctt tttc■acgt
224101 ggacgagtgt gagggtaacc accgctgcca gcatggctgc cagaacatca ttgggggcta
224161 caggtgcagc tgcccccagg gctacctcca gcactaccag tggaaccagt gtgttggcaa■
224221 gtaacttttc ctcactctca agatgcatgg ctatcaggtc ctatgaagca aaacactgct
                                    X
                                    X
                                    X
```

X

# EXON 63 (298 bp)

```
preceding intron phase: 1
ttattttgcctgcag<-flank
frame 1 (1):   D  E  N  E  C  L  S  A  H  I  C  G  G  A  S  C  H  N  T  L  G  S  Y  K  C  M  C  P  A  G
19498521       ATGAAAACGAATGCCTCAGCGCTCACATCTGCGGAGGAGCCTCCTGTCACAACACCCTGGGGAGCTACAAGTGCATGTGTCCCGCCGGCT
7953           ..........................................................................................
               D  E  N  E  C  L  S  A  H  I  C  G  G  A  S  C  H  N  T  L  G  S  Y  K  C  M  C  P  A  G
7953           ..........................................................................................
               D  E  N  E  C  L  S  A  H  I  C  G  G  A  S  C  H  N  T  L  G  S  Y  K  C  M  C  P  A  G

frame 1 (1):   F  Q  Y  E  Q  F  S  G  G  C  Q  D  I  N  E  C  G  S  A  Q  A  P  C  S  Y  G  C  S  N  T
19498431       TCCAGTATGAACAGTTCAGTGGAGGATGCCAAGACATCAATGAATGTGGCTCTGCGCAGGCCCCCTGCAGCTATGGCTGTTCCAATACCG
8043           ..........................................................................................
               F  Q  Y  E  Q  F  S  G  G  C  Q  D  I  N  E  C  G  S  A  Q  A  P  C  S  Y  G  C  S  N  T
8043           ..........................................................................................
               F  Q  Y  E  Q  F  S  G  G  C  Q  D  I  N  E  C  G  S  A  Q  A  P  C  S  Y  G  C  S  N  T

                                                     intron phase: 2
                                                     flank->gtaagcagtgctctt
frame 1 (1):   E  G  G  Y  L  C  G  C  P  P  G  Y  F  R  I  G  Q  G
19498341       AGGGCGGTTACCTGTGTGGCTGTCCACCTGGTTACTTCCGCATAGGCCAAGG
8133           ....................................................
               E  G  G  Y  L  C  G  C  P  P  G  Y  F  R  I  G  Q  G
8133           ....................................................
               E  G  G  Y  L  C  G  C  P  P  G  Y  F  R  I  G  Q  G
```

```
229021 ccttcctgag agcctagctg agggccagct ggccggcagc aagtggccag atccaatgtc
229081 ctcaatagaa atctctggct gctgccacac atgccgcttc ttattttgcc tgc■atgaa
229141 aacgaatgcc tcagcgctca catctgcgga ggagcctcct gtcacaacac cctggggagc
229201 tacaagtgca tgtgtcccgc cggcttccag tatgaacagt tcagtggagg atgccaagac
229261 atcaatgaat gtggctctgc gcaggccccc tgcagctatg gctgttccaa taccgagggc
229321 ggttacctgt gtggctgtcc acctggttac ttccgcatag gccaagg■a agcagtgctc
229381 ttcctggtca tggttggaga ttctttcatt cgtaatataa ttaagtatac tgaactcaaa
229441 attacctgtc ctagcagagg agaaccatgc tttttgtaat cctaaaatta attccagtta
```

X
X
X
X
X
X
X

# EXON 64 (243 bp)

```
preceding intron phase: 2
ccccttctgctgcag<-flank
frame 2 (1):   H  C  V  S  G  M  G  M  G  R  G  N  P  E  P  P  V  S  G  E  M  D  D  N  S  L  S  P  E  A
19495497      GCACTGTGTTTCTGGAATGGGCATGGGCCGAGGAAACCCAGAGCCACCTGTCAGTGGTGAAATGGATGACAATTCACTCTCCCCAGAGGC
8185          ..........................................................................................
               H  C  V  S  G  M  G  M  G  R  G  N  P  E  P  P  V  S  G  E  M  D  D  N  S  L  S  P  E  A
8185          ..........................................................................................
               H  C  V  S  G  M  G  M  G  R  G  N  P  E  P  P  V  S  G  E  M  D  D  N  S  L  S  P  E  A

                                                                          intron phase: 0
                                                                          flank->gtgggtcagaagtta
frame 2 (1):   C  Y  E  C  K  I  N  G  Y  P  K  R  G  R  K  R  R  S  T  N  E  T  D  A  S  N  I  E
19495407      TTGTTACGAGTGTAAGATCAATGGCTACCCCAAACGGGGCAGGAAACGGAGAAGCACAAACGAAACTGATGCCTCCAATATCGAG
8275          ....................................................................................
               C  Y  E  C  K  I  N  G  Y  P  K  R  G  R  K  R  R  S  T  N  E  T  D  A  S  N  I  E
8275          ....................................................................................
               C  Y  E  C  K  I  N  G  Y  P  K  R  G  R  K  R  R  S  T  N  E  T  D  A  S  N  I  E
```

```
232081 ggcttcacca gaactcttgt accacctacc ttgtcttccc attctaatga aaaacatcta
232141 tgctcccctt ctgctgc■g cactgtgttt ctggaatggg catgggccga ggaaacccag
                >>>>MS 64 F>>>>
232201 agccacctgt cagtggtgaa atggatgaca attcactctc cccagaggct tgttacgagt
232261 gtaagatcaa tggctacccc aaacggggca ggaaacggag aagcacaaac gaaactgatg
232321 cctccaatat cgag■gggt cagaagttag tttctcctga tgtctcctgt ggtggaaagc
```

```
                                                 <<<<<<<<MS 64 R <<<<<<
232381 ccttccagat tcctgtggtt tcctccaagg atgctccaaa gtgtgaaaaa gctccccagg
232441 gagaaactcc agacattccc tgagctctag gctgtatttt acaagaggct gtggggcttt
232501 ctggagtttc ttgctgcttt tcgggagtgg ccatttgaag gttttcacat acttgctcat
232561 ttcctctgtc tcgccaaaaa aaagtcaaat acccacccaa agtctgctga aaggataagg
232621 tctgagatcc aggacagcaa gtgggcagaa tagcgaagtg acatagtgct ccctacatag
232681 gaatgccaag ttctgaaagc tagaggctcc tctactgaaa agagccatgt ggagtgtgag
232741 ggaacttgga ttgtagacaa aaataacgtg gagctgaagc agcctggttt cctttgccta
232801 cggctgctgc agcctttcag tcaactcaga agtgcacaga tattcaccca gacctctggc
232861 tctgggccca tagaatccag cacatgtggt cccagagagg caagtaagaa gccaggtgaa
232921 aagcattgaa tattagcaga tcatcacagt caggaagtag gctcataggt gtgtccttag
232981 gagaagatgg ttttacgtca atgaatcaaa ttgaaatagt ctggttcttc gtattggtgc
233041 tgcaacccat ctgctgtcat ttaagaaatg agataagtct ttggaaagaa cataatctat
233101 taaaacagtg gttctcaaac tttagcatgc atctgaatca cctactaggt cttgttaaaa
233161 ttgattgttg ggagacatcc cagggtttct gattcagtag ttctggggtg atgccgagaa
233221 tttgcatttc taacagattc ccaggtgatg ctgatgcagc ttgttcaggg actacatttt
233281 gagacctcca gatacaaatg atttcaacct gcctttcttc ctgacatcag ttaatatttt
233341 caaatattac aaatatgtgc caatttaata cacttgtggt ctaaacaaaa tgctttcaat
233401 attgtgtatg cagcataagg cagaaaattg tattagtgtg aaatttgagt catttttttct
```

## Exon 65 ( 283 and 232 bp)

```
preceding intron phase: 0
catgtgtctttccag<-flank
frame 3 (1):  D Q S E T E A N V S L A S W D V E K T A I F A F N I S H V S
19494133      GATCAGTCTGAGACAGAAGCCAATGTGAGTCTTGCAAGTTGGGATGTTGAGAAGACAGCCATCTTTGCTTTCAATATTTCCCACGTCAGT
8360          ..........................................................................................
              D Q S E T E A N V S L A S W D V E K T A I F A F N I S H V S
8360          ..........................................................................................
              D Q S E T E A N V S L A S W D V E K T A I F A F N I S H V S

frame 3 (1):  N K V R I L E L L P A L T T L T N H N R Y L I E S G N E D G
19494043      AACAAGGTTCGAATCCTAGAACTCCTTCCAGCTCTTACAACTCTGACGAATCACAACAGATACTTGATCGAATCTGGAAATGAAGATGGC
8450          ..........................................................................................
              N K V R I L E L L P A L T T L T N H N R Y L I E S G N E D G
8450          ..........................................................................................
              N K V R I L E L L P A L T T L T N H N R Y L I E S G N E D G

frame 3 (1):  F F K I N Q K E G I S Y L H F T K K K P V A G T Y S L Q I S
19493953      TTCTTTAAAATCAACCAAAAGGAAGGGATCAGCTACCTCCACTTCACAAAGAAGAAGCCAGTGGCTGGAACCTATTCATTACAAATCAGT
8540          ..........................................................................................
              F F K I N Q K E G I S Y L H F T K K K P V A G T Y S L Q I S
8540          ..........................................................................................
              F F K I N Q K E G I S Y L H F T K K K P V A G T Y S L Q I S

frame 3 (1):  S T P L Y K K K E L N Q L E D K Y D K D Y L S G E L G D N L
19493863      AGTACTCCACTTTATAAAAAGAAAGAACTTAACCAACTAGAAGACAAATATGACAAAGACTACCTCAGTGGTGAACTGGGTGATAATCTG
8630          ..........................................................................................
              S T P L Y K K K E L N Q L E D K Y D K D Y L S G E L G D N L
8630          ..........................................................................................
              S T P L Y K K K E L N Q L E D K Y D K D Y L S G E L G D N L

frame 3 (1):  K M K I Q V L L H *
19493773      AAGATGAAAATCCAGGTTTTGCTTCATTAATTCACCATCCAGAGACCAAATAATTAAAAGAAAAACAAATATAGATAGGTAGAACTATAT
8720          ..........................................................................................
              K M K I Q V L L H *
8720          ...........................................-..............................................
              K M K I Q V L L H *
```

```
233461 ttaatatgag agctaagtgg catatgtaca ttgtatttaa catattgcca tgtgtctttc
233521 c  gatcagt ctgagacaga agccaatgtg agtcttgcaa gttgggatgt tgagaagaca
233581 gccatctttg ctttcaatat ttcccacgtc agtaacaagg ttcgaatcct agaactcctt
233641 ccagctctta caactctgac gaatcacaac agatacttga tcgaatctgg aaatgaagat
233701 ggcttcttta aaatcaacca aaaggaaggg atcagctacc tccacttcac aaagaagaag
       ccagtggctg gaacctattc
233761 ccagtggctg gaacctattc attacaaatc agtagtac c cactttataa aaagaaagaa
                                         ss461315 C/T rs363847
233821 cttaaccaac tagaagacaa atatgacaaa gactacctca gtggtgaact gggtgataat
```

233881 ctgaagatga aaatccaggt tttgcttcat [illegible]ttcacca tccagagacc aaataattaa
**ss461316** **T/C rs363848**

233941 aagaaaaaca aatatagata ggtagaacta tattttcccc caatcagaat catcatatca
234001 taggtacaat ctttcaccaa gtaaatttgt ataaataagc actattcttt gtattaccaa
234061 agcaaggtac aggtgactac cctagttcaa aacaaccact ttctcaggct tctcatgtgt
234121 gtagctaagc taccttgtca tatgtgttga ttcttgaaaa ctgggacgtg tatttccatt
234181 gggggttggc catttatgct gacatgccat ccttccagca aacgtacggg aatgtgcttt
234241 caattgatgg actactctat tttttgcaaa tttgtaaact ttgcttctcc aaatacaagt
234301 actaggttgt ccatttatgg tacctatttg gtgctagtaa attttcaaac tagatttata
234361 aatgcactgt aatatgtaca caacttagaa accaaattac aagtattcag ttccaatact
234421 tcattaattt caatcaacca aagttagttc agtagcttat ctcagttatg agtataatac
234481 attacatgta aattaagtgt gtgtatactg taatcgtgct attttttatc attgaaacat
234541 ttataaacta gaataataat gcccttaatg tgagggtttg taatggtgct tattaagacc
234601 aaagacttgt taaatgtata caccaagtgg taatgaaatt tcggtgactg gcccacacgt
234661 gcatagaggt ctgggaggac caggaaacag cctcagtggc cagaggatca ccagtgcatc
234721 cttcatcaca gcatgtgcaa tatgccaaga ttaccctcgg tcattcctgt caacaagggg
234781 tcaatgtcat aaatgtcaca ataaaacaat ctcttctttt ttttagttta ccccttggct
234841 ttgtgttctt gcatggattt ggggttggag gggccattcc ggaggctaaa taaagtctcc
234901 tggatttaaa tta