

https://dspace.org/

Genetic susceptibility to fetal alcohol syndrome in the Northern Cape coloured population: Potential roles of astrotactin and reelin

Macaulay, Shelley

2007-02-19T12:39:25Z
http://hdl.handle.net/10539/2049

Downloaded from a DSpace repository. https://dspace.org/

GENETIC SUSCEPTIBILITY TO FETAL ALCOHOL SYNDROME IN THE NORTHERN CAPE COLOURED POPULATION: POTENTIAL ROLES OF *ASTROTACTIN* AND *REELIN*

Shelley Macaulay

A dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, in fulfilment of the requirements for the degree of Master of Science in Medicine

Johannesburg, 2006

DECLARATION

I, Shelley Macaulay, declare that this dissertation is my own work. It is being submitted for the degree of Master of Science in Medicine in the branch of Human Genetics, in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other university.

……...….

……...day of……., 2006.

DEDICATION

This dissertation is dedicated to the three most special people in my life, my mother, father and sister who have given me constant love, encouragement and guidance. Together they have been my pillars of strength and have helped me achieve my goals and dreams. For all those reasons and more, I am eternally grateful.

ABSTRACT

Fetal alcohol spectrum disorder (FASD) encompasses a range of conditions induced by prenatal alcohol exposure. Fetal alcohol syndrome (FAS) is the most severe of these conditions. FAS is characterised by discriminating facial features along with growth deficiencies and central nervous system abnormalities.

FASD is a growing concern in South Africa, particularly in the Northern and Western Cape Provinces. In the Northern Cape, astounding prevalence rates of 122 and 73.8 per 1000 school entry children have been established for the towns of De Aar and Upington respectively.

Studies involving twin concordance research and animal models have indicated that there is a genetic influence contributing towards FAS susceptibility in individuals. FAS is considered a complex disease whereby both genetic and environmental factors interact in disease pathogenesis. For this reason a case-control study involving the investigation of appropriate candidate genes was conducted.

The neuronal migration pathway in the developing brain is targeted by prenatal alcohol exposure. The *astrotactin* (*ASTN*) and *reelin* (*RELN*) genes were selected for investigation based on their fundamental role in neuronal migration. A FAS case-control study involving 45 cases and 112 controls was conducted on the Northern Cape Coloured population.

Four single nucleotide polymorphisms (SNPs) including missense and non-coding variants were selected within *ASTN* and four missense SNPs were selected within *RELN*. The study aimed to determine the genotype and allele frequencies of the variants within the case and control groups and to assess whether any association between the gene variants and the predisposition to FAS existed. Statistical analyses indicated a significant genotypic association (P= 0.049) between *RELN's* rs607755 marker; the C/T genotype was more likely to be found amongst controls thus inferring a possible protective effect

against FAS. A logistic regression model supported the above association by indicating the C/T genotype as being independently significant (P= 0.026).

The most limiting factor of this study was the small sample size and consequent lack of power to detect genes with minor effects. It would therefore be suggested that the study be repeated once a larger sample size has been established. A larger sample size would increase the chances of detecting true associations between genes of minor effect and FAS, thus minimising false-positive associations from arising.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank the following people:

- My supervisor Professor Michèle Ramsay for her advice and encouragement regarding my project over the past two years.

- My co-supervisor Professor Denis Viljoen for allowing me to sit in on the examinations of the FAS children during the field trips to Upington and for sharing his expertise on the clinical aspects of the disease with me.

- Dr. Lize van der Merwe for her expertise in Biostatistics. I am so grateful for her help and input on the statistical analyses of my results.

- Zané Lombard, my colleague and friend, for her constant help, friendship and words of encouragement over the past two years.

- Dhamari Naidoo for her help and advice and for the many laughs shared between her, Zané and myself over our several lunch and coffee sessions.

- Mom, Dad and Casey for their endless love and support throughout the good times and the bad times. Thank you for never failing to believe in me and for helping me believe in myself. I could not have done it without the three of you!

TABLE OF CONTENTS

Page

DECLARATION ii
DEDICATION iii
ABSTRACT iv
ACKNOWLEDGEMENTS vi
TABLE OF CONTENTS vii
LIST OF FIGURES x
LIST OF TABLES xiii
LIST OF ABBREVIATIONS xiv

CHAPTER 1: INTRODUCTION 1

1.1 Fetal Alcohol Syndrome 1
1.1.1 An Overview 1
1.1.2 Clinical Features 3
1.1.2.1 Growth Deficiency 4
1.1.2.2 Central Nervous System Dysfunction 4
1.1.2.3 Facial Dysmorphism 5
1.1.2.4 Other Anomalies of FAS 6
1.1.3 Diagnostic Criteria of FAS 6
1.1.3.1 Additional Terminology 8
1.1.4 Other Maternal Risk Factors 8
1.2 Alcohol Consumption in South Africa 10
1.2.1 FAS Prevalence: South Africa versus Other Countries 11
1.3 Neurological Effects of Prenatal Ethanol Exposure 15
1.3.1 Areas of the Brain Susceptible to Alcohol Induced Damage 17
1.3.2 Structural Damage 18
1.3.3 Neuropsychological Impairments 20
1.3.3.1 Learning and Memory 20
1.3.3.2 Executive Functioning 21
1.3.3.3 Motor Control 21
1.3.3.4 Psychosocial and Behavioural Discrepancies 22
1.3.4 Interference at the Cellular Level 22
1.4 Neuronal Migration 24
1.4.1 The Effect of Alcohol on Neuronal Migration 25
1.5 The Genetics Behind Fetal Alcohol Syndrome 25
1.5.1 Twin Concordance Studies 26
1.5.2 Animal Models 27
1.6 Complex Diseases 29
1.6.1 Linkage Studies 30
1.6.2 Association Studies 31
1.6.2.1 Case-control Approach 33
1.6.2.2 Family-based Approach 34
1.6.3 Previous Candidate Gene Association Studies on FAS 35
1.7 The *Astrotactin* Gene 37
1.7.1 *Astrotactin* as a Candidate Gene for FAS 38

1.8 The *Reelin* Gene 39
1.8.1 *Reelin* as a Candidate Gene for FAS 42
1.9 Hypothesis 43
1.10 Aims and Objectives 43

CHAPTER 2: SUBJECTS AND METHODS 44

2.1 Subject Recruitment 44
2.2 DNA Extraction 45
2.3 DNA Quantification and Storage 45
2.4 Candidate Gene Choice 46
2.5 Selection of Single Nucleotide Polymorphisms 46
2.5.1 *Astrotactin* SNPs 47
2.5.2 *Reelin* SNPs 48
2.6 Primer Design 49
2.7 Techniques Employed 50
2.7.1 Polymerase Chain Reaction 50
2.7.2 Restriction Fragment Length Polymorphisms 51
2.7.3 Amplification Refractory Mutation System 51
2.7.4 Agarose Gel Electrophoresis 52
2.7.5 DNA Sequencing 54
2.7.6 Standard Laboratory Procedures 55
2.8 Analysis of *Astrotactin* and *Reelin* SNPs 56
2.8.1 *Astrotactin* SNPs 56
2.8.1.1 rs6413830 57
2.8.1.2 rs12118933 58
2.8.1.3 rs2228956 59
2.8.1.4 rs11587640 61
2.8.2 *Reelin* SNPs 62
2.8.2.1 rs607755 62
2.8.2.2 rs3025963 64
2.8.2.3 rs3025966 65
2.8.2.4 rs362691 66
2.9 Statistical Analyses 69
2.9.1 Allele and Genotype Frequencies 69
2.9.2 Hardy-Weinberg Equilibrium 69
2.9.3 Testing for Association 70
2.9.3.1 Fisher's Exact Test 70
2.9.3.2 Haplotype Analysis 70
2.9.3.3 Logistic Regression 71
2.9.4 Linkage Disequilibrium 72
2.9.5 Post-test Power Calculations 73

CHAPTER 3: RESULTS 74

3.1 Genotyping of *Astrotactin* SNPs 74
3.1.1 rs6413830 74
3.1.2 rs12118933 77
3.1.3 rs2228956 80

3.1.4 rs11587640 83
3.2 Genotyping of *Reelin* SNPs 86
3.2.1 rs607755 86
3.2.2 rs3025963 89
3.2.3 rs3025966 91
3.2.4 rs362691 93
3.3 Statistical Analyses 98
3.3.1 Genotype and Allele Frequencies 98
3.3.2 Hardy-Weinberg Equilibrium 100
3.3.3 Association Testing 101
3.3.4 Haplotype Analysis 102
3.3.5 Logistic Regression 103
3.3.6 Linkage Disequilibrium 105
3.3.6.1 *Astrotactin* Markers 105
3.3.6.2 *Reelin* Markers 106

CHAPTER 4: DISCUSSION 107

4.1 Project Findings 107
4.1.1 Laboratory Results 107
4.1.2 Comparison of Genotype and Allele Frequencies in Different Populations 108
4.1.3 Impact of Admixture on Identifying Genetic Susceptibility Factors 110
4.1.4 The Effect of Small Sample Size 111
4.1.5 Inferred Associations Between the Markers and Disease Susceptibility 112
4.1.6 Interaction Between the Markers 112
4.1.7 Decision Not to Correct for Multiple Testing 113
4.2 Project Limitations 115

CHAPTER 5: CONCLUSION AND FUTURE PROSPECTS 117

REFERENCES 120

APPENDIX A Ethics Clearance Certificates 128

APPENDIX B SNP Diagrams and Genotyping Results 131

APPENDIX C DNA Ladders and Laboratory Protocols 138

APPENDIX D Reagents 144

LIST OF FIGURES

Page

Figure 1.1: The typical facial features associated with fetal alcohol syndrome 5
Figure 1.2: An index family from the De Aar FAS study in the Northern Cape 14
Figure 1.3: The various stages of organ development and the critical periods at which the fetus is most susceptible to birth defects induced by teratogens 16
Figure 1.4: The areas of the brain most seriously affected by prenatal alcohol exposure 17
Figure 1.5: The difference in size and structure between the brain of a six week old FAS child and an unaffected child of the same age 20
Figure 1.6: Dizygotic twin sisters 27
Figure 1.7: The similarities in the clinical presentation of FAS in the mouse versus the human 28
Figure 1.8: The chromosomal location of the *astrotactin* gene, 1q25.2 37
Figure 1.9: The chromosomal location of the *reelin* gene, 7q22 39
Figure 1.10: Schematic illustration of normal versus reelin deficient cortical development 41
Figure 2.1: A map of South Africa indicating the geographical positioning of Upington and De Aar 44
Figure 2.2: A schematic representation of the selected *ASTN* SNPs orientated on the reverse strand 48
Figure 2.3: A schematic representation of the chosen *RELN* SNPs orientated on the reverse strand 49
Figure 2.4: The three possible genotypes at rs6413830 following *Alu*I digestion 58
Figure 2.5: The expected genotype results at rs12118933 following *Msp*I digestion 59
Figure 2.6: The expected band pattern on a gel of the genotypes at rs2228956 following *Hha*I digestion 60
Figure 2.7: The three predicted genotypes at rs11587640 generated by *Sbf*I digestion 61
Figure 2.8: The expected fragments after *Pvu*II digestion of the PCR products containing rs607755 63

Figure 2.9: Expected band sizes for the different rs3025963 genotypes generated by *Mva*12691 digestion 65
Figure 2.10: The three possible genotypes at rs3025966 resulting from *Mva*12691 digestion 66
Figure 3.1: 203 bp PCR products encompassing *ASTN's* rs6413830 SNP run on a 3% agarose gel 74
Figure 3.2: Digestion of rs6413830 PCR products with *Alu*I viewed on a 4% agarose gel 75
Figure 3.3: Electropherograms (forward strands) confirming genotyping results obtained from *Alu*I enzyme digestion of rs6413830 76
Figure 3.4: PCR products of 593 bp encompassing *ASTN's* rs12118933 SNP run on a 3% agarose gel 77
Figure 3.5: Digestion of *ASTN's* rs12118933 PCR products with *Msp*I, samples viewed on a 4% agarose gel 78
Figure 3.6: Electropherograms confirming genotyping results for rs12118933 79
Figure 3.7: 381 bp PCR products encompassing *ASTN's* rs2228956 SNP run on a 3% agarose gel 80
Figure 3.8: Digestion of *ASTN*'s rs2228956 PCR products with *Hha*I viewed on a 4% agarose gel 81
Figure 3.9: Sequencing results confirming the rs2228956 genotypes 82
Figure 3.10: PCR products of 172 bp in size encompassing *ASTN's* rs11587640 SNP run on a 3% agarose gel 83
Figure 3.11: 4% agarose gels showing digestion of *ASTN's* rs11587640 PCR products with *Sbf*I 84
Figure 3.12: 529 bp PCR products encompassing *RELN's* rs607755 SNP run on a 3 % agarose gel 86
Figure 3.13: Digestion of *RELN*'s rs607755 PCR products with *Pvu*II, products viewed on a 4% agarose gel 87
Figure 3.14: Sequencing results confirming the three genotypes obtained for rs607755 88
Figure 3.15: PCR products of 285 bp in size encompassing *RELN's* rs3025963 SNP run on a 3% agarose gel 89

Figure 3.16: Digestion of *RELN's* rs3025963 PCR products with *Mva*12691, products viewed on a 4% agarose gel 90
Figure 3.17: 226 bp PCR products encompassing *RELN's* rs3025966 SNP run on a 3% agarose gel 91
Figure 3.18: Digestion of *RELN's* rs3025966 PCR products with *Mva*12691, products viewed on a 4% agarose gel 92
Figure 3.19: PCR products from *RELN's* rs362691 ARMS system shown on a 3% agarose gel 93
Figure 3.20: PCR products of 441 bp generated for the purpose of rs362691 sequencing, run on a 3% agarose gel 94
Figure 3.21: Electropherograms of the forward and reverse strands of control sample NPC005, C/C homozygote for rs362691 95
Figure 3.22: Electropherograms of rs362691 C/G heterozygotes 96
Figure 3.23: A heat map indicating the degree of LD between the *ASTN* SNPs 105
Figure 3.24: A heat map indicating the degree of LD between the *RELN* SNPs 106
Figure B1: Diagrammatic legend explaining Ensembl's annotated gene diagrams 131
Figure B2: The positioning of all recorded SNPs in the *ASTN* gene on chromosome one 132
Figure B3: The positioning of all recorded SNPs in the *RELN* gene on chromosome seven 133

LIST OF TABLES

Page

Table 1.1: Centers for Disease Control criteria for the diagnosis of FAS with confirmed or unknown maternal drinking history 7
Table 1.2: The effects of alcohol on the developmental stages of the human CNS 23
Table 1.3: The advantages and disadvantages of linkage studies 31
Table 1.4: The advantages and disadvantages of case-control studies 34
Table 2.1: SNPs in the *ASTN* gene chosen for investigation 47
Table 2.2: SNPs in the *RELN* gene chosen for investigation 48
Table 3.1: Genotype and allele frequencies for *ASTN* markers: cases and controls 98
Table 3.2: Genotype and allele frequencies for *RELN* markers: cases and controls 99
Table 3.3: *P*-values for the HWE test in cases and controls for *ASTN* 100
Table 3.4: *P*-values for the HWE test in cases and controls for *RELN* 100
Table 3.5: *P*-values for association tests for cases versus controls at the allelic and genotypic level for *ASTN* 101
Table 3.6: *P*-values for association tests for cases versus controls at the allelic and genotypic level for *RELN* 101
Table 3.7: Comparison of FAS cases versus controls: Haplotype Scores, *P*-values and Haplotype Frequencies (hf) for *ASTN* 102
Table 3.8: Comparison of FAS cases versus controls: Haplotype Scores, *P*-values and Haplotype Frequencies (hf) for *RELN* 103
Table 3.9: Final logistic regression model for FAS cases versus controls 104
Table 4.1: Comparison of genotype and allele frequencies in three different populations for some of the *ASTN* and *RELN* SNPs investigated 109
Table B1: Genotyping data generated for the *ASTN* and *RELN* SNPs investigated 134

LIST OF ABBREVIATIONS

A	adenine
aa	amino acid
ADH	alcohol dehydrogenase
ADHD	attention deficit hyperactivity disorder
AIC	Akaike information criterion
ALD	admixture linkage disequilibrium
ALDH	aldehyde dehydrogenase
ARBD	alcohol related birth defects
ARMS	amplification refractory mutation system
ARND	alcohol related neurological defects
ASTN	*astrotactin* gene
BMI	body mass index
BSA	bovine serum albumin
C	cytosine
CDC	Centers for Disease Control
CNS	central nervous system
cSNP	coding single nucleotide polymorphism
CYP2E1	cytochrome P450 2E1
ddH_20	double distilled water
DNA	deoxyribose nucleic acid
dNTPs	deoxynucleotide triphosphates
ddNTPs	dideoxynucleotide triphosphates
DZ	dizygous
EDTA	ethylene diamine tetraacetic acid
EGF	epidermal growth factor
f	frequency
FAE	fetal alcohol effects
FARR	Foundation for Alcohol Related Research
FAS	fetal alcohol syndrome
FASD	fetal alcohol spectrum disorder

FNIII	fibronectin III
G	guanine
HAS	high-alcohol-sensitive
HCl	hydrochloric acid
hf	haplotype frequency
HWE	Hardy-Weinberg equilibrium
ID	identity
i.e.	that is to say
IgG	immunoglobulin G
IOM	Institute of Medicine
IQ	intelligence quotient
LAS	low-alcohol-sensitive
LD	linkage disequilibrium
$MgCl_2$	magnesium chloride
MRI	magnetic resonance imaging
MRC	Medical Research Council
MZ	monozygous
NaOH	sodium hydroxide
NCBI	National Center for Biotechnology Information
NHLS	National Health Laboratory Service
NIAAA	National Institute for Alcohol Abuse and Alcoholism
NIDA	National Institute on Drug Abuse
NPC	code for Northern Cape controls
NPF	code for Northern Cape FAS cases
PCR	polymerase chain reaction
P-value	probability value
RELN	*reelin* gene
RFLP	restriction fragment length polymorphism
SNP	single nucleotide polymorphism
SPECT	single photon emission computed tomography
T	thymine
TBE	tris borate EDTA

TDT	transmission disequilibrium test
UCSC	University of Santa Cruz
USA	United States of America
UV	ultraviolet

UNITS AND SYMBOLS

$	dollar
%	percentage
µl	micro litre
bp	base pair
Da	Daltons
D’	D prime (LD unit)
g	gram
kb	kilobase
l	litre
mg	milligram
mins	minutes
ml	millilitre
ng/µl	nanograms per micro litre
nm	nanometres
oz	ounces
rpm	rates per minute
secs	seconds
V/cm	volts per centimetre
^{0}C	degrees Celsius
3’ UTR	3 prime untranslated region

AMINO ACID CODES

A alanine	**C** cysteine	**D** aspartic acid	**E** glutamic acid
F phenylalanine	**G** glycine	**H** histidine	**I** isoleucine
K lysine	**L** leucine	**M** methionine	**N** asparagine
P proline	**Q** glutamine	**R** arginine	**S** serine
T threonine	**V** valine	**W** tryptophan	**Y** tyrosine

CHAPTER 1: INTRODUCTION

1.1 FETAL ALCOHOL SYNDROME

1.1.1 An Overview

The term "fetal alcohol syndrome" (FAS) was first used in 1973 by Jones and colleagues whilst studying the offspring of severely alcoholic mothers (Jones et al. 1973). This however, was not the first study conducted on children of alcoholic mothers. In France, five years prior to the work carried out by Jones et al., Lemoine and colleagues performed a similar study and observed the same dysmorphic symptoms in the offspring of alcoholic mothers (reviewed in Viljoen, 1991).

Growth retardation, central nervous system (CNS) damage and distinguishing craniofacial features are the most commonly reported abnormalities associated with FAS (Viljoen, 1999; Warren and Li, 2005). Alcohol is referred to as a "teratogen" as it has the ability to cause birth defects (Goodlett and Horn, 2001). Of all the features of FAS, the teratogenicity of *in utero* alcohol exposure on the developing central nervous system appears to be the most severe and permanent and thus, consumption of alcohol by pregnant mothers has been described as being the most common cause of nonhereditary mental retardation amongst several communities worldwide. In South Africa, FAS is a severe and escalating problem (Viljoen, 1991; Viljoen, 1999).

Ethanol can freely cross the placenta, which results in the blood alcohol levels of the fetus being comparable to those of the mother (Thomas and Riley, 1998). Knowledge about the mother's drinking habits is of importance during diagnosis and the severity of the disease is proportional to the amount of alcohol consumed. No 'safe' level of drinking during pregnancy has been reported. Pregnant women are advised to abstain from all alcohol throughout the nine months of pregnancy (Viljoen, 1991). An average of more than one drink (0.5 oz) per day or in the case of binge drinking, more than five drinks per occurrence, has been described as being potentially harmful to an unborn fetus (Sokol et al. 2003).

Binge drinking has been shown to be more damaging towards the CNS than drinking a constant amount of alcohol over a long time span. Binge drinking results in a rapid increase in blood alcohol levels as opposed to a gradual increase seen with constant drinking over a long period of time (reviewed in Guerri, 2002).

Heavy drinking has been described as being six or more drinks or more than 100 ml of absolute alcohol per day (Viljoen, 1991). It is well recognised that heavy alcohol consumption at any time during a pregnancy can cause serious problems involving CNS development and cognitive functioning.

FAS is a severe and irreversible disease, which reduces the chances of the individual leading a normal life. Serious social and medical implications arise from FAS. Affected individuals have learning disabilities and are therefore compromised with regards to education. The disease itself can pose huge expenses upon the state in the form of medical care, special educational support and lost productivity. One of the most recent cost estimates from a National Institute on Drug Abuse (NIDA) and National Institute on Alcohol Abuse and Alcoholism (NIAAA) study indicated that the cost of FAS to the United States was more than $4 billion in 1998. It was also estimated that the lifetime cost for each affected individual was $2 million in the year 2002 (Lupton et al. 2004). In South Africa the effective costs would be lower since medical services and remedial education are not available for the majority of affected individuals.

Certain studies have focused on the adverse life outcomes that occur amongst individuals affected by prenatal alcohol exposure. Drug and alcohol abuse appear to be problematic amongst FAS individuals. Another well documented adverse life outcome is disruptive and criminal behaviour. Individuals prenatally exposed to alcohol tend to be disorderly in a schooling environment and often find themselves in trouble with the law (Chaudhuri, 2000).

Studies like those mentioned above emphasise the importance of an early diagnosis when prenatal alcohol exposure has occurred in order to effectively cater and care for the affected child's needs (Streissguth et al. 2004). Not only is FAS a severe medical

concern, it is a social one too (Chaudhuri, 2000) that requires early health education and intervention. A sound knowledge of the diagnostic criteria and a better understanding of the molecular basis of the disease can together help manage and hopefully prevent increasing numbers of cases worldwide.

1.1.2 Clinical Features

The severity of clinical symptoms following prenatal alcohol exposure can differ significantly based on the amount of alcohol consumption and the gestational timing at which the alcohol was consumed. The differences in these two factors can result in varying degrees of clinical manifestations. Fetal alcohol syndrome is the most severe of fetal alcohol diagnoses which all fall under the broader term of fetal alcohol spectrum disorder (FASD). In 1978 researchers suggested that drinking during the month preceding recognition of pregnancy appeared to produce the most significant relationship between maternal alcohol consumption and adverse fetal outcome (Hanson et al. 1978).

Heavy drinking during the first trimester of pregnancy affects the facial and structural features of the unborn fetus; drinking during the second trimester appears to increase the rates of spontaneous abortions and third trimester drinking has a significant effect upon growth. However, because the brain is continuously developing throughout the gestation period, drinking at any time during pregnancy can cause neurodevelopmental, intellectual and behavioural abnormalities (reviewed in May et al. 2004).

Clinically, FAS presents with distinct morphological abnormalities involving the head, face, limbs and other organs such as the heart, liver and eyes. Growth and varying degrees of mental retardation are other major defining characteristics of the disease (Jones et al. 1973; Chaudhuri, 2000).

1.1.2.1 Growth Deficiency

Pre and postnatal growth deficits are widely observed in children exposed to alcohol *in utero*. Height and weight are discriminating factors as well as head circumference. FAS individuals tend to be below the tenth percentile for any measurement regarding weight, height and head circumference for gestational age (Kvigne et al. 2004). Microcephaly is a distinguishing feature of FAS, whereby the head circumference is below the norm for the individuals' age. Microcephaly is indicative of deficient brain growth (Clarren and Smith, 1978).

1.1.2.2 Central Nervous System Dysfunction

The developing CNS is highly sensitive and vulnerable to the harmful effects of alcohol. Of all the classic features, mental retardation is the most consistent and detrimental factor of FAS even though the degree to which it occurs can vary. In 1980, it was reported that FAS is the third most common disorder in which mental retardation is a principle factor (Shaywitz et al. 1980). An overall lowering of intelligence is commonly reported in those affected with a mean intelligence quotient (IQ) of 65 (Clarren and Smith, 1978; Viljoen, 1999).

Along with a below average IQ, FAS children are inclined to have behavioural problems, developmental delays, speech and language impediments and learning disabilities. The neuropsychological implications of prenatal alcohol exposure are described in more detail later on in this chapter. With regards to behaviour, FAS children are usually hyperactive with short attention spans and commonly suffer from attention deficit disorder (Kvigne et al. 2004). Adolescents and adults with FAS regularly demonstrate poor socialisation and communication skills (Mattson et al. 2001).

1.1.2.3 Facial Dysmorphism

The facial abnormalities are the most pronounced indicators of the disease; typically the middle third of the face is affected. A classic case would present with narrow palpebral fissures (eye openings), a long upper lip with an indistinct philtrum (groove between the top lip and the nose) and a thin vermilion border, epicanthic folds (skin folds over the eyes), a flattened nasal bridge and a short upturned nose as well as micrognathia (abnormal smallness of the jaw, especially of the mandible) (Viljoen, 1999). There are varying degrees of the clinical features in individuals and in particular, the severity of the facial features tend to diminish with increased age (Chaudhuri, 2000). Figure 1.1 illustrates the diagnostic facial features as well as other associated facial features of FAS.

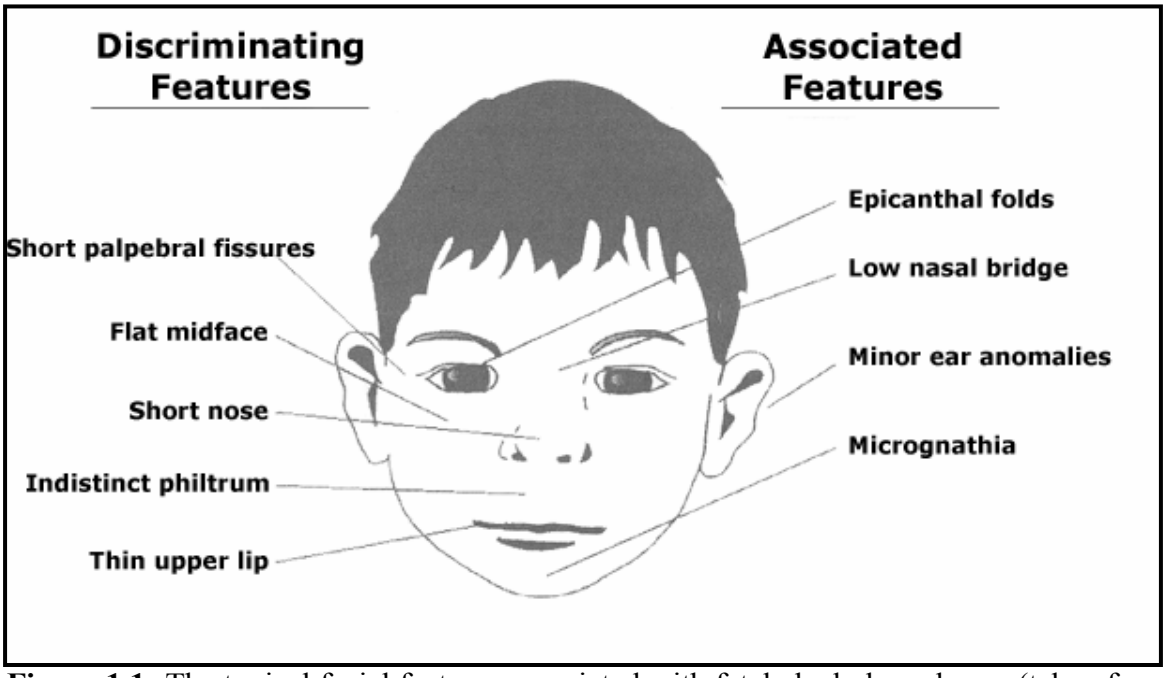


Figure 1.1: The typical facial features associated with fetal alcohol syndrome (taken from Minnesota Department of Health website).

1.1.2.4 Other Anomalies of FAS

Although not always observed, cardiac, skeletal and other organ manifestations have frequently been reported in FAS subjects (Kvigne et al. 2004). Cleft lip and palate, neural tube defects, genital abnormalities, renal and liver defects, heart murmurs, fifth finger hypoplasia, abnormal palmer creases especially a "hockey stick" configuration of the upper palmer crease and a variety of radiological alterations are some of the numerous minor deficits resulting from exposure to alcohol prenatally (Clarren and Smith, 1978).

1.1.3 Diagnostic Criteria of FAS

The diagnosis of FAS is easier to make when an individual is between the ages of four and fourteen years than at other ages. It is during these years that the features of FAS are most apparent. FAS diagnosis on newborns or adults is more complicated, as the features are either under developed (newborns) or have disappeared or become less distinct (adults) (Lupton et al. 2004).

Diagnosis of fetal alcohol syndrome is based upon three major proposed criteria (Sokol and Clarren, 1989). The diagnostic criteria are as follows:

1. Prenatal and/or postnatal growth deficiency (weight or length or both, below the tenth percentile for the gestational age).
2. Central nervous system irregularities, including developmental delays, neurological problems, intellectual impairment and structural abnormalities.
3. Characteristic facial features including short palpebral fissures, a thin upper lip and an elongated, flattened midface as well as an indistinct philtrum.

Together with the existence of the above features, confirmation of heavy maternal drinking (more than 90-100 mg of absolute alcohol per week) during pregnancy should ideally be obtained in order to make a successful diagnosis of FAS (Viljoen, 1999). Table 1.1 illustrates the necessary characteristics required for a sound diagnosis of FAS.

Table 1.1: Centers for Disease Control criteria for the diagnosis of FAS with confirmed or unknown maternal drinking history (modified from Welch-Carre, 2005)

Facial Features	Growth Deficiencies	CNS Abnormalities
All 3 must be present: • Smooth philtrum • Thin vermilion • Small palpebral fissures	Height, weight, or both, at or below the 10th percentile, documented at any one point in time.	**Must have 1 of the following:** **1. Structural** Head circumference at or below the 10th percentile, adjusted for age and sex. **2. Neurological** Neurological problems not due to postnatal insult or fever, or other soft neurological sign outside normal limits. **3. Functional** Below average performance as evidenced by: Global cognitive or intellectual deficits representing multiple domains of deficit with performance below the 3rd percentile. Functional deficits below the 16th percentile in at least 3 of the following areas: • Cognitive or developmental deficits • Executive functioning deficits • Motor functioning delays • Problems with attention and hyperactivity • Social skills • Sensory problems, language and memory deficits

1.1.3.1 Additional Terminology

Fetal alcohol spectrum disorder (FASD) is a term encompassing all conditions arising from prenatal alcohol exposure. Fetal alcohol syndrome falls at the severe end of the spectrum (Hoyme et al. 2005). Not all episodes of drinking during pregnancy result in full-blown FAS affected infants. There are other categories a child can be classified into depending on which features are or are not present. As defined by the Institute of Medicine (IOM), 1996, FASD refers collectively to FAS, partial FAS, alcohol related birth defects (ARBD) and alcohol related neurodevelopmental disorder (ARND) (Jacobson and Jacobson, 2002).

Partial FAS is used to describe children who display some of the facial features of FAS as well as some of the physical growth or neurodevelopmental abnormalities. ARBD is a term used to describe an individual with confirmed maternal drinking during pregnancy who has some of the structural organ abnormalities associated with prenatal alcohol exposure such as cardiac, renal and skeletal defects. ARND refers to children who have confirmed prenatal alcohol exposure but do not have the phenotypic criteria necessary for a FAS diagnosis; instead they display neurodevelopmental deficiencies typical to FAS children (Jacobson and Jacobson, 2002).

1.1.4 Other Maternal Risk Factors

It is clearly understood that heavy drinking during pregnancy is the primary risk factor for FAS. However, the pattern of drinking, especially heavy episodic (binge) drinking is particularly associated with the disease (May et al. 2004). Based on studies carried out on mothers of FAS children, other maternal risk factors have been pinpointed in association with the high prevalence rate of FAS in certain populations.

A study carried out by Viljoen et al. (2000) focused on mothers of FAS children from a community of relatively low socioeconomic status in the Western Cape of South Africa. Mothers of FAS children had heavier drinking patterns, began consuming alcohol at an earlier age and had higher rates of drinking and alcohol abuse within their extended

families when compared to control mothers. It appeared as though the majority of mothers of FAS children had mothers themselves (maternal grandmothers to FAS children) who drank heavily. This differed significantly from the control group where maternal grandmothers were considered “non-drinkers” or “light drinkers.”

Tobacco smoking during pregnancy was far higher in case mothers compared to control mothers. Besides these factors, this study found significant differences between the case and control individuals regarding education and religiosity. Mothers of FAS children had received notably less formal education, calculated to be about 28% less than control mothers and were less involved in religious practices such as church attendance and prayer (Viljoen et al. 2002).

Kvigne et al. (2003) carried out a study in the United States, on Northern Plains Indian mothers of FAS children. Like the results generated by Viljoen et al. (2002), this investigation also illustrated that the mean educational level of case mothers was 9.8 years which was significantly lower than the 11.0 years calculated for control mothers. Case mothers in this investigation had far higher rates of heavy drinking compared to control mothers. Furthermore, most case grandmothers had alcohol use documented in their medical records whereas only one half of control grandmothers had alcohol use recorded in their files.

Kvigne et al. (2003) also found that mothers of FAS children were slightly older than the control mothers. In addition to the other findings the investigators found that within medical records, case mothers had mental health problems, especially depression, injuries (including alcohol related ones), sexual abuse and alcohol related problems, such as cirrhosis, recorded significantly more often than in the medical records of control mothers (Kvigne et al. 2003).

May et al. (2004) indicated that maternal age at time of pregnancy is definitely a variable of maternal risk for FAS. Older women, classified by Viljoen et al. (2001) as being greater than 27 years, are at higher risk of producing FAS children, which could be linked to more years of drinking compared to younger women. It has also been suggested that nutrition

and body size (Body Mass Index (BMI)) could be important risk factors for FAS. The poorer the nutrition of the individual, the lower their body size and therefore BMI. Smaller individuals have less body mass to assist in metabolising the consumed alcohol, which might make them more susceptible to producing FAS offspring (May et al. 2004).

These studies illustrate that certain social and maternal factors have been seen to accompany the presence of FAS within a family. Sufficient education seems to be a key factor that is lacking in mothers of FAS children and this probably goes hand-in-hand with low socioeconomic status. Poor nutrition and small body mass are possibly closely associated with poorer living conditions due to low socioeconomic status. Living within a family that drinks heavily, especially having a mother that drinks, appears to negatively influence the drinking habits of mothers with FAS children in that they are more likely to start drinking at an early age and become heavy drinkers themselves.

1.2 ALCOHOL CONSUMPTION IN SOUTH AFRICA

In certain regions of South Africa, especially in wine-growing areas, alcohol abuse and consequently FAS, is a severely escalating problem. Studies conducted in South Africa have focused on the Western and Northern Cape Provinces as well as the Gauteng Province (Croxford and Viljoen, 1999).

The Western Cape, in particular, is renowned for its grape-growing and wine production and most reports on alcohol consumption and FAS have focused on communities within this province. The predominant ethnic group of the Western Cape is Coloured, a population that arose over 300 years ago from admixture between the original local inhabitants, the Khoisan, and colonisers of European decent as well as local black Africans that originally migrated from Central Africa (Viljoen et al. 2001). The demographics for the Western Cape Province population in the year 2000 were 57% Coloured, 18% Black, 24% White and 1% other (May et al. 2000). Alcohol consumption and FAS prevalence are relatively high amongst the Coloured population (Viljoen et al. 2001).

Traditionally, farm workers were partially paid for their labour in the form of wine, which was consumed daily. This method of part remuneration of farm labourers was known as the "dop" system (May et al. 2000). Despite being outlawed today, the origins of excessive alcohol consumption and abuse amongst farm workers of wine-growing regions, have their roots in the fore mentioned "dop" system. Alcohol still appears to be valued and favoured, and problematic drinking behaviour is widespread amongst the local population of workers (Viljoen et al. 2002).

Along with the legacy of the "dop" system, another attribute to excessive alcohol consumption in certain regions within South Africa is the ease at which alcohol can be obtained. Accessibility to alcohol has increased in the form of illegal, informal bars known as "shebeens" where cheap and inferior quality alcohol can be purchased (May et al. 2000).

A study carried out in the year 2000 reported 87% of farm workers in the Western Cape region as being problem drinkers. For many communities in South Africa where poverty is prevalent and where drinking is customary, alcohol is still a major form of recreation (Viljoen et al. 2002).

1.2.1 FAS Prevalence: South Africa versus Other Countries

One of the early reports on FAS in South Africa was written in 1978, where four infants born to alcoholic mothers in Cape Town were examined and clinical features consistent with those reported a few years earlier by Jones et al. (1973) were noted (Beyers and Moosa, 1978). FAS has therefore been reported in South Africa for a number of years but it was not until May et al. (2000) carried out an epidemiological study that the actual prevalence of the disease was noted.

The United States prevalence rate for FAS ranges from 0.5 - 2.0 per 1000 births (May et al. 2004). Other studies have investigated the prevalence of FAS in various regions of certain countries. A study carried out in Roubaix, France, indicated an estimated incidence

of 2.3 per 1000 births and two studies conducted in Stockholm and Göteborg, Sweden, revealed 2.51 and 1.61 per 1000 births estimates respectively (Sampson et al. 1997).

An investigation into alcohol consumption by pregnant women in three regions of the Western Cape revealed that 42.8% of pregnant women attending antenatal clinics admitted to consuming alcohol during their pregnancy which indicated that the FAS incidence in this province was expected to be high (Croxford and Viljoen, 1999). In the year 2000, May et al. reported a community of low socioeconomic status in the Western Cape Province of South Africa as having the highest reported prevalence rate of FAS worldwide, a rate of 39.2 – 46.4 per 1000 births (May et al. 2000).

Research undertaken by the Foundation for Alcohol Related Research (FARR) in South Africa has indicated that the FASD prevalence rate in Wellington, a town in the Western Cape, is actually rising each year. The rates of FASD were established in school entry children during the years 1997, 1999 and 2001. Rates of 46, 75 and 88 per 1000 children were reported for the respective years (Viljoen, 2005). These studies illustrate the effects of the drinking patterns developed amongst communities within the wine-growing regions; alcohol consumption is severely high and as a result FAS is endemic.

The FAS prevalence rates mentioned above are exceptionally high. However, researchers were unsure whether the rate amongst other communities of low socioeconomic status in South Africa would indicate similar findings or whether the results were exclusively associated with the wine-growing regions of the Western Cape (Centers for Disease Control, 2003). Therefore, in 2001, a study was conducted on four communities around Johannesburg in the Gauteng Province where wine production is absent. Methods similar to those employed in the Western Cape study were used and results revealed a FAS prevalence rate of 19 per 1000. Despite not being as high as the rate obtained for the Western Cape, 19 per 1000 births is still extremely high when compared to figures from the rest of the world. This study emphasised that although FAS is endemic amongst some communities in the wine-growing regions it is not exclusive to them; instead it appears to be a serious public health problem in South Africa (Centers for Disease Control, 2003).

Currently studies are underway in the Northern Cape Province of South Africa, particularly in the towns of Upington and De Aar where FAS is significantly present. The majority of the inhabitants of these areas are also Coloured with some Black individuals and a few White residents. Even though grape-growing and wine production is not as prevalent as in the Western Cape, the majority of the inhabitants in these two towns are poverty stricken and culturally alcohol plays a major role in recreation. An astounding FASD prevalence rate of 122 per 1000 school entry children has been established for the town of De Aar, the highest rate ever reported. The Upington epidemiology study has revealed a prevalence rate of 73.8 per 1000 school entry children (Viljoen, 2005).

The photographs seen in figure 1.2 were taken as part of the Northern Cape De Aar project. The extent to which FAS is found within one South African family is illustrated in (a), the three children standing in the photo were all diagnosed with FAS. According to the mother, she stopped drinking during her pregnancy with the fourth child (seen in her arms), however, this child did not fall within the study group and was therefore not examined for an accurate diagnosis to be made. The front view and profile of one of the affected children indicating the facial features of FAS can be seen in (b) and (c) respectively.







Figure 1.2: An index family from the De Aar FAS study in the Northern Cape. All three children standing are affected with FAS, the mother admitted to dinking throughout all three pregnancies **(a)**. The front **(b)** and side **(c)** profiles of one of the FAS affected children. Some discriminating facial features include a long smooth upper lip and smooth philtrum as well as short palpebral fissures. Note the flattened nasal bridge and relatively small mandible (photos courtesy of Leigh-Anne Fourie, FARR, with permission from the family).

1.3 NEUROLOGICAL EFFECTS OF PRENATAL ETHANOL EXPOSURE

Human gestation is separated into two main phases, the embryonic stage, which is up to eight weeks of gestation, and the fetal stage, which is from eight weeks up until birth. It is during the embryonic period in particular that abnormalities are readily induced by certain drugs within the mother's bloodstream or through her diet (Michaelis and Michaelis, 1994).

During embryological and fetal development different organs develop intermittently (figure 1.3). As described previously, agents that are able to cause birth defects are referred to as teratogens. Most teratogens tend to be selective as to which organs they affect. This selectivity is dependent upon factors such as timing of exposure of the embryo to the teratogen, the quantity of the teratogen taken by the pregnant mother and the sensitivity of each developing organ's cells to the harmful effect of the teratogen (reviewed in Michaelis and Michaelis, 1994).

Based on figure 1.3, the heart starts developing from the middle of the third week up until the end of the eighth week. It is from the middle of week three until the middle of week six that the heart is most sensitive to the effects of teratogens. From the middle of week six until the end of week eight the heart is still undergoing development but should it be exposed to teratogens during this period, physiological and minor as opposed to major anomalies would persist.

The CNS on the other hand develops throughout the nine months of pregnancy and is therefore constantly vulnerable to the harmful effects of teratogens, such as alcohol, over the full 38 weeks. The human developing brain undergoes a rapid surge of growth during the third trimester and it is during this period that the brain is especially susceptible to ethanol-induced damage resulting in microcephaly and behavioural deficits (reviewed in Guerri, 2002). Based on figure 1.3, from week three up until the end of week sixteen the CNS is particularly sensitive and any insult during this period would result in major structural irregularities. After week sixteen and up until the end of gestation, the CNS is prone to physiological and minor structural problems (table 1.2).

Heavy and frequent alcohol consumption during the first trimester has been shown to predominantly affect the facial and structural features whereas drinking during the third trimester has an adverse effect on growth (May et al. 2004). It is therefore not surprising that maternal alcohol ingestion during any time of pregnancy can be detrimental to the unborn child.

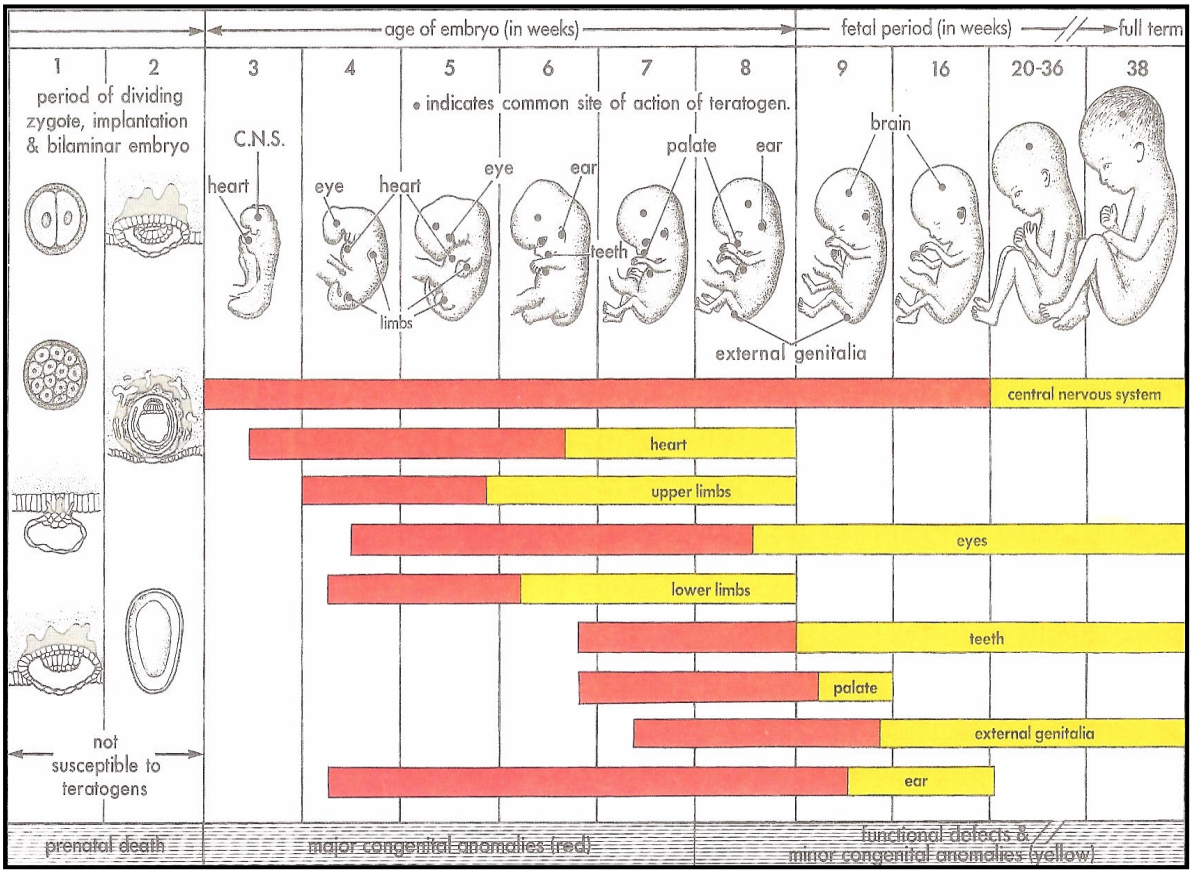


Figure 1.3: The various stages of organ development and the critical periods at which the fetus is most susceptible to birth defects induced by teratogens. Red denotes the highly sensitive periods where major defects may be produced. Yellow indicates the stages that are less sensitive to teratogens and where minor defects may be induced (taken from Moore and Persaud, 1993).

1.3.1 Areas of the Brain Susceptible to Alcohol Induced Damage

Damage caused to the developing brain by prenatal alcohol exposure is serious and permanent. The neurological effects are irreversible and will remain throughout a person's life (Chaudhuri, 2000). Autopsy reports and neuroimaging techniques such as magnetic resonance imaging (MRI) and single photon emission computed tomography (SPECT) have helped researchers understand the structural abnormalities of the brain caused by prenatal alcohol exposure. These findings have in turn lent some insight into what neuropsychological and behavioural problems manifest as a result of the structural changes (Clarren et al. 1978).

The areas of the brain most vulnerable to alcohol's toxicity are illustrated in figure 1.4. Most investigations have focused on the corpus callosum, cerebellum, basal ganglia and hippocampus (Mattson et al. 2001). The frontal lobes and hypothalamus have also been viewed as being sensitive areas to alcohol's teratogenicity (Kellerman, 2003).

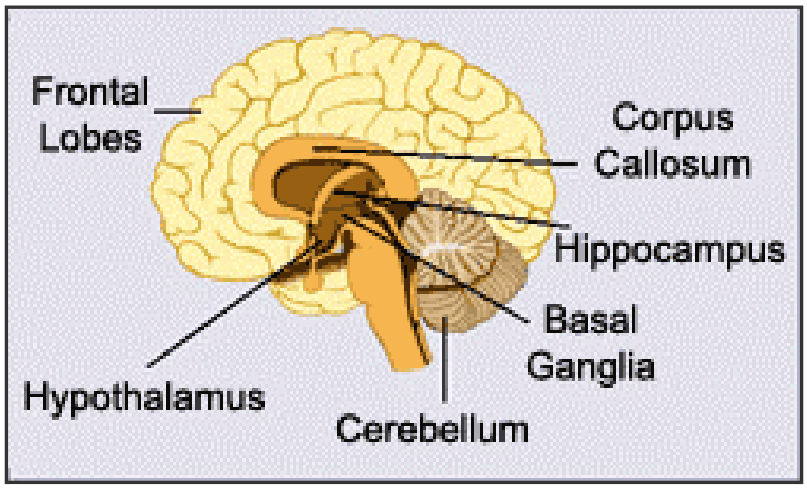


Figure 1.4: The areas of the brain most seriously affected by prenatal alcohol exposure (Kellerman, 2003).

The **corpus callosum** is a mass of intertwined nerve fibres that connects the two hemispheres of the brain. It is responsible for passing information from side to side with the right side controlling impulse and feelings while the left side governs logic.

Abnormalities in the corpus callosum have negative effects on attention span, reading, verbal and learning abilities and overall intellectual performance (Mattson et al. 2001).

The **basal ganglia** are a group of nerve cell clusters found deep within the white matter of the cerebrum. Within these clusters the caudate nucleus, putamen and globus pallidus can be found. These three sub-structures are involved in spatial memory, perception and thinking. The caudate nucleus in particular governs some advanced cognitive functions such as the ability to switch modes, planning ahead and problem solving (Mattson et al. 2001; NIAAA, 2000).

The **cerebellum** is located at the base of the brain and is especially affected by prenatal alcohol exposure. This region is involved in motor and cognitive skills such as attention as well as balance and co-ordination. Damage to this area of the brain has been coupled with learning problems as well as balance and co-ordination concerns (Mattson et al. 2001; NIAAA, 2000).

The key function of the **hippocampus** is consolidation of memories and any insult on this region would result in deficits in memory functions such as the ability to store new memories (Mattson et al. 2001). The **fontal lobes** oversee decision making and impulses. The prefrontal cortex is responsible for controlling the executive functions such as problem solving, regulation of emotion, motivation and inhibition. The **hypothalamus** rules temperature as well as the feeling of pain and appetite (Kellerman 2003).

1.3.2 Structural Damage

One of the diagnostic characteristics of FAS is microcephaly, which means ‘small head,’ one that has a circumference at or below the tenth percentile (table 1.1). Microcephaly is nearly always associated with an abnormally undersized brain, most often referred to as micrencephaly. The overall size of a FAS brain is usually significantly reduced compared to an unaffected brain (figure 1.5) making micrencephaly a considerably common finding (Clarren et al. 1978).

Studies on the brains of FAS individuals have revealed that damage is extensive and involves irregularities in the formation of the actual brain tissue itself, both the grey as well as the white matter regions are affected (Archibald et al. 2001). The cortex of the brain is composed of grey matter, which consists of nerve cell bodies, dendrites, and axons that are not surrounded by a myelin sheath. Grey matter actually appears grey in colour. White matter on the other hand consists of myelinated nerve fibres (the axons) which connect nerve cells with each other; the myelin sheaths give this tissue a white appearance (Mattson et al. 2001; Campbell et al. 1999).

Postmortem studies of FAS brains have reported cerebral and cerebellar disorganisation with strange arrangements of neurons in the cortex. Another interesting finding that has been documented in a number of FAS cases is the presence of neuroglial heterotopias which are abnormal neural and glial tissue covering segments of the brain surface (Clarren et al. 1978).

The corpus callosum is an area of the brain that has often been reported as being malformed in studies involving MRI, SPECT and autopsies on FAS individuals. Hypoplasia (underdevelopment), as well as partial to complete agenesis (absence) of the corpus callosum has been documented in a number of FAS cases (Swayze et al. 1997). Reduction in the basal ganglia and cerebellum volume have also been seen within FAS brains (Archibald et al. 2001). Cortical atrophy (degeneration), delayed myelination of the white matter (Riikonen et al. 1999) and hydrocephaly (Clarren et al. 1978) are other features that have been documented in FAS individuals.

The FAS brain in figure 1.5 is a particularly severely affected one. As mentioned previously the degree of FASD in a child, and therefore the structural damage caused by alcohol to the child's brain, can vary depending on factors such as alcohol dosage and time during pregnancy at which alcohol was consumed (Clarren et al. 1978). One cannot then assume that all FAS affected individuals' brains physically appear like the one presented in figure 1.5 as there is no specific pattern of brain damage and the extent of the abnormalities is not consistent in all FAS cases (NIAAA, 2000). One can however confidently say that the brain is most definitely targeted by alcohol and causes varying

degrees of physical abnormalities and as a consequence, psychological problems in a child.

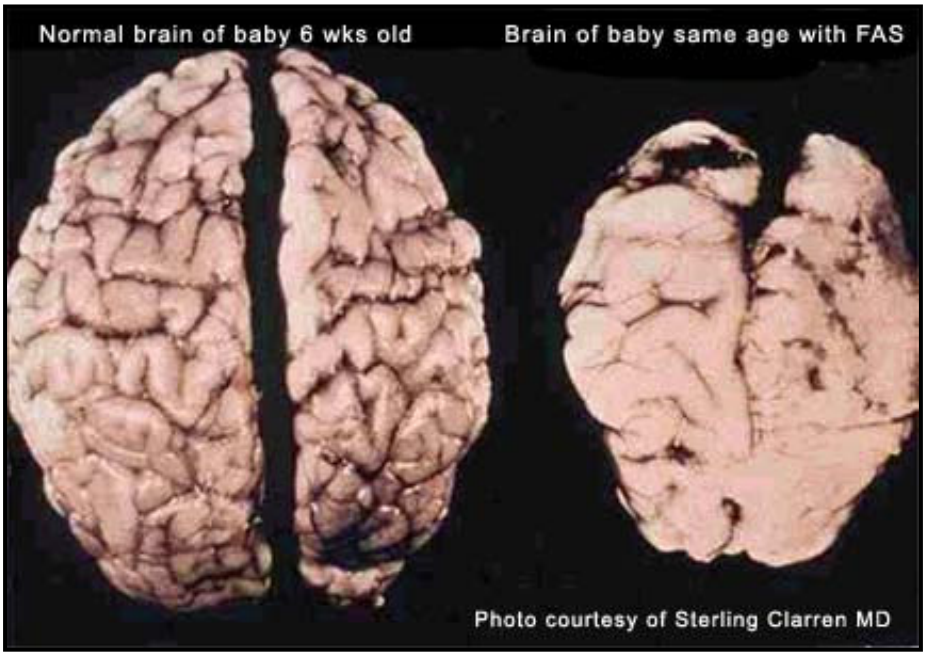


Figure 1.5: The difference in size and structure between the brain of a six week old FAS child and an unaffected child of the same age (Kellerman, 2003).

1.3.3 Neuropsychological Impairments

FAS individuals tend to exhibit cognitive and fine and gross motor performance impairments. The various areas of performance that are affected are discussed below:

1.3.3.1 Learning and Memory

Studies performed on FAS individuals have indicated that they show impairment in certain areas of memory. In particular these people tend to have difficulty in learning new information (Mattson et al. 2001). With regards to memory, studies have highlighted that retention and recognition memory are relatively unaffected in most FAS cases but declarative memory, whereby individuals have to give verbal feedback on a previously learnt task is often impaired (Jacobson and Jacobson, 2002). Based on the various functions of the different regions in the brain that are most susceptible to alcohol damage

(figure 1.4), impairments in learning and memory are indicative of damage to the cerebellum and hippocampus respectively (Mattson et al. 2001).

1.3.3.2 Executive Functioning

Executive functioning encompasses all higher level cognitive skills such as the ability to plan ahead, solve problems and think conceptually. Children exposed to alcohol prenatally usually have executive functioning deficits that can cause negative impacts on their lives. Individuals with prenatal alcohol exposure are more likely to act on impulse instead of first considering what consequences their actions might entail and they also exhibit difficulty in problem solving and planning ahead. It is evident that any individual experiencing problems such as the ones mentioned would more than likely struggle in a mainstream schooling environment as they would require special attention and extra help (Mattson et al. 2001). Impairments in the mentioned executive functioning activities indicate damage to the frontal lobes of the brain, particularly to the prefrontal cortex (Kellerman, 2003).

1.3.3.3 Motor Control

Motor control involves feedback from the body's peripheral nervous system, involving sensory organs such as the eyes and ears, to the CNS. The vestibular system is located in the inner ear and is also important in motor control as it controls one's sense of balance. Tests on children exposed to alcohol prenatally have revealed abnormalities in the development of motor control. It has also been suggested that balance problems in these individuals could possibly be attributable to direct deficits of the CNS as opposed to the peripheral nervous system (NIAAA, 2000). Direct damage to the cerebellum is responsible for balance deficits (Mattson et al. 2001).

1.3.3.4 Psychosocial and Behavioural Discrepancies

Various studies have indicated that children exposed to alcohol prenatally experience behavioural problems that can be disruptive in their homes, schools and social environments. Behaviour wise, attention deficits and hyperactivity have been associated with alcohol exposure in children (Mattson et al. 2001). Riikonen et al. (1999) reported an interesting finding when examining the brains of ten FAS subjects through MRI and SPECT. Normal left-right dominance was lacking in the frontal area of the brain. This is the area of the brain that is affected in attention deficit hyperactivity disorder (ADHD).

Psychiatric disorders and antisocial behaviour are common amongst children exposed to alcohol. Their socialisation and communication skills tend to be poor and depression and anxiety have been documented in a number of subjects. The behaviour of children prenatally exposed to alcohol has been documented as being characteristic of people with autism in that they exhibit deficits in social relations and communication (NIAAA, 2000).

1.3.4 Interference at the Cellular Level

Understanding the means to alcohol induced fetal damage is multifaceted as the cellular and molecular processes that occur during normal development of the CNS are still being explored. Alcohol has the ability to interfere directly with the survival of brain cells; it can induce cell death or affect the actual cellular processes of certain CNS cells (Goodlett and Horn, 2001). Table 1.2 indicates the various stages of CNS development as well as the effects alcohol has on each stage.

A number of cellular and chemical processes have been well documented as being disrupted by prenatal ethanol exposure; these processes include cell proliferation, migration and adhesion as well as neurotransmission (Guerri, 2002). Normal fetal development consists of three main stages of events. Firstly, cell division and proliferation occur followed by the second stage involving cell growth and differentiation. The second stage involves the specialisation of cells making them more specific in organisation and purpose. The third and final event involves the migration of maturing cells to their final

destinations within the developing embryo. Once they have reached their destination they will remain there and form a connection with already present surrounding cells (Michaelis and Michaelis, 1994).

Table 1.2: The effects of alcohol on the developmental stages of the human CNS (Guerri, 2002)

	1st month gestation	**2nd trimester gestation**	**3rd trimester to 2nd year of life**
Process of development	• Neural tube formation • Neural death	• Continuation of brain differentiation • Appearance of stem cells and/or radial glia • Neuronal proliferation and migration	• Brain growing at its fastest rate • Massive neural cell death • Astrogliogenesis and myelin development • Functional neural connections • Differentiation of the cerebellum
Effects of alcohol	• Increased neural crest cell death • Severe neural tube defects such as hydrocephaly and craniofacial abnormalities	• Abnormal radial glia • Abnormal cell migration • Neural cell loss	• Prominent microcephaly • Abnormal astroglial development • Increase in apoptosis and necrosis • Alterations in neural connections • Alterations in the cerebellum

Neuronal damage and loss appear to be the predominant outcomes of alcohol exposure to the developing CNS. An increase in programmed cell death (apoptosis) is well established as being one of the effects of prenatal alcohol exposure at the cellular level. The overall reduction in size and the partial to complete agenesis of certain regions in the brain of FAS individuals result from the loss of neurons through apoptosis (Olney, 2004). At the cellular level, cell survival, migration, adhesion and differentiation are altered. At the

molecular level, growth factors as well as neurotransmitters are affected by alcohol (Goodlett and Horn, 2001).

1.4 NEURONAL MIGRATION

During brain development, neurons are programmed to move to precise locations where they serve a highly specific role in the general organisational activity of the brain. New neurons are generated in the germinal layers of the neural tube and it is through this tube that they migrate to their final destinations in the various areas of the CNS. Once positioned, they form synaptic connections with neighbouring cells. Completion of this process results in a correctly developed and organised brain (Sobeih and Corfas, 2002). There are six distinct layers in the forebrain in which neurons settle, specific classes of cells reside in specific layers, this is referred to as neurogenesis. Neurogenesis takes place in an "inside-out" manner in that the earliest generated neurons are situated in the deepest layers and the later generated neurons reside in the layers closest to the surface. CNS migration can be separated into three distinct steps, firstly, establishment of cell identity, secondly directed migration and thirdly, assembly into dense neuronal layers (Hatten, 1999).

Neurons migrate in their postmitotic phase in one of two manners, radial or tangential migration. Radial migration involves movement of the neurons along glial fibres whereas cells migrating tangentially are attached to one another as opposed to glial cells and therefore form chains of migrating neurons (Sobeih and Corfas, 2002). Radial glial-guided migration is the predominant pathway in neuronal migration although recent studies have indicated that different neuronal populations use different modes of migration (reviewed in Sobeih and Corfas, 2002).

Radial migration involves glial cells. Glial cells are more abundant in the CNS than neurons and during the process of radial migration, they form a scaffold that spans the entire thickness of the developing CNS, it is along these glial fibres that neurons attach and travel to their precise locations in the developing brain. Once the process of migration

is complete, glial cells change form and become star-shaped astrocytes and the migrating neurons are locked in position (Hatten, 1999; Dirks, 2001).

1.4.1 The Effect of Alcohol on Neuronal Migration

The cellular process of cell migration is a definite target of prenatal ethanol exposure. The brain abnormalities seen in FAS children such as hydrocephalus, partial to complete agenesis of the corpus callosum, microcephaly, neuroglial heterotopias and abnormal arrangements of neurons are all indicative of errors in migration of neuronal and glial elements (Clarren et al. 1978; Hirai et al. 1999). It has also been reported that irregular neuronal migration and a 36% reduction in the amount of glial cells from the somatosensory cortex have been found in animals prenatally exposed to alcohol (reviewed in Guerri et al. 2001).

Glial cells are adversely affected by alcohol. Studies have indicated that glial cells exposed to alcohol transform into astrocytes prematurely. Consequently, later generated neurons, which would migrate and form the superficial layers of the cerebral cortex, fail to migrate to their proper locations and end up in abnormal positions. This would explain the disorganisation of neurons found in FAS brains (Goodlett and Horn, 2001; Welch-Carre, 2005). Table 1.2 indicates the stages of fetal development where neuronal migration and astroglial viability is interfered with as a result of prenatal ethanol exposure.

1.5 THE GENETICS BEHIND FETAL ALCOHOL SYNDROME

For a number of years researchers have been searching for genes involved in alcoholism. It has been reported that approximately 50 to 60% of the discrepancies in alcohol dependence can be attributed to genetic factors (Dick and Foroud, 2002). A likely assumption is that mothers of FAS children actually suffer from alcoholism. This therefore makes it exceptionally difficult to distinguish the genetic factors that contribute to alcoholism from those that cause susceptibility towards FAS. Certain environmental

and circumstantial factors such as availability of alcohol and poor social and economic status definitely increase the pervasiveness of alcoholism and consequently FAS (Viljoen et al. 2002). Factors like dose and timing of exposure contribute to the variation in diagnoses between children exposed to alcohol prenatally. It is however likely that there is a strong genetic influence contributing towards susceptibility for this syndrome in individuals (Thomas et al. 1998). Not all women who consume alcohol during pregnancy produce children with symptoms of fetal alcohol damage. This therefore suggests the likelihood of a role for either the mother's, child's or both individuals' genotypes in susceptibility towards, or possibly protection against FAS (Stoler et al. 2002).

1.5.1 Twin Concordance Studies

Studies indicating a genetic contribution towards alcohol induced birth defects have involved twin concordance studies. Streissguth and Dehaene (1993) carried out twin concordance studies with both monozygous (MZ) (identical twins, same genetic make-up) and dizygous (DZ) (nonidentical twins, different genetic make-up) twins of alcoholic mothers. Sixteen pairs of twins, of which five were MZ and eleven were DZ formed the study group. All twin pairs had been exposed to excessive amounts of alcohol prenatally.

The study revealed that five MZ twin pairs were equally affected by the prenatal alcohol exposure and were diagnosed as having FAS. Of the eleven DZ pairs, only seven pairs showed concordance for FAS. The other four DZ pairs showed discordance. In one of those discordant twin pairs, one twin was clinically diagnosed as having FAS whereas the other twin was diagnosed with what was then referred to as "fetal alcohol effects" (FAE). This term is no longer used in FASD diagnosis; nowadays this child would most probably be classified as partial FAS, ARND or ARBD (Hoyme et al. 2005). Nevertheless, a diagnosis of FAE indicated that this twin did not present with all the characteristics necessary for a clear FAS diagnosis. In another two of the discordant DZ pairs, one twin had FAE and the other twin was not affected at all (Streissguth and Dehaene, 1993).

Regardless of the fact that alcohol exposure within twin pairs was similar, alcohol effects are expressed more uniformly in MZ than DZ twins. Streissguth and Dehaene's (1993)

study supports a likely role of the fetal genome as playing an important role in protecting against or increasing susceptibility towards alcohol induced damage. It also highlights the varying control the genes may have on an individual in the expression of the harmful effects of alcohol (Streissguth and Dehaene, 1993).

Figure 1.6 is a photograph of a South African dizygotic twin pair examined by Professor Denis Viljoen. Their mother drank heavily during her pregnancy; one twin was born with all the characteristics of FAS whereas the other twin was unaffected.



Figure 1.6: Dizygotic twin sisters. The twin on the right has a clear FAS phenotype whereas the twin on the left was unaffected by the prenatal exposure to alcohol (photo courtesy of Prof. Denis Viljoen, FARR, with permission from the parents).

1.5.2 Animal Models

Various animals such as rodents, zebra fish, pigs and sheep have been used as models to explore the effects of prenatal alcohol exposure. These models have helped researchers understand the many means by which alcohol induces fetal damage, in relation to timing of exposure during the different stages of fetal development as well as dosage and pattern of alcohol administered (Cudd, 2005). The mouse in particular has proven to be an extremely good model for prenatal alcohol effect studies. The clinical presentations of

FAS in the mouse are remarkably similar to those seen in the human. Typical facial dysmorphisms, growth retardation and CNS malfunctioning are all seen in experimental mice models after prenatal alcohol exposure (figure 1.7) (Sulik, 2005).

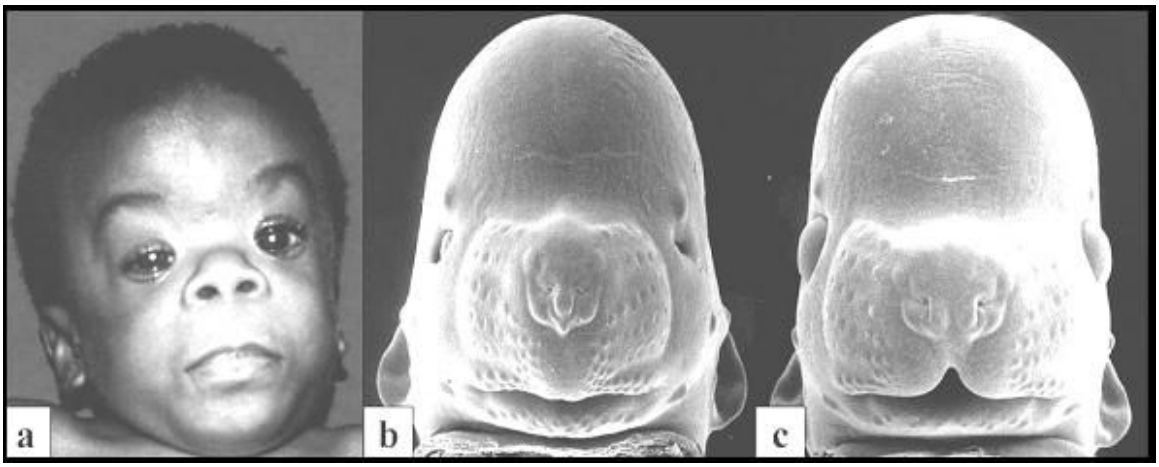


Figure 1.7: The similarities in the clinical presentation of FAS in the mouse versus the human. A child affected with FAS **(a)** has the same craniofacial features, including microcephaly, short palpebral fissures, a long upper lip and a smooth philtrum when compared to a mouse fetus whose mother was treated with ethanol on day 7 of her pregnancy (equivalent to week 3 of human gestation) **(b)**. A normal mouse fetus of the same developmental stage is present for comparison **(c)** (taken from Sulik, 2005).

Different mouse strains have also been used to explore the genetically related responses to alcohol induced fetal damage. One such study looked at three inbred mouse strains namely, C57BL/6J, DBA/2J and A/J to establish whether they vary in their susceptibilities to ethanol teratogenesis. These strains have long been used in alcohol research and they are the founding parental stocks of the most extensively genetically mapped recombinant inbred strains BXD and AXB/BXA. The findings revealed that both the extent and pattern of alcohol induced fetal damage differed by strain. Prenatal ethanol exposure increased the percentage of abnormalities for the C57BL/J and A/J strains (when compared to their relevant maltose exposed controls) but the D2 strain showed no significant difference when compared to its maltose exposed controls. The results also indicated that C57BL/J6 litters had a significant increase of digit, kidney and skeletal malformations whereas A/J litters only appeared to have increased skeletal malformations. These results point towards the presence of a strong genetic component in the susceptibility to alcohol's damaging effects upon fetal development (Boehm et al. 1997).

Another study used rats to model the responses of two genetically different strains to prenatal alcohol exposure. High-alcohol-sensitive (HAS) and low-alcohol-sensitive (LAS) rat pups were exposed to the same amount of ethanol. These two groups of rats are distinguished based on their sensitivity to the hypnotic effect of alcohol. After ethanol exposure, HAS rats were classified as being hyperactive compared to their controls. However, LAS rats did not differ in activity levels when compared to their controls (Thomas et al. 1998).

These types of findings support the suggestion that the genome does have significant control over predisposing to or protecting against alcohol induced fetal damage. This therefore implies that FAS is a complex multifactorial disease, one in which both environmental and genetic factors play a role.

1.6. COMPLEX DISEASES

Genetic research has been successful in identifying genes involved in Mendelian genetic diseases, whereby a mutation in a single gene is responsible for a particular disease. FAS falls under the group of complex diseases. Complex diseases do not follow Mendelian inheritance; no single gene is responsible for disease pathogenesis instead, both environmental and genetic factors play a role, these diseases are termed “multifactorial”. With complex diseases there are often multiple genetic factors contributing which interact with each other as well as with environmental factors. Schizophrenia, diabetes mellitus, hypertension and the common cancers are a few examples of complex diseases in which genetic components as well as environmental components contribute to disease pathogenesis (Risch and Merikangas, 1996; Mueller and Young, 2003).

For most complex diseases, the genetic component is usually the cause of an inherited predisposition or genetic susceptibility towards the disease. The detection of the genetic components of complex diseases has proved to be rather difficult as there are usually several genes of varying effect involved (Kwon and Goate, 2000). The search for

susceptibility genes in multifactorial disorders involves two different approaches, linkage studies and association studies (Kwon and Goate, 2000).

1.6.1 Linkage Studies

Linkage-based studies have been very successful in mapping single gene disorders. Linkage analysis makes use of affected pairs of siblings or extended families where the disease is prevalent. The basis of linkage studies is to analyse the degree of sharing of marker alleles between siblings or relatives presenting with the same trait (Schork, 1997; Dean, 2003). The approach involves studying the segregation of polymorphic markers, usually through microsatellite typing, from each chromosome in affected families until a marker is identified that co-segregates with the disease more often than what is expected to occur by chance. This then indicates that the marker and the disease are likely to be linked (Mueller and Young, 2003).

Linkage analysis has its pitfalls when it comes to studying complex diseases. Firstly, the method has limited power to detect genes of modest effect and as multifactorial disorders have multiple genes exerting small but relevant effects, this type of study would fail to identify these genes. Secondly, linkage studies are not designed to evaluate many gene and environmental effects concurrently and thirdly, sample collection of large disease affected families is time consuming and costly. The most significant problem with sample collection for these types of studies is heterogeneity. Families from different ethnic backgrounds and environmental exposure can enter the study causing heterogeneity and thus obscuring the signal of a given gene effect (Schork, 1997).

The sib-pair approach has been applied to a number of complex disorders such as cancers and asthma and, despite some obviously relevant loci being identified in these studies, most researchers have been unable to identify a locus by stringent criteria (reviewed in Dean, 2003). Linkage analysis is an excellent approach with high power for identifying rare high-risk alleles but for detection of complex disease alleles, association studies are far more powerful (Carlson et al. 2004). Table 1.3 lists the advantages and disadvantages of linkage studies.

Table 1.3: The advantages and disadvantages of linkage studies

Study	Advantages	Disadvantages
Linkage studies	• High power for detecting rare high-risk alleles • No need to understand pathogenesis of the disease in detail	• Low power for detecting rare risk alleles with moderate effects • Have to recruit family members

1.6.2 Association Studies

Association studies seek to test the association between gene associated deoxyribose nucleic acid (DNA) variants, referred to as a polymorphism, and a disease. Polymorphisms arise through evolution as a result of mutations, which attain significant frequency in populations. If the variant allele is more common in affected individuals than unaffected individuals then one could deduce that either there is a true association between the polymorphism and the disease and that the genetic variant causes susceptibility towards the disease or, it could be that the polymorphism is in linkage disequilibrium (LD) with the true susceptibility allele at a nearby locus (Schork, 1997).

Association studies involve the candidate gene approach that can be assessed in one of two ways, the case-control method or the family-based method. Unlike linkage analysis where the entire genome in scanned in an unbiased manner without any presumptions about the role of a particular gene, the candidate gene approach allows the investigation into whether or not there is an association between an "educated guess" about the genetic basis of a disease and its pathogenesis (Kwon and Goate, 2000).

For a candidate gene analysis, it is imperative that researchers have some knowledge about the underlying mechanisms of the disease so that they are able to choose a relevant gene, one that would seem logically associated with the disease (Kwon and Goate, 2000). Selecting a candidate gene involves knowledge of the gene and its protein product and its potential role in the pathways and mechanisms involved in the pathogenesis of the

disease. For example, in FAS one might look at the various cellular and molecular pathways affected by prenatal alcohol exposure and then look for genes associated with those particular pathways.

Once the candidate gene has been selected, one needs to select which particular DNA variants within the gene will be assessed. There are different types of DNA variations namely, microsatellite repeats, insertions/deletions and single nucleotide polymorphisms (SNPs) (Romero et al. 2002).

SNPs are the most numerous types of polymorphisms in the genome and were thought to occur every 500-1000 base pairs (bp) (Daly and Day, 2001). More recent research has reported SNP occurrence every 100-300 bp (http://www.ncbi.nlm.nih.gov/SNP) and the HapMap project has recorded one SNP per 279 bp (The International HapMap Consortium, 2005). SNPs are variations in DNA sequence where one of the four nucleotides is replaced with another and where the least frequent allele occurs at a frequency of 1% or more in the general population (Daly and Day, 2001).

SNPs can be classified as coding or non-coding depending on where exactly in the genome they are located. Non-coding SNPs can occur in the promoter region, the introns, 5' and 3' untranslated regions as well as intergenic regions. Coding SNPs occur in the exons of a gene. A SNP that affects an amino acid is referred to as a coding SNP or cSNP. Coding SNPs can either be synonymous, non-synonymous or nonsense mutations. Synonymous SNPs change the codon into another that codes for the same amino acid. Non-synonymous SNPs, also referred to as missense mutations, on the other hand change the codon to one that codes for a different amino acid thus potentially altering the protein structure and function (Romero et al. 2002). Nonsense mutations change the codon into a "stop codon", UAG, UAA or UGA. This can lead to nonsense mediated RNA decay whereby the mRNA carrying the premature termination codon is degraded *in vivo* (Strachan and Read, 2004).

Based on this knowledge, SNPs that are more likely to have functional consequences, such as coding SNPs, are usually the choice of SNP for candidate gene studies. It is also

important for researchers to look at whether or not a SNP is in linkage disequilibrium to other SNPs nearby, either in the same gene or a neighbouring gene. As well as looking at any possible association between individual SNPs and a disease, haplotype analysis, which involves looking at a combination of SNPs together, is also an important feature of candidate gene association studies (Daly and Day, 2001).

1.6.2.1 Case-control Approach

A case-control study involves DNA samples from affected individuals (cases) as well as matched controls (unaffected individuals). A case-control investigation will compare allelic and genotypic frequencies of a SNP in the two different groups. It is most important that the control subjects are matched to the cases based on ethnicity, age, sex and geographical positioning. Controls cannot be related to any of the case individuals. Mismatch between the two groups, known as population stratification, could lead to a false-positive association (Brookes, 1999).

Case-control studies depend on the frequency of the variant allele in order to determine how many cases and controls are required to attain adequate statistical power to observe a difference in frequency between the two groups. This is not always possible as frequency data are not always available. A 2:1 ratio of controls to cases is a widely accepted ratio in case-control studies especially when the number of cases is limited due to difficulties in finding patients or the disease being very rare (Daly and Day, 2001).

It is important to have a sufficient sample size in a case-control study. If the genetic contribution to a disease is small, large sample sizes will be necessary to exhibit an association between the polymorphism and the disease. Similarly, if the genetic contribution to a disease were large, a smaller sample size would be sufficient (Romero et al. 2002). Table 1.4 lists the advantages and disadvantages of case-control studies.

Table 1.4: The advantages and disadvantages of case-control studies

Study	Advantages	Disadvantages
Case-control studies	• Simple approach • Cost effective • Easier to obtain cases and control samples compared to family-based studies • High power for detecting rare risk alleles with small to moderate effect	• Have to have a good understanding of disease pathogenesis and the possible biological pathways involved in order to choose good candidate genes • Good understanding of candidate genes, their functions and various polymorphisms is required • Population stratification if controls are not well matched to cases • Large sample size required if genetic contribution to disease is small

1.6.2.2 Family-based Approach

The family-based approach to association studies involves the use of cases and their parents or unaffected siblings who act as internal controls. The use of either parents or unaffected siblings enables one to rule out population stratification problems (Romero et al. 2002). The most frequently used test in family-based studies is the transmission disequilibrium test (TDT). TDT compares the frequency of transmission of alleles from heterozygous parents to affected offspring (reviewed in Dean, 2003). It is irrelevant whether or not the parents are affected. It is imperative that at least one parent be heterozygous for the polymorphism of interest (Warren and Li, 2005). The test therefore classifies the transmitted alleles to affected offspring as "cases" and the nontransmitted alleles as "controls." If an association were seen, the high-risk alleles would be expected to be amongst those transmitted as opposed to those not transmitted (Romero et al. 2002).

One of the drawbacks of TDT testing is the actual recruitment of parents or unaffected siblings. Collecting DNA samples from parents when the affected individual is still a child is not as problematic as collecting DNA from parents when the disease onset is late in life as parents might be deceased by then (Daly and Day, 2001). Deviations from the original TDT test include the use of single parents and unaffected siblings (Dean, 2003). The sib-TDT is employed when one or both parents are missing (reviewed in Strachan and Read, 2004).

Both case-control and family-based approaches assume the SNP being tested does contribute to genetic susceptibility or that it is in LD with the true susceptibility allele. It is important to test SNPs at regular intervals as LD is thought to be maintained over genomic regions of about 5000-50000 bp of DNA (Mathew, 2001). The HapMap project has however clearly demonstrated that LD varies greatly across the genome and between populations (McVean et al. 2005; http://www.hapmap.org).

1.6.3 Previous Candidate Gene Association Studies on FAS

A few case-control association studies investigating the role of certain genetic polymorphisms in the protection against or susceptibility towards alcohol related birth defects have been conducted. Up until now, the focus has only been on genes involved in the actual alcohol metabolism pathway. Ethanol is metabolised by alcohol dehydrogenase (ADH) and cytochrome P450 2E1 (CYP2E1) into acetaldehyde, which is further metabolised by aldehyde dehydrogenase (ALDH) into acetate (McCarver et al. 1997).

The *ADH2* gene has been of particular interest in studies investigating the genetics behind alcohol related birth defects (McCarver et al. 1997). *ADH2* is polymorphic and as a result there are different isoforms of the enzyme arising from functional polymorphisms in the *ADH2* gene. The different variants catalyse the oxidation of ethanol at different rates (Viljoen et al. 2001). The *ADH2*1*, *ADH2*2* and *ADH2*3* alleles have been associated with FAS and ARBD in certain populations. The frequencies of these alleles differ by race, the *ADH2*3* allele has been reported as being unique to African Americans (McCarver et al. 1997).

One of the first ARBD candidate gene investigations was performed by McCarver et al. (1997). The study took place in Detroit, United States of America (USA) and focussed on African American women and their infants. The investigators reported on the *ADH2*3* allele and found that it occurred more frequently in control mothers and infants thus contributing to, or being a marker for a protective effect against ARBD among African Americans. *ADH2*3* metabolises alcohol at an increased efficiency compared to the other isoenzymes and so it was suggested that the protective element is secondary to the more effective metabolism (McCarver et al. 1997).

Stoler et al. (2000) carried out a second study involving the *ADH2* gene in the African American population in Boston, USA. These investigators also reported on the *ADH2*3* allele and contrary to the findings of McCarver et al. (1997), found that the allele was more prevalent in mothers who admitted to high alcohol use during pregnancy as well as in affected infants. This investigation concluded that the *ADH2*3* allele actually increased the risk of fetal damage induced by prenatal alcohol exposure.

In South Africa, Viljoen et al. (2001) carried out a similar genetics study in the Coloured population of the Western Cape Province. The investigators examined the *ADH2* polymorphisms in FAS affected children and their mothers as well as a control group. This study indicated that the *ADH2*2* allele was significantly more common in the control group thus either conferring protection or acting as a marker against FAS. The *ADH2*3* allele frequency was not significantly different between the study groups.

These mentioned genetic studies have been the first few to investigate the link between genetic polymorphisms and the effects of prenatal alcohol exposure. The ethanol metabolism pathway is just one of many pathways that should be explored for evidence of a genetic role in FAS. The best way to start analysing genes that may contribute to FAS susceptibility is to focus on pathways affected by alcohol such as neuronal migration and apoptosis.

1.7 THE *ASTROTACTIN* GENE

The *astrotactin* (*ASTN*) gene was discovered in 1996 by scientists at The Rockefeller University and Howard Hughes Medical Institute in New York, USA. The chromosomal location of *ASTN* in the human is 1q25.2 (figure 1.8). The structure of the astrotactin protein (ASTN) includes a signal peptide, a recognised transmembrane domain, three epidermal growth factor-like (EGF-like) repeats and two fibronectin type III (FNIII) repeats (Fink et al. 1997). EGF repeats are found in the extracellular domains of many proteins and partake in intercellular signalling during neuronal development. FNIII repeats are known to be present in axonal glycoproteins of the immunoglobulin G (IgG) superfamily. The arrangement and quantity of the EGF and FNIII repeats suggests that ASTN is unique compared to any other cellular adhesion or signalling molecule (Zheng et al. 1996).

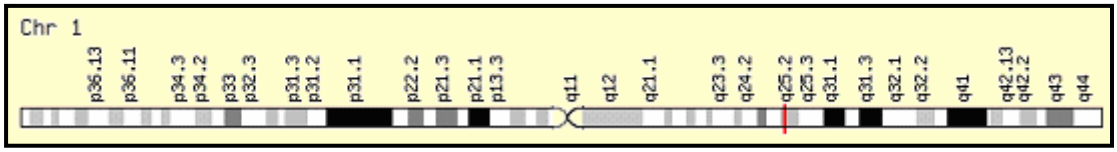


Figure 1.8: The chromosomal location of the *astrotactin* gene, 1q25.2 (taken from http://www.genecards.org).

The astrotactin family is one consisting of neuronal ligands which are neuronal adhesion molecules necessary for radial glial-guided neuronal movement in cortical areas of the developing brain (Koirala and Ko, 1998). ASTN is described as a glycoprotein and is the first protein shown to operate directly in neuronal migration along the glial fibres (Fink et al. 1997).

Expression studies have indicated that ASTN is expressed by postmitotic neuronal precursors in the cerebrum, cerebellum, hippocampus and olfactory bulb where neuronal migration forms the various layers of the brain. High levels of ASTN have been documented in developing brains but very low levels of the protein are present in adult brains (Hatten, 1999). *In vitro* studies demonstrated that young cerebellar granule neurons exposed to antibodies blocking ASTN cease to migrate along the glial fibres (Zheng et al.

1996). Removal of ASTN reduces the rate of neuronal movement by 60% (Pearlman et al. 1998).

The *ASTN* gene is 303.82 kb in size and is comprised of 23 exons, its orientation is on the reverse strand. The ASTN protein consists of 1310 amino acids and is said to have a molecular weight of 145670 Da (http://www.genecards.org).

A second *Astrotactin* gene, *ASTN2,* has been discovered (Sklar et al. 2002). *ASTN2* is located on the ninth human chromosome at position 9q33.2 and appears to have a similar role to chromosome one's *ASTN* by mediating neuronal-glial contact (Frosk et al. 2002). However, more detailed studies are required to gain a better understanding of the specific function of this gene. Due to the discovery of this second gene, the gene located on chromosome 1q25.2 is either referred to as *ASTN* or *ASTN1*.

1.7.1 *Astrotactin* as a Candidate Gene for FAS

Fink et al. (1997) described *ASTN* as "an excellent candidate gene for neuronal migration defects in humans." There are a number of brain defects that arise as a result of aberrant neuronal migration. Failure to form normal convolutions (gyria) in the brain is caused by abnormal migration. Disorganisation of the laminar structure of the brain cortex as well as the presence of neuronal heterotopias whereby subpopulations of neurons have moved to inappropriate places, result from neuronal migration malfunctioning (Fink et al. 1997). It is a known fact that prenatal alcohol exposure disrupts the process of neuronal migration. Neuronal heterotopias, disorganisation of the brain cortex and an overall reduction in brain size are common findings of FAS brains (Clarren et al. 1978; Hirai et al. 1999) thus supporting the appropriateness of *ASTN* as a candidate gene for FAS research.

Interestingly, deletions at the 1q25 position, the chromosomal location of *ASTN*, have been linked to micrencephaly (reduced brain size) in humans (Hatten, 1999). Considering micrencephaly is a predominant feature of FAS, *ASTN* appears to be a suitable candidate gene choice.

Astrotactin knockout mice have given researchers insight into the importance of this gene during brain development. Mice lacking the *Astn* gene tend to have slowed neuronal migration. In addition to slowed neuronal migration, a decrease in the binding process of the neurons to the glial scaffold is apparent and the development of the type of neurons known as Purkinje cells is altered. *Astn* null mice possess cerebella that are 10% smaller than wild type mice and are inherently less able to carry out tasks involving balance and co-ordinated movement (Adams et al. 2002). The cerebellar cortex controls movement and balance, thus the findings of Adams et al. (2002) support the presence of astrotactin protein in the cerebellum. Co-ordination and fine motor skills are poor in FAS individuals (Mattson et al. 2001; NIAAA, 2000) once again supporting the choice of *ASTN* as a candidate gene in FAS genetic studies.

1.8 THE *REELIN* GENE

The *reelin* (*RELN*) gene is located on human chromosome seven at position 7q22 (figure 1.9) (Pearlman et al. 1998). RELN is an extracellular matrix protein that contains a signal peptide followed by an N-terminal sequence and a hinge region. The hinge region is positioned upstream from eight RELN repeats of 350-390 amino acids (reviewed in Fatemi, 2005).

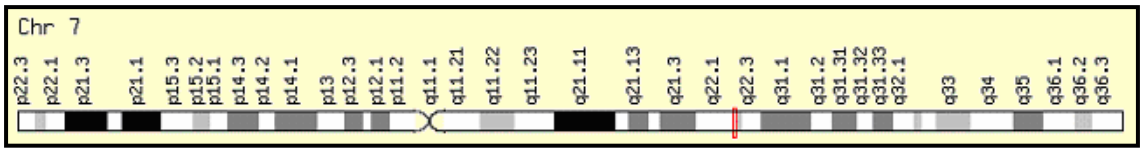


Figure 1.9: The chromosomal location of the *reelin* gene, 7q22 (taken from http://www.genecards.org).

RELN plays an important role in regulating radial neuronal migration in the developing neocortex as well as other areas of the brain such as the cerebellum and hippocampus (Wood, 2005). RELN protein is secreted by Cajal-Retzius neurons at the cerebral cortical surface. Expression of RELN peaks during the time at which neurons are generated and begin their migration (Lombroso, 1998).

Like the *ASTN* gene, the *RELN* gene is orientated on the reverse strand. The RELN protein is 517.73 kb in size with 24 exons and consists of 3460 amino acids. Its molecular weight is 3889399 Da (http://www.genecards.org; http://www.ensembl.org). The amino acid sequence of RELN is comparable to that of proteins secreted into the extracellular matrix surrounding neurons which assist neurons as they migrate (Lombroso, 1998).

For many years prior to the cloning of the *RELN* gene, *reeler* mice provided researchers with insight into the molecular basis of CNS development (Lombroso, 1998). *Reeler* mice are autosomal recessive for a spontaneous mutation in the *RELN* gene (reviewed in Honda et al. 2003). These mice lack *RELN* function and as a result, fail to form the necessary layers in certain brain regions, especially the cerebral and cerebellar cortices. The RELN protein is therefore vital for correct formation of the cortical laminae (reviewed in Hatten, 1999).

As previously described, normal layering of the cerebral cortex in the developing brain occurs in an "inside-out" manner in that the earliest generated neurons are situated in the deepest layers and the later generated neurons settle in the more superficial layers (Hatten, 1999). The *reeler* mutation causes the earliest generated neurons to migrate to the surface, the next generation of neurons then settle directly below and the final neurons form the deepest cortical layer (figure 1.10). The neurons therefore settle in an extremely disorganised, non-laminar manner (Lombroso, 1998). RELN is not produced by or distributed along glial cells. The initial phases of neuronal migration seem unaffected in *reeler* mice. Abnormalities lie in the termination of migration and layer formation (reviewed in Pearlman et al. 1998).

The precise manner in which RELN acts during neuronal migration has been disputed for many years by a number of researchers. Several suggestions have been made and studies have been performed to investigate RELN's exact mode of action. One of the most recognised proposed modes of action by RELN is that it acts as a neuronal migration stop signal by allowing neurons to detach from the glial cells once they have reached their proposed destination (Lombroso, 1998).

RELN expression is high in the developing brain but RELN levels have been observed in the postnatal and adult brains well after neuronal migration is completed. Postnatally high expression has been noted in the hippocampus, cerebellum and olfactory bulb. This indicates that RELN is not only a crucial protein in the embryological period but its presence is required for the growth and synaptic remodelling in the adult brain (reviewed in Zhang et al. 2002).

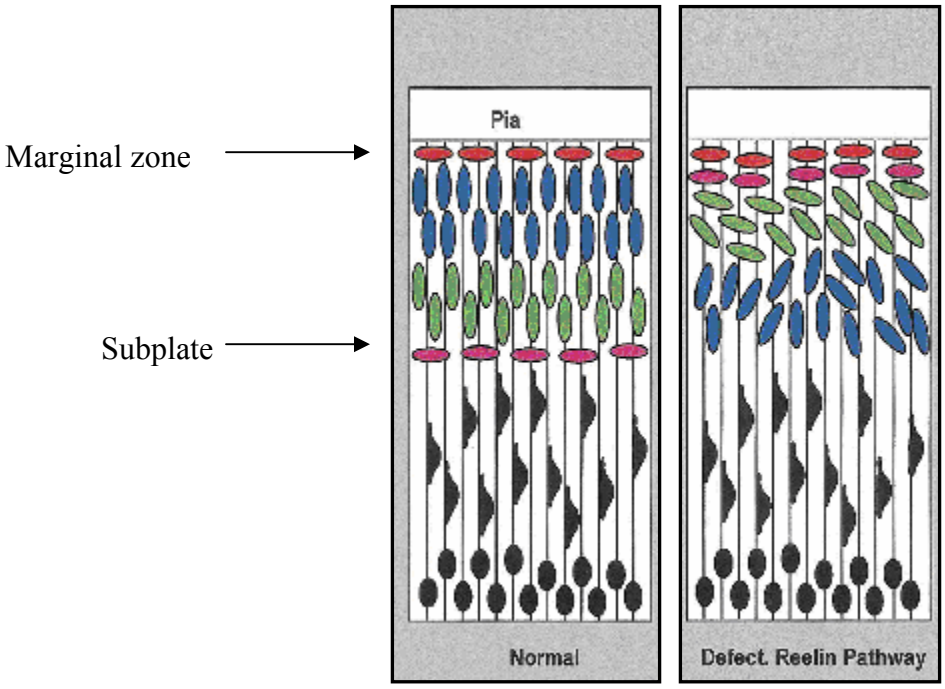


Figure 1.10: Schematic illustration of normal versus reelin deficient cortical development. In the normal cortex, neurons proliferate in the ventricular zone and migrate along radial fibres. Early postmigratory neurons settle horizontally in the marginal zone (red) and subplate (pink). Reelin produced in the marginal zone (red neurons) helps organise postmigratory neurons into layers. Early born cortical neurons (green) are found in the depth of the cortex and younger cells (blue) more superficially ("inside-outside" gradient). In the absence of reelin signalling, cortical neurons (green) settle at an angle in the cortex and displace the early contingent of neurons (red and pink) all together in the marginal zone; the next generation (blue) cannot cross the first one and the gradient is directed from outside to inside (adapted from Lambert de Rouvroit and Goffinet, 2001).

The pathological role of the *RELN* gene has been implicated in neuropsychiatric disorders such as autism (Bonora et al. 2003), schizophrenia, bipolar disorder, depression and lissencephaly (reviewed in Fatemi, 2005).

1.8.1 *Reelin* as a Candidate Gene for FAS

As described, RELN plays an essential role in the correct formation of the various regions of the brain by assisting the neuronal migration process. Postmortem studies on FAS brains have often reported disorganisation of neurons and structural alterations (Goodlett and Horn, 2001; Welch-Carre, 2005). Mutations in the *RELN* gene have been connected to considerable learning disabilities, underdeveloped cerebella and cognitive impairment in both man and mouse (reviewed in Fatemi, 2005). FAS individuals suffer from learning problems and cognitive deficits (Welch-Carre, 2005) and MRI and postmortem studies have indicated overall hypoplasia of the brain, in particular reduced cerebellar volume (Riikonen et al. 1999).

Research has also focussed on investigating a possible association between *RELN* mutations and autism (Bonora et al. 2003). This adds to the reasons for choosing *RELN* in a FAS genetics study as some of the behavioural aspects of FAS children have been characterised as being similar to autistic children (NIAAA, 2000).

Reeler mice show neurological symptoms such as tremors, dystonia and ataxia (loss of muscle tonicity and co-ordination respectively) as well as abnormal neuron positioning throughout the cerebral cortex, cerebellum and hippocampus (reviewed in Bonora et al. 2003). Newborn FAS infants have been described as being “tremulous” and although relatively rare, seizures have been reported in the neonatal period of FAS individuals (Clarren and Smith, 1978).

The correlations between the phenotype caused by *RELN* mutations and the neurological abnormalities seen in FAS individuals support the choice of *RELN* as a candidate gene for a FAS genetic study.

1.9 HYPOTHESIS

Certain single nucleotide polymorphisms in and around the *astrotactin* and *reelin* genes may have an association with, or be markers for, susceptibility towards or protection against fetal alcohol syndrome in the Coloured population of the Northern Cape Province in South Africa.

1.10 AIMS AND OBJECTIVES

The primary objective of this study was to examine genetic variation associated with the *ASTN* and *RELN* genes and to assess their potential role in the development of FAS in the Coloured population of certain regions in South Africa's Northern Cape Province. The aims of the study involved examining the genotype, haplotype and allele frequencies in DNA samples taken from FAS children in the Upington and De Aar areas and comparing them to geographically and ethnically matched control samples.

Specific objectives:

i. To determine the genotype and allele frequencies of the polymorphic variants of the *ASTN* and *RELN* genes in the samples.
ii. To analyse whether or not there is an association between *ASTN* and *RELN* polymorphisms and the predisposition to FAS.

CHAPTER 2: SUBJECTS AND METHODS

2.1 SUBJECT RECRUITMENT

Subjects for this study resided in the towns of Upington and De Aar in the Northern Cape Province of South Africa (figure 2.1).



Figure 2.1: A map of South Africa indicating the geographical positioning of Upington and De Aar (http://www.theodora.com/maps).

FAS cases were identified through physical and neurological examinations performed by trained medical doctors, psychometrists and psychologists. Most FAS cases were young children of school entry age. Informed consent to take venous blood samples from FAS cases was obtained from the parents. Blood from 45 FAS cases, all from the Coloured population, was collected in total. Of the 45 cases, 24 were female and 21 male.

Control individuals were recruited from the same geographical regions. The ethnic origins of the individual's parents and grandparents were recorded. However, no information was obtained on whether control individuals were or were not exposed to alcohol *in utero*. Although it would be ideal to have controls that were exposed to alcohol prenatally and did not develop FAS, it is exceptionally difficult and ethically challenging to achieve this. Only individuals with at least three Coloured grandparents were included. Control

individuals had to be 18 years old or older for informed consent purposes. In total 112 control subjects, 57 females and 55 males were collected for this study. These individuals did not undergo any physical or neurological examinations. They were considered to be a random sample of the study population. From each case and control individual, 10 ml of blood was drawn into EDTA (anticoagulant) tubes. Case and control samples were numbered consecutively. "NPF" was used as the code for cases and "NPC" was used for controls. The blood was stored in a cooler bag and sent to the Molecular Genetics Laboratory at the National Health Laboratory Service (NHLS) in Braamfontein, Johannesburg.

Previous ethics approval was granted for blood extraction by the Research Ethics Committee of the University of the Witwatersrand, reference numbers: M02-10-41 and M03-10-20 (Appendix A).

2.2 DNA EXTRACTION

Once the blood samples had reached the laboratory they were frozen at -20^0C. After defrosting at room temperature, DNA was extracted using the Flexigene DNA extraction kit manufactured by Qiagen. Appendix C describes the protocol.

2.3 DNA QUANTIFICATION AND STORAGE

Once extracted and resuspended in the provided buffer, the concentration of each DNA sample was quantified using the Nanodrop ND-1000 spectrophotometer. DNA concentration is determined by measuring the absorbance of ultraviolet (UV) light at a wavelength of 260 nm. As well as the concentration being determined by the Nanodrop, the purity of the DNA was also measured at a wavelength of 280 nm. The 260/280 ratio is indicative of the DNA purity. A ratio of 1.8 indicates pure DNA; a result of less than one would suggest protein contamination. Working stocks of each sample were made at a concentration of 100 ng/µl using double distilled water (ddH_2O). Working stocks were

kept at 4^0C. For each case and control sample, 200 µl of undiluted DNA was stored at -20^0C and the remainder of the extracted DNA was kept at -70^0C. The quality of the extracted DNA was of a very high standard which made genotyping of all 45 cases and 112 controls easily achievable.

2.4 CANDIDATE GENE CHOICE

Astrotactin and *reelin* were selected for this particular FAS genetics study. Text mining was the major procedure employed to select suitable candidate genes. Numerous journal articles and literature reviews covering topics such as FAS, the developing CNS, neuronal migration and the genes involved in these processes were read. Due to their crucial role in the process of neuronal migration the *ASTN* and *RELN* genes were decided upon for investigation. The Ensembl ID for the *ASTN* gene is ENSG00000152092, transcript ENST00000361833. The *RELN* gene ID on Ensembl is ENSG000001899056, transcript ENST00000339713. Ethics approval (protocol M040611) was obtained from the University of the Witwatersrand's Research Ethics Committee for the investigation of *ASTN* and *RELN* SNPs in the Northern Cape DNA samples (Appendix A).

2.5 SELECTION OF SINGLE NUCLEOTIDE POLYMORPHISMS

The Ensembl database (http://www.ensembl.org) was accessed to retrieve a list of all reported SNPs within each of the candidate genes. The database listed whether the SNPs were present in exons, introns or untranslated regions (UTRs). The functionality, chromosomal and amino acid (aa) positioning was also specified. In addition to that, the database provided information on whether or not a SNP had been verified by frequency studies or cluster analysis or whether it was merely suspected.

In cases where SNPs had been verified, the allele and genotype frequencies across several populations were stated. In order to locate the respective SNPs within the genomic sequence of each gene, the genomic sequences for *ASTN* and *RELN* were downloaded

from the University of Santa Cruz (UCSC) Genome Bioinformatics website (http://genome.ucsc.edu).

2.5.1 *Astrotactin* SNPs

Table 2.1 lists the four *ASTN* SNPs chosen for investigation. (See Appendix B, figure B2 for a schematic diagram of all SNPs present in *ASTN*). The three non-synonymous SNPs were chosen due to their potential for functional impact; a change in amino acid can cause serious alterations to the structure of the protein. An additional SNP, the 3' UTR SNP was selected, as frequency data was available for comparison. Figure 2.2 illustrates the chosen SNPs on chromosome one indicating their base pair positioning. "Status" represents whether or not the SNP has been proven in certain populations or if it is just suspected

Table 2.1: SNPs in the *ASTN* gene chosen for investigation

Ensembl ID	Alleles (forward strand)	SNP type	Chromosomal positioning	SNP Status*	aa change	aa site
rs6413830	G → A	3 prime UTR	173564209	Verified	-	-
rs12118933	G → C	Non-synonymous, coding	173565154	Suspected	R/G	1285
rs2228956	T → C	Non-synonymous, coding	173595524	Verified	H/R	939
rs11587640	G → T	Non-synonymous, coding	173733291	Suspected	L/M	282

* The term "SNP status" refers to the extent to which the SNP has been investigated. If "verified", it means that it has been confirmed in one or more populations and genotype frequencies are available. If it is "suspected" to occur it has not yet been validated through research in human populations.

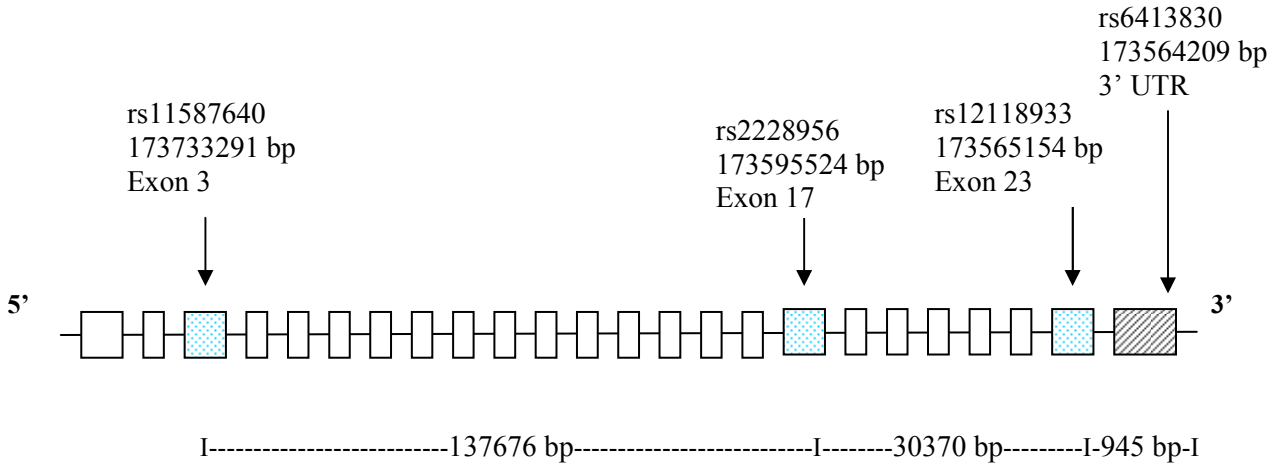


Figure 2.2: A schematic representation of the selected *ASTN* SNPs orientated on the reverse strand. The base pair position of each SNP on chromosome one is indicated as well as the base pair distance between them.

2.5.2 *Reelin* SNPs

Table 2.2 indicates the four *RELN* SNPs selected for the study (see Appendix B, figure B3 for a schematic diagram of all SNPs present in *RELN*). All four SNPs are non-synonymous coding SNPs that cause amino acid changes. Figure 2.3 illustrates the base pair positioning on chromosome seven of the selected SNPs.

Table 2.2: SNPs in the *RELN* gene chosen for investigation

Ensembl ID	Alleles (reverse strand)	SNP type	Chromosomal positioning	SNP Status*	aa change	aa site
rs607755	T → C	Non-synonymous, coding	102983905	Verified	L/S	199
rs3025963	A → T	Non-synonymous, coding	102845204	Suspected	E/V	973
rs3025966	T → G	Non-synonymous, coding	102864505	Suspected	A/C	852
rs362691	C → G	Non-synonymous, coding	102845112	Verified	L/V	1004

* The term "SNP status" refers to the extent to which the SNP has been investigated. If "verified", it means that it has been confirmed in one or more populations and genotype frequencies are available. If it is "suspected" to occur it has not yet been validated through research in human populations.

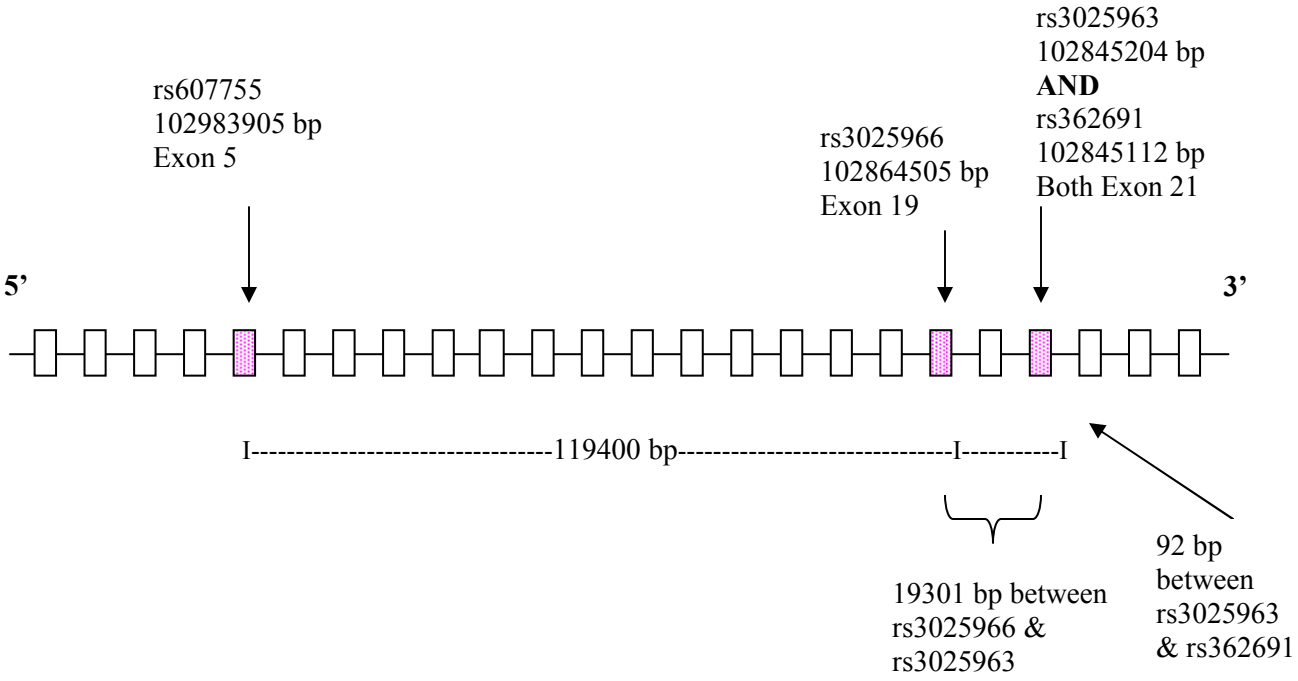


Figure 2.3: A schematic representation of the chosen *RELN* SNPs orientated on the reverse strand. The base pair position of each SNP on chromosome seven is indicated as well as the distance between them.

2.6 PRIMER DESIGN

In order to amplify the DNA region flanking each SNP by means of the polymerase chain reaction (PCR), oligonucleotide primers that are complementary to the region need to be designed. Primer design was performed both manually, by viewing the genomic sequence on either side of each SNP, and with the assistance of a bioinformatical program, Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi).

The number of primers designed per SNP was dependent upon the proposed mode of detection of the alleles. All but one of the eight SNPs investigated were restriction fragment length polymorphisms (RFLPs) (described below). For RFLPs only two primers are required, a forward and a reverse primer. When a SNP cannot be detected by an RFLP method another mode of detection is required. This study chose the amplification refractory mutation system (ARMS) whereby three primers are required for allele

detection per sample and an additional set off primers is required to serve as an internal PCR control. In total, ARMS requires five primers (Ye et al. 2001; http://www.ich.ucl.ac.uk/cmgs/arms99.htm).

When designing primers several factors including the GC content and consequently melting temperature of the primers were taken into consideration. Once primers had been designed they were verified against the human genome using Ensembl's BLAST function. This function allowed one to visualise at what locations in the human genome the primers bound to. Ideally, primers should only amplify one product, the desired DNA sequence containing the SNP. In order to achieve this, adding some base pairs onto or subtracting a few from one or both primers increases specificity.

Once a pair of primers had been selected, they were entered into a program called UCSC In-Silico PCR (http://www.genome.ucsc.edu) where a virtual PCR reaction would take place. The program reported the melting temperature of each primer and the size of the PCR product yielded by each reaction. This program confirmed that only one product was expected per reaction and also gave an indication as to what range the annealing temperature should fall into.

2.7 TECHNIQUES EMPLOYED

2.7.1 Polymerase Chain Reaction

PCR is a process that allows selective exponential amplification of a specific DNA sequence. It is a rapid and relatively simple procedure that is highly sensitive. Only a very small amount of starting DNA is required for a PCR reaction. Each PCR reaction requires a forward and reverse oligonucleotide primer, a thermostable DNA polymerase (isolated from the heat-stable bacterium *Thermophilus aquaticus* and therefore referred to as *Taq* polymerase), deoxynucleotide triphosphates (dNTPs) and double stranded DNA. Magnesium chloride ($MgCl_2$) and a buffer are also added to the reaction to supply optimal conditions for the polymerase. Each PCR reaction consists of several cycles of

temperature fluctuations that result in DNA denaturation, primer annealing, primer extension and hence DNA synthesis (Mueller and Young, 2003).

For each of the *ASTN* and *RELN* RFLPs basic PCR was used to amplify the regions flanking the SNPs. For *RELN's* rs362691 an ARMS PCR system was designed to detect the two alleles. Appendix C lists the ingredients used for the PCR reactions used in this study.

2.7.2 Restriction Fragment Length Polymorphisms

RFLPs are SNPs whose alternate alleles either introduce or remove a cleavage site for a specific restriction enzyme. RFLPs can be detected after PCR through the digestion of the PCR product with the appropriate restriction enzyme. RFLPs are effective and usually simple and detection of genotypes is relatively rapid and easily distinguishable. Temperatures at which digestion takes place and the amount of time required for appropriate digestion varies amongst the different enzymes. Digestion products are viewed by gel electrophoresis. The web-based program, known as NEBcutter (http://tools.neb.com/NEBcutter2/), by New England Biolabs, was used to detect which restriction enzyme sites were present around each SNP. Once an enzyme had been selected, NEBCutter performed a virtual digest and listed the digestion product sizes.

2.7.3 Amplification Refractory Mutation System

The principle behind an amplification refractory mutation system (ARMS) is the distinction of the different alleles of a SNP through the use of primers that differ by a single nucleotide at the 3' terminus. An ARMS primer is designed so that it can differentiate between templates that vary at a single nucleotide residue. A typical ARMS procedure consists of two separate reactions and utilises three primers. One reaction involves a primer whose 3' terminal residue is complementary to the "wild type" allele and the other reaction involves a primer whose nucleotide at the 3' end is complementary to the "mutant" allele. A third primer is designed which is constant and complementary to the DNA template in both reactions. Two reactions are therefore carried out per DNA

sample. If a sample is homozygous for the "wild type" or "mutant allele," amplification will occur in one reaction only. If however, a sample is heterozygous in that it has both the "wild type" and "mutant allele," amplification will take place in both reactions (http://www.ich.ucl.ac.uk/cmgs/arms99.htm). Results of each reaction are detected by gel electrophoresis.

ARMS primers are usually 30 bases in length and the GC content should be about 50%. It is important that the nucleotide mismatch is always at the 3' terminal end of a primer. When referring to the alleles involved in a SNP (the inherent mismatch), there are two main types of mismatches, strong and weak. A strong mismatch is a SNP involving a G→A or a C→T transition and a weak mismatch is one that involves a C→A or a G→T transversion. In order to make priming more specific in an ARMS reaction, it has been suggested that a second mismatch (an artificial mismatch) be introduced into the allele specific primers. This second mismatch usually occurs one or two bases from the 3' end. When introducing a second mismatch it is important to take into consideration what type of mismatch, strong or weak, the SNP is classified as. An initial strong mismatch will require a weak second mismatch whereas a weak inherent mismatch will call for a strong second artificial mismatch (Ye et al. 2001).

2.7.4 Agarose Gel Electrophoresis

Electrophoresis is a process by which charged molecules migrate through an electric field. The rate of migration is dependent upon the size of the molecules and their electric charge. Gel electrophoresis involves loading the DNA mixed with a loading buffer which weights the DNA and prevents it from floating out of the wells, into prepared wells of a gel. A gel consists of a network of fibrils and the researcher can control the pore size during gel preparation. The DNA molecules migrate through the pores of the gel under the influence of an electric current, the negatively charged molecules move from the negative to the positive electrode. Larger fragments will migrate at a slower rate than smaller fragments thus separating fragments by size. The loading buffer (which includes a tracking dye) is coloured and allows one to visually track the movement of dye through the gel. Ethidium bromide, which is an intercalating agent, is added to gel mixtures.

Ethidium bromide binds to the DNA and fluoresces under ultraviolet light allowing one to visualise the DNA fragments (Madigan et al. 2000). A DNA ladder with known standard sizes is always loaded in one of the wells of a gel. This ladder is used to estimate the size the DNA fragments.

Agarose gels are prepared with powdered agarose and 1 X Tris Borate EDTA (TBE). The mixture is heated until the agarose has completely dissolved, and once slightly cool it is poured into a horizontal tray where combs are inserted to create wells (see Appendix D). The pore size of gels can vary depending on agarose concentration. High percentage gels have smaller pores than low percentage gels. The DNA molecules suspended in a gel migrate through the pores at rates dependent upon their molecular weight. Small molecules will migrate faster than large ones. High percentage gels are therefore useful when separating very small fragments as the decreased pore size slows the migration of the small DNA molecules down thus separating them effectively.

Agarose gel electrophoresis was used to separate and estimate the size of the DNA fragments generated by PCR and restriction enzyme digestion. Five microlitres of each PCR product was mixed with Ficoll dye and then run on a 3% agarose gel and 15 µl of each digestion reaction was mixed with Ficoll dye and run on a 4% agarose gel. All gels were run at a voltage of 4 V/cm. Depending on what sizes were expected, the 1Kb Plus (from Invitrogen) or HyperLadder V (from Bioline) DNA markers were used as size standards. Appendix C lists the fragment sizes of each DNA marker and Appendix D indicates how the 1Kb Plus DNA ladder and Ficoll dye were prepared. HyperLadderV was manufactured in a ready-for-use state.

2.7.5 DNA Sequencing

DNA sequencing was utilised to confirm the RFLP and ARMS results. The ABI 3130 Genetic Analyzer (capillary sequencer) by Applied Biosystems that applies the Sanger dideoxy-chain termination method was utilised.

Prior to cycle sequencing the PCR product was purified. This purification step was performed using the Nucleospin Extract II kit by Macherey-Nagel. The cycle sequencing reaction is similar to a PCR reaction in that it requires a thermostable DNA polymerase and cycles of fluctuating temperatures resulting in DNA denaturation, primer annealing and DNA synthesis. However, there are some significant differences between cycle sequencing and conventional PCR. Firstly, cycle sequencing requires only one primer, either the forward or reverse as opposed to both. This results in linear amplification of DNA versus PCR's exponential amplification. Secondly, dideoxynucleotides (ddNTPs) labelled with fluorophores are added to serve as chain terminators, these differ from dNTPs in that they lack a 3'-hydroxyl group. The incorporation of a ddNTP into a growing DNA chain causes termination of chain synthesis, as a 5' to 3' phosphodiester bond cannot be formed without a 3' hydroxyl. The reaction involving the incorporation of ddNTPs will therefore produce molecules that vary in length. Each of the four ddNTPs is labelled with a different fluorophore (Wilson and Walker, 1995; Strachan and Read, 2004).

Before a cycle sequencing product can be size fractionated it was purified, vacuum dried and denatured. Once these three processes have been completed the cycle sequencing product is ready to be loaded into the automated ABI 3130 Genetic Analyzer. Purification was carried out using the Qiagen Di-ex spin columns.

The 3130 instrument requires samples to be loaded in 96-well plates. Minute volumes of each DNA sample are sucked up into individual capillary tubes filled with gel. Electrophoresis occurs and the DNA fragments migrate through the tubes from the cathode to the anode. During electrophoresis the different sized DNA products pass by an optical sensor that detects which ddNTP terminated them by identifying the different

fluorescence signals emitted. The output from the optical sensor is recorded onto a computer and an intensity profile is created (Wilson and Walker, 1995). An electropherogram is the physical output of sequence data that illustrates the different base specific intensity profiles.

For each SNP, two samples of each genotype found were sequenced in order to confirm the RFLP and ARMS results. For every sample, a forward sequence reaction as well as a reverse sequence reaction was done. Occasionally, only one of the reactions worked. In the results section both the forward and reverse sequences are presented if both reactions were successful otherwise the results of one of the reactions are shown. See Appendix C for the cycle sequencing protocol and purification steps.

2.7.6 Standard Laboratory Procedures

For each RFLP the same course of action was employed in order to detect the various genotypes. This included:

i. PCR reaction to generate desired PCR product
ii. Agarose gel electrophoresis confirming the correct size of the PCR product
iii. Specific restriction enzyme digestion
iv. Agarose gel electrophoresis to detect the digestion products (different genotypes)
v. DNA sequencing to confirm genotypes generated by enzyme digestion

ARMS detection involved the following procedures:

i. Two PCR reactions per sample (differing in the forward primer used)
ii. Agarose gel electrophoresis to detect whether or not a PCR product was produced in each reaction
iii. New PCR reaction using an upstream primer and the common reverse primer (no control primers) to generate a single PCR product encompassing the SNP site for sequencing purposes
iv. DNA sequencing to confirm the different genotypes obtained

2.8 ANALYSIS OF *ASTROTACTIN* AND *REELIN* SNPs

The genotypes obtained for the *ASTN* and *RELN* SNPs in every case and control subject are listed in table B1 in Appendix B.

2.8.1 *Astrotactin* SNPs

The first step of analysis for each SNP involved performing a temperature gradient to detect the optimal primer annealing temperature. All *ASTN* primers were synthesised by Inqaba Biotech and separate temperature gradients for each primer pair were carried out on the 2720 thermal cycler (Applied Biosystems). The same temperature gradient was used on all four primer pairs; the gradient was centred at 57^0C and fluctuated by 7^0C in either direction. Control DNA was used for each temperature gradient. Once PCR products from the temperature gradients were run on 3% agarose gels at 4 V/cm, the brightest and most defined band of the correct size determined the optimal annealing temperature.

Interestingly, all four *ASTN* PCR reactions performed optimally at an annealing temperature of 63.4^0C. All four PCR reactions were carried out on the GeneAmp thermal cycler (Eppendorf). The following PCR protocol was therefore used for all four reactions:

Initial denaturation	95^0C	7 mins	
Denaturation	94^0C	30 secs	30 cycles
Annealing	63.4^0C	30 secs	
Extension	72^0C	40 secs	
Final extension	72^0C	5 mins	
Hold	15^0C	∞	

2.8.1.1 rs6413830

ASTN's rs6413830 involves a G→A transition in the 3' UTR. The positioning of the SNP on the forward strand and the areas upon which the primers anneal are indicated below in green:

```
5'CCAAGCCTTACCTGTCCAAGGTAAGAGGTGTGGGGGTTAAAGATAATCCAGGCACAT
GGATATATATTGGTAGGCTGGAGAAGTCTATCCAAAGGAAGCATGGTTTTTTTAAAAAA
CAATAAACTCCAAGTCTCCTCTTGAATGTGGACTGCTAACTAGTCTCCCAGGTAGACTT
TTCTGAAGGTGCAGTTGACAGATGGCAT 3'
```

Forward Primer: 5' CCAAGCCTTACCTGTCCAAG 3'
Reverse Primer: 5'ATGCCATCTGTCAACTGCACCTTCAGAAAA 3'

A PCR reaction incorporating the above primers yielded a PCR product of 203 bp. The size of the PCR product was determined on a 3% agarose gel run at 4 V/cm. Prior to loading of the sample, 5 µl of Ficoll dye was added to 5 µl of PCR product. HyperLadder V was loaded as a DNA marker alongside the PCR products. Once the electrophoresis run was completed, the gel was viewed using the gel documentation system (Syngene).

This particular SNP is an RFLP, the presence of the "A" allele creates a recognition site, "AGCT" for the enzyme *Alu*I. Digestion reactions totalled 30 µl. Each reaction contained 20 µl of PCR product, 0.5 µl (5 units) *Alu*I (manufactured by Fermentas), 3 µl 10x Tango™ Buffer and 6.5 µl ddH_20. The digestion reaction took place at 37^0C for 2.5 hours and digestion products were run against HyperLadder V on a 4% agarose gel and viewed using the gel documentation system.

This particular enzyme does not have star activity and so over digestion was not a concern. It did however require thermal inactivation at 65^0C for 20 minutes. The expected results following *Alu*I digestion are illustrated in figure 2.4.

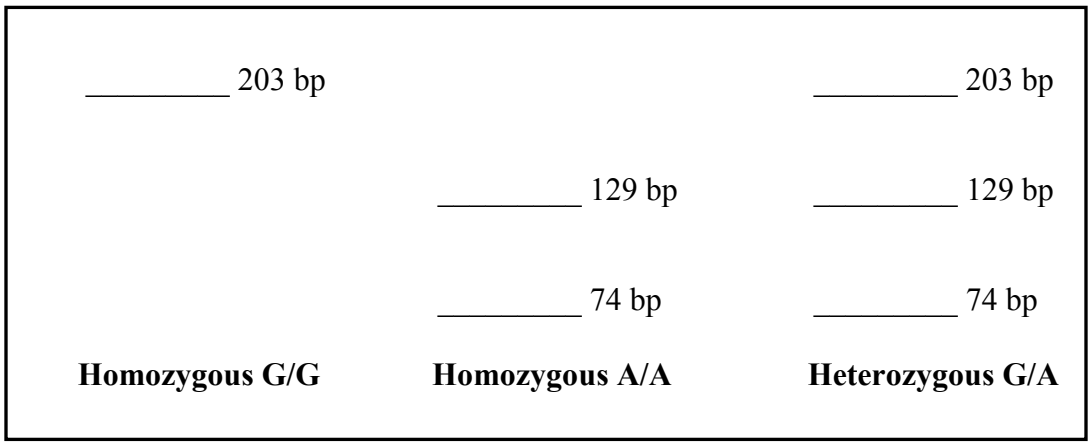


Figure 2.4: The three possible genotypes at rs6413830 following *Alu*I digestion.

2.8.1.2 rs12118933

ASTN's rs12118933 involves a G→C transition within exon 23 that causes an amino acid change. The positioning of the SNP on the forward strand and the areas upon which the primers anneal are indicated below in blue:

5′ TGCGGCAGGCCTAACAACTTTCTGGCCTCACCTCTGCTCCCAGTTGCAATGGATTTA
GGAACTGAGTGAGCACAGGGTGCCTACTTAATAGACTTTTCATGGGGAGCACGGCAGAG
CTCTTGAGCATTTTGGTAAACTTCTCAGGGAAAAAAAAATGAATGGAACAAATTAATGG
CAAAGCAAACCCCAAAGTAATCCTCTAAAGAAATATGGCACTGCATGAAGCCACTGGCT
GGCAGATTTCCCAATCATGCAGATGGAGAACATCATACTGAAGAAGTTCTGCAGGGAGC
AAGGATCACTTTCTCCCTGGTTTCGATGTTGCTGTGGAGGTTTTCATGTTCCATTAAAA
AATATTAAAAATCCACAGACCAACCCAGATGGATCCCTCCTCTTTCCTACTTCATTCTG
GCAGCAGCTCCCTGGCCTTATGGTGCTAGATCTCTTTGCTGTCCCCATAGTCGTTGTAG
GGGATACTCAGGGTCTGCTCCTCACACGTCTTCCTGAGGTCCCGGCTGAGCTCCGCCCA
GTCAAGTCCGTAGGGTTTGATCTCGCTGTAGCGGCACCCCAGGTACTTGAGTGAGCTGC
GCCGC **3′**

Forward Primer: 5' TGCGGCAGGCCTAACAACT 3'
Reverse Primer: 5' GCGGCGCAGCTCACTCAAG 3'

The product formed from a PCR reaction involving the above primers was 593 bp in size. PCR products were viewed on a 3% agarose gel using the 1 Kb Plus DNA marker. The presence of the "G" allele produces a "CCGG" recognition site for the enzyme *Msp*I.

Once it was confirmed that the correct PCR product had been amplified, digestion with *Msp*I followed. Each digestion reaction contained 10 µl of PCR product, 0.5 µl (5 units) *Msp*I (Amersham Biosciences), 3 µl 10x Buffer T, 3 µl bovine serum albumin (BSA) and 3.5 µl ddH_20. The optimal temperature for this digestion reaction was 37^0C and the incubation period was 3-5 hours. Digestion products were confirmed on a 4% agarose gel and their sizes were estimated with the use of the 1Kb Plus DNA ladder. The three possible genotypes generated from *Msp*I digestion are illustrated in figure 2.5.

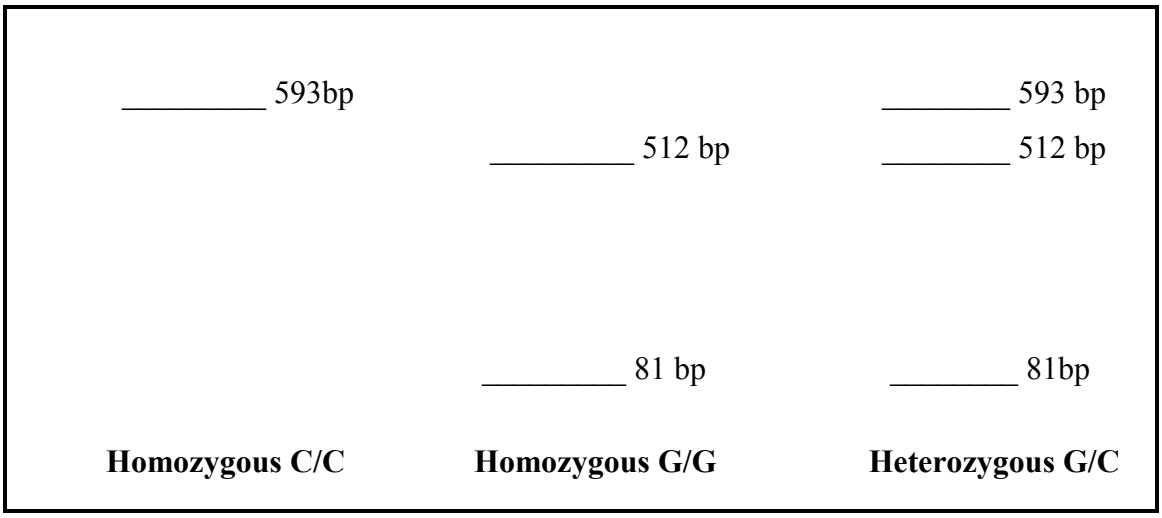


Figure 2.5: The expected genotype results at rs12118933 following *Msp*I digestion.

2.8.1.3 rs2228956

ASTN's rs2228956 is present in exon 17 of the *ASTN* gene and involves a T→C base transition, which results in an amino acid change. Depicted below is the SNP position and the DNA sequences complementary to the designed primers on the forward strand of the *ASTN* gene:

5′ CCGGCTCAGCAGGGGTGTCAGGGCTGGATGTCACATGACACAGGGGGCAGGACGAGG
GGCATCGTCCCTTGCTGTGGCACTCCATGCGGACTCCAGCCGCCATGTGCTTGGTGCCG
GAGTCTGACAAGCTGGTGATGTATTCTGGGAATGTCAGCACCTTGGGGTCTCTTTCCCG
CTCCTCAGACTCATCTGATGGGGAGTTGCCTGCAGACACAAAATGGAAAGAAACAGTGA
AGGAGTCAAAAAAGTAAGGACTATCATGACCTCGTGGCCCCTGAGTGCTGGGGTCCAAT
CTTTGGGATCAGAAACAGCCACTGAAGGGGTAAGAATGTGATTCTGCCATCCCACTTAA
TTTGGGGCACCCAACCTCCCAGCGACACA **3′**

Forward primer: 5' CCGGCTCAGCAGGGGTGTC 3'

Reverse primer: 5' TGTGTCGCTGGGAGGTTGG 3'

A product of 381 bp was amplified using these primers. As before, PCR products were checked on a 3% agarose gel. HyperLadder V was used as the DNA marker. Rs2228956 is also an RFLP, when the "C" allele is present *Hha*I recognises the "GCGC" sequence and cleaves it.

Digestion of the 381 bp PCR product involved 20 µl of PCR product, 0.5 µl (5 units) of *Hha*I (Fermentas), 3 µl 10x Tango™ Buffer and 6.5 µl ddH_20. The *Hha*I reaction was incubated at 37^0C for 3 hours after which the enzyme was inactivated at 80^0C for 20 minutes. Digestion products were run on a 4% agarose gel at 4 V/cm against HyperLadder V. The expected results for the *Hha*I digestion are indicated below in figure 2.6.

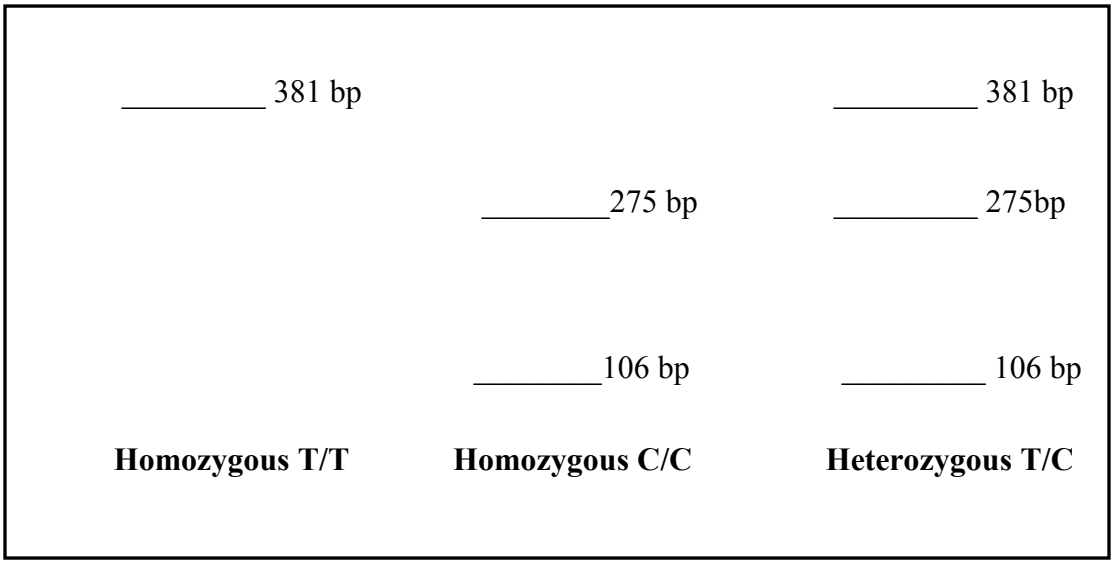


Figure 2.6: The expected band pattern on a gel of the genotypes at rs2228956 following *Hha*I digestion.

2.8.1.4 rs11587640

ASTN's rs11587640 is a non-synonymous coding SNP located in exon 3. The allele transition involved is G→T. The position of the SNP on the forward strand and the areas of primer binding are depicted in purple below:

```
5′CTGGTGTGAGGTCCATCCCCGACTTTTCATTGCAGCCCTGCAGGAGTCGAGGGTGC
GCGTGACCTGGCTGGCAAAGTCCTCCCCTCCGTTCATGCACTCCCTCTGCAGATGGTGG
CGCAGATCAGTGATGTCATACTCATAGCCGTCCAGGATAGGTGTCTCCCGGATGCT 3′
```

Forward Primer: 5' CTGGTGTGAGGTCCATCCCCG 3'
Reverse Primer: 5' AGCATCCGGGAGACACCTAT 3'

A PCR product of 172 bp was synthesised from the above primers. All PCR products were viewed as described previously. This SNP is an RFLP and can therefore be detected by enzymatic digestion. The "G" allele creates a recognition site, "CCTGCAGG" for *Sbf*I. A digestion reaction involved, 20 µl of PCR product, 0.5 µl (5 units) *Sbf*I, 3 µl 10x NEBuffer 4 and 6.5 µl ddH_20. New England Biolabs manufactured *Sbf*I and NEBuffer 4. The reaction was incubated at 37^0C for 3 to 16 hours (see results section) after which the enzyme was inactivated for 20 minutes at 65^0C. The predicted results after *Sbf*I digestion are specified in figure 2.7 below.

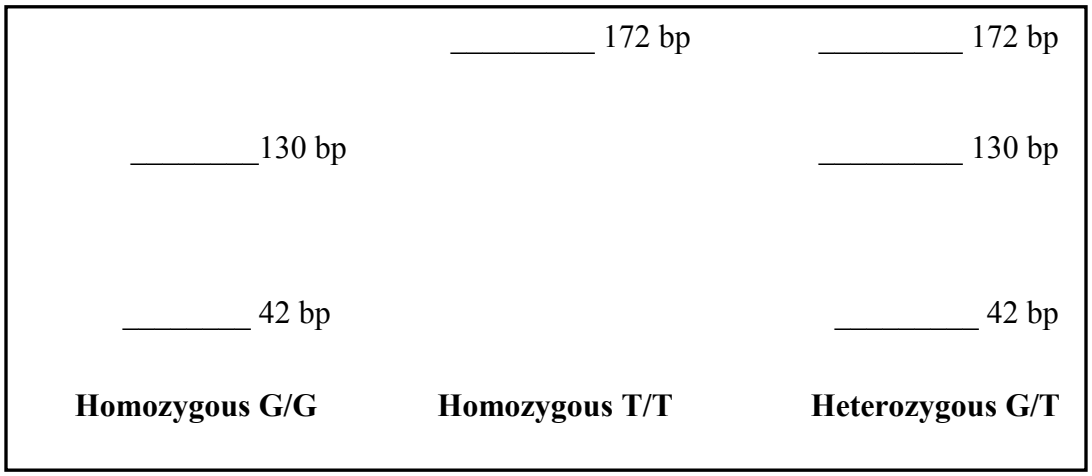


Figure 2.7: The three predicted genotypes at rs11587640 generated by *Sbf*I digestion.

2.8.2 *Reelin* SNPs

Unlike the *ASTN* SNPs, the PCR protocols for the four *RELN* SNPs all differed from each other. The individual PCR programmes are listed under each separate heading below. Inqaba Biotech manufactured the *RELN* primers. All PCR reactions were optimised with a temperature gradient using the Applied Biosystems 2720 thermal cycler. Three of the four *RELN* SNPs were RFLPs and were therefore detected through enzyme digestion. For the three RFLPs, 5 µl of each PCR product was loaded with 5 µl of Ficoll dye onto 3% agarose gels run at 4 V/cm. Digestion products were run on 4% agarose gels at the same voltage. I Kb Plus DNA marker or HyperLadder V were used depending on the size of the DNA fragments. All gels were viewed using the Syngene gel documentation system. The ARMS system only required a 3% agarose gel upon which the PCR products were run.

2.8.2.1 rs607755

The rs607755 SNP entails a non-synonymous T→C alteration. The SNP is highlighted below and the positioning of the primer pair on the reverse strand is illustrated in green:

5′GTCCCGAATAGCCTCTCCTTGTAGGTTGGTGCTGATTTTGGCCTTCCTGCCTTG
CCTGCCTAGTGATAGCTTCATAGATTTCCTTTTGGGGTCTTTGGTTTTCCCAACTT
CTAGGTGTTTAAACCGCAGAGCTTTTTACCCTTGTCCCTTAGTAGTTTGAGGGACA
AAGCAAATATAATCAACTGCTCATTTAAAATTTAGTTTTAGGATACATTATTCTTC
ATAAATACAGCATGTTGGCACTTGTGTTTAGCTAAATGTACTGTTGTAGTATATTT
CTTTCCTAAATACTTCCTTTTTCTTTTAGCTGAAATACATAGTGACAGCATTATCC
TGAGAGATGACTTTGACTCCTACCACCAACTGCAATTAAATCCAAATATATGGTAA
GTTACAGGAACATTCCTCACTTTGGTTTGCTTAATGTAACAAATGATGCTGACTGC
TCCTTTAAGGAGCAGTTACTTAAAATTAAGTTGATCTTTAGTGCTAAGCTATCAGT
GAGACTTGTTAAGCTATGGAACTGGGC **3′**

Forward Primer: 5’ GTCCCGAATAGCCTCTCCTT 3’

Reverse Primer: 5’ GCCCAGTTCCATAGCTTAACA 3’

A temperature gradient was performed to find the optimum annealing temperature for the PCR reaction. The gradient was centred at 58^0C and fluctuated by 5^0C in either direction.

An optimum annealing temperature of 62.6^0C was obtained and the PCR reaction yielded a 529 bp product. The PCR parameters were set at:

Initial denaturation	95^0C	7 mins	
Denaturation	94^0C	30 secs	30 cycles
Annealing	62.6^0C	30 secs	
Extension	72^0C	40 secs	
Final extension	72^0C	5 mins	
Hold	15^0C	∞	

The change of allele from a T→C results in a recognition sequence, "CAGCTG" for *Pvu*II. This enzyme was used to digest the 529 bp PCR product at 37^0C for 2 hours. Each digestion reaction included 20 µl of PCR product, 0.5 µl (5 units) *Pvu*II, 3 µl Buffer G and 6.5 µl ddH_20. The enzyme and buffer were both manufactured by Fermentas. The expected DNA fragments resulting from digestion with *Pvu*II, are illustrated in figure 2.8.

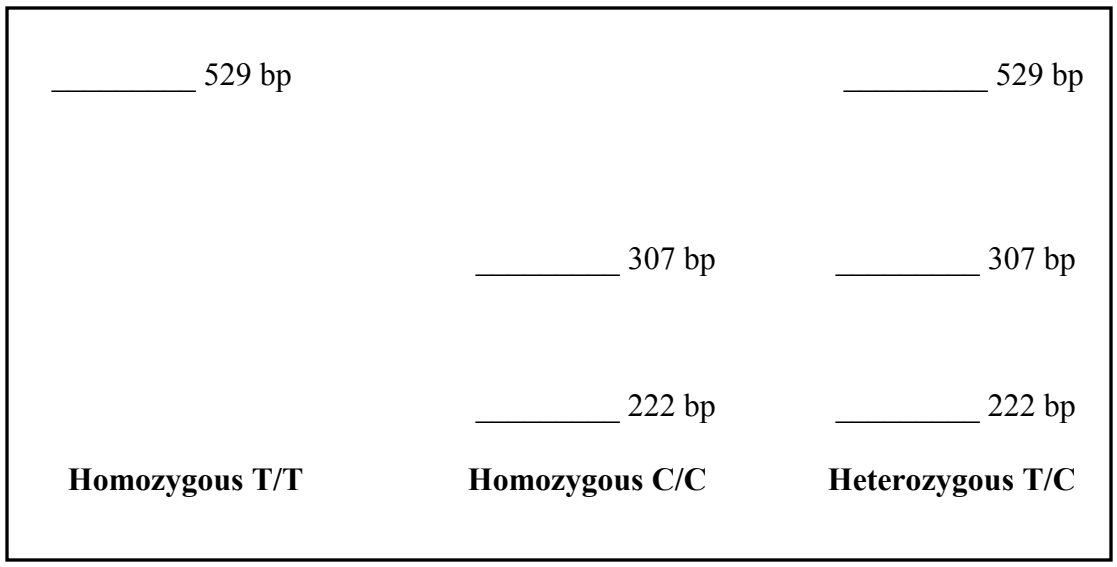


Figure 2.8: The expected fragments after *Pvu*II digestion of the PCR products containing rs607755.

2.8.2.2 rs3025963

RELN's rs3025963 consists of an A→T transition. This SNP is also an RFLP, when the "A" allele is present a recognition site, "GAATGCN" for the enzyme *Mva*12691 is produced. *Mva*12691 is the Fermentas isoschizomer for *Bsm*I. The areas of primer binding and the position of the SNP on the reverse strand can be seen in pink below:

```
5' TGTATGTGCGAGTATTGGGAAATACTCAATTTCTTTATGGCACCTCATTTGAAATTA
TTTCTTACATAATTTTAAGACTTTCTATGAAGTTTTAGCTTCATGGAAAGAAAAATCAT
GGATTGTTTCCATATCACAAATAGTTAAGATTATGACATTAGATTTCTCTGAGGTCTTT
TTCACTTACAGTGGGTGAGTAGACATATTTCACCTGAGGAAAACAGAAGCTGGGTTTGA
TTTGTGCCTAACTTTCTTTCAGGAATGCCTTCCAAGTATGCCAAGTTGTCA 3'
```

Forward Primer: 5' TGTATGTGCGAGTATTGGGAAA 3'

Reverse Primer: 5' TGACAACTTGGCATACTTGGA 3'

A temperature gradient was performed on the above primers; the middle temperature was set at 59.5^{0}C and fluctuated by 5^{0}C in both directions. An optimum annealing temperature of 63.1^{0}C was obtained.

The PCR parameters for the amplification of this SNP were set at:

Initial denaturation	95^{0}C	7 mins	
Denaturation	94^{0}C	30 secs	30 cycles
Annealing	63.1^{0}C	30 secs	
Extension	72^{0}C	40 secs	
Final extension	72^{0}C	5 mins	
Hold	15^{0}C	∞	

A PCR product of 285 bp was produced. Once the size of the product had been confirmed, 0.5 µl (5 units) of *Mva*129691, 3 µl of 10x Buffer R and 6.5 µl ddH_20 were added to 20 µl of PCR product and incubated at 37^{0}C for 2 hours. Fermentas supplied both the enzyme and buffer. The expected band patterns after digestion are illustrated in figure 2.9.

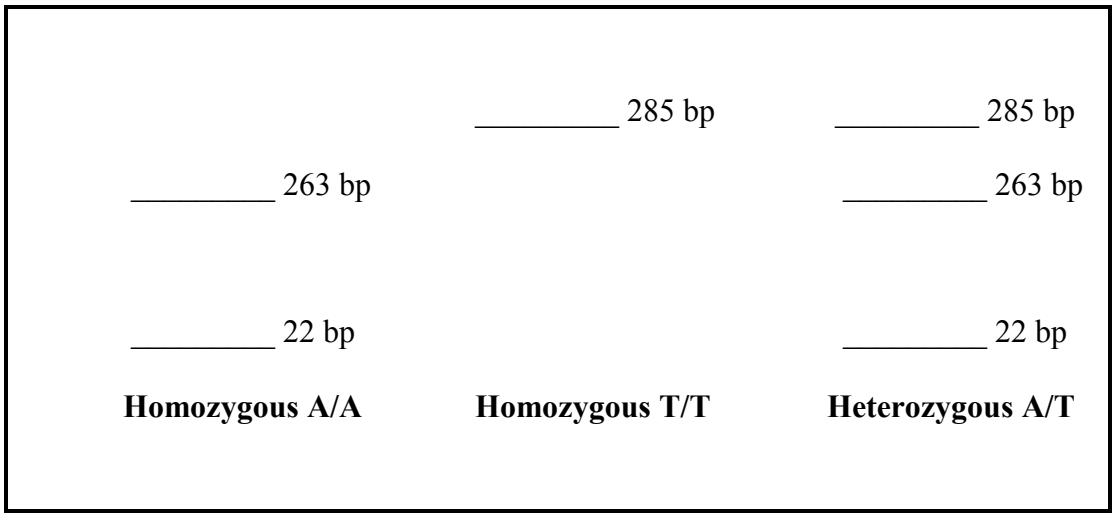


Figure 2.9: Expected band sizes for the different rs3025963 genotypes generated by *Mva*12691 digestion.

2.8.2.3 rs3025966

SNP rs3025966 involves a T→G allele conversion. It too is an RFLP whose "G" allele produces an *Mva*12691 recognition site (5' "NGCATTC" 3'). The areas of primer binding and the position of the SNP on the genomic sequence's reverse strand are indicated in red below:

5′CTACCAGGTGATGCAAAGCAGTTTGGAATTCAGTTCAGATGGTGGCAACCGTATCATT
CTTCCCAGAGAGAAGATGTATGGGCTATTGATGAGATTATCATGACATCTGTGCTTTTCA
ACAGCATTAGTCTTGACTTTACCAATCTTGTGGAGGTCACTCAGTCTCTGGGATTCTACC
TTGGAAATGTTCAGCCATACTGTGGCCACGACTGGACCCTTTGGTAAG 3′

Forward Primer: 5' CTACCAGGTGATGCAAAGCA 3'

Reverse Primer: 5' CTTACCAAAGGGTCCAGTCG 3'

The temperature gradient performed on the above primers fluctuated by 5^0C in both directions around a central temperature of 58^0C. An annealing temperature of 57.8^0C was optimum for the reaction.

The PCR parameters for the reaction were set at:

Initial denaturation	95^0C	7 mins	
Denaturation	94^0C	30 secs	30 cycles
Annealing	57.8^0C	30 secs	
Extension	72^0C	40 secs	
Final extension	72^0C	5 mins	
Hold	15^0C	∞	

A 226 bp product was generated using the PCR parameters described above. In order to detect which allele of rs3025966 was present in each DNA sample an *Mva*12691 digest was set up following the same protocol used for rs3025963. The expected DNA bands for each genotype are shown in figure 2.10.

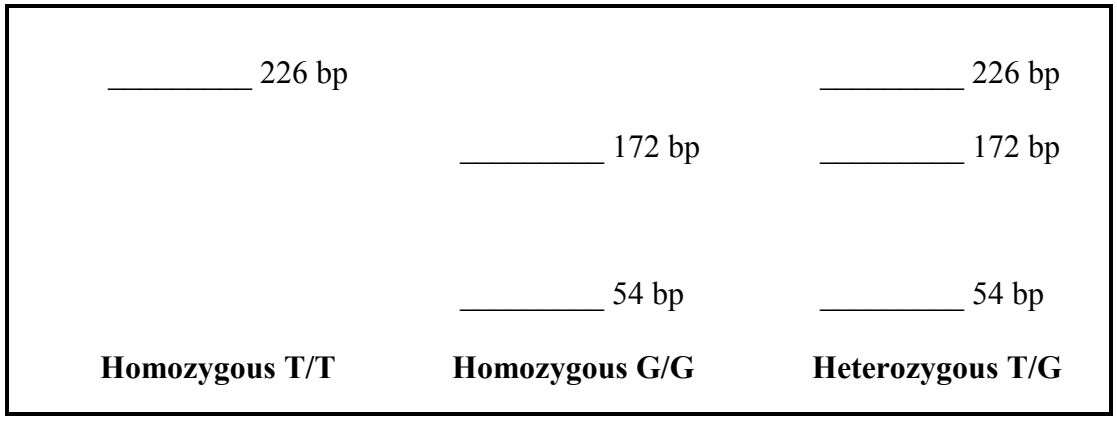


Figure 2.10: The three possible genotypes at rs3025966 resulting from *Mva*12691 digestion.

2.8.2.4 rs362691

The alleles contributing to rs362691 are C or G. For this particular SNP no restriction enzyme recognition sequence was present with either allele in place. An ARMS system was therefore designed to detect which allele was present in each sample. Two forward primers were designed, one detecting the "C" allele and the other detecting the "G" allele. The "C" or "G" allele was present at the 3' terminus of each primer. In addition to the

differing base at each 3' end, an artificial mismatch was incorporated three bases upstream from the 3' end to increase primer specificity.

In the reverse strand sequence seen below the first 30 underlined bases represent the positioning of the forward primers. Due to the C→G change being a weak mismatch, a strong second mismatch was required. The "T" allele positioned three nucleotides up from the 3' terminus was converted to a "C" in both forward primers. A common reverse primer indicated in light blue below was used for both reactions:

5' AGTTTACACAGTGGAGGAGAGTCATAGTGCTTCTTCCCCAGAAAACTTGGTGAGATG
ACATCTTTAGATATGGCTGTCACTTTAACTTTGACTATCCCTTTTGATTCAATTCTGTC
TCTATTGATTCAAGTATGACTTGAATATGAAGGCAAGTTAAGCACAATAGCATGATGTT
TATTCTTTTTATTTTATTATCAAAAGACCAAAACCTTTCATTTATGTGAATGATATTTG
AAAAAGAGAGCTAATAACAAATATGAAAATGTGGTTTCCAAACTGTCCTAGCCTTCTCT
GACTTTCCTACTTGCCCCAGCTTTTTCTTTTTATGGGAGAACTCAATGAGTCGGTTACT
TTCAGT 3'

Forward Primer for the "C" allele: 5'AGTTTACACAGTGGAGGAGAGTCATAGCGC 3'
Forward Primer for the "G" allele: 5'AGTTTACACAGTGGAGGAGAGTCATAGCGG 3'
Common Reverse Primer: 5' ACTGAAAGTAACCGACTCATTGAGTTCTCC 3'

A set of control primers was required for the ARMS system in order to confirm whether or not each PCR reaction was successful. Control primers should ideally amplify a sequence of DNA on another chromosome. The ARMS primers targeted chromosome seven and so control primers, which amplified a product of 172 bp on chromosome one were used.

A temperature gradient that incorporated the control primers, the common reverse primer and the forward primer for the "C" allele was performed. The middle temperature was set at 63^0C and increased and decreased by 5^0C. An optimum temperature of 63.4^0C was reached. The other PCR parameters were kept the same as those used in previous PCR reactions except that the final extension time was increased to seven minutes.

Two reactions were carried out per sample, a "C allele" reaction and a "G allele" reaction. Both reactions contained the control primers and the common reverse primer but differed in the forward primer used. In the case of a C/C homozygote a band of 358 bp as well as the 172 bp control band would be amplified from the "C allele" reaction only. The "G allele" reaction should only have produced the control band of 172 bp. If however, a sample were C/G heterozygous then both reactions should have produced a 358 bp product in addition to the control product of 172 bp.

The PCR parameters was set as:

Initial denaturation	95^0C	7 mins	
Denaturation	94^0C	30 secs	30 cycles
Annealing	63.4^0C	30 secs	
Extension	72^0C	40 secs	
Final extension	72^0C	7 mins	
Hold	15^0C	∞	

For sequencing purposes of the rs362691 genotypes, another forward primer located upstream from the SNP had to be designed. The following primer was designed and used to produce one product of 441 bp using the same PCR parameters as above:

```
5′GTGCCTAACTTTCTTTCAGGAATGCCTTCCAAGTATGCCAAGTTGTCAGGAATTTAC
ATCAGCAAGTATTTACCATGCCAGTGAGTTTACACAGTGGAGGAGAGTCATAGTGCTTC
TTCCCCAGAAAACTTGGTGAGATGACATCTTTAGATATGGCTGTCACTTTAACTTTGAC
TATCCCTTTTGATTCAATTCTGTCTCTATTGATTCAAGTATGACTTGAATATGAAGGCA
AGTTAAGCACAATAGCATGATGTTTATTCTTTTTATTTTATTATCAAAAGACCAAAACC
TTTCATTTATGTGAATGATATTTGAAAAAGAGAGCTAATAACAAATATGAAAATGTGGT
TTCCAAACTGTCCTAGCCTTCTCTGACTTTCCTACTTGCCCCAGCTTTTTCTTTTTATG
GGAGAACTCAATGAGTCGGTTACTTTCAGT 3′
```

Upstream Forward Primer: 5' GTGCCTAACTTTCTTTCAGGAATGCCT 3'

Common Reverse Primer: 5' ACTGAAAGTAACCGACTCATTGAGTTCTCC 3'

2.9 STATISTICAL ANALYSES

The genotyping data generated by means of the techniques described above were statistically analysed in order to determine whether any significant results, which might indicate a role of the SNP in FAS, were obtained. Once genotyping data had been generated for each SNP, genotype and allele frequencies were calculated by simple gene counting. The Chi-square test was performed manually and results were then confirmed using the "Genetics" package of the computer program called "R". Due to the manual results and those generated by the computer being similar, the computer-generated results were used in this study. Computer-based statistical programs were run by Dr. Lize van der Merwe of the Biostatistics Unit at the Medical Research Council (MRC) in Cape Town, South Africa. R: A language and environment for statistical computing (R Development Core Team, 2005) and the R "Genetics" package (Warnes and Leisch, 2005) were utilised for the majority of the tests unless otherwise stated. *P*-values were calculated for all tests. A *P*-value smaller then 0.05 indicated a significant effect.

2.9.1 Allele and Genotype Frequencies

Allele and genotype frequencies were calculated for each *ASTN* and *RELN* SNP that data had been generated for. Case and control frequencies were calculated separately and frequency values were rounded to three decimal places. Rounding results in the frequency values do not always add up to one.

2.9.2 Hardy-Weinberg Equilibrium

Hardy-Weinberg equilibrium (HWE) holds in a large randomly mating population where migration and natural selection are absent. When a population is in HWE the genotype frequencies are expected to remain constant from generation to generation and at any locus, the genotype frequencies are a function of allele frequencies. HWE is an important phenomenon in genetics and it is the norm for geneticists to check whether observed genotypes conform to the expectations of HWE. Any deviations from HWE may suggest inbreeding, population stratification or sampling or genotyping error. Where deviations

from HWE are present in case individuals (those affected with a particular disease), an association between the marker and disease susceptibility might be implied (Wigginton et al. 2005).

Chi-square (X^2) analysis was used to assess whether the groups demonstrated HWE. These calculations were performed manually and then confirmed using the computer program "R". Only the *P*-values are reported in the dissertation. If the *P*-value for a specific locus is above 0.05, the populations from which the individuals originated is said to be in HWE at that locus, whereas any *P*-values below 0.05 indicate Hardy-Weinberg disequilibrium. The equation used to calculate X^2 is as follows:

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

2.9.3 Testing for Association

To test whether a SNP or several SNPs together are associated with the disease three tests were performed; the Fisher's Exact test, haplotype analysis and logistic regression.

2.9.3.1 Fisher's Exact Test

The Fisher's Exact test was performed to compare counts of genotypes or alleles at a specific locus between cases and controls. Exact *P*-values calculated and significant associations have *P*-values of less than 0.05 (Fisher, 1922). Fisher's Exact calculations were carried out using the "Genetics" package of the computer program "R."

2.9.3.2 Haplotype Analysis

Since families were not analysed and some haplotypes cannot be determined with certainty, statistical inference was used to calculate the most likely haplotypes. The R package used to generate haplotypes for the *ASTN* and *RELN* genes was "haplo.stats" (Sinnwell and Schaid, 2005). Haplotypes have to be on the same chromosome and for this reason *ASTN* SNPs were analysed together and *RELN* SNPs were analysed together. The

program reported which haplotype for each gene was more common amongst the two study groups. A negative haplotype score indicated that the haplotype was more common amongst controls whereas a positive haplotype score specified association of the haplotype with cases. A global simulated *P*-value was established by haplo.stats, a *P*-value of less than 0.05 suggested there was a significant difference between the haplotype frequencies inferred in the cases versus the controls.

2.9.3.3 Logistic Regression

The idea behind a logistic regression model is to describe the differences between two groups, in this instance, cases and controls, as a function of other variables, in this situation, genotypes at the individual SNPs. The results generated from a logistic regression analysis indicate what the probability of being a case or a control is as a function of the SNP genotypes. The coefficients of the model are estimated. A negative coefficient is associated with being a control whereas a positive estimate is associated with being a case.

A global *P*-value is presented for the model. This is an indication of how well the model fits the observed data. A *P*-value is also presented for each possible term inside the model. If the genotype has a *P*-value less than 0.05, it means that the probability at that genotype is independently significantly different from the probability at the baseline genotype for that SNP. Independently significant means after adjusting for all other SNPs (in this instance) in the model.

A backward stepwise procedure was used to select a logistic regression model. The procedure started with a logistic regression model expressing the probability of being a case or a control as a function of the genotypes at all the SNPs. The Akaike information criterion (AIC) is a measurement that describes how well the model fits the data, the lower the AIC value the better the fit of the model. The AIC is used to determine when the best model has been reached: the procedure discards terms until the lowest possible AIC is reached, indicating the best combination of SNPs that differentiate between cases and controls. In this study the initial variables entered to determine the best fitting model were

affection status, sex of individual and genotypes at all informative *ASTN* and *RELN* SNPs. The R package "Mass" (Venables and Ripley, 2002) was used for this analysis.

2.9.4 Linkage Disequilibrium

LD is usually described as being the non-random association of neighbouring alleles. These correlations are said to indicate haplotypes that have descended from distinct, ancestral chromosomes. A measure of LD therefore denotes whether or not a region of the genome cosegregates more often than expected with another region, thus behaving dependently or completely independently. Sometimes this association may be due to factors like the physical distance between the loci (Reich et al. 2001; Ardlie et al. 2002).

There are several measures of LD, one of which is *D'*. A *D'* value of one would indicate complete LD suggesting that genetic markers are tightly linked with one another. A *D'* value of zero corresponds to no LD between markers. It is however well recognised that unlinked alleles may show LD due to population phenomena such as population stratification or selection.

The equation used to calculate D' is: $\boldsymbol{D' = D/D_{max}}$ whereby *D* equals the difference between the observed frequency of a two-locus haplotype and the expected frequency if the alleles were segregating randomly. An equation used to describe *D* is as follows:

$$D_{ij} = P_{ij} - p_i q_i$$

P_{ij} is the observed haplotype frequency at locus "i" and locus "j" and $p_i q_i$ represents the expected frequencies at locus "i" and locus "j" (Hartl and Clark, 1989).

D is then divided by its maximum possible value, given the allele frequencies at two loci (Ardlie et al. 2002). D_{max} is the maximum value that *D* can reach with the given allele frequencies (Lenwontin, 1964).

"Since $P_{12} = p_2q_1 - D \geq 0$, it follows that $D \leq p_2q_1$. Therefore, the maximum possible value of D (D_{max}) is given by: $D_{max} = \min(p_1q_2, p_2q_1)$ whereby "min" is interpreted as "the smaller of " (Hartl and Clark, 1989)."

The calculations of D, D_{max} and subsequently D' were performed using the R "Genetics" package.

2.9.5 Post-test Power Calculations

Where a significant association was seen between a genetic marker and disease susceptibility, a post-test power calculation was conducted. A post-test power value measures a new study's ability, with the same sample size, to find and label as statistically significant, a specific effect. Post-test power depends on the specific effect size observed in one's study. A post-test power value indicates the probability of a true positive association, the higher the power value the more likely one is to observe a difference with a small effect size. An ideal post-test power value is around 75%, this would suggest that the chance of one finding the same significant association in a new study using the same sample size is relatively high (Long et al. 1997).

CHAPTER 3: RESULTS

3.1 GENOTYPING OF *ASTROTACTIN* SNPs

The process of determining the cases' and controls' genotypes for the *ASTN* SNPs under investigation involved two main steps: PCR followed by restriction enzyme digestion. Confirmatory sequencing was done on a few samples to validate the genotypes. The following gel pictures and electropherograms represent the findings for SNPs rs6413830, rs12118933, rs2228956 and rs11587640. Gel photographs are depicted below in black and white, the white illustrating the DNA products. Electropherograms are depicted in colour with each base represented by a different colour.

3.1.1 rs6413830

ASTN's rs6413830 causes a G to A amino acid change. Figure 3.1 illustrates the PCR products of 203 bp incorporating this SNP. Figure 3.2 depicts the visual genotyping results after *Alu*I digestion of the 203 bp products. Sequencing was performed to confirm the results observed through *Alu*I digestion. These sequencing results are illustrated as a series of electropherograms in figure 3.3.

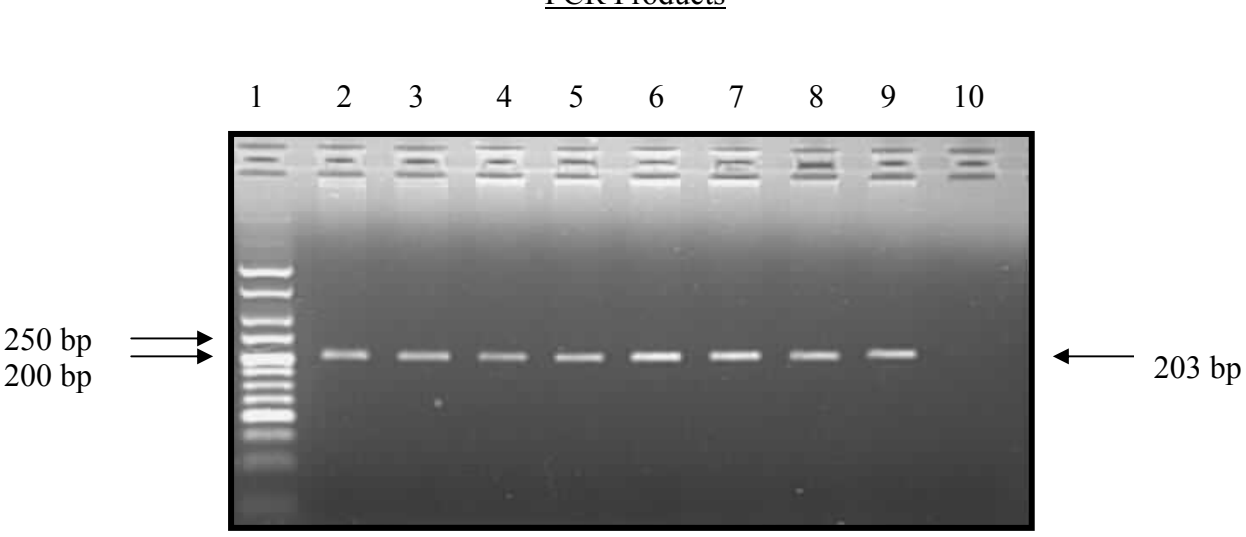


Figure 3.1: 203 bp PCR products encompassing *ASTN's* rs6413830 SNP run on a 3% agarose gel. **Lane 1:** HyperLadder V. **Lanes 2-9:** PCR products from case and control samples. **Lane 10:** Negative control (no DNA).

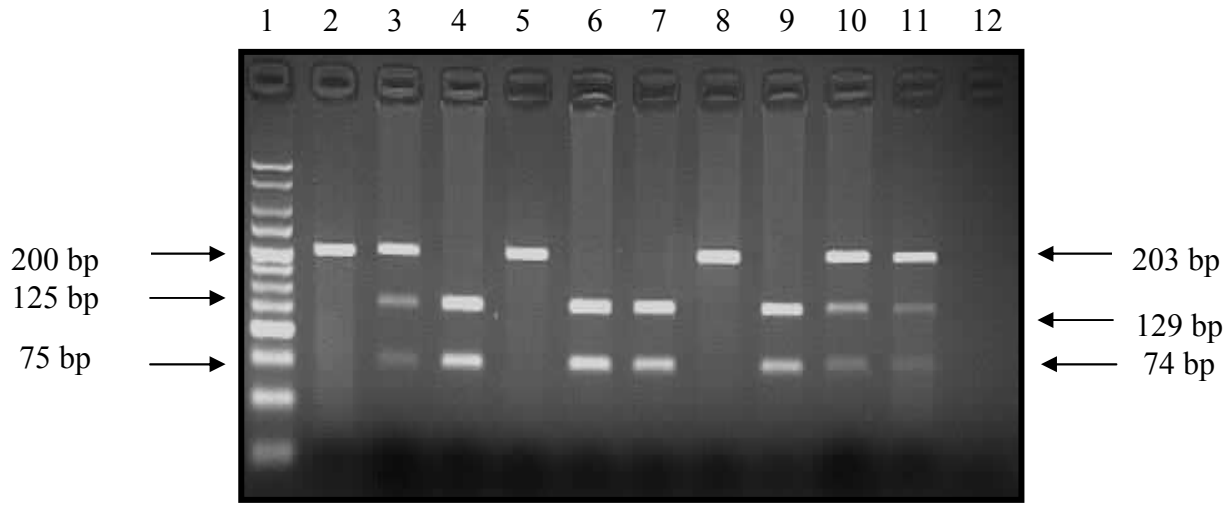


Figure 3.2: Digestion of rs6413830 PCR products with *Alu*I viewed on a 4% agarose gel. **Lane 1:** HyperLadder V. **Lane 2:** Uncut PCR product (203 bp). **Lane 3:** Positive G/A heterozygote. **Lane 4:** Positive A/A homozygote. **Lane 5:** Positive G/G homozygote. **Lanes 6-11:** Control sample digests. **Lane 12:** Negative control (no DNA). "Positive" genotypes were confirmed by sequencing.

Sequencing Results

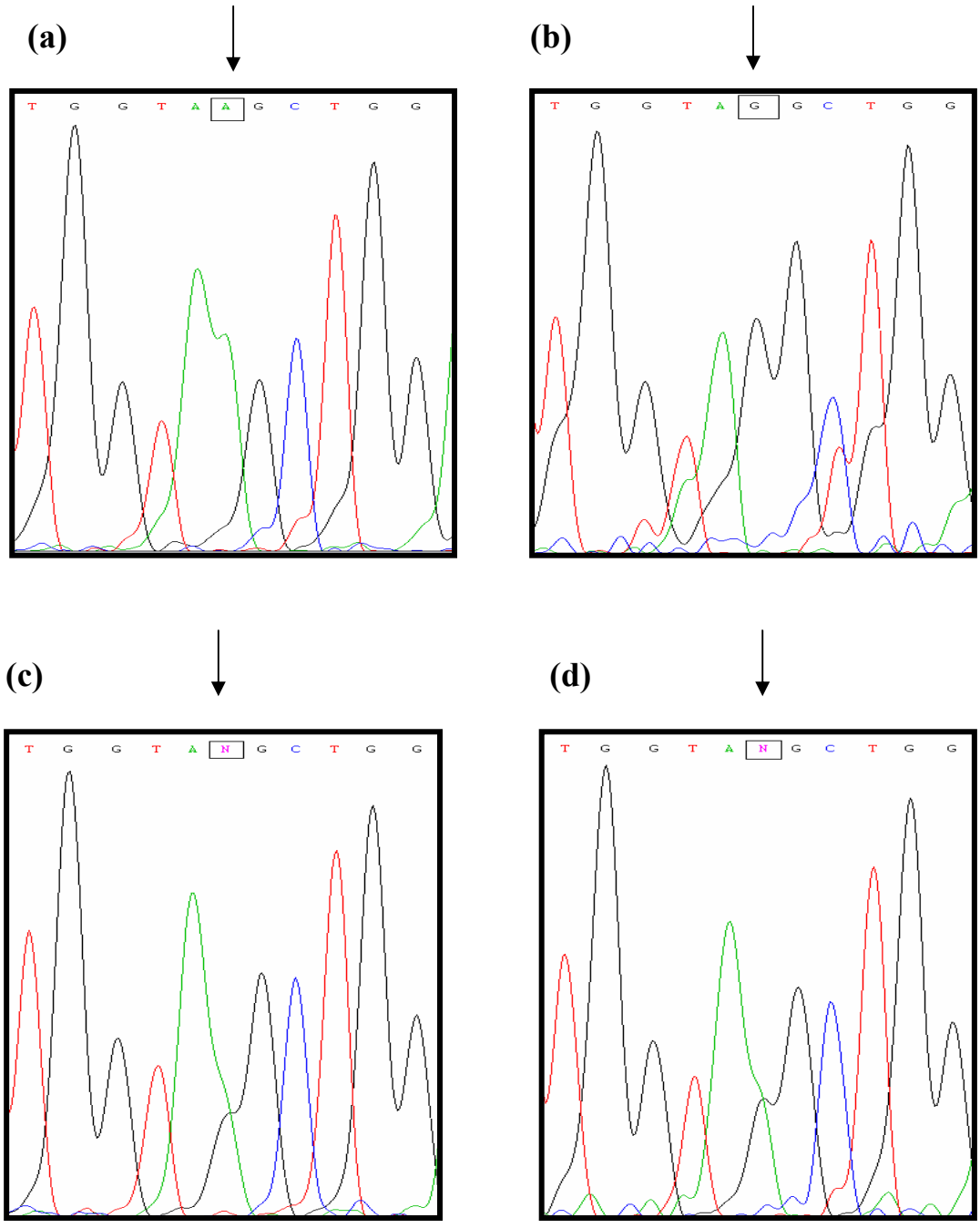


Figure 3.3: Electropherograms (forward strands) confirming genotyping results obtained from *Alu*I enzyme digestion of rs6413830. The arrows indicate the SNP position. **A/A homozygote:** Control sample NPC004 **(a)**. **G/G homozygote:** Control sample NPC018 **(b)**. **G/A heterozygotes:** DNA samples sequenced were NPC021 and NPC027 respectively **(c & d)**.

3.1.2 rs12118933

ASTN's rs12118933 involves a G to C base alteration and is responsible for an amino acid change. Figure 3.4 depicts PCR products of 593 bp within which the SNP is included. Figure 3.5 illustrates the various genotypes indicated by the different banding patterns produced after the PCR products were digested with *Msp*I. Electropherograms confirming the genotype results produced by *Msp*I digestion are represented in Figure 3.6

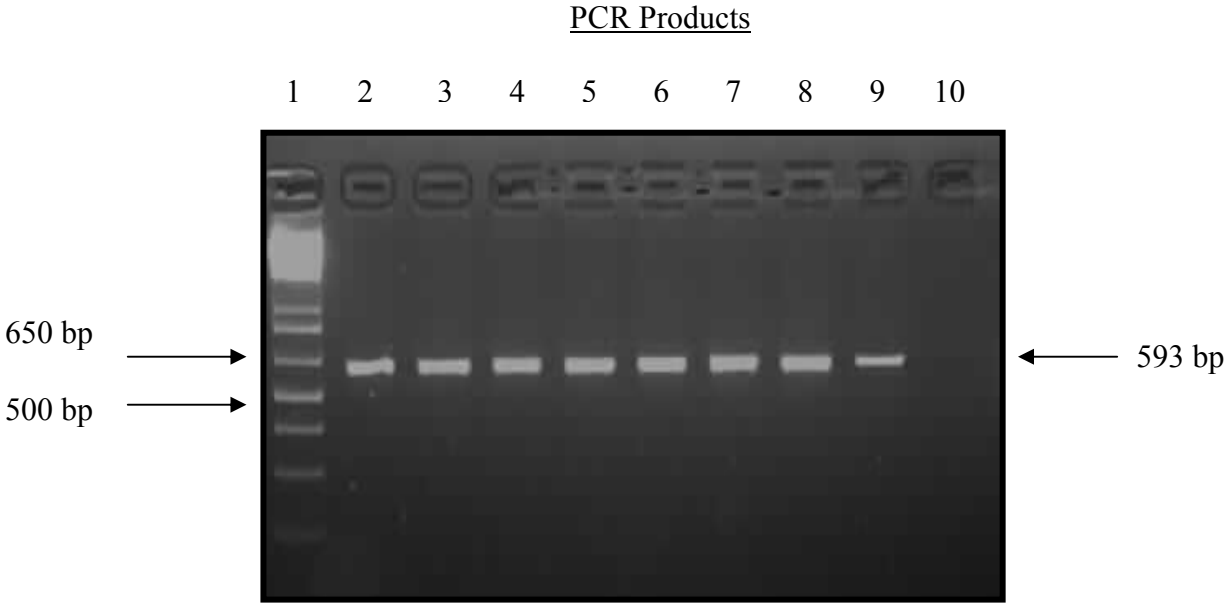


Figure 3.4: PCR products of 593 bp encompassing *ASTN's* rs12118933 SNP run on a 3% agarose gel. **Lane 1:** 1 Kb Plus DNA Ladder. **Lanes 2-9:** PCR products for case and control samples. **Lane 10:** Negative control (no DNA).

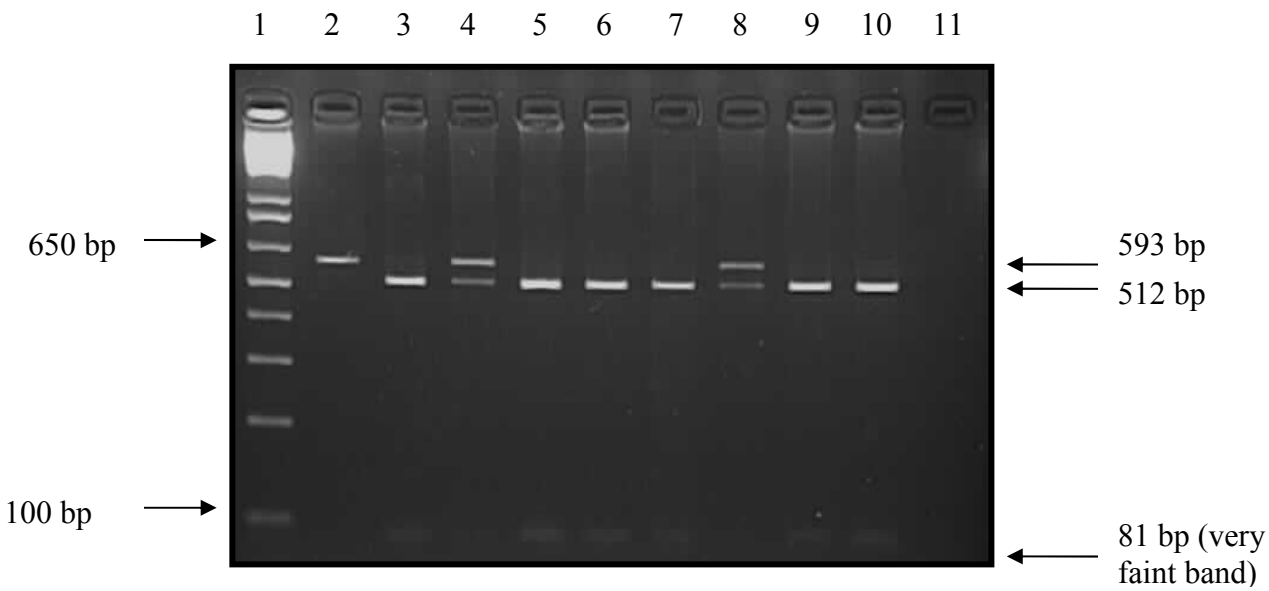


Figure 3.5: Digestion of *ASTN's* rs12118933 PCR products with *Msp*I, samples viewed on a 4% agarose gel. **Lane 1:** 1 Kb Plus DNA Ladder. **Lane 2:** Uncut PCR product (593 bp). **Lane 3:** Positive G/G homozygote. **Lane 4:** Positive G/C heterozygote. **Lanes 5-10:** Control sample digests. **Lane 11:** Negative control (no DNA). No C/C homozygotes were identified in either cases or controls. "Positive" genotypes were confirmed by sequencing.

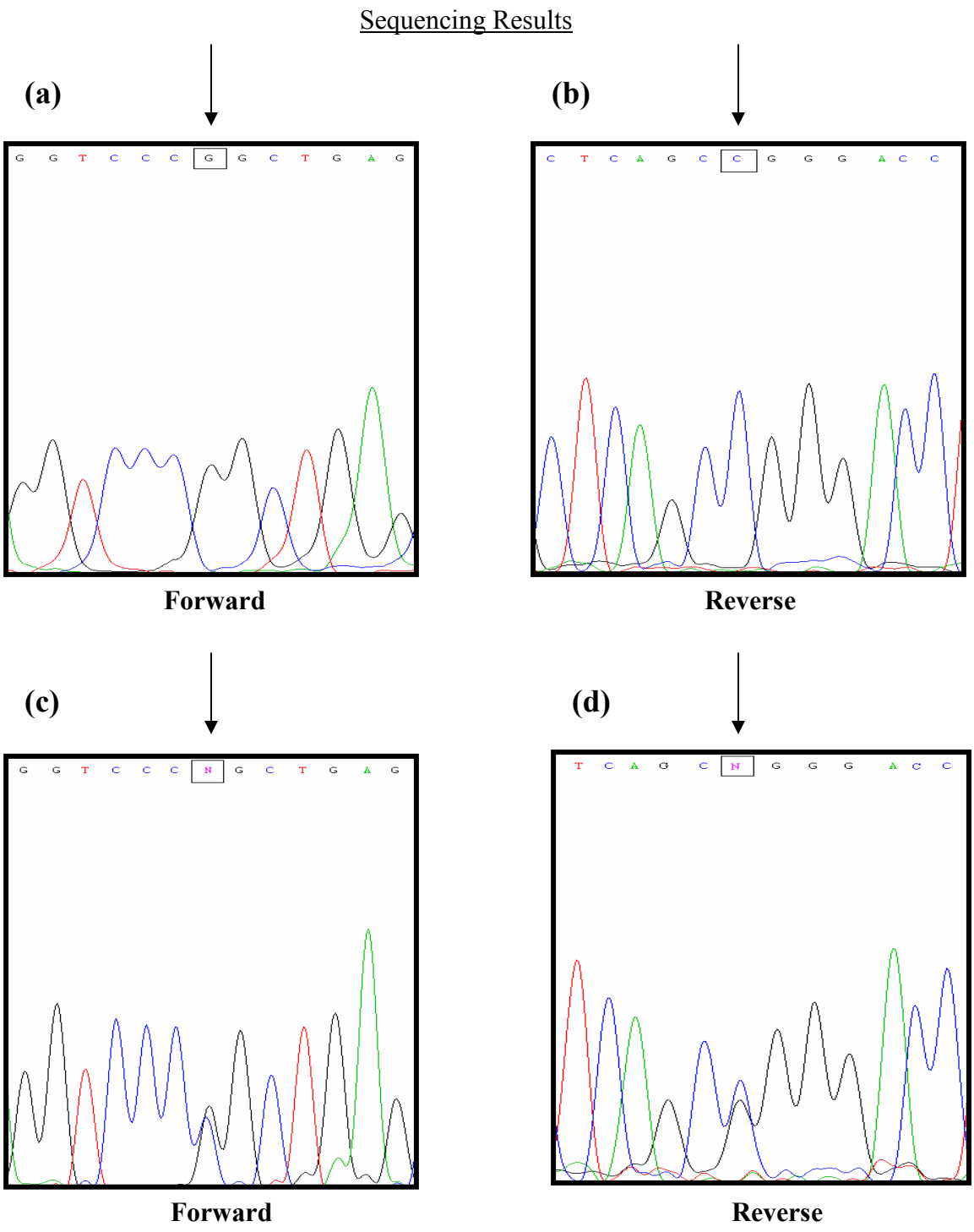


Figure 3.6: Electropherograms confirming genotyping results for rs12118933. Arrows indicate SNP position. **G/G homozygote:** Forward **(a)** and reverse **(b)** strands of sample NPC001 indicating a "G" and reverse compliment "C" allele respectively at the SNP position. **G/C heterozygote:** Forward **(c)** and reverse **(d)** strands of sample NPF017. At position "N", the forward strand indicates a G/C heterozygote (black and blue double peak) and the reverse strand confirms this.

3.1.3 rs2228956

Rs2228956 involves a T to C base change in the *ASTN* gene. Figure 3.7 below illustrates the PCR products of 381 bp incorporating the SNP. Once the size of the PCR products had been confirmed they were digested with *Hha*I to generate different band patterns depicting the various genotypes (figure 3.8). Figure 3.9 illustrates the sequencing results produced of the various genotypes.

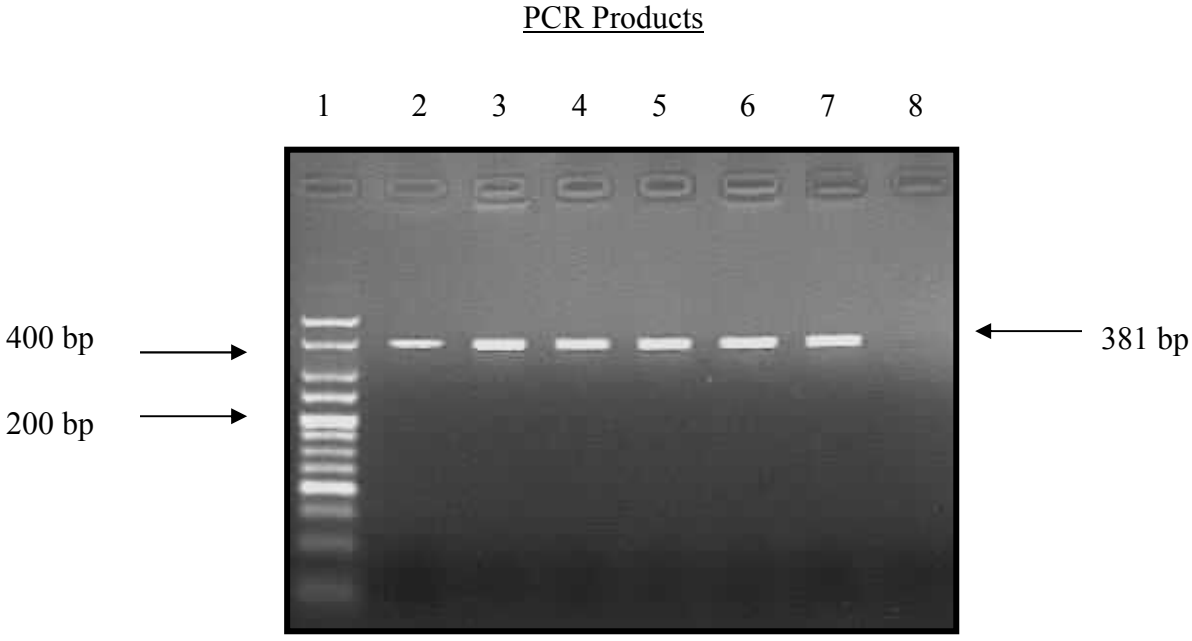


Figure 3.7: 381 bp PCR products encompassing *ASTN's* rs2228956 SNP run on a 3% agarose gel. **Lane 1:** HyperLadder V. **Lanes 2- 7:** PCR products for case and control samples. **Lane 8:** Negative control (no DNA).

*Hha*I Digestion

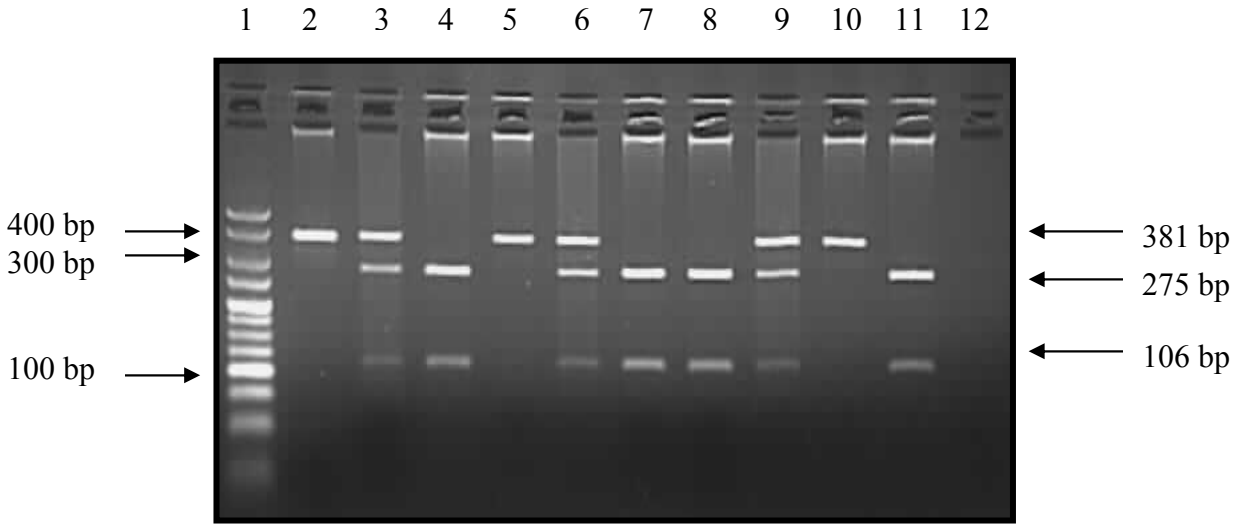
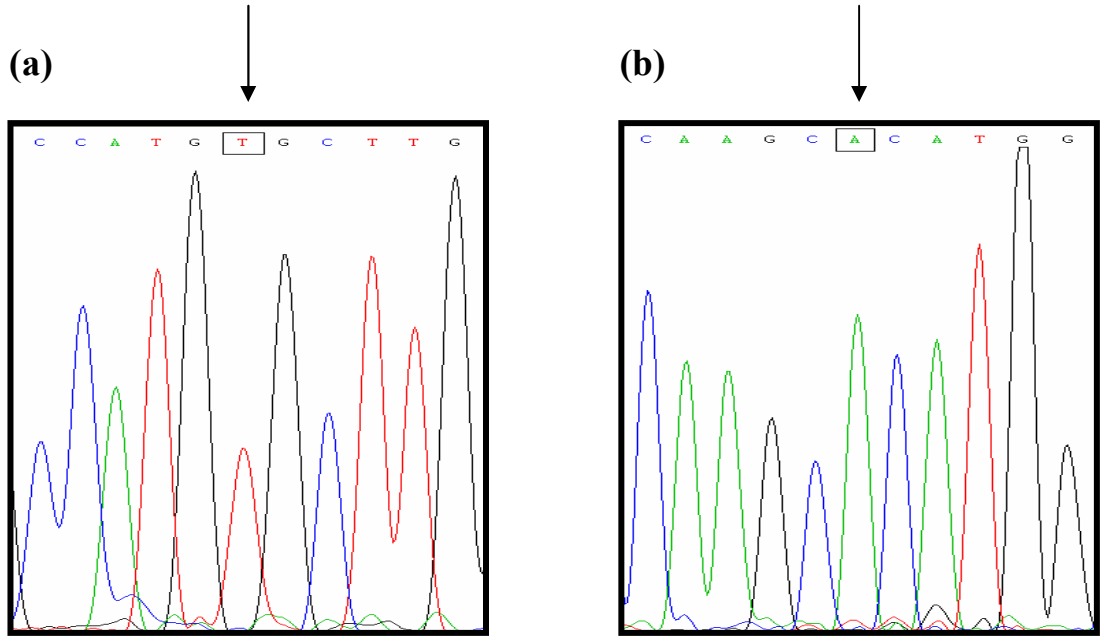


Figure 3.8: Digestion of *ASTN*'s rs2228956 PCR products with *Hha*I viewed on a 4% agarose gel. **Lane 1:** HyperLadder V. **Lane 2:** Uncut PCR product (381 bp). **Lane 3:** Positive T/C heterozygote. **Lane 4:** Positive C/C homozygote. **Lane 5:** Positive T/T homozygote. **Lanes 6-11:** Control sample digests. **Lane 12:** Negative control (no DNA). "Positive" genotypes were confirmed by sequencing.

Sequencing Results



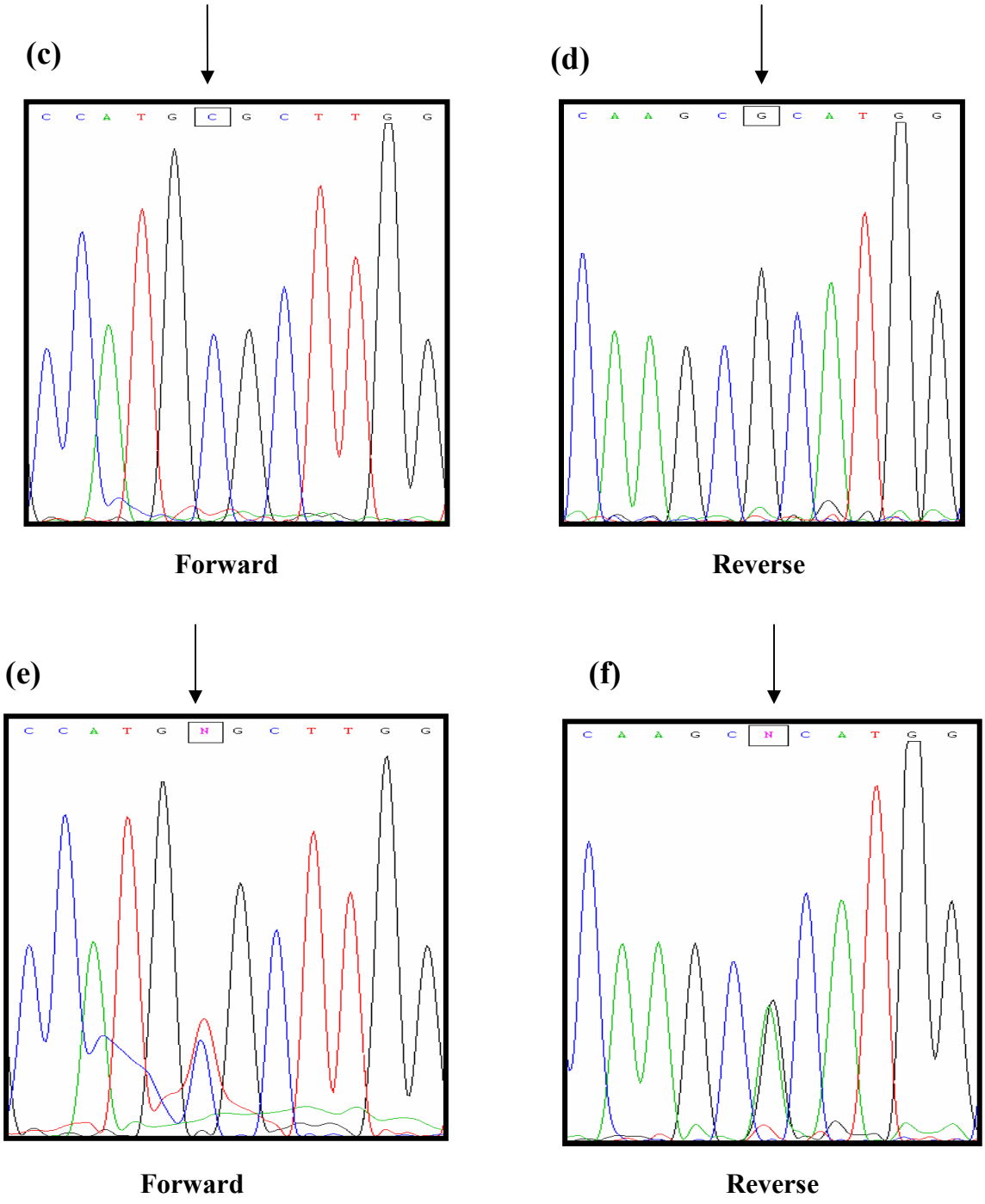


Figure 3.9: Sequencing results confirming the rs2228956 genotypes. The arrows indicate the SNP position. **T/T homozygote:** Forward **(a)** and reverse **(b)** strands of control sample NPC029. The forward strand indicates a T/T homozygote and the reverse strand confirms this by showing a single "A" allele peak (reverse compliment) at the same point. **C/C Heterozygote:** Forward **(c)** and reverse **(d)** strands of control sample NPC022. The forward strand shows a clear "C" allele and the reverse strand indicates the compliment "G" allele. **T/C Heterozygote:** Forward **(e)** and reverse **(f)** strands of control sample NPC059. The forward strand indicates a T/C heterozygote and the reverse strand confirms this by showing a green ("A") and black ("G") double peak at the same point.

3.1.4 rs11587640

ASTN's rs11587640 consists of a G to T base change. Figure 3.10 shows the generated PCR products (172 bp) incorporating the SNP. Figure 3.11 is a gel picture that visualises the unfavourable results obtained after *Sbf*I digestion of the PCR products. Due to the digestion problems this SNP was abandoned.

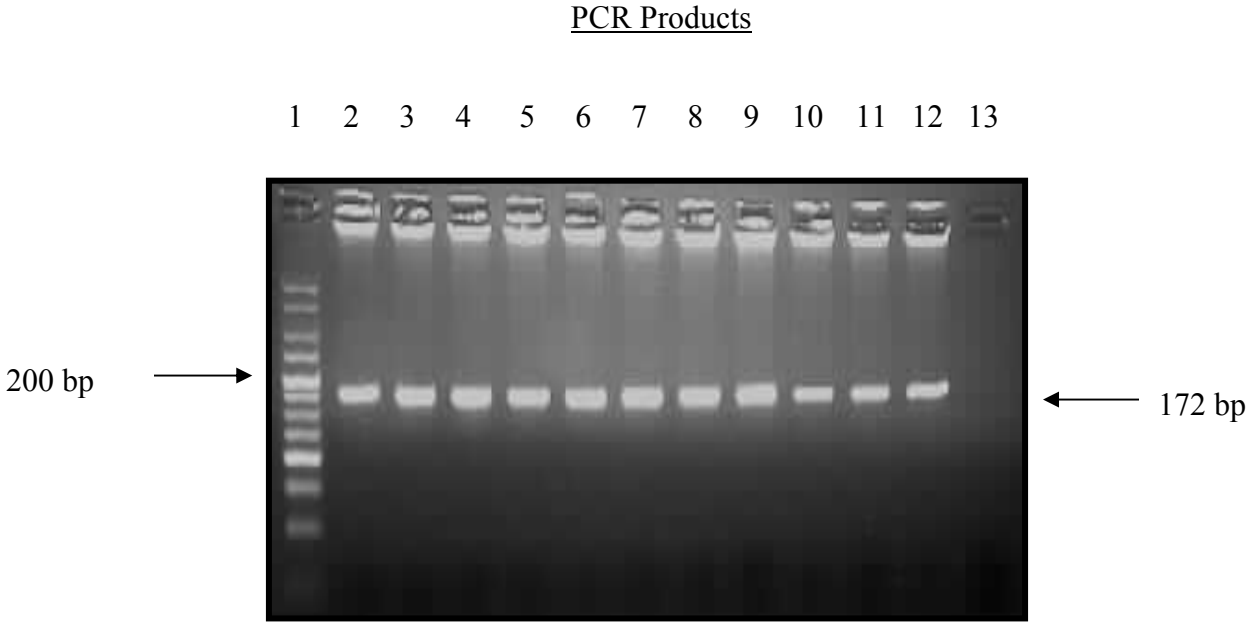


Figure 3.10: PCR products of 172 bp in size encompassing *ASTN's* rs11587640 SNP run on a 3% agarose gel. **Lane 1:** HyperLadder V. **Lanes 2-12:** PCR products for case and control samples. **Lane 13:** Negative control (no DNA).

*Sbf*I Digestion

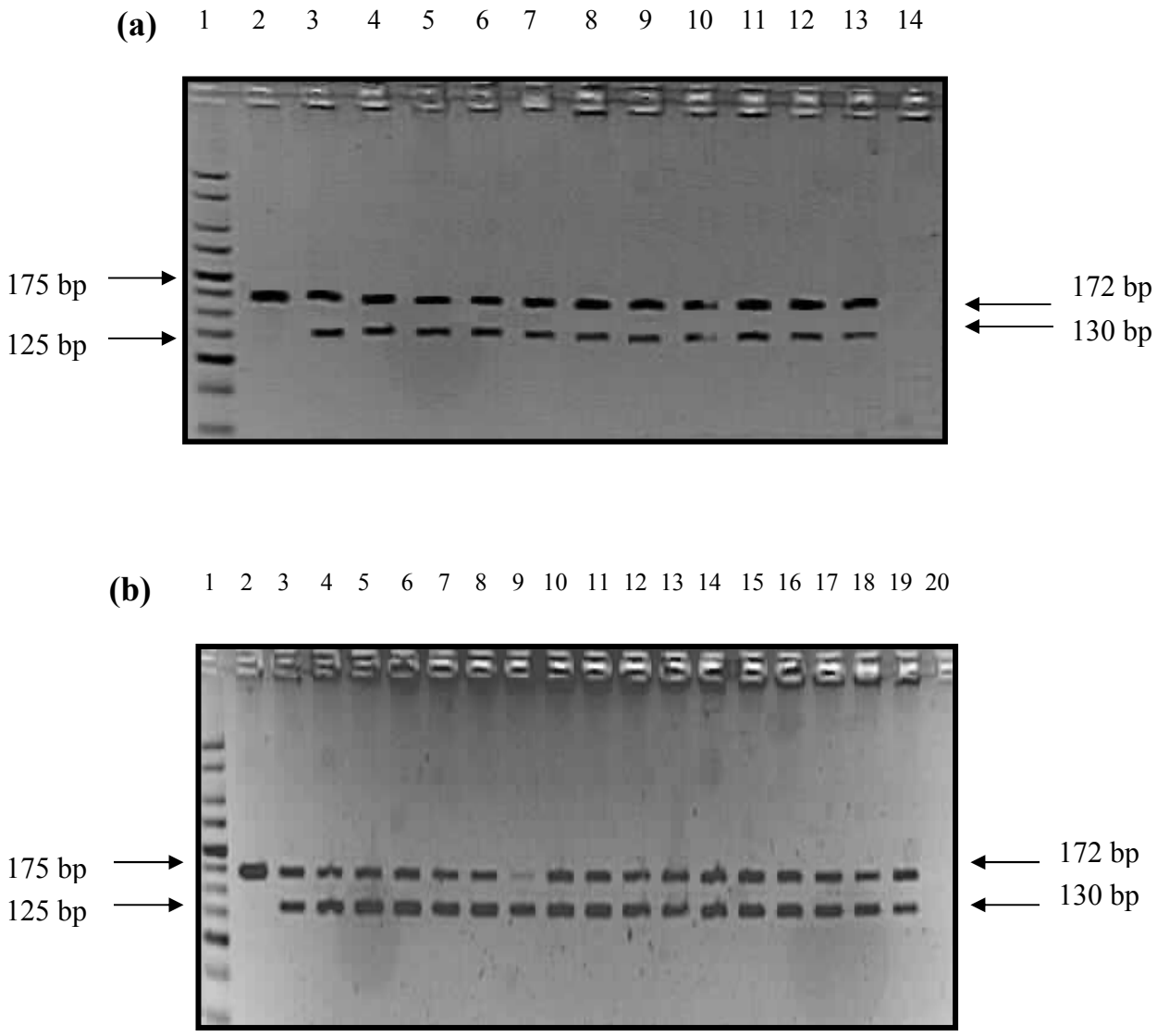


Figure 3.11: 4% agarose gels showing digestion of *ASTN's* rs11587640 PCR products with *Sbf*I. Gel **(a)** shows digestion of case samples whereas gel **(b)** shows digestion of control samples. For both gels, **lane 1:** HyperLadder V. **Lane 2:** Uncut PCR product of 172 bp. The last lane on each gel contains a negative control (no DNA). The majority of samples abnormally appear to be G/T heterozygotes with the sample in lane 9(b) representing a partial digest. Due to these problems and failure to optimise the detection method this SNP was abandoned.

The PCR reaction for rs11587640 was successful in amplifying the desired 172 bp product from each sample. However, identification of the different alleles by *Sbf*I digestion appeared to be problematic in that ambiguous results were obtained. *Sbf*I manufactured by

New England Biolabs was used. This enzyme failed to produce feasible results even after 16 hour digestions as recommended by the manufacturer. On a 4% agarose gel, *Sbf*I appeared to have cut, however, G/T heterozygotes were seen for all cases and the majority of controls. A sample representing possible incomplete digestion was also observed (lane 9 of figure 3.11(b)). This is unlikely to be a true reflection of the 157 individuals genotyped (112 controls and 45 cases). In a population, one would expect to locate some definite homozygotes, G/G or T/T or both. When contacted, the manufacturers stated that *Sbf*I is strongly inhibited by salts such as residual salts from commercial DNA purification kits.

3.2 GENOTYPING OF *REELIN* SNPs

Much like the protocols and processes performed involving the *ASTN* SNPs, the processes performed on the *RELN* SNPs involved PCR followed by restriction enzyme digestion and confirmatory sequencing. One another technique was employed for one of the *RELN* SNPs and that was the specialised PCR reaction known as ARMS. The following gel pictures and electropherograms represent the findings for SNPs rs607755, rs3025963, rs3025966 and rs362691.

3.2.1 rs607755

RELN's rs607755 involves the alleles T or C. Figure 3.12 illustrates the PCR products within which the SNP is present. Figure 3.13 is a gel picture depicting the genotyping results obtained after digestion of the PCR products with *Pvu*II. Figure 3.14 is a series of electropherograms that were generated to confirm the genotyping findings obtained by enzyme digestion.

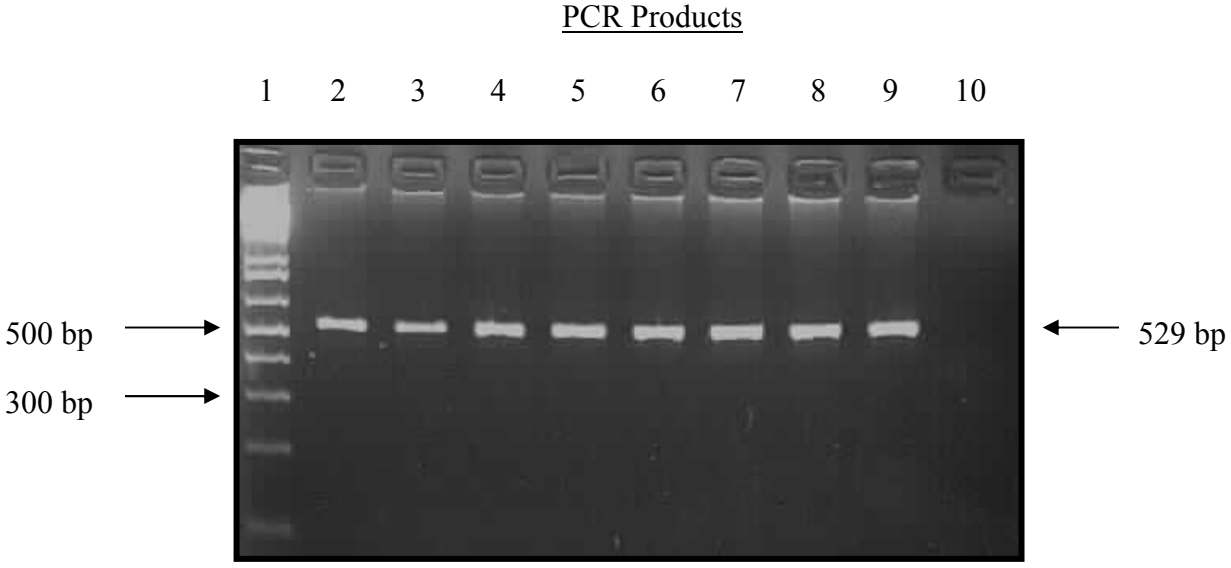


Figure 3.12: 529 bp PCR products encompassing *RELN's* rs607755 SNP run on a 3% agarose gel. **Lane 1:** 1 Kb Plus DNA Ladder. **Lanes 2- 9:** PCR products for case and control samples. **Lane 10:** Negative control (no DNA).

*Pvu*II Digestion

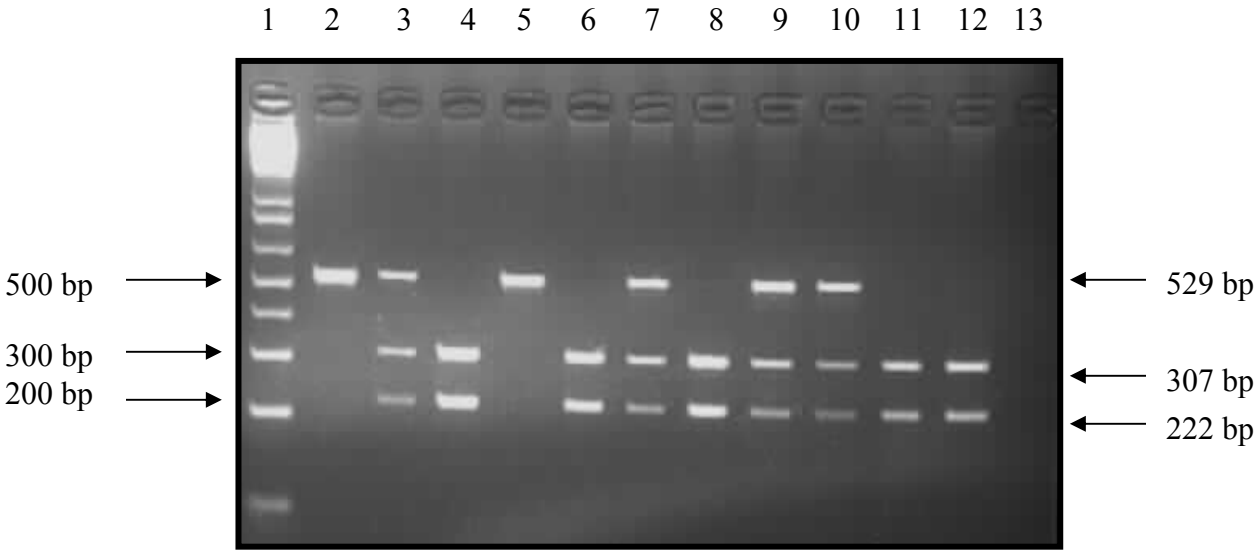
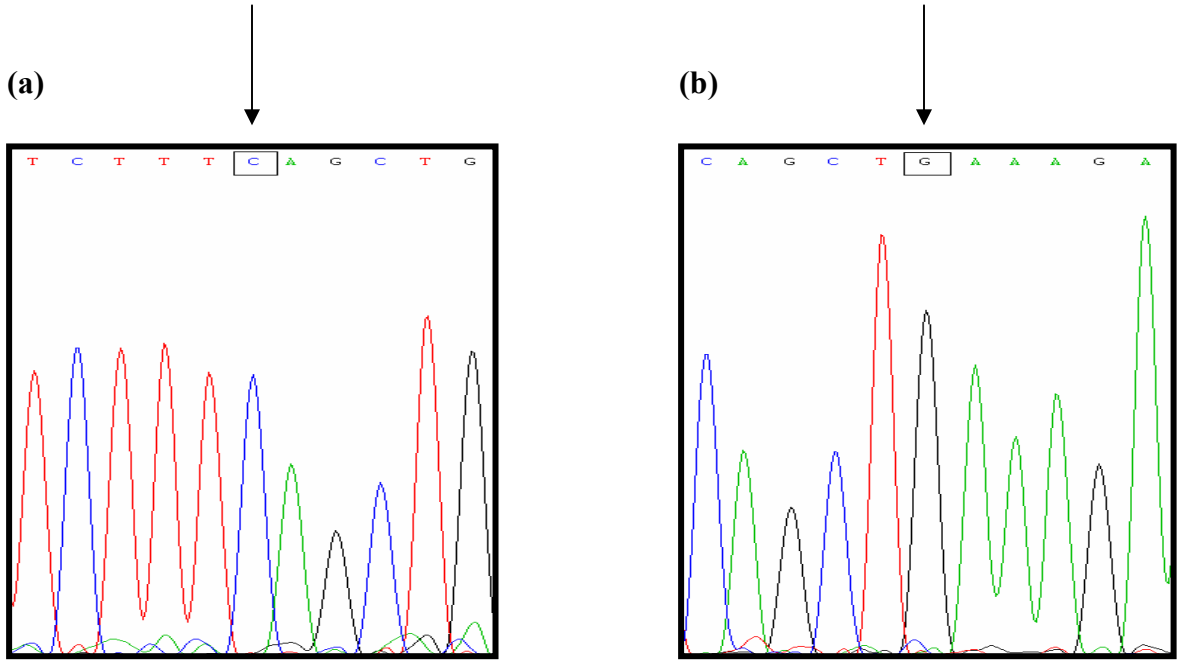


Figure 3.13: Digestion of *RELN*'s rs607755 PCR products with *Pvu*II, products viewed on a 4% agarose gel. **Lane 1:** I Kb Plus DNA Ladder, **Lane 2:** Uncut PCR product (529 bp), **Lane 3:** Positive T/C heterozygote. **Lane 4:** Positive C/C homozygote. **Lane 5:** Positive T/T homozygote. **Lanes 6-12:** Control sample digests. **Lane 13:** Negative control (no DNA). "Positive" genotypes were confirmed by sequencing.

Sequencing Results



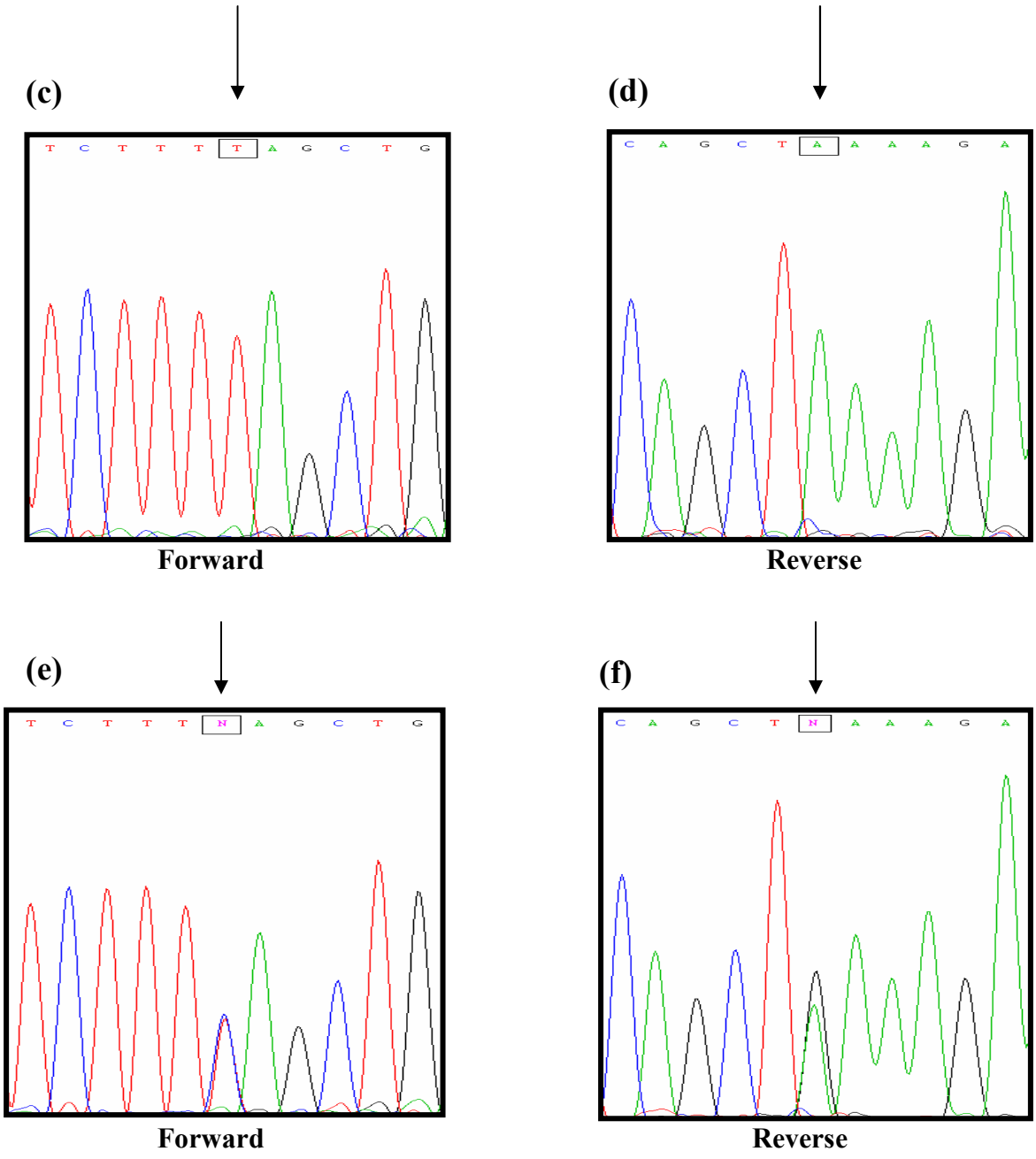


Figure 3.14: Sequencing results confirming the three genotypes obtained for rs607755. **C/C Homozygote:** Forward **(a)** and reverse **(b)** strands of NPC038. A single blue peak indicative of a "C" allele is evident on the forward strand and a single black peak indicative of a "G" allele is seen on the reverse strand (arrows indicate SNP position). **T/T Homozygote:** Forward **(c)** and reverse **(d)** strands of sample NPC051. The Forward strand indicates single "T" allele at the SNP position and the reverse strand confirms this by showing a single reverse compliment "A" allele at the same position. **C/T Heterozygote:** Forward **(e)** and reverse **(f)** strands of NPC104. A red and blue double peak, indicative of both "C" and "T" alleles can be seen on the forward strand and a complimentary black and green double peak, illustrating the presence of both "G" and "A" alleles is present on the reverse strand.

3.2.2 rs3025963

Rs3025963 involves an A to T base change. Figure 3.15 below is a gel picture that illustrates the 285 bp PCR products incorporating rs3025963. Figure 3.16 depicts the results obtained after digestion of the PCR products with *Mva*12691. SNP rs3025963 proved uninformative, no variation was seen in either cases or controls after 60 controls and 25 cases were typed. No further analysis could be carried out on this marker.

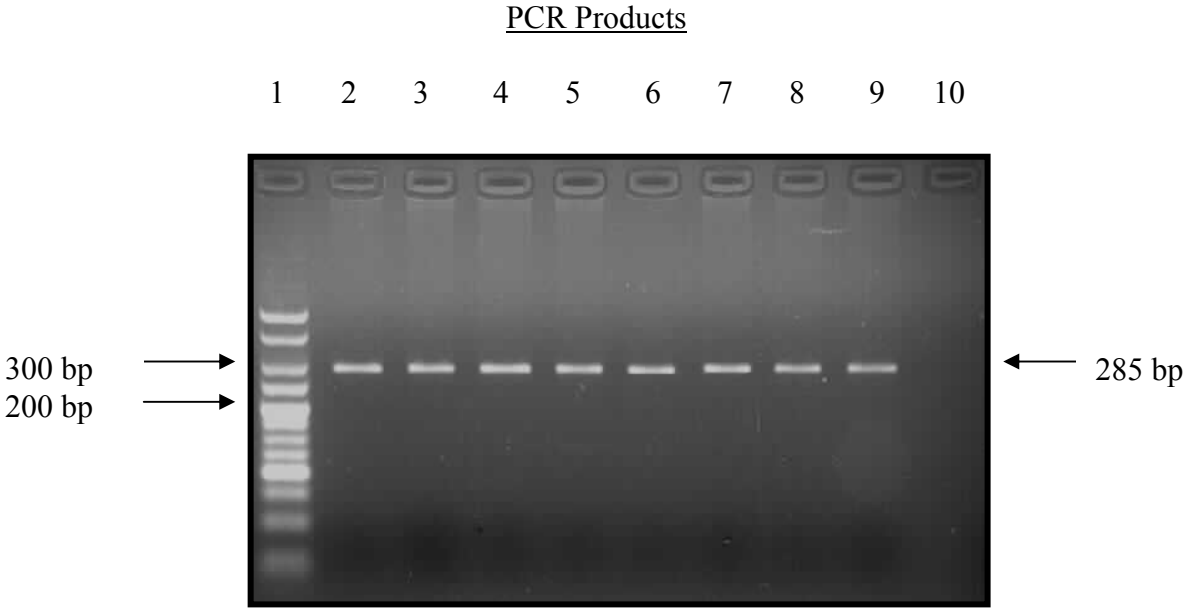


Figure 3.15: PCR products of 285 bp in size encompassing *RELN's* rs3025963 SNP run on a 3% agarose gel. **Lane 1:** HyperLadder V. **Lanes 2-9:** PCR products for case and control samples. **Lane 10:** Negative control (no DNA).

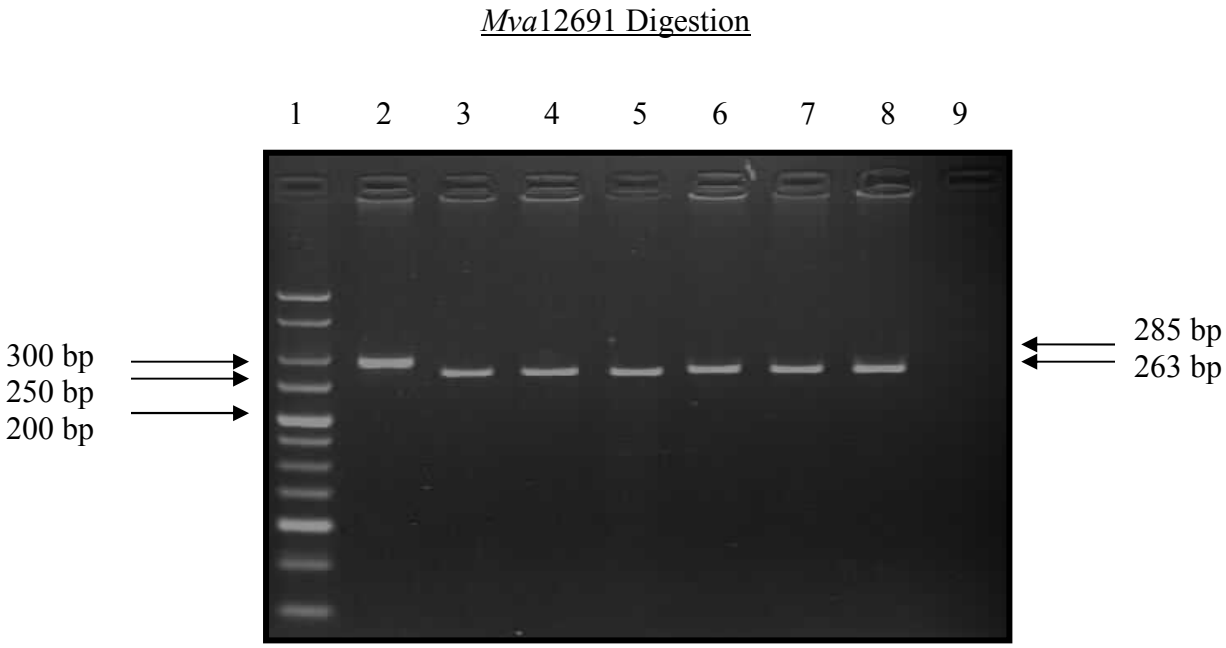


Figure 3.16: Digestion of *RELN's* rs3025963 PCR products with *Mva*12691, products viewed on a 4% agarose gel. **Lane 1:** HyperLadder V, **Lane 2:** Uncut PCR product (285 bp), **Lanes 3-8:** Control sample digests of 263 bp in size (indicating the presence of the "A" allele). The 22 bp fragment would have run off the gel. **Lane 9:** Negative control (no DNA).

3.2.3 rs3025966

RELN's rs3025966 involves a T or G allele. Figure 3.17 illustrates a gel picture of the PCR products that incorporate the SNP. Figure 3.18 illustrates the results obtained after *Mva*12691 digestion of the 226 bp PCR products. On digestion of *RELN* marker rs3025966 the same problem observed in rs3025963 occurred. Rs3025966 proved to be uninformative. Due to no allelic variation being seen in either cases or controls no further analysis was carried out on this SNP.

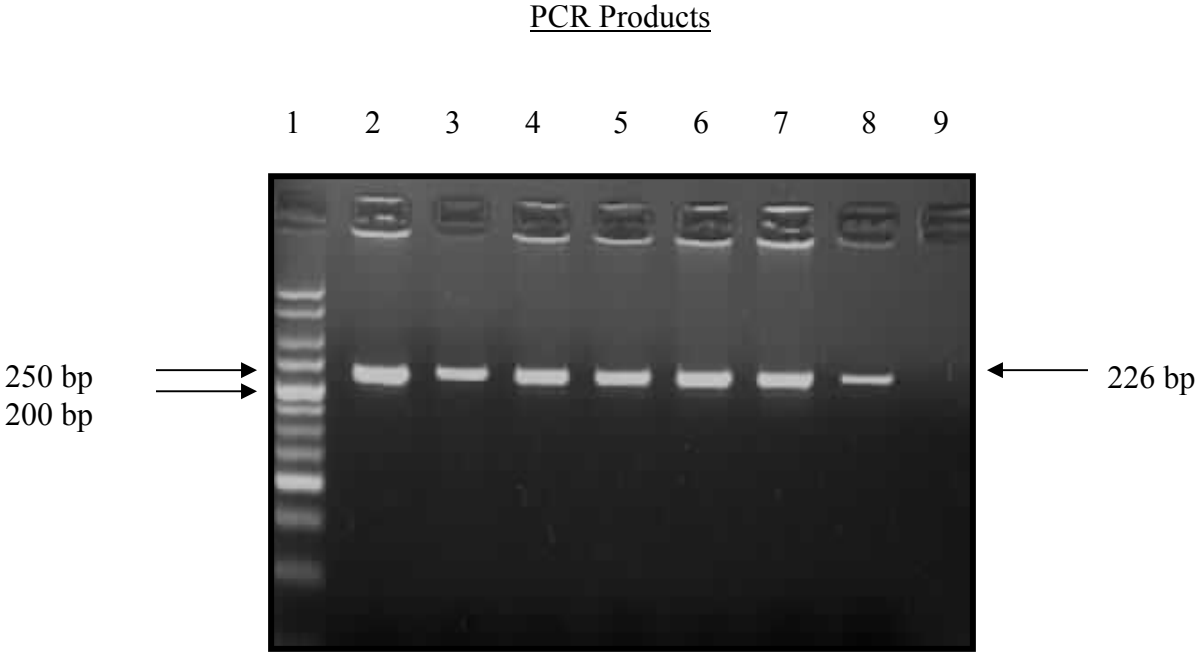


Figure 3.17: 226 bp PCR products encompassing *RELN's* rs3025966 SNP run on a 3% agarose gel. **Lane 1:** HyperLadder V. **Lanes 2-8:** PCR products for case and control DNA samples. **Lane 9:** Negative control (no DNA).

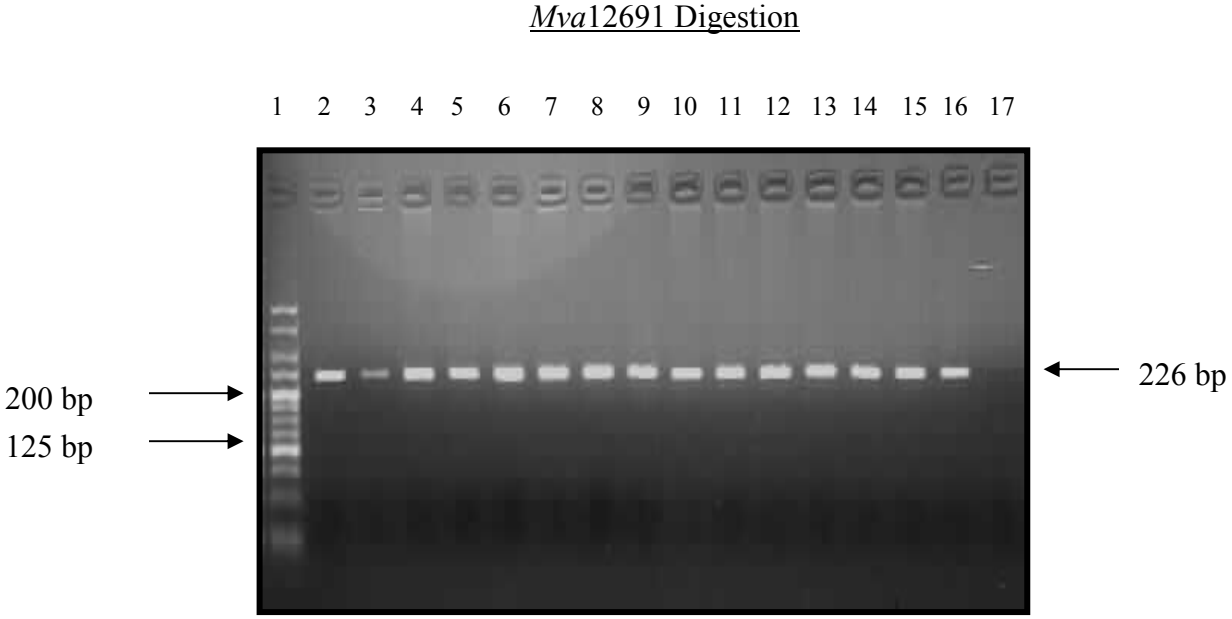


Figure 3.18: Digestion of *RELN's* rs3025966 PCR products with *Mva*12691, products viewed on a 4% agarose gel. **Lane 1:** HyperLadder V. **Lane 2:** Uncut PCR product (226 bp). **Lanes 3-16:** Case samples after digestion, note cutting of the PCR products has not occurred indicating the absence of the "G" allele at the SNP position. **Lane 17:** Negative control (no DNA).

Digestion of rs3025966 required the same *Mva*12691 enzyme used in rs3025963 digestion. Based on the results seen in figure 3.16 it was evident that the enzyme was indeed cutting as a change in size from 285 bp to 263 bp was observed. In figure 3.18 no enzymatic cleaving occurred indicating the presence of the "T" allele in all samples.

3.2.4 rs362691

Rs362691 is not an RFLP and so an ARMS system was employed. Figure 3.19 illustrates the results obtained from the ARMS reaction. In order to confirm the results obtained by ARMS, sequencing was conducted. However, prior to sequencing, new PCR products amplified by only one set of primers (no control primers) had to be generated. Figure 3.20 illustrates the PCR products used for sequencing and figure 3.21 depicts the sequencing results obtained.

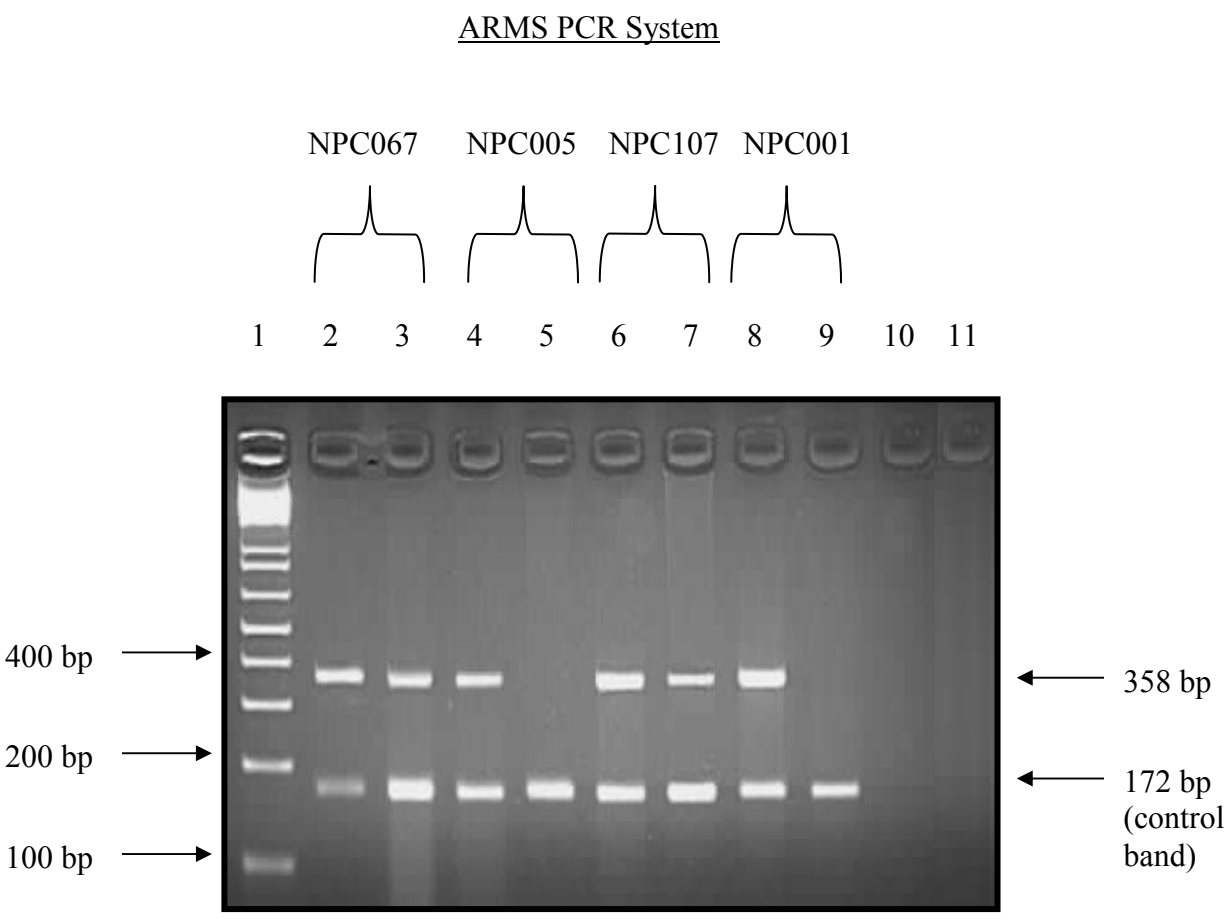


Figure 3.19: PCR products from *RELN's* rs362691 ARMS system shown on a 3% agarose gel. **Lane 1:** 1 Kb Plus DNA Ladder. **Lanes 2 and 3:** Control sample NPC067, C/G heterozygote (lane 2= products from "C" allele reaction & lane 3= products from "G" allele reaction). **Lanes 4 and 5:** Control sample NPC005, C/C homozygote (lane 4= products from "C" allele reaction & lane 5= products from "G" allele reaction). **Lanes 6 and 7:** Control sample NPC107, a C/G heterozygote (lane 6= products "C" allele reaction & lane 7= products from "G" allele reaction). **Lanes 8 and 9:** Control sample NPC 001, a C/C homozygote (lane 8= products from "C" allele reaction & lane 9= products from "G" allele reaction). **Lane 10:** Negative control for the "C" allele Master Mix (no DNA). **Lane 11:** Negative control (no DNA) for the "G" allele Master mix.

Only C/C homozygotes and C/G heterozygotes were observed in the cases and controls for rs362691. In order to confirm the results obtained by the ARMS PCR, sequencing of the two genotypes was carried out.

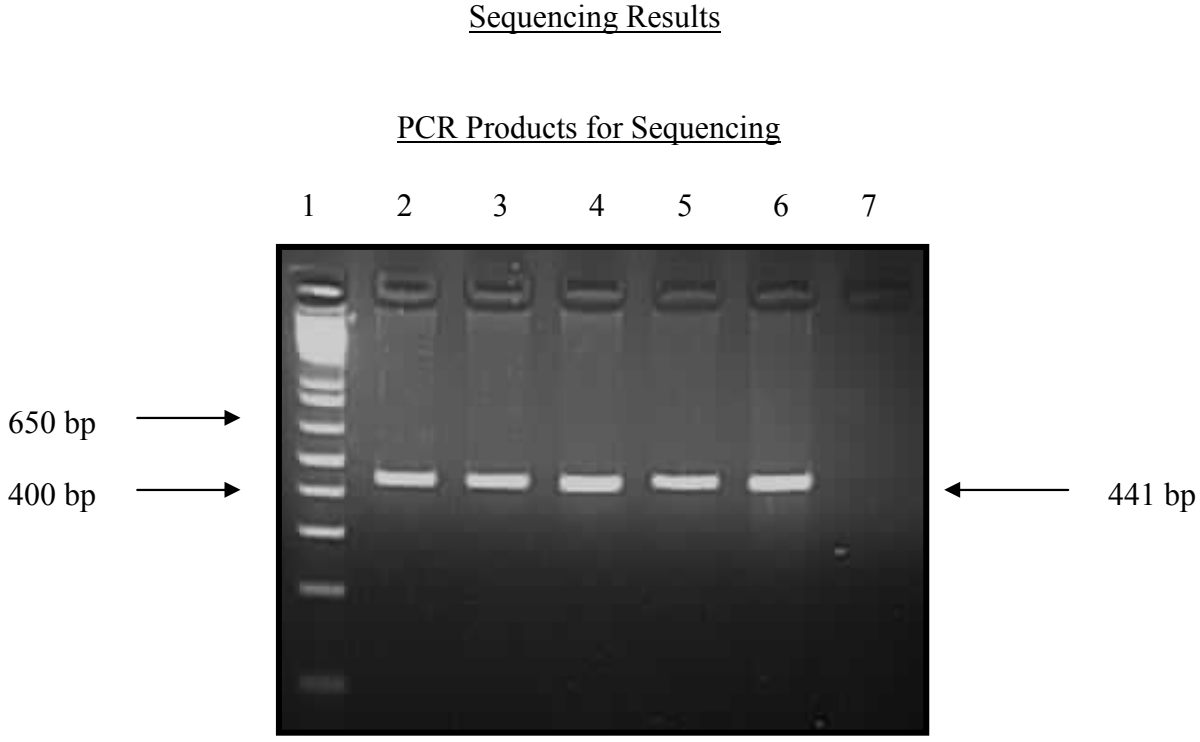


Figure 3.20: PCR products of 441 bp generated for the purpose of rs362691 sequencing, run on a 3% agarose gel

This SNP proved to be problematic during sequencing. Despite electropherogram data being generated for the samples sequenced, the sequencing results were obscured by what appeared to be a contaminating sequence. On the 3% agarose gels, only one band was visible for PCR products (figure 3.20) it therefore did not appear as though any other region of the genome was being amplified. To rule out experimental error with the cycle sequencing and the product purification steps, the SNP was sequenced on three different occasions. For each sequencing run new PCR products were amplified. All three attempts produced electropherograms that had the same in appearance. Figure 3.21 illustrates the sequencing results obtained.

Sequencing Results

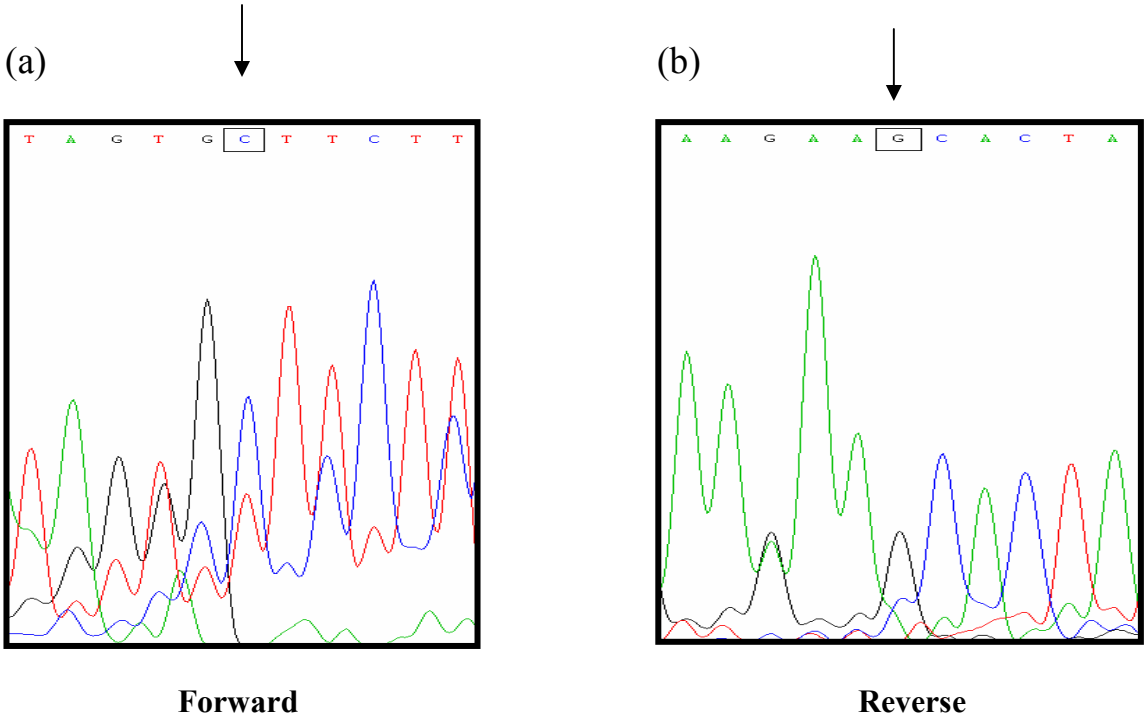


Figure 3.21: Electropherograms of the forward and reverse strands of control sample NPC005, C/C homozygote for rs362691 (see figure 3.19). On the forward strand **(a)**, the red peak below the blue "C" allele peak at the site of the SNP (indicated by the arrow) appears to be caused by the neighbouring "T" allele. The reverse strand **(b)** indicates an expected black "G" allele peak at the SNP position. The blue peak beneath the "G" allele also appears to be caused by the neighbouring "C" allele.

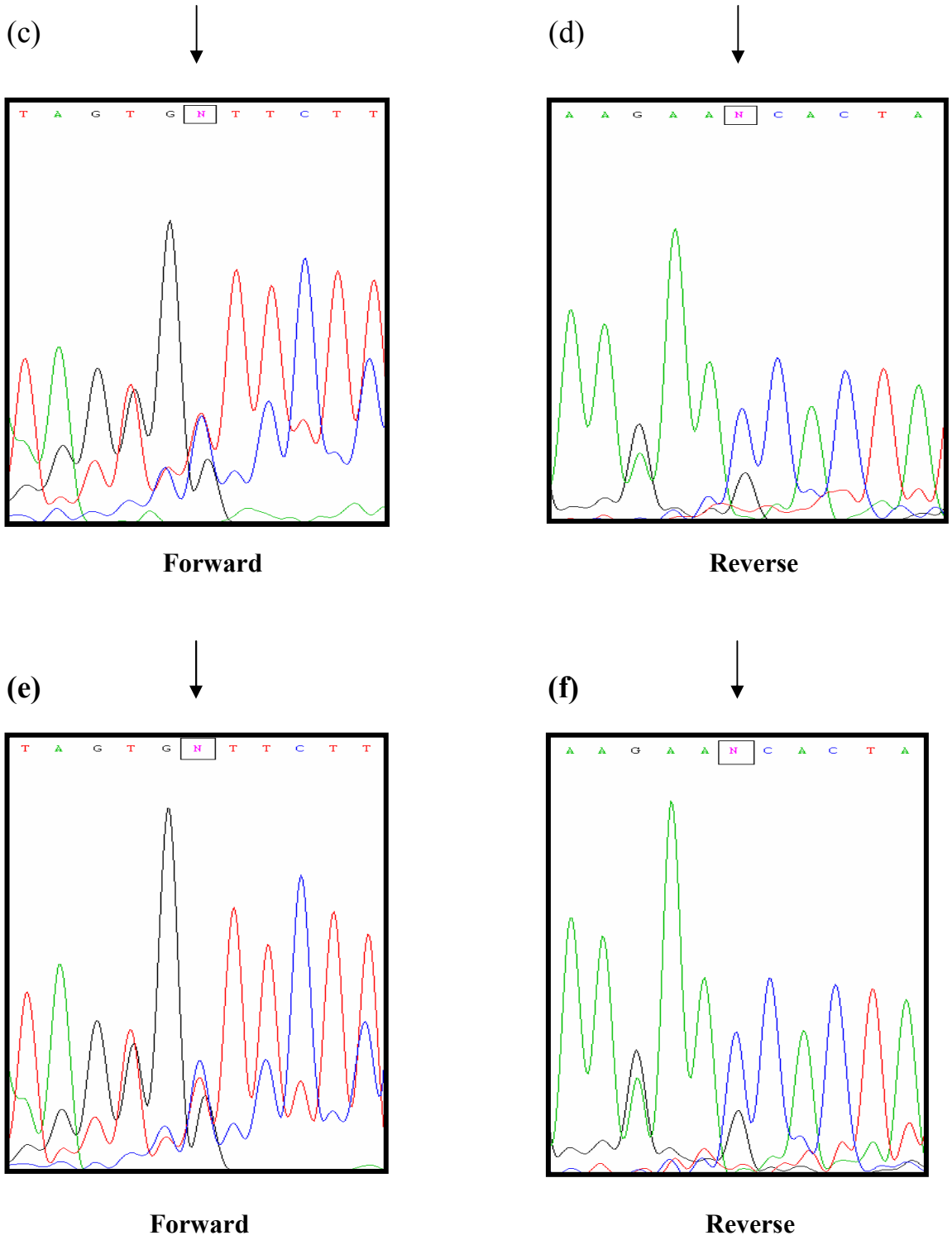


Figure 3.22: Electropherograms of rs362691 C/G heterozygotes. The forward **(c)** and reverse **(d)** strands of sample NPC067 and the forward **(e)** and reverse **(f)** strands of NPC107. At the position of the SNP indicated by the arrow on the forward strands **(c & e)**, a blue peak is visible with a black peak underneath it. The red peak overlapping the blue peak most probably results from the "T" allele to the right of the SNP. The reverse strands **(d & f)** show similar results with a blue peak as well as a black peak being present at the SNP site.

Based on the sequencing data obtained from rs362691, it is evident that the quality of the electropherograms is not ideal and it appears as though the signals from the different bases are being carried over onto the neighbouring bases. As mentioned previously, sequencing was carried out on three separate occasions using fresh PCR products every time and on all three occasions the same peak pattern was seen in each electropherogram. It is possible that the primer could have been defective in that there were two populations of oligonucleotides in that primer mix but one population could have been one base shorter than the other.

If one compares a sequenced C/C homozygote (figure 3.21) to a sequenced C/G heterozygote (figure 3.22) it is evident that the underlying black "G" allele peak present in the heterozygote samples is absent in the C/C homozygote electropherograms. This therefore confirms that the samples NPC067 and NPC107 are indeed C/G heterozygotes as shown by the ARMS PCR system.

3.3 STATISTICAL ANALYSES

3.3.1 Genotype and Allele Frequencies

Table 3.1: Genotype and allele frequencies for *ASTN* markers: cases and controls

		CASES		CONTROLS	
SNP	**Genotype/allele**	***f***	***n***	***f***	***n***
rs6413830	G/G	0.644	29	0.589	66
	G/A	0.333	15	0.357	40
	A/A	0.022	1	0.053	6
	Total:		**45**		**112**
	G	0.811	73	0.768	172
	A	0.189	17	0.232	52
	Total:		**90**		**224**
rs12118933	G/G	0.956	43	0.973	109
	G/C	0.044	2	0.027	3
	C/C	0.000	0	0.000	0
	Total:		**45**		**112**
	G	0.978	88	0.987	221
	C	0.022	2	0.013	3
	Total:		**90**		**224**
rs2228956	C/C	0.822	37	0.830	93
	T/C	0.178	8	0.152	17
	T/T	0.000	0	0.018	2
	Total:		**45**		**112**
	C	0.911	82	0.906	203
	T	0.089	8	0.094	21
	Total:		**90**		**224**

n, number of observations for genotypes or alleles; *f*, frequency

On superficial scrutiny the results depicted in the table above appear to show that there is no striking difference between the case and control groups in the frequencies of each *ASTN* SNP. Allele and genotype frequencies for both groups are very similar.

Table 3.2: Genotype and allele frequencies for *RELN* markers: cases and controls

		CASES		CONTROLS	
SNP	**Genotype/allele**	***f***	***n***	***f***	***n***
rs362691	C/C	0.978	44	0.902	101
	C/G	0.022	1	0.098	11
	G/G	0.000	0	0.000	0
	Total:		**45**		**112**
	C	0.989	89	0.951	213
	G	0.011	1	0.049	11
	Total:		**90**		**224**
rs607755	C/C	0.489	22	0.313	35
	T/C	0.333	15	0.545	61
	T/T	0.178	8	0.143	16
	Total:		**45**		**112**
	C	0.344	59	0.585	131
	T	0.656	31	0.415	93
	Total:		**90**		**224**

n, number of observations for genotypes or alleles; *f*, frequency

On initial visual scrutiny table 3.2 shows similar frequencies for cases and controls regarding rs362691. There is a clear difference however, between the cases and the controls for rs607755; the T/C genotype accounts for the majority of the 112 controls (n = 61) followed by the C/C genotype (n = 35) and lastly the T/T genotype (n = 16). However, in the 45 cases the most common genotype was C/C (n = 22) followed by the T/C genotype (n = 15) and finally the T/T genotype (n = 8). Association tests were performed to confirm the above observations.

3.3.2 Hardy-Weinberg Equilibrium

The Hardy-Weinberg test was done to see whether the study groups conformed to HWE at the various loci and to detect if any sampling or genotyping error were present.

Table 3.3: *P*-values for the HWE test in cases and controls for *ASTN*

Marker	Cases	Controls
rs6413830	1.000	1.000
rs12118933	1.000	1.000
rs2228956	0.238	1.000

Table 3.4: *P*-values for the HWE test in cases and controls for *RELN*

Marker	Cases	Controls
rs362691	1.000	1.000
rs607755	0.245	0.0975

The *P*-values generated for the HWE tests for the *ASTN* and *RELN* SNPs are all above 0.05. It is therefore unlikely that there are any sampling or genotyping errors, inbreeding or population stratification. These results indicate that the study groups together form a fair representation of the Northern Cape Coloured community with respect to these two genes.

3.3.3 Association Testing

The tables below indicate the results generated from the Fisher's exact test, which looked for association between the polymorphic variants and the susceptibility to FAS.

Table 3.5: *P*-values for association tests for cases versus controls at the allelic and genotypic level for *ASTN*

Marker	Allelic Association	Genotypic Association
rs6413830	0.453	0.747
rs12118933	0.627	0.625
rs2228956	1.000	0.904

The *P*-values obtained for the allelic and genotypic associations presented in table 3.5 are all above 0.05 and indicate that there are no significant associations.

Table 3.6: *P*-values for association tests for cases versus controls at the allelic and genotypic level for *RELN*

Marker	Allelic Association	Genotypic Association
rs362691	0.190	0.181
rs607755	0.254	0.049

Table 3.6 lists the *P*-values obtained for the allele and genotype tests performed on *RELN's* rs362691 and rs607755. Rs362691 shows no association with the disease at either the allelic or genotypic level, rs607755 on the other hand has a result of 0.049 ($P<0.05$) indicating significant genotypic association.

A post-test power calculation was carried out for rs607755's significant result; a post-test power value of 0.58 was obtained. This value suggests a moderate chance (58%) of one observing the significant effect in a new study using the same sample size.

3.3.4 Haplotype Analysis

The haplotype analysis involved analysing whether any of the inferred haplotypes where significantly associated with either the case or control groups. Tables 3.7 and 3.8 list the results.

Table 3.7: Comparison of FAS cases versus controls: Haplotype Scores, *P*-values and Haplotype Frequencies (hf) for *ASTN*

Global *P*-value = 0.9

	rs6413830	rs12118933	rs2228956	Hap. Score	*P*-value	Control hf	Case hf
1	A	G	C	-0.762	0.472	0.210	0.170
2	A	G	T	-0.417	0.770	0.020	0.020
6	G	G	T	-0.182	0.826	0.070	0.060
5	G	G	C	0.709	0.473	0.680	0.730
3	G	C	C	0.000	0.000	0.010	0.010
4	G	C	T	0.000	0.000	0.000	0.010

A negative Hap-score would suggest that a particular inferred haplotype is associated with controls (i.e. is more common amongst controls) whereas a positive Hap-score would indicate the inferred haplotype is associated with cases.

The analysis revealed that the inferred GGC haplotype (harbouring alleles, rs6413830^G, rs12118933^G and rs2228956^C) is the most common haplotype for the *ASTN* gene as it is predicted to be present in 68% of controls and 73% of cases. The global *P*-value generated for *ASTN* is 0.9, indicating that there is no significant difference between the cases and controls regarding the inferred haplotypes of the *ASTN* gene. In addition, each Hap-score has a *P*-value of more than 0.05 and so none of the inferred haplotypes can be said to be associated with either cases or controls.

Table 3.8: Comparison of FAS cases versus controls: Haplotype Scores, *P*-values and Haplotype Frequencies (hf) for *RELN*

Global *P*-value = 0.269

	rs362691	rs607755	Hap. Score	*P*-value	Control hf	Case hf
4	G	T	-1.176	0.284	0.032	0.011
2	C	T	-0.798	0.444	0.383	0.333
1	C	C	1.412	0.144	0.567	0.656
3	G	C	-	-	0.017	0.000

The global *P*-value obtained for *RELN's* inferred haplotype analysis is 0.269 indicating no significant difference between the two study groups. The most common haplotype was CC (alleles rs362691^C and rs607755^C) with 56.7% of controls and 65.6% of cases having this inferred haplotype. None of the inferred haplotypes can be said to be associated with either cases or controls as all *P*-values are above 0.05.

3.3.5 Logistic Regression

The final logistic regression model contained the following variables: the two *RELN* SNPs and affection status (FAS or control). All the *ASTN* SNPs were discarded and there was no difference between sex of the cases and controls.

A significant overall *P*-value of 0.032 was obtained for the model, which indicates that the whole model with all the predictors is significant in distinguishing between the cases and the controls. Table 3.9 gives the estimates of the coefficients in the model, their standard errors and corresponding *P*-values for the logistic regression model of cases versus controls.

Table 3.9: Final logistic regression model for FAS cases versus controls

***P*-value for entire model** = 0.032

Coefficients	Estimate	Standard Error	*P*-value
Intercept	-0.425	0.274	0.121
RELN rs607755 C/C	0.000	-	-
RELN rs607755 C/T	-0.890	0.399	0.026
RELN rs607755 T/T	-0.131	0.520	0.802
RELN rs362691 C/C	0.000	-	-
RELN rs362691 C/G	-1.490	1.071	0.164

Table 3.9 indicates that the C/T and T/T genotypes from rs607755 and the C/G genotype from rs362691 are all associated with being a control individual. From the individual *P*-values for each genotype however, only rs607755's C/T has a *P*-value of less than 0.05.

This suggests that the probability of being a FAS case with the C/T genotype is independently significantly different from the baseline genotype C/C and is more likely to be found amongst control subjects thus conferring a protective role against FAS. This supports the genotypic association seen between this marker and the disease in table 3.6.

3.3.6 Linkage Disequilibrium

3.3.6.1 Astrotactin Markers

Out of the four *ASTN* SNPs, genotypic information was gained for only three. The pairwise linkage disequilibrium heat map below indicates the degree of linkage between the SNPs based on the results generated by the control group. *ASTN1a* = rs6413830, *ASTN1b* = rs12118933 and *ASTN1c* = 2228956. The distance between *ASNT1a* and *ASTN1c* is 31.3kb. The colour key below indicates D' values. High LD is seen between *ASTN1a* and *ASTN1b* (D'= 0.978). *ASTN1b* and *ASTN1c* are also in very high LD (D'= 0.946) whereas *ASTN1a* and *ASTN1c* are in low LD (D'= 0.018) with each other.

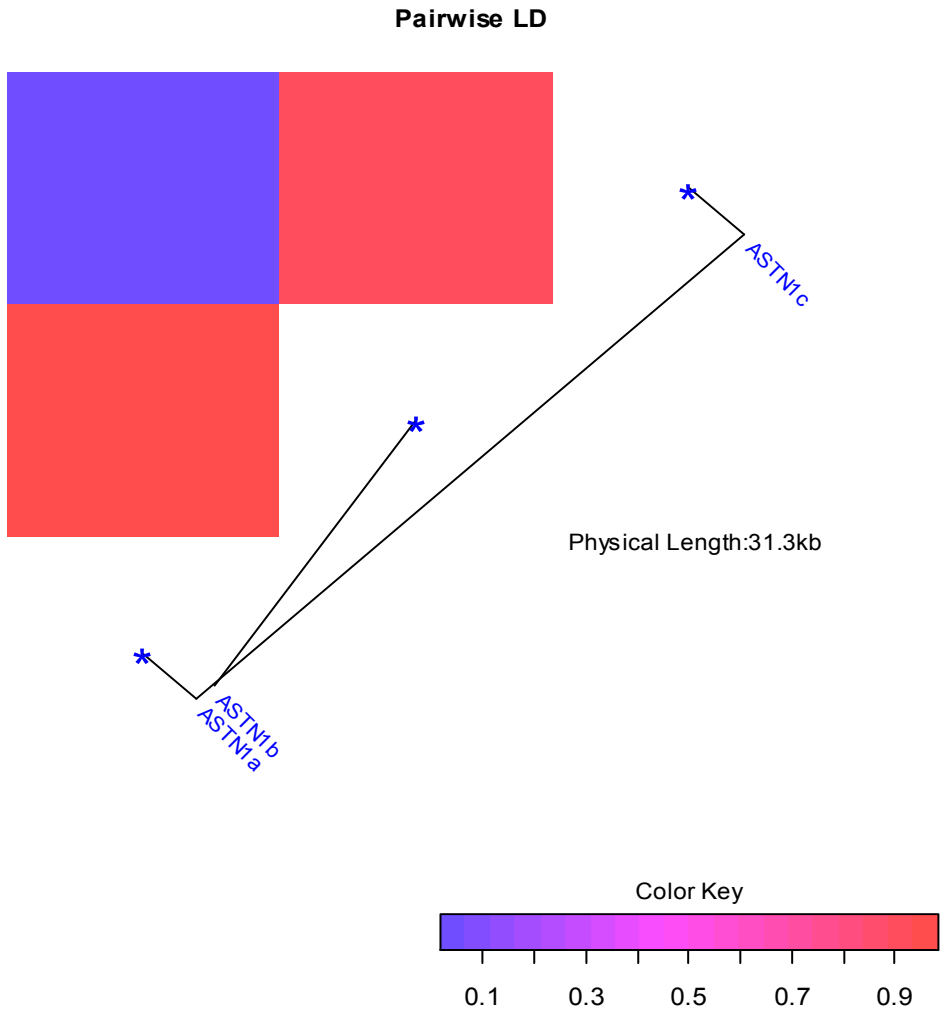


Figure 3.23: A heat map indicating the degree of LD between the *ASTN* SNPs.

3.3.6.2 Reelin Markers

Of the four SNPs chosen in the *RELN* gene, two of them proved to be monomorphic and so informative data was collected for two SNPs, rs362691 and rs607755. The heat map below indicates the degree of linkage disequilibrium between markers *RELNa* (rs362691) and *RELNb* (rs607755) based on the control samples' results. The distance between the SNPs is 138.8kb. The two SNP are in moderate LD (D'= 0.396) with each other.

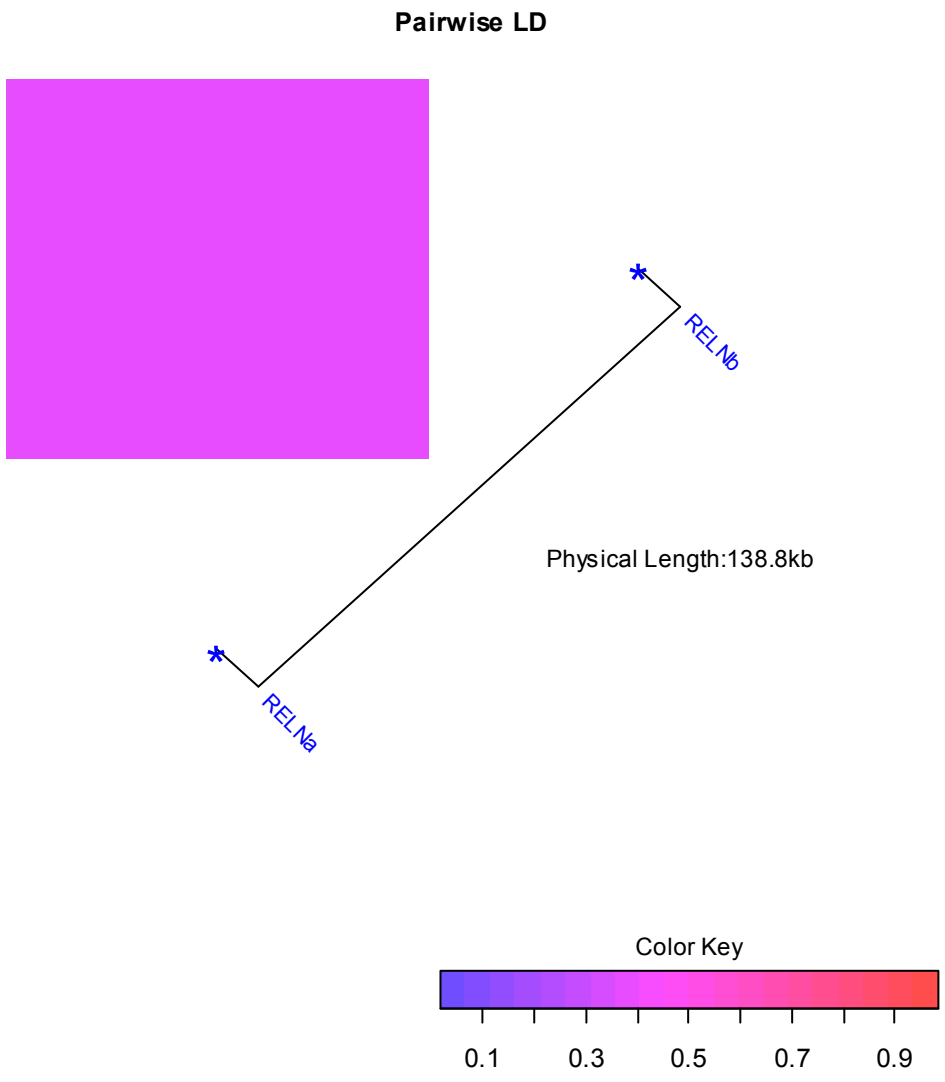


Figure 3.24: A heat map indicating the degree of LD between the *RELN* SNPs.

CHAPTER 4: DISCUSSION

4.1 PROJECT FINDINGS

In this particular FAS genetics study the focus was on genes involved in the neuronal migratory pathway. Up until now, the few studies that have been conducted on the genetics behind FAS and other alcohol related birth defects have focussed on alcohol's metabolic pathway. It is well known that alcohol affects the process of neuronal migration in the developing brain and so this pathway, and the genes involved, appeared to be suitable for investigation in a case-control association study on FAS.

Astrotactin and *Reelin* were selected as candidate genes based on the crucial role they both play in neuronal migration. Also, the phenotypic presentation of *ASTN* and *RELN* defects are similar to those seen in FAS individuals.

4.1.1 Laboratory Results

The majority of the laboratory analysis conducted on the four *ASTN* and four *RELN* SNPs was successful. PCR reactions were successful and the desired products were obtained in all instances. Problems arose in one of the *ASTN* SNPs (rs11587640) that required a highly sensitive restriction enzyme, *Sbf*I, for digestion. Residual salts present in the DNA (a result of the commercial DNA extraction kit used) inhibited the enzyme. An isoschizomer for *Sbf*I was located but was not available in South Africa. It was therefore decided to abandon this SNP from further analysis.

Another problem encountered was the discovery of uninformative SNPs. Two of the *RELN* SNPs chosen for investigation, rs3025963 and rs3025966, proved to be uninformative with no genetic variation seen at all in either cases or controls. These uninformative SNPs were reported on the Ensembl database as being "suspected' with no frequency data ever having been generated and recorded. Both these SNPs were considered coding, non-synonymous SNPs. Most researchers select non-synonymous

SNPs over those considered synonymous or intronic, as they are more likely to have an effect on the functionality of the protein. *RELN's* rs3025966 was selected as a SNP for investigation in April 2005, in August of that year it had been removed from the *RELN* SNP list on Ensembl and then two months later it had been placed back on the list but on this occasion it was referred to as a "stop gained SNP, (Y/*)." A stop gained SNP is one where the allele change forms a stop codon thus terminating translation. This emphasises the importance of validated data being recorded and available for public access through databases. It should be a guideline for future studies that only SNPs that have been validated, and for which frequency data have been generated, should be used experimentally. Alternatively, researchers could identify *de novo* SNPs in the populations on which the studies are done, especially if they have not previously been well studied.

4.1.2 Comparison of Genotype and Allele Frequencies in Different Populations

Genotype and allele frequencies were calculated for those SNPs that had sufficient genotypic information. The T/C genotype for *RELN's* rs607755 was significantly more common in the control group when compared to the case group (0.545 vs. 0.333). Although this SNP gives rise to a missense mutation (leucine to serine), no known functional difference has been attributed to it. The change from leucine to serine involves the protein changing from being uncharged to charged polar which could indicate a likely functional effect. It is not clear how the heterozygous state is associated with protectiveness against FAS but one possible explanation would be heterozygous advantage. However, the relevance of such an explanation to a monomeric extracellular protein is not obvious.

The allele and genotype frequencies calculated for the various SNPs are important and interesting data that can be entered into a database such as Ensembl or the National Center for Biotechnology Information's (NCBI's) SNP database known as "dbSNP." Table 4.1 indicates genotype and allele frequencies recorded on Ensembl for some of the *ASTN* and *RELN* SNPs as well as the frequency data generated from this study. HapMap data is available for all five markers investigated in this study. The HapMap frequency data is comparable to the Ensembl data (http://www.hapmap.org).

Table 4.1: Comparison of genotype and allele frequencies in three different populations for some of the *ASTN* and *RELN* SNPs investigated

			Population Frequencies		
Gene	**SNP and strand alleles correspond to**	**Genotype/ Allele**	**African American* (n = 23)**	**European, American Descent* (n = 24)**	**South African, Northern Cape Coloured ** (n = 112)**
ASTN	rs6413830 (forward strand)	G/G	0.435	0.292	0.589
		G/A	0.478	0.583	0.357
		A/A	0.087	0.125	0.053
		A	0.326	0.417	0.232
		G	0.674	0.583	0.768
	rs2228956 (forward strand)	C/C	0.591	0.750	0.830
		T/C	0.318	0.208	0.152
		T/T	0.091	0.042	0.018
		C	0.750	0.854	0.906
		T	0.250	0.146	0.094
RELN	rs362691 (reverse strand)	C/C	0.826	0.750	0.902
		C/G	0.174	0.167	0.098
		G/G	0.000	0.083	0.000
		C	0.913	0.833	0.951
		G	0.087	0.167	0.049
	rs607755 (reverse strand)	C/C	0.348	0.250	0.313
		T/C	0.174	0.625	0.545
		T/T	0.478	0.125	0.143
		C	0.435	0.562	0.585
		T	0.565	0.438	0.415

* Taken from the Ensembl database

** Based on controls only

In certain instances only two of the three possible genotypes were observed (tables 3.1 and 3.2). Where genotype and allele data for other populations is available, one can compare whether or not the genotypes not observed in this study were reported in other populations and at what frequency. Genotype and allele frequencies were available for African Americans and Americans of European descent for markers rs2228956 and rs362691. Although genotype and allele frequencies do vary amongst different populations, the comparative values merely give an indication of what one could expect from a population that is considered admixed and comprised of African and European ancestry. From table

4.1 it is evident that the T/T genotype of rs2228956 occurs at a low frequency, 0.091 and 0.042 in the other two populations. This would indicate that the frequency findings from this study follow the genotype distribution in the African American and European populations. Like the Coloured population, the G/G genotype for rs362691was not observed in the African American population and occurs at a low frequency of 0.083 in the European population. These comparative frequencies support the likelihood that the G/G genotype will not be observed in the Northern Cape Coloureds, given the sample size. If one had a larger sample size, one might locate the genotype at a very low frequency. In the Coloured population the "G" allele occurs at a frequency of 0.049 and so one would expect a G/G frequency of $(0.049)^2 = 0.002$. Therefore, if the population was in HWE and the sample size was large enough one would expect to see the G/G genotype.

4.1.3 Impact of Admixture on Identifying Genetic Susceptibility Factors

The population used for this study is considered to be an admixed one. Admixture arises from interbreeding between different populations. Admixed populations can be most useful in candidate gene studies as they tend to exhibit different disease and marker allele frequencies compared to the parental populations (Wright et al. 1999). The process of admixture generates LD between both linked and unlinked loci that differ in allele frequencies in the parental populations. The degree of difference between the parental allele frequencies, the level of admixture and the time since admixture occurred are all factors affecting admixture linkage disequilibrium (ALD). Admixture is a continuous process that tends to create disequilibrium that decays rapidly for unlinked markers (within two to four generations) but slowly for linked markers. Admixture is considered a continuous process in two ways, internally and externally. Internal admixture involves random mating between members of the same admixed population whereas external admixture results from random mating between members of an admixed population and members of different populations. Admixed populations are of particular usefulness in LD mapping studies (Halder and Shriver, 2003).

One of the most significant disadvantages of using an admixed population is population stratification. It is unlikely that an admixed population will consist of individuals of the

same admixture levels thus resulting in population stratification, which can cause false-positive associations between genetic markers and a given disease. It is advised that researchers using admixed populations in their investigations correct for admixture in order to prevent false-positive associations from arising. However, in order to correct for admixture, it is important for a researcher to know the proportional contributions of the parental populations and also to know the level of stratification caused by admixture in the population under investigation (Halder and Shriver, 2003).

With regards to the current investigation involving the Northern Cape Coloured population, one is unable to state the contribution of the ancestral populations as the admixed population arose hundreds of years ago from many different populations including the European, Khoisan and Central African black immigrants. Also, admixture in this case is an ongoing process and so the degree of it cannot be assessed. For these reasons this study did not statistically correct for admixture effects and so population stratification could have been a confounding factor.

In order to rule out population stratification, inbreeding, and genotyping errors in genetic association studies, the Hardy-Weinberg equilibrium (HWE) test is employed. HWE was used to test whether all *ASTN* and *RELN* markers under investigation conformed to Hardy-Weinberg expectations. Based on the results presented in tables 3.4 and 3.5, all *P*-values were above 0.05 indicating that all markers were in HWE. This confirmed that the genotyping and sampling procedures were efficient and that it was unlikely that major confounding population substructure existed.

4.1.4 The Effect of Small Sample Size

A small sample size results in a low power, which decreases one's chance of observing an association between markers of minor effect and disease susceptibility. With the significant genotypic association obtained for rs607755 (P= 0.049) a post-test power calculation generated a value of 0.58. Although this is not a poor post-test power value, an ideal value would lie between 75-80% to indicate a high chance of one detecting this association in a new study using the same sample size. One must therefore bear in mind

that it might not be a true positive association as a result of the low post-test power value. It is impossible for one to conclude that the other markers are definitely not involved in the disease, they might have a subtle role in disease pathogenesis or protection but due to the small sample size and therefore lack of adequate statistical power, no association was observed. If a gene plays a major role a smaller sample size should still allow one to observe it. However, if a gene plays a minor role, it would only be observed with a large sample size.

4.1.5 Inferred Associations Between the Markers and Disease Susceptibility

The haplotype analysis revealed that cases and controls do not differ significantly for either gene (tables 3.7 and 3.8). *ASTN's* global simulated *P*-value was 0.90 and *RELN's* was 0.269; neither was below 0.05. The analysis revealed that the GGC haplotype (which harboured alleles, rs6413830^G, rs12118933^G and rs2228956^C) was the most common haplotype for the *ASTN* gene as it was present in 68% of controls and 73% of cases. For *RELN*, the most common haplotype was CC (alleles rs362691^C and rs607755^C) with 56.7% of controls and 65.6% of cases having this haplotype.

The backward stepwise logistic regression analysis revealed that the two *RELN* SNPs best fit the "FAS versus controls" model. The *P*-value generated from the overall model was significant (*P*= 0.032) which indicates that the entire model significantly distinguishes between cases and controls. An independent significant association was seen between rs607755's C/T genotype (*P*= 0.026) and control subjects thus suggesting it may have a minor protective role against FAS.

4.1.6 Interaction Between the Markers

The results generated from the pairwise linkage disequilibrium (LD) tests indicated that *ASTN's* rs6413830 and rs12118933 were tightly linked. High LD was also seen between rs12118933 and rs2228956. Rs6413830 and rs2228956 were in very low LD (figure 3.23). When markers are in high LD they are thought to be inherited together and any effect they have will be a combined one. However, markers that are situated close together are often

in LD due to their proximity, especially in recently admixed populations. The *ASTN* SNPs reported to be in high LD are situated closely together therefore indicating that they could possibly be inherited together or the high *D'* values are simply due to their proximity to one another. The two *RELN* SNPs were shown to be in moderate LD suggesting the two SNPs act independently from one another (figure 3.24).

4.1.7 Decision Not to Correct for Multiple Testing

An issue that could possibly be raised from this genetics study is the absence of correction for multiple testing. The topic of correcting for multiple testing in a genetics study is a controversial one that has been seriously debated (Benjamini et al. 2001). The major concern for studies employing multiple statistical tests is false-positive associations. There are several ideas and suggestions as to how and why one should correct for multiple testing but there is no single correct method that is appropriate in all situations.

Van den Oord (2005) mentioned that adjustments for multiple testing are rarely seen in candidate gene studies where a relatively small number of genetic markers are typed. This could be due to genetic researchers being hesitant to correct for multiple testing as it may reduce the power of a study and therefore lower the chances of being able to detect any associations. It has also been raised that the methods of correcting for multiple testing are inappropriate for genetic studies as genetic markers are often analysed in a variety of ways (individually or simultaneously as haplotypes) and therefore tests performed cannot always be considered independent (van den Oord, 2005).

Several methods have been used to limit false discoveries such as the Bonferroni correction, the sequential Bonferroni (Gordi and Khamis, 2004) and the False Discovery Rate (FDR) (Benjamini et al. 2001). One of the most commonly used procedures is the Bonferroni correction. The Bonferroni correction involves dividing the α value (significance level), usually set at 0.05, by the total number of tests performed. This generates a corrected α value that the *P*-values are compared against to test for any significant association. This test has the reputation of being too conservative (Nyholt, 2001).

The problem of false-positives arising is most severe when the various tests are independent as opposed to being correlated. Correlated tests are less likely to form false discoveries. If markers are in significant LD with one another, their association results with the disease will be correlated. It has been suggested that where significant LD exists between markers it is possibly better to just report the disease-marker association results without correction and then state the LD association between the markers (Nyholt, 2001).

For this FAS genetics study involving *ASTN* and *RELN* markers, correction for multiple testing was not performed for the following reasons. Firstly, there is no definite correcting procedure that has been universally accepted and utilised in the field of Human Genetics for these types of studies. Secondly, correction for multiple testing considers independent variables as the number of tests; this study involved both independent and dependent variables. *ASTN's* rs6413830 and rs12118933 were considered to be in high LD and so were markers rs12118933 and rs2228956 (figure 3.1). However, not all markers of this study were in LD with one another, which resulted in some independent and some correlated variables being present.

Thirdly, if one were to ignore the possible correlated variables and count individual tests for this study, the number would be enormous considering the various statistical analyses that were conducted. A large number of tests would result in an inconsequential α and true associations with factors that make a minor contribution to the phenotype may be missed. A fourth and final reason for not correcting for multiple testing in this study was the fact that the candidate genes were not randomly chosen, they were selected based on their appropriateness for this particular study. One would expect a higher number of spurious associations to be seen between a disease and randomly selected genes. However, *ASTN* and *RELN* play a definite role in the neuronal migration pathway and were selected based on the phenotypic similarities seen between animal models when these genes or gene regions are mutated and FAS individuals. The likelihood of finding a true association would therefore be high.

4.2 PROJECT LIMITATIONS

Like most research investigations this study had its limitations. A confounding factor to this study might be alcoholism itself. The reason for this is that children inherit half their genes from their mothers and there is a high likelihood that mothers of FAS children are alcoholics. One could therefore be detecting genes associated with alcoholism instead of those associated with susceptibility to FAS.

The greatest limiting factor was the small sample size and consequent lack of adequate power. FAS is prevalent in relatively isolated, rural areas of the Northern Cape and so collecting case samples is dependent upon several logistical factors. Diagnosing FAS individuals does not happen on a continuous basis. The trained medical team from FARR and Johannesburg's Department of Human Genetics attend the Northern Cape clinics once a month for about five days at a time.

Collection of FAS samples is therefore highly dependent upon a number of people including the professionals responsible for making a positive diagnosis and those in charge of obtaining informed consent and taking blood samples. In addition to that, the lengthy assessment necessary for confirming or ruling out a FAS diagnosis in an individual allows only a few patients to be seen per visit. The turn-around time from diagnosis to collection of a blood sample to receiving the sample in Johannesburg is relatively slow and limits the number of case subjects obtained. A larger sample size would increase the power of the study and heighten the chances of one observing an association between a marker of a gene that might only have a subtle effect on the susceptibility for FAS.

Another problem observed was the method of SNP detection. This study pointed out how not all information presented on genetics/bioinformatics databases is valid. Selecting and analysing SNPs that have not been verified with frequency data can often be a costly waste of time.

The final limiting factor is the actual time taken to analyse a small number of SNPs in candidate genes. Genotyping SNPs manually in a low throughput laboratory as described previously is a lengthy process and therefore limits the number of candidate genes and markers one can investigate per study. In high throughput laboratories thousands of SNPs can be typed in one day due to the automated process that is carried out by costly equipment.

CHAPTER 5: CONCLUSION AND FUTURE PROSPECTS

This FAS genetics study involving *ASTN* and *RELN* yielded only one significant genotypic association between *RELN's* rs607755 and FAS. The logistic regression model supported this by indicating that the heterozygous state of rs607755 (C/T) was more likely to be associated with control individuals thus inferring a possible protective effect against the disease. One of the limitations of this study was the lack of adequate power related to the small sample size, due to this one cannot completely rule out any possible role of *ASTN* and the other *RELN* marker or other closely linked genetic variants in disease susceptibility.

Regarding the frequency results generated for the SNPs in the control group of the Northern Cape Coloured population, these will be submitted to Ensembl along side the frequencies for the African American and European populations. These would be good comparative data representing a truly admixed population and confirming the polymorphic nature of the SNPs.

For future studies a more efficient SNP selection method should be employed to prevent uninformative SNPs from being examined in the laboratory. Information from the HapMap project may be considered for future work but used with caution in populations that have not yet been studied. The International HapMap Project was initiated with the aim of constructing a public database of common variation in the human genome and was launched in 2002. The first phase of the project has just been completed; over one million SNPs in 269 DNA samples from four different populations have been recorded (The International HapMap Consortium, 2005). In essence the project aims to map the "structure of allelic associations" present in the human genome (McVean et al. 2005). The HapMap database (http:www/hapmap.org) indicates haplotype blocks and SNPs that are in high LD with one another. This information will be useful for candidate gene studies as researchers will be able to select SNPs with good coverage across the candidate genes. Also, the information presented in the HapMap database has been confirmed and so it is unlikely for researchers to choose uninformative SNPs for their investigations but, as

mentioned above, this information should be used with caution as only a few populations have been typed.

In order to improve future FAS genetics studies researchers should consider the fine categorisation of case individuals into more stringent phenotype groups. FAS diagnosis is dependent upon the presence of the typical diagnostic features listed in table 1.1; however, the degree of severity in each case individual can vary. It might be of considerable value to divide cases into different categories based on the measurements taken to confirm a diagnosis. In addition to that, cases that present with other anomalies that are not always associated with FAS such as the cardiac and skeletal defects should be categorised separately. As described previously, complex disorders like FAS arise due to the collective effect of a number of susceptibility DNA variants interacting together with the environment. Some genetic variants are likely to have a very subtle role in disease pathogenesis and so the more stringent grouping of cases, based on their phenotype, could be useful for detection of associations within genes of small effect.

Genes in the neuronal migration pathway remain an excellent choice of candidate genes to investigate in FAS. There are many other genes, besides *ASTN* and *RELN*, from which investigators could select additional candidates. One suggestion could be to look at polymorphisms present the receptors of *ASTN* and *RELN*. Despite not finding a highly significant association between the genes and FAS, *ASTN* and *RELN* should be investigated further in a FAS genetics study focussing on neuronal migration. Once several more samples from the same population have been collected the current study should be extended. It is possible that with an increased sample size and increased power, more associations could be found between *ASTN*, *RELN* and FAS. Due to not correcting for multiple testing in this study, repeating the investigation and seeing whether or not the same association between *RELN's* rs607755 materialises could point towards the original finding as being either a true association or a false-positive.

Candidate-gene studies involving complex diseases are far more complicated than those involving Mendelian diseases and are therefore still in their infancy. The study under review emphasises the difficulties in conducting a genetics study on a complex disorder.

Together with the availability of more accurate human genetic information (such as the HapMap), researchers should take into account the outcomes and limitations of small genetic studies and use them to develop better means of candidate gene selection and investigation. Improved bioinformatical methodology possibly involving text mining and the linking of terms associated with the disease and specific genes, could minimise the number of candidate gene choices and assist in selecting reputable candidate genes for case-control studies involving complex diseases.

With regards to FAS in particular, prevention is undoubtedly the best cure. If a mother abstains from alcohol during her pregnancy the risk of producing a FAS child or a child affected by any other FASD condition is zero. However, the importance of studies such as the one discussed in this dissertation lies in the understanding of the molecular basis behind the disease. A better understanding of FAS at the molecular and cellular level could aid researchers in developing strategies to minimise the effects of the disease and possibly protect the unborn child from the severity of alcohol's teratogenicity through supplements, much like folic acid protects against neural tube defects. In addition to that, pinpointing genetic markers could allow for the development of a set of genetic tests as markers for the disease in the future.

REFERENCES

Adams, N.C., Tomoda, T., Cooper, M., Dietz, G. and Hatten, M.E. (2002) Mice that lack astrotactin have slowed neuronal migration. *Development* **129**, 965-72.

Archibald, S.L., Fennema-Notestine, C., Gamst, A., Riley, E.P., Mattson, S.N. and Jernigan, T.L. (2001) Brain dysmorphology in individuals with severe prenatal alcohol exposure. *Dev Med Child Neurol* **43**, 148-54.

Ardlie, K.G., Kruglyak, L. and Seielstad, M. (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* **3**, 299-309.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. and Golani, I. (2001) Controlling false discovery rate in behavior genetics research. *Behav Brain Res* **125**, 279-84.

Beyers, N. and Moosa, A. (1978) The fetal alcohol syndrome. *S Afr Med J* **54**, 575-8.

Boehm, S.L. 2nd, Lundahl, K.R., Caldwell, J. and Gilliam, D.M. (1997) Ethanol teratogenesis in the C57BL/6J, DBA/2J, and A/J inbred mouse strains. *Alcohol* **14**, 389-95.

Bonora, E., Beyer, K.S., Lamb, J.A., Parr, J.R., Klauck, S.M., Benner, A., Paolucci, M., Abbott, A., Ragoussis, I., Poustka, A., Bailey, A.J. and Monaco, A.P. (2003) Analysis of reelin as a candidate gene for autism. *Mol Psychiatry* **8**, 885-92.

Brookes, A.J. (1999) The essence of SNPs. *Gene* **234**, 177-86.

Campbell, N.A., Reece, J.B. and Mitchell, L.G. (1999) Biology. Fifth edition. California: Addison-Wesley Longman, Inc. 978.

Carlson, C.S., Eberle, M.A., Kruglyak, L. and Nickerson, D.A. (2004) Mapping complex disease loci in whole-genome association studies. *Nature* **429** , 446-52.

Centers for Disease Control (2003) Fetal alcohol syndrome- South Africa, 2001. *MMWR* **52**, 660-2.

Chaudhuri, J.D. (2000) Alcohol and the developing fetus--a review. *Med Sci Monit* **6**, 1031-41.

Clarren, S.K., Alvord, E.C. Jr, Sumi, S.M., Streissguth, A.P. and Smith, D.W. (1978) Brain malformations related to prenatal exposure to ethanol. *J Pediatr* **92**, 64-7.

Clarren, S.K. and Smith, D.W. (1978) The fetal alcohol syndrome. *N Engl J Med* **298**, 1063-7.

Croxford, J. and Viljoen, D. (1999) Alcohol consumption by pregnant women in the Western Cape. *S Afr Med J* **89**, 962-5.

Cudd, T.A. (2005) Animal model systems for the study of alcohol teratology. *Exp Biol Med (Maywood)* **230**, 389-93.

Daly, A.K. and Day, C.P. (2001) Candidate gene case-control association studies: advantages and potential pitfalls. *Br J Clin Pharmacol* **52**, 489-99.

Dean, M. (2003) Approaches to identify genes for complex human diseases: lessons from Mendelian disorders. *Hum Mutat* **22**, 261-74.

Dick, D.M. and Foroud, T. (2002) Genetic strategies to detect genes involved in alcoholism and alcohol- related traits. *Alcohol Res Health* **26**, 172-80.

Dirks, P.B. (2001) Glioma migration: clues from the biology of neural progenitor cells and embryonic CNS cell migration. *J Neurooncol* **53**, 203-12.

Fatemi, S.H. (2005) Reelin glycoprotein: structure, biology and roles in health an disease. *Mol Psychiatry* **10**, 251-7.

Fink, J.M., Hirsch, B.A., Zheng, C., Dietz, G., Hatten, M.E. and Ross, M.E. (1997) Astrotactin (ASTN), a gene for glial-guided neuronal migration, maps to human chromosome 1q25.2. *Genomics* **40**, 202-5.

Fisher, R.A. (1922) On the interpretation of χ^2 from contingency tables, and the calculation of P. *J Roy Statist Soc* **85**, 87-94.

Frosk, P., Weiler, T., Nylen, E., Sudha, T., Greenberg, C.R., Morgan, K., Fujiwara, M. and Wrogemann, K. (2002) Limb-girdle muscular dystrophy Type 2H associated with mutation in *TRIM32* , a putative E3-ubiquitin-ligase gene. *Am J Hum Genet* **70**, 663-72.

Goodlett, C.R. and Horn, K.H. (2001) Mechanisms of alcohol-induced damage to the developing nervous system. *Alcohol Res Health* **25**, 175-84.

Gordi, T. and Khamis, H. (2004) Simple solution to a common statistical problem: Interpreting multiple tests. *Clin Ther* **25**, 780-6.

Guerri, C. (2002) Mechanisms involved in central nervous system dysfunctions induced by prenatal ethanol exposure. *Neurotox Res* **4**, 327-35.

Guerri, C., Pascual, M. and Renau-Piqueras, J. (2001) Glia and fetal alcohol syndrome. *Neurotoxicology* **22**, 593-9.

Halder, I. and Shriver, M.D. (2003) Measuring and using admixture to study the genetics of complex diseases. *Hum Genomics* **1** , 52-62.

Hanson, J.W., Streissguth, A.P. and Smith, D.W. (1978) The effects of moderate alcohol consumption during pregnancy on fetal growth and morphogenesis. *J Pediatr* **92**, 457-60.

Hartl, D.L. and Clark, A.G. (1989) Principles of population genetics. Second edition. Massachusetts: Sinauer Associates, Inc.

Hatten, M.E. (1999) Central nervous system neuronal migration. *Annu Rev Neurosci* **22**, 511-39.

Hirai, K., Yoshioka, H., Kihara, M., Hasegawa, K., Sawada, T. and Fushiki, S. (1999) Effects of ethanol on neuronal migration and neural cell adhesion molecules in the embryonic rat cerebral cortex: a tissue culture study. *Brain Res Dev Brain Res* **118**, 205-10.

Honda, T., Tabata, H. and Nakajima, K. (2003) Cellular and molecular mechanisms of neuronal migration in neocortical development. *Semin Cell Dev Biol* **14**, 169-74.

Hoyme, H.E., May, P.A., Kalberg, W.O., Kodituwakku, P., Gossage, J.P., Trujillo, P.M., Buckley, D.G., Miller, J.H., Aragon, A.S., Khaole, N., Viljoen, D.L., Jones, K.L. and Robinson, L.K. (2005) A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: clarification of the 1996 institute of medicine criteria. *Pediatrics* **115**, 39-47.

Jacobson, J.L. and Jacobson, S.W. (2002) Effects of prenatal alcohol exposure on child development. *Alcohol Res Health* **26**, 282-6.

Jones, K.L., Smith, D.W., Ulleland, C.N. and Streissguth, P. (1973) Pattern of malformation in offspring of chronic alcoholic mothers. *Lancet* **1**, 1267-71.

Koirala, S. and Ko, C.-P. (1998) The sixth annual university of Southern California neuroscience symposium, Los Angeles, November 6 1998. *CNS Drug Reviews* **4**, 380-3.

Kvigne, V.L., Leonardson, G.R., Borzelleca, J., Brock, E., Neff-Smith, M. and Welty, T.K. (2003) Characteristics of mothers who have children with fetal alcohol syndrome or some characteristics of fetal alcohol syndrome. *J Am Board Fam Pract* **16**, 296-303.

Kvigne, V.L., Leonardson, G.R., Neff-Smith, M., Brock, E., Borzelleca, J. and Welty, T.K. (2004) Characteristics of children who have full or incomplete fetal alcohol syndrome. *J Pediatr* **145**, 635-40.

Kwon, J.M. and Goate, A.M. (2000) The candidate gene approach. *Alcohol Res Health* **24**, 164-8.

Lambert de Rouvroit, C. and Goffinet, A.M. (2001) Neuronal migration. *Mech Dev* **105**, 47-56.

Lenwontin, R.C. (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genet* **49**, 49-67.

Lombroso, P.J. (1998) Development of the cerebral cortex: III. The reeler mutation. *J Am Acad Child Adolesc Psychiatry* **37**, 333-4.

Long, A.D., Grote, M.N. and Langley, C.H. (1997) Genetic analysis of complex diseases. *Science* **275**, 1328; author reply 1329-30.

Lupton, C., Burd, L. and Harwood, R. (2004) Cost of fetal alcohol spectrum disorders. *Am J Med Genet C Semin Med Genet* **127**, 42-50.

Madigan, M.T., Martinko, J.M. and Parker, J. (2000) Brock, biology of microrganims. New Jersey: Prentice Hall, Inc. 170-2.

Mathew, C. (2001) Science, medicine, and the future: Postgenomic technologies: hunting the genes for common disorders. *BMJ* **322**, 1031-4.

Mattson, S.N., Schoenfeld, A.M. and Riley, E.P. (2001) Teratogenic effects of alcohol on brain and behavior. *Alcohol Res Health* **25**, 185-91.

May, P.A., Brooke, L., Gossage, J.P., Croxford, J., Adnams, C., Jones, K.L., Robinson, L. and Viljoen, D. (2000) Epidemiology of fetal alcohol syndrome in a South African community in the Western Cape Province. *Am J Public Health* **90**, 1905-12.

May, P.A., Gossage, J.P., White-Country, M., Goodhart, K., Decoteau, S., Trujillo, P.M., Kalberg, W.O., Viljoen, D.L. and Hoyme, H.E. (2004) Alcohol consumption and other maternal risk factors for fetal alcohol syndrome among three distinct samples of women before, during, and after pregnancy: the risk is relative. *Am J Med Genet* **127C**, 10-20.

McCarver, D.G., Thomasson, H.R., Martier, S.S., Sokol, R.J. and Li, T.K. (1997) Alcohol dehydrogenase-2*3 allele protects against alcohol related birth defects among African Americans. *J Pharmacol Exp Ther* **283**, 1095-101.

McVean, G., Spencer, C.C.A. and Chaix, R. (2005) Perspectives on human genetic variation from the HapMap project. *PLOS Genetics* **1**, 413-8.

Michaelis, E.K. and Michaelis, M.L. (1994) Cellular and molecular bases of alcohol's teratogenic effects. *Alcohol Health Res World* **18**, 17-21.

Moore, K.L. and Persaud, T.V.N. (1993) The developing human: clinically orientated embryology. Fifth edition. Philadelphia: W.B. Saunders.

Mueller, R.F. and Young, I.D. (2003) Emery's elements of medical genetics. Eleventh edition. London: Churchill Livingstone. 23, 58-9, 14, 127-33.

Neese, S., La Grange, L., Trujillo, E. and Romero, D. (2004) The effects of ethanol and silymarin treatment during gestation on spatial working memory. *BMC Complement Altern Med* **4**, 4

NIAAA (2000) Prenatal exposure to alcohol. *Alcohol Res Health* **24**, 32-41.

Nyholt, D.R. (2001) Genetic case-control association studies- correcting for multiple testing. *Human Genetics* **109**, 564-65.

Olney, J.W. (2004) Fetal alcohol syndrome at the cellular level. *Addict Biol* **9**, 137-49; discussion 151.

Pearlman, A.L., Faust, P.L., Hatten, M.E. and Brunstrom, J.E. (1998) New directions for neuronal migration. *Curr Opin Neurobiol* **8**, 45-54.

R Development Core Team (2005) R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R. and Lander, E.S. (2001) Linkage disequilibrium in the human genome. *Nature* **411**, 199-204.

Riikonen, R., Salonen, I., Partanen, K. and Verho, S. (1999) Brain perfusion SPECT and MRI in foetal alcohol syndrome. *Dev Med Child Neurol* **41**, 652-9.

Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* **273**, 1516-7.

Romero, R., Kuivaniemi, H., Tromp, G. and Olson, J. (2002) The design, execution, and interpretation of genetic association studies to decipher complex diseases. *Am J Obstet Gynecol* **187**, 1299-312.

Sampson, P.D., Streissguth, A.P., Bookstein, F.L., Little, R.E., Clarren, S.K., Dehaene, P., Hanson, J.W. and Graham, J.M. Jr (1997) Incidence of fetal alcohol syndrome and prevalence of alcohol-related neurodevelopmental disorder. *Teratology* **56**, 317-26.

Schork, N.J. (1997) Genetics of complex disease: approaches, problems, and solutions. *Am J Respir Crit Care Med* **156**, S103-9.

Shaywitz, S.E., Cohen, D.J. and Shaywitz, B.A. (1980) Behavior and learning difficulties in children of normal intelligence born to alcoholic mothers. *J Pediatr* **96**, 978-82.

Sinnwell, J.P. and Schaid, D.J. (2005) Haplo.stats: Statistical analysis of haplotypes with traits and covariates when linkage phase is ambiguous. R package version 1.2.2.

Sklar, P., Gabriel, S.B., McInnis, M.G., Bennett, P., Lim Y-M. , Tsan, G., Schaffner, S., Kirov, G., Jones, I., Owen, M., Craddock, N., DePaulo, J.R. and Lander, E.S. (2002) Family-based association study of 76 candidate genes in bipolar disorder: BDNF is a potential risk locus. *Mol Psychiatry* **7**, 579-93.

Sobeih, M.M. and Corfas, G. (2002) Extracellular factors that regulate neuronal migration in the central nervous system. *Int J Dev Neurosci* **20**, 349-57.

Sokol, R.J. and Clarren, S.K. (1989) Guidelines for use of terminology describing the impact of prenatal alcohol on the offspring. *Alcohol Clin Exp Res* **13**, 597-8.

Sokol, R.J., Delaney-Black, V. and Nordstrom, B. (2003) Fetal alcohol spectrum disorder. *JAMA* **290**, 2996-9.

Stoler, J.M., Ryan, L.M. and Holmes, L.B. (2002) Alcohol dehydrogenase 2 genotypes, maternal alcohol use, and infant outcome. *J Pediatr* **141**, 780-5.

Strachan, T. and Read, A.P. (2004) Human molecular genetics. Third edition. London: Garland Science. 182-5, 445.

Streissguth, A.P., Bookstein, F.L., Barr, H.M., Sampson, P.D., O'Malley, K. and Young, J.K. (2004) Risk factors for adverse life outcomes in fetal alcohol syndrome and fetal alcohol effects. *J Dev Behav Pediatr* **25**, 228-38.

Streissguth, A.P. and Dehaene, P. (1993) Fetal alcohol syndrome in twins of alcoholic mothers: concordance of diagnosis and IQ. *Am J Med Genet* **47**, 857-61.

Sulik, K.K. (2005) Genesis of alcohol-induced craniofacial dysmorphism. *Exp Biol Med (Maywood)* **230**, 366-75.

Swayze, V.W. 2nd, Johnson, V.P., Hanson, J.W., Piven, J., Sato, Y., Giedd, J.N., Mosnik, D. and Andreasen, N.C. (1997) Magnetic resonance imaging of brain anomalies in fetal alcohol syndrome. *Pediatrics* **99**, 232-40.

The International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* **43**, 1299-320.

Thomas, J.D., Melcer, T., Weinert, S. and Riley, E.P. (1998) Neonatal alcohol exposure produces hyperactivity in high-alcohol-sensitive but not in low-alcohol-sensitive rats. *Alcohol* **16**, 237-42.

Thomas, J.D. and Riley, E.P. (1998) Fetal alcohol syndrome: does alcohol withdrawal play a role? *Alcohol Health Res World* **22**, 47-53.

van den Oord, E.J.C.G. (2005) Controlling false discoveries in candidate gene studies. *Mol Psychiatry* **10**, 230-31.

Venables, W.N. and Ripley, B.D. (2002) Modern applied statistics with S. Fourth edition. New York: Springer.

Viljoen, D. (1999) Fetal alcohol syndrome. *S Afr Med J* **89**, 958-60.

Viljoen, D., Croxford, J., Gossage, J.P., Kodituwakku, P.W. and May, P.A. (2002) Characteristics of mothers of children with fetal alcohol syndrome in the Western Cape Province of South Africa: a case control study. *J Stud Alcohol* **63**, 6-17.

Viljoen, D.L. (1991) The fetal alcohol syndrome. *CME* **9**, 783-90.

Viljoen, D.L. (2005) Foundation for Alcohol Related Research, chairman's report for the year ended 31 March 2005.

Viljoen, D.L., Carr, L.G., Foroud, T.M., Brooke, L., Ramsay, M. and Li, T.K. (2001) Alcohol dehydrogenase-2*2 allele is associated with decreased prevalence of fetal alcohol syndrome in the mixed-ancestry population of the Western Cape Province, South Africa. *Alcohol Clin Exp Res* **25**, 1719-22.

Warnes, G. and Leisch, F. (2005) Genetics: Population genetics. R package version 1.1.4.

Warren, K.R. and Li, T.K. (2005) Genetic polymorphisms: impact on the risk of fetal alcohol spectrum disorders. *Birth Defects Res A Clin Mol Teratol* **73**, 195-203.

Welch-Carre, E. (2005) The neurodevelopmental consequences of prenatal alcohol exposure. *Adv Neonatal Care* **5**, 217-29.

Wigginton, J.E., Cutler, D.J. and Abecasis, G.R. (2005) A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* **76**, 887-93.

Wilson, K. and Walker, J. (1995) Practical biochemistry and techniques. Fourth edition. New York: Cambridge University Press. 142-7.

Wood, H. (2005) Neuronal migration. Reelin out the hindbrain nuclei. *Nat Rev Neurosci* **6**, 261.

Wright, A.F., Carothers, A.D. and Pirastu, M. (1999) Population choice in mapping genes for complex diseases. *Nat Genet* **23**, 397-404.

Ye, S., Dhillon, S., Ke, X., Collins, A.R. and Day, I.N. (2001) An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res* **29**, E88-8.

Zhang, H., Liu, X., Zhang, C., Mundo, E., Macciardi, F., Grayson, D.R., Guidotti, A.R. and Holden, J.J. (2002) Reelin gene alleles and susceptibility to autism spectrum disorders. *Mol Psychiatry* **7**, 1012-7.

Zheng, C., Heintz, N. and Hatten, M.E. (1996) CNS gene encoding astrotactin, which supports neuronal migration along glial fibers. *Science* **272**, 417-9.

Electronic References

dbSNP website: http://www.ncbi.nlm.nih.gov/SNP

Ensembl website: http://www.ensembl.org
[Last accessed December 2005].

GeneCards website: http://www.genecards.org
[Last accessed April 2005].

HapMap Homepage: http://www.hapmap.org
[Last accessed 2 August 2006]

Kellerman, T. (2003) "FAS and the brain." http://www.come-over.to/FAS/FASbrain.htm
[Last accessed 28 November 2005].

Minnesota Department of Health. (2003) "Identification of at-risk drinking with women of childbearing age." http://www.health.state.mn.us/fas/at-risk-drinking/section2.html

[Last accessed 30 November 2005].

MRCPath Preparation course 1999-00 Mutation Detection Systems- Assignments for 21.9.99: http://www.ich.ucl.ac.uk/cmgs/arms99.htm
[Last accessed July 2005].

NEBCutter website: http://tools.neb.com/NEBcutter2/index.php
[Last accessed October 2005].
Primer3 website: http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi
[Last accessed August 2005].

Quick maps: http://www.theodora.com/maps [Last accessed December 2005].

UCSC genome browser: http://genome.ucsc.edu [Last accessed May 2005].

APPENDIX A

ETHICS CLEARANCE CERTIFICATES

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

COMMITTEE FOR RESEARCH ON HUMAN SUBJECTS (MEDICAL)
Ref: R14/49 Lombard

CLEARANCE CERTIFICATE **PROTOCOL NUMBER** M03-10-20

PROJECT	Fetal Alchol Syndrome in the Northern Cape Province, Investigation of Population Structure and-Admixture in the Coloured Community and Collection of Control Subjects
INVESTIGATORS	Z Lombard
DEPARTMENT	School of Pathology, NHLS
DATE CONSIDERED	03-10-31
DECISION OF THE COMMITTEE	Approved unconditionally

Unless otherwise specified the ethical clearance is valid for 5 years but may be renewed upon application
This ethical clearance will expire on 1 January 2008.

DATE 03-12-07 CHAIRMAN .....(Professor P E Cleaton-Jones)

* Guidelines for written "informed consent" attached where applicable.

c c Supervisor: Prof M Ramsay
Dept of School of Pathology , NHLS
Works2\Iain0015\HumEth97.wdb\M 03-10-20

==========

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and ONE COPY returned to the Secretary at Room 10001, 10th Floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. I agree to a completion of a yearly progress form. I/we agree to inform the Committee once the study is completed.

DATE 3/6/2004 SIGNATURE .....

PLEASE QUOTE THE PROTOCOL NO IN ALL QUERIES .: M 03-10-20
PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

q

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

~~COMMITTEE FOR RESEARCH ON HUMAN SUBJECTS (MEDICAL)~~
Ref: R14/49 Viljoen/Ramsay

DIVISION OF THE DEPUTY REGISTRAR (RESEARCH)
RECEIVED 2002 -12- 12
FILE:
ACTION:

CLEARANCE CERTIFICATE PROTOCOL NUMBER M02-10-41

PROJECT A Search for Gentic Factors Predisposing to Fetal alcohol Syndrome in South African Populations

INVESTIGATORS Prof's D/M Viljoen/Ramsay

DEPARTMENT School of Pathology, NHLS

DATE CONSIDERED 02-10-25

DECISION OF THE COMMITTEE Approved unconditionally

Unles otherwise specified the ethical clearance is valid for 5 years but may be renewed upon application
This ethical clearance will expire on 30 July 2007.

DATE 02-12-02 CHAIRMAN .....(Professor P E Cleaton-Jones)

* Guidelines for written "informed consent" attached where applicable.

c c Supervisor: Prof D Viljoen
Dept of School of Pathology, NHLS
Works2\Iain0015\HumEth97.wdb\M 02-10-41

==========

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and ONE COPY returned to the Secretary at Room 10001, 10th Floor, Senate House, University.

I/we fully understand the conditions under which ~~I am~~/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. I agree to a completion of a yearly progress form. I/we agree to inform the Committee once the study is completed.

DATE 17/1/03 SIGNATURE M. Ramsay / D Viljoen

PLEASE QUOTE THE PROTOCOL NO IN ALL QUERIES .: M 02-10-41

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Macaulay

CLEARANCE CERTIFICATE **PROTOCOL NUMBER M040611**

PROJECT Assess the Role of the Astrotactin 1 Gene and Other Genes in Neural Development, in the Development of Fetal Alcohol Syndrome Within the Mixed Ancestry Population of Southern Africa

INVESTIGATORS Ms S Macaulay

DEPARTMENT Human Genetics

DATE CONSIDERED 04.06.25

DECISION OF THE COMMITTEE* of South Africa: APPROVED

UNCONDITIONALLY

Unless otherwise specified this ethical clearance is valid for 5 years and may be renewed upon application.

DATE 04.06.28 **CHAIRPERSON** .....
(Professor PE Cleaton-Jones)

*Guidelines for written 'informed consent' attached where applicable

cc: Supervisor : Prof M Ramsay

S Macaulay M. Ramsay

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10005, 10th Floor, Senate House, University.
I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to a completion of a yearly progress report.**

APPENDIX B

SNP DIAGRAMS AND GENOTYPING RESULTS

- **All recorded SNPs in the *ASTN* and *RELN* genes**

The schematic illustration below (figure B1) is the legend for figures B2 and B3. It explains what the different colours and symbols represent.

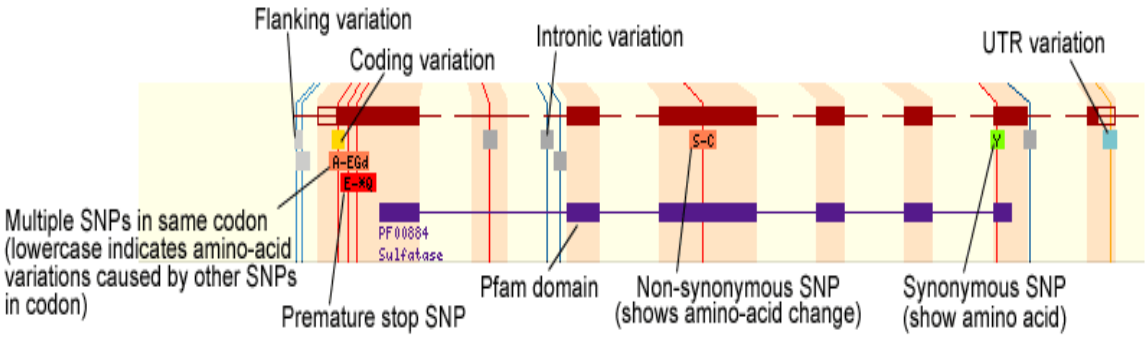


Figure B1: Diagrammatic legend explaining Ensembl's annotated gene diagrams.

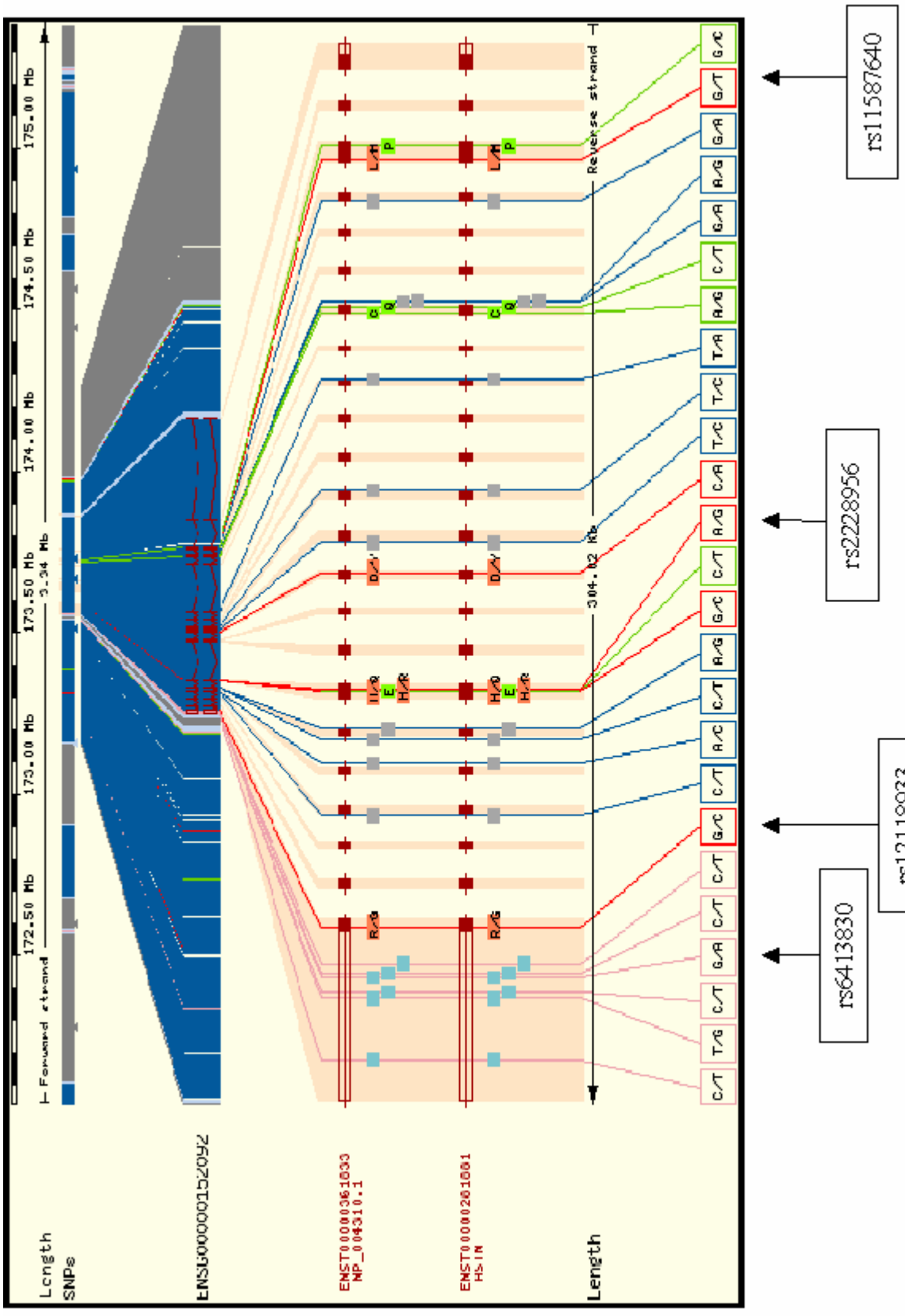


Figure B2: The positioning of all recorded SNPs in the *ASTN* gene on chromosome one (taken from Ensembl).

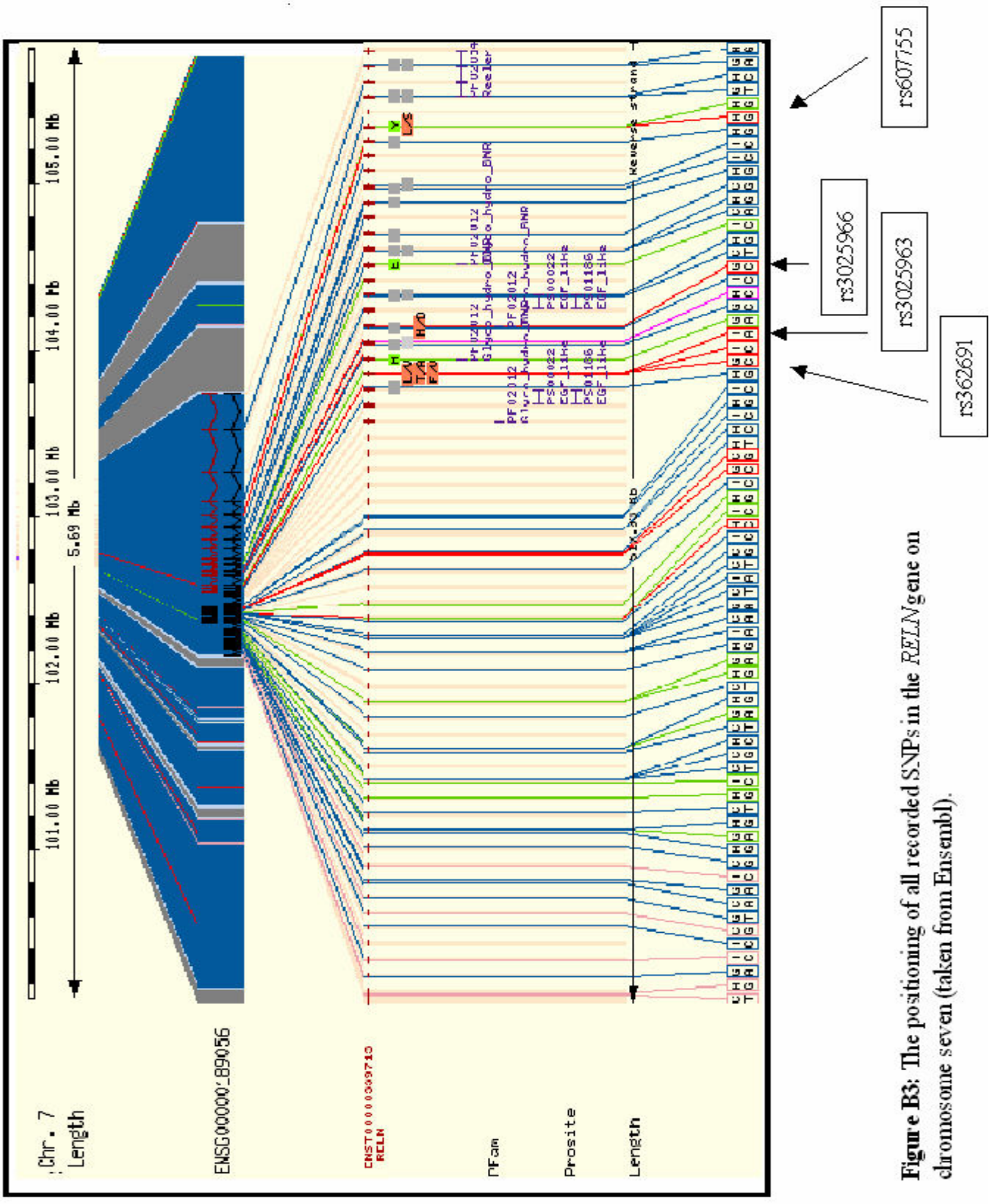


Figure B3: The positioning of all recorded SNPs in the *RELN* gene on chromosome seven (taken from Ensembl).

- **Genotyping Data Generated**

Table B1: Genotyping data generated for the *ASTN* and *RELN* SNPs investigated

Code	Sex	Case / Control	*ASTN* rs6413830	*ASTN* rs12118933	*ASTN* rs2228956	*RELN* rs362691	*RELN* rs607755
NPC 001	F	Control	G/G	G/G	C/C	C/C	T/C
NPC 003	M	Control	G/A	G/G	C/C	C/G	T/C
NPC 004	M	Control	A/A	G/G	C/C	C/C	T/T
NPC 005	F	Control	A/A	G/G	C/C	C/C	C/C
NPC 006	F	Control	G/G	G/G	T/C	C/C	T/C
NPC 007	F	Control	G/G	G/G	C/C	C/C	T/C
NPC 008	M	Control	A/A	G/G	C/C	C/C	T/C
NPC 009	F	Control	G/G	G/G	C/C	C/C	T/C
NPC 010	M	Control	G/A	G/G	C/C	C/C	T/C
NPC 011	F	Control	G/G	G/C	C/C	C/C	C/C
NPC 012	F	Control	G/G	G/G	C/C	C/C	T/C
NPC 013	F	Control	G/G	G/G	C/C	C/C	C/C
NPC 014	F	Control	G/G	G/G	C/C	C/C	T/C
NPC 015	M	Control	G/G	G/G	C/C	C/C	T/C
NPC 017	F	Control	G/A	G/G	C/C	C/C	T/C
NPC 018	F	Control	G/G	G/G	C/C	C/C	T/C
NPC 019	M	Control	G/G	G/G	T/C	C/C	T/T
NPC 021	F	Control	G/A	G/G	T/C	C/C	T/C
NPC 022	M	Control	G/A	G/G	C/C	C/C	T/T
NPC 023	M	Control	G/G	G/G	C/C	C/C	C/C
NPC 024	F	Control	G/G	G/G	C/C	C/G	T/T
NPC 025	M	Control	G/A	G/G	C/C	C/C	T/C
NPC 026	F	Control	G/G	G/G	C/C	C/C	T/C
NPC 027	F	Control	G/A	G/G	C/C	C/C	C/C
NPC 028	F	Control	A/A	G/G	C/C	C/C	T/T
NPC 029	F	Control	G/G	G/G	T/T	C/C	T/C
NPC 030	F	Control	G/G	G/G	C/C	C/C	C/C
NPC 031	F	Control	G/A	G/G	C/C	C/C	T/C
NPC 032	M	Control	G/G	G/G	C/C	C/C	T/C
NPC 033	M	Control	G/G	G/G	C/C	C/C	T/T
NPC 034	F	Control	G/A	G/G	T/C	C/C	T/C
NPC 035	M	Control	G/G	G/G	C/C	C/C	T/C
NPC 036	F	Control	G/G	G/G	C/C	C/C	C/C
NPC 037	F	Control	G/G	G/G	C/C	C/C	C/C
NPC 038	F	Control	G/G	G/G	C/C	C/C	C/C
NPC 039	F	Control	G/A	G/G	C/C	C/C	C/C
NPC 040	F	Control	G/G	G/C	C/C	C/C	T/C
NPC 041	F	Control	G/G	G/G	C/C	C/C	T/C
NPC 042	M	Control	G/G	G/G	T/C	C/C	T/C
NPC 043	M	Control	G/G	G/G	C/C	C/C	T/C
NPC 044	F	Control	G/A	G/G	C/C	C/C	C/C
NPC 045	F	Control	G/A	G/G	C/C	C/C	T/C
NPC 046	M	Control	G/G	G/G	C/C	C/C	T/C

NPC 048	M	Control	G/G	G/G	C/C	C/C	T/C
NPC 049	M	Control	G/G	G/G	C/C	C/C	T/C
NPC 050	M	Control	G/G	G/G	C/C	C/G	C/C
NPC 051	F	Control	G/G	G/G	C/C	C/C	T/T
NPC 052	F	Control	G/G	G/G	C/C	C/C	C/C
NPC 053	M	Control	G/A	G/G	C/C	C/G	T/T
NPC 054	M	Control	G/A	G/G	C/C	C/C	T/C
NPC 055	M	Control	A/A	G/G	C/C	C/C	T/C
NPC 056	F	Control	G/G	G/G	C/C	C/C	T/T
NPC 057	F	Control	G/A	G/G	T/C	C/C	T/C
NPC 058	F	Control	G/A	G/G	C/C	C/C	T/C
NPC 059	F	Control	G/A	G/G	T/C	C/C	T/C
NPC 060	F	Control	G/G	G/G	T/C	C/C	T/C
NPC 061	M	Control	G/A	G/G	C/C	C/C	C/C
NPC 062	F	Control	G/A	G/G	T/T	C/C	C/C
NPC 063	F	Control	G/A	G/G	C/C	C/C	C/C
NPC 064	M	Control	G/G	G/G	C/C	C/G	T/C
NPC 065	F	Control	G/G	G/G	C/C	C/G	T/C
NPC 066	F	Control	G/G	G/G	C/C	C/G	T/C
NPC 067	M	Control	G/G	G/G	T/C	C/G	T/C
NPC 068	M	Control	G/A	G/G	C/C	C/C	T/C
NPC 069	M	Control	G/G	G/G	C/C	C/C	T/T
NPC 070	F	Control	G/G	G/G	T/C	C/C	C/C
NPC 071	M	Control	G/G	G/G	C/C	C/G	C/C
NPC 072	M	Control	G/A	G/G	C/C	C/C	T/T
NPC 073	M	Control	G/A	G/G	T/C	C/C	T/C
NPC 074	M	Control	G/G	G/C	C/C	C/C	T/T
NPC 075	F	Control	G/G	G/G	C/C	C/C	T/T
NPC 077	M	Control	G/G	G/G	C/C	C/C	C/C
NPC 079	M	Control	G/G	G/G	C/C	C/C	C/C
NPC 080	M	Control	G/A	G/G	C/C	C/C	T/C
NPC 081	F	Control	G/G	G/G	C/C	C/C	C/C
NPC 082	F	Control	G/G	G/G	C/C	C/C	C/C
NPC 083	F	Control	G/A	G/G	C/C	C/C	C/C
NPC 084	F	Control	G/G	G/G	T/C	C/C	C/C
NPC 085	F	Control	G/G	G/G	C/C	C/C	C/C
NPC 086	M	Control	G/A	G/G	C/C	C/C	T/C
NPC 087	F	Control	G/A	G/G	T/C	C/C	T/C
NPC 088	F	Control	G/G	G/G	C/C	C/C	T/C
NPC 089	F	Control	G/A	G/G	C/C	C/C	C/C
NPC 090	M	Control	G/G	G/G	C/C	C/C	T/C
NPC 091	M	Control	G/G	G/G	C/C	C/C	C/C
NPC 092	M	Control	G/G	G/G	C/C	C/C	C/C
NPC 093	F	Control	G/A	G/G	C/C	C/C	C/C
NPC 094	F	Control	G/A	G/G	T/C	C/C	T/C
NPC 095	M	Control	G/G	G/G	C/C	C/C	C/C
NPC 096	M	Control	G/G	G/G	C/C	C/C	T/C
NPC 097	F	Control	G/G	G/G	C/C	C/C	T/C

NPC 098	F	Control	G/G	G/G	C/C	C/C	T/C
NPC 099	F	Control	G/G	G/G	C/C	C/C	C/C
NPC 100	F	Control	G/G	G/G	C/C	C/C	T/C
NPC 101	M	Control	G/G	G/G	C/C	C/C	T/C
NPC 102	M	Control	G/G	G/G	C/C	C/C	T/C
NPC 103	M	Control	G/A	G/G	C/C	C/C	C/C
NPC 104	M	Control	A/A	G/G	C/C	C/C	T/C
NPC 105	M	Control	G/A	G/G	C/C	C/C	T/T
NPC 106	M	Control	G/G	G/G	C/C	C/C	C/C
NPC 107	M	Control	G/G	G/G	T/C	C/G	T/C
NPC 108	F	Control	G/A	G/G	C/C	C/C	T/C
NPC 110	M	Control	G/A	G/G	C/C	C/C	T/T
NPC 111	M	Control	G/A	G/G	C/C	C/C	T/T
NPC 112	M	Control	G/A	G/G	C/C	C/C	T/C
NPC 113	M	Control	G/A	G/G	C/C	C/C	T/C
NPC 114	M	Control	G/G	G/G	C/C	C/C	T/C
NPC 115	F	Control	G/A	G/G	C/C	C/G	T/C
NPC 116	M	Control	G/G	G/G	C/C	C/C	T/C
NPC 117	M	Control	G/G	G/G	T/C	C/C	T/C
NPC 118	M	Control	G/A	G/G	T/C	C/C	C/C
NPC 119	M	Control	G/A	G/G	C/C	C/C	C/C
NPF 001	F	Index case	A/A	G/G	C/C	C/C	T/T
NPF 010	M	Index case	G/G	G/G	C/C	C/C	C/C
NPF 013	M	Index case	G/G	G/G	C/C	C/C	T/C
NPF 017	M	Index case	G/G	G/C	C/C	C/C	C/C
NPF 018	F	Index case	G/A	G/G	T/C	C/C	C/C
NPF 022	F	Index case	G/A	G/G	C/C	C/C	C/C
NPF 024	F	Index case	G/G	G/G	C/C	C/C	T/C
NPF 026	M	Index case	G/G	G/G	C/C	C/C	T/C
NPF 028	F	Index case	G/A	G/G	C/C	C/C	C/C
NPF 031	M	Index case	G/A	G/G	C/C	C/C	C/C
NPF 034	M	Index case	G/G	G/G	C/C	C/C	T/C
NPF 036	M	Index case	G/A	G/G	C/C	C/C	T/C
NPF 038	M	Index case	G/G	G/G	C/C	C/C	C/C
NPF 042	M	Index case	G/G	G/G	T/C	C/C	C/C
NPF 044	F	Index case	G/G	G/G	T/C	C/C	T/C
NPF 045	M	Index case	G/G	G/G	C/C	C/C	T/T
NPF 047	F	Index case	G/G	G/G	C/C	C/C	C/C
NPF 049	F	Index case	G/G	G/G	C/C	C/C	T/T
NPF 053	M	Index case	G/G	G/G	C/C	C/C	C/C
NPF 054	F	Index case	G/A	G/G	C/C	C/C	C/C
NPF 059	F	Index case	G/A	G/G	C/C	C/C	C/C
NPF 060	M	Index case	G/A	G/G	T/C	C/C	T/T
NPF 062	M	Index case	G/A	G/G	C/C	C/C	T/C
NPF 065	M	Index case	G/G	G/G	C/C	C/C	T/C
NPF 066	F	Index case	G/G	G/G	C/C	C/C	C/C
NPF 069	F	Index case	G/G	G/G	T/C	C/C	C/C

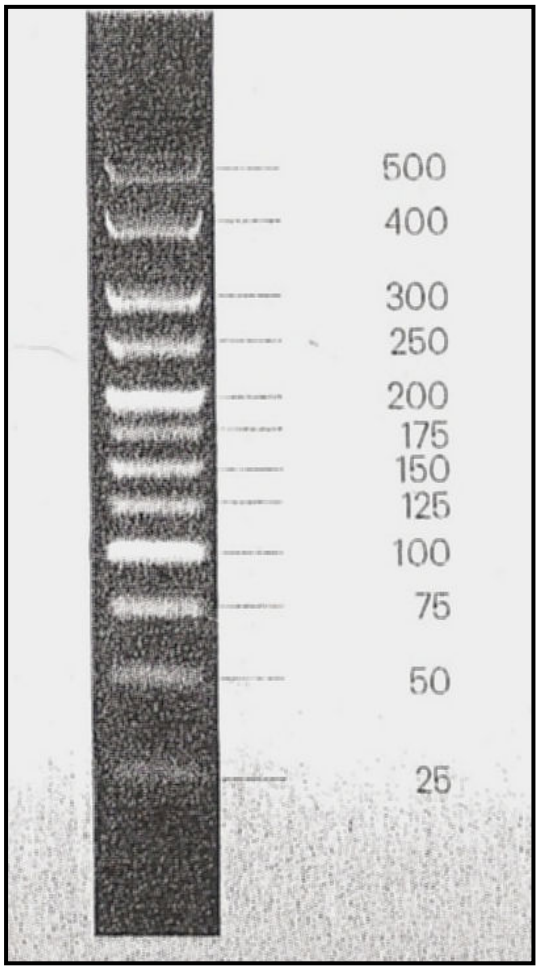
NPF 072	F	Index case	G/G	G/C	T/C	C/C	T/T
NPF 078	F	Index case	G/G	G/G	C/C	C/C	T/C
NPF 082	M	Index case	G/G	G/G	C/C	C/C	T/T
NPF 083	F	Index case	G/G	G/G	C/C	C/C	T/T
NPF 085	M	Index case	G/A	G/G	C/C	C/C	C/C
NPF 086	F	Index case	G/G	G/G	C/C	C/C	C/C
NPF 089	F	Index case	G/G	G/G	T/C	C/G	T/T
NPF 091	F	Index case	G/A	G/G	T/C	C/C	C/C
NPF 093	M	Index case	G/A	G/G	C/C	C/C	T/C
NPF 094	F	Index case	G/G	G/G	C/C	C/C	T/C
NPF 095	M	Index case	G/A	G/G	C/C	C/C	C/C
NPF 098	M	Index case	G/G	G/G	C/C	C/C	C/C
NPF 099	F	Index case	G/G	G/G	C/C	C/C	C/C
NPF 102	F	Index case	G/G	G/G	C/C	C/C	C/C
NPF 104	F	Index case	G/G	G/G	C/C	C/C	T/C
NPF 106	F	Index case	G/A	G/G	C/C	C/C	T/C
NPF 108	F	Index case	G/G	G/G	C/C	C/C	C/C
NPF 110	M	Index case	G/G	G/G	C/C	C/C	T/C
NPF 112	M	Index case	G/A	G/G	C/C	C/C	T/C

APPENDIX C

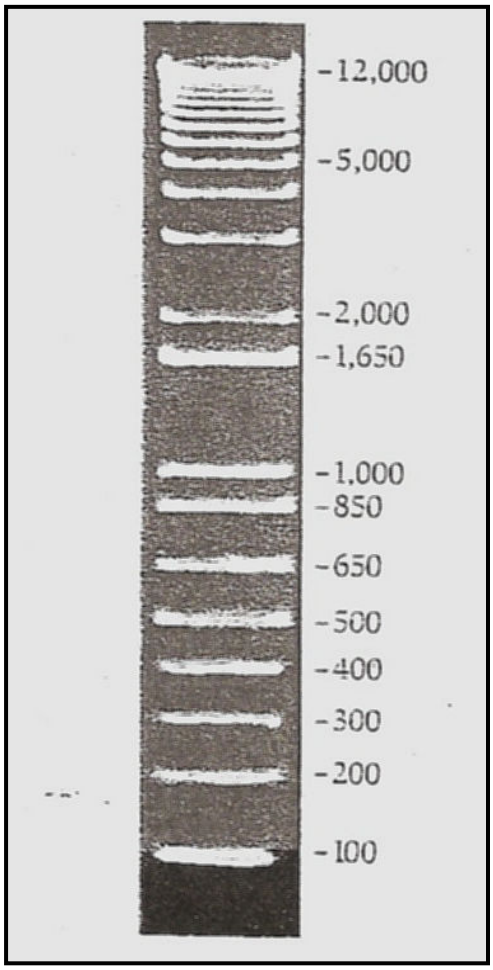
DNA LADDERS AND PROTOCOLS

1) DNA Ladders used (Band sizes shown on agarose gels in base pairs)

HyperLadder V (Bioline)



1 KB Plus DNA Ladder (Invitrogen)



2) Flexigene DNA Extraction Kit Protocol

- **Reagents required per 8 ml blood sample:**

Reagent	Volume (ml)
Buffer FG1	20.0
Buffer FG2/ protease	4.0
100% isopropanol	4.0
70% ethanol	4.0
Buffer FG3	0.8

- **Reagent Preparation:**

i. Qiagen protease

- resuspend protease in 1.4 ml Buffer FG3

ii. FG2/Qiagen protease mix

- to be prepared an hour before extraction
- 4 ml FG2 and 40 µl protease

- **Protocol:**

- Add 20 ml buffer FG1 into a 50 ml NUNC tube
- Add 8 ml whole blood and invert 5 times
- Centrifuge for 5 minutes at 3000 rpm
- Discard the supernatant and invert the tube on paper for 2 minutes
- Add 4 ml FG2/protease and vortex until completely homogenised
- If traces of jelly-like consistency are still visible add 1 ml FG2 and vortex again
- Invert tube 3 times
- Incubate at 65^0C for 10 minutes
- Colour change from red to olive green should occur indicating complete protein digestion, leave for longer if colour change does not occur

- Add 4 ml 100% isopropanol and mix thoroughly by inversion until DNA becomes visible in threads
- Centrifuge for 3–5 minutes at 3000 rpm
- Discard the supernatant and invert the tube on paper for 2 minutes
- Add 4 ml 70% ethanol and vortex for 5 seconds
- Centrifuge for 3-5 minutes at 3000 rpm
- Discard the supernatant and invert the tube on paper for 2 minutes
- Air-dry the pellet for about 5 minutes
- Add 0.8 ml FG3 and vortex for 5 seconds at low speed
- Incubate for 1 hour at 65^0C or overnight at room temperature

3) PCR reagents for a standard PCR reaction

Reagent	Volume (µl)
DNA (100 ng/µl)	1.0
AmpliTaq Gold (5 units/µl; Applied Biosystems, Roche)	0.2
10x PCR Gold Buffer (Applied Biosystems, Roche)	2.5
$MgCl_2$ (25 mM; Applied Biosystems, Roche)	2.0
dNTP mix (1.25 mM; see Appendix D)	2.5
Forward primer (10 µM; see Appendix D for primer dilutions)	1.0
Reverse primer (10 µM; see Appendix D for primer dilutions)	1.0
ddH_20	14.8
Total	**25.0**

4) PCR reagents for the ARMS PCR reaction (rs362691)

Reagent	Volume (µl)
DNA (100 ng/µl)	1.0
AmpliTaq Gold (5 units/µl; Applied Biosystems, Roche)	0.2
10x PCR Gold Buffer (Applied Biosystems, Roche)	2.5
$MgCl_2$ (25 mM; Applied Biosystems, Roche)	2.0
dNTP mix (1.25 mM; see Appendix D)	2.5
Forward primer (either detecting the "mutant" or "wild-type" allele) (10 µM; see Appendix D for primer dilutions)	1.0
Reverse primer (10 µM; see Appendix D for primer dilutions)	1.0
Control Forward primer (10 µM; see Appendix D for primer dilutions)	1.0
Control Reverse primer (10 µM; see Appendix D for primer dilutions)	1.0
Spermidine (1/40 dilution of a 0.1 M stock; see Appendix D)	2.5
ddH_20	10.3
Total	**25.0**

5) PCR product clean-up using the NucleoSpin® Extract II kit (Macherey-Nagel)

- Adjust PCR product's volume to 50 µl with TE buffer (pH=7.5) (30 µl TE + 20 µl PCR product)
- Mix 2 volumes of NT buffer with 1 volume of PCR product (i.e. 100 µl NT + 50 µl PCR product)
- Place column in collection tube and load sample onto column
- Centrifuge for 1 minute at 14 000 rpm
- Discard flow-through and place column back into collection tube
- Pipette 600 µl NT3 buffer onto column
- Centrifuge for 1 minute at 14 000 rpm
- Discard flow-through and place column back in to collection tube
- Centrifuge for 2 minutes at 14 000 rpm
- Discard flow-through

- Incubate columns for 2-5 minutes at 70^0C with the caps open
- Place columns into clean Eppendorf tubes
- Pipette 20 µl ddH_20 (volume must be equal to starting volume)
- Incubate samples for 1 minute at room temperature
- Centrifuge for 1 minute at 14 000 rpm
- Discard column

6) Cycle sequencing PCR reagents

Reagent	Volume (µl)
Cleaned PCR product	6
BigDye Terminator v3.1 (Applied Biosystems)	4
BigDye Terminator v3.1 5X sequencing buffer (Applied Biosystems)	2
Primer (forward or reverse; 10 µM; see Appendix D for dilutions)	1
ddH_20	7
Total	**20**

- **Cycle Sequencing Parameters**

Denaturation	96^0C	30 secs	25 cycles
Annealing	50^0C	15 secs	
Extension	60^0C	4 mins	
Hold	4^0C	∞	

7) Cycle sequencing product clean-up using the Qiagen DyeEx 2.0 spin kit

- Gently vortex each spin column to remove air bubbles
- Slightly loosen the column cap and snap off the bottom of each column tube
- Place column in a collection tube
- Centrifuge for 3 minutes at 3000 rpm
- Discard flow-through and place column in a clean collection tube
- Load entire cycle sequencing product onto the middle of the column
- Centrifuge for 3 minutes at 3000 rpm
- Discard column

Sample was then vacuum dried for 1 hour in the 5301 Concentrator (vacuum dryer) manufactured by Eppendorf.

8) Preparation for the ABI 3130 Genetic Analyzer

- Add 10 µl HiDi Formamide (Applied Biosystems) to each vacuum dried sample
- Load products into a 96-well sequencing plate
- Place plate in a 2720 thermal cycler for 2 minutes to denature the DNA
- Place plate on ice
- Load plate into the 3130 Genetic Analyzer (Applied Biosystems)
- Select Z_Seq_POP7_36_UltraSeq for sequences less than 500 bp or Z_Seq_POP7_36_RapidSeq for longer sequences
- Select analysis protocol 3130pop7_BDTV3_kb
- Follow the prompts on the screen and enter the identity of each product in each well. Sequencing can then begin

APPENDIX D

REAGENTS

dNTP mix (Bioline)

A 1 ml working solution containing 1.25 mM of each dNTP was made as follows:

12.5 µl of 100 mM dATP + 112.5 µl ddH_20
12.5 µl of 100 mM dCTP + 112.5 µl ddH_20
12.5 µl of 100 mM dGTP + 112.5 µl ddH_20
12.5 µl of 100 mM dTTP + 112.5 µl ddH_20

} Mix together and add 500 µl ddH_20

0.5 M EDTA (pH 8.0)

93.06 g EDTA dihydrate (Research Organics)
100 ml ddH_20
pH to 8 with 5 M NaOH pellets (SMM Chemicals)

Ficoll dye

25 g sucrose (Invitrogen Life Technologies)
5 ml 0.5 M EDTA
0.05 g bromophenol blue (Merck)
5 g Ficoll (Sigma-Aldrich)
Add ddH_20 to make a total volume of 50 ml

1 Kb Plus DNA ladder

250 µl 1 Kb Plus ladder (Invitrogen Life Technologies)
125 µl Ficoll dye (Sigma-Aldrich)
2.1 ml TE buffer

Primer dilutions

All primers were resuspended in a certain volume of ddH_20 as stated by the manufacturer (Inqaba Biotech) to form 100 µM stock solutions. 1 in 10 dilutions (10 µM) of each primer were made for PCR reactions:

- 10 µl of primer stock (100 µM)
- 90 µl ddH_20
- All primer stocks and dilutions were stored at -20^0C

Spermidine

The following protocol was used to make a 0.1 M stock solution:

1 g spermidine (254.63 g/mol; Merck)

40 ml ddH_20

1 in 40 Spermidine dilution

1 µl of a 0.1 M spermidine stock solution

39 µl ddH_20

10 x TBE buffer

432 g Tris (Roche)

220 g Boric acid (Promega)

29.7 g EDTA dihydrate (Research Organics)

Add ddH_20 and make up to a final volume of 4 *l*

Autoclave before use

1M Tris-HCl

12.11 g Tris (Roche)

pH to 8.0 and make up to 100 ml with ddH_20

1 x TE buffer (pH 7.5)

1 ml 1 M Tris-HCl

200 µl 0.5 M EDTA

Make up to 100 ml with ddH_20

3% Agarose gel

12 g Agarose (Hispanagar)

400 ml 1x TBE buffer

Heat in the microwave until fully dissolved

Once slightly cooled, add 12 µl of a 10 mg/µl ethidium bromide stock (Sigma)

Pour the gel mix into a gel tray

4% Agarose gel

16 g Agarose (Hispanagar)

400 ml 1x TBE buffer

Heat in the microwave until fully dissolved

Once slightly cooled, add 12 µl of a 10 mg/µl ethidium bromide stock (Sigma)

Pour the gel mix into a gel tray