DETERMINING THE MOLECULAR BASIS OF SPINAL MUSCULAR ATROPHY IN THE BLACK SOUTH AFRICAN POPULATION

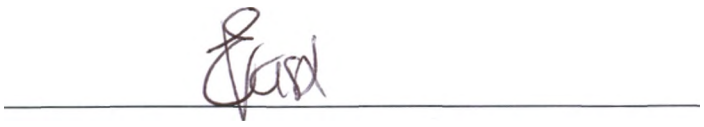
Elana Vorster

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

Johannesburg, 23 June 2017

DECLARATION

I, Elana Vorster, declare that this dissertation is my own, unaided work (refer to Appendix L for Turn-it-in plagiarism reports). It is being submitted for the Degree of Masters of Science in Medicine (MSc (Med)) in Human Genetics at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



Elana Vorster

Signed: 23rd day of June 2017 in Johannesburg

This dissertation is dedicated to my father-in-law, Piet Vorster and my dear friend, Eve Atwell, who do not allow their neuromuscular diseases to hold them back, they truly live life fully. Your courage is inspirational.

Spinal muscular atrophy

My microscope focusses on the stars,
flickering in the night sky of cosmic cells.
The tools of a half-scientist, half-Sangoma:
hiking boots, notepad and butterfly net.
I pray: Oh Ancestors! Let me inherit your wisdom...

The wind fasciculates through the limbs of your trees,
the bush veld is diagnosed with atrophy.
Neurons short circuit like cricket song
and fatigue pounces on its helpless prey -
paralyses your muscles, truncates your days.

Even though your lungs are ventilated,
the camp fire of your heart stutters on.
Surely, we are more than our neurology?
Constructed from more than dust and biology?

The answers escape me like whispers, like waves...

The insomniac birds experiment with lyrics
to be sung on a sun rise stage.
The sky mutates, hours rearrange
and the day's fate is chosen.

PRESENTATIONS ARISING FROM THIS RESEARCH PROJECT

- 10 April 2013: Division of Human Genetics - Project proposal presentation
- 17 August, 2015: South African Society of Human Genetics (SASHG) conference > oral presentation
  - **Title:** Determining the molecular basis of spinal muscular atrophy (SMA) in the black South African (SA) population
  - Won joint-second prize for oral presentation
- 16 March, 2016: Division of Human Genetics - Research update presentation
- 11 May, 2016: Division of Human Genetics - Final presentation of project
- 1 September, 2016: Health Sciences Research Day > oral presentation
  - **Title:** Determining the molecular basis of spinal muscular atrophy (SMA) in the black South African (SA) population

Additional project outcomes

- I supervised Miss Haseena Khan's Honours project on SMA titled: Investigating the molecular aetiology of silent carriers of spinal muscular atrophy in the black South African population.
- Article in preparation (to be published 2017) – Spinal muscular atrophy in the black South African population: a matter of rearrangement?
- In the past three years of working on this project, I have developed a keen interest in science communication. I wrote an article about my master's project and it was accepted as part of the 2014 Mail & Guardian Science Voices supplement. The article can be accessed from http://mg.co.za/article/2014-08-18-unique-local-genetic-problem-needs-a-unique-local-solution).
- I have also consequently developed a blog about my SMA research project which can be accessed from https://therusticlab.wordpress.com/.
- I have also advised MRC Holland on the improvement and development of a new version of the SALSA P021-A2 MLPA kit.

ABSTRACT

Spinal muscular atrophy (SMA) is an autosomal recessive, neuromuscular disorder, characterised by muscle atrophy and impaired mobility. A homozygous deletion of exon 7 of the survival motor neuron 1 gene (*SMN1*) is the main cause of SMA in ~94% of patients worldwide but only accounts for 51% of South African (SA) black patients. *SMN1* and its highly homologous pseudogene, *SMN2,* are located in a complex duplicated region on chromosome 5q13. Unusual copy number variations (CNVs) have been reported in black patients, suggesting the presence of complex pathogenic rearrangements. The aim of this project was to investigate the disease mechanism in the black SA population further. MLPA (multiplex ligation-dependent amplification) testing was performed on 172 unrelated black individuals referred for SMA testing. Of these, 50 had a homozygous deletion of *SMN1*, exon 7; 50 had a homozygous deletion of *SMN2*, exon 7 and 72 had no homozygous deletions, but a strong clinical suspicion of SMA. Furthermore, 122 black individuals negative for SMA were tested. For comparison, 68 white individuals (30 with a homozygous deletion of *SMN1;* 8 with a homozygous deletion of *SMN2*, exon 7 and 30 negative for SMA) were tested. No clear pathogenic CNV pattern was identified in black patients. Only 8.3% (6/72) of patients with a strong clinical suspicion had a heterozygous deletion of *SMN1*, exon 7 which is lower than previous SA reports of 69.5%. Multiple (>2) copies of *SMN1*, exon 7 were observed in 50.8% (62/122) of black negative controls which could mask the presence of silent carriers and potential pathogenic CNVs. This study re-emphasises the hypervariability and the limited understanding of CNVs of the SMN region in black populations. While MLPA has been shown to be an appropriate technique to identify SMA carriers in the white population, it has limited clinical utility in the black SA population.

ACKNOWLEDGEMENTS

I would like to thank:

- All the patients whose DNA samples were included in this study. Research on SMA in black patients will continue and I sincerely hope our laboratory will be able to design a more accurate diagnostic test in the future.
- Prof Amanda Krause for sharing her wisdom, experience and ideas in such a gracious way. I truly admire your passion and dedication.
- Fahmida Essop for her remarkable friendship and incredible emotional support. You have been a wonderful mentor to me for the past 10 years.
- Prof John Rodda for his ongoing collaboration and his help in identifying appropriate patients for this study.
- The National Health Laboratory Service (NHLS) for allowing me the use of their facilities for this project.
- My NHLS molecular diagnostic colleagues for their continuous support and for sharing laughs and takeaways to keep me motivated.
- Andrew May for sharing his knowledge on statistical analysis, writing using Word and plagiarism checks. I truly appreciate your patience, time and wisdom.
- The NHLS Research Trust and the WITS Faculty Research Council for their generous funding of this project.
- MRC Holland who has generously sponsored 800 free MLPA probe mix reactions for this project. I am extremely grateful for their support and interest.
- My parents: Dad, Mams and Ma Elmare for their infinite support, understanding and supernatural motivation.
- My dear friends, Yolande and Eve and my siblings, Jowan and Anja, for sharing my passion for science and fuelling the fires of curiosity. I hope to have many more philosophical discussions with you about our role in the universe!
- My husband, Cobus, for making this degree possible by being my cheerleader, my superhero, my psychologist, my IT guy and my audience. Thank you for your enduring patience.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND SYMBOLS

A	Adenine
ACMG	The American College of Medical Genetics
ACOG	The American College of Obstetricians and Gynecologists
AHGVP	African Human Genome Variation project
ANOVA	Analysis Of Variance
BDNF	Brain-derived Neurotrophic Factor
bp	base pairs
C	Cytosine
CCD	charge-coupled device
cm	centimetre
CNS	central nervous system
CNV	copy number variation
Ctrl	control
ddH_2O	deionised, distilled water
ddNTPs	dideoxyribonucleotide trIphosphates
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotide triphosphates
DQ	dosage quotient
EDTA	ethylene-diamine-tetra-acetate
EtBr	ethidium bromide
Etc.	et cetera
F	forward primer
F	ANOVA test statistic
g	grams
G	Guanine
GTF2H2	General Transcription Factor IIH Subunit 2
H	Kruskall-Wallis test statistic unit
ID	intellectual disability
IHC	inter-homologous chromosomes
INHC	inter-non-homologous chromosomes

kb	kilobases
kDa	kilo Dalton
L	litre
MANOVA	Multivariate Analysis Of Variance
Mg^{2+}	Magnesium ion
$MgCl_2$	Magnesium chloride
min	minutes
ml	millilitre
MLPA	multiplex ligation-dependent probe amplification
mm	millimeter
mM	millimolar
mRNA	messenger RNA
NaCl	Sodium Chloride
NAHR	Non-allelic homologous recombination
ND	nanodrop
ng	nanograms
NAIP	NLR Family Apoptosis Inhibitory Protein
NGRL	National Genetics Reference Laboratory
NHLS	National Health Laboratory Service
nm	nanometer
OMIM	Online Mendelian Inheritance in Man
p	page number
PCR	polymerase chain reaction
PLS3	Plastin 3
pmol	pico mol
qPCR	quantitative/real-time PCR
qRT-PCR	reverse transcription qPCR
R	reverse primer
RAD17	cell cycle checkpoint protein
RNA	ribonucleic acid
rpm	revolutions per minute
SAHGP	The Southern African Human Genome project
SDS	Sodium Dodecyl Sulphate

SERF1A/B	Small EDRK-Rich Factor 1A/B
SMA	spinal muscular atrophy
SMN	Survival Motor Neuron
SMN1	Survival Motor Neuron 1 gene
SMN2	Survival Motor Neuron 2 gene
s	seconds
T	Thymine
TBE	Tris Borate EDTA
TE	Tris-EDTA
Tris	Tris (hydroxymethyl) methylamine
U	units
µl	micro litre
V	volts
vs	versus
UV	ultraviolet radiation
X^2	Chi-squared test statistic unit

LIST OF DEFINITIONS

U/U^B	Subject Group A: **b**lack patients who previously tested negative for the homozygous deletion of *SMN1*, exon 7 and who are clinically suggestive of SMA. These patients have an assumed "**U**nidentified mutation/**U**nidentified mutation" genotype and form the main focus of this study.
N/N^B	Subject Group B: these **b**lack individuals do not have SMA and act as a black negative control group. They have an assumed "**N**egative/**N**egative" unaffected genotype.
N/N^W	Subject Group C: these **w**hite individuals do not have SMA and act as a white negative control group. They have an assumed "**N**egative/**N**egative" unaffected genotype.
$M_1/M_1{}^B$	Subject Group D: these **b**lack individuals have the common homozygous deletion of *SMN1*, exon 7 (mutation 1) and are therefore confirmed to be affected with SMA. They have an assumed "**M**utation **1**/**M**utation **1**" genotype.
$M_1/M_1{}^W$	Subject Group E: these **w**hite individuals have the common homozygous deletion of *SMN1*, exon 7 (mutation 1) and are therefore confirmed to be affected with SMA. They have a "**M**utation **1**/**M**utation **1**" genotype.
$M_2/M_2{}^B$	Subject Group F: these **b**lack individuals have a homozygous deletion of *SMN2*, exon 7 (mutation 2). This deletion is not thought to cause SMA directly. They have a "**M**utation **2**/**M**utation **2**" genotype.
$M_2/M_2{}^W$	Subject Group G: these **w**hite individuals have a homozygous deletion of *SMN2*, exon 7 (mutation 2). This deletion is not thought to cause SMA directly. They have a "**M**utation **2**/**M**utation **2**" genotype.
The Division	The National Health Laboratory Service (NHLS) and University of the Witwatersrand (WITS), School of Pathology, Division of Human Genetics, Johannesburg, South Africa

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 1 INTRODUCTION

Spinal muscular atrophy (SMA) is an autosomal recessive neuromuscular disorder which is as common as albinism in the black South African (SA) population. SMA is caused by mutations in the survival motor neuron 1 gene (*SMN1*) which is located in a complex region on chromosome 5q13. A common homozygous deletion of *SMN1*, exon 7 accounts for the majority of SMA cases worldwide (~94%), but has been observed at a much lower frequency in black SA cases (51%).

The main aim of this project was to investigate copy number variation (CNV) using multiplex ligation dependent probe amplification (MLPA) to try and identify potential pathogenic CNVs which could contribute to the disease mechanism of SMA in the black SA population. Chapter one of this dissertation will explain the clinical presentation, disease epidemiology, genetics and disease mechanism of SMA. Furthermore, the aims and objectives of this project will be introduced.

Please note that all mutations and variations will be described in this dissertation using the nomenclature guidelines of the human genome variation society (HGVS) (den Dunnen *et al.*, 2016).

1.1 Clinical presentation of SMA

SMA is characterised by the progressive degeneration of anterior horn cells (lower motor neurons) of the spinal cord, causing symmetrical muscle atrophy, weakness and paralysis. Historically, SMA was categorised into four clinical subtypes (SMA I – IV), ranging in severity, maximum muscle activity achieved and age of onset, although it has been suggested that the SMA phenotype rather spans a continuum (Prior *et al.*, 2004). Pearn *et al.* recommended the importance of interpreting the phenotype and clinical outcome of affected patients within the context of their family history (Pearn *et al.*, 1973).

SMA type I (previously known as Werdnig-Hoffman syndrome) is the most severe form, with onset usually at birth or before six months with an average lifespan of two years. Pearn *et al.* showed that the age of onset is highly correlated with the lifespan

of an affected child and that the acute infantile type is a distinctive sub-type of SMA (Pearn *et al.*, 1973). A further subdivision of SMA type I was suggested by Thomas and Dubowitz who showed that an onset of disease before two months is correlated with a distinct earlier age of death (Thomas and Dubowitz, 1994). An additional Type 0 with prenatal onset has also been proposed (MacLeod *et al.*, 1999).

SMA type II is an intermediate form with an onset between six and eighteen months with survival beyond four years and was first described by Fried and Emery. Patients with SMA type II are able to sit but not to walk (Fried and Emery, 1971).

SMA type III is a mild form with onset after twelve months and a relatively normal lifespan. It was first described by Kugelberg and Welander. Patients with SMA type III are able to sit and walk, but have muscle weakness which might lead to a loss of mobility in later years (Kugelberg and Welander, 1956).

SMA type IV is the mildest form with adult onset and a normal lifespan and was first described by Pearn *et al.* The median age of onset was 35 years and patients lost some mobility with most patients needing assistance with walking 20 years after initial diagnosis (Pearn *et al.*, 1978). Table 1.1 compares the clinical symptoms of the different types of SMA.

Table 1.1 Summary of clinical findings of SMA subtypes

Phenotype	Age of Onset	Life Span	Motor Milestones	Other Findings
SMA 0	Prenatal	2-6 months	None achieved	
SMA I	<6 months	Most often ≤2 years, but may live longer	Do not walk; sit with support only	- Mild joint contractures - Normal or minimal facial weakness - Variable suck and swallow difficulties - Muscle twitching/ fasciculations
SMA II	6-18 Months	70% alive at age 25 years	- Do not walk - independent sitting when placed	Postural tremor of fingers
SMA III	>12 months	Normal	Independent ambulation	Muscle weakness
SMA IV	Adulthood	Normal	Normal	Muscle weakness

Adapted from Butchbach and Burghes, 2004.

1.1.1 SMA in SA

A study performed in Durban, Kwazulu-Natal suggested that the clinical presentation of SMA in black SA patients differs from worldwide reports with more frequent involvement of facial muscles in the severe infantile form leading to expressionless facies. In addition they observed a 1:1.7 ratio of affected female to male patients, although this could have been due to a small sample size (n=45) (Moosa and Dawood, 1990). These results were challenged by a group situated in Cape Town who reported that black patients with SMA (n=12) did not differ genetically or phenotypically from international reports (Wilmshurst *et al.*, 2002).

Moosa and Dawood also reported a paucity of family history. Only 9% of patients reported a positive family history suggesting either a reluctance to share family information due to cultural reasons or the possibility that SMA could be more sporadic in the black SA population (Moosa and Dawood, 1990).

1.1.2 Disease epidemiology

SMA has been reported to be the second most common autosomal recessive disorder in Caucasian individuals after cystic fibrosis. The predicted birth incidence of SMA varies between 1 in 6 000 and 1 in 10 000 with a carrier frequency estimated at 1 in 40 to 1 in 60 worldwide (Hendrickson *et al.*, 2009). In a previous study performed at the Division of Human Genetics, molecular diagnostic laboratory, National Health Laboratory Service Johannesburg (NHLS) and the University of the Witwatersrand (WITS), henceforth referred to as "the Division", the birth incidence of SMA in black SA patients was estimated to be much higher at 1 in 3574. This indicates that SMA may have a higher birth incidence than albinism (birth incidence: 1 in 3900) in the black SA population. The carrier rate was estimated at 1/50 (Labrum *et al.*, 2007).

1.2 The genetics of SMA

1.2.1 The structure of the SMN region

The SMN1 and SMN2 genes are located in a large 500 kb inverted duplicated region on chromosome 5q13. The region consists of multiple pseudogenes (Selig *et al.*, 1995), repetitive sequences (Bürglen *et al.*, 1996) and retrotransposon-like elements

(Francis *et al.*, 1995). As a result of the complexity and hypervariability of the *SMN* region, there is no current complete and accurate map of the *SMN* region. Figure 1.1 shows a proposed map of the SMN region constructed using current map information from the Ensembl genome browser (Ensembl genome browser, 2016).

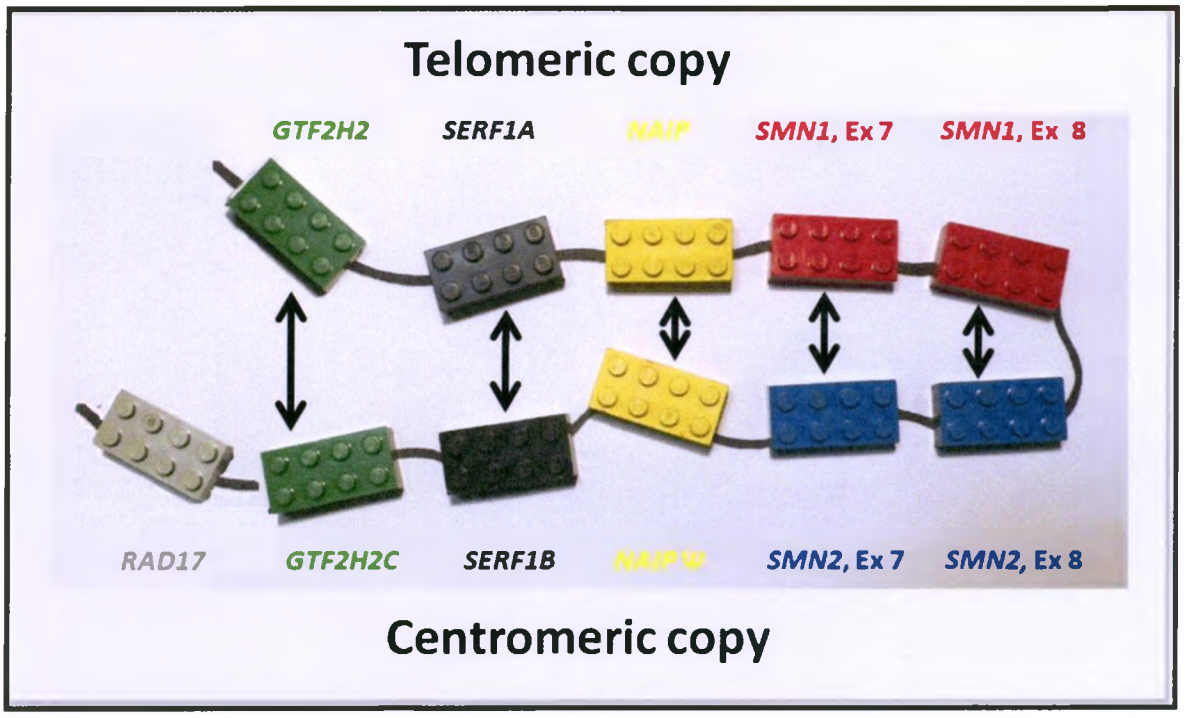


Figure 1.1: Proposed order of genes in the SMN region on chromosome 5q13

Differently coloured Lego blocks represent genes in the telomeric region and their corresponding pseudogene in the centromeric region. Arrows indicate regions with similar sequence homology where potential crossovers could occur, resulting in CNVs. Please note: only *SMN1/2*, exons 7 and 8 have been included. Full SMN1/2 genes are not shown.

SMA is caused by mutations within the SMN1 gene. Mutations within the SMN2 gene and other genes located in the SMN region, *NAIP*, *SERF1A*, *RAD17* and *GTF2H2* are not thought to cause SMA directly, but could modify disease severity (Jedrzejowska *et al.*, 2008; Noguchi *et al.*, 2016; Schwartz *et al.*, 1997; Theodorou *et al.*, 2015).

1.2.2 The SMN1 and SMN2 genes

The survival motor neuron 1 gene (*SMN1*) was discovered by Lefebvre *et al.* in 1995, when his team identified *SMN1* deletions and disruptions in 226 of the 229 patients with SMA investigated (Lefebvre *et al.*, 1995). *SMN1* (OMIM #600354; NCBI gene ID:

6606), also known as telomeric *SMN* and a highly homologous pseudogene, *SMN2*, (OMIM #601627; NCBI gene ID: 6607) also known as centromeric *SMN*, consist of nearly identical ~28 kb sequences, each containing nine exons (historically numbered: exons 1, 2a, 2b, 3-8) and each encoding the 294 amino acid survival motor neuron (SMN) protein (NCBI, 2016; OMIM, 2016).

SMN1 and *SMN2* differ in only five nucleotides of sequence, with two nucleotide differences located in exons 7 and 8 and three nucleotide differences located in introns 6 and 7 (SMN1, SMN2; NCBI gene ID: 6606, 6607) (NCBI, 2016). The critical difference between *SMN1* and *SMN2* is a silent C to T transition at cDNA position 840 (c.840C>T) in *SMN2*, resulting in the exclusion of exon 7 during splicing and causing the majority of *SMN2* transcripts to be truncated and unstable (Lorson *et al.*, 1999; Monani *et al.*, 1999).

1.2.2.1 The SMN1 gene

A homozygous deletion of *SMN1*, exon 7 is reported to cause SMA in ~94% of patients with SMA worldwide (Hendrickson *et al.*, 2009). In contrast, only 51% of SA SMA cases have been reported to be caused by a homozygous deletion of *SMN1*, exon 7 (Labrum *et al.*, 2007, Stevens *et al.*, 1999). An *SMN1* deletion in conjunction with a second mutation, results in a compound heterozygote pattern (two different mutations, one on each chromosome) and accounts for an additional 2-5% of patients with SMA worldwide (Wirth, 2000). The heterozygous deletion of *SMN1*, exon 7 rate in black SA patients with SMA who tested negative for the homozygous *SMN1*, exon 7 deletion, was reported to be much higher at 69.5% (Labrum *et al.*, 2007) and will be further discussed in section 1.3.1.

Currently 101 mutations in *SMN1* are reported in the Human Genome Mutation Database (HGMD) including missense, nonsense, small intragenic deletions and insertions and splice mutations (Human Genome Mutation Database, 2016).

Diagnostic testing for deletions of *SMN1*, exon 7

Polymerase chain reaction (PCR) and restriction enzyme digestion analysis was originally used as the main diagnostic assay worldwide to detect homozygous

deletions of *SMN1*, exon 7. This assay involves amplifying *SMN1*, exon 7 and *SMN2*, exon 7 simultaneously with the same set of primers on conventional PCR analysis, followed by restriction enzyme digestion to distinguish between *SMN1* and *SMN2* at the c.840C>T critical difference in *SMN2* (van der Steege *et al.*, 1996). This assay is limited as it is not quantitative and cannot detect heterozygous deletions of *SMN1*, exon 7 (de Souza Godinho *et al.*, 2012). Figure 1.2 demonstrates the concept of the current restriction enzyme based diagnostic assay offered at the Division.

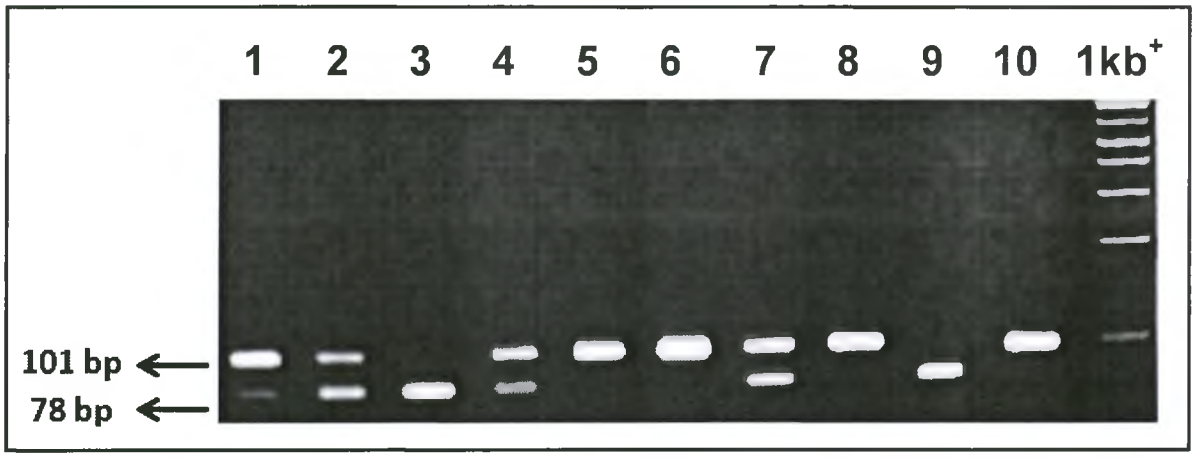


Figure 1.2: SMA diagnostic assay offered at the Division of Human Genetics

This procedure involves simultaneous PCR amplification of exon 7 of the *SMN1* and *SMN2* genes followed by restriction enzyme digestion with the Hinf I enzyme, which cuts the *SMN1* gene at the c.840C>T critical position. The *SMN2* gene remains undigested. The resulting PCR products are separated on a 3% agarose gel using gel electrophoresis and the *SMN1* and *SMN2* genes can be distinguished by size. *SMN1* is represented by the undigested 101bp PCR product and *SMN2* is represented by the 78 and 23 bp digested PCR products. The 23 bp PCR product is too small to visualise on a 3% agarose gel but is not necessary for diagnosis. Lanes 1, 2, 4 and 7: negative for homozygous *SMN1/SMN2*, exon 7 deletions (lane 1 appears to have an unequal dosage of *SMN1* vs *SMN2* but the actual copy number cannot be established by gel electrophoresis). Lanes 3 and 9: homozygous deletion of *SMN2*, exon 7; Lanes 5, 6 and 8: homozygous deletion of *SMN1*, exon 7, confirming a diagnosis of SMA. Lane 10 is an uncut control (no enzyme added) and lane 11 contains the molecular weight marker ($1kb^+$).

In recent years, quantitative techniques such as MLPA and quantitative/real-time PCR (qPCR) have simplified the detection of heterozygous deletions of *SMN1*, exon 7 in Caucasian populations. MLPA and qPCR approaches use probes designed specifically to the critical c.840C>T difference in order to distinguish between *SMN1*

and *SMN2*. The copy number of these genes are determined by comparing the fluorescent signal of *SMN1* and *SMN2* to the fluorescent signal generated by internal reference regions (de Souza Godinho *et al.*, 2012; Maranda *et al.*, 2012). Currently, carrier testing for the heterozygous deletion of *SMN1*, exon 7 (telomeric) can be offered to white family members, at risk. Diagnostic testing for the heterozygous *SMN1*, exon 7 (telomeric) deletion is not offered routinely at the Division.

1.2.2.2 The SMN2 gene

Homozygous deletions of *SMN2*, exon 7 are thought to not be pathogenic (Schwartz *et al.*, 1997) but are commonly encountered in black SA patients referred for SMA testing (Labrum *et al.*, 2007; Stevens *et al.*, 1999). Section 3.1.3 discusses detailed statistics from an audit of SMA referrals, done at the Division and will expand on this observation. A number of studies have suggested that *SMN2* acts as a disease modifying gene as SMA disease severity is inversely correlated with *SMN2* copy number (Feldkötter *et al.*, 2002; McAndrew *et al.*, 1997; Wirth *et al.*, 1999). SMA has been reported to be heterogeneous within families. Jedrzejowska *et al.* (2008) reported three families with asymptomatic family members who had homozygous deletions of *SMN1*, exon 7 and three to five copies of *SMN2* which seemed to partly compensate for the lack of SMN protein produced from *SMN1*.

Mice have only one functional SMN gene and a homozygous deletion of this gene is lethal due to massive embryonic cell death. The introduction of an SMN2 human transgene rescues mice with homozygous *SMN* deletions from lethality and furthermore, an increased SMN2 transgene copy number is associated with a decreased severity of the disease (Hsieh-Li *et al.*, 2000; Monani *et al.*, 1999).

Section 1.2.3 further explains the transcription and translation of *SMN1* and *SMN2* and the mechanism by which multiple *SMN2* copies can decrease disease severity.

Gene conversion events resulting in an increased number of SMN2 gene copies have been suggested to be associated with a milder form of SMA such as SMA types II and III (Campbell *et al.*, 1997; Talbot *et al.*, 1997) and will be further explained in Section 1.2.2.3.

1.2.2.3 Gene conversion between *SMN1* and *SMN2*

Non-allelic homologous recombination (NAHR) involves the crossover of two homologous sequences which are not located at the same locus. NAHR is one of the many mechanisms utilised to repair double stranded breaks in DNA but it also plays a frequent role in the formation of chromosomal rearrangements. NAHR between two similar sequences could lead to the unidirectional transfer of a nucleotide sequence from a "donor" DNA strand to an "acceptor" DNA strand. Mismatched repair of the heteroduplex formed by this transfer could ultimately lead to gene conversion. Gene conversion events can take place on the same chromatid, between sister chromatids and even between homologous and non-homologous chromosomes (Chen *et al.*, 2010). Figure 1.3 showcases some of the different mechanisms of gene conversions.

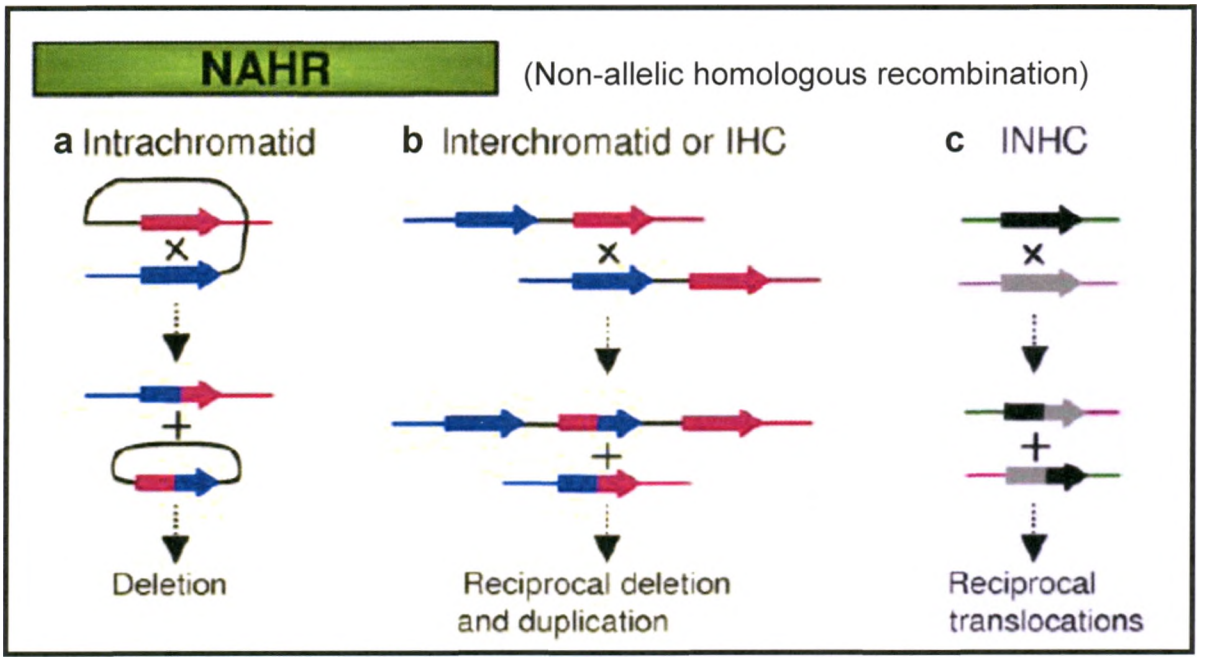


Figure 1.3: Mechanisms of gene conversion

This figure illustrates three mechanisms by which non-allelic homologous recombination could cause gene conversion and potential CNVs: **a.)** Homologous recombination on the same chromatid (e.g. between the SMN1 and SMN2 gene on a single chromatid); **b.)** homologous recombination between two different chromatids (e.g. between the SMN1 and SMN2 genes on different chromosomes); **c.)** recombination between inter-nonhomologous chromosomes (not a known mechanism in SMA). The two different colours (red and blue) represent two homologous sequences from different origins. Adapted from Chen et al. (2010).

Lefebre *et al.* (1995) reported 93% of *SMN1* deletions to span both exons 7 and 8 but in 5.6% of patients, a deletion of *SMN1*, exon 7 with the presence of exon 8 was identified. This finding can be explained by gene conversion events resulting in a hybrid SMN1 gene consisting of a converted exon 7 from *SMN2* and a true copy of exon 8 from *SMN1* (Hahnen *et al.*, 1996; Lefebvre *et al.*, 1995).

Figure 1.4 shows an example of how gene conversion can take place due to a crossover event between *SMN1* and *SMN2* on a single chromosome (intrachromatid NAHR).

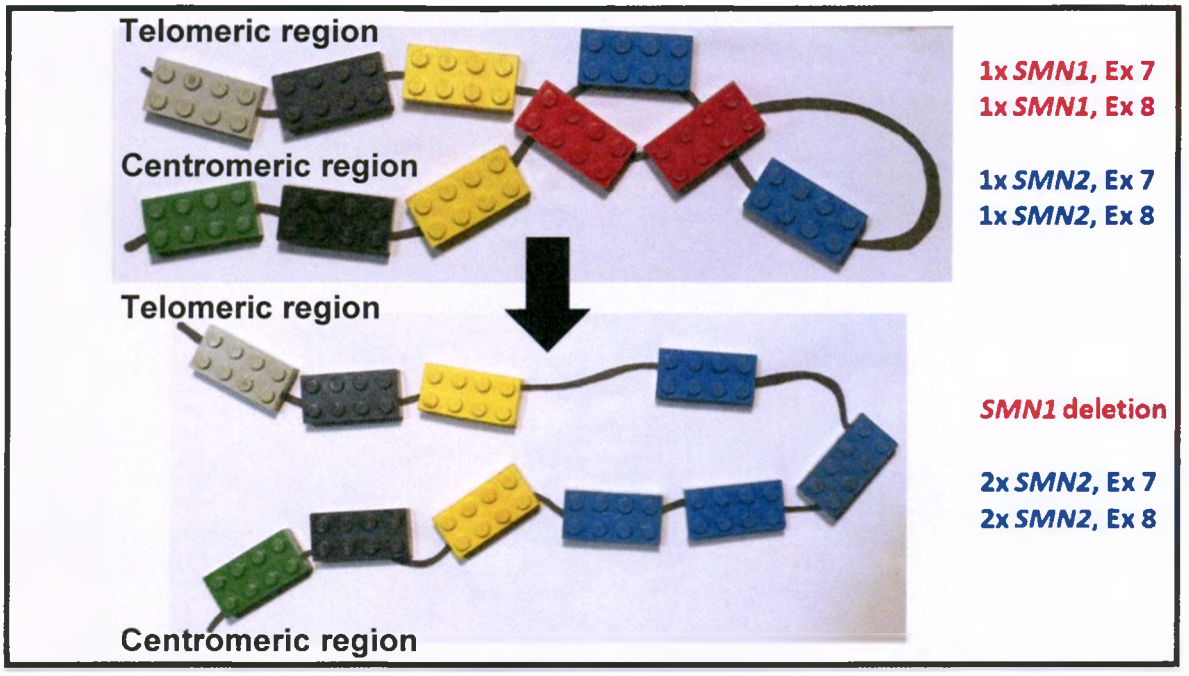


Figure 1.4: A crossover event leading to gene conversion between *SMN1* and *SMN2*. This illustration shows in a single chromosome, how a potential crossover between the centromeric & telomeric copies of SMN could take place due to the high homology of this region. *SMN1*, exons 7 & 8, indicated in red, are converted to *SMN2*, exons 7 & 8, indicated in blue, resulting in the loss of *SMN1* and in the production of multiple *SMN2* copies.

Recombinations between *SMN1* and *SMN2* could potentially interrupt a critical region of *SMN1*, leading to the loss of full-length functional SMN transcripts. Patients with conversions will therefore test positive for *SMN1*, exon 7 deletions. SMA type II and III patients have been shown to have *SMN1* conversions instead of deletions, resulting in a higher copy number of *SMN2*, which has been associated with a milder phenotype (Campbell *et al.*, 1997). Figure 1.5 further illustrates the mechanism by which different types of gene conversions can take place between *SMN1* and *SMN2*.

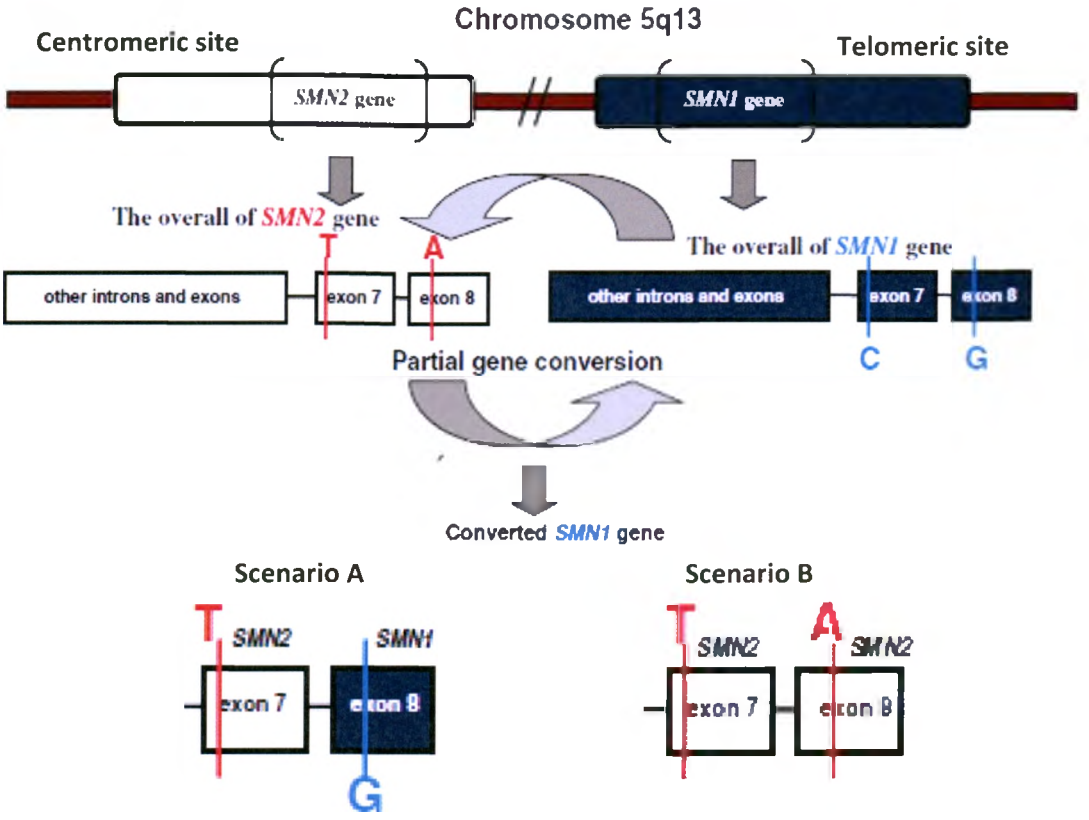


Figure 1.5: Possible gene conversions between *SMN1* and *SMN2*, exons 7 and 8. Although multiple conversions are possible between *SMN1* and *SMN2*, exons 7 and 8, only scenarios A and B will result in a functional loss of *SMN1*, exon 7, leading to SMA if combined with a second SMA-causing mutation. Both scenarios involve a conversion from the C allele at the critical position 840 in exon 7 of *SMN1* to a T allele. *SMN1* is essentially converted to *SMN2* and leads to a loss of full-length functional SMN transcripts. Adapted from Wang *et al.*, 2010.

Figure 1.6 illustrates how a deletion or gene conversion of *SMN1*, exon 7 could result in a loss of functional SMN transcripts, resulting in SMA.

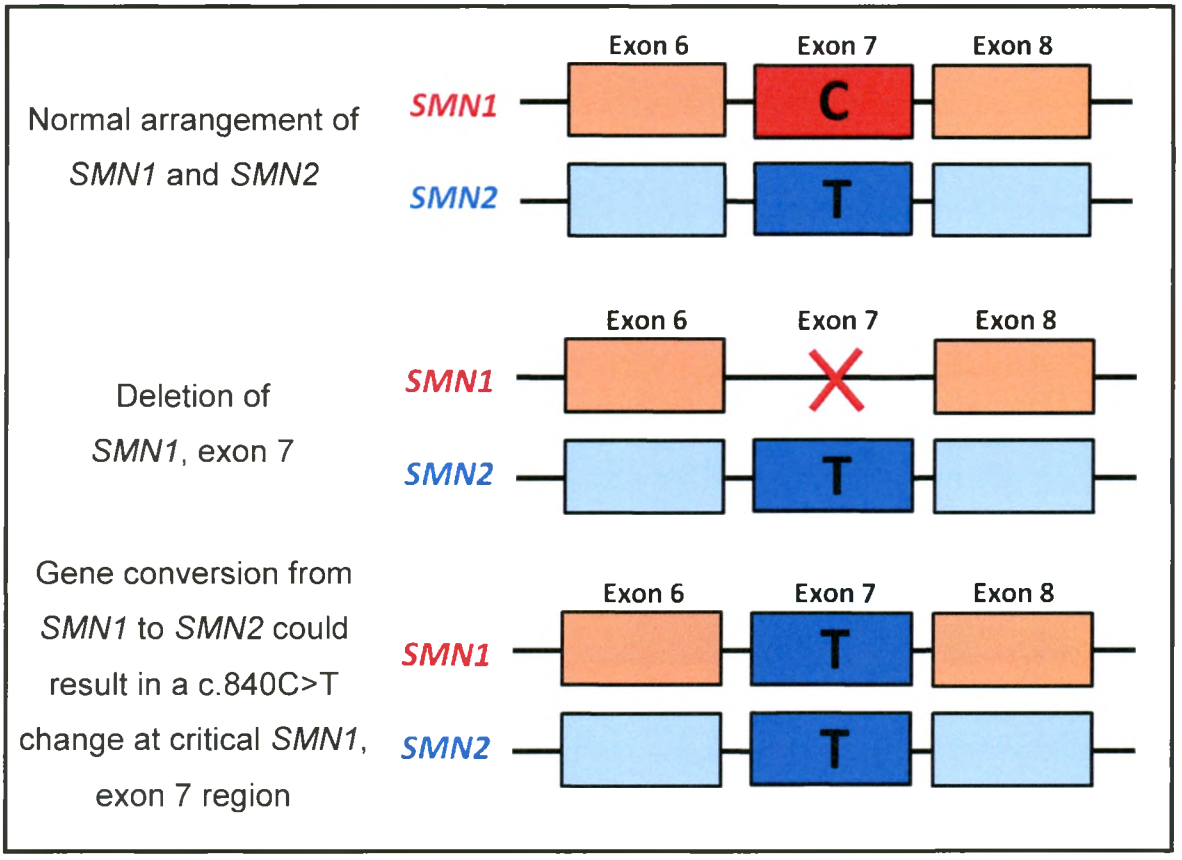


Figure 1.6: Loss of *SMN1*, exon 7 critical region due to deletion or gene conversion.

The c.840C allele is important for correct splicing of exon 7 and production of full-length SMN transcript. Once the C allele is deleted or converted, splicing of exon 7 is compromised and SMN transcripts become unstable and will be degraded rapidly.

1.2.2.4 Copy number variation of the SMN1 and SMN2 genes

Approximately 4% of American and Canadian individuals have been found to have two SMN1 gene copies on a single chromosome in addition to a chromosome with a deletion of *SMN1*, exon 7. These individuals are in effect silent carriers (also known as 2:0 carriers), since they have the ability to pass on a deletion chromosome to subsequent generations. Carrier testing is compromised since MLPA cannot distinguish between two copies of *SMN1*, one copy present on each chromosome or two copies of *SMN1* present on a single chromosome in conjunction with 0 copies on the second chromosome (McAndrew *et al.*, 1997). It is recommended that potential carriers with two copies of *SMN1* need to be analysed in a family context to try and clarify the phase of these CNVs and accurately assign carrier status. Figure 1.7 illustrates this finding.

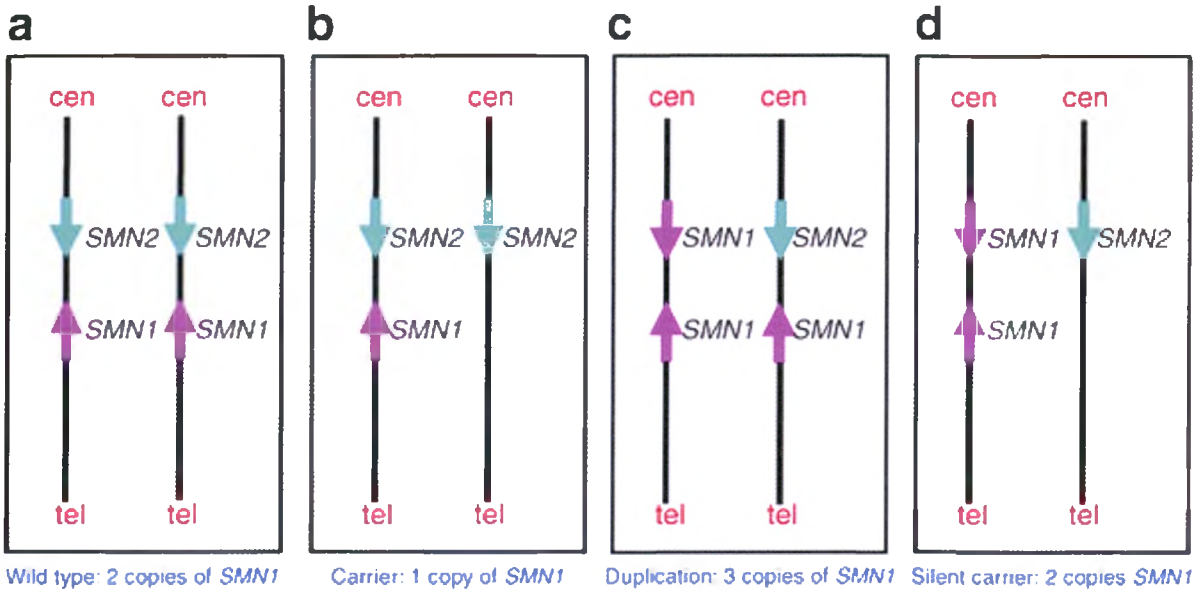


Figure 1.7: The outcomes of various *SMN1* and *SMN2* copy number combinations

This figure illustrates the clinical outcomes of various combinations of *SMN1* and *SMN2* alleles. The purple arrows represent *SMN1* and the blue arrows represent *SMN2*. *SMN2* formed due to a duplicated inversion of *SMN1*, hence the arrows in opposite orientation (Lefebvre *et al.*, 1995). Example (a) shows the classic configuration of unaffected individuals with two copies of *SMN1* and *SMN2*, with one copy of each gene on each chromosome (1:1 *SMN1* genotype). Example (b) shows an individual who has a heterozygous deletion of *SMN1* and is therefore a classic carrier of SMA (1:0). Example (c) shows an unaffected individual who has three copies of *SMN1*, with 2 copies on one chromosome and one copy on the second chromosome (2:1). The third *SMN1* copy could have arisen from a gene conversion from *SMN2* to *SMN1*. Example (d) shows an individual who has two copies of *SMN1* on one chromosome in conjunction with zero copies on the second chromosome (2:0). This individual is a silent carrier, since he/she can pass on a deletion chromosome to subsequent generations. Illustration modified from Luo *et al.*, (2014).

1.2.3 The SMN transcript, protein and function

The SMN protein is present in both the cytoplasm and nucleus of all cells, but is particularly abundant in motor neurons. The SMN protein's main function involves the assembly of small nuclear ribonucleoprotein (snRNP) complexes important for pre-messenger RNA splicing (Markowitz *et al.*, 2012). Mutations within the Tudor domain (amino acids 90-160) of the SMN protein reduce its ability to interact with other proteins and to oligomerise which may cause the biochemical defects observed in SMA (Bühler *et al.*, 1999).

Only 20% of the total full-length SMN (FL-SMN) transcript is produced from the *SMN2* gene, which partly compensates for the lack of FL-SMN transcript produced from *SMN1* in patients with SMA but does not produce sufficient SMN protein levels in motor neurons (Zheleznyakova *et al.*, 2011). Individuals with homozygous deletions of *SMN1* but with five copies of *SMN2* have been shown to have sufficient FL-SMN protein produced from *SMN2* (Prior *et al.*, 2004). See Figure 1.8 for a detailed illustration of SMN protein production.

Many clinical trials are aimed at increasing the total full length transcripts of *SMN2* (Finkel *et al.*, 2012). Genetic therapies for SMA are further discussed in section 1.4.

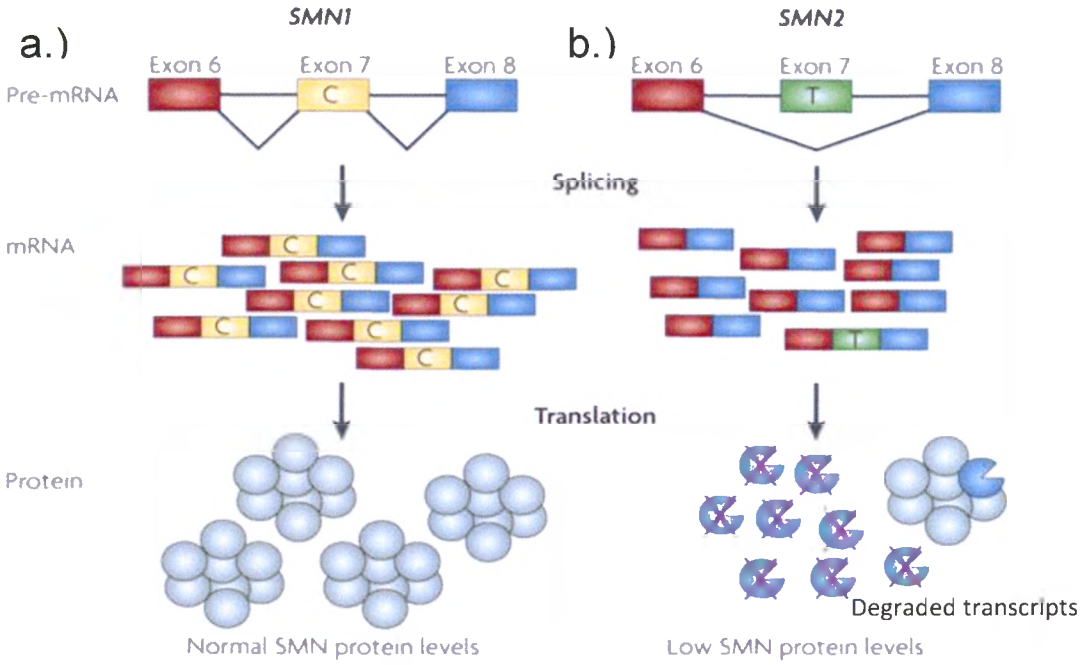


Figure 1.8: SMN protein production from the *SMN1* and *SMN2* genes

a.) *SMN1* produces full-length SMN transcripts containing exon 7 (FL-SMN) leading to the production of normal levels of the SMN protein. **b.)** A single base pair change from a C to a T allele in exon 7 of *SMN2* affects protein splicing and consequently, the majority (~80%) of *SMN2* transcripts lack exon 7 (SMN-Δ7), are unstable and are therefore degraded rapidly. A small amount (~20%) of FL-SMN is produced from *SMN2* but this amount is usually not enough to compensate for the lack of FL-SMN transcripts produced from *SMN1* in patients with SMA. Illustration modified from Burghes and Beattie (2009).

1.2.4 Other genes and variants associated with SMA

It has been suggested that the heterogeneity seen among different family members with similar genetic profiles and different disease severities, could potentially be due to the effect of modifier genes in addition to *SMN2* (Jedrzejowska *et al.*, 2009). Other genes located in the duplicated SMN region at chromosome 5q13.2 include *NAIP*, *GTF2H2, RAD17* and *SERF1A* and their multiple pseudocopies. Due to the lack of understanding about the physical structure and orientation of these genes in the SMN region, it is very difficult to completely understand the involvement of these genes in the SMA disease mechanism. Analysing multiple genes in the SMN region could potentially identify interactions between genes and their associated pseudogenes.

NOTE

Analysis of the SMN region proved to be very complicated and for the sake of simplicity and clarity, it is recommended that focus should be placed on the sections focussed on the functional telomeric SMN region including *SMN1*, exons 7, 8 and *NAIP* (1.2.2-1.2.3 and 1.2.4.1). Neighbouring genes provide some insight into the nature of large CNVs of the SMN region but no clear pathogenic CNVs of these regions were identified. Sections 1.2.4.2-1.2.4.5 explains the role of additional neighbouring genes, *GTF2H2*, *RAD17* and *SERF1A/1B* and other modifying factors.

1.2.4.1 The NAIP and NAIPΨ genes

The NLR family, apoptosis inhibitory protein *(NAIP),* also known as baculoviral IAP repeat-containing protein 1 (*BIRC1*) gene, has a size of 56 632 kb and encodes a protein of 1 403 amino acids (*NAIP*; NCBI gene ID: 4671) (NCBI, 2016). The NAIP protein is involved in the suppression of apoptosis. Various transcripts and partial and complete pseudogenes have been observed. A pseudogene, *NAIPΨ*, is located in the centromeric copy of the SMN region and differs from *NAIP* by lacking exons 4 and 5 (Banyer *et al.*, 1998).

A more severe SMA phenotype has been associated with large deletions of *SMN1*, exon 7 including the *NAIP* gene (Jedrzejowska *et al.*, 2009; Miskovic *et al.*, 2011; Watihayati *et al.*, 2009). Patients affected with SMA, type II have been shown to have smaller deletions, with at least one copy of the NAIP gene present, compared with patients affected with SMA type 1 (Samilchuk *et al.*, 1996). Qu *et al.* (2015) showed that a combined increased copy number of *SMN2* and *NAIP* are associated with a larger effect on disease severity than each gene alone. This is supported by a recent study performed by Noguchi *et al.* (2016) where the telomeric *NAIP* copy number could be correlated with SMA type. The centromeric *NAIPΨ* copy number seemed to stay stable regardless of SMA type. A similar association is seen for the GTF2H2 gene, as discussed in section 1.2.4.2.

1.2.4.2 The telomeric and centromeric GTF2H2 genes

The general transcription factor IIH subunit 2 (telomeric-*GTF2H2*) gene, also known as the p44 subunit of basal transcription factor II gene (*BTF2p44*) has a size of 1 951 bp and encodes a protein of 395 amino acids (*GTH2F2*; NCBI gene ID: 2966) (NCBI, 2016). The GTF2H2 protein is involved in transcription and DNA repair. Various transcripts and a centromeric pseudogene (centromeric-*GTF2H2*) have been reported. The exact role of *GTF2H2* in SMA remains unknown. The GTF2H2 gene has also been shown to act as a negative modifier of SMA (Theodorou *et al.*, 2015). Noguchi *et al.* (2016) reported the copy number of telomeric-*GTF2H2* to be correlated with SMA type. The copy number of centromeric-*GTF2H2* seemed to stay stable regardless of SMA type. A similar association is seen for the NAIP gene, discussed in section 1.2.4.1.

Two proposed base pair differences between centromeric-*GTF2H2* and telomeric-*GTF2H2* can be used to distinguish between these two genes. The c.453A>G (p. Ile151Met; rs200357275) variant in exon 7 has been associated with centromeric-*GTF2H2* and the c.706C>G (p.Leu236Val; rs201102513) variant in exon 10, has been associated with telomeric-*GTF2H2*. These variants can therefore be used to determine CNVs of centromeric-*GTF2H2* and telomeric-*GTF2H2*, respectively (Carter *et al.*, 1997).

1.2.4.3 The RAD17 gene

The RAD17 checkpoint clamp loader component (*RAD17*) gene has a size of 45 505 bp and encodes a protein of 370 amino acids (*RAD17*; NCBI gene ID: 5884) (NCBI, 2016). The RAD17 protein is a cell cycle checkpoint protein involved in cell cycle arrest and subsequent DNA repair. Various transcripts and pseudogenes have been reported. *RAD17* is located 300kb away from *SMN1* and *RAD17* deletions could indicate the extent of homozygous deletions including *SMN1*, exon 7 (SALSA MLPA probemix P021-A2 SMA product description (MRC Holland, 2016)).

1.2.4.4 The SERF1A and SERF1B genes

The small EDRK-rich factor 1A (SERF1A) and SERF1B genes (*SERF1B*; NCBI gene ID: 728492), also respectively known as telomeric and centromeric *SERF*, have

identical nucleotide sequences but can be distinguished by their different upstream sequences. *SERF1A* and *SERF1B* both have a size of 17 858 bp and encode a protein of 110 amino acids (NCBI, 2016). The SERF protein is thought to act as a modifier of amyloid fibre assembly (Falsone, *et al.*, 2012).

Arkblad *et al.* (2009) reported that *SERF1A* copy number correlated with *SMN2* copy number with a specificity of 98% and Noguchi *et al.* (2016) showed that *SERF1A* copy number correlates with disease severity. An inverse correlation was found between *SMN2*, *SERF1B* and *NAIP* copy number and disease severity. *SMN2* and *SERF1A* were shown to have a joint modifying effect on SMA phenotype (Brkušanin *et al.*, 2015).

1.2.4.5 Other disease modifying factors

The p.Gly287Arg (c.859G>C) variant in *SMN2* has been shown to act as a splice enhancer increasing the amount of full-length SMN transcripts produced from *SMN2* in patients with homozygous deletions of *SMN1*, exon 7 (Prior *et al.*, 2009). An increased expression of plastin 3 (PLS3) transcripts have been associated with a milder form of SMA in postpubertal girls, suggesting that there might be a gender/age-based effect on patients with SMA (Stratigopoulos *et al.*, 2010). Recently, seven novel target genes (*BMP4*, *SERPINE1*, *GATA6*, *PTGS2*, *BCL22*, *IL6* and *CNTN1*) have been found to influence the severity of SMA by microarray RNA transcriptional analysis (Yang *et al.*, 2016).

1.3 SMA in African populations

1.3.1 SMA in the black SA population

Only 51% of black SA patients with SMA have been shown to have the common homozygous deletion of *SMN1*, exon 7 (Labrum *et al.*, 2007, Stevens *et al.*, 1999). Black SA patients, who previously tested negative for the common homozygous *SMN1*, exon 7 deletion and who were thought to be clinically suggestive of SMA after examination by a paediatric neurologist, were investigated for *SMN1* copy number using an in-house dosage assay.

It was previously reported that one copy only of *SMN1,* exon 7 was deleted in 69.5% of black SA patients. The heterozygous *SMN1*, exon 7 deletion frequency in the SA population seemed much higher compared to other related studies (2-5%), supporting the diagnosis of SMA in these patients and suggesting that SMA is probably due to unidentified mutations in this region (Labrum *et al.*, 2007). The heterozygous rate may have been overestimated and will be further discussed in section 4.2.4.2.

In the same study, the frequency of SMA was reported to be high in SA populations with an estimated birth incidence of 1 in 3 574 and a carrier rate of 1 in 50 in the black population. The incidence of SMA in SA black individuals was reported to be almost as common as that for albinism (1 in 3 900) (Labrum *et al.*, 2007). The carrier rates may have been overestimated and will be further discussed in section 4.2.1.2.

Furthermore, a high frequency of 31.5% of black SA patients with SMA were shown to have smaller deletions including *SMN1*, exon 7 but with exon 8 present, due to gene conversions (Stevens *et al.*, 1999). Additional evidence for this hypothesis was reported by Labrum *et al.* (2007) who observed a lower frequency of large deletions spanning *SMN1*, exons 7, 8 and the *NAIP* gene in black SA patients with SMA (9.8%) when compared to white SA patients with SMA (41.7%).

These results were disputed by a study performed in the Western Cape on 30 patients (of which 12 were of black ancestry), showing that black SA patients affected with SMA do not differ genetically or phenotypically from international reports (Wilmshurst *et al.*, 2002).

MLPA, a robust quantitative detection method, has simplified the detection of heterozygote deletions of *SMN1*, exon 7 and allows for an unbiased detection of copy number of various genes located in the SMN region. A pilot study performed at the Division using MLPA, showed unusual CNVs of the *SMN* region in black patients (Vorster, Essop & Krause, 2011).

1.3.2 SMA in other African populations

A study performed on various North-American population groups, showed an unusually high frequency of multiple copies of *SMN1* in the African American population when compared to other North-American populations (Hendrickson *et al.*,

2009). A study performed on unaffected individuals from various sub-Saharan African populations (Kenyan, Malian and Nigerian) confirmed this observation and showed a higher frequency of multiple copies of *SMN1* and deletions of *SMN2* than European populations (Sangaré *et al.*, 2014).

SMA was previously thought to be rare in African populations with limited studies performed in Northern Africa (Tunisia, Egypt, Nigeria, Algeria and Senegal) but this was likely due to an underestimation (Mrad *et al.*, 2006; Ndiaye *et al.*, 2002; Pelleboer *et al.*, 1989; Shawky *et al.*, 2001; Tazir and Geronimi, 1990). SMA is one of the most common reasons for referral for diagnostic testing at the Division. Multiple *SMN1* copies could potentially confound diagnostic and carrier testing and will be further discussed in sections 4.2.1.2 and 4.2.4.3.

1.4 Genetic therapies for SMA

Current research has moved away from diagnosing SMA to looking for potential mechanisms by which SMA can be clinically treated. Of the six programs hosted by the Cure SMA clinical trial programs, one uses gene therapy technology, two are aimed at protecting the nerves or muscles and three are designed to manipulate the *SMN2* gene to increase SMN production (Cure SMA 2016).

In December 2016, the Food and Drug Administration (FDA) announced their approval of Spinraza, the first therapy approved for SMA. An antisense oligonucleotide called Nusinersen targets the intronic splicing silencer N1 (ISS-N1) and has been shown to fully correct the faulty splicing of exon 7 from SMN2, restoring normal levels of SMN protein in all SMA types (Glascock et al., 2017; Singh et al., 2015).

1.5 Aim and objectives

It has been hypothesised that complex population-specific rearrangements of the SMN region could cause SMA in the black SA population (Labrum *et al.*, 2007; Vorster, Essop & Krause, 2011). This hypothesis is supported by studies performed on African American and sub-Saharan African populations showing a higher variability

of the SMN region in individuals of African ancestry (Hendrickson *et al.*, 2009; Sangaré *et al.*, 2014).

Due to the discrepancies observed in previous SA studies (Labrum *et al.*, 2007; Wilmshurst *et al.*, 2002) and in order to investigate potential novel complex rearrangements of the SMN region, quantitative MLPA testing was proposed to be a useful technique to determine the copy number of various genes in the SMN region and the potential interaction between them.

Genomic rearrangement can be described as mutations involving deletion, duplication, insertion, inversion and translocation. Genomic rearrangements can alter the copy number of genes or disrupt genes and could therefore be pathogenic or benign. Rearrangements are caused by both homologous and non-homologous recombination mechanisms, with homologous recombination being more commonly involved with recurrent rearrangements (Lupski and Stankiewicz, 2005). Homologous recombination as a mechanism of gene conversion in the SMN region is described in section 1.2.2.3.

The main aim of this research project was to further investigate the molecular basis of SMA and to attempt to understand the possible disease mechanism of SMA in the black SA population, using newer, more reliable technologies. This could potentially improve the diagnostic, carrier and prenatal testing and genetic counselling for the 49% of black individuals clinically affected with SMA and their families, who test negative for the homozygous deletion of *SMN1,* exon 7 and for whom definitive diagnostic testing is currently unavailable.

The specific objectives for this project were:

1.) To create an SMA database and perform a retrospective audit of patients, referred to the Division for SMA testing, from September, 1991 to October, 2015, to identify patients belonging to different patient categories (M_1/M_1^B, M_2/M_2^B, M_1/M_1^W, M_2/M_2^W and U/U^B).

2.) To perform MLPA analysis on patients, who previously tested negative for the homozygous deletion of *SMN1*, exon 7 and who had features clinically suggestive of SMA (U/U^B), to try to identify potential pathogenic CNVs.

3.) To perform comparative MLPA analysis on black individuals referred for SMA testing with known *SMN1* and *SMN2* CNVs (M_1/M_1^B and M_2/M_2^B) and black individuals not affected by SMA (N/N^B), to identify common CNVs associated with SMA in the black SA population.

4.) To perform MLPA analysis on white patients referred for SMA testing with known *SMN1* and *SMN2* CNVs (M_1/M_1^W and M_2/M_2^W) and unaffected white individuals (N/N^W), as a comparison to black individuals (M_1/M_1^B and M_2/M_2^B and N/N^B).

5.) To construct pedigrees of M_1/M_1^B and N/N^B families to investigate common pathogenic and non-pathogenic CNV patterns within the black SA population.

6.) To analyse data generated by MLPA using statistical analysis in order to identify statistically significant associations between CNVs and SMA disease status to better understand the genomic architecture of the SMN region.

7.) To validate and optimise qPCR as an alternative diagnostic technique to detect *SMN1* and *SMN2* copy number.

CHAPTER 2 SUBJECTS AND METHODS

This chapter will introduce the subjects and controls selected for this project and explain the different methods employed to determine CNV patterns. Furthermore, this chapter will outline the strategy for testing.

The project focussed primarily on the black SA population. Patient request forms are often not completed sufficiently and information on the clinical symptoms, family history (background) or ethnicity of patients are frequently omitted. If the ethnic group was not indicated on the request from, the ethnic group was assigned using the family surname as a proxy, if possible. This method is by no means accurate since the black SA population consists of various ethnic groups from various geographical locations and surnames do not always correlate consistently with ethnicity. It is possible that there could be genetic differences related to the disease etiology of SMA among different African ethnic groups. Panel testing using ancestry-informative markers are not available for more detailed ancestry testing in South Africa to determine the different tribal origins of black SA individuals, e.g. Xhosa, Zulu, Sotho, etc. Furthermore, this kind of testing would be costly and not practical in a diagnostic setting.

The ethnicity of SA patients of white, Indian and mixed ancestry origin was determined similarly using the surname and geographical location of the patient as a guide, where possible.

2.1 Audit of SMA referrals: Sep 1991 – Oct 2015

A retrospective audit of patients, referred to the Division for SMA testing, from September, 1991 to October, 2015, a period of more than twenty years was performed. Data were collected from request forms (including date collected, date of birth, gender, ethnicity, referring doctor, location) and laboratory reports (homozygous and heterozygous (where available) deletion of *SMN1*, exon 7 results).

An SMA database was set up to identify patients belonging to different patient categories (M_1/M_1^B, M_2/M_2^B, M_1/M_1^W, M_2/M_2^W and U/U^B), further defined in section 2.5.

2.2 Subjects and controls

Comparisons of CNV patterns were performed among seven different patient groups which will be further defined in subsequent paragraphs. The patient and subject groups have been defined in the list of definitions at the beginning of the dissertation (p.xv) and consist of:

- **Group A (U/U^B):** black patients who presented with symptoms clinically suggestive of SMA and who previously tested negative for the homozygous deletion of *SMN1*, exon 7 on the diagnostic assay. These **b**lack patients have an assumed U/U^B (**U**nidentified mutation/**U**nidentified mutation) genotype.
- **Group B (N/N^B):** **B**lack controls negative for SMA (individuals with no family history of SMA). These black controls have an assumed N/N^B (**N**egative/**N**egative) genotype.
- **Group C (N/N^W):** **W**hite controls negative for SMA (individuals with no family history of SMA). These white controls have an assumed N/N^W (**N**egative/**N**egative) genotype.
- **Group D (M_1/M_1^B):** **B**lack individuals who were previously identified to have a homozygous deletion of *SMN1*, exon 7 on the diagnostic assay and were therefore confirmed to be affected with SMA. These individuals have a M_1/M_1^B (**M**utation **1**: deletion of *SMN1*, exon 7 mutation/**M**utation **1**: deletion of *SMN1*, exon 7 mutation) genotype.
- **Group E (M_1/M_1^W):** **W**hite individuals who were previously identified to have a homozygous deletion of *SMN1*, exon 7 on the diagnostic assay and were therefore confirmed to be affected with SMA. These individuals have a M_1/M_1^W (**M**utation **1**: deletion of *SMN1*, exon 7 mutation/ **M**utation **1**: deletion of *SMN1*, exon 7 mutation) genotype.
- **Group F (M_2/M_2^B):** **B**lack individuals who were previously identified to have a homozygous deletion of *SMN2*, exon 7 on the diagnostic assay. These individuals have a M_2/M_2^B (**M**utation **2**: deletion of *SMN2*, exon 7 mutation/ **M**utation **2**: deletion of *SMN2*, exon 7 mutation) genotype.
- **Group G (M_2/M_2^W):** **W**hite individuals who were previously identified to have a homozygous deletion of *SMN2*, exon 7 on the diagnostic assay. These individuals have a M_2/M_2^W (**M**utation **2**: deletion of *SMN2*, exon 7 mutation/ **M**utation **2**: deletion of *SMN2*, exon 7 mutation) genotype.

Each group will be further defined in sections 2.2.1 – 2.2.2.

2.2.1 Group A: U/U^B individuals

U/U^B individuals were identified in collaboration with the Clinical Section of the Division and Professor John Rodda, head of the Paediatric Neurology Department, Chris Hani Baragwanath and Charlotte Maxeke Academic Hospitals. In total, 72 U/U^B individuals were identified, nine of whom had muscle biopsies, suggestive of SMA. MLPA analysis was performed on these individuals to identify additional potential pathogenic CNV patterns. DNA samples of family members of U/U^B individuals were not available. Group A is the main focus of this research project.

2.2.2 Subject groups used for comparison

2.2.2.1 Control groups negative for SMA

Group B: N/N^B individuals

Sixty one black families referred to the Division for diagnostic testing for other genetic conditions, excluding SMA or any other muscular disorders, were identified (N/N^B families). DNA samples of these families are stored in the DNA bank of the Division. In order to be included, DNA had to be available from two unrelated parents and at least one child. Siblings were included, where available.

MLPA analysis was performed on family members of all 61 N/N^B families (200 individuals). The unaffected parents of these families were then used as negative controls in this study and consisted of a total of 122 unrelated N/N^B individuals. Haplotypes were constructed from the MLPA data and family pedigrees (see section 2.3.4 for an explanation of haplotype analysis) in order to investigate the phase of potential common non-pathogenic CNV patterns and to identify the normal variation of the SMN region in the black SA population.

Group C: N/N^W individuals

In order to compare the typical non-pathogenic CNV patterns of black and white individuals, 30 random unrelated N/N^W individuals were tested. N/N^W families were not included since haplotype analysis was mainly aimed at determining the phase of multiple copies of *SMN1* which are common in African populations but reported to be rare in Caucasian populations.

2.2.2.2 Subject groups with a homozygous deletion of *SMN1*, exon 7

Group D: M_1/M_1^B individuals

Twenty five unrelated black families (71 individuals), previously referred for SMA testing, were identified (M_1/M_1^B families). In order to be included in the study, families had to consist of at least two individuals, including at least one affected individual who previously tested positive for the homozygous deletion of *SMN1*, exon 7 on the diagnostic assay. For example, some families consisted of only one affected child and a single parent and other families consisted of a pair of affected siblings. These M_1/M_1^B families included 24 parents in 21 families who were all assumed to be obligate carriers. This assumption might not be entirely accurate in families where the parents have more than one copy of *SMN1*, exon 7 since novel mutation events could explain the results in affected individuals.

MLPA analysis was performed on available family members of all 25 M_1/M_1^B families and haplotypes were constructed from the MLPA data and family pedigrees (see section 2.3.4 for an explanation of haplotype analysis). One proband of each of these 25 M_1/M_1^B families in addition to 50 unrelated M_1/M_1^B individuals were included in Group D (75 individuals in total). These M_1/M_1^B individuals were tested with MLPA in order to investigate the molecular structure of pathogenic CNV patterns including the extent of homozygous deletion of *SMN1*, exon 7 and potential gene conversion events. If a clear common pathogenic deletion CNV pattern were to be identified in M_1/M_1^B individuals, perhaps other pathogenic CNV patterns could be inferred in U/U^B individuals with a heterozygous deletion of *SMN1*, exon 7 by analysing the underlying CNV patterns overlapping the common homozygous deletion of *SMN1*, exon 7 patterns.

Group E: M_1/M_1^W individuals

For comparison, 30 random unrelated M_1/M_1^B individuals were tested to compare the molecular structure of pathogenic CNV patterns between M_1/M_1^B and M_1/M_1^W individuals. Once again, M_1/M_1^W families were not included.

2.2.2.3 Subject groups with a homozygous deletion of *SMN2*, exon 7

Group F: M_2/M_2^B individuals

Fifty unrelated M_2/M_2^B individuals were included in Group F and tested on MLPA to determine the underlying molecular structure of this common CNV and to understand the interaction between the *SMN1* and *SMN2* genes.

Group G: M_2/M_2^W individuals

For comparison, eight random unrelated M_2/M_2^W individuals were tested on MLPA to compare the molecular structure of this CNV between M_2/M_2^B and M_2/M_2^W individuals. Only eight M_2/M_2^W individuals were included in this group, since they were the only M_2/M_2^W individuals who have been identified to have the homozygous deletion of *SMN2*, exon 7. The frequency of this common CNV differs significantly between black and white individuals and will be further discussed in section 4.2.1.3.

Figure 2.1 gives an overall description of the subjects and controls included in this study and the testing strategy of this project.

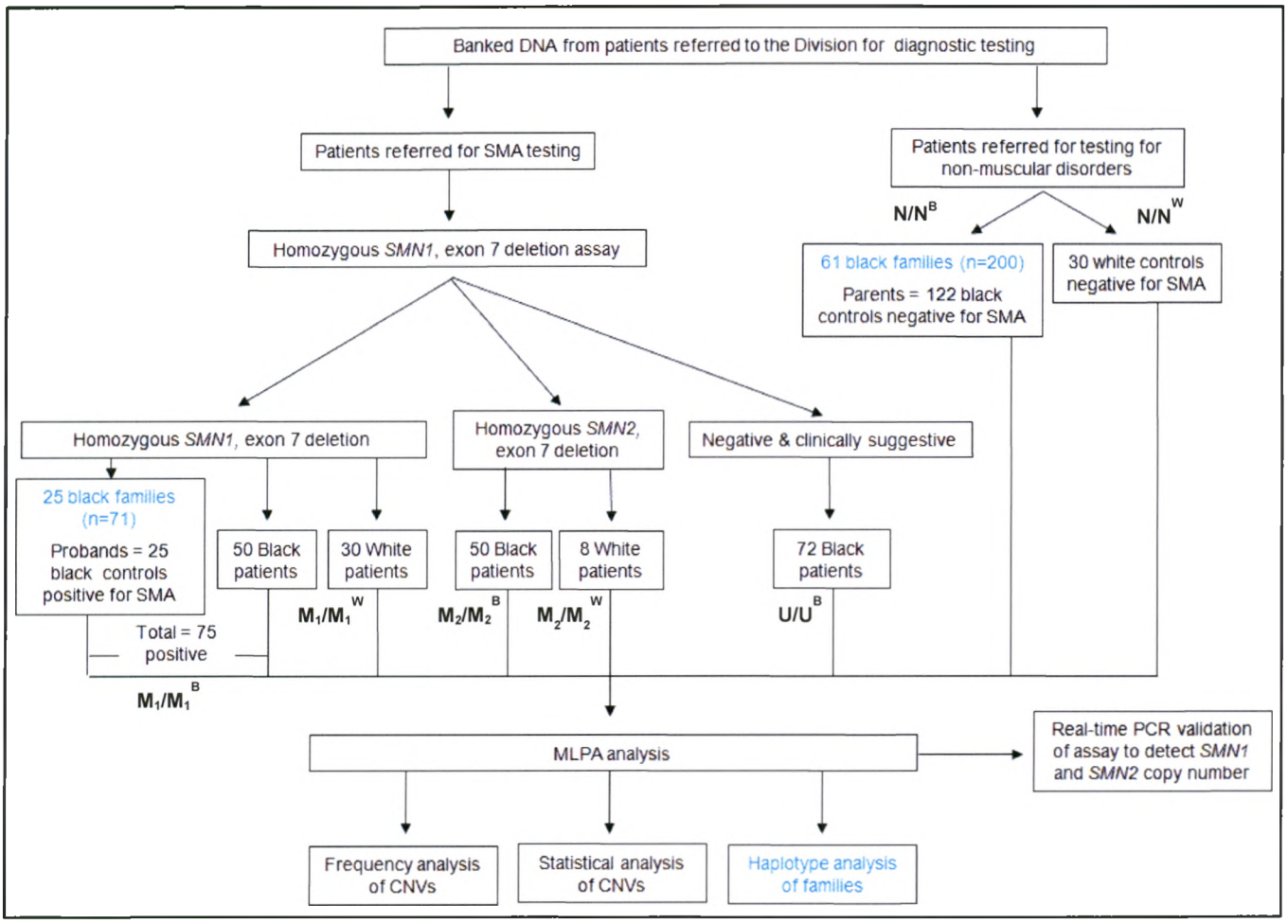


Figure 2.1: The strategy of testing, summarising the patients and controls included in this study. Families are indicated in blue.

2.3 Methods

2.3.1 DNA Extraction

Genomic DNA was extracted from whole blood using the salting out method (Miller *et al.*, 1988), a commercial DNA extraction kit (High Pure PCR Template Preparation Kit, Roche Diagnostics) or in the case of chorionic villus sampling (CVS) and amniocyte material, the phenol-chloroform extraction method (Barker, 2004). All samples have been processed and extracted in a diagnostic setting, with stringent quality assurance.

2.3.2 MLPA using the P021 kit

MLPA is a robust PCR application, which utilises fluorescently labelled probes specific to multiple targets, which are amplified with an universal primer pair in a single reaction. By comparing the copy number of PCR products observed in a patient sample with endogenous reference probes and external control samples, quantitative changes in DNA fragments can be determined. Schouten *et al.* (2002) first described the MLPA technique to determine abnormal CNVs in tumour samples, trisomies and even to detect point mutations. MLPA is the gold standard technique used in exon deletion and duplication detection in many diagnostic tests such as Duchenne/Becker muscular dystrophy (Abbs *et al.*, 2010).

MLPA consists of five main steps:

1. DNA denaturation and hybridisation with multiple gene-specific probes
2. Ligation of the two separate parts of each probe
3. PCR amplification of the probe regions
4. Separation of PCR products by electrophoresis
5. Fragment analysis

Each probe consists of two parts (A and B), located one base pair away from each other. Both parts A and B contain a hybridisation sequence (indicated in blue in Figure 2.2) which is specific to the target sequence of interest and a primer binding region where the universal PCR primers will bind during PCR amplification (indicated in black in Figure 2.2). In addition, part B contains a "stuffer" sequence (indicated in green in Figure 2.2), which differs in length for each probe to ensure that different PCR products can be distinguished by length (MRC-Holland, 2016). Figure 2.2 illustrates the various steps involved in MLPA analysis.

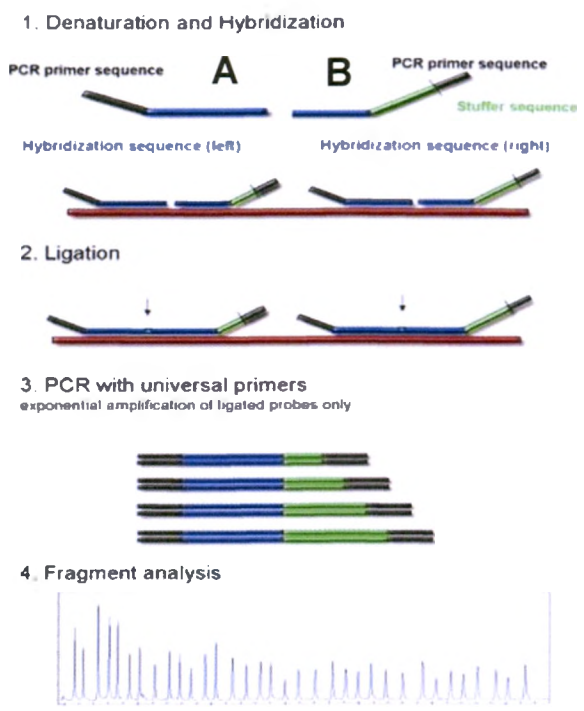


Figure 2.2: Summary of the MLPA protocol (Adapted from MRC-Holland, 2016)

1. Denaturation and hybridisation
Sample DNA is denatured and probes hybridise to the regions of interest overnight. An MLPA probe consists of 2 parts (A and B). The hybridisation sequences, indicated in blue, are specific to the region of interest.

2. Ligation
A ligation enzyme is added to the reaction to join parts A and B of each probe to make a single probe.

3. PCR
Universal primers bind to the PCR primer sequences, indicated in black. It is important to note that amplification of the probes and not the DNA takes place.

4. Fragment analysis
Fragment separation and visualisation are performed on a genetic analyzer. Probes are distinguished by the difference in PCR product size.

5. Analysis
The fluorescence quantity of each probe is compared to the fluorescence quantity of the internal control probes and external negative control samples, in order to determine the copy number of these regions

2.3.2.1 The P021 probe mix (version A2)

The MLPA P021 probe mix (MRC Holland, Amsterdam, Netherlands) consists of a multiplex of 46 probes, consisting of seven "DQ" dosage quality assurance probes; two sex-chromosome specific probes (for gender determination); 22 internal reference probes, specific to various chromosomal regions and not associated with SMA; 15 probes specific to the SMN region (eight targeting the *SMN1* and *SMN2* genes and seven probes targeting neighbouring genes).

MLPA probes specific to *SMN1* and *SMN2*

The MLPA P021 probe mix is mainly designed to detect *SMN1* and *SMN2*, exon 7 copy numbers, but since deletions have been observed in exons 1, 4, 6 and 8 (Arkblad *et al.*, 2006), probes binding to these exons have also been included. Probes have been designed to target the critical one base pair difference between *SMN1* and *SMN2* in exon 7 and can therefore distinguish between exon 7 of *SMN1* and *SMN2*. Similarly, probes have been designed to target a one base pair difference between *SMN1* and *SMN2*, exon 8 and can therefore distinguish between exon 8 of *SMN1* and *SMN2*.

Due to identical sequences of exons 1, 4, 6 and another region of exon 8 of the *SMN1* and *SMN2* genes, this kit cannot distinguish between *SMN1* and *SMN2* for these regions and will give a combined copy number result. This essentially means the copy number in a patient who is unaffected by SMA and who has the expected two copies of these regions would be 4 (two copies of *SMN1* and two copies of *SMN2*). An absence of these probes could therefore represent a deletion in either or both copies of the *SMN1* and *SMN2*, which complicates analysis. It is important to keep in mind that heterozygous deletions of these probes (representing a deletion of 2 out of 4 copies) in conjunction with an *SMN1*, exon 7 or *SMN2*, exon 7 deletion could suggest a full deletion of the *SMN1* or *SMN2* genes as shown in figure 2.3. These regions are very difficult to analyse and results for these probes need to be interpreted along with the results for the *SMN1* and *SMN2* specific probes.

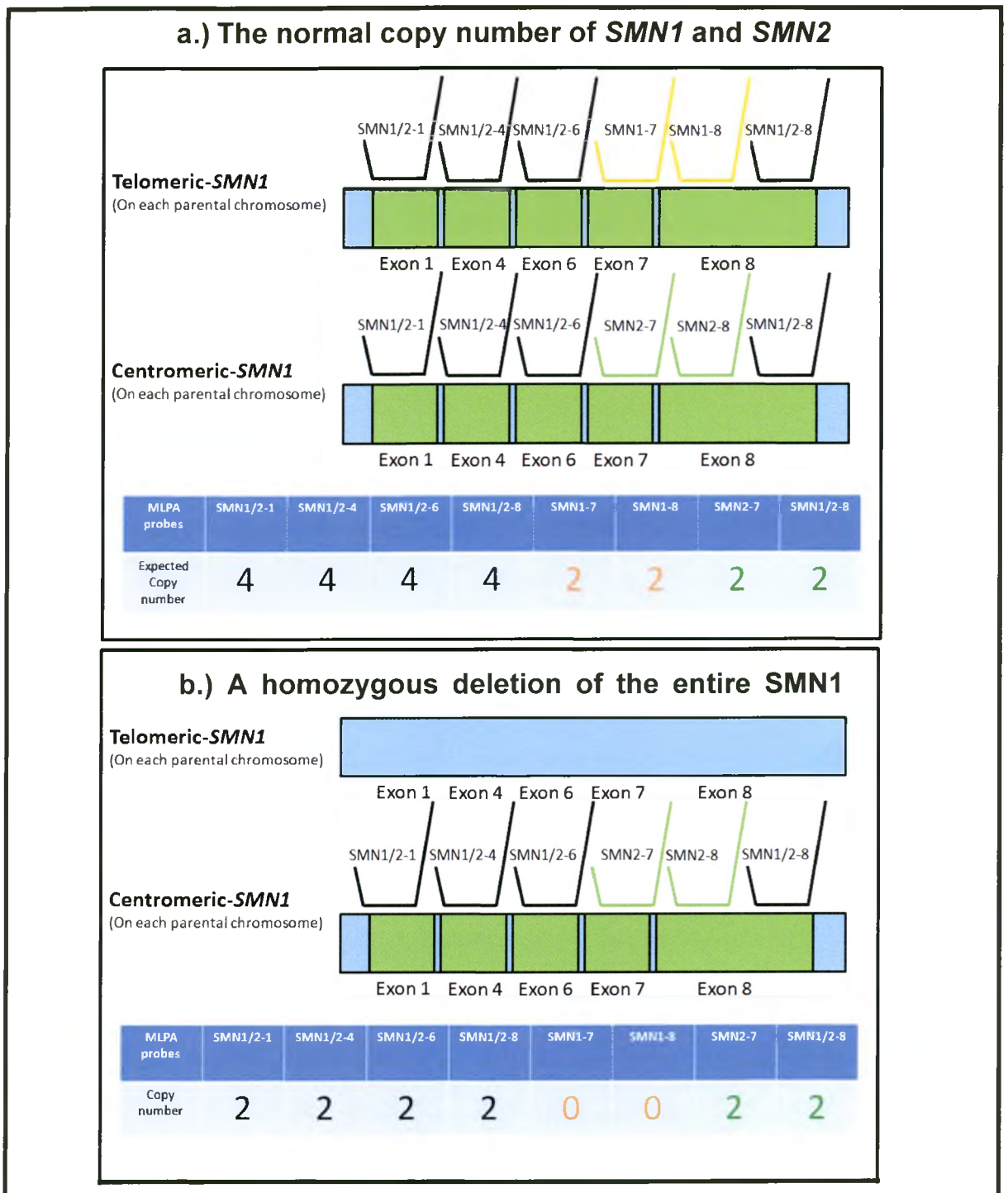


Figure 2.3: The relationship between copy number and the *SMN1/2* probes

a.) The location of the different *SMN1/2* MLPA probe-binding regions is indicated in a patient who has the expected two copies of *SMN1* & *SMN2* (one on each parental chromosome). The red probes are specific to *SMN1*, exons 7 and 8 and the blue probes are specific to *SMN2*, exon 7 and 8. Two copies of each of these regions are expected. The purple probes bind to both *SMN1* & *SMN2*, exons 1, 4, 6 & 8. Four copies of each of these regions are expected.

b.) In a patient who has a homozygous deletion of *SMN1*, no probes will bind to this gene, leading to 0 copies of *SMN1*, exons 7 & 8 and only 2 copies of *SMN1* & *SMN2*, exons 1, 4, 6 & 8, now only representing the two copies of centromeric *SMN2*.

Probes specific to neighbouring genes in the SMN region have been included to assist with determining the extent of deletions. These genes are regarded as potential disease-modifying genes (He *et al.*, 2013). Due to the high homology and the presence of multiple pseudogenes of the SMN region, probes have once again been designed to target single nucleotide differences between different pseudogenes, where possible, to provide a more detailed picture of the SMN region.

MLPA probes specific to the *NAIP* and *NAIPΨ* regions

A probe specific to exon 13 present in both the *NAIP (BIRC1)* gene and its pseudogene, *NAIPΨ,* binds to both genes and will give a combined copy number result, similarly to *SMN1/2*, exons 1, 4, 6 and 8. The copy number in a patient who has the expected two copies of these regions would be four copies (two copies of *NAIP* and two copies of *NAIPΨ*). An absence of these probes could therefore represent a deletion in either or both of the *NAIP* and *NAIPΨ* genes. An additional probe specific to exon 5 only binds to *NAIP* which (exon 5 is deleted in the *NAIPΨ* gene), which co-locates with *SMN1*. These regions are very difficult to interpret. A heterozygous deletion of *NAIP/NAIPΨ* exon 13 (representing a deletion of two out of four copies expected for exon 13) in conjunction with a *NAIP*, exon 5 deletion could suggest a full deletion of the *NAIP* gene as illustrated by figure 2.4. Alternatively a heterozygous deletion of *NAIP/NAIPΨ* exon 13 only, could suggest a full deletion of the *NAIPΨ* gene.

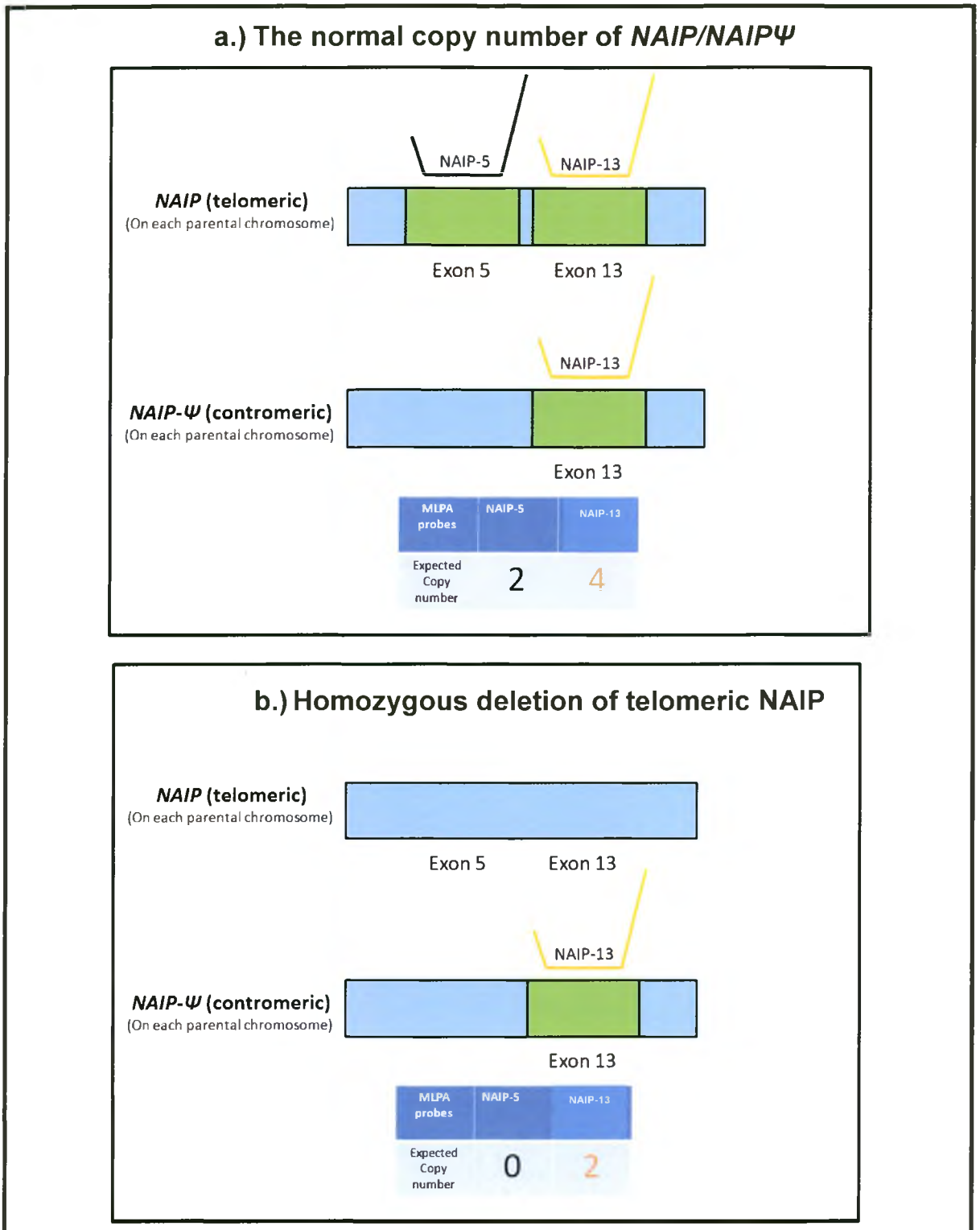


Figure 2.4: The relationship between copy number and the *NAIP* probes

a.) The location of the different *NAIP* MLPA probe-binding regions is indicated in a patient who has the expected two copies of *NAIP* & *NAIP-ψ* (one on each parental chromosome). The purple probe binds to exon 13 of both *NAIP* & *NAIP-ψ*. Four copies of this region are expected. The red probe binds to exon 5, only present in *NAIP*. Two copies of this region are expected.

b.) In a patient who has a homozygous deletion of *NAIP*, no probes will bind to this gene, leading to zero copies of NAIP-5 and only two copies of NAIP-13, now only representing the two copies of *NAIP-ψ*.

NOTE

The SMN region and therefore the design of the P021 MLPA kit are complex and it is recommended that the reader should focus on the functional telomeric probe regions, *SMN1*, exons 7, 8 and *NAIP*, exon 5. Probes specific to the neighbouring genes in the SMN region provide some insight into the nature of large CNVs of the SMN region but the main findings of this dissertation are focussed on *SMN1*, exons 7, 8 and *NAIP.*

MLPA probes specific to the centromeric-*GTF2H2* and telomeric-*GTF2H2* regions

A probe specific to exon 5 binds to both GTF2H2 gene variants (telomeric-*GTF2H2*, co-locating with *SMN1* and centromeric-*GTF2H2*, proposed to co-locate with *SMN2*) and will therefore give a combined result, similarly to *SMN1/2*, exons 1, 4, 6 and 8 and *NAIP/NAIPΨ*, exon 13. The copy number in a patient who has the expected two copies of these regions would be four copies (two copies of telomeric-*GTF2H2* and two copies of centromeric-*GTF2H2*). An absence of these probes could therefore represent a deletion in either or both of the telomeric-*GTF2H2* and centromeric-*GTF2H2* genes. Figure 2.5 further explains the relationship of these probes with the copy number of the GTF2H2 genes.

In addition, the P021-A2 kit includes two probes targeting the two base pair differences between telomeric-*GTF2H2* and centromeric-*GTF2H2*. One probe (GTF2H2-8) is specific to the c.453A>G (p. Ile151Met; rs200357275) variant in exon 7 which has been associated with centromeric-*GTF2H2*. In contrast, the other probe (GTF2H2-11) is specific to the c.706C>G (p.Leu236Val; rs201102513) variant in exon 10, which has been associated with telomeric-*GTF2H2* as explained in section 1.2.4.2 (Carter *et al.*, 1997). The exact role and orientation of telomeric-*GTF2H2* and centromeric-*GTF2H2* in SMA are unknown. Figure 2.5 shows a visual representation of the binding regions of the three *GTF2H2* MLPA probes.

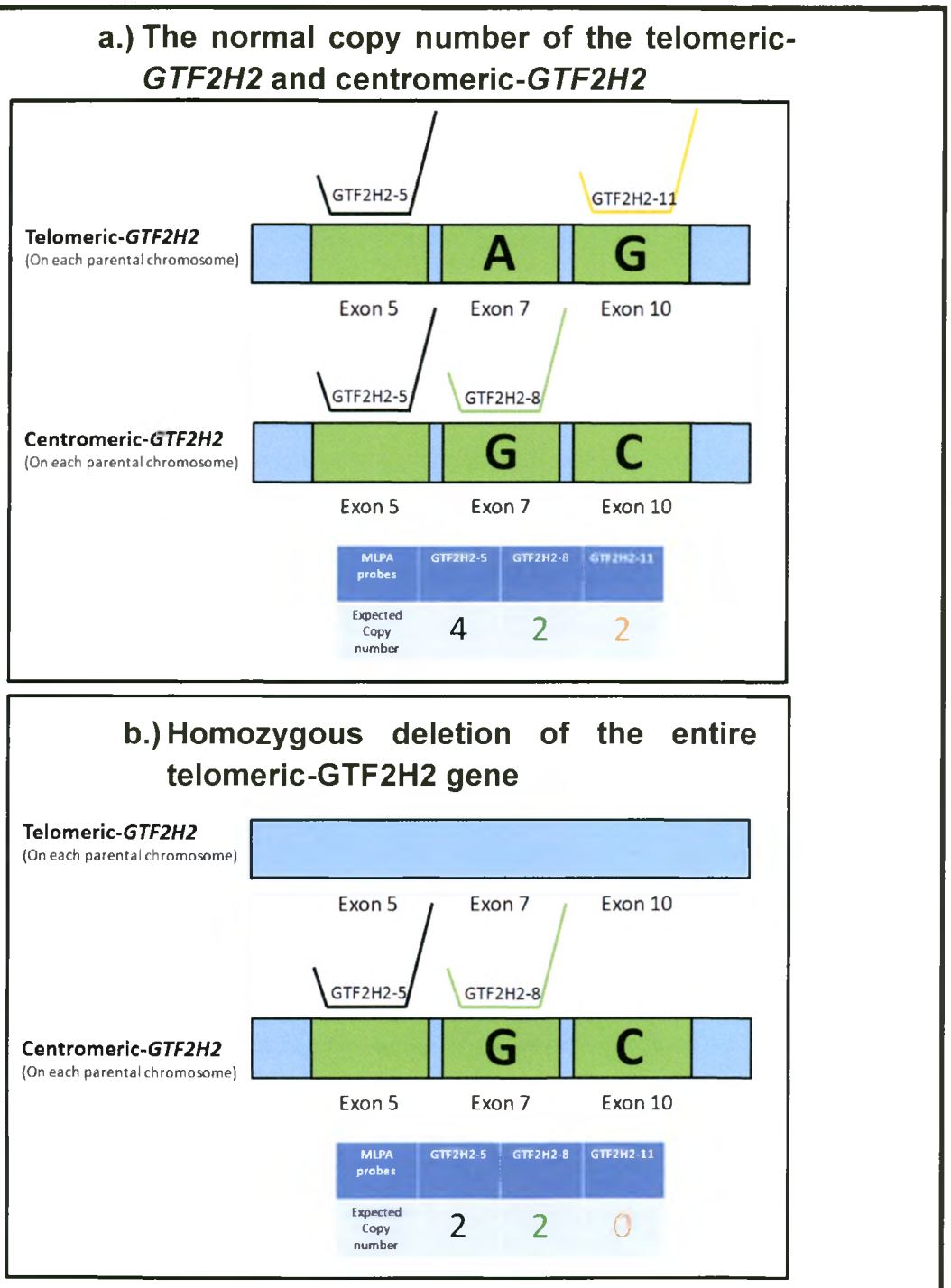


Figure 2.5: The relationship between copy number and *GTF2H2* probes

a.) The location of the different *GTF2H2* MLPA probe-binding regions is indicated in a patient who has the expected two copies of telomeric-*GTF2H2* & centromeric-*GTF2H2* (one on each parental chromosome). The purple probe binds to exon 5 of both telomeric-*GTF2H2* and centromeric-*GTF2H2*. Four copies of this region are expected. The blue probe is specific to the G allele at the c.453 location in exon 7 which has been associated with centromeric-*GTF2H2* and the red probe is specific to the G allele at the c.706 location in exon 11, which has been associated with telomeric-*GTF2H2*. Two copies of each of these regions are expected.

b.) In a patient who has a homozygous deletion of telomeric-*GTF2H2*, no probes will bind to this gene, leading to zero copies of GTF2H2-11 and only two copies of GTF2H2-5, now only representing the two copies of centromeric-*GTF2H2*.

MLPA probes specific to the SERF1A and RAD17 regions

Probes specific to the SERF1A (co-locating with *SMN1*) and RAD17 genes are also included and could provide some information on the extent of large CNVs of *SMN1* and *SMN2*.

Appendix A shows a table summarising the probes present in the P021-A2 kit and Appendix B includes a step-by-step worksheet to set up the P021-A2 SMA MLPA assay. For more information, the SALSA MLPA probemix P021-A2 SMA (Description version 32; 14-10-2014) can be consulted (MRC-Holland, 2016).

2.3.2.2 MLPA data analysis

MLPA was performed in batches of 60 samples using the Applied Biosystems (ABI) 9700 thermal cycler and fragment separation was performed using the ABI Genetic Analyzer 3130xl (Applied Biosystems, Foster City, CA, USA). The ABI Foundation Data Collection and Genemapper data packages, compatible with the ABI 3130xl Genetic Analyzer, were used to collect raw MLPA data (Applied Biosystems, Foster City, CA, USA). Dosage analysis was performed, using the freely available Coffalyser software (MRC Holland, Amsterdam, Netherlands) to quantify CNVs. The Coffalyser software has several quality assurance checks to verify the quality of the DNA used and to ensure that the results obtained are reliable. Intra-sample normalisation compares the intensity of each probe's fluorescent amplification product to the intensity of a set of internal reference probes (located across several chromosomes) in a single patient. Inter-sample normalisation compares the relative fluorescent intensity of each probe's fluorescent amplification product of each patient with a set of known negative controls (who have two copies of each probe). The intensity of a specific probe's amplification is measured by the height of the peak of the associated PCR product.

The profile of a patient with no CNVs will be identical to the reference samples with 100% of the expected copy number (two) and will have a dosage quotient (DQ) of 1. Figure 2.6 represents the P021-A2 SMA MLPA profile of an N/N^{W} individual (two copies present of each probe), generated by Coffalyser.

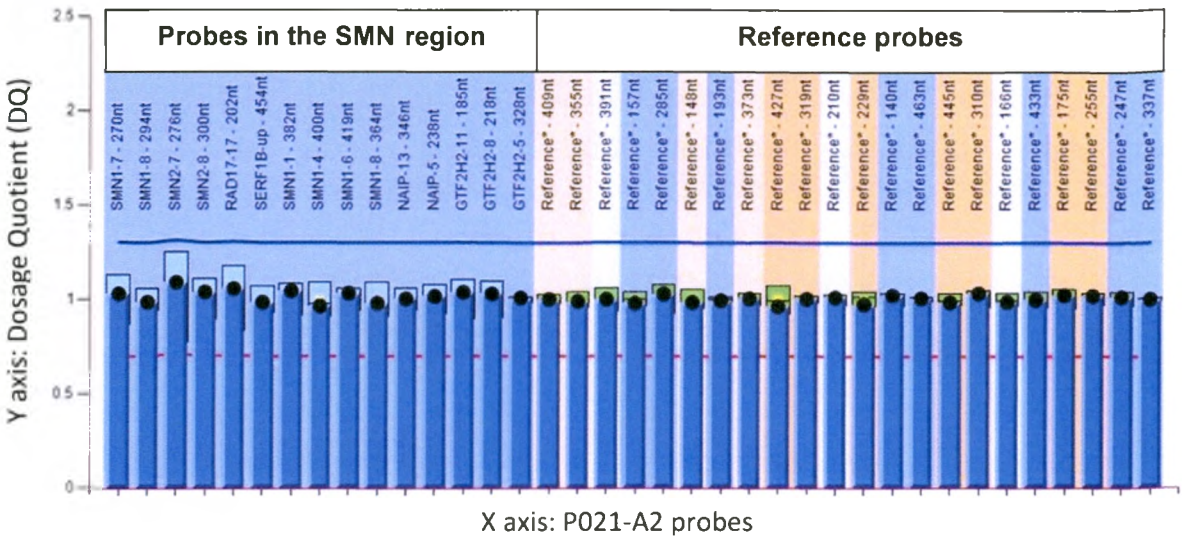


Figure 2.6 Typical P021-A2 MLPA profile of a negative control

A DQ value of 1 represents the expected number of copies of this region. A DQ value of 1 represents four copies of the universal probes, specific to both the telomeric and centromeric copies of a gene: SMN1/2, exons 1, 4, 6 & 8; NAIP-13 and GTF2H2-5. A DQ value of 1 represents two copies of the rest of the probes.

Table 2.1 lists the DQ ranges used in this study to determine the copy number of a specific probe region. These values were adapted from the recommended ranges in the MLPA Protocol One-tube MDP-v005 (MRC Holland, 2016) and on experience gained from performing various MLPA assays on different genetic disorders in a diagnostic setting. These ranges are more inclusive than the original ranges and could potentially influence the accuracy of results.

It is important to note that the copy number allocation differs between probes which are specific to a single region with an expected total of two copies per region, one per parental chromosome (table 2.1a) and probes which are specific to two identical regions located in both the centromeric and telomeric regions (*SMN1* and *SMN2*, exons 1, 4, 6 and 8; *NAIP/NAIPΨ*, exon 13 and *GTFTH2*, exon 5), with an expected total of four copies per region, two per parental chromosome (table 2.1b). When the copy number of a region falls within the equivocal range (in the case of one, three, five and seven copies of the probe regions specific to two regions), it may be difficult to determine the exact copy number.

Table 2.1 The relationship of the MLPA DQ with the probe region copy number

2.1a.) Normal probes specific to a single region (one on each parental chromosome)

Dosage Quotient (DQ) distribution	Copy number
0 « DQ < 0.35	0
0.35 « DQ < 0.65	1
0.65 « DQ < 1.35	2
1.35 « DQ < 1.65	3
1.65 « DQ < 2.35	4

To be applied to the majority of probes which targets a single gene or region. Two copies of these regions are expected in a normal white individual.

2.1b.) Universal probes specific to two regions (two on each parental chromosome)

Dosage Quotient (DQ) distribution	Copy number
0 « DQ < 0.35	0
0.35 « DQ < 0.65	2
0.65 « DQ < 1.35	4
1.35 « DQ < 1.65	6
1.65 « DQ < 2.35	8

To be applied to the *SMN1/2*, exons 1, 4, 6 & 8; *NAIP/NAIPΨ*, exon 13 and *GTF2H2*, exon 5 probes which targets combined telomeric and centromeric regions. Four copies of these regions are expected in a normal white individual.

MLPA results were analysed using two approaches: a.) by statistical analysis to compare dosage trends between different subject groups and to determine significant differences among patient and subject groups (further explained in section 2.3.3) and secondly b.) by haplotype analysis of pedigrees of N/N^B and $M_1/M_1{}^B$ families to determine common chromosomal CNV patterns / haplotypes in the black SA population (further explained in section 2.3.4).

2.3.3 Statistical analysis

Statistical analysis, using Statistica (Dell, version 12.7) and Real Statistics Using Excel (Real-Statistics, 2016) software, was performed on the seven patient groups A-G (as defined in section 2.2).

After performing distribution-of-fit tests, it was determined that the seven patient and subject groups do not conform to a normal distribution and non-parametric analysis (which does not require data to conform to a normal distribution) would be more appropriate when performing statistical analysis. The non-parametric Kruskal-Wallis

(the non-parametric counterpart of the Analysis Of Variance (ANOVA) statistical test) and median test (a more crude measurement of the Kruskal-Wallis ANOVA) were performed to look for significant differences among the seven different patient and subject groups.

The parametric Multivariate Analysis Of Variance (MANOVA) statistical test was used to confirm results observed using the non-parametric Kruskal-Wallis and median tests.

Due to the highly complex nature of the data (data of 15 MLPA probes across seven patient and subject groups with potential associations among probes), multivariate statistical analysis needed to be implemented. In order to perform valuable statistical analysis, the right questions need to be asked. Due to a limited understanding of the SMN region and CNVs in SA populations, statistical analyses were mainly exploratory to see if any obvious associations could be identified among different patient and subject groups. All statistical tests were performed at an α value of 0.05. The Statistica statistical output is available from the CD accompanying this dissertation and the statistical output of RealStats have been included in Appendix F.

2.3.4 Haplotype analysis using family pedigrees and MLPA results

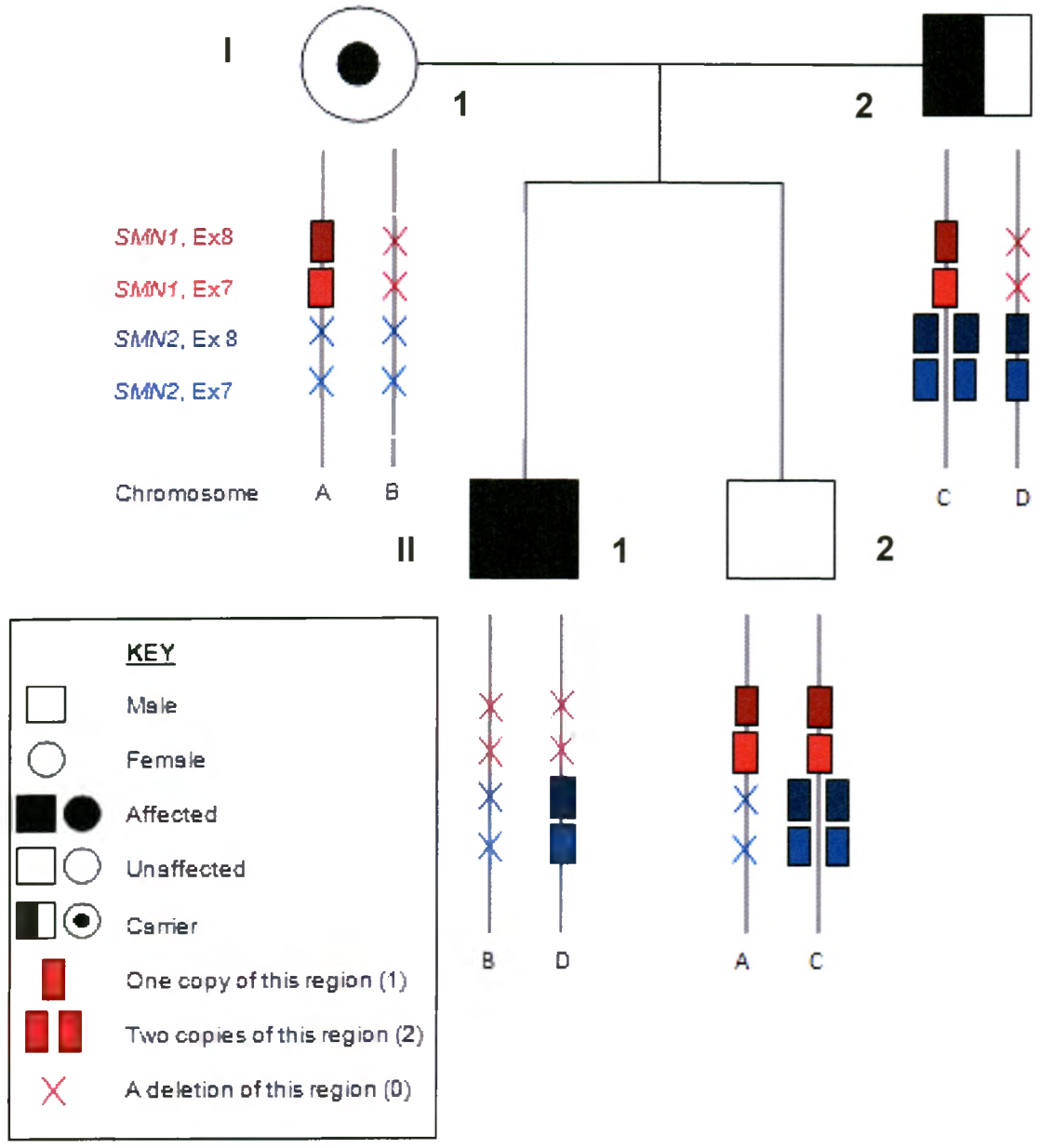
Pedigrees were constructed and used to haplotype MLPA data collected for each of the 15 MLPA probes specific to the SMN region for each of the 61 N/N^B and 25 $M_1/M_1{}^B$ families.

Haplotype phasing software was not used to determine the chromosomal phase of different CNVs due to the complexity of the region which resulted in various possibilities within families. By manually interrogating each SMN probe using raw MLPA data, we attempted to determine the chromosomal phase of different CNVs within a family. The aim of this exercise was to try and determine common pathogenic haplotypes in $M_1/M_1{}^B$ individuals and haplotypes which form part of the normal variation in N/N^B individuals. If common haplotypes could be identified, it was hoped that it could assist in the interpretation of CNV patterns in U/U^B families.

Family pedigrees were categorised as informative if a clear pattern of inheritance from parents to their children could be established with certainty for each of the MLPA probes (i.e. the phase of the CNVs were correctly determined). Family pedigrees were categorised as uninformative if the phase of the CNVs could not be correctly determined due to multiple possible combinations which could explain the inheritance within a family or discrepant results.

If the genotype of one or more family members did not fit within the context of the family across multiple non-contiguous probes, these families were reported to have discrepant results. Discrepant results could arise due to potential novel mutation events, rearrangement of the order of genes, technical MLPA faults or family members who are not related as specified. It is unlikely that discrepant results are due to technical errors with MLPA analysis since all discrepant results were reloaded on the ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) to confirm results and these results were usually not equivocal (borderline between two different copy numbers, as categorised in table 2.1 in section 2.3.2.2) making it unlikely that these copy numbers were classified incorrectly. If the CNV pattern of a single family member did not fit within the context of the family for a single probe and the MLPA result was confirmed on a reload of the sample on the 3130xl Genetic Analyzer and the result was not equivocal, these families were reported to have potential novel mutation events.

Non-pathogenic haplotypes were constructed from parental samples from informative N/N^B families. Pathogenic haplotypes were constructed from probands from informative $M_1/M_1{}^B$ families. Figure 2.7 shows an example of an informative pedigree of a $M_1/M_1{}^B$ family, constructed from MLPA data from the *SMN1*, exon 7 and 8 and *SMN2*, exon 7 and 8 regions. Clear haplotypes could be established and the phase of these haplotypes in this family could be accurately determined.



***SMN1*, exon 7 deletion chromosomes**	**Number of copies**			
	***SMN1*, exon 7**	***SMN1*, exon 8**	***SMN2*, exon 7**	***SMN2*, exon 8**
B	0	0	0	0
D	0	0	1	1

Figure 2.7 Example of an informative pedigree of a family with a history of SMA. The red blocks represent *SMN1*, exons 7 and 8 and the blue blocks represent *SMN2*, exons 7 and 8. The affected child (II-1) has 0 copies of the *SMN1*, exon 7 and 8 region, indicating a homozygous deletion of these regions. Since the affected child had to inherit a deletion chromosome from each parent (I-1 & I-2), the phase of the CNVs could be determined. Child II-2 inherited one *SMN1*, exon 7 copy from each parent and is therefore not a carrier of SMA. The table indicates the resulting pathogenic haplotype identified in the affected child.

NOTE

The main focus of this project is on the MLPA analysis of relevant patient and subject groups (Groups A - G). The next few sections explain additional techniques which could be utilised for SMA diagnostic testing in the future. These techniques do not form part of the main project but proofs of concept have been developed for these techniques.

The design and optimisation of an alternative technique to MLPA, a DNA-based real-time PCR (qPCR) assay is explained in section 2.3.5.1. RNA-based techniques, including the optimisation of RNA extraction and reverse transcription as well as the design of a real-time reverse transcription PCR (qRT-PCR) assay are explained in sections 2.3.5.2 and 2.3.5.3, respectively.

2.3.5 Real-time PCR

qPCR has been found to be a reliable, accurate, rapid and user-friendly technique in identifying *SMN1* and *SMN2* copy numbers as an alternative to the PCR restriction enzyme SMA assay and MLPA (de Souza Godinho *et al.*, 2012; Maranda *et al.*, 2012) and has been implemented as a diagnostic assay in many laboratories worldwide.

There are two general qPCR detection methods: one using double strand DNA binding agents and a second using gene specific fluorescent probes with gene specific probe assays being much more accurate and reliable (Pfaffl 2004). There are two main methods of analysing real-time PCR data: absolute and relative quantification. Absolute quantification relates the fluorescent PCR amplification signal to the input copy number using a calibration curve while relative quantification measures the change in copy number versus a reference gene, similarly to intra-sample normalisation performed in MLPA. Various mathematical models are available to calculate the relative copy number/expression using threshold values (C_t). Both absolute and relative qPCR are performed as part of the ABI 7900HT-fast real-time PCR system software (Sequence Detection Software (SDS) and RQ-Manager) (Applied Biosystems, Foster City, CA, USA).

Two types of real-time PCR assays were designed as part of this project. The first DNA-based assay (real-time PCR (qPCR)) is specific for genomic quantification of *SMN1*, exon 7 and *SMN2*, exon 7 and was designed and optimised as part of this study as a cheaper, less laborious and faster alternative to MLPA. This assay was performed using the ABI 7900HT-fast real-time PCR system (Applied Biosystems, Foster City, CA, USA). The objective of this assay is to determine the *SMN1* and *SMN2* genomic copy number, simultaneously in a single assay (further explained in 2.3.51).

The second RNA-based assay (real-time reverse transcription PCR (qRT-PCR)) is specific for the quantification of the three different RNA transcripts of *SMN1* and *SMN2*: full length-SMN1 (FL-SMN1), full length-SMN2 (FL-SMN2) and SMN2 transcripts lacking exon 7 (SMNΔ7). This assay was designed but not optimised as part of this project to detect the expression of the SMN protein in U/U^B individuals. RNA extraction was validated as a proof of concept. The SMN transcript expression levels will be further investigated in a future MSc (Med) project.

2.3.5.1 Genomic real-time PCR (qPCR) – an alternative to MLPA

This qPCR assay for genomic quantification of *SMN1*, exon 7 and *SMN2*, exon 7 has been designed based on SMA Taqman probe assays previously described (de Souza Godinho *et al.*, 2012; Maranda *et al.*, 2012) and the primer and probe sequences are summarised in table 2.2.

Similarly to the P021-A2 MLPA probe mix, differently labelled fluorescent probes are designed to target the single critical nucleotide difference between exon 7 of *SMN1* and *SMN2* (c.840C>T), making it possible to distinguish between these two genes. A single primer pair (SMN-F and SMN-R) amplifies the *SMN1* and *SMN2* genes simultaneously followed by hybridisation with two differently fluorescently labelled probes, FAM-SMN1-Ex7 and VIC-SMN2-Ex7 binding to the *SMN1* and *SMN2* genes, respectively. Primers and probes specific to the 5' UTR region of *CFTR* (7q31) amplify a region of 62bp that acts as an endogenous control, similar to the reference probes in an MLPA kit.

Copy numbers are determined by comparing the fluorescence generated by the *SMN1* and *SMN2* probes with the fluorescence generated by the CFTR probe and compared to known control samples. See Figure 2.8 for an illustration of the binding sites of the *SMN1*, *SMN2* and *CFTR* primers and probes. See Appendix C for the SMA genomic qPCR worksheet.

Table 2.2: Primer and probe sequences of the DNA-based *SMN1* and *SMN2* qPCR assay

Gene	Primer	Sequence (5'-3')	Probe	Sequence (5'-3')
SMN1 144bp	SMN-F	CTTGTGAAACAAAATGCTTTTTAACATCCAT	SMN1-Ex7	FAM-R-TTTTGTCTGAAACCC
	SMN-R	GAATGTGAGCACCTTCCTTCTTTTT		
SMN2 144bp	Will be amplified by SMN-F/R primers		SMN2-Ex7	VIC-ATTTTGTCTAAAACCC
CFTR 62bp	CFTR-F	CAACCTGCCTTCTCTGGGAAT	CFTR	NED-CTGCTGCCTGAACAT
	CFR-R	CAAGCCTGGCAATAAACAATGA		

The one nucleotide difference (c.840C>T) between *SMN1* and *SMN2* is highlighted in blue.

Table 2.3: Primer and probe sequences of the RNA-based SMN expression qRT-PCR assay

Trans-cript	Primers	Sequence (5'-3')	Probe	Sequence (5'-3')
FL-SMN1 73bp	FL-SMN-F	TACATGAGTGGCTATCATACTGGCTA	FL-SMN1	NED-TATGGGTTTCAGACAAA
	FL-SMN-R	AATGTGAGCACCTTCCTTCTTTTT		
FL-SMN2 73bp	Same as FL-SMN1		FL-SMN2	VIC-ATATGGGTTTTAGACAAAA
Δ7SMN 78bp	Δ7SMN-F	CTGATGCTTTGGGAAGTATGTTAATT	Ex7del-SMN	NED-CATGGTACATGAGTGGCTA
	Δ7SMN-R	CCAGCATTTCTCCCATATAATAGCCAGTA		
GAPDH 74bp	GAPDH-F	GGGTGTGAACCATGAGAAGTATGA	GAPDH	FAM-CAAGATCATCAGCAATGC
	GAPDH-R	CTAAGCAGTTGGTGGTGCAGG		

The one nucleotide difference (c.840C>T) between FL-SMN1 and FL-SMN2 is highlighted in blue.

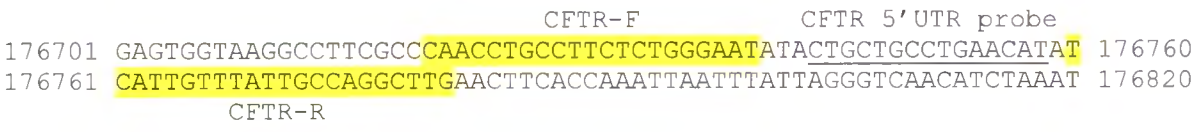
The *SMN1* target gene (144 bp)

(Ensembl genomic sequence: ENSG00000172062)

The *SMN2* target gene (144 bp)

(Ensembl genomic sequence: ENSG00000205571)

The *CFTR* endogenous control gene (62bp)



(Ensembl genomic sequence: ENSG00000001626) (Ensembl, 2016)

Figure 2.8: Location of *SMN1/SMN2* relative quantification primers and probes
The pink highlighted regions represent the location of the universal SMN1/2 primers (these primers bind to both *SMN1* and *SMN2*) and the yellow highlighted regions represent the location of the CFTR endogenous control primers. The underlined regions represent the location of the relevant qPCR probes. The blue highlighted region represents the critical nucleotide difference (c.840C>T) between *SMN1* and *SMN2* as well as their relevant probes.

2.3.5.2 RNA extraction and reverse transcription

RNA degrades rapidly and therefore fresh blood samples have to be collected from positive and negative controls in special formulated Tempus blood RNA collection tubes (Applied Biosystems, Foster City, CA, USA) to stabilise RNA and to preserve RNA expression patterns (Duale *et al.*, 2012). As part of validation, RNA was extracted using the Tempus Spin RNA Isolation Kit (Applied Biosystems, Foster City, CA, USA) using purification filters. See Tempus Blood RNA Tube and Tempus Spin

RNA Isolation Kit protocol for a step-by-step guide to RNA extraction using the Tempus blood RNA collection tubes (Thermo Fisher, 2016).

RNA was converted to its DNA complement (cDNA) using the two-step ImProm-II Reverse Transcription System (Promega, Madison, WI, USA). This technique involves denaturation of the RNA, followed by addition of a reverse transcription reaction containing an oligo (dT) primer and Taq polymerase (for further instructions see ImProm-II Reverse Transcription System). cDNA was stored for future qRT-PCR setup. This technique has been validated as part of this MSc project, as a proof of concept and will be applied in a future MSc (Med) project (see Appendix D for the Reverse transcription worksheet and step-by-step procedure).

2.3.5.3 Real-time reverse transcription PCR (qRT-PCR) – determining SMN expression

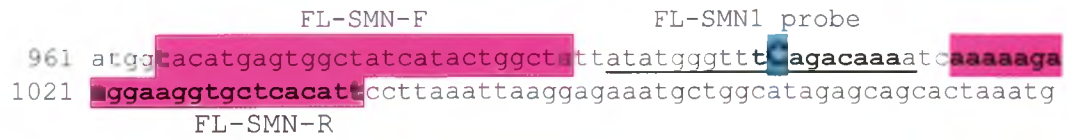
qRT-PCR is a particularly sensitive method for the quantification of gene expression. A qRT-PCR SMN expression assay was designed but not optimised as part of this project as a proof of concept. A future MSc (Med) project will involve collecting fresh blood and extracting RNA from patients to quantify the expression of *SMN1* and *SMN2* transcripts and comparing results with genomic CNV patterns. A qRT-PCR assay was designed based on assays described by Tiziano, Pinto, Fiori *et al.* (2010) and Zheleznyakova *et al.* (2011) to assess the quantity of SMN transcripts and to distinguish between the three SMN transcripts: FL-SMN1, FL-SMN2 and SMNΔ7 (summarised in Table 2.3, p.43).

Differently labelled fluorescent probes are designed to target the single nucleotide difference in exon 7 of the FL-SMN1 and FL-SMN2 transcripts (c.840C>T), making it possible to distinguish between these two transcripts and determining the genetic origin of the FL-SMN transcripts. A single primer pair amplifies the FL-SMN1 and FL-SMN2 transcripts simultaneously, followed by the three different fluorescent probes, FL-SMN1, FL-SMN2 and Ex7del-SMN, binding to the SMN1, SMN2 and Δ7SMN transcripts, respectively. The Ex7del-SMN probe will not bind to the FL-SMN1/2 transcripts since the primers have been designed to overlap the exon 6-exon 8 boundary and will therefore only bind to SMN transcripts excluding exon 7. Similarly,

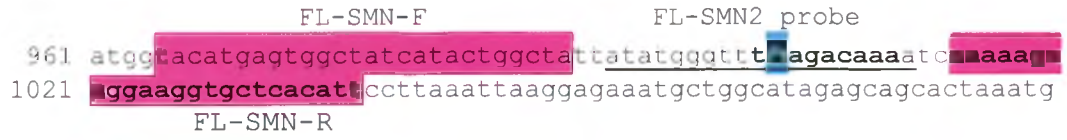
the primers and/or probes of the SMN expression qRT-PCR assay have been designed to overlap exon-exon boundaries to ensure that genomic DNA (which contains introns) will not be amplified and interfere with results.

Primers and probes specific to the GAPDH transcripts amplify a region that acts as an endogenous control. Expression levels of the different transcripts are determined comparing the fluorescence generated by the FL-SMN1, FL-SMN2 and Ex7del-SMN probes with the fluorescence generated by the GAPDH probe and compared to known control samples. See Figure 2.9 for an illustration of the binding sites of the FL-SMN1, FL-SMN2, Ex7del-SMN and GAPDH primers and probes.

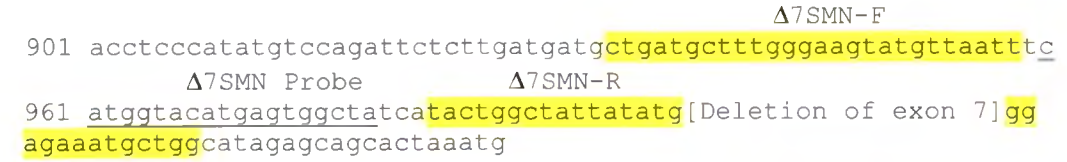
The full length SMN1 transcript (73 bp)



The full length SMN2 transcript (73bp)



The SMNΔ7 transcript (78bp)



The *GAPDH* endogenous control transcript (74 bp)

NCBI Reference Sequence: NM_000344.3, NM_017411.3, CCDS Sequence: 8549.1 (NCBI, 2016)

Figure 2.9: Location of SMN expression qRT-PCR primers and probes

The pink highlighted regions represent the location of the universal FL-SMN1/2 primers (these primers bind to both the FL-SMN1 and FL-SMN2 transcripts). The yellow highlighted regions represent the location of the primers specific to the Δ7SMN transcript (the Δ7SMN-R primer overlaps the exon 6-exon 8 cDNA boundary). The green highlighted regions represent the location of the GAPDH endogenous control primers. The underlined regions represent the locations of the relevant qRT-PCR probes. The blue highlighted regions represent the critical difference between the FL-SMN1 and FL-SMN2 transcripts and their relevant probes.

2.3.6 Complimentary DNA (cDNA) sequencing

An SMN cDNA sequencing assay using primers described by (Yu-Jin *et al.*, 2012) also has been designed as part of this MSc project as a proof of concept. The forward primer's (SMN-cDNA-F) sequence is: 5'-ATCCGCGGGTTTGCTATG-3 and the reverse primer's (SMN-cDNA-R) sequence is: 5'-ACTGCCTCACCACCGTGCTGG-3'. The primer pair produces a PCR product of 1177bp. Figure 2.10 illustrates the binding sites of the cDNA primers. A future MSc (Med) project may involve collecting blood from patients with heterozygous deletions of *SMN1*, exon 7 for RNA extraction and

performing cDNA sequencing in order to identify additional mutations. It is important to note that the cDNA sequence of *SMN1* and *SMN2* are identical except for the c.840C>T critical difference and therefore the gene origin of any pathogenic mutations identified cannot be determined.

NCBI Reference Sequence: NM_000344.3, NM_017411.3V (NCBI, 2016)

Figure 2.10: Location of SMN cDNA primers

The pink highlighted regions represent the location of the universal SMN-cDNA primers (these primers bind to both the SMN1 and SMN2 transcripts). The blue highlighted region represents the critical difference between the FL-SMN1 and FL-SMN2 transcripts. PCR product size: 1177bp.

2.4 ETHICS

An ethics application was approved unconditionally by the WITS Medical Human Research Ethics Committee (ethics clearance number: M130950, awarded 21/10/2013, valid for five years). See Appendix E for a copy of the signed ethics clearance certificate.

3 RESULTS

This chapter will present and explain the results of this project. Results from the audit on patients referred for diagnostic SMA testing will be shown and compared to data from the MLPA analysis. MLPA results were analysed using two different approaches. Firstly, MLPA data were analysed statistically to explore potential significant differences of CNV patterns among the different patient and subject groups studied and secondly, haplotypes were constructed in black families to investigate the inheritance pattern of CNVs in the SMN region in the black SA population.

The results of the audit will be summarised and explained in section 3.1 and the MLPA results for patient groups A-G will be presented and compared in section 3.2. Refer to section 2.2 and the list of definitions at the beginning of the dissertation for definitions. Section 3.3 will summarise results from general statistical analysis and section 3.4 will summarise and compare haplotypes from N/N^B and $M_1/M_1{}^B$ families. Provisional results as part of a proof of concept for RNA extraction and real-time PCR technology will be summarised in sections 3.5 and 3.6, respectively.

3.1 Results from audit on patients referred for diagnostic SMA testing

3.1.1 Descriptive statistics

In total 1 268 patients were referred for diagnostic SMA testing, of which 42.5% (539/1268) were female, 52.1% (661/1268) were male and for 5.4% (68/1268) the gender was not provided. Of these patients, 78.2% (991/1268) were black, 15.1% (192/1268) white, 3.9% (50/1268) Indian, 1.3% (16/1268) of mixed ancestry and 1.5% (19/1268) with an unknown ethnicity.

Twenty six individuals (10 black, 13 white, two mixed ancestry, one Indian) were referred for carrier testing and a total of 49 prenatal tests were performed in 38 families (including 21 white families, 13 black families and four Indian families).

Fifty patients referred for diagnostic SMA testing (41 black, six white, one mixed ancestry, two Indian) who previously tested negative for the homozygous deletion of SMN1, exon 7 were tested for a heterozygous deletion of SMN1, exon 7.

3.1.2 Homozygous deletions of *SMN1*, exon 7

Homozygous deletions of *SMN1*, exon 7 (telomeric) were identified in 31.6% (313/991) of black patients, 32.3% (62/192) of white patients, 32% (16/50) of Indian patients and 37.5% (6/16) of patients of mixed ancestry. These patients were all therefore confirmed to be affected with SMA. The rates of positive diagnosis did not differ significantly between genders (χ^2=0.6; p=0.44) and different ethnic population groups (χ^2=0.28; p=0.96) (see Appendix F, Statistical test output 1).

3.1.3 Homozygous deletions of *SMN2*, exon 7

Homozygous deletions of *SMN2*, exon 7 (centromeric) were identified in 12.4% (123/991) of black patients, 4.7% (9/192) of white patients, 4% (2/50) of Indian patients and 18.8% (3/16) of patients with mixed ancestry. There is a significantly higher percentage of *SMN2*, exon 7 deletions in black patients when compared to white patients (χ^2=11.64; p=0.000645). Patients of mixed ancestry did not differ significantly from black patients (χ^2=0.51; p=0.47). This finding will be further discussed in section 4.2.1.3 (see Appendix F, Statistical test output 2). Figure 3.1 summarises the homozygous *SMN1* and *SMN2*, exon 7 deletion percentages in SA populations.

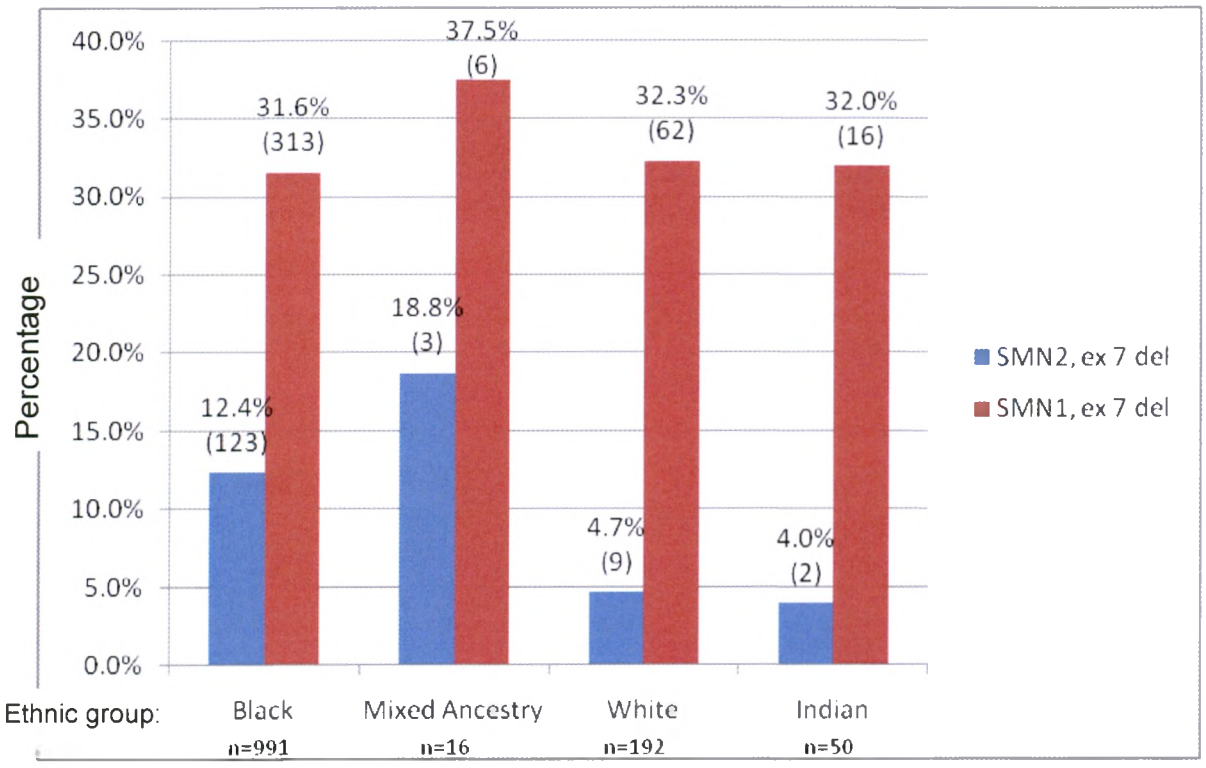


Figure 3.1: A bar graph comparing the percentages of homozygous *SMN1* and *SMN2*, exon 7 deletions in SA patients referred for SMA testing
The percentage of *SMN1*, exon 7 deletions do not differ significantly among different population groups. The percentage of *SMN2*, exon 7 deletions differ significantly, due to a high frequency of deletions in the black and mixed ancestry population groups.

3.1.4 Carrier, additional diagnostic and prenatal testing for SMA

A total of 30.8% (8/26) of individuals referred for carrier testing was found to be carriers of the common homozygous deletion *SMN1*, exon 7. Carriers of the deletion of *SMN1*, exon 7 were identified in 25% (3/12) of black individuals, 27.3% (3/11) of white individuals, 0% (0/1) of Indian individuals and 100% (2/2) of mixed ancestry individuals tested.

Fifty patients, who previously tested negative for the homozygous deletion of *SMN1*, exon 7, were tested for a heterozygous deletion of *SMN1*, exon 7 of which 24% (12/50) were found to test positive on the previously used dosage assay or MLPA. Heterozygous *SMN1*, exon 7 deletions were identified in 26.8% (11/41) of black individuals, 0% (0/6) of white individuals, 0% (0/2) of Indian individuals and 100% (1/1) of mixed ancestry individuals tested. The accuracy of the heterozygous deletion of *SMN1*, exon 7 results is discussed in sections 4.2.1.2 and 4.2.4.2.

Since SMA is an autosomal recessive disorder, two carriers have a risk of 25% with each pregnancy of having an affected child. A positive result (a homozygous deletion of *SMN1*, exon 7) was found in 22.4% (11/49) of prenatal tests, which compares well with the expected rate. *SMN1*, exon 7 homozygous deletions were identified in 18.8% (3/16) of black individuals, 19.2% (5/26) of white individuals, 33.3% (2/6) of Indian individuals and 100% (1/1) of mixed ancestry individuals tested.

The SMA database used for the audit is available on the CD accompanying this dissertation.

3.2 MLPA Results

This section will summarise and compare the MLPA results of this project. Comparisons will be made between black and white individuals from three different categories (negative controls (Groups B and C); individuals with homozygous deletions of *SMN1*, exon 7 (Groups D and E) and individuals with homozygous deletions of *SMN2*, exon 7 (Groups F and G)). Finally comparisons will be made between Group A and other black subject categories (Groups B, D and F). Refer to section 2.2 for definitions of the patient and subject groups.

Section 3.2.1 will compare results of subject groups B (N/N^B) and C (N/N^W) to establish normal CNV patterns in the black and white SA populations. Section 3.2.2 will compare results of subject groups D (M_1/M_1^B) and E (M_1/M_1^W) in order to determine and compare pathogenic haplotypes related with the homozygous deletion *SMN1*, exon 7 in black and white individuals with SMA. Section 3.2.3 will compare results of subject groups F (M_2/M_2^B) and G (M_2/M_2^W) to further understand the role of this common CNV in the black and white SA populations. Finally section 3.2.4 will compare results of subject group A (U/U) with the other sample groups to try to identify potential novel pathogenic CNVs.

Each comparison will look at the following different sections of the SMN region:

- ***SMN1*, exons 7, 8 and *NAIP*, exon 5**
- *SMN2*, exons 7 and 8
- Combined regions: *SMN1/2*, exons 1, 4, 6 and 8; *NAIP/NAIPΨ*, exon 13 and telomeric/centromeric-*GTF2H2*, exon 5
- Neighbouring genes: *SERF1B* and *RAD17*

NOTE

The SMN region is complex and due to the complex nature of the statistical analysis between multiple control and patient groups among multiple regions, it is recommended that the reader should focus on results and discussions pertaining to the functional telomeric SMN region including *SMN1*, exons 7, 8 and *NAIP*, exon 5. The neighbouring regions provide some insight into the nature of large CNVs of the SMN region but the main findings of this dissertation are focussed on *SMN1*, exons 7, 8 and *NAIP*.

Each sample group's results were analysed statistically as explained in section 2.3.3. The raw DQ value data generated by MLPA and Statistica output are available on the CD accompanying this dissertation.

3.2.1 Comparative analysis of Groups B (N/N^B) and C (N/N^W)

Groups B (N/N^B) and C (N/N^W) were tested to determine a baseline CNV pattern to compare with CNV patterns in other subject groups and to compare CNV trends of the black and white SA populations with worldwide reports. The results summarised in this section will be further discussed in section 4.2.1.

The non-parametric Kruskal-Wallis and median tests were performed using the Statistica software to determine which probes had a significant difference between N/N^B and N/N^W individuals. A significant difference ($p<0.05$) was observed for the majority of MLPA probes: *SMN1-7*, *SMN1-8*, *SMN2-8* (Kruskal-Wallis test only), *SMN1-1*, *SMN1-4*, *NAIP5*, *NAIP-13*, *GTF2H2-5*, *GTF2H2-8* and *GTF2H2-11)*, suggesting the presence of different population-specific CNVs and will be further discussed in the next few sections. Table 3.1 summarises the results of N/N^B and N/N^W individuals. The number and percentage of individuals who fall into different

CNV categories (one-ten copies) are indicated for each P021-A2 MLPA probe. Statistically significant differences are highlighted. See Figure 3.2 for examples of MLPA results.

Table 3.1: Comparison of CNVs between N/N^B and N/N^W individuals. Significant differences ($p<0.05$) between Groups B and C are highlighted.

Nr of copies	Group N/N^B: n=122 N/N^W: n=30	P021-A2 probes									Nr of copies of universal probe regions						
		RAD17	SERF1B	SMN2 Exon 7	SMN2 Exon 8	SMN1 Exon 7	SMN1 Exon 8	NAIP-5	GTF2H2-11 rs200357275	GTF2H2-8 rs201102513		SMN1/2 Exon 1	SMN1/2 Exon 4	SMN1/2 Exon 6	SMN1/2 Exon 8	NAIP-13	GTF2H2-5
0	N/N^B	0 (0.0%)	0 (0.0%)	15 (12.3%)	33 (27.0%)	0 (0.0%)	1 (0.8%)	1 (0.8%)	11 (9.0%)	0 (0.0%)	0	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	N/N^W	0 (0.0%)	0 (0.0%)	2 (6.7%)	2 (6.7%)	0 (0.0%)	0 (0.00%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.00%)	0 (0.0%)	0 (0.0%)
1	N/N^B	0 (0.0%)	1 (0.8%)	20 (16.4%)	55 (45.1%)	0 (0.0%)	0 (0.0%)	4 (3.3%)	37 (30.3%)	0 (0.0%)	2	14 (11.5%)	9 (7.4%)	3 (2.5%)	7 (5.7%)	9 (7.4%)	30 (24.6%)
	N/N^W	0 (0.0%)	0 (0.0%)	8 (26.7%)	12 (40.%)	0 (0.0%)	0 (0.00%)	1 (3.3%)	5 (16.7%)	0 (0.0%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (6.7%)	0 (0.0%)	0 (0.0%)
2	N/N^B	104 (85.2%)	120 (98.4%)	73 (59.8%)	31 (25.4%)	60 (49.2%)	54 (44.3%)	71 (58.2%)	74 (60.7%)	111 (91.0%)	4	106 (86.9%)	112 (91.8%)	118 (96.7%)	114 (93.4%)	112 (91.8%)	91 (74.6%)
	N/N^W	30 (100%)	30 (100%)	18 (60.0%)	14 (46.7%)	29 (96.7%)	28 (93.3%)	25 (83.3%)	25 (83,3%)	24 (80.0%)		30 (100%)	30 (100%)	30 (100%)	28 (93.3%)	30 (100%)	30 (100%)
3	N/N^B	17 (13.9%)	1 (0.8%)	10 (8.2%)	3 (2.5%)	21 (17.2%)	52 (42.6%)	29 (23.8%)	0 (0.0%)	8 (6.6%)	6	2 (1.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.8%)
	N/N^W	0 (0.0%)	0 (0.0%)	2 (6.7%)	2 (6.7%)	1 (3.3%)	2 (6.7%)	4 (13.3%)	0 (0.0%)	5 (16.7%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.00%)	0 (0.0%)	0 (0.0%)
4	N/N^B	1 (0.8%)	0 (0.0%)	3 (2.5%)	0 (0.0%)	35 (28.7%)	14 (11.5%)	14 (11.5%)	0 (0.0%)	3 (2.5%)	8	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	N/N^W	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (37.50%)	0 (0.0%)	0 (0.0%)	1 (3.3%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.00%)	0 (0.0%)	0 (0.0%)
5	N/N^B	0 (0.0%)	0 (0.0%)	1 (0.8%)	0 (0.0%)	6 (4.9%)	1 (0.8%)	3 (2.5%)	0 (0.0%)	0 (0.0%)	10	0 (0.0%)	1 (0.8%)	1 (0.8%)	1 (0.8%)	1 (0.8%)	0 (0.0%)
	N/N^W	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.00%)	0 (0.0%)	0 (0.0%)	0 (0.00%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.00%)	0 (0.0%)	0 (0.0%)

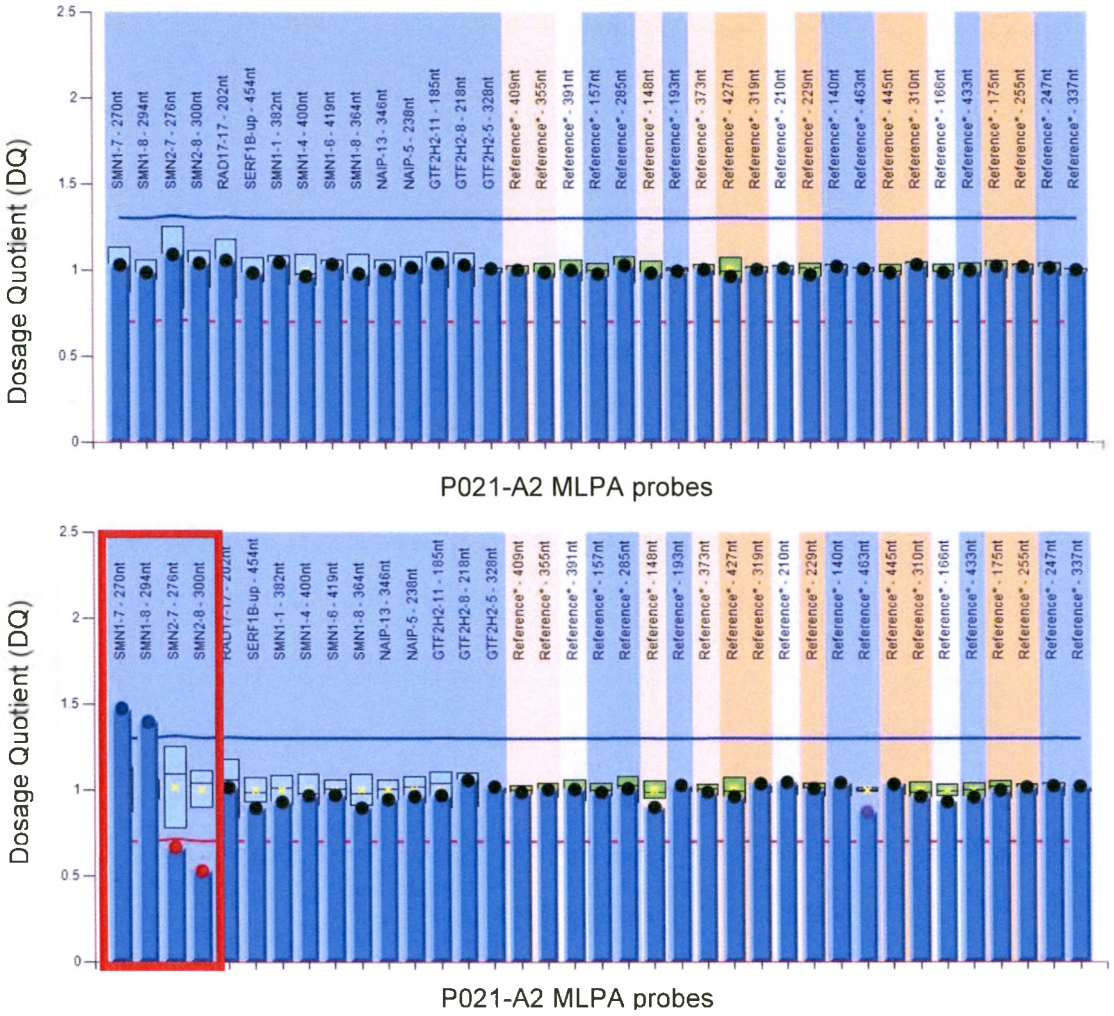


Figure 3.2: Examples of MLPA results of N/N^{B} and N/N^{W} individuals

a.) An example of an MLPA result from a N/N^{W} individual

All the probes have a ratio of 1, representing two copies of each of these regions. As can be seen from comparing this MLPA profile with figure b.), N/N^{B} individuals appear to have a higher variability and multiple copies of probes, when compared to N/N^{W} individuals.

b.) An example of an MLPA result from a N/N^{B} individual

All the reference probes have a ratio of 1, representing two copies of each of these regions.

The *SMN1*, exon 7 (SMN1-7) and exon 8 (SMN1-8) probes have a ratio of ~1.5 representing three copies of these regions, in conjunction with a ratio of ~0.5 of the *SMN2*, exon 7 (SMN2-7) and exon 8 (SMN2-8) probes, representing a heterozygous deletion, possibly due to a conversion from *SMN2*, exon 7, 8 to *SMN1*, exon 7, 8.

3.2.1.1 Telomeric SMN region: *SMN1*, exons 7, 8 and *NAIP*, exon 5

The SMN1 gene is the functional gene involved in SMA. The *SMN1*, exon 7 copy number was found to differ statistically between N/N^B and N/N^W individuals ($H=32.7$; $p<0.0001$). In this study, 50.8% (62/122) of N/N^B individuals were found to have multiple (three-six) copies of *SMN1*, exon 7. These results stand in sharp contrast to trends observed in N/N^W individuals, with only 3.3% (1/30) with multiple copies of *SMN1*, exon 7, supporting the hypothesis that black African individuals have multiple copies of *SMN1* more commonly than white individuals as previously reported by Hendrickson *et al.* and Sangaré *et al.* (Hendrickson *et al.*, 2009; Sangaré *et al.*, 2014).

Additionally the *SMN1*, exon 8 and *NAIP*, exon 5 copy numbers differed significantly between N/N^B and N/N^W individuals ($H=14.3$; $p=0.0002$ and $H=7.2$; $p=0.0071$, respectively). As seen with the *SMN1*, exon 7 region; 54.9% (67/122) of N/N^B individuals were found to have multiple copies of *SMN1*, exon 8 and 37.7% (46/122) were found to have multiple copies of *NAIP*, exon 5, which both co-locate with *SMN1* in the telomeric SMN region.

Once again, these results stand in contrast to trends observed in N/N^W individuals, with only 6.7% (2/30) of individuals with multiple copies of *SMN1*, exon 8 and only 13.3% (4/30) with multiple copies of *NAIP*, exon 5. As can be seen from figure 3.3, N/N^B individuals more often have multiple copies of the telomeric region (*SMN1*, exon7 and exon 8 and *NAIP*, exon 5) than N/N^W individuals.

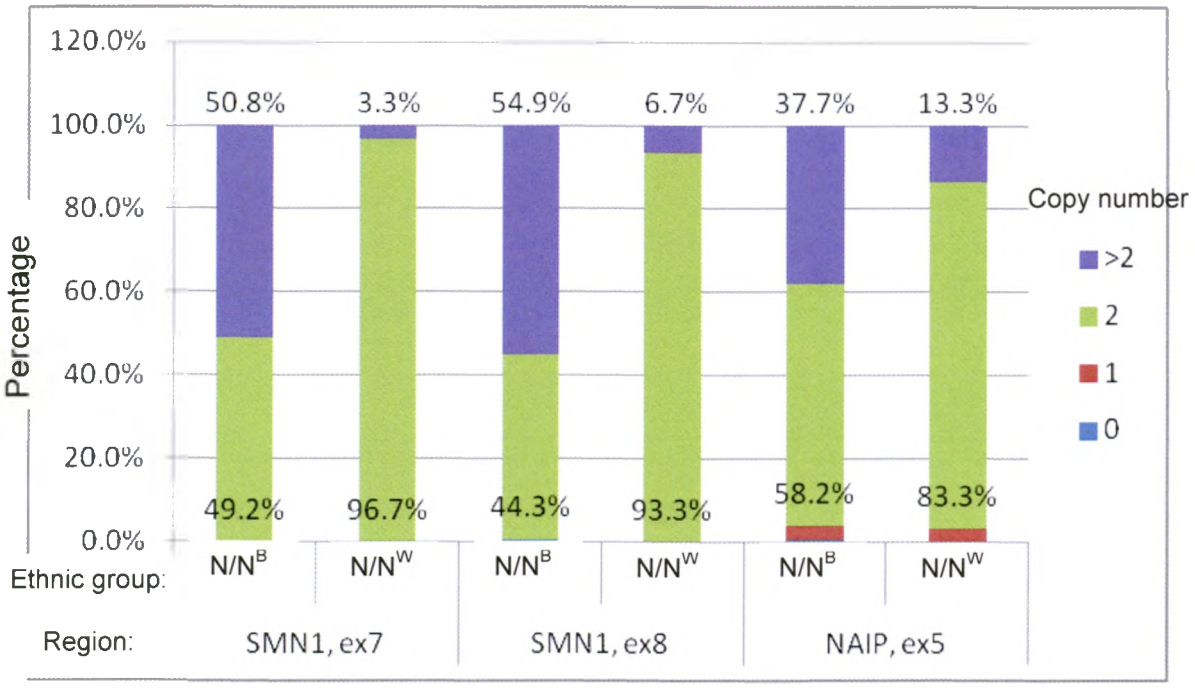


Figure 3.3: Comparison of the copy numbers of *SMN1*, exons 7 and 8 and *NAIP*, exon 5 between N/N^B and N/N^W individuals

The *SMN1*, exon 7 and 8 copy numbers do not correlate fully suggesting that some of these copies may not be contiguous and this concept will be further explained in section 4.2.1.5.

The *SMN1*, exon 7 results of N/N^B individuals and N/N^W individuals were compared to international data from North-America, Asia, Europe and Sub-Saharan Africa as summarised in table 3.2 and figure 3.4. There was no significant difference between the results of N/N^B individuals and individuals from Sub-Saharan Africa: Mali (χ^2=0.75; p=0.69), Nigeria (χ^2=5.45; p=0.07) and Kenya (χ^2=4.62; p=0.1) and African-American populations (χ^2=1.88; p=0.4). There also was no significant difference between results of Sub-Saharan African and African American populations (χ^2=6; p=0.2). See Appendix F, statistical output 3 for a breakdown of the statistical results.

There was no significant difference between the results of N/N^W individuals and individuals from Europe: Germany (χ^2=0.89; p=0.64), France (χ^2=0.73; p=0.69) and Sweden (χ^2=0.75; p=0.57) and American populations (χ^2=1.34; p=0.51). There also was no significant difference between results of European and American populations (χ^2=0.77; p=0.94). See Appendix F, statistical output 4 for a breakdown of the statistical results.

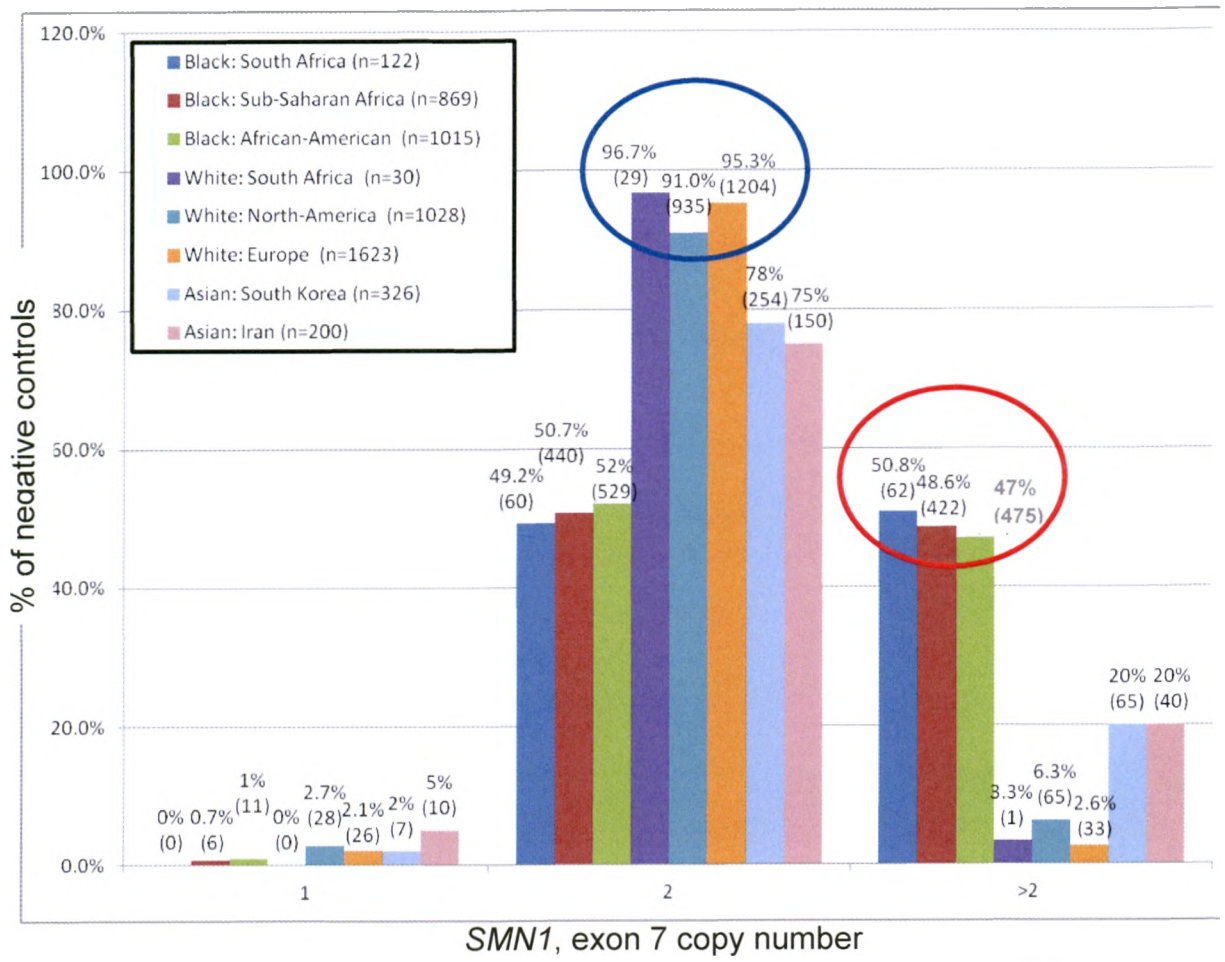
Table 3.2: *SMN1* copy number across various geographical populations

SMN1 copy number	SA		Sub-Saharan Africa			North-America		Europe			Asia	
	Black (N/N^B) n=122	White (N/N^W) n=30	Mali n=628	Nigeria n=120	Kenya n=120	African Americans n=1015	Caucasians n=1028	Germany n=140	France n=621	Sweden n=502	South-Korea n=326	Iran n=200
1	0% (0)	0% (0)	0.5% (3)	2% (2)	1% (1)	1% (11)	2.7% (28)	3% (4)	2% (13)	2% (9)	2% (7)	5% (10)
2	49.2 % (60)	96.7 % (29)	47% (295)	60% (72)	61% (73)	52% (529)	91% (935)	94% (132)	96% (593)	95% (479)	78% (254)	75% (150)
>2 (3-4)	50.8 % (62)	3.3% (1)	52.5 % (330)	38% (46)	38% (46)	47% (475)	6.3% (65)	3% (4)	2% (15)	3% (14)	20% (65)	20% (40)
Reference	This study		Sangaré *et al.*, 2014			Hendrickson *et al.*, 2009		Feldkötter *et al.*, 2002	Corcia *et al.*, 2012		Lee *et al.*, 2004	Hasanzad *et al.*, 2009

Adapted from (Sangaré *et al.*, 2014). High frequencies of multiple *SMN1*, exon 7 copies highlighted.

Figure 3.4 shows that N/N^W individuals have a much lower percentage of multiple *SMN1*, exon 7 copies of 3.3%, which is comparable to previous reports of 6.3% (65/1028) in white North-American populations (Hendrickson *et al.*, 2009) and a combined percentage of 2.6% (33/1263) in European populations (Germany, France and Sweden) (Corcia *et al.*, 2012; Feldkötter *et al.*, 2002) when compared to N/N^B individuals of which 50.8% have multiple (3-6) copies of *SMN1*, exon 7, which is similar to previous reports of 46.8% (475/1015) in African-American individuals (Hendrickson *et al.*, 2009) and a combined percentage of 48.6% (422/868) in sub-Saharan African populations (Mali, Nigeria and Kenya) (adapted from Sangaré *et al.*, 2014).

See section 4.2.1.1 for a detailed discussion of these results.



Graph produced from combined data of this study and a study performed by Sangaré et al., 2014.

Figure 3.4: Comparison of the *SMN1*, exon 7 copy number distribution of N/N^B and N/N^W individuals with various international studies

The blue circle indicates a similar percentage of white individuals with two copies of *SMN1*, exon 7 in SA, North-America and Europe. The red circle indicates a similar percentage of black individuals with multiple (more than two) copies of *SMN1*, exon 7 in SA, Sub-Saharan Africa and North-America.

3.2.1.2 Heterozygous *SMN1*, exon 7 deletion frequency

No N/N^B (0/122) and N/N^W (0/30) individuals appeared to have a detectable heterozygous deletion of *SMN1*, exon 7 (1:0 genotype), usually accepted to be carriers of SMA. Previous studies reported on a carrier rate of 1/50 and 1/23 in the black and white SA populations, respectively (Labrum *et al.*, 2007). These differences in carrier rates are most probably due to small sample sizes and the accuracy of the previous method used to detect the heterozygous deletion of *SMN1*, exon 7 and will be further discussed in section 4.2.1.2.

3.2.1.3 Centromeric SMN region: *SMN2*, exon 7 and 8

The *SMN2*, exon 7 copy number was not found to differ significantly between N/N^B and N/N^W individuals (H=0.09; p=0.7652). The majority of N/N^B and N/N^W individuals had two copies of *SMN2*, exon 7; 59.8% (73/122) and 60% (18/30), respectively.

The homozygous deletion of *SMN2*, exon 7 rate did not differ significantly (χ^2=0.2; p=0.7) between referred white patients included in the audit (4.7% (9/192)) and N/N^W individuals in this study (6.7% (2/30)). In contrast, the homozygous *SMN2*, exon 7 deletion rate differed significantly (χ^2=16.3; p=0.0000535) between black patients included in the audit (12.4% (123/991)) and N/N^B individuals included in this study (27% (33/122)) and could be due to a higher frequency of the homozygous deletion of *SMN2*, exon 7 rate in N/N^B individuals and will be further discussed in section 4.2.1.3.

The *SMN2*, exon 8 copy number was found to differ significantly between N/N^B and N/N^W individuals (H=11.1; p=0.0009). N/N^B individuals had a higher rate of homozygous *SMN2*, exon 8 deletions (27% (33/122) than N/N^W individuals (6.7% (2/30). This finding could be due to gene conversions from *SMN2*, exon 8 to *SMN1*, exon 8 resulting in hybrid genes consisting of *SMN1*, exon 7 and *SMN2*, exon 8 as explained in section 1.2.2.3. Deletions of these hybrid genes would result in a loss of *SMN1*, exon 7 in conjunction with *SMN2*, exon 8 (further discussed in sections 3.2.4.3 and 4.2.1.3).

Figure 3.5 compares the CNV results between N/N^B and N/N^W individuals, showing a different distribution of CNVs in *SMN2* exons 7 and 8 across both population groups.

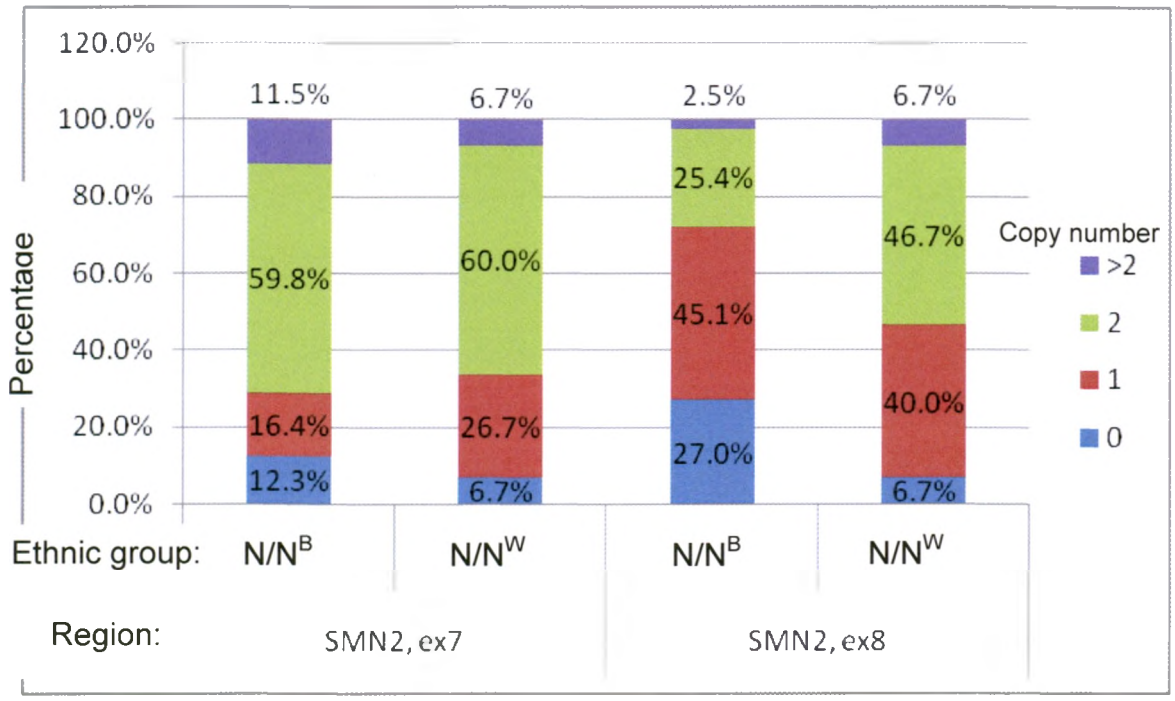


Figure 3.5: Comparison of the copy numbers of *SMN2*, exons 7 and 8 between N/N^B and N/N^W individuals

Section 3.1 has a detailed summary of results from the audit and the statistical results are available from Appendix F, statistical test output 5.

3.2.1.4 CNVs of combined regions: *SMN1/2*, exons 1, 4, 6 and 8; *NAIP*, exon 13 and *GTF2H2*, exon 5

The probes discussed in this section, bind to both the telomeric and centromeric copy of each of these genes (identical in probe binding regions) and will therefore give a combined result. Four copies represent a full complement of these regions.

The *SMN1/2*, exons 1 and 4 copy numbers were found to differ statistically between N/N^B and N/N^W individuals (H=7.5; p=0.0061 and H=6; p=0.0143, respectively). The *SMN1/2*, exons 6 and 8 copy numbers were not found to differ significantly between N/N^B and N/N^W individuals (H=0.4; p=0.5046 and H=0.002; p=0.9649, respectively). Although the *SMN1/2*, exons 1, 4, 6 and 8 regions mostly consisted of the full complement of four copies, deletions of these regions (50% (2/4 copies of each probe region deleted) were more frequently observed in N/N^B individuals (exon 1 (11.5% (14/122), exon 4 (7.4% (9/144)), exon 6 (2.5% (3/122) and exon 8 (5.7% (7/122)) when compared to N/N^W individuals (exon 1, 4 and 6: 0% (0/30), exon 8 (6.67%

(2/30)). The *SMN1/2*, exons 1, 4, 6 and 8 copy numbers do not correlate fully which, once again, may suggest that these exon copies may not all be contiguous. This finding is further dicussed in section 4.2.4.3.

The *GTF2H2*, exon 5 and *NAIP/NAIPΨ*, exon 13 copy numbers were found to differ statistically between N/N^B and N/N^W individuals (H=14.70; p=0.001 and H=8.10; p=0.0043, respectively). Although the *GTF2H2-5* and *NAIP/NAIPΨ*, exon 13 regions mostly consisted of the full complement of four copies, once again deletions (50% (2/4) of copies deleted) of these regions were more frequently observed in N/N^B individuals (24.6% (30/122) and 7.4% (9/122) respectively) than N/N^W individuals (0% (0/30) and 0% (0/30) respectively). Figure 3.6 shows the higher frequency of deletions of the *SMN1/2*, exons 1, 4, 6 and 8; *GTF2H2*, exon 5 and *NAIP/NAIPΨ*, exon 13 regions in N/N^B individuals when compared to N/N^W individuals.

Centromeric-*GTF2H2* (*GTF2H2-8*) and telomeric-*GTF2H2* (*GTF2H2-11*) copy numbers were found to differ significantly between Groups B (N/N^B) and C (N/N^W) (H=10.60; p=0.001 and H=9.10; p=0.003, respectively) and will be further explained and compared in Section 3.2.2.4.

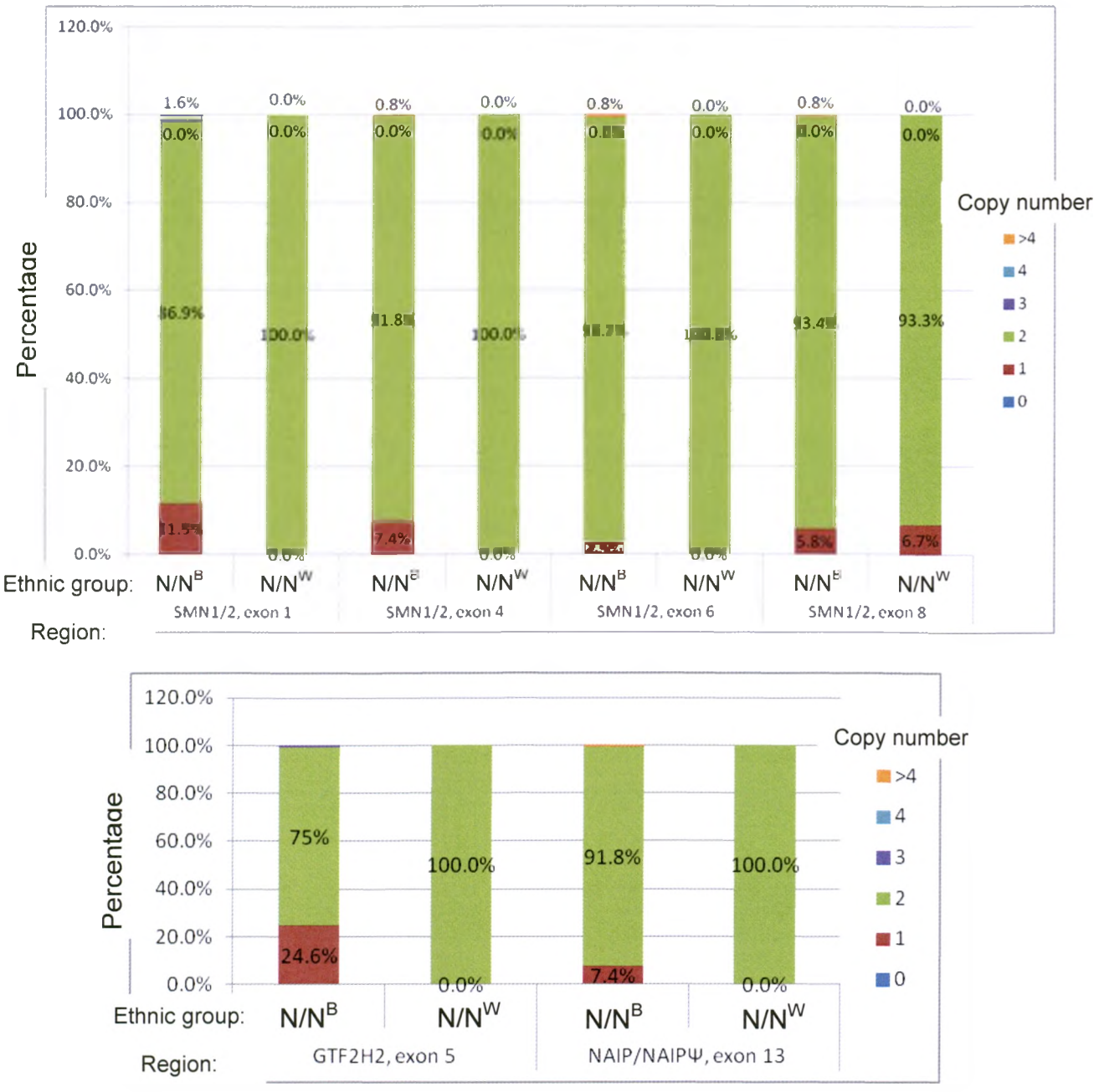


Figure 3.6: Comparison of the copy numbers of *SMN1/2*, exons 1, 4, 6 and 8; *GTF2H2*, exon 5 and *NAIP/NAIPΨ*, exon 13 between N/N^B and N/N^W individuals

3.2.1.5 CNVs of neighbouring genes: *SERF1B* and *RAD17*

The *SERF1B* and *RAD17* copy numbers were not found to differ significantly between Groups B (N/N^B) and C (N/N^W) (H=0.80; p=0.367 and H=0.00; p=0.967, respectively). Two copies of *SERF1B* were present in 98.4% (120/122) of N/N^B individuals. Multiple copies (more than 2) of *RAD17* were observed in 14.8% (18/122) of N/N^B individuals. No N/N^W individuals had deletions/duplications extending into *SERF1B* or *RAD17*,

suggesting that N/N^{B} individuals have a higher variability of these regions than N/N^{W} individuals.

3.2.1.6 Difference in variance between N/N^{B} and N/N^{W} individuals

The Kruskal-Wallis test is a non-parametric test of variance. N/N^{B} individuals had a significantly higher variability than N/N^{W} individuals for the majority (67.7% (10/15)) of MLPA probes: *SMN1*, exon 7; *SMN1*, exon 8; *NAIP5*; *SMN2*, exon 8; *NAIP13*; *SMN1/2*, exon 1; *SMN1/2*, exon 4; *GTF2H2-5*; *GTF2H2-8* and *GTF2H2-11*. Figure 3.7 shows scatterplots illustrating a strongly significant difference in variance between N/N^{B} and N/N^{W} individuals for the telomeric *SMN1* region. The significance and consequences of this observation is discussed in section 4.2.1.4. Significant results across other SMN regions are not shown here but are available in Appendix F, statistical test output 6.

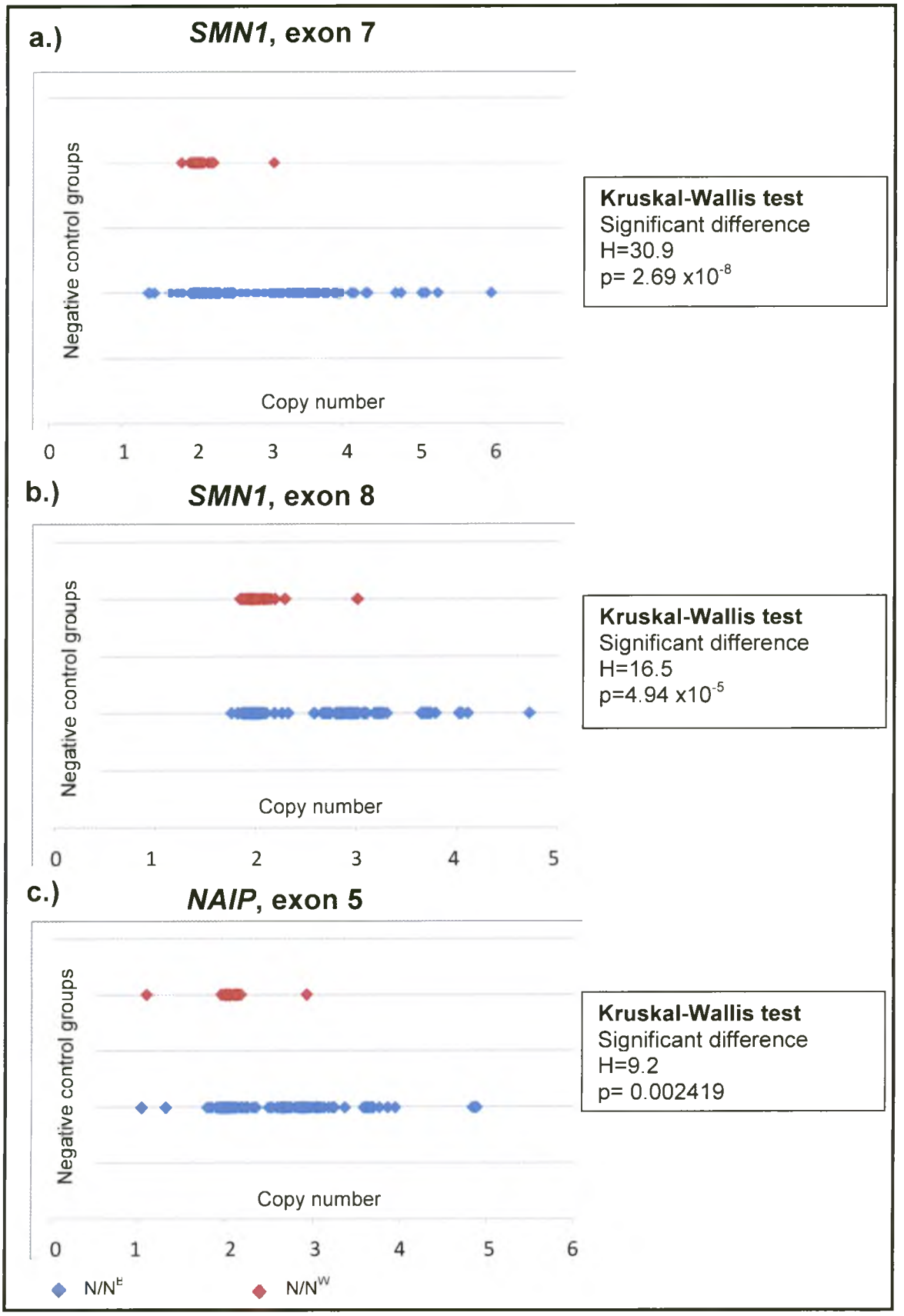


Figure 3.7: Scatterplot comparison of probe variances of a.) *SMN1*, exon 7; b.) *SMN1*, exon 8 and c.) *NAIP*, exon 5 between N/N^{B} and N/N^{W} individuals

3.2.2 Comparative analysis of Groups D (M_1/M_1^B) and E (M_1/M_1^W)

M_1/M_1^B and M_1/M_1^W individuals were tested by MLPA to confirm previous diagnostic results obtained by the diagnostic assay (explained in section 1.2.2.1) and to determine and compare pathogenic CNV profiles between these two population groups. The results summarised in this section will be further discussed in section 4.2.2.

The non-parametric Kruskal-Wallis and median tests were performed using the Statistica software to determine which probes had a different distribution between M_1/M_1^B and M_1/M_1^W individuals. A significant difference ($p<0.05$) was observed for the majority of MLPA probes: SMN1-8, *SMN2-7* (H=9, p=0.0027), *SMN1-4* (H=4.4, p=0.0355), *SMN1-6* (H=4.2, p=0.0399), *NAIP5* (H=5.6, p=0.0181), *NAIP13* (H=13.8, p=0.0002), *SERF1B* (H=3.9, p=0.0480), *GTF2H2-5* (H=10.8, p=0.0010), *GTF2H2-8* (H=12, p=0.0005), and *GTF2H2-11* (H=10.7, p=0.0011) and will be further discussed in the next few sections.

Table 3.3 summarises the results of M_1/M_1^B and M_1/M_1^W individuals. The number and percentage of individuals who fall into different CNV categories (one-ten copies) are indicated for each P021-A2 MLPA probe. Statistically significant differences are highlighted. See Figure 3.8 for typical examples of MLPA results of M_1/M_1^B and M_1/M_1^W individuals.

3.2.2.1 Telomeric SMN region: *SMN1*, exon 7, exon 8 and *NAIP*, exon 5

The SMN1 gene is the functional gene involved in SMA. As expected, MLPA analysis confirmed *SMN1*, exon 7 homozygous deletions in all 75 M_1/M_1^B individuals and all 30 M_1/M_1^W individuals.

The *SMN1*, exon 8 copy numbers were found to differ statistically between M_1/M_1^B and M_1/M_1^W individuals (H=7.8, p=0.0053). This significant difference could be due to homozygous *SMN1*, exon 8 deletions being more common in M_1/M_1^W individuals (zero copies: 80% (24/30); one copy: 13.3% (4/30)) than M_1/M_1^B individuals (zero copies: 50.7% (38/75); one copy: 29.3% (22/75)).

Table 3.3: Comparison of CNVs in M_1/M_1^B and M_1/M_1^W individuals. Significant differences ($p<0.05$) are highlighted.

Nr of copies	Ethnicity M_1/M_1^B: n=75 M_1/M_1^W: n=30	P021-A2 probes									Nr of copies of universal probe regions	SMN1/2 Exon 1	SMN1/2 Exon 4	SMN1/2 Exon 6	SMN1/2 Exon 8	NAIP-13	GTF2H2-5
		RAD17	SERF1B	SMN2 Exon 7	SMN2 Exon 8	SMN1 Exon 7	SMN1 Exon 8	NAIP-5	GTF2H2-11 rs200357275	GTF2H2-8 rs201102513							
0	M_1/M_1^B	0 (0.0%)	0 (0.0%)	0 (0.0%)	8 (10.7%)	75 (100%)	38 (50.7%	29 (38.7%)	56 (74.7%)	0 (0.0%)	0	1 (1.3%)	1 (1.3%)	1 (1.3%)	0 (0.0%)	8 (10.7%)	5 (6.7%)
	M_1/M_1^W	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	30 (100%)	24 (80.0%)	12 (40.0%)	15 (50.0%)	0 (0.0%)		2 (6.7%)	2 (6.7%)	2 (6.7%)	2 (6.7%)	3 (10.0%)	0 (0.0%)
1	M_1/M_1^B	0 (0.0%)	26 (34.7%)	4 (5.3%)	15 (20.0%)	0 (0%)	22 (29.3%)	16 (21.3%)	8 (10.7%)	6 (8.0%)	2	45 (60.0%)	46 (61.3%)	47 (62.7%)	46 (61.3%)	46 (61.3%)	50 (66.7%)
	M_1/M_1^W	0 (0.0%)	8 (26.7%)	2 (6.7%)	5 (16.7%)	0 (0.0%)	4 (13.3%)	17 (56.7%)	12 (40.0%)	0 (0.0%)		10 (33.3%)	10 (33.3%)	10 (33.3%)	11 (36.7%)	7 (23.3%)	15 (50.0%)
2	M_1/M_1^B	75 (100.0%)	49 (65.3%)	63 (84.0%)	49 (65.3%)	0 (0%)	12 (16.0%)	22 (29.3%)	11 (14.7%)	67 (89.3%)	4	29 (38.7%)	28 (37.3%)	27 (36.0%)	29 (38.7%)	21 (28.0%)	20 (26.7%)
	M_1/M_1^W	30 (100.0%)	22 (73.3%)	17 (56.7%)	13 (43.3%)	0 (0.0%)	2 (6.7%)	1 (3.3%)	3 (10.0%)	29 (96.7%)		18 (60.0%)	18 (60.0%)	18 (60.0%)	17 (56.7%)	20 (66.7%)	15 (50.0%)
3	M_1/M_1^B	0 (0.0%)	0 (0.0%)	5 (6.7%)	3 (4.0%)	0 (0%)	2 (2.7%)	7 (9.3%)	0 (0.0%)	2 (2.7%)	6	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	M_1/M_1^W	0 (0.0%)	0 (0.0%)	10 (33.3%)	11 (36.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (3.3%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
4	M_1/M_1^B	0 (0.0%)	0 (0.0%)	3 (6.7%)	0 (0.0%)	0 (0%)	1 (1.3%)	1 (1.3%)	0 (0.0%)	0 (0.0%)	8	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	M_1/M_1^W	0 (0.0%)	0 (0.0%)	1 (3.3%)	1 (3.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
5	M_1/M_1^B	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	10	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	M_1/M_1^W	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

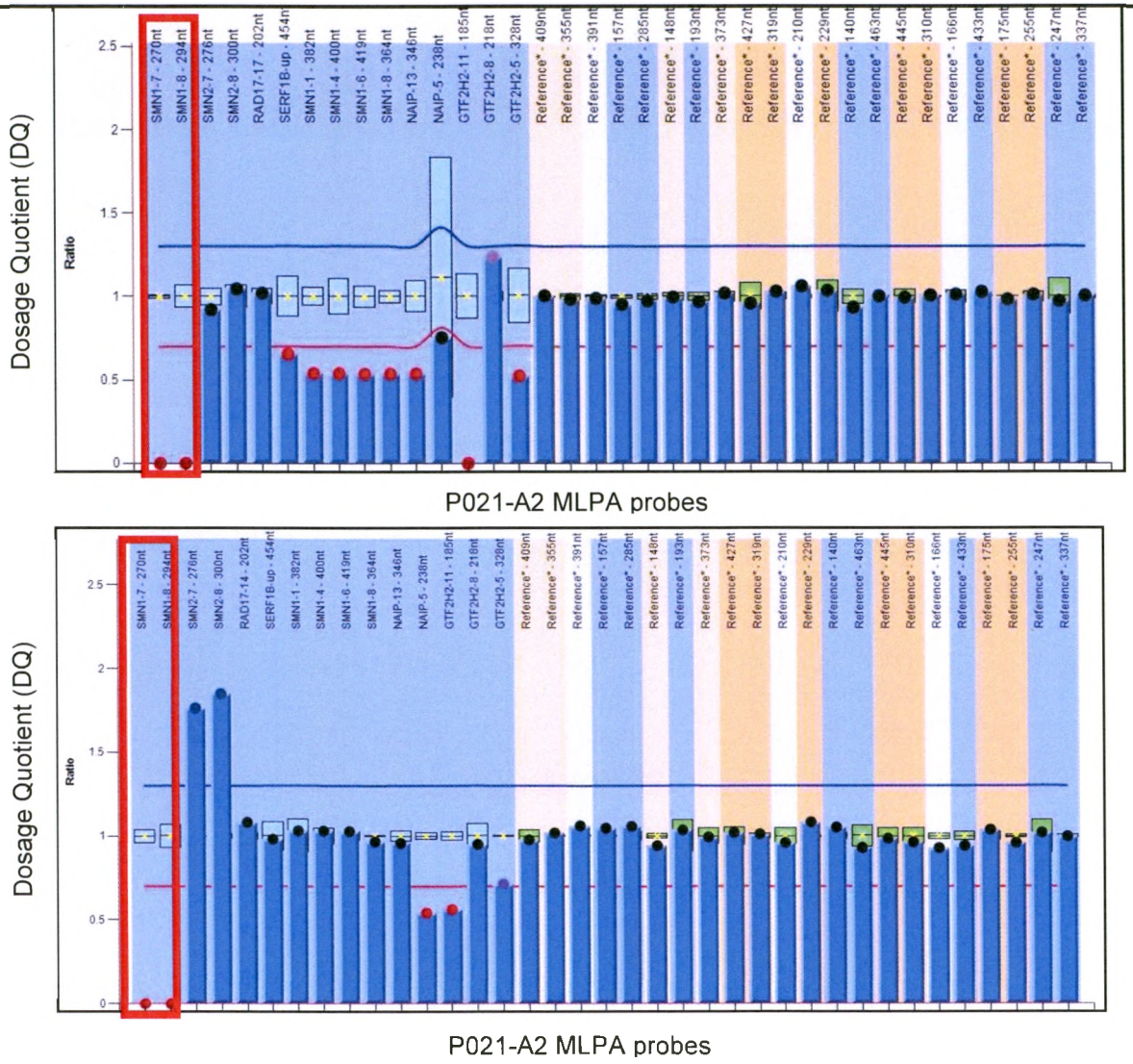


Figure 3.8: Examples of MLPA results in M_1/M_1^B and M_1/M_1^W individuals

a.) An example of an MLPA result for a M_1/M_1^B individual

All the reference probes have a ratio of 1, representing two copies of each of these regions. The reference probes are used to normalise SMN probe results. The *SMN1*, exon 7 (SMN1-7) and exon 8 (SMN1-8) probes have a ratio of 0 (indicated with a red box) representing a homozygous deletion of these regions. This deletion appears to stretch into exons 1 (SMN1-1), 4 (SMN1-4), 6 (SMN1-6), and 8 (SMN1-8) of *SMN1* since the probes have a ratio of ~0.5 most probably representing a homozygous deletion of *SMN1*. These probes cannot distinguish between *SMN1* and *SMN2*. The deletion further appears to extend into *GTF2H2* (ratio of 0.5).

b.) An example of an MLPA result for a M_1/M_1^W individual

The *SMN1*, exon 7 (SMN1-7) and exon 8 (SMN1-8) probes have a ratio of 0 representing a homozygous deletion (indicated with a red box) The *SMN2*, exon 7 (SMN2-7) and exon 8 (SMN2-8) probes have a ratio of ~1.75 representing four copies of these regions, possibly due to a full conversion from *SMN1*, exon 7, 8 to *SMN2*, exon 7, 8. The *SMN1/2* probes have a ratio of 1, representing four copies suggesting that the *SMN1*, exon 7 deletion is limited to *SMN1*, exon 7, therefore likely being a conversion rather than a large deletion of the entire *SMN1* gene.

Table 3.4 compares the homozygous *SMN1*, exon 7, 8 and *NAIP*, exon 5 deletion results of the current study with data from a previous SA study (Labrum *et al.*, 2007). Results from this study did not differ significantly from results from the previous study in either M_1/M_1^W (χ^2=0.9, p=0.8) or M_1/M_1^B individuals (χ^2=1.6, p=0.7) but this could be due to an overlap of 27 M_1/M_1^B individuals between these two studies (see Appendix F, statistical test output 9 for an explanation of the calculations).

When only looking at deletion results of *SMN1*, exon 7 and 8 and *NAIP*, it appears as if M_1/M_1^W individuals have larger deletions including *SMN1*, exon 7 and 8 and *NAIP* more frequently with a frequency of 36.7-43.5%, than M_1/M_1^B individuals with a frequency of 21.3-28%. This phenomenon will be further discussed in section 4.2.2.1.

Table 3.4: The profile of homozygous deletions of *SMN1*, exon 7 in SA populations. Highligted results indicate a higher frequency of large deletions in M_1/M_1^B

Study	**Labrum *et al.* (2003)**		**This study**	
Patient category	**M_1/M_1^W (n=23)**	**M_1/M_1^B (n=47)**	**M_1/M_1^W (n=30)**	**M_1/M_1^B (n=75)**
SMN1, exon 7 deletion	5 (21.7%)	18 (38.3%)	7 (23.3%)	29 (38.6%)
SMN1, exon 7 and 8 deletions	8 (34.8%)	15 (31.9%)	11 (36.7%)	17 (22.7%)
SMN1, exon 7 and *NAIP* deletions (*SMN1*, exon 8 is present)	0 (0%)	4 (8.5%)	1 (3.3%)	8 (10.7%)
SMN1, exon 7 and 8 and *NAIP* deletions	10 (43.5%)	10 (21.3%)	11 (36.7%)	21 (28%)

Furthermore *SMN1*, exon 7 and 8 copy numbers do not correlate fully suggesting that these copies may not be contiguous.

Additionally the *NAIP*, exon 5 copy numbers were found to differ statistically between M_1/M_1^B and M_1/M_1^W individuals (H=5.6, p=0.0181). Heterozygous deletions of *NAIP*, exon 5 (which co-locates with *SMN1*) were more common in M_1/M_1^W individuals (1 copy: 56.7% (17/30)) than M_1/M_1^B individuals (1 copy: 21.3% (16/75)).

Figure 3.9 shows a comparison of *SMN1*, exon 8 and *NAIP* CNVs between M_1/M_1^B and M_1/M_1^W individuals, showing a higher frequency of deletions in M_1/M_1^W individuals.

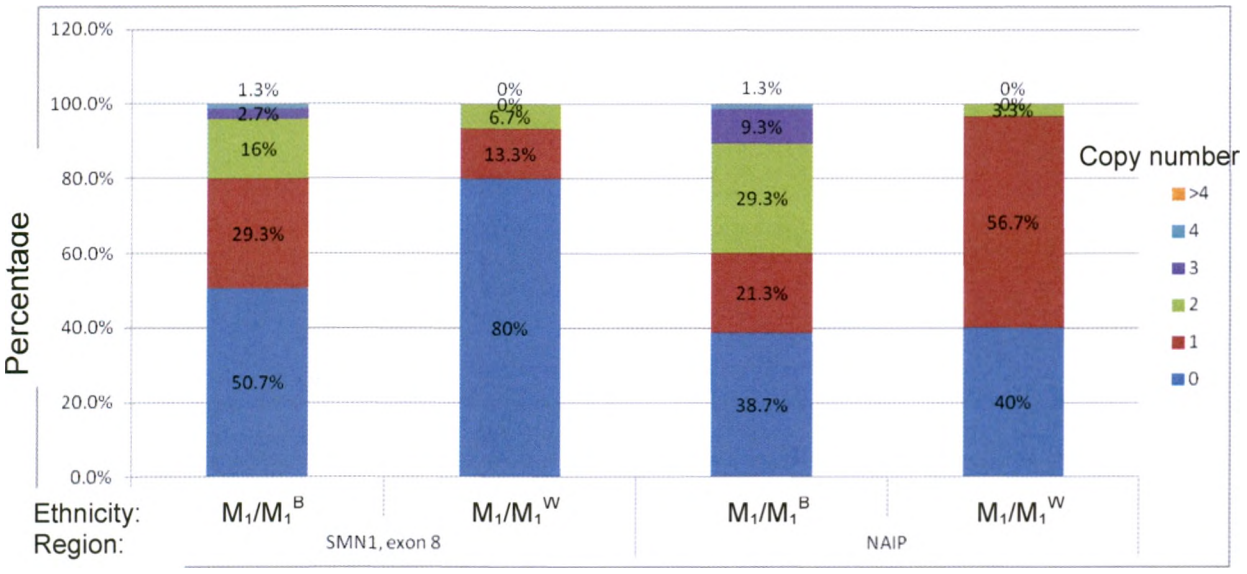


Figure 3.9: Comparison of the copy numbers of *SMN1*, exon 8 and *NAIP* CNVs between M_1/M_1^B and M_1/M_1^W individuals

3.2.2.2 Centromeric SMN region: *SMN2*, exon 7 and 8

The *SMN2*, exon 7 copy number was found to differ significantly between M_1/M_1^B and M_1/M_1^W individuals (H=9, p=0.0027), likely due to a higher frequency of multiple copies of this region in M_1/M_1^W individuals (three copies: 33.3% (10/30), four copies: 3.3% (1/30)) when compared to M_1/M_1^B individuals (three copies: 6.7% (5/75), four copies: 4% (3/75)). Multiple copies of *SMN2*, exon 7 in conjunction with homozygous deletions of *SMN1*, exon 7 suggest gene conversion from *SMN1*, exon 7 to *SMN2*, exon 7 being more common in M_1/M_1^W individuals.

The *SMN2*, exon 8 copy number was not found to differ significantly between M_1/M_1^B and M_1/M_1^W individuals (H=3.1; p=0.0788), although once again a higher frequency of multiple copies of this region was observed in M_1/M_1^W individuals (three copies: 36.7% (11/30), four copies: 3.3% (1/30)) when compared to M_1/M_1^B individuals (three copies: 4% (4/75), four copies: 0% (0/75)). Once again the *SMN2*, exon 7 and 8 copy numbers do not correlate possibly due to these regions not being contiguous and will be further discussed in sections 3.2.4.3 and 4.2.1.3.

Figure 3.10 shows a comparison of *SMN2*, exons 7 and 8 CNVs between M_1/M_1^B and M_1/M_1^W individuals with a higher frequency of multiple copies of these regions observed in M_1/M_1^W individuals when compared to M_1/M_1^B individuals.

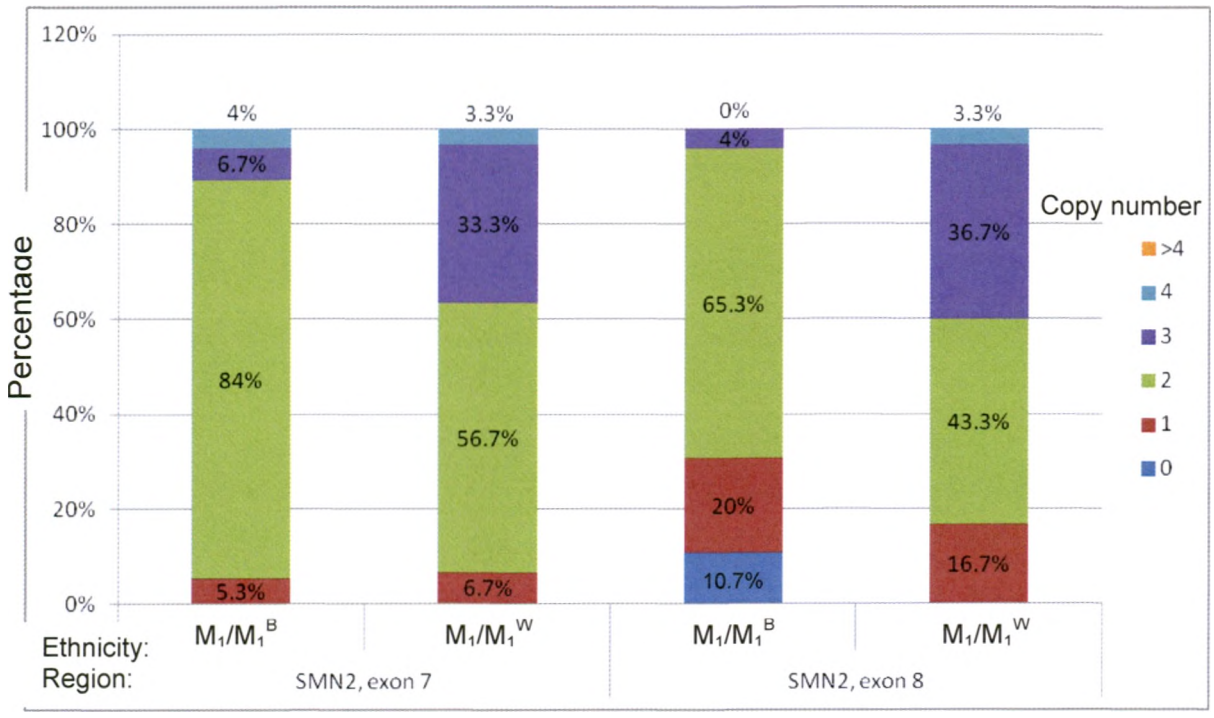


Figure 3.10: Comparison of the copy numbers of *SMN2*, exons 7 and 8 between M_1/M_1^B and M_1/M_1^W individuals

Interestingly, a homozygous deletion of *SMN2*, exon 8 is more commonly observed in 10.7% (8/75) of M_1/M_1^B individuals in comparison to 0% (0/30) in M_1/M_1^W individuals. This finding could potentially be caused by the formation of hybrid genes during gene conversion, as explained in section 3.2.1.3.

To summarise, by just looking at the *SMN1* and *SMN2* regions, it appears as if M_1/M_1^W individuals have a higher frequency of large *SMN1* deletions including *SMN1*, exon 8 and *NAIP*, exon 5 as well as a higher frequency of gene conversions from *SMN1* to *SMN2* resulting in multiple copies of *SMN2* when compared to M_1/M_1^B individuals.

3.2.2.3 CNVs of combined regions: *SMN1/2*, exons 1, 4, 6 and 8; *NAIP/NAIPΨ*, exon 13 and telomeric/centromeric *GTF2H2*, exon 5

The probes discussed in this section, bind to both the telomeric and centromeric copy of each of these genes (identical in probe binding regions) and will therefore give a combined result. Four copies represent a full complement of these regions.

The *SMN1/2*, exons 4 and 6 copy numbers were found to differ statistically between M_1/M_1^B and M_1/M_1^W individuals (H=4.4, p=0.0355 and H=4.2, p=0.0399, respectively). The *SMN1/2*, exons 1 and 8 copy numbers were not found to differ significantly between M_1/M_1^B and M_1/M_1^W individuals (H=1; p=0.3188 and H=1.3; p=0.2485, respectively). Although the *SMN1/2*, exons 1, 4, 6 and 8 regions mostly consisted of the full complement of four copies, a homozygous deletion of these regions (2/4 copies deleted) of *SMN1* and/or *SMN2* were more frequently observed in M_1/M_1^B individuals (exon 1: 60% (45/75), exon 4: 61.3% (46/75), exon 6: 62.7% (47/75) and exon 8: 61.3% (46/75)) than M_1/M_1^W individuals (exon 1, 4 and 6: 31.3% (10/30) and exon 8: 36.7% (11/30)). Once again, the *SMN1/2*, exons 1, 4, 6 and 8 copy numbers do not correlate fully with each other or with the *SMN1*, exon 7 and 8 copy numbers in M_1/M_1^B or M_1/M_1^W individuals suggesting that these copies may not be contiguous (discussed in section 4.2.1.3).

The *GTF2H2*, exon 5 and *NAIP/NAIPΨ*, exon 13 copy numbers were found to differ statistically between M_1/M_1^B and M_1/M_1^W individuals (H=10.8, p=0.001 and H=13.8, p=0.0002, respectively). Although the *GTF2H2*, exon 5 and *NAIP/NAIPΨ*, exon 13 regions mostly consisted of four copies, a full gene deletion (2/4 copies deleted) of *GTF2H2*, exon 5 and *NAIP/NAIPΨ*, exon 13 were more frequently observed in M_1/M_1^B individuals (66.7% (50/75) and 61.3% (46/75) respectively) than M_1/M_1^W individuals (50% (15/30) and 23.3% (7/30) respectively). Figure 3.11 shows the higher frequency of deletions of the *SMN1/2*, exons 1, 4, 6 and 8; *GTF2H2*, exon 5 and *NAIP/NAIPΨ*, exon 13 regions in M_1/M_1^B individuals when compared to M_1/M_1^W individuals.

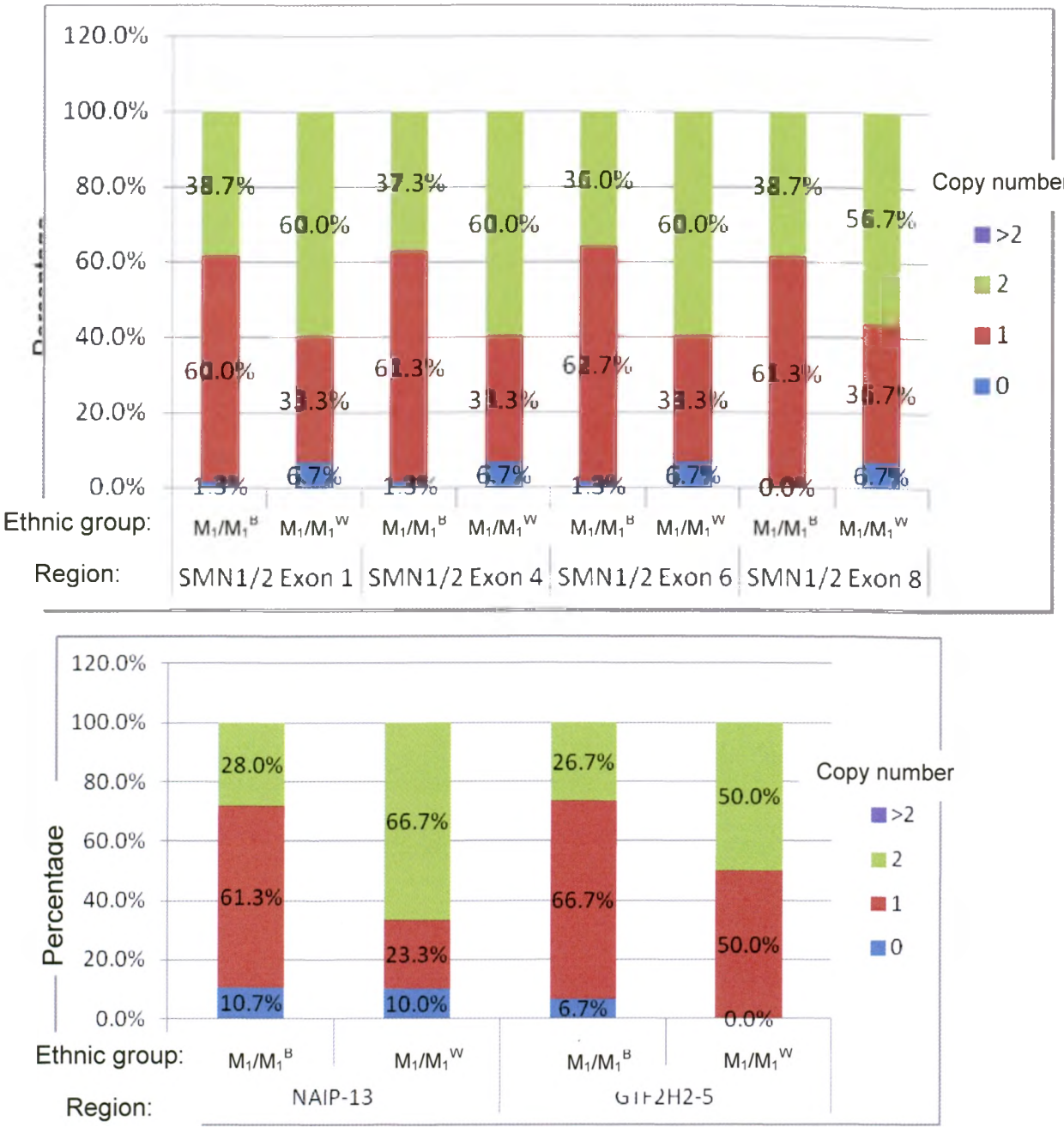


Figure 3.11: Comparison of the copy numbers of *SMN1/2*, exons 1, 4, 6 and 8; *GTF2H2*, exon 5 and *NAIP/NAIPΨ*, exon 13 between M_1/M_1^B and M_1/M_1^W individuals

Centromeric-*GTF2H2* (*GTF2H2-8*) and telomeric-*GTF2H2* (*GTF2H2-11*) copy numbers were found to differ significantly between M_1/M_1^B and M_1/M_1^W individuals (H=12, p=0.0005 and H=10.7, p=0.0011, respectively) and will be further explained and compared in Section 3.2.4.4.

3.2.2.4 CNVs of neighbouring genes: *SERF1B* and *RAD17*

SERF1B copy numbers were found to differ significantly between M_1/M_1^B and M_1/M_1^W individuals (H=3.9, p=0.0480), most likely due to a higher frequency of heterozygous deletions of *SERF1B* in 34.7% (26/75) of M_1/M_1^B individuals compared to 0% (0/30) in M_1/M_1^W individuals (illustrated in figure 3.12).

There was no difference in copy number of *RAD17* between M_1/M_1^B and M_1/M_1^W individuals.

Figure 3.12: Comparison of the copy numbers of *SERF1B* between M_1/M_1^B and M_1/M_1^W individuals

Results of the *SMN1/2*, exons 1, 4, 6 and 8; *GTF2H2*, exon 5; *NAIP/NAIPΨ*, exon 13 and *SERF1B* copy numbers suggest that large homozygous deletions of *SMN1*, exon 7 extending into neighbouring genes in the SMN region is more common in M_1/M_1^B individuals when compared to M_1/M_1^W individuals. These results appear to contrast results from section 3.2.2.1, which suggests that M_1/M_1^W individuals have larger homozygous deletions of *SMN1*, exon 7 extending into *SMN1*, exon 8 and *NAIP*, exon 5 when compared to M_1/M_1^B individuals. This inconsistency will be explained and discussed in section 4.2.2.1.

3.2.3 Comparative analysis of Groups F (M_2/M_2^B) and G (M_2/M_2^W)

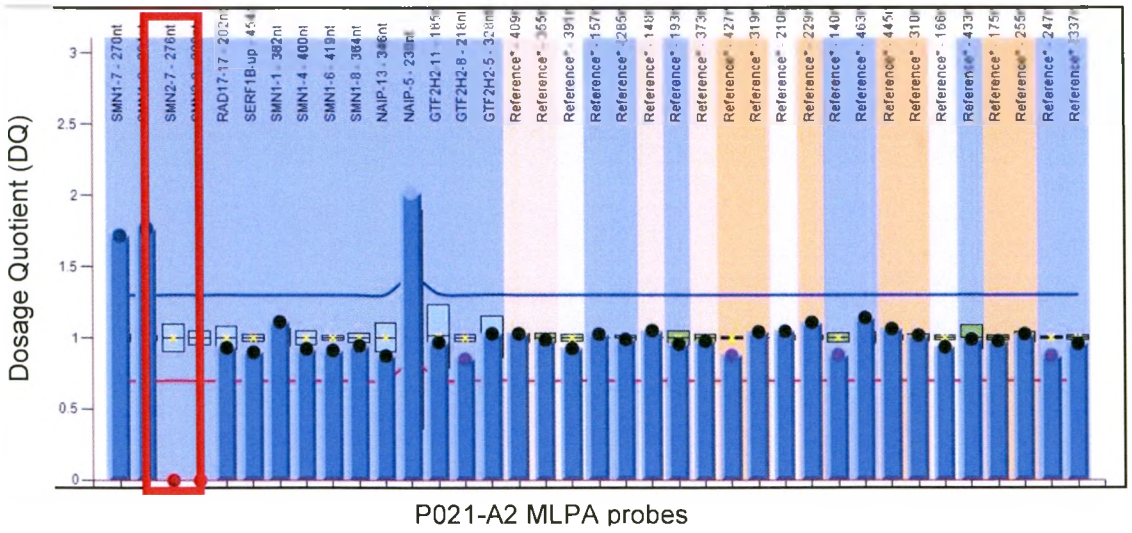
M_2/M_2^B and M_2/M_2^W individuals were tested by MLPA to confirm previous diagnostic results obtained by the diagnostic assay (explained in section 1.2.2.1); to determine the underlying molecular structure of this common CNV and to understand the interaction between the SMN1 and SMN2 genes. The results summarised in this section will be further discussed in section 4.2.3.

The non-parametric Kruskal-Wallis and median tests were performed using the Statistica software to determine which probes had a significant difference between M_2/M_2^B and M_2/M_2^W individuals. A significant difference was observed for only three MLPA probes: *NAIP-5*, *NAIP-13* and *GTF2H2-8* and will be further discussed in the next few sections.

Table 3.5 summarises the results of M_2/M_2^B and M_2/M_2^W individuals. The number and percentage of individuals who fall into different CNV categories (one-ten copies) are indicated for each P021-A2 MLPA probe. Statistically significant differences are highlighted. See Figure 3.13 for examples of MLPA results.

Table 3.5: Comparison of CNVs in M_2/M_2^B and M_2/M_2^W individuals. Significant differences ($p<0.05$) are highlighted.

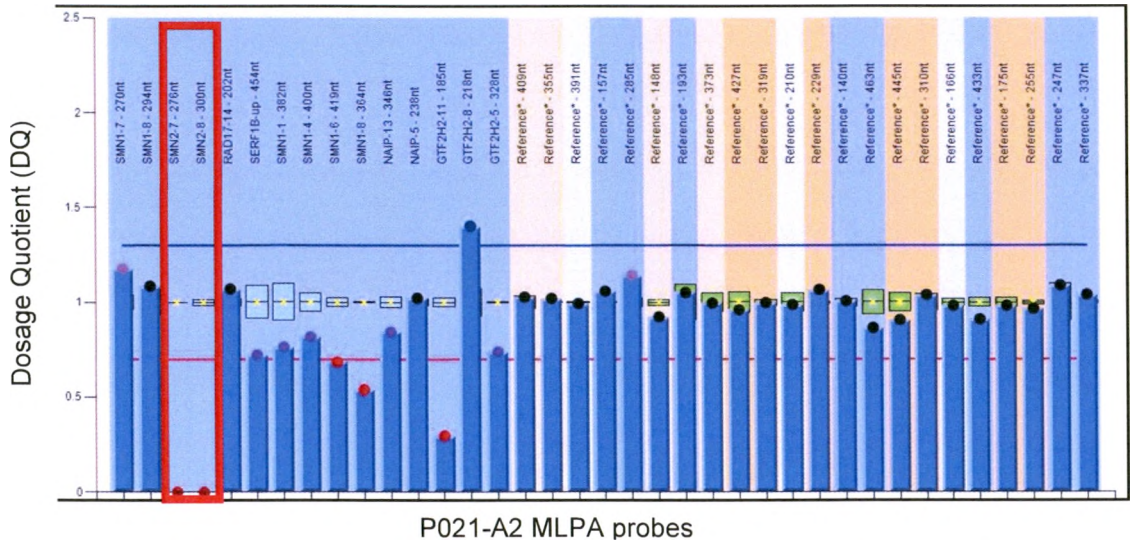
Nr of copies	Ethnicity M_2/M_2^B n=122 M_2/M_2^W n=30	P021-A2 probes															
		RAD17	SERF1B	SMN2 Exon 7	SMN2 Exon 8	SMN1 Exon 7	SMN1 Exon 8	NAIP-5	GTF2H2-11 rs200357275	GTF2H2-8 rs201102513	Nr of copies of universal probe regions	SMN1/2 Exon 1	SMN1/2 Exon 4	SMN1/2 Exon 6	SMN1/2 Exon 8	NAIP-13	GTF2H2-5
0	M_2/M_2^B	0 (0.0%)	0 (0.0%)	50 (100.0%)	49 (98.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (6.0%)	0 (0.0%)	0	1 (2.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	M_2/M_2^W	0 (0%)	0 (0%)	8 (100%)	8 (100%)	0 (0%)	0 (0%)	0 (0%)	1 (12.5%)	0 (0%)		0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
1	M_2/M_2^B	0 (0.0%)	10 (20.0%)	0 (0.0%)	1 (2.0%)	2 (4.0%)	0 (0.0%)	0 (0.0%)	13 (26.0%)	0 (0.0%)	2	9 (18.0%)	16 (32.0%)	17 (34.0%)	16 (32.0%)	22 (44.0%)	3 (6.0%)
	M_2/M_2^W	0 (0%)	1 (12.5%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (12.5%)	3 (37.5%)	0 (0%)		1 (12.5%)	1 (12.5%)	1 (12.5%)	5 (62.5%)	0 (0%)	0 (0%)
2	M_2/M_2^B	50 (100.0%)	40 (80.0%)	0 (0.0%)	0 (0.0%)	16 (32.0%)	21 (42.0%)	5 (10.0%)	34 (68.0%)	50 (100.0%)	4	39 (78.0%)	34 (68.0%)	33 (66.0%)	34 (68.0%)	28 (56.0%)	47 (94.0%)
	M_2/M_2^W	8 (100%)	7 (87.5%)	0 (0%)	0 (0%)	5 (62.5%)	5 (62.5%)	5 (62.5%)	3 (37.5%)	7 (87.5%)		7 (87.5%)	7 (87.5%)	7 (87.5%)	3 (37.5%)	8 (100%)	7 (87.5%)
3	M_2/M_2^B	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	11 (22.0%)	14 (28.0%)	22 (44.0%)	0 (0.0%)	0 (0.0%)	6	1 (2.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	M_2/M_2^W	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (12.5%)	1 (12.5%)	1 (12.5%)		0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (12.5%)
4	M_2/M_2^B	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	19 (38.0%)	15 (30.0%)	16 (32.0%)	0 (0.0%)	0 (0.0%)	8	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	M_2/M_2^W	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (37.5%)	3 (37.5%)	1 (12.5%)	0 (0%)	0 (0%)		0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
5	M_2/M_2^B	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (4.0%)	0 (0.0%)	7 (14.0%)	0 (0.0%)	0 (0.0%)	10	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	M_2/M_2^W	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)		0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)



a.) An example of an MLPA result for a M_2/M_2^B individual

All the reference probes have a ratio of 1, representing two copies of each of these regions. The reference probes are used to normalise SMN probe results.

The *SMN2*, exon 7 (SMN2-7) and exon 8 (SMN2-8) have a ratio of 0 representing a homozygous deletion of these regions. The *SMN1*, exon 7 (SMN1-7), exon 8 (SMN1-8) and *NAIP-5* probes have a ratio of ~1.75 representing four copies of these regions, possibly due to a conversion from the centromeric region: *SMN2*, exon 7, 8_and *NAIP-ψ* to the telomeric region: *SMN1*, exon 7, 8 and *NAIP*.



b.) An example of an MLPA result for a M_2/M_2^W individual

The *SMN2*, exon 7 (SMN2-7) and exon 8 (SMN2-8) probes have a ratio of 0 representing a homozygous deletion of these regions. There is high variability observed in some of the other probes, specifically the GTF2H2-11and SMN1/2, exons 4 and 6 regions, which appear to have one copy each.

Figure 3.13: Examples of MLPA results in M_2/M_2^B and M_2/M_2^W individuals

3.2.3.1 Telomeric SMN region: *SMN1*, exons 7, 8 and *NAIP*, exon 5

The SMN1 gene is the functional gene involved in SMA. CNVs of *SMN1*, exon 7 and 8 were found not to differ statistically between M_2/M_2^B and M_2/M_2^W individuals (H=0.97; p=0.3245 and H=0.08; p-0.7819, respectively), although a higher frequency of multiple copies of *SMN1*, exon 7 were observed in M_2/M_2^B individuals (three copies: 22% (11/50), four copies: 38% (19/50), five copies: 4% (2/50)) when compared to M_2/M_2^W individuals (three, five copies: 0% (0/8), four copies: 37.5% (3/8)). Similarly, a higher frequency of multiple copies of *SMN1*, exon 8 were observed in M_2/M_2^B individuals (three copies: 28% (14/50), four copies: 30% (15/50), five copies: 0% (0/50)) when compared to M_2/M_2^W individuals (three, five copies: 0% (0/8), four copies: 37.5% (3/8)). Again, *SMN1*, exon 7 and 8 copy numbers do not correlate fully in M_2/M_2^B individuals, suggesting that these copies may not be contiguous (discussed in section 4.2.4.3).

The *NAIP*, exon 5 copy number was found to differ statistically between M_2/M_2^B and M_2/M_2^W individuals (H=9.6; p=0.0019). As expected, a higher frequency of multiple copies of *NAIP*, exon 5 were observed in M_2/M_2^B individuals (three copies: 44% (22/50), four copies: 32% (16/50), five copies: 14% (7/50)) when compared to M_2/M_2^W individuals (three, four copies: 12.5% (1/8), five copies: 0% (0/8)). Figure 3.14 illustrates differences in CNV of the telomeric region between M_2/M_2^B and M_2/M_2^W individuals, showing a higher frequency of multiple copies of this region in M_2/M_2^B individuals.

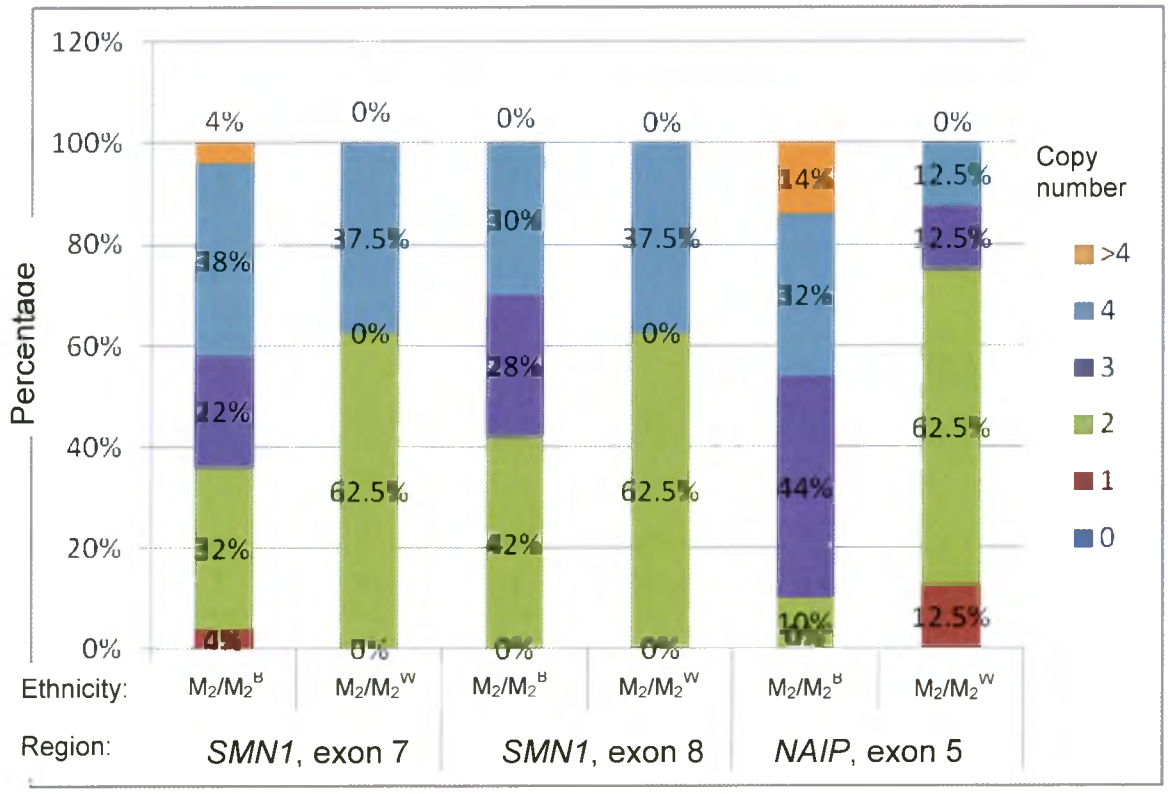


Figure 3.14: Comparison of the copy numbers of *SMN1*, exon7 and exon 8 and *NAIP*, exon 5 between M_2/M_2^B and M_2/M_2^W individuals

3.2.3.2 Centromeric SMN region: *SMN2*, exon 7 and 8

MLPA analysis confirmed homozygous *SMN2*, exon 7 deletions in 100% (50/50) of M_2/M_2^B individuals and 100% (8/8) of the 8 M_2/M_2^W individuals identified. Homozygous deletions of *SMN2*, exon 8 were detected in 98% (49/50) of M_2/M_2^B individuals and 100% (8/8) of M_2/M_2^W individuals. CNVs of *SMN2*, exons 7 and 8 were not found to differ statistically between M_2/M_2^B and M_2/M_2^W individuals (H=2.4; p=0.1252 and H=0.2; p=0.6921, respectively).

A higher frequency of multiple copies of the telomeric region (*SMN1*, exons 7, 8 and *NAIP*, exon 5) in conjunction with homozygous deletions of the centromeric region (*SMN2*, exons 7 and 8) in M_2/M_2^B individuals suggests a possible gene conversion from *SMN2*, exons 7, 8 to *SMN1*, exons 7, 8 and will be further discussed in section 4.2.3.2.

3.2.3.3 CNVs of combined regions: *SMN1/2*, exons 1, 4, 6 and 8; *NAIP/NAIPΨ*, exon 13 and telomeric/centromeric *GTF2H2*, exon 5

The probes discussed in this section, bind to both the telomeric and centromeric copy of each of these genes (identical in probe binding regions) and will therefore give a combined result. Four copies represent full complement of these regions.

CNVs of *SMN1/2*, exons 1 (H=0.07; p=0.7904), 4 (H=2.5; p=0.1107), 6 (H=0.8; p=0.3579) and 8 (H=0.01; p=0.9030) were not found to differ statistically between M_2/M_2^B and M_2/M_2^W individuals. Although, the *SMN1/2*, exons 1, 4, 6 and 8 regions mostly consisted of the full complement of four copies, a full gene deletion (two of four copies deleted) of *SMN1* and/or *SMN2* was observed in M_2/M_2^B individuals (exon 1: 18% (9/50), exon 4: 32% (16/50), exon 6: 34% (17/50) and exon 8: 32% (16/50) and M_2/M_2^W individuals (exon 1, 4 and 8: 12.5% (1/8), exon 6: 62.5% (5/8).

CNVs of *NAIP/NAIPΨ*, exon 13 were found to differ statistically between M_2/M_2^B and M_2/M_2^W individuals (H=5.6; p=0.0177), most likely due to full gene deletions (two of four copies deleted) being more frequently observed in M_2/M_2^B (44% (22/50) than M_2/M_2^W individuals (0% (0/8)). This is not surprising since *NAIPΨ* is thought to co-locate with *SMN2*.

CNVs of *GTF2H2*, exon 5 were not found to differ statistically between M_2/M_2^B and M_2/M_2^W individuals (H=1.98; p=0.1596). M_2/M_2^W individuals seemed to have multiple copies of *GTF2H2*, exon 5 more often (12.5% (1/8)) than M_2/M_2^B individuals (0% (0/50)), although this result is likely to have been influenced by the small sample size of M_2/M_2^W individuals.

CNVs of centromeric-*GTF2H2* (*GTF2H2-8*) were found to differ significantly between M_2/M_2^B and M_2/M_2^W individuals (H=10.2, p=0.0014) and CNVs of telomeric-*GTF2H2 (GTF2H2-11*) were not found to differ significantly between M_2/M_2^B and M_2/M_2^W individuals (H=0.7, p=0.4124) and will be further explained and compared in section 4.2.4.4.

Figure 3.15 shows the higher frequency of deletions of *NAIP/NAIPΨ*, exon 13 in M_2/M_2^B when compared to M_2/M_2^W individuals and a higher frequency of multiple

copies of *GTF2H2*, exon 5 in M_2/M_2^W individuals when compared to M_2/M_2^B individuals.

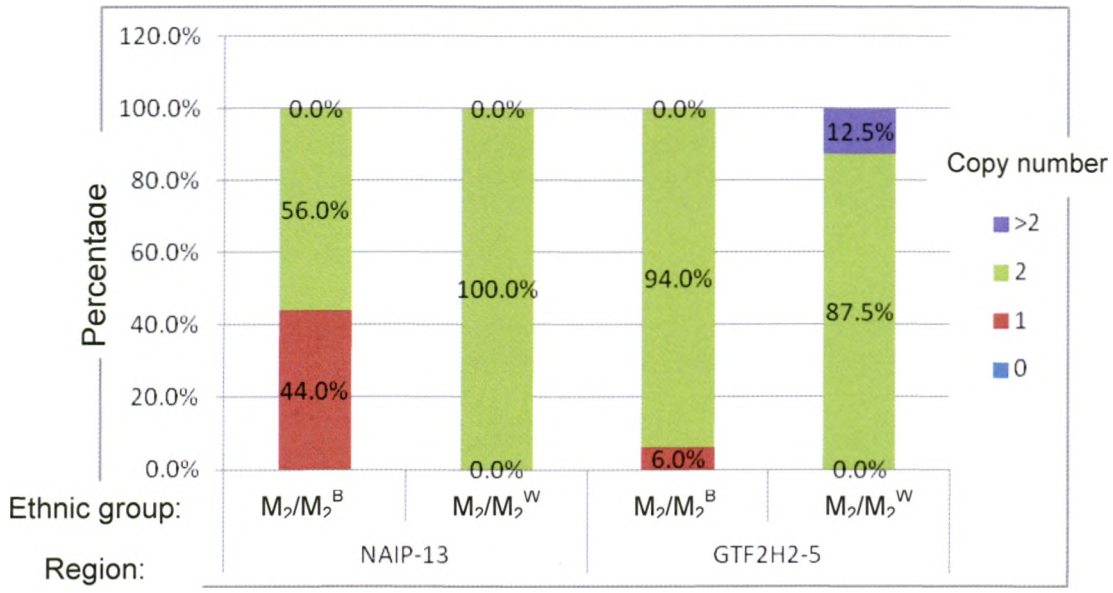


Figure 3.15: Comparison of the copy numbers of *NAIP/NAIPΨ*, exon 13 and *GTF2H2*, exon 5 between M_2/M_2^B and M_2/M_2^W individuals

3.2.3.4 CNVs of neighbouring genes: *SERF1B* and *RAD17*

The *SERF1B* copy numbers were not found to differ significantly between M_2/M_2^B and M_2/M_2^W individuals (H=0.02; p=0.8855). All M_2/M_2^B and M_2/M_2^W individuals had two copies of *RAD17* (100% (50/50) and 100% (8/8), respectively).

3.2.4 Comparative analysis of Group A (U/U^B) with other black subject groups (Groups B (N/N^B), D (M_1/M_1^B) and F (M_2/M_2^B)

U/U^B individuals were tested on MLPA to look for a potential novel and common disease-causing mutation and to investigate the hypothesis of pathogenic rearrangements of the SMN region forming part of the SMA disease mechanism in the black SA population.

The non-parametric Kruskal-Wallis and median tests were performed using Statistica to determine which probes had a significant difference between U/U^B individuals and other black subject groups (N/N^B, $M_1M_1^B$ and $M_1M_1^B$ individuals). Table 3.6 summarises the statistical comparisons between these subject groups.

Significant differences ($p<0.05$) between U/U^B and $M_1M_1^B$ individuals were observed for all (15/15) MLPA probes, suggesting that hypothesised novel pathogenic CNVs are distinct from the common homozygous deletion of *SMN1*, exon 7. Similarly, significant differences ($p<0.05$) between U/U^B and $M_2M_2^B$ individuals were observed for the majority of MLPA regions (11/15).

U/U^B individuals more closely resembled N/N^B individuals, however significant differences ($p<0.05$) between U/U^B and N/N^B individuals were observed for only five (5/15) MLPA probe regions:

- *SMN1*, exon 7
- *NAIP*, exon 5
- *NAIP/NAIPΨ*, exon 13
- Telomeric-*GTF2H2 (GTF2H2-11)*
- *GTF2H2*, exon 5

The results for *SMN1*, exon 7 will be further discussed in section 3.2.4.1; the results for *NAIP-5* and *NAIP/NAIPΨ*, exon 13 will be further discussed in section 3.2.4.2 and results for *GTF2H2*, exon 5 and telomeric-*GTF2H2* will be further discussed in section 3.2.4.4. The significance of these differences will be further discussed in section 4.2.4.

Table 3.7 summarises and compares the results of U/U^B and N/N^B individuals. The number and percentage of individuals who fall into different CNV categories (one-ten copies) are indicated for each P021-A2 MLPA probe. Statistically significant differences are highlighted. See Figure 3.16 for examples of MLPA results.

Table 3.6: Significant differences in CNVs between patient group U/U^B and black subject groups (M_1/M_1^B, M_2/M_2^B and N/N^B)

Sample group	*RAD17*	*SERF1B*	*SMN2*-7	*SMN2*-8	*SMN1*-7	*SMN1*-8	*NAIP*-5	GTF2H2-11	GTF2H2-8	SMN1/2 Exon 1	SMN1/2 Exon 4	SMN1/2 Exon 6	SMN1/2 Exon 8	NAIP-13	GTF2H2-5
$M_1M_1^B$	p=0.0005 H=12.2	p<0.0001 H=77.7	p=0.0123 H=6.3	p<0.0001 H=28.2	p<0.001 H=126.2	p<0.0001 H=70.6	p<0.0001 H=47.5	p<0.0001 H=65	p<0.0001 H=24.9	p<0.0001 H=56.6	p<0.0001 H=63.2	p<0.001 H=80.2	p<0.0001 H=64.2	p<0.001 H=81.8	p<0.0001 H=66.6
$M_2M_2^B$	p<0.0001 H=51.1	p<0.0001 H=19.7	p<0.0001 H=76.7	p<0.0001 H=77	p=0.0031 H=8.7	p=0.1484 H=2.1	p<0.0001 H=45.1	P=0.2831 H=1.2	p<0.0001 H=24.2	p=0.8836 H=0.02	p=0.0006 H=11.9	p<0.0001 H=27.7	p<0.0001 H=22.7	p<0.0001 H=30.4	p=0.1760 H=1.8
N/N^B	p=0.7576 H=0.1	p=0.5812 H=0.4	p=0.5783 H=0.3	p=0.812 H=0.06	p=0.0201 H=5.4	p=0.3946 H=0.7	p=0.0028 H=8.9	p=0.0052 H=7.8	p=0.5381 H=0.4	p=0.1528 H=2	p=0.924 H=0.009	p=0.3764 H=0.8	p=0.1285 H=2.3	p=0.0204 H=5.4	p=0.0003 H=13.3

Statistically significant differences between U/U^B individuals and other black subject groups are highlighted.

Table 3.7: Comparison of CNVs in U/U^B and N/N^B individuals. Significant differences ($p<0.05$) between U/U^B and N/N^B individuals are highlighted.

Nr of copies	Groups U/U^B (n=72) N/N^B (n=122)	P021-A2 probes									Nr of copies of universal probe regions	SMN1/2 Exon 1	SMN1/2 Exon 4	SMN1/2 Exon 6	SMN1/2 Exon 8	NAIP-13	GTF2H2-5
		RAD17	SERF1B	SMN2 Exon 7	SMN2 Exon 8	SMN1 Exon 7	SMN1 Exon 8	NAIP-5	GTF2H2-11 rs200357275	GTF2H2-8 rs201102513							
0	U/U^B	0 (0.0%)	0 (0.0%)	10 (13.9%)	17 (23.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	5 (6.9%)	0 (0.0%)	0	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.4%)	2 (2.8%)
	N/N^B	0 (0.0%)	0 (0.0%)	15 (12.3%)	33 (27.0%)	0 (0.0%)	1 (0.8%)	1 (0.8%)	11 (9.0%)	0 (0.0%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
1	U/U^B	0 (0.0%)	1 (1.4%)	8 (11.1%)	33 (45.8%)	6 (8.3%)	5 (6.9%)	5 (6.9%)	14 (19.4%)	0 (0.0%)	2	4 (5.6%)	4 (5.6%)	3 (4.2%)	3 (4.2%)	5 (6.9%)	9 (12.5%)
	N/N^B	0 (0.0%)	1 (0.8%)	20 (16.4%)	55 (45.1%)	0 (0.0%)	0 (0.0%)	4 (3.3%)	37 (30.3%)	0 (0.0%)		14 (11.5%)	9 (7.4%)	3 (2.5%)	7 (5.7%)	9 (7.4%)	30 (24.6%)
2	U/U^B	71 (98.6%)	70 (97.2%)	51 (70.8%)	19 (26.4%)	38 (52.8%)	31 (43.1%)	42 (58.3%)	49 (68.1%)	65 (90.3%)	4	68 (94.4%)	68 (94.4%)	69 (95.8%)	68 (94.4%)	66 (91.7%)	60 (83.3%)
	N/N^B	104 (85.2%)	120 (98.4%)	73 (59.8%)	31 (25.4%)	60 (49.2%)	54 (44.3%)	71 (58.2%)	74 (60.7%)	111 (91.0%)		106 (86.9%)	112 (91.8%)	118 (96.7%)	114 (93.4%)	112 (91.8%)	91 (74.6%)
3	U/U^B	1 (1.4%)	1 (1.4%)	3 (4.2%)	3 (4.2%)	11 (15.3%)	24 (33.3%)	25 (34.7%)	4 (5.6%)	7 (9.7%)	6	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.4%)
	N/N^B	17 (13.9%)	1 (0.8%)	10 (8.2%)	3 (2.5%)	21 (17.2%)	52 (42.6%)	29 (23.8%)	0 (0.0%)	8 (6.6%)		2 (1.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.8%)
4	U/U^B	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	16 (22.2%)	12 (16.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	8	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.4%)	0 (0.0%)	0 (0.0%)
	N/N^B	1 (0.8%)	0 (0.0%)	3 (2.5%)	0 (0.0%)	35 (28.7%)	14 (11.5%)	14 (11.5%)	0 (0.0%)	3 (2.5%)		0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
5	U/U^B	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	10	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	N/N^B	0 (0.0%)	0 (0.0%)	1 (0.8%)	0 (0.0%)	6 (4.9%)	1 (0.8%)	3 (2.5%)	0 (0.0%)	0 (0.0%)		0 (0.0%)	1 (0.8%)	1 (0.8%)	1 (0.8%)	1 (0.8%)	0 (0.0%)

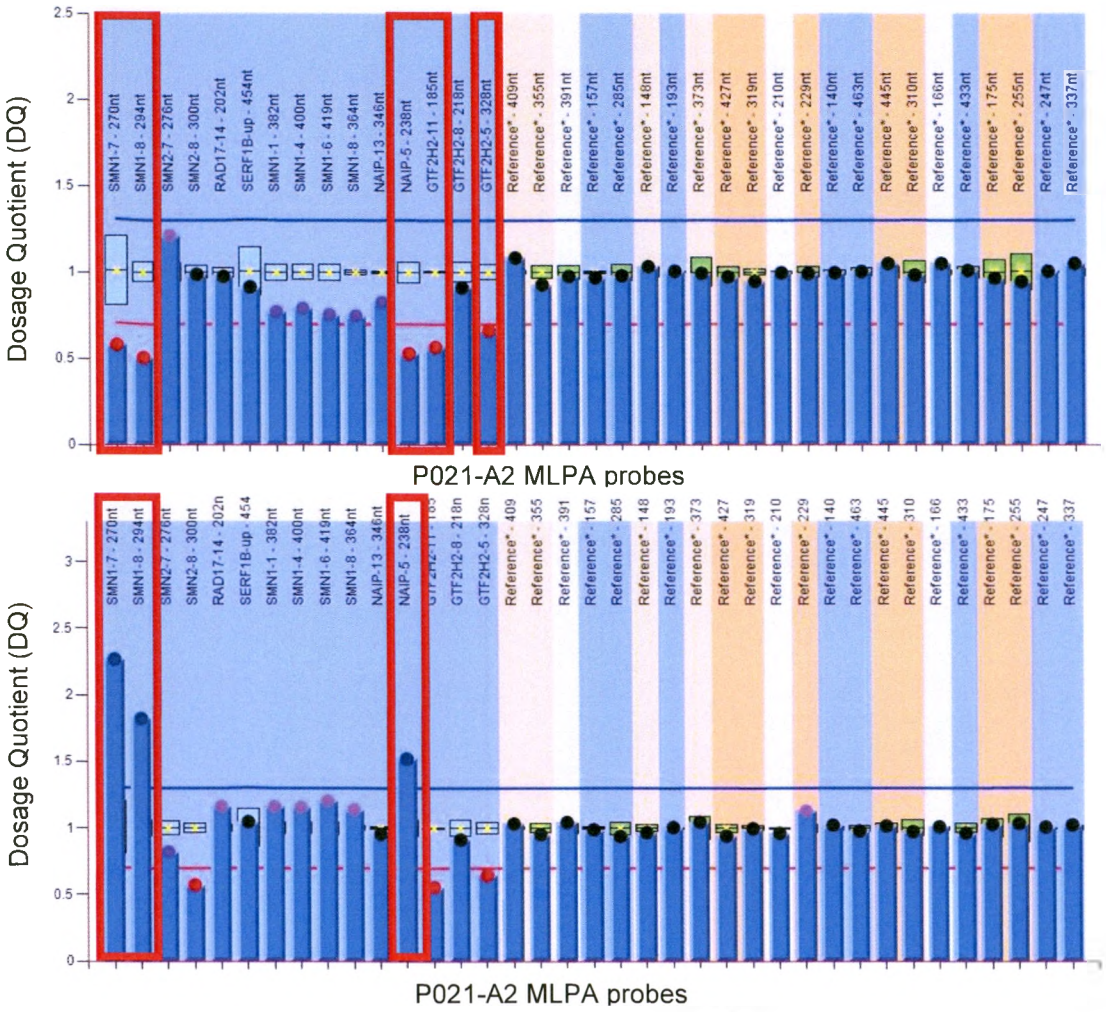


Figure 3.16: Examples of MLPA results in U/U^{B} individuals

a.) An example of a heterozygous *SMN1*, exon 7 deletion MLPA result in an U/U^{B} individual

All the reference probes have a ratio of 1, representing two copies of each of these regions. The reference probes are used to normalise SMN probe results.

The *SMN1*, exon 7 (SMN1-7), exon 8 (SMN1-8), *NAIP-5, GTF2H2-5* and*GTF2H2-11* gene probes have a ratio of ~0.5 representing a large heterozygous deletion of these regions, supporting a diagnosis of SMA, with the second pathogenic mutation remaining unidentified.

b.) An example of an MLPA result with multiple *SMN1* copies in an U/U^{B} individual

The *SMN1*, exon 7 (SMN1-7), exon 8 (SMN1-8) and *NAIP-5* probes have a ratio of ~1.5-2.5 representing three to four copies of these regions. The SMN2, exon 8 (SMN2-8) and the GTF2H2-5 and GTF2H2-11 probes have a ratio of 0.5, representing a heterozygous deletion of these regions po67yyyyyyyyyyyyyyyyssibly representing the presence of a large rearrangement, causing SMA in this patient.

3.2.4.1 Telomeric SMN region: *SMN1*, exons 7 and 8

The SMN1 gene is the functional gene involved in SMA. Multiple copies (more than two copies) were observed for *SMN1*, exon 7 (38.9% (28/72)) and *SMN1*, exon 8 (50% (37/72)) in U/U^B individuals which is similar to results of N/N^B individuals (refer to section 3.2.1.1).

The significant difference of the *SMN1*, exon 7 probe between between U/U^B and N/N^B individuals could be attributed to the presence of heterozygous *SMN1*, exon 7 deletions in 8.3% (6/72) of U/U^B individuals, not observed in any N/N^B individuals (0% (0/122); section 3.2.1.2). This result stands in contrast to a previous South-African study that reported a heterozygous *SMN1*, exon 7 deletion rate of 69.5% (16/23) (Labrum *et al.*, 2007) and will be further discussed in section 4.2.4.2.

3.2.4.2 CNVs of *NAIP*, exon 5 and *NAIP/NAIPΨ*, exon 13

The copy numbers of *NAIP*, exon 5 and the combined *NAIP/NAIPΨ*, exon 13 were both shown to differ significantly between U/U^B and N/N^B individuals. The only obvious difference is a larger number of multiple copies of *NAIP*, exon 5 in N/N^B individuals (three copies: 11.5% (14/122) and four copies: 2.5% (3/122)) compared to U/U^B individuals (three, four copies: 0% (0/72)).

The *NAIP/NAIPΨ*, exon 13 results represent a combined *NAIP* and *NAIPΨ* copy number with a full complement of four copies. No clear significant differences can be seen.

Figure 3.17 compares the CNVs of *NAIP*, exon 5 & *NAIP/NAIPΨ*, exon 13 between U/U^B and N/N^B individuals. These results will be further discussed in section 4.3.3.

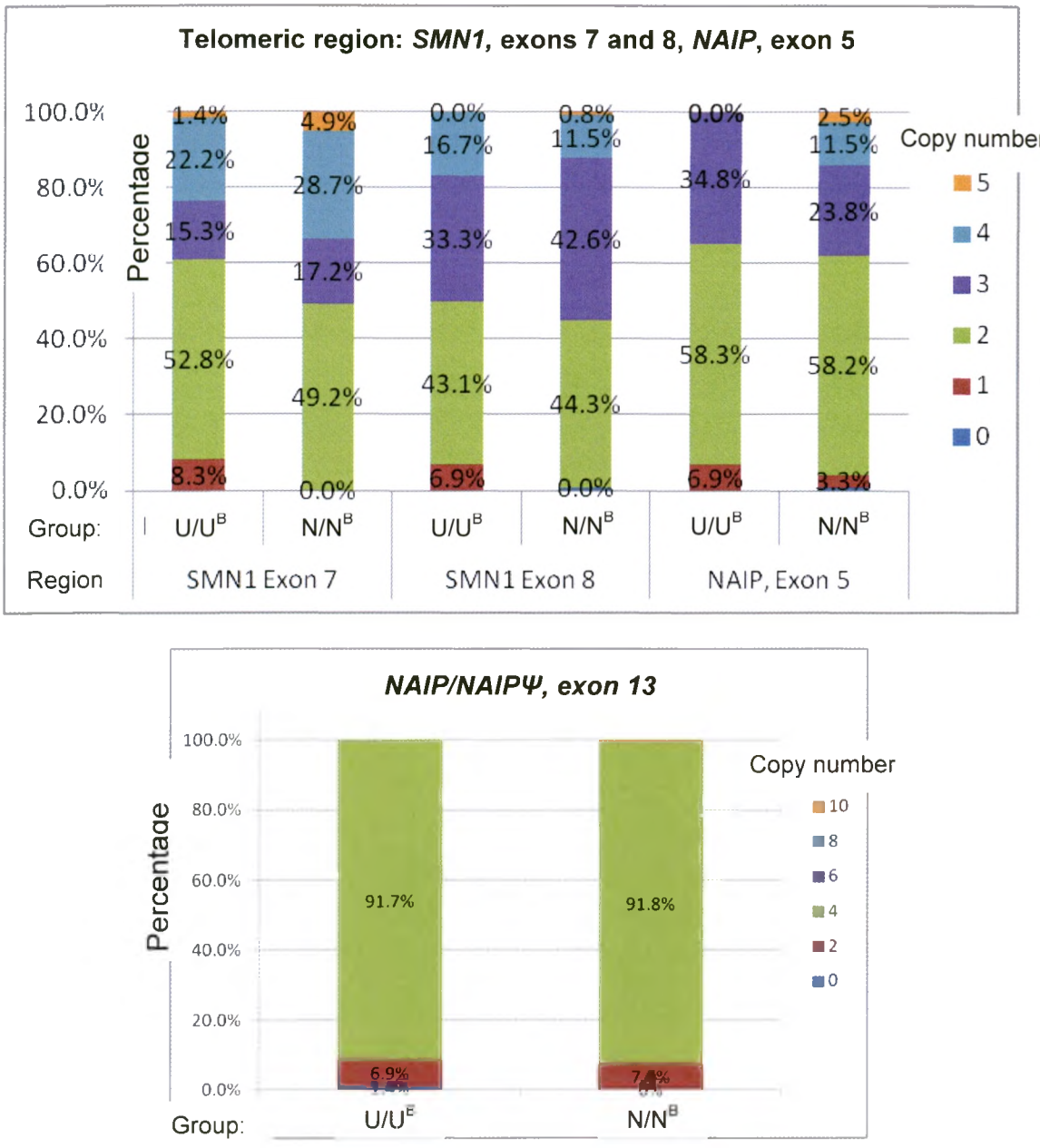


Figure 3.17: Bar charts comparing the copy numbers of *NAIP*, exon 5 & *NAIP/NAIPΨ*, exon 13 between U/U^B and N/N^B individuals

3.2.4.3 Centromeric SMN region: *SMN2*, exon 7 and 8

Homozygous deletions of *SMN2*, exon 7 were observed in 13.9% (10/72) of U/U^B individuals which is similar to frequencies observed in black patients referred for diagnostic testing for SMA: 12.4% (123/991) (see section 3.1 for a detailed summary of the results from the audit). Once again, the question arises whether these deletions might form part of the general CNV patterns observed in the black SA population. The homozygous *SMN2*, exon 7 deletion rate did not differ significantly between black

patient samples included in the audit and U/U^B individuals in this study (χ^2=0.13; p=0.9). Once again, copy numbers of SMN2, exons 7 and 8 do not correlate as seen in N/N^B individuals (further discussed in section 4.2.1.3). For a summary of the statistical calculations, see Appendix F, statistical test output 5.

These conversions are also observed in $M_1/M_1{}^B$ and U/U^B individuals (refer to sections 3.2.2.2 and 3.2.4.3, respectively) and are compared in table 3.8. Gene conversions (explained in section 1.2.2.3) from *SMN2*, exon 8 to *SMN1*, exon 8 would result in a loss of *SMN2*, exon 8 and an increased copy number of *SMN1*, exon 8.

Table 3.8: The copy numbers of *SMN2*, exons 7 and 8 do not correlate

Region	Copy number	N/N^B	$M_1/M_1{}^B$	U/U^B
SMN2, exon7	0 (homozygous deletion)	12.3%	0%	13.9%
	1 (heterozygous deletion)	16.4%	5.3%	11.1%
SMN2, exon8	0 (homozygous deletion)	27%	10.7%	23.6%
	1 (heterozygous deletion)	45.1%	20%	45.8%

The highlighted cells indicate a higher deletion frequency of exon 8 when compared to exon 7. Compiled from data from tables 3.1, 3.3 and 3.7.

3.2.4.4 CNVs of telomeric-*GTF2H2* (c.706C>G, rs200357275) and centromeric-*GTF2H2* (c.453A>G, rs201102513)

The CNVs of the centromeric-*GTF2H2* (c.453A>G; *GTF2H2-8*, rs201102513) variant were shown to differ significantly among sample groups (χ^2=25.2, p=0.00032) (see Appendix F, statistical test output 7 for an explanation of the calculations). This could be due to a deletion of the centromeric-*GTF2H2* region being present in 4% (6/150) of *SMN1*, exon 7 deletion chromosomes; not seen in any of the other sample groups and do not represent any major observations.

The CNVs of the telomeric-*GTF2H2* (c.706C>G; *GTF2H2-11*, rs200357275) variant were shown to differ significantly among sample groups (χ^2=277.7, p=0.482x10^{-57}) (see Appendix F, statistical test output 7 for an explanation of the calculations). The c.706C>G variant represents the presence of telomeric-*GTF2H2*. Deletions of the c.706C>G variant seems to be associated specifically with three different subject groups: $M_1/M_1{}^B$, $M_1/M_1{}^W$ as expected and surprisingly, N/N^W.

See Table 3.9 for a summary of results of these two variant genotypes for all the different subject groups. The significance of these two variants and their associations with different subject groups will be further discussed in section 4.2.4.4.

Table 3.9: Genotype results for telomeric-*GTF2H2* and centromeric-*GTF2H2*

Patient group	Ethnicity	Deletion of centromeric-*GTF2H2* & telomeric-*GTF2H2* (combined result) (*GTF2H2-5*)	Centromeric-*GTF2H2* (*GTF2H2-8*, rs201102513, c.453A>G)		Telomeric-*GTF2H2* (*GTF2H2-11*, rs200357275, c.706C>G)	
			Presence of region (G allele)	Absence/ deletion of region (A allele)	Presence of region (G allele)	Absence/ deletion of region (C allele)
M_1/M_1^B n=150	Black	40% (60)	96% (144)	4% (6)	20% (30)	80% (120)
M_1/M_1^W n=60	White	25% (15)	100% (60)	0% (0)	30% (18)	70% (42)
M_2/M_2^B n=100	Black	3% (3)	100% (100)	0% (0)	81% (81)	19% (19)
M_2/M_2^W n=16	White	0% (0)	100% (16)	0% (0)	68.8% (11)	31.25% (5)
N/N^B n=244	Black	12.3% (30)	100% (244)	0% (0)	75.8% (185)	24.2% (59)
N/N^W n=60	White	0% (0)	100% (60)	0% (0)	8.3% (5)	91.7% (55)
U/U^B n=144	Black	9% (13)	100% (144)	0% (0)	83.3% (120)	16.7% (24)

The major allele in each group is shaded.

3.3 General statistical analysis

In order to test the accuracy of non-parametric statistical tests used to analyse MLPA results, a parametric MANOVA statistical test was performed. A MANOVA statistical test uses the "group means less total means" value to showcase the more obvious

major differences between subject groups for each MLPA probe. A value of zero represents the expected two copies and no major fluctuations in copy number. A positive value represents a higher copy number of a region (multiple copies) and a negative value represents a lower copy number of a region (deletions). These results are only intended to show major significant differences.

MANOVA analysis showed strongly significant differences among the seven subject groups. Figure 3.18 summarises the MANOVA results in a graph.

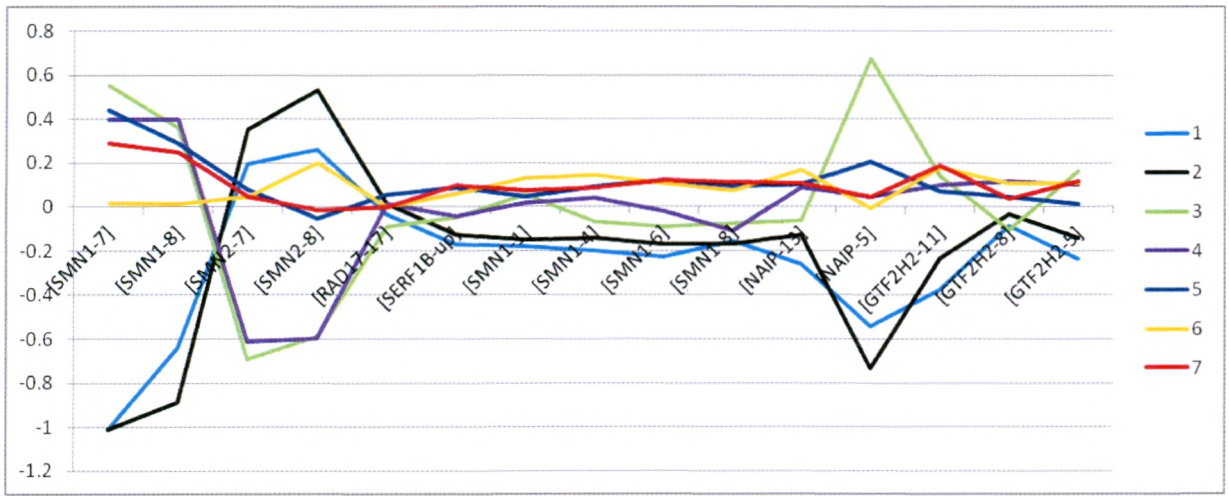


Figure 3.18: Comparison of MLPA probe group means across 7 patient groups
Each colour represents a different subject group - 1: M_1/M_1^B, 2: M_1/M_1^W; 3: M_2/M_2^B; 4: M_2/M_2^W; 5: N/N^B; 6: N/N^W and 7: U/U^B individuals. The X axis represents the different MLPA probes and the Y axis represents the "group means less total means" value. A value of zero represents no fluctuation/variation from the normal expected two copies; a value greater than zero represents an increased copy number and a value smaller than zero represents a decrease in copy number of these regions.

The N/N^W profile seems to cluster around zero, which represents the two expected copies of each probe, showing no major fluctuations among different MLPA probes. The N/N^B profile seems to have a higher value for the *SMN1*, exons 7 and 8 probes, which is supported by the observation of multiple copies of these regions.

As expected both the M_1/M_1^B and M_1/M_1^W profiles have a lower value for the telomeric region (*SMN1*, exons 7, 8 and *NAIP*, exon 5) when compared to other groups due to deletions of these regions in conjunction with higher values of the centromeric region (*SMN2*, exons 7 and 8) possibly due to gene conversion events.

As expected, the M_2/M_2^B and M_2/M_2^W subject groups have a lower value for the centromeric region (*SMN2*, exons 7, 8) when compared to other groups due to

deletions of these regions in conjunction with higher values of the telomeric region (*SMN1*, exons 7 and 8) possibly due to gene conversion events.

The only major difference between these two groups is the *NAIP*, exon 5 value which is much higher in the M_2/M_2^B profile when compared to the M_2/M_2^W profile, suggesting the presence of multiple *NAIP*, exon 5 copies in M_2/M_2^B individuals not seen in M_2/M_2^W individuals.

The MANOVA statistical test, in summary, therefore confirms major statistical differences between subject groups observed using non-parametric statistical tests used to analyse MLPA results and therefore adds confidence to the results.

3.4 Haplotype analysis

Pedigrees were drawn for all 86 black families: 25 families positive for the homozygous deletion of *SMN1*, exon 7 (M_1/M_1^B families) and 61 families negative for SMA (N/N^B families), in order to attempt to determine the phase of CNVs within families. M_1/M_1^B families were analysed to possibly construct common pathogenic haplotypes associated with SMA and N/N^B families were analysed to possibly construct non-pathogenic haplotypes.

The haplotypes were constructed using current map information from the Ensembl genome browser (Ensembl genome browser, 2016) as illustrated in section 1.2.1, figure 1.1. As explained in section 1.2.2, the exact location and order of genes in the SMN region is not completely understood, and may influence haplotype results.

Some examples of family case studies will be discussed. All the pedigrees are available on the CD disk that accompanies this dissertation.

3.4.1 N/N^B families

Only 31.7% (19/60) of N/N^B families were completely informative where the phase of CNVs could be determined with absolute certainty (figure 3.19 shows an example). For 60% (36/60) of N/N^B families, multiple combinations were possible, due to the presence of multiple copies of one or more probe regions. The exact locations of these multiple copies are not known. Figure 3.20 shows examples of N/N^B family

pedigrees with multiple gene copies which showcase the challenges of constructing clear pedigrees. Discrepant results were observed in 5% (3/60) of N/N^B families either due to non-paternity or novel deletion/duplication events in the proband. Figure 3.21 shows an example of an N/N^B family with discrepant results. Two N/N^B families had clear novel results with a new mutation rate of 3.3% (2/60). Figure 3.22 shows an example of an N/N^B family with novel results. Section 2.3.4 explains the requirements of family pedigrees to be classified as having discrepant or novel results.

3.4.1.1 Non-pathogenic haplotypes

In total, 19 N/N^B families, consisting of 38 unrelated parents were found to be informative, from which 76 haplotypes were constructed. In total, 35 unique haplotypes were identified, emphasising the high variability of this region, and these data are summarised in Appendix G. Even though no clear pattern emerged, some generalisations could be drawn.

The telomeric region: *SMN1*, exons 7, 8 and *NAIP*, exon 5

Multiple copies (more than two) of *SMN1*, exon 7 and 8 were observed in 19.7% (15/76) of haplotypes and multiple copies of *NAIP*, exon 5 were observed in 21.1% (16/76) of haplotypes, as expected from results observed in N/N^B individuals (section 3.2.1.1). The majority of these 15 haplotypes (80% (12/15)) had a combined *SMN1*, exon 7 and 8 duplication. The remaining 20% (3/15) of haplotypes had a larger duplication extending into *NAIP*, exon 5.

No N/N^B family had family members with a clear silent carrier genotype (2:0, 3:0, 4:0, 5:0, 6:0 etc.; defined in section 1.2.2.4). Four individuals from N/N^B families had uninformative results for *SMN1*, exon 7, where the individual could either have one copy of *SMN1*, exon 7 on each of his/her chromosomes (1:1 genotype) or two copies of *SMN1*, exon 7 on one chromosome in conjunction with no copies of *SMN1*, exon 7 on the other chromosome (2:0 silent carrier genotype).

The centromeric region: *SMN2*, exons 7 and 8

Deletions of *SMN2*, exon 7 were observed in 31.6% (24/76) and deletions of *SMN2*, exon 8 were observed in 36.8% (28/76) of haplotypes. Once again the copy number of *SMN2*, exons 7 and 8 do not fully correlate with each other.

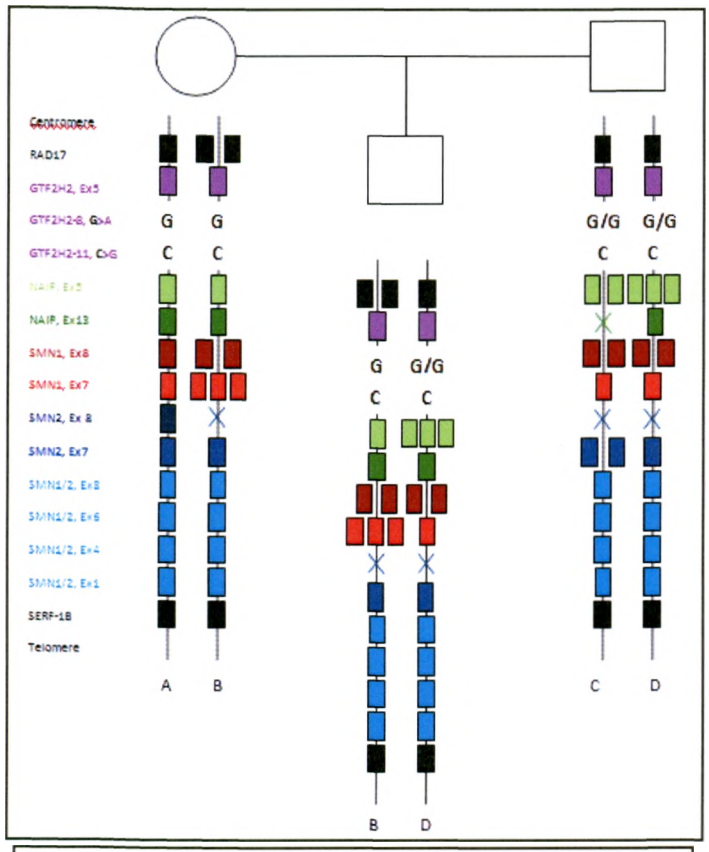
A deletion of *SMN2*, exon 7, together with a deletion of *SMN2*, exon 8, was observed in 26.3% (20/76) of haplotypes, suggesting a true deletion in these cases, although larger deletions of the rest of the *SMN1* and/or *SMN2* exons were not observed in these cases. Only 7.9% (6/76) of haplotypes had a full deletion of *SMN2* exon 7 and 8 in addition to four copies of *SMN1*, exon 7 and 8, suggesting a full gene conversion from *SMN2* to *SMN1*.

Combined regions: *SMN1/2*, exons 1, 4, 6 and 8; *GTFTH2*, exon 5 and *NAIP*, exon 13 regions

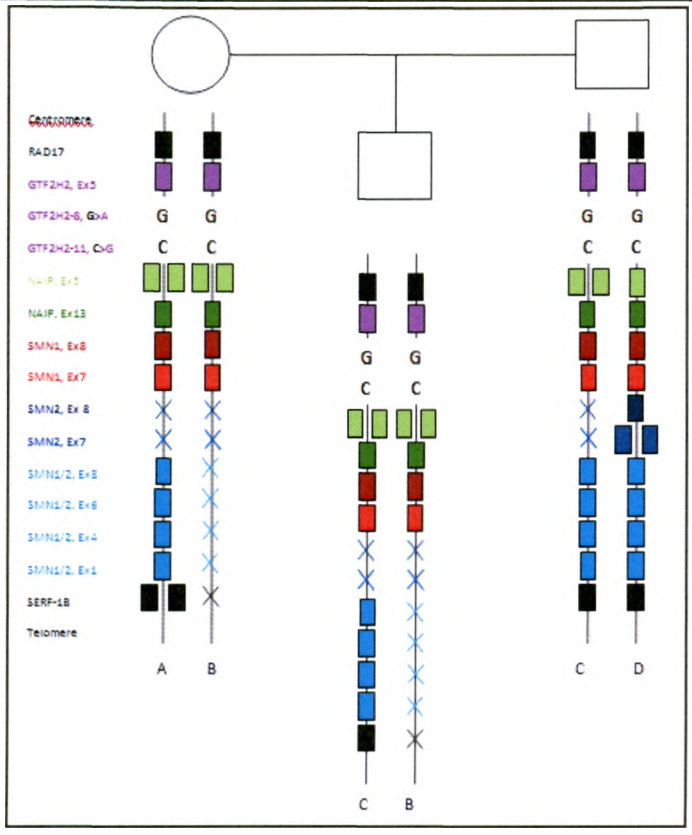
As explained in section 2.3.2.1, probes specific to the combined *SMN1/2*, exons 1, 4, 6 and 8, *GTFTH2*, exon 5 and *NAIP*, exon 13 regions cannot distinguish between the telomeric and centromeric copies of these regions and will give a combined result. CNV results of these regions need to be analysed in context of the CNVs observed in each family. Only 2.6% (2/76) of haplotypes with deletions of *SMN2*, exons 7 and 8, included deletions of *SMN1* and/or *SMN2* exons 1, 4, 6 and 8 suggesting full gene deletions of *SMN2* in these cases. Deletions of *GTF2H2*, exon 5 were observed in 6.6% (5/76) of haplotypes, suggesting a deletion of centromeric-*GTF2H2*.

Telomeric-*GTF2H2* and centromeric-*GTF2H2*

All 76 haplotypes had a full complement (one copy per chromosome) of telomeric-*GTF2H2*. A full complement (one copy per chromosome) of centromeric-*GTF2H2* was observed in 85.5% (65/76) of haplotypes.



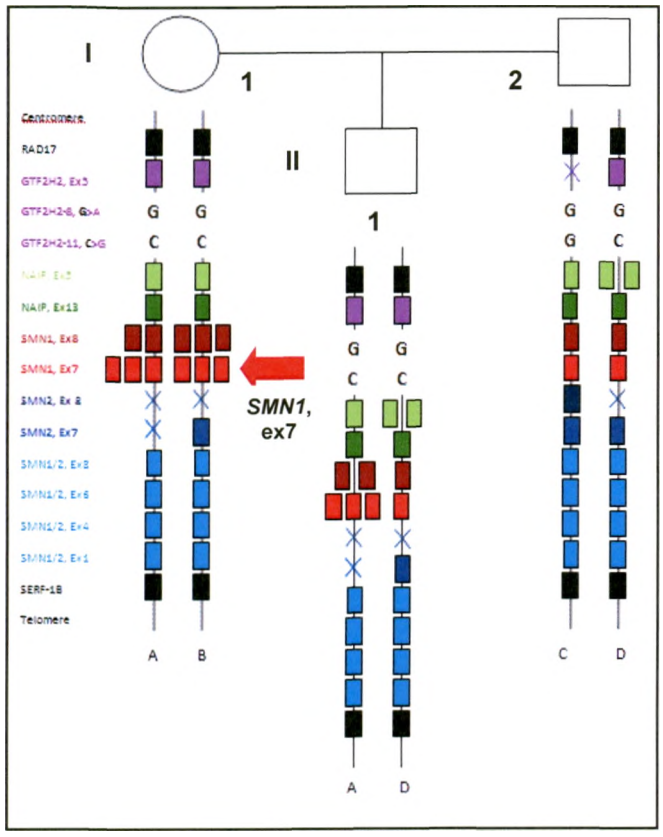
a.) This pedigree shows an example of a clear informative pedigree where the phase of the SMN region could be accurately determined, despite multiple copies of *NAIP*, exon 5; *SMN1*, exons 7 and 8 and *RAD17*.



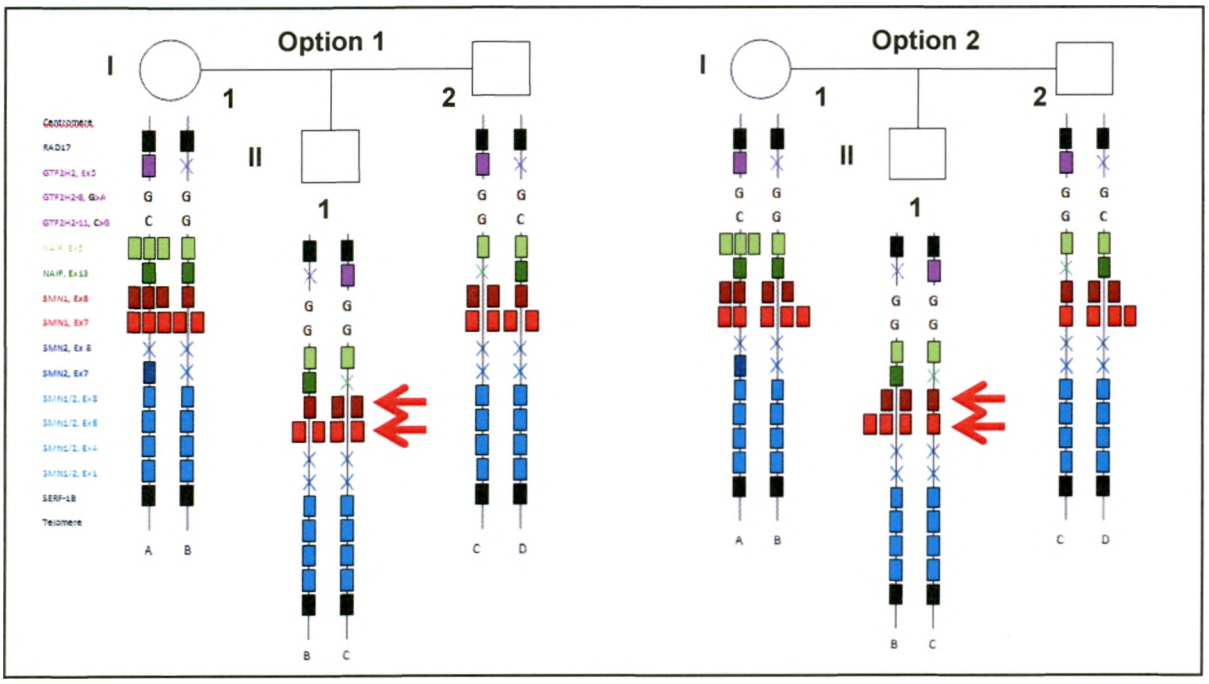
b.) This pedigree shows another example of a clear informative pedigree where the phase of the SMN region could be accurately determined. This pedigree is also an example of a family with a large deletion chromosome (chromosome B). Interestingly, all three family members have two copies of *SMN1*, exon 7 and therefore are not at risk of SMA. The mother (I-1) and child (II-1) both have homozygous deletions of *SMN2*, exons 7 and 8. This is a good example of the high variability in the SMN region in the black negative population.

Figure 3.19: Informative pedigrees of N/N^{B} families

The different coloured boxes represent the copy number of different probe regions. One coloured box indicates a single copy of a region. The *SMN1/2*, exons 1, 4, 6 and 8; *NAIP*, exon 13 and *GTF2H2*, exon 5 regions represent combined results for the telomeric and centromeric copies of these regions. An X indicates a deletion of the specific region.

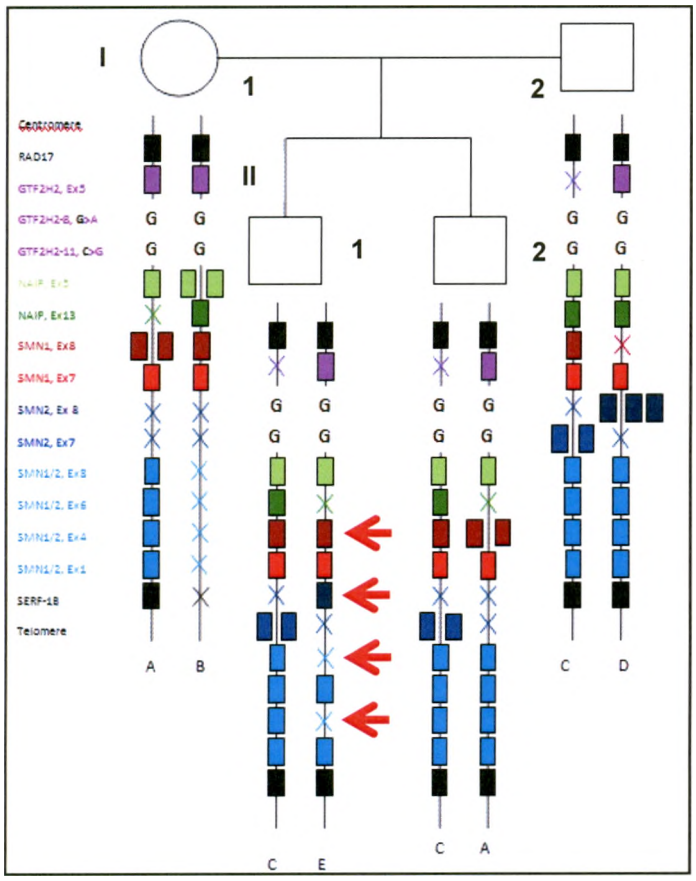


a.) This pedigree shows an example of an uninformative pedigree where the phase of the SMN region could not be accurately determined due to multiple copies of *SMN1*, exons 7 and 8. The mother (I-1) has the largest number of copies (six) of *SMN1*, exon 7 observed in this population group. She has a heterozygous deletion of *SMN2*, exon 7 and a homozygous deletion of *SMN2*, exon 8, which could be indicative of a gene conversion from *SMN2* to *SMN1*.



b.) Multiple CNV combinations are possible in this family. This picture showcases two potential combinations of many. Differences are indicated with arrows. The phase of the chromosomes cannot be accurately assigned due to the presence of multiple gene copies.

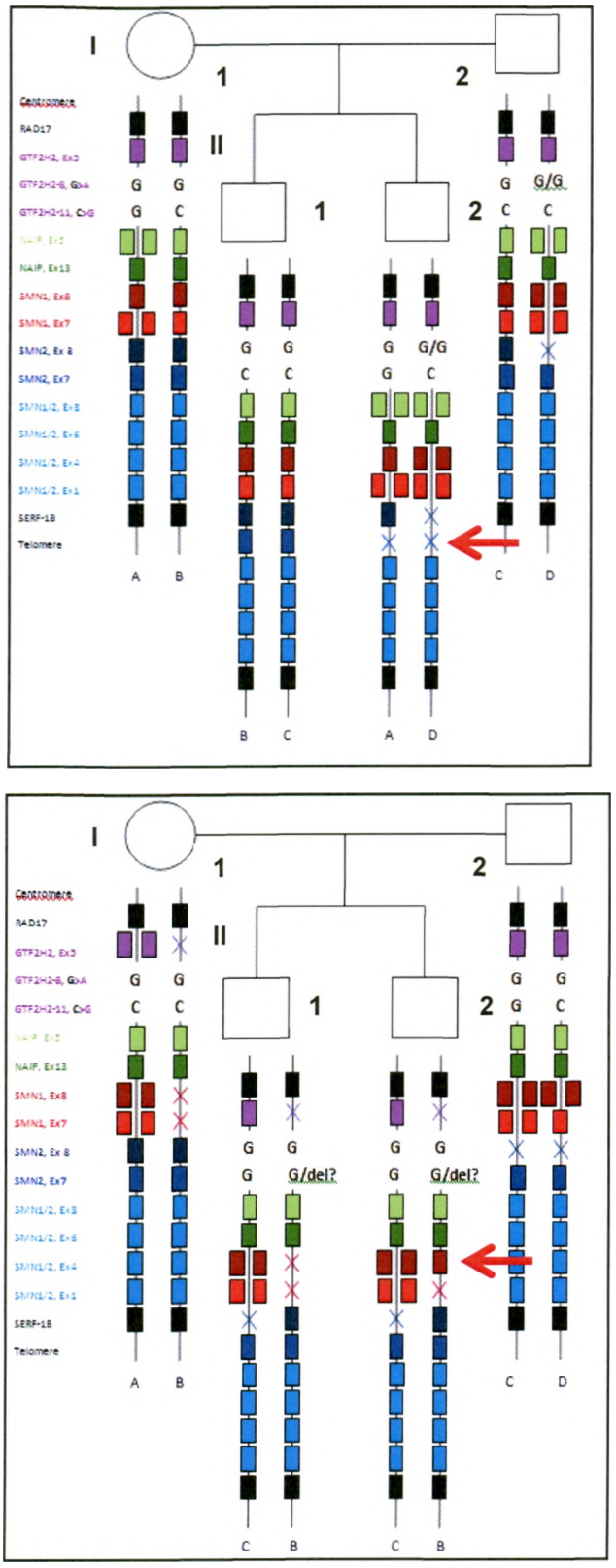
Figure 3.20: Uninformative pedigrees of N/N^{B} families with multiple gene copies. The different coloured boxes represent the copy number of different probe regions. One coloured box indicates a single copy of a region. The *SMN1/2*, exons 1, 4, 6 and 8; *NAIP*, exon 13 and *GTF2H2*, exon 5 regions represent combined results for the telomeric and centromeric copies of these regions. An X indicates a deletion of the specific region.



This pedigree shows an example of a family with discrepant results. Potential novel events could have taken place in child 1 (II-1). In order for the child to fit within the context of the family, four novel mutation events are needed to have taken place: a deletion of *SMN1/2*, exons 4 and 8 in addition to a deletion of *SMN1*, exon 8 and a duplication of *SMN2*, exon 8. This could potentially indicate a novel conversion of *SMN1*, exon 8 to *SMN2*, exon 8 in child II-1. It is also possible that the mother in this pedigree (I-1) is not the actual biological mother of child II-1. The mother's CNV profile suggests that she could be the true biological mother of child II-2.

Figure 3.21: Pedigree of a N/N^{B} family with discrepant results

The different coloured boxes represent the copy number of different probe regions. One coloured box indicates a single copy of a region. The *SMN1/2*, exons 1, 4, 6 and 8; *NAIP*, exon 13 and *GTF2H2*, exon 5 regions represent combined results for the telomeric and centromeric copies of these regions. An X indicates a deletion of the specific region.



a.) The two children (II-1 and II-2) appear to have inherited different chromosomes from both parents. Child II-2 seems to have inherited chromosome D from his father (I-2) with a novel deletion of *SMN2*, exon 7 (indicated with an arrow). The deletion can be explained either due to a de novo deletion event in the child or non-paternity. Multiple combinations are possible in this uninformative family pedigree due to multiple gene copies.

b.) The two children (II1 and II2) appear to have inherited the same chromosomes from both parents (chromosomes B and C). Child II2 seems to have inherited chromosome B from his mother (I1) with a novel duplication of *SMN1*, exon 8 (indicated with an arrow). The duplication either arose as a result of a novel duplication event in the child or perhaps non maternity, although this is unlikely since the children have inherited the same CNV pattern. Multiple combinations are possible in this uninformative family pedigree due to multiple gene copies.

Figure 3.22: Pedigrees of N/N^{B} families with novel mutation events

The different coloured boxes represent the copy number of different probe regions. One coloured box indicates a single copy of a region. The *SMN1/2*, exons 1, 4, 6 and 8; *NAIP*, exon 13 and *GTF2H2*, exon 5 regions represent combined results for the telomeric and centromeric copies of these regions. An X indicates a deletion of the specific region.

3.4.2 M_1/M_1^B families

Only 44% (11/25) of M_1/M_1^B families were completely informative where the phase of CNVs could be assigned with certainty (figure 3.23 shows some examples). For 48% (12/25) of these M_1/M_1^B families, multiple combinations of CNVs were possible, due to the presence of multiple copies of one or more probes (figure 3.24 shows an example). CNV data on additional relatives could potentially improve the assignment of phase in these families. Discrepant results were observed in 8% (2/25) of M_1/M_1^B families which could be due to a variety of causes such as non-paternity or novel deletion or duplication events in the proband (figure 3.25 shows pedigrees of these two families). Section 2.3.4 explains the requirements of family pedigrees to be classified as having discrepant or novel results.

3.4.2.1 Pathogenic *SMN1*, exon 7 deletion haplotypes

In total, 22 pathogenic haplotypes were constructed from probands from M_1/M_1^B families. Of these, 17 unique haplotypes were identified and are summarised in Appendix H. Even though no clear pathogenic CNV pattern associated with SMA emerged, some generalisations could be drawn.

The telomeric region: *SMN1*, exons 7, 8 and *NAIP*, exon 5

It was determined that 63.6% (14/22) of haplotypes had larger deletions including *SMN1*, exon 8 and 31.8% (7/22) had even larger deletions including *SMN1*, exon 8 and *NAIP*, exon 5 (which are both thought to co-locate with *SMN1*).

Five unrelated individuals from M_1/M_1^B families, exon 7 had a confirmed 2:0 silent carrier genotype (as explained in section 1.2.2.4). Figure 3.23b shows an example of such a family (individuals I1, I2 and II 2). The parents and fetus II in this family are all silent carriers of SMA since they have the ability to pass on an *SMN1*, exon 7 deletion to subsequent generations.

Two unrelated individuals from M_1/M_1^B families had uninformative results for *SMN1*, exon 7, where the individual could either have one copy of *SMN1*, exon 7 on each of his/her chromosomes (1:1 genotype) or two copies of *SMN1*, exon 7 on one chromosome in conjunction with no copies of *SMN1*, exon 7 on the other

chromosome (2:0 genotype) (figure 3.24, individual I2 shows an example of such a case).

The centromeric region: *SMN2*, exons 7 and 8

An *SMN1*, exon 7 deletion in conjunction with an *SMN2*, exon 8 deletion was observed in 18.2% (4/22) of haplotypes, suggesting partial gene conversion from *SMN2*, exon 8 to *SMN1*, exon 8 on these chromosomes.

Only 1 haplotype included deletions of both *SMN1* exon 7 and 8 in conjunction with 4 copies of both *SMN2*, exon 7 and 8, suggesting a full gene conversion from *SMN1* to *SMN2*. Two haplotypes had a large deletion encompassing both *SMN1*, exon 7 and 8, as well as *SMN2*, exon 7 and 8. One of these major deletions included *SMN1* and/or *SMN2* exons 1, 4, 6 and 8, *NAIP*, exons 5 and 13, *SERF1B* and *GTF2H2*, exon 5. It is important to note that no homozygous deletions of both *SMN1*, exon 7 and 8, in addition to *SMN2*, exon 7 and 8 have been observed in any of the patient and subject groups, possibly being incompatible with life.

Combined regions: *SMN1/2*, exons 1, 4, 6 and 8; *GTFTH2*, exon 5 and *NAIP*, exon 13 regions

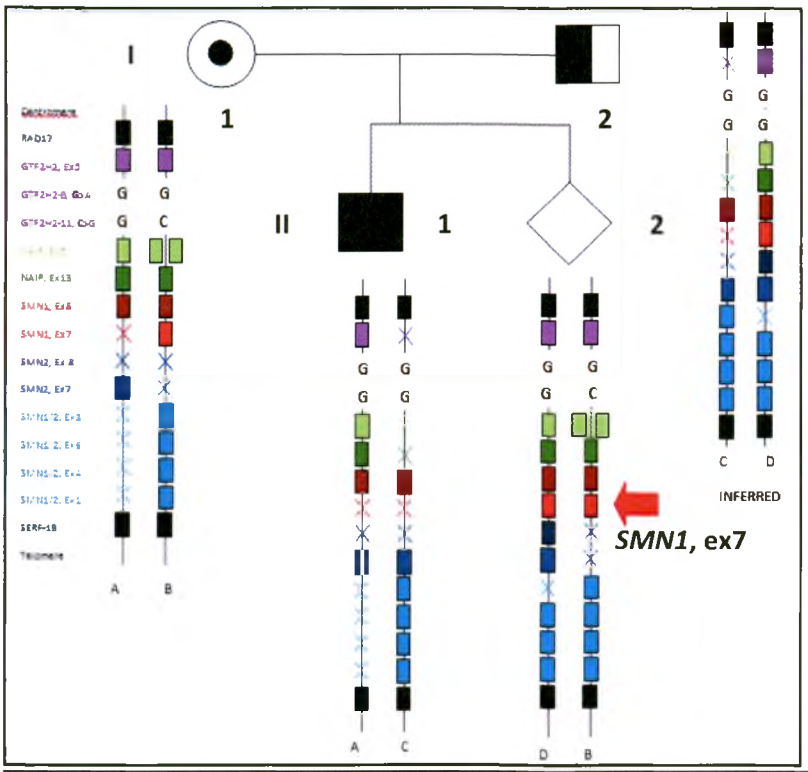
Large deletions of *SMN1* and/or *SMN2*, exons 1, 4, 6 and 8 regions were observed in 36.4% (8/22) of M_1/M_1^B family haplotypes, suggesting full gene deletions of *SMN1* in these cases. A further 27.3% (6/22) of haplotypes had *GTF2H2*, exon 5 deletions and 13.6% (3/22) had *SERF1A* deletions.

Telomeric-*GTF2H2* and centromeric-*GTF2H2*

A full complement (1 copy per chromosome) of centromeric-*GTF2H2* was observed in 95.5% (21/22) of M_1/M_1^B family haplotypes. Deletions of telomeric-*GTF2H2* were observed in 77.3% (17/22) of haplotypes, confirming an association of *SMN1*, exon 7 deletions with telomeric-*GTF2H2* deletions as discussed in section 3.2.4.4.

3.4.2.2 Non-pathogenic haplotypes

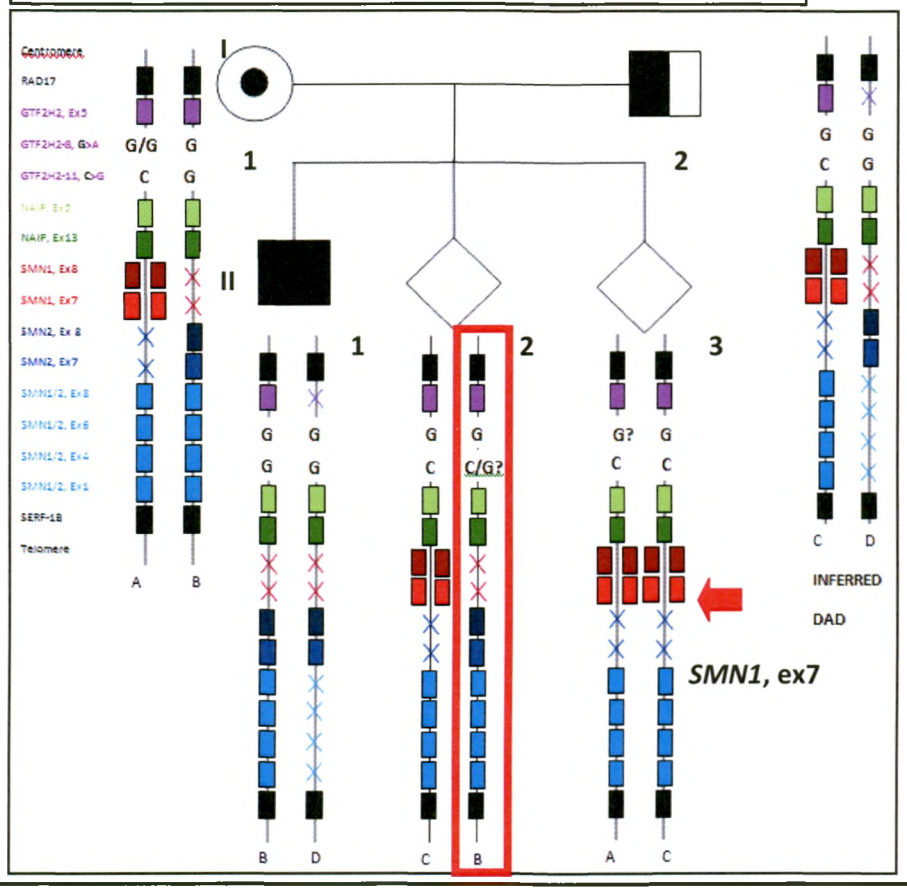
Twenty non-deletion (negative) haplotypes were identified from the parents of affected probands. The number of haplotypes does not correspond to the 22 pathogenic haplotypes identified because the fathers of two families were not available and the paternal haplotypes could not be inferred due a lack of family data. The negative parental haplotypes from M_1/M_1^B families were compared to negative haplotypes from N/N^B families (described in section 3.4.1) using a MANOVA statistical test and interestingly were found to differ significantly for a single probe, *NAIP*, exon 5 ($p=0.01086$; $F=6.75$). The non-deletion (negative) haplotypes are available from Appendix I and the statistical calculations are summarised in Appendix F, statistical test output 8 and will be discussed further in section 4.3.3.



a.) Informative family pedigree

Individual II-1 is affected with SMA caused by a homozygous deletion of *SMN1*, exon 7 and fetus II-2 is not affected with SMA.

This pedigree shows an example of a clear informative inheritance pattern where the phase of the SMN region could be accurately determined.

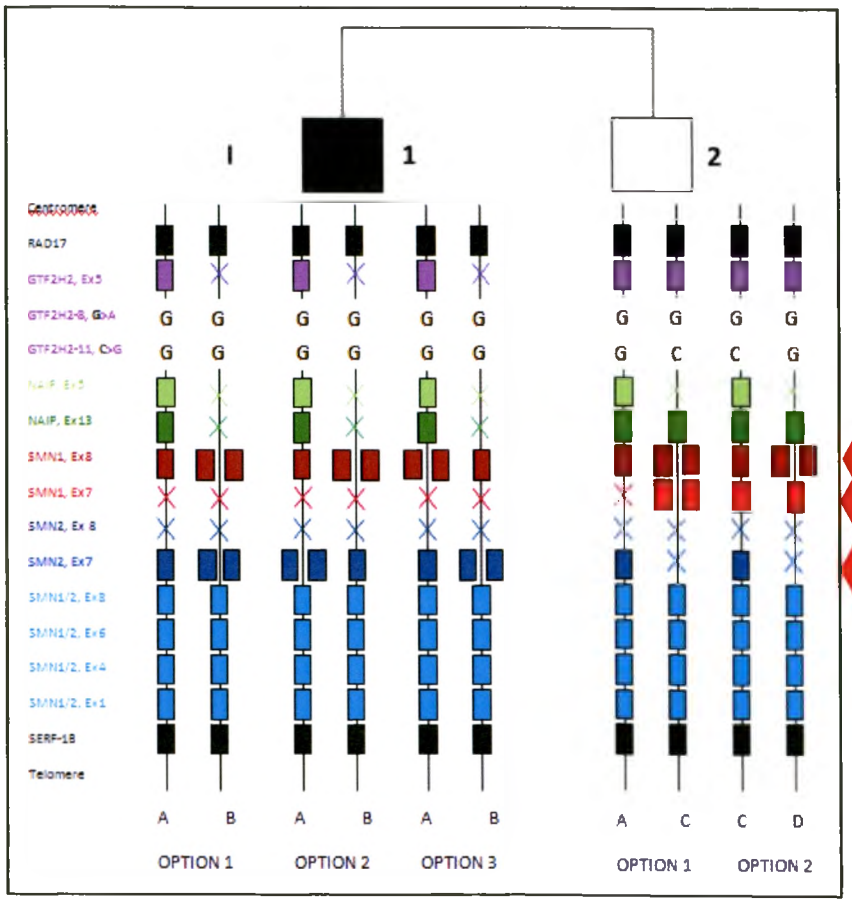


b.) Prenatal testing in an informative family

This pedigree shows an example of an informative pedigree where the phase of the SMN region could be accurately determined. The SMA status was accurately determined in two prenatal tests. Fetus II-2 has two copies and fetus II-3 has four copies of *SMN1*, exon 7. Both fetuses tested negative for SMA. Fetus II-2 is a silent carrier (indicated with a red block) with two copies of *SMN1* on a single chromosome in conjunction with zero copies on the other chromosome.

Figure 3.23: Pedigrees of informative M_1/M_1^B families

The different coloured boxes represent the copy number of different probe regions. One coloured box indicates a single copy of a region. The *SMN1/2*, exons 1, 4, 6 and 8; *NAIP*, exon 13 and *GTF2H2*, exon 5 regions represent combined results for the telomeric and centromeric copies of these regions. An X indicates a deletion of the specific region.



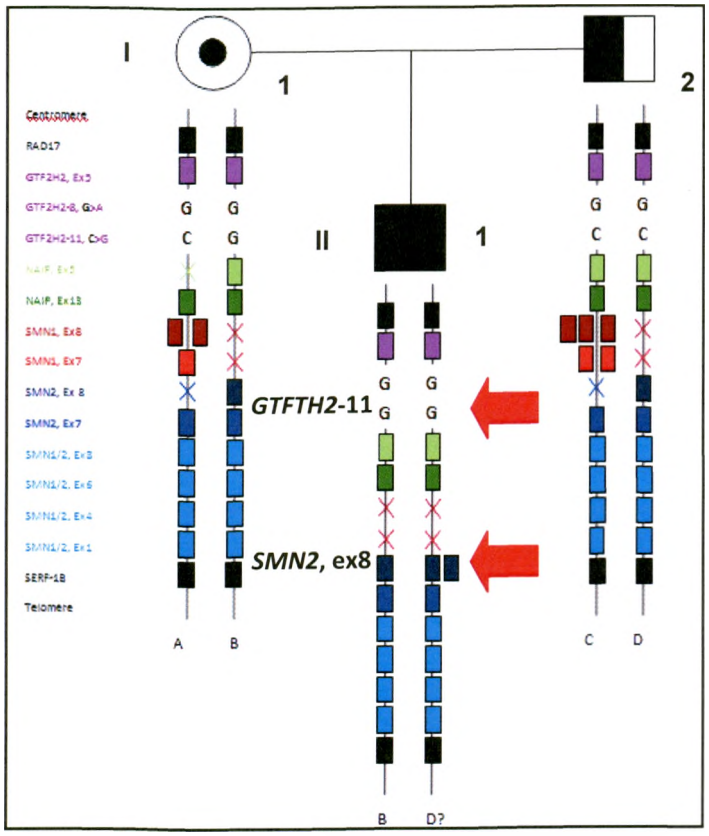
This pedigree shows an example of an uninformative pedigree where the phase of the SMN region could not be accurately determined due to the presence of multiple gene copies (specifically *SMN1*, exon 7 and *SMN2*, exon 8). There are multiple potential options/scenarios which could explain the results in this family. Note that there are no data available on the parents which could potentially assist with the CNV inheritance pattern.

Individual I-1 is affected with SMA and has a homozygous deletion of *SMN1*, exon 7 but the phase of the multiple copies of *SMN1*, exon 7 and *SMN2*, exon 8 is unknown. Individual I-2 is not affected with SMA and has two copies of *SMN1*, exon 7. This individual could either have one copy on each chromosome 5 (1:1) or have two copies on one chromosome 5 (2:0), meaning that he could be a silent carrier of SMA.

This is also an example of a possible gene conversion since *SMN1*, exon 7 and *SMN2*, exon 8 are homozygously deleted in conjunction with multiple copies of *SMN2*, exon 7 and *SMN1*, exon 8. It appears as if a hybrid *SMN1*, exon 7 and *SMN2*, exon 8 gene was deleted.

Figure 3.24: Pedigree of an uninformative M_1/M_1^B family with multiple gene copies

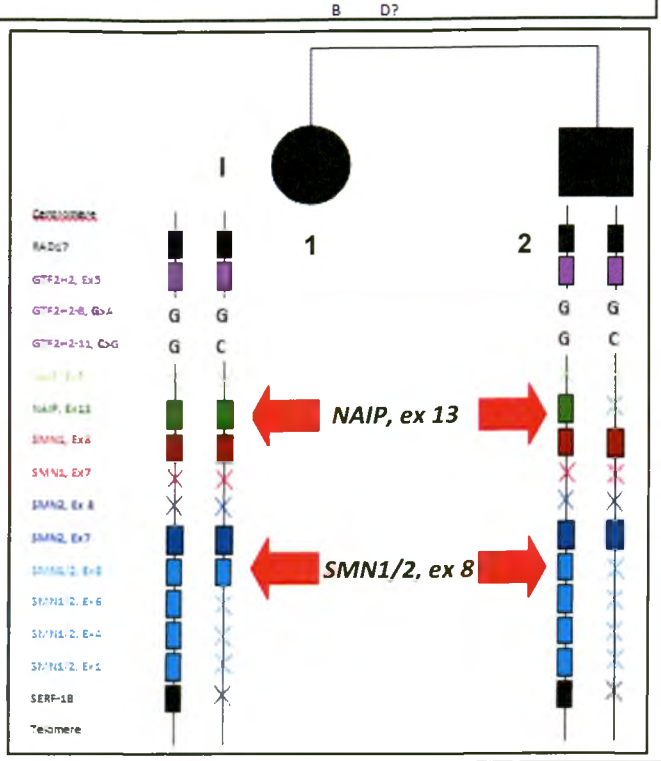
The different coloured boxes represent the copy number of different probe regions. One coloured box indicates a single copy of a region. The *SMN1/2*, exons 1, 4, 6 and 8; *NAIP*, exon 13 and *GTF2H2*, exon 5 regions represent combined results for the telomeric and centromeric copies of these regions. An X indicates a deletion of the specific region.



a.) The affected child (II-1) has 3 copies of *SMN2*, exon 8 which do not seem to fit in context of the parental chromosomes. One copy was inherited from the mother (I-1) and at least one copy from the father (I-2).

A second discrepancy is seen at the *GTFTH2-11* probe. The affected child is homozygous for a G allele/deletion at this position. One G allele was inherited from the mother. The father is homozygous for the C allele and could not have passed on a G allele.

The third copy of *SMN2*, exon 8 and the additional G allele in the child is most likely explained by non-paternity.



b.) The two affected children (I-1 and I-2) have discrepant results even though they are expected to share the same mutant chromosomes. The affected girl (I-1) has two copies of SMN1/2, exon 8 while her affected brother has only one copy of this region. In addition, the girl has two copies of *NAIP*, exon 13 whereas the brother has only one copy. These discrepancies could be due to novel events in one / both children. Non-paternity cannot be excluded, although it is unlikely that the children have two different biological fathers who both happen to be carriers of a deletion *SMN1*, exon 7.

Figure 3.25: Pedigrees of uninformative M_1/M_1^B families with discrepant results

The different coloured boxes represent the copy number of different probe regions. One coloured box indicates a single copy of a region. The *SMN1/2*, exons 1, 4, 6 and 8; *NAIP*, exon 13 and *GTF2H2*, exon 5 regions represent combined results for the telomeric and centromeric copies of these regions. An X indicates a deletion of the specific region.

3.5 RNA extraction

Since patient blood samples often take days and sometimes weeks to be transported to the Division for diagnostic testing, a reliable and simple method is needed to protect the global RNA expression pattern at the time of blood collection.

RNA was successfully extracted from blood collected in Tempus Blood RNA tubes (as described in section 2.3.5.2) which contain a stabilising agent which deactivates natural RNase enzymes in the blood and preferentially precipitates RNA molecules. RNA expression profiles are said to be stable in Tempus Blood RNA Tubes for up to five days at room temperature and for at least seven days at 4°C. Since DNA yield improves during salting out DNA extraction if the blood has been frozen, a Tempus Blood RNA tube was stored at -70°C for two months to test long term storage. Figure 3.26 shows the gel electrophoresis results, which confirms the quality of the RNA at storage with different conditions. Good RNA quality was confirmed in all samples using the Nanodrop spectrophotometer (Nanodrop, Wilmington, Delaware, USA). A 260/280 ratio of ~2.1 indicates good quality RNA.

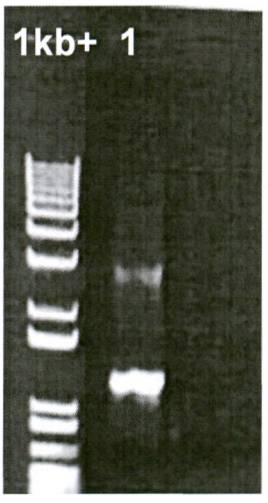


Figure 3.25a

28S rRNA

18S rRNA

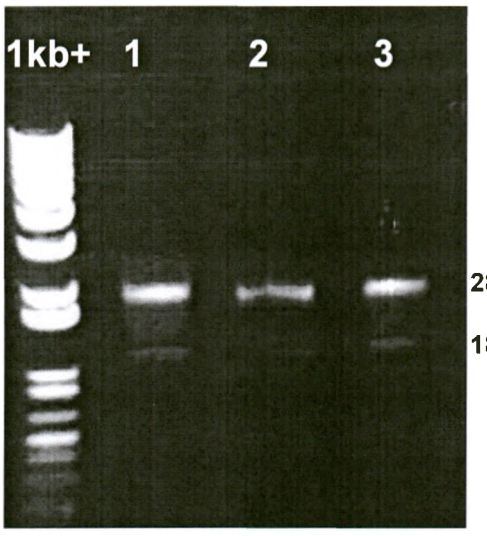


Figure 3.25b

28S rRNA

18S rRNA

a.) Lane 1: RNA extracted from Tempus Blood RNA tube immediately after collection.

b.) Lane 1: RNA extracted from Tempus Blood RNA tubes left for five days at room temperature, Lane 2: RNA extracted from Tempus Blood RNA tubes left for seven days at 4°C, Lane 3: RNA extracted from Tempus Blood RNA tubes left for two months at -70°C

Figure 3.26: Gel electrophoresis results illustrating RNA extraction

3.6 Provisional real-time PCR results

qPCR was performed in order to optimise a DNA-based assay to identify *SMN1* and *SMN2* copy numbers as a quicker and cheaper alternative to MLPA analysis (described in section 2.3.5.1). The assay was designed to distinguish between exon 7 of *SMN1* and *SMN2* using differently labelled fluorescent probes and comparing the fluorescence with a *CFTR* endogenous control in a single multiplex reaction. Individuals with known *SMN1* and *SMN2* copy numbers, determined by MLPA analysis, were used as controls to optimise the assay. All samples were set up in triplicate.

The *SMN1*, exon 7 simplex assay shows some promise with *SMN1* fluorescent signals increasing exponentially to the *SMN1* copy number, although there seems to be some overlap in results of different copy numbers. The *SMN2* copy number could not be accurately determined using this assay and therefore the assay needs further optimisation. Figure 3.27 shows a snapshot of the *SMN1*, exon 7 qPCR results.

Once the *SMN1*, exon 7 and *SMN2*, exon 7 multiplex is fully optimised, it could be used as an alternative diagnostic test for SMA. This assay could be useful for diagnostic and prenatal testing as it will identify homozygous deletions of *SMN1*, exon 7. Zero *SMN1*, exon 7 copies could either represent a failed qPCR reaction or a homozygous deletion of *SMN1*, exon 7. Zero *SMN1*, exon 7 copies in combination with one or more *SMN2*, exon 7 copies indicates that the PCR reaction worked and that the patient tested positive for a homozygous deletion of *SMN1*, exon 7. Zero copies of *SMN1*, exon 7 and SMN2, exon 7, indicates that the PCR reaction failed. It is unlikely that *SMN1*, exon 7 and *SMN2*, exon 7 will both be homozygously deleted since it might be incompatible with life as explained in section 3.4.2.1.

The multiplex assay would also be useful in identifying carriers, since it is quantitative and would be able to detect heterozygous deletions of *SMN1*, exon 7. Silent carriers with two or more copies of *SMN1*, exon 7 will not be identified using this assay, unless analysed within a family context. Furthermore, the *SMN2*, exon 7 copy number can be determined in patients who are affected with SMA in order to determine their appropriateness for inclusion in clinical trials which increase the production of FL-

SMN transcripts from multiple SMN2 gene copies (as explained in section 1.2.3). This assay is not able to identify other pathogenic CNVs.

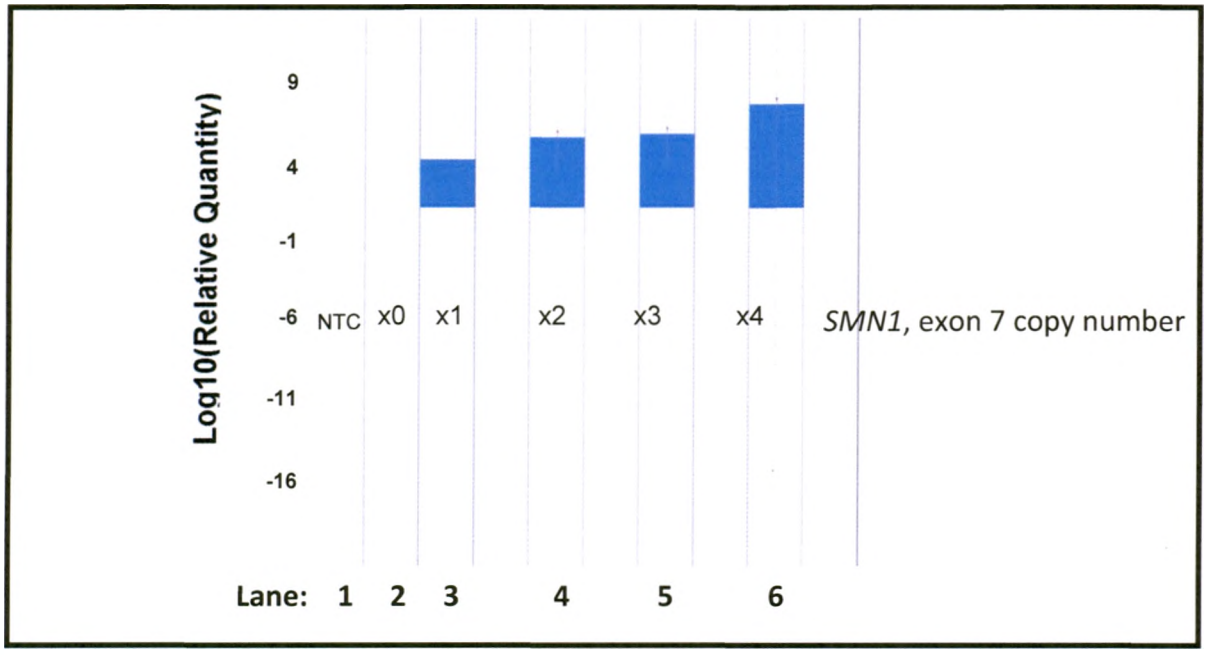


Figure 3.27: qPCR results indicating *SMN1* copy number

The X axis represents different patients with known copy SMN1/2 numbers. The *SMN1*, exon 7 copy number is indicated (e.g. four copies = x4). The Y axis represents the relative fluorescence as compared with the *CFTR* endogenous control. Lane 1: NTC (no template control); Lane 2: patient with zero *SMN1*, exon 7 copies; Lane 3: patient with one *SMN1*, exon 7 copy; Lane 4: patient with two *SMN1*, exon 7 copies; Lane 5: patient with three *SMN1*, exon 7 copies; Lane 6: patient with four *SMN1*, exon 7 copies.

Table 3.10 shows a comparison of the cost for three different techniques (the current diagnostic homozygous *SMN1*, exon 7 deletion assay, MLPA and qPCR). The current diagnostic homozygous *SMN1*, exon 7 deletion assay (described in section 1.2.2.1) is much cheaper than MLPA and qPCR although MLPA and qPCR have the advantage of being quantitative techniques, which would be able to determine the copy number of *SMN1*, exon 7 and *SMN2*, exon 7. This cost comparison will be further discussed in section 4.5.

Table 3.10: Comparison of cost of the diagnostic homozygous *SMN1*, exon 7 deletion assay, qPCR and MLPA

Assay/technique	Reagent cost per patient	Labour cost per patient	Final cost per patient
Diagnostic homozygous *SMN1*, exon 7 deletion assay	R11.46	R77.77 4.5 hours/batch over 2 days	R89.23
qPCR (set up in triplicate)	R16.26	R77.77 3.5 hours/batch on 1day	R94.03
MLPA	R314.62	R148.24 7 hours/batch over 2-3 days	R462.86

A batch consists of ten patient samples and positive and negative controls

See Appendix K for the calculations of this costing comparison.

4 DISCUSSION

A major limitation of previous quantitative studies of the SMN1 gene performed in African-American (Hendrickson *et al.*, 2009) and sub-Saharan African populations (Sangaré *et al.*, 2014) was that CNV analysis was performed in unaffected individuals only. This study focuses on comparing CNVs in black individuals who are negative for SMA (N/N^B) in order to identify non-pathogenic CNVs as well as individuals with known homozygous *SMN1* and *SMN2*, exon 7 deletions (M_1/M_1^B and M_2/M_2^B, respectively) and patients who are clinically suggestive of SMA (U/U^B) in order to delineate potential pathogenic CNVs.

The results of the audit performed on samples referred to the Division for SMA testing will be discussed in section 4.1. General trends in CNVs determined by MLPA analysis for each population group (black and white SA populations) as well as between different subject groups (Groups A - G) will be discussed in sections 4.2.1 to 4.2.4. The results of pedigree and haplotype analysis in M_1/M_1^B and N/N^B families will be discussed in section 4.3.

As a proof of concept, the provisional results for RNA analysis and qPCR will be discussed in sections 4.4 and 4.5, respectively. The implications for diagnostic testing will be discussed in section 4.6 and finally, the challenges and limitations of this study will be explored in section 4.7.

4.1 Audit on patients referred for diagnostic SMA testing

4.1.1 Homozygous deletions of *SMN1*, exon 7

Homozygous *SMN1*, exon 7 (telomeric) deletions were identified in an average of 31.7% of patients referred for SMA testing across all population groups (refer to section 3.1 for results) and did not differ significantly among different population groups. Currently a diagnosis of SMA therefore cannot be confirmed or refuted in 68.3% of referred patients across all population groups.

Since the *SMN1*, exon 7 homozygous deletion frequency does not appear to be different in black patients when compared to other SA population groups, it is possible

that black patients who test negative on the diagnostic assay do not in fact have SMA, but may be presenting with other similar neuromuscular disorders. These results appear to contradict previous SA reports of a lower homozygous deletion of *SMN1*, exon 7 frequency observed in black SA SMA cases (51%) when compared to frequencies worldwide (~94%) but this previous frequency was generated from results of a cohort of patients who were assessed by a paediatric neurologist and determined to be clinically suggestive of SMA (Labrum *et al.*, 2007).

Many of the request forms for testing for SMA have little or no clinical information and often samples are referred for multiple genetic tests with overlapping clinical symptoms with SMA such as hypotonia (e.g. Prader Willi syndrome and myotonic dystrophy) questioning whether these patients are being "over-serviced" and whether appropriate testing is being done.

The positive pickup rate for patients affected by skeletal muscle disorders is higher using NGS strategies when compared to gene-by-gene analysis (Ankala et al., 2015; Nigro and Savarese, 2016). In a South African context, referrals for neuromuscular genetic testing are not always appropriate and many patients are not referred by neurology clinics, which might result in delayed diagnoses and a high cost associated with gene-by-gene testing.

Furthermore, 16 additional genes with overlapping phenotypes to SMA have been shown to be associated with non-5q forms of SMA (Peeters et al., 2014). It may be more appropriate to perform testing by a NGS neuromuscular panel first to exclude other related neuromuscular diseases and other causes of SMA before continuing SMA testing in individuals who test negative for the homozygous *SMN1*, exon 7 deletion. This will need to be evaluated for clinical utility before being instituted as a diagnostic test, as NGS panel testing is expensive.

4.1.2 Homozygous deletions of *SMN2*, exon 7

Homozygous *SMN2*, exon 7 (centromeric) deletions were significantly more common in the black and mixed ancestry populations (12.4% and 18.8%, respectively) when compared to the white and Indian populations (4.7% and 4%, respectively); (results are available in section 3.1.3). Interestingly, the frequency of homozygous *SMN2*,

exon 7 deletions was even higher in N/N^B individuals (27%); (refer to section 3.2.1.3 for results of N/N^B individuals).

The homozygous *SMN2*, exon 7 deletion frequency discrepancy between M_2/M_2^B and N/N^B individuals, could potentially be indicative of two different types of CNVs involving homozygous *SMN2*, exon 7 deletions. A non-pathogenic CNV resulting in a homozygous *SMN2*, exon 7 deletion could form part of the normal variation present in the black SA population. A different pathogenic CNV, perhaps involving the homozygous loss of genes converted from *SMN1* to *SMN2*, would also result in a homozygous *SMN2*, exon 7 deletion and may be the cause of SMA in M_2/M_2^B individuals. SMN1 genes could potentially be masquerading as SMN2 genes.

4.2 MLPA analysis – general trends in CNVs

The next few sections will discuss the results described in section 3.2

Black SA individuals were found to have a high variability in CNV when compared to white SA individuals across multiple MLPA probe regions in the SMN region. More specifically, multiple copies of the telomeric region including *SMN1*, exon 7, exon 8 and *NAIP*, exon 5 were more commonly found in black individuals when compared to white individuals which could complicate diagnostic, carrier and prenatal testing in the black SA population. Results of this study emphasises the limited knowledge we have of the complex architecture of the SMN region particularly in the hypervariable black SA population. The next few sections will look at these outcomes in more detail.

Section 4.2.1 will discuss results of the comparative analysis of Groups B (N/N^B) and C (N/N^W); section 4.2.2 will discuss results of the comparative analysis of Groups D (M_1/M_1^B) and E (M_1/M_1^W); section 4.2.3 will discuss results of the comparative analysis of Groups F (M_2/M_2^B) and G (M_2/M_2^W) and lastly, section 4.2.1 will discuss results of the comparative analysis of Group A (U/U^B) with Groups B (N/N^B), D (M_1/M_1^B) and F (M_2/M_2^B).

4.2.1 Groups B (N/N^B) and C (N/N^W)

4.2.1.1 Multiple copies of the telomeric region (*SMN1*, exons 7 and 8 and *NAIP*, exon 5) were observed in N/N^B individuals and could complicate analysis

This section will discuss the MLPA results from section 3.2.1. In this study, 50.8% of N/N^B individuals were found to have multiple (3-6) copies of *SMN1*, exon 7, which is similar to previous reports in African-American individuals (Hendrickson *et al.*, 2009) and sub-Saharan African populations (Mali, Nigeria and Kenya) (adapted from Sangaré *et al.*, 2014).

In contrast, N/N^W individuals were found to have a lower percentage of multiple *SMN1*, exon 7 copies of 3.3%, which is similar to previous reports in white North-American populations (Hendrickson *et al.*, 2009) and European populations (Germany, France and Sweden) (Corcia *et al.*, 2012; Feldkötter *et al.*, 2002). Table 3.2 and Figure 3.4, section 3.2.1.1 shows a detailed comparison between various international studies.

The high rates of multiple copies of *SMN1*, exon 7 in N/N^B individuals could be a consequence of the hypervariability of the SMN region in the black SA population, (further discussed in sections 4.3.2 and 4.7.1).

Similarly, multiple copies of *SMN1*, exon 8 and *NAIP*, exon 5 were more frequently observed in N/N^B individuals (54.9% and 37.7%, respectively) when compared to N/N^W individuals (6.7% and 13.3%, respectively) (described in section 3.2.1.1). It is interesting to note that the copy numbers of *SMN1*, exon 7 and 8 and *NAIP*, exon 5 do not correlate, which could indicate the presence of non-contiguous copies of these genes (further discussed in section 4.2.4.3).

Multiple *SMN1*, exon 7 copies confound analysis since the location, phase and functionality of these multiple gene copies are unknown. Multiple *SMN1*, exon 7 copies could complicate carrier testing by masking silent carrier profiles (further discussed in section 4.2.1.2) and carriers of potential unidentified pathogenic CNVs (further discussed in section 4.2.4.3).

4.2.1.2 Silent carriers of SMA may be masked by the presence of multiple *SMN1*, exon 7 copies in N/N^B individuals

A study performed on unaffected individuals from various sub-Saharan African populations (Kenyan, Malian and Nigerian) using qPCR reported a slightly lower carrier frequency of deletions of *SMN1*, exon 7 (0.5-2%) in sub-Saharan populations when compared to carrier frequencies of 1-5% also determined using qPCR in the European and Asian populations (Sangaré *et al.*, 2014).

Interestingly, no heterozygous deletions of *SMN1*, exon 7 were identified in either N/N^B or N/N^W individuals in this study (section 3.2.1.2). These rates are much lower than the previously predicted SA carrier rate of 1/50 in the black population and 1/23 in the white population (Labrum *et al.*, 2007). A small sample size (n=30) could have caused this discrepancy in N/N^W individuals in this study. The discrepancy in N/N^B individuals could be due to two different reasons. Firstly, MLPA analysis is a very robust technique, which has built-in statistical tests and extensive normalisation, which are likely to yield more reliable and accurate results than the previously used in-house dosage system (Labrum *et al.*, 2007). In support of this hypothesis, four individuals previously reported to have heterozygous deletions of *SMN1*, exon 7 on the dosage assay were retested on MLPA and all four individuals tested negative for heterozygous deletions of *SMN1*, exon 7 (results in Appendix J). Two of these individuals had three copies of *SMN1*, exon 7, which were mistaken for one copy on the previous dosage assay.

Secondly, it is possible that there is a higher frequency of silent carriers (2:0, 3:0, 4:0, 5:0, 6:0, etc.) in the black SA population, not detectable by either of these two assays. Multiple copies of *SMN1*, exon 7 could mask carrier profiles. MLPA cannot provide information on the location or phase of multiple *SMN1*, exon 7 copies on an individual's chromosomes and therefore these multiple copies could be located on a single chromosome, resulting in a silent carrier profile. Figure 4.1 illustrates this concept and section 1.2.2.4 defines the term "silent carrier".

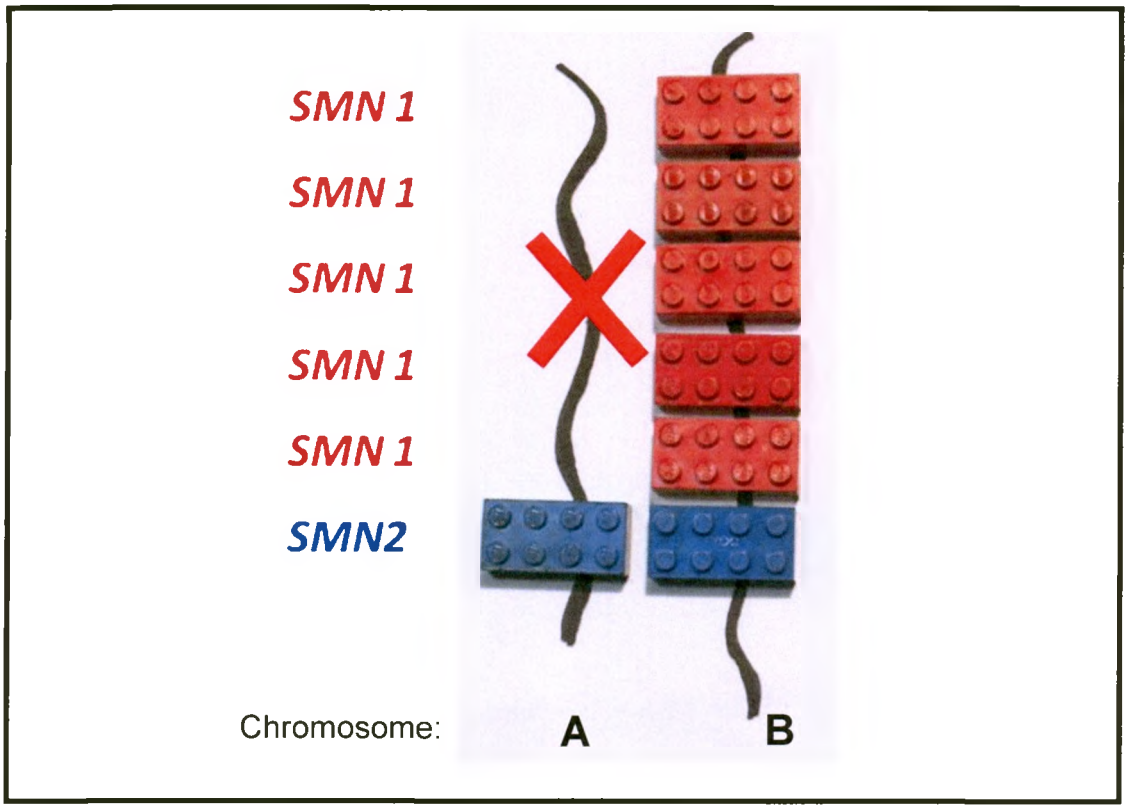


Figure 4.1: The presence of multiple copies of *SMN1* could mask deletion chromosomes in silent carriers of the *SMN1*, exon 7 deletion
The red Lego blocks represent *SMN1*, the blue Lego blocks represent *SMN2* and a cross represents a deletion. Two chromosomes are shown here (A + B). This individual has two copies of *SMN2*, one copy on each chromosome (A + B) and five copies of *SMN1*, all located on chromosome B in conjunction with a deletion of *SMN1* on chromosome A. This individual is therefore a silent carrier since he/she has a risk of passing on the deletion chromosome (A) to the next generation even though he/she has multiple copies of *SMN1*, exon 7.

4.2.1.3 Homozygous *SMN2*, exons 7 and 8 deletions could form part of the normal variation in N/N^B individuals

The high frequency of homozygous *SMN2*, exon 7 (centromeric) deletions in N/N^B individuals (27%) when compared to N/N^W individuals (6.7%) suggests that these deletions could form part of non-pathogenic CNVs present in the black SA population.

Furthermore, the copy numbers of *SMN2*, exons 7 and 8 do not correlate, suggesting that these two regions may not be contiguous due to possible gene conversions from *SMN2*, exon 8 to *SMN1*, exon 8 which would result in a loss of *SMN2*, exon 8 and an increased copy number of *SMN1*, exon 8 (see section 3.2.4.3, table 3.8 for a summary of results).

4.2.1.4 A high variability in copy number of *SMN1/2* exons and neighbouring genes could be indicative of large CNVs in N/N^{B} individuals

Although rare, N/N^{B} individuals seemed to have larger deletions and duplications stretching into the rest of the SMN region than N/N^{W} individuals, suggesting that potential large rearrangements of the SMN region could form part of the general variability of the SMN region in the black SA population (refer to section 3.2.1.4).

Since the *SMN2,* exon 7 and 8 regions have more frequent deletions than the *SMN1*, exon 7 and 8 regions, it is hypothesised that the deletions observed in the *SMN1/2*, *NAIP/NAIPΨ*, exon 13 and *GTF2H2*, exon 5 combined regions most probably represent deletions of *SMN2* extending into the rest of the centromeric SMN region. This hypothesis cannot be confirmed as the P021-A2 MLPA kit cannot distinguish between the centromeric and telomeric copies of these regions.

The SMN region is too large to be cloned in its entirety but Yu-Jin *et al.* described a technique by which the cDNA of patients affected with SMA was subcloned. SMN1 and SMN2 subclones were differentiated by restriction enzyme digest and the clones were sequenced in order to identify subtle SMA-causing mutations (Yu-Jin et al., 2012). It may be possible to use this technique, to perform specific analysis of *SMN1* and *SMN2* and potentially of *NAIP/NAIPΨ* and telomeric/centromeric-*GTF2H2.* The complexity of the SMN region, the presence of partial genes and other rearrangements in this region may make this extremely challenging.

4.2.1.5 Multiple copies of *SMN1*, exon 7 may be incomplete, non-contiguous gene copies which may not be functional

N/N^{B} individuals had a significantly higher variability than N/N^{W} individuals for the majority of MLPA probes (section 3.2.1.6). Whereas the copy number of *SMN1*, exons 7, 8 and *NAIP*, exon 5 of N/N^{W} individuals seem to cluster at two copies, as expected; the copy number of *SMN1*, exons 7, 8 and *NAIP*, exon 5 of N/N^{B} individuals varies extensively from one to six (refer to figure 3.7 in section 3.2.1.6). Perhaps some of these *SMN1* gene copies are partial disconnected gene copies which may be

non-functional due to interruptions in the coding region and will be further discussed in section 4.2.4.3. See figure 4.2 for an illustration of the possible composition of CNVs of the SMN region in N/N^{B} individuals.

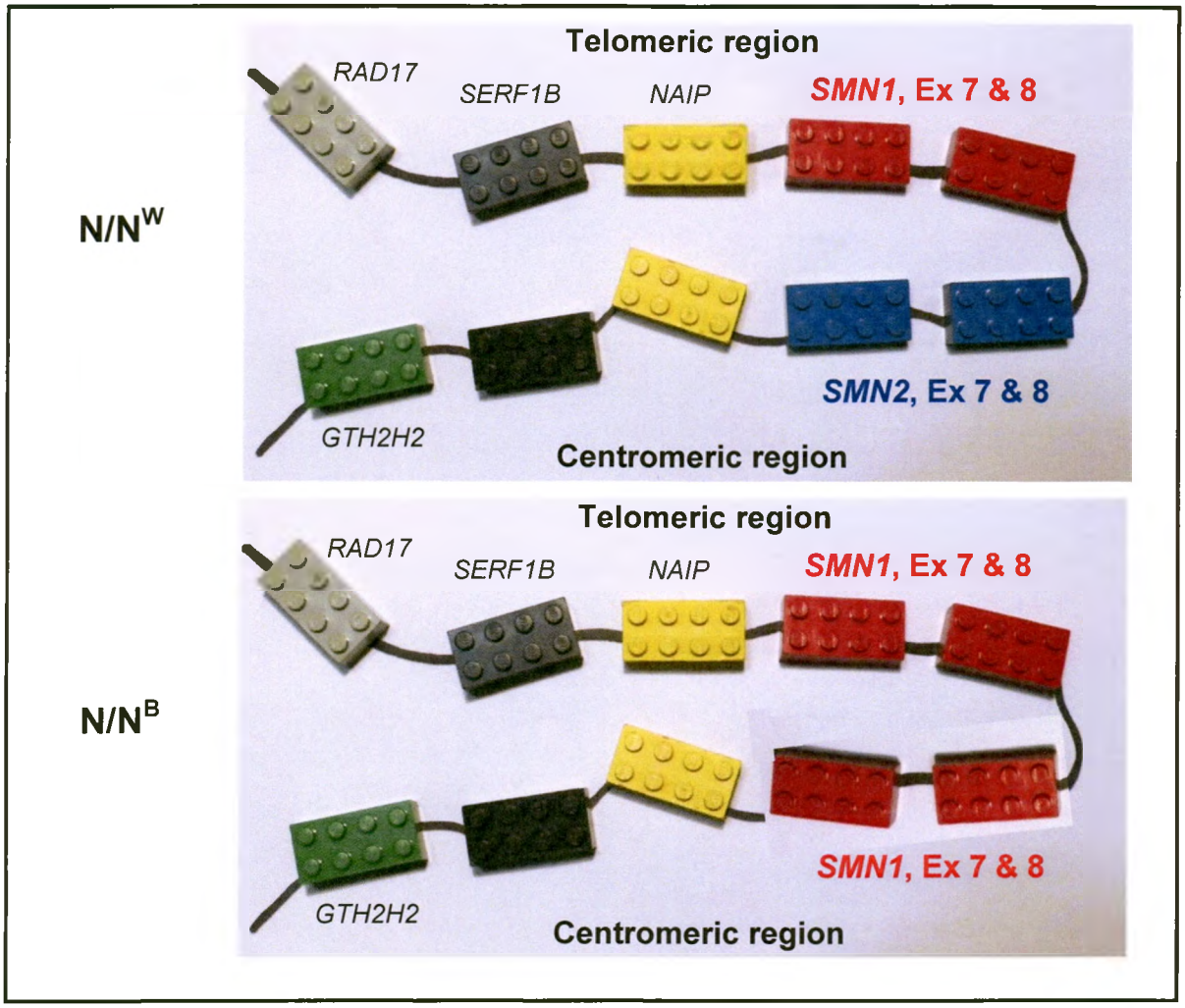


Figure 4.2: The SMN region in N/N^{B} individuals is more variable than N/N^{W} individuals The red Lego blocks represent the *SMN1* gene and the blue Lego blocks represent the *SMN2* gene. N/N^{B} individuals more frequently have multiple copies of *SMN1*, exon 7 in conjunction with *SMN2*, exon 7 deletions than N/N^{W} individuals.

4.2.2 Groups D ($M_1/M_1{}^{B}$) and E ($M_1/M_1{}^{W}$)

This section will discuss the MLPA results from section 3.2.2. As expected, MLPA analysis confirmed homozygous deletions of *SMN1*, exon 7 in all $M_1/M_1{}^{B}$ and $M_1/M_1{}^{W}$ individuals (section 3.2.2.1).

4.2.2.1 In contrast to previous SA reports, large deletions extending into the rest of the SMN region appear to be more common in M_1/M_1^B individuals

Homozygous deletions of *SMN1,* exon 8 have been reported to be associated with 93% of homozygous deletions of *SMN1,* exon 7 (Lefebvre *et al.*, 1995). Furthermore, *NAIP* deletions have been associated with *SMN1* deletions in 67% of SMA type I patients (Roy *et al.*, 1995). A previous SA study proposed that M_1/M_1^W individuals had larger homozygous deletions of *SMN1,* exon 7 also encompassing *SMN1*, exon 8 and *NAIP* more often than M_1/M_1^B individuals (Labrum *et al.*, 2007). Furthermore, M_1/M_1^B individuals were reported to have a homozygous *SMN1*, exon 7 deletion in conjunction with a homozygous *NAIP* deletion (*SMN1*, exon 8 is present), suggestive of gene conversion (discussed in section 1.2.2.3) from *SMN1,* exon 7 to *SMN2,* exon 7, more often than M_1/M_1^W individuals (Labrum *et al.*, 2007; Stevens *et al.*, 1999).

When only looking at information from *SMN1*, exon 7, exon 8 and the *NAIP* gene, a lower frequency of large deletions and a higher frequency of potential gene conversions (*SMN1*, exon 7 deletion in conjunction with a *NAIP* deletion) were predicted in M_1/M_1^B individuals when compared to M_1/M_1^W individuals (table 3.4, section 3.2.2.1) which supports the results of previous SA studies. MLPA probes cover a larger part of the SMN region than previous studies, giving us additional clues on the extent of CNVs.

NAIP and *GTF2H2* deletions have been observed in patients with SMA (Carter *et al.*, 1997; Roy *et al.*, 1995)) and could therefore provide some information on the extent of *SMN1*, exon 7 deletions. A higher frequency of homozygous deletions into the rest of *SMN1/2* (exons 1, 4, 6 and 8), *NAIP/NAIPΨ*, exon 13 and *GTF2H2*, exon 5 were observed in M_1/M_1^B individuals (60%, 61.3%, 62.7% 61.3%, 61.3%, 66.7% respectively) compared to M_1/M_1^W individuals (31.3%, 31.3%, 31.3%, 36.7%, 23.3%, 50%); (refer to sections 3.2.2.3 and 3.2.2.4). Since the *SMN1,* exon 7 and 8 regions have more frequent deletions than the *SMN2*, exon 7 and 8 regions, it is hypothesised that the deletions observed in the *SMN1/2*, *NAIP*, exon 13 and *GTF2H2*, exon 5 combined regions most probably represent deletions of the SMN1 gene extending into the rest of the telomeric SMN region. This hypothesis cannot be confirmed as the

P021-A2 MLPA kit cannot distinguish between the centromeric and telomeric copies of these regions.

Furthermore, heterozygous deletions of *SERF1B* were higher in M_1/M_1^B individuals (34.7%) compared to M_1/M_1^W individuals (0%). The frequency of *SERF1B* deletions in M_1/M_1^B and M_1/M_1^W individuals are much lower than reports of 94% in European patients with *SMN1* deletions (Scharf *et al.*, 1998). The small sample size could have confounded results of M_1/M_1^W individuals in this study and a different underlying pathogenic CNV pattern may explain this discrepancy in M_1/M_1^B individuals in this study. This discrepancy could also be due to the use of different technologies to determine copy number (microsatellite analysis in the European study vs MLPA in this study).

4.2.2.2 Gene conversions between *SMN1* and *SMN2*, exons 7 and 8 could explain the lack of correlation of copy number of exons within the same genes

Deletions of *SMN2*, exon 8 were observed more often in M_1/M_1^B individuals (zero copies: 10.7%, one copy: 20%) than M_1/M_1^W individuals (zero copies: 0%, one copy: 16.7%). It is possible that some of these *SMN2*, exon 8 copies may have been converted from *SMN2*, exon 8 to *SMN1*, exon 8, resulting in hybrid genes consisting of *SMN1*, exon 7 and *SMN2*, exon 8. Deletions of these hybrid genes would result in a loss of both *SMN1*, exon 7 and *SMN2*, exon 8, as proposed in section 4.2.2.1. The presence of such hybrid genes could mask larger deletions of the SMN1 gene.

All of these observations suggest that large deletions are more common in M_1/M_1^B individuals than M_1/M_1^W individuals. Gene conversion from *SMN2*, exon 8 to *SMN1*, exon 8, also seen in N/N^B individuals (table 3.4, section 3.2.2.1), could have confounded previous SA reports (Labrum *et al.*, 2007; Stevens *et al.*, 1999) and created the false impression of smaller deletions in M_1/M_1^B individuals when compared to M_1/M_1^W individuals.

In support of this hypothesis, the CNVs of *SMN2*, exons 7 and 8 do not correlate in N/N^B, M_1/M_1^B and N/N^B individuals (as shown in table 3.8, section 3.2.4.3), suggesting that these two regions may not be contiguous due to gene conversion (as discussed

in section 4.2.1.3). Once again due to limited knowledge available on the structure and orientation of genes in the SMN region, it was not possible to determine underlying pathogenic CNV patterns in M_1/M_1^B individuals.

4.2.2.3 In contrast to previous SA reports gene conversion from *SMN1*, exon 7 to *SMN2*, exon 7 appear to be more common in M_1/M_1^W individuals

In contrast, M_1/M_1^W individuals had a higher frequency of multiple copies of *SMN2*, exon 7 (three copies: 33.3%, four copies: 3.3%) when compared to M_1/M_1^B individuals (three copies: 6.7%, four copies: 4%), suggesting that gene conversion from *SMN1*, exon 7 to *SMN2*, exon 7 might be more common in M_1/M_1^W individuals (section 3.2.2.2). See figure 4.3 for an illustration of this hypothesis.

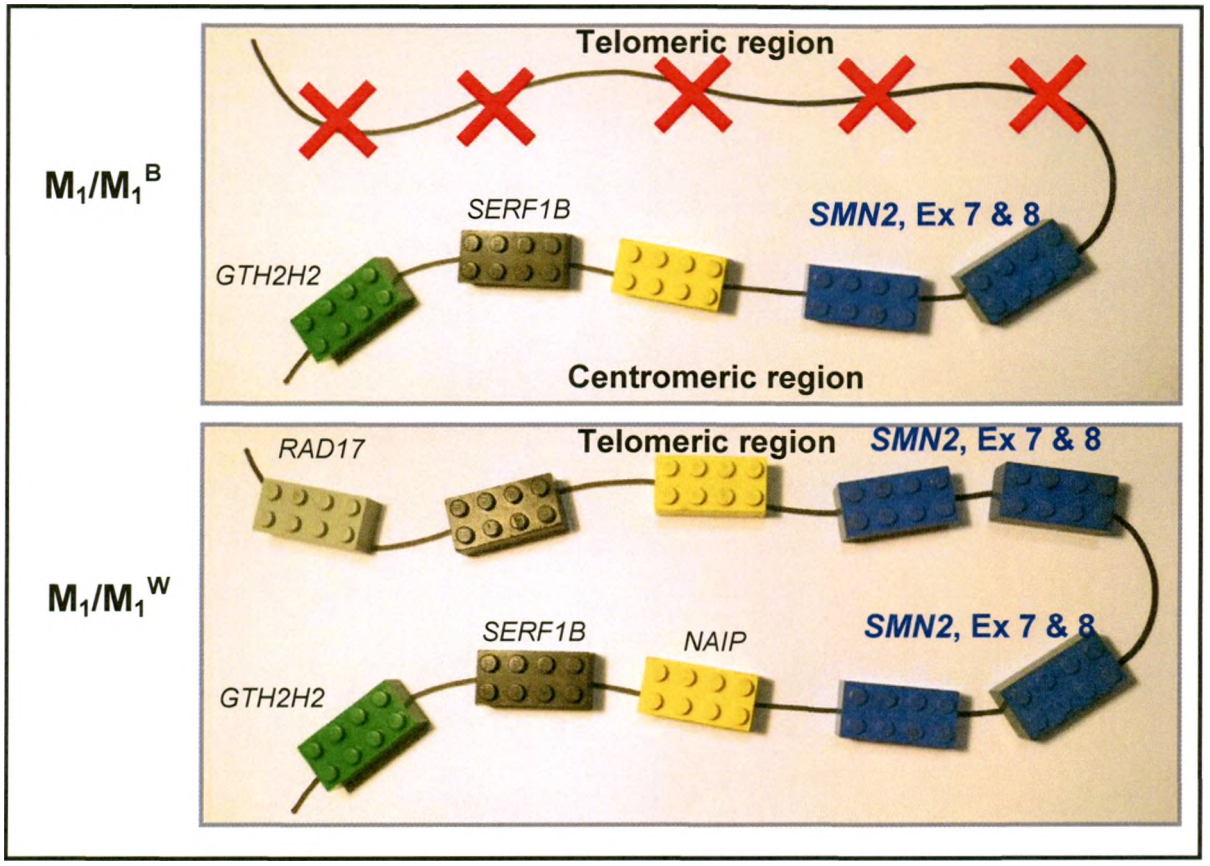


Figure 4.3: Homozygous *SMN1*, exon 7 deletions appear to be caused by different mechanisms in $M_1M_1^B$ and $M_1M_1^W$ individuals. The blue Lego blocks represent the *SMN2* gene and a cross represents a deletion. $M_1M_1^B$ individuals appear to have large deletions of *SMN1*, exon 7 including *SMN1*, exons 1, 4, 6 and 8, *GTF2H2* and *SERF1B* more often than $M_1M_1^W$ individuals. $M_1M_1^W$ individuals appear to have gene conversion events from *SMN1* to *SMN2*, resulting in multiple copies of the *SMN2* region, more often than $M_1M_1^B$ individuals. Both deletions and conversions of SMN1, exon 7 results in the loss of this region.

4.2.2.4 Due to a different genetic composition of the SMN region, M_1/M_1^B and M_1/M_1^W individuals could have different clinical phenotypes

True large deletions appear to be more common in M_1/M_1^B individuals whereas gene conversions seem to be more common in M_1/M_1^W individuals in this study.

Gene conversion from *SMN1*, exon 7 to *SMN2*, exon 7 could result in a loss of *SMN1*, exon 7 and a corresponding increase in *SMN2* copy number and since disease severity is inversely correlated with *SMN2* copy number (McAndrew *et al.*, 1997), these patients are predicted to have a milder form of SMA (Campbell *et al.*, 1997). Patients with large scale homozygous deletions including *SMN1*, *NAIP* and

GTF2H2 have been shown to develop a severe form of SMA within two months and die by the age of two (He *et al.*, 2013) and therefore appear to have a more severe clinical picture.

Unfortunately there was limited clinical information available on the subtypes of SMA for black and white patients referred for SMA testing to the Division. Black patients affected with SMA have not been reported to be more severely affected than white patients but they do appear to have a different phenotype with more frequent involvement of the facial muscles (Moosa and Dawood, 1990), potentially due to the different genetic background observed in this study.

4.2.3 Groups F (M_2/M_2^B) and G (M_2/M_2^W)

This section will discuss the MLPA results from section 3.2.3. MLPA analysis confirmed homozygous deletions of *SMN2*, exon 7 in all M_2/M_2^B and M_2/M_2^W individuals (see section 3.2.3.2).

4.2.3.1 The homozygous *SMN2*, exon 7 deletion forms part of the normal genetic variability of the SMN region in the black SA population

A similar frequency of N/N^B and U/U^B individuals tested positive for a homozygous *SMN2*, exon 7 deletion (12.3% and 13.9% respectively) suggesting that these deletions may form part of the normal background variability in the black SA population and are probably not useful in predicting the disease status in U/U^B individuals. Interestingly, black patients referred for SMA testing to the Division had a higher homozygous *SMN2*, exon 7 deletion frequency, which could potentially be indicative of a pathogenic CNV, also resulting in an apparent homozygous *SMN2*, exon 7 deletion. The possibility of overlapping CNVs complicates the accurate interpretation of homozygous *SMN2*, exon 7 deletions.

4.2.3.2 Gene conversions between *SMN1* and *SMN2*, exons 7 and 8 could explain the lack of correlation of copy number of exons within the same genes and the presence of multiple copies of *SMN1*, exons 7 and 8 and *NAIP*, exon 5

A higher frequency of multiple copies of *SMN1*, exon 7 was observed in M_2/M_2^B individuals (three copies: 22%, four copies: 38%, five copies: 4%) when compared to M_2/M_2^W individuals (three, five copies: 0%, four copies: 37.5%), suggesting that homozygous *SMN2*, exon 7 deletions are likely to be due to gene conversion from *SMN2*, exon 7 to *SMN1*, exon 7, ultimately leading to the loss of *SMN2*, exon 7 in M_1/M_1^B individuals. Similarly, a higher frequency of multiple copies of *SMN1*, exon 8 was observed in M_2/M_2^B individuals (three copies: 28%, four copies: 30%) when compared to M_2/M_2^W individuals (three, five copies: 0%, four copies: 37.5%) and a high frequency of multiple copies of *NAIP*, exon 5 was observed in M_2/M_2^B individuals (three copies: 44%, four copies: 32%, five copies: 14%) when compared to M_2/M_2^W individuals (three, four copies: 12.5%). Again, *SMN1*, exon 7 and 8 and *NAIP*, exon 5 copy numbers do not correlate fully in M_2/M_2^B individuals, suggesting that these multiple copies may not be contiguous, as explained by section 4.2.1.5. Refer to section 3.2.3.1 for a more detailed summary of the results.

4.2.3.3 Large deletions extending into the rest of the SMN region appear to be more common in M_2/M_2^B individuals

M_2/M_2^B individuals appear to have a slighter higher number of chromosomes with homozygous deletions extending into the rest of the *SMN1* and/or *SMN2* genes, exons 1, 4, 6 and 8 and neighbouring genes, *NAIP/NAIPΨ* (18%, 32%, 34%, 32%, 44%, respectively) than M_2/M_2^W individuals (12.5%, 12.5%, 12.5%, 62.5%, 0%, respectively), supporting the hypothesis of large deletions/rearrangements being more common in black patients than white patients (section 4.2.2.1). See figure 4.4 for an illustration of the possible mechanism causing homozygous *SMN2*, exon 7 deletions in black patients.

Since the *SMN2*, exons 7 and 8 regions have more frequent deletions than the *SMN1*, exon 7 and 8 regions, it is hypothesised that the deletions observed in the SMN1/2, *NAIP*, exon 13 and *GTF2H2*, exon 5 combined regions most probably represent deletions of the SMN2 gene into the rest of the centromeric SMN region. This hypothesis cannot be confirmed as the P021-A2 MLPA kit cannot distinguish between the centromeric and telomeric copies of these regions.

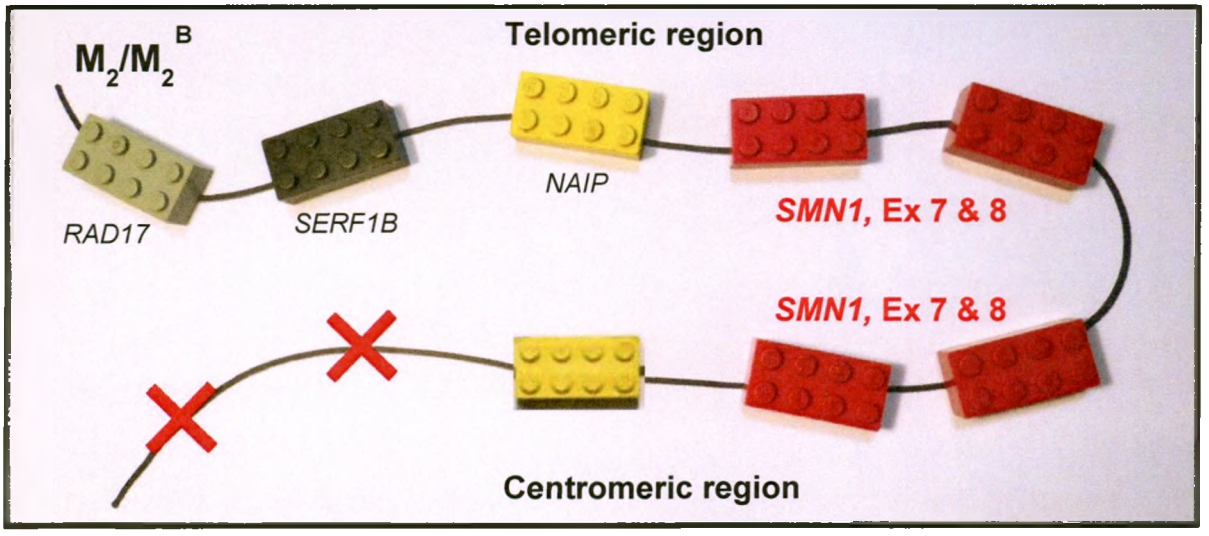


Figure 4.4: A proposed mechanism of homozygous *SMN2*, exon 7 deletions in $M_2M_2^B$ individuals. The red Lego blocks represent *SMN1* and a cross represents a deletion. $M_2M_2^B$ individuals appear to have large deletions of *SMN2*, exon 7, potentially due to gene conversion from the *SMN2* to the *SMN1* region, resulting in multiple copies of the *SMN2* region (*SMN2*, exon 7, 8 and *NAIP*, exon 5). Homozygous deletions of *SMN2*, exon 7 appear to be large in $M_2M_2^B$ individuals including *SMN2*, exons 1, 4, 6 and 8, and neighbouring genes more often than seen in $M_2M_2^W$ individuals.

Even though no clear common CNVs could be identified in M_1/M_1^B and M_1/M_1^W individuals, testing for *SMN2*, exon 7 copy numbers may be useful in individuals who are interested in knowing if they will be suitable for enrolment in clinical therapies which are based on increasing full-length SMN transcripts by utilising multiple copies of *SMN2*, exon 7 (Cure SMA 2016).

4.2.4 Group A (U/U^B)

This section will discuss the MLPA results from section 3.2.4.

4.2.4.1 No novel pathogenic CNVs were identified in U/U^B individuals

It was hypothesised that pathogenic rearrangements of the SMN region could potentially form part of the SMA disease mechanism in black SA patients (Labrum *et al.*, 2007; Vorster, Essop & Krause, 2011). The presence of large complex rearrangements in the black SA population is supported by the high variability and lack of correlation of copy number between different exons and genes seen in black SA individuals (further discussed in sections 4.2.1 and 4.7.1) even though no novel pathogenic CNVs were identified in U/U^B individuals.

U/U^B individuals differed significantly from $M_1/M_1{}^B$ and $M_2/M_2{}^B$ individuals for the majority of MLPA probes. Homozygous deletions of *SMN2*, exon 7 in U/U^B individuals were present at a frequency of 13.9% and have been discussed in section 4.2.3.1.

U/U^B individuals differed significantly from N/N^B individuals for only five probes: *SMN1*, exon 7, *NAIP*, exon 5, *GTF2H2*, exon 5 and combined telomeric/centromeric regions: *NAIP/NAIPΨ*, exon 13 and *GTF2H2*, exon 11, which will be discussed in the next few sections. Results are available from section 3.2.4.

4.2.4.2 A lower frequency of heterozygous *SMN1*, exon 7 deletions was observed in U/U^B individuals than the previous SA study

Black SA patients have always appeared to have a different SMA mutation profile than other populations described in the literature. Previously, U/U^B individuals appeared to have a much higher frequency of the heterozygous *SMN1*, exon 7 deletion when compared to international studies (69.5% vs 2-5%) (Labrum *et al.*, 2007; Wirth, 2000).

The significant difference of *SMN1*, exon 7 results between U/U^B and N/N^B individuals, can be attributed to a higher frequency of heterozygous *SMN1*, exon 7

deletions in U/U^{B} individuals in this study. In contrast to previous South-African reports of heterozygous *SMN1*, exon 7 deletions being present in 69.5% of U/U^{B} individuals (Labrum *et al.*, 2007), only 8.3% of U/U^{B} individuals were confirmed to have heterozygous deletion of *SMN1*, exon 7 in this study.

Only one mutation, a heterozygous deletion of *SMN1*, exon 7, was identified in these individuals and could be supportive of a diagnosis of SMA. The second disease-causing mutation, if present, remains unidentified in these U/U^{B} individuals.

The discrepancy of the heterozygous *SMN1*, exon 7 rate between the previous SA study and this project could firstly be due to MLPA being a more reliable technique than the previously used in-house dosage system (as explained in section 4.2.1.2) and secondly the presence of multiple copies of *SMN1*, exon 7 could mask the actual heterozygous *SMN1*, exon 7 deletion rate in U/U^{B} individuals (further discussed in section 4.2.4.3). Finally, it is also possible that some of these patients do not have SMA but a related neuromuscular disease. The results of non-SMA patients could confound the true heterozygous *SMN1*, exon 7 deletion rate.

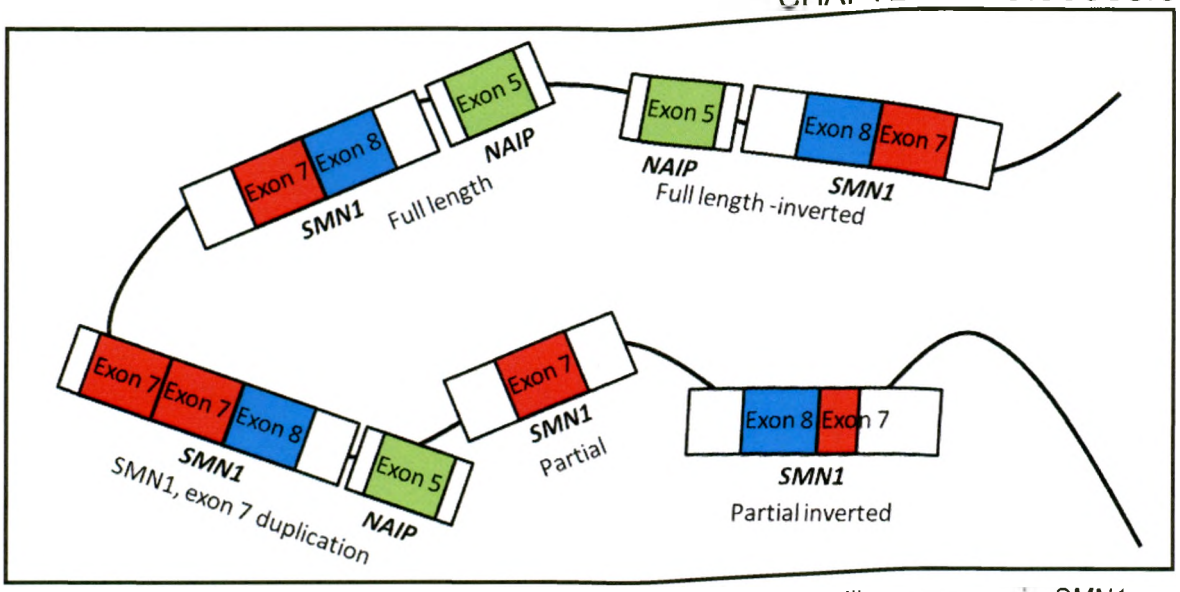
4.2.4.3 Multiple copies of the telomeric region *SMN1*, exons 7 could mask pathogenic CNVs in U/U^{B} individuals

Multiple copies of *SMN1*, exon 7 were observed in 38.9% of U/U^{B} individuals, similar to the rate observed in N/N^{B} individuals (50.8%). It is possible that these multiple copies may be non-contiguous, partial copies, which may not be functional and could mask pathogenic CNVs of the SMN gene. This hypothesis is supported by the observation that the copy number of *SMN1*, exons 7 and 8 and *NAIP*, exon 5 do not correlate fully with each other or with the *SMN1*, exons 1, 4, 6 and 8 copy numbers as illustrated in section 3.2.2.3.

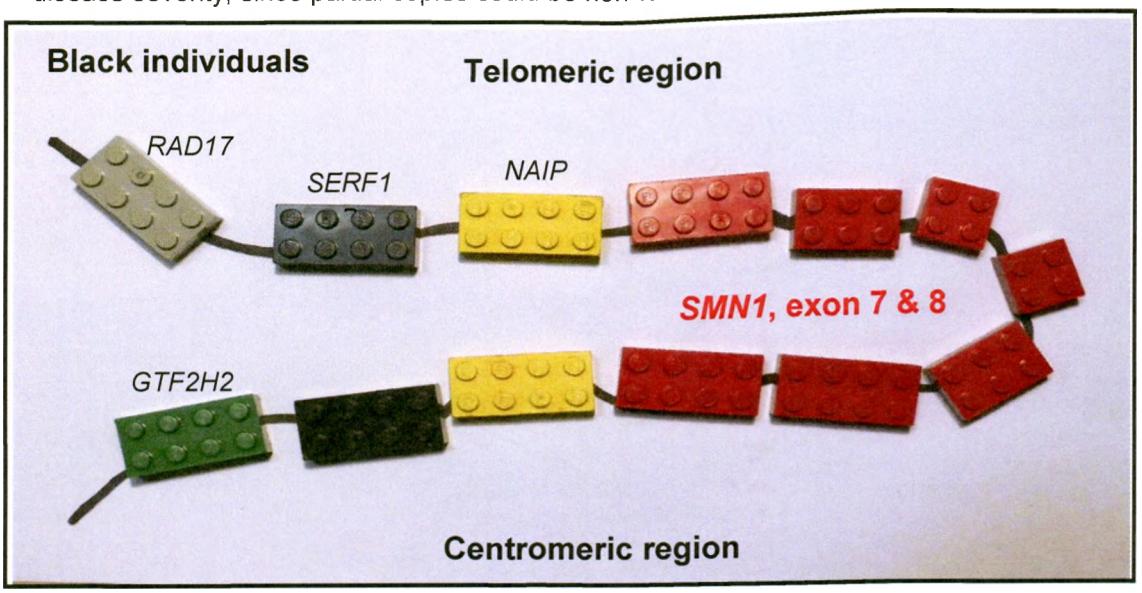
MLPA testing cannot give us information about the functionality of potential multiple, partial copies, suggesting that functional studies (RNA expression and protein studies) may be more informative about the levels of full-length protein present in affected individuals and will be further discussed in section 5.5.1.

Furthermore, the exact orientation of genes in the SMN region and therefore the orientation of MLPA probe regions are not well understood and could confound analysis. Results of this study did not provide additional information on the structure of the SMN region.

See figure 4.5 for an illustration of the possible mechanism of multiple *SMN1*, exon 7 copies in black SA individuals.



a.) Multiple partial copies of the SMN region are present. This will appear as six *SMN1* gene copies on MLPA. The copy number cannot be interpreted as having an effect on disease severity, since partial copies could be non-functional.



b.) Multiple partial copies of the *SMN1* region, indicated by red Lego blocks are present.

Figure 4.5: Multiple partial non-coding gene copies could result in an apparent increased copy number

4.2.4.4 The copy number of centromeric-*GTF2H2* and telomeric-*GTF2H2* are not associated with SMA in either the white or black SA populations

Probes specific to the c.706C>G (telomeric-*GTF2H2)* and c.453A>G (centromeric-*GTF2H2*) variants were included in the P021-A2 MLPA kit and are used as proxies to distinguish between the telomeric and centromeric copies of the *GTF2H2* gene as described in a single publication (Carter *et al.*, 1997). Limited research has been done

on the association of these genes to SMA, questioning the appropriateness of their inclusion in this MLPA kit.

Even though significant differences were observed for the *GTF2H2*, exon 5 and *GTF2H2*, exon 11 regions between U/U^B and N/N^B individuals, no major conclusions could be drawn. This difference alone does not suggest any clear pathogenic rearrangement in U/U^B individuals.

The centromeric-*GTF2H2* copy number does not differ among different subject and population groups, questioning whether this variant has any association with SMA (see table 3.9 in section 3.2.4.4).

Deletions of telomeric-*GTF2H2* were associated with M_1/M_1^B homozygous *SMN1*, exon 7 (telomeric) deletions in M_1/M_1^B (80%) and M_1/M_1^W individuals (70%) as expected as deletions of at least one copy of telomeric-*GTF2H2* have been observed in patients with SMA (Carter *et al.*, 1997).

It is surprising to note that deletions of telomeric-*GTF2H2* were observed in M_1/M_1^W individuals as well as N/N^W individuals, questioning the association of this deletion with *SMN1*, exon 7 deletions in white populations.

The copy number of telomeric-*GTF2H2* does not differ among N/N^B, M_2/M_2^B or U/U^B individuals, suggesting that this probe is not useful in providing information on pathogenic rearrangements in the black SA population. Testing for the presence of telomeric-*GTF2H2* might not have any clinical utility but to provide information on *SMN1*, exon 7 deletions extending into neighbouring genes in the black SA population.

In summary, the c.453A>G and c.706C>G variants are not useful in predicting SMA disease status in the black or white SA population and do not seem to add any value to the P021-A2 MLPA kit. The inclusion of probes specific to these two gene variants does not seem appropriate or useful, even for Caucasian populations.

As a side note, MLPA is not an ideal technique to test for the presence of single nucleotide differences. An absence of fluorescent probe signal could either be caused by actual deletions of this region or a change from a G to a C allele, essentially converting telomeric-*GTF2H2* to centromeric-*GTF2H2* and leading to the loss of

telomeric-*GTF2H2* or potentially large rearrangements of the SMN region could influence results.

There were no families available of U/U^B individuals. Determining the phase of CNVs in these families could potentially assist with the identification of novel pathogenic CNV patterns.

4.3 Haplotype and CNV pattern analysis

4.3.1 Multiple *SMN1*, exon 7 copies complicate haplotype construction in the black SA population

As part of a previous study performed in the Division, linked marker analysis, using two intragenic and six extragenic microsatellite markers across the *SMN1* gene, was performed to see if a common chromosomal background could be established in U/U^B individuals. No clear haplotype or common allele was identified and it was reported that it was particularly difficult to construct haplotypes (Labrum *et al.*, 2007).

Similarly, in this study, multiple gene copies in the black SA population complicated haplotype construction. Only 44% of M_1/M_1^B families and 31.7% of N/N^B families were completely informative where the phase of the haplotype could be determined with absolute certainty (see section 2.3.4 for an explanation of family pedigree construction and haplotype analysis).

The orientation of genes in the SMN region is not known (explained by figure 1.1 in section 1.2.1) and this study could not predict the orientation of MLPA probe regions even though the phase could be determined. Furthermore, each probe region was analysed as if it was a single locus, in order to simplify analysis but this may not be completely accurate since CNVs include multiple interacting regions. Without an accurate map of the SMN region, the true copy number and breakpoints of CNVs cannot be determined.

Results for N/N^B families are summarised in section 3.4.1 and results for M_1/M_1^B families are summarised in section 3.4.2.

4.3.2 A high variability of the SMN region in the black SA population could lead to novel changes

Seventeen unique pathogenic haplotypes were identified in probands of M_1/M_1^B families and 35 additional unique haplotypes were identified in N/N^B individuals emphasising the high variability of the SMN region in the black SA population.

Discrepant results (apparent non-Mendelian inheritance) were observed in 8% of M_1/M_1^B families and 5% of N/N^B families, potentially due to a variety of causes such as the parents in these families not being the actual biological parents of the children; proposed novel deletion or duplication events in the proband or the unknown orientation of genes and multiple copies within this family confounding results.

Potential novel events (sporadic deletions or duplications) were observed in 3.3% (2/61) of N/N^B families. A new mutation rate of 3.3% is not unexpected as novel mutations have been reported at a rate of 2% due to the high instability of this region (Wirth *et al.*, 1997). Moosa and Dawood reported a paucity of family history in black SA families affected by SMA (discussed in section 1.1) potentially due to SMA being more sporadic in this population (Moosa and Dawood, 1990). Sporadic mutations could be a consequence of novel gene conversion and rearrangement events. Table 4.1 summarises these results.

Table 4.1: A comparison of haplotype results of M_1/M_1^B and N/N^B families

Black families	Informative results	Uninformative results		
		Multiple combinations possible	Discrepant results	Clear novel events
M_1/M_1^B families (n=25)	44% (11/25)	48% (12/25)	8% (2/25)	-
N/N^B families (n=60)	31.7% (19/60)	60% (36/60)	5% (3/60)	3.3% (2/60)

4.3.3 Multiple *NAIP*, exon 5 copies observed in N/N^B families

The significant difference of *NAIP*, exon 5 and *NAIP/NAIPΨ*, exon 13 results between U/U^B and N/N^B individuals, can be attributed to a higher frequency of multiple copies (>2) of *NAIP*, exon 5 in N/N^B individuals when compared to U/U^B individuals (see section 3.2.4.2 for results). Perhaps multiple copies of *NAIP* have a protective effect in the black SA population. Caution is advised with this hypothesis since the role of *NAIP* in the SMA disease mechanism needs to be further investigated and should not be viewed in isolation. Figure 4.6 illustrates this association.

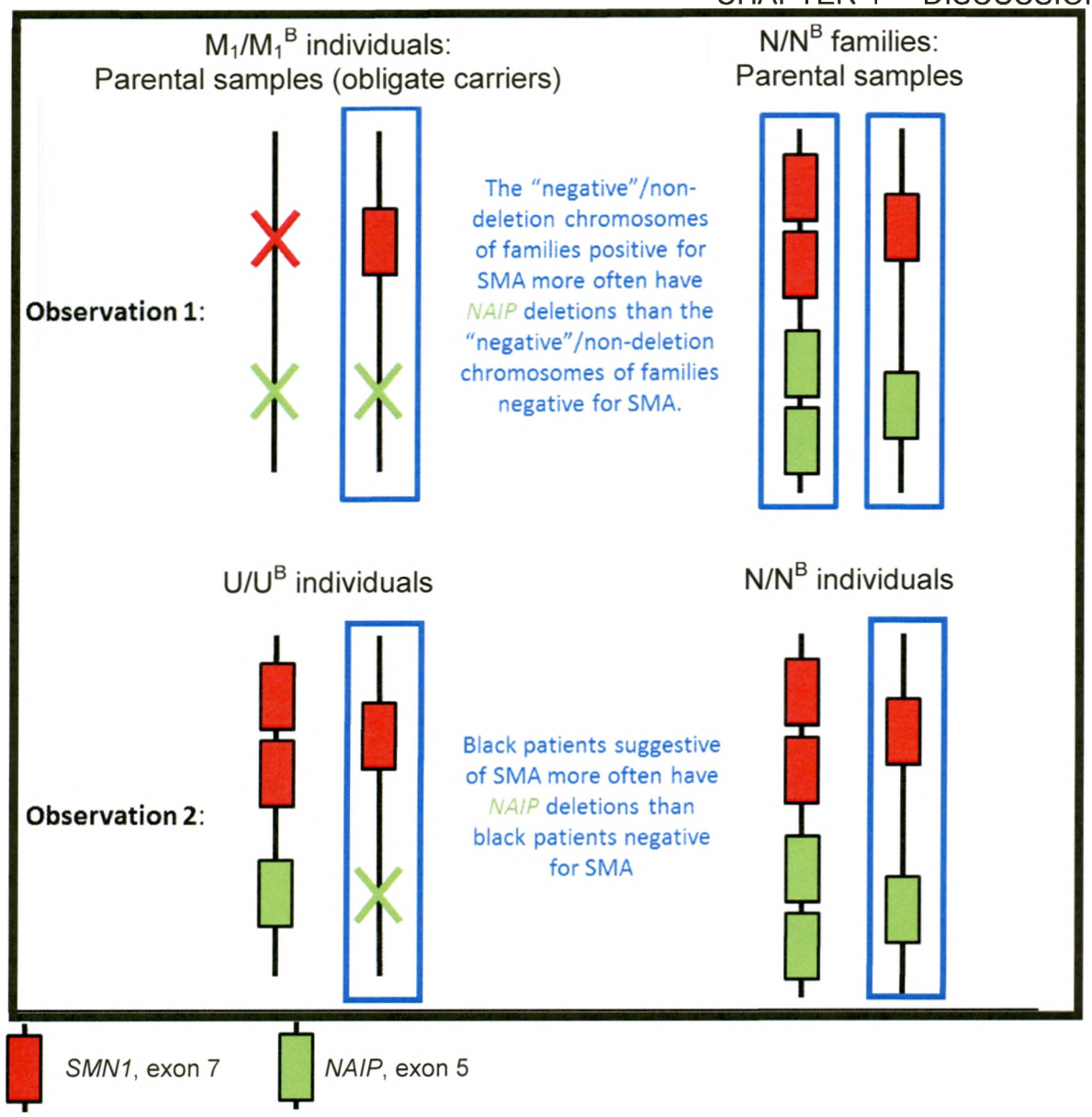


Figure 4.6: The association of *NAIP*, exon 5 copy number with disease status

N/N^B and U/U^B individuals tend to have multiple copies of *SMN1*, exon 7 (represented by red blocks). N/N^B individuals tend to have multiple copies of *NAIP* 7 (represented by green blocks). Deletions are indicated by crosses and appear to be common in M_1/M_1^B and U/U^B individuals.

4.3.4 Silent carriers of SMA may be masked by the presence of multiple *SMN1*, exon 7 copies in N/N^B individuals

This hypothesis has been introduced in section 4.2.1.2. Five unrelated individuals from M_1/M_1^B families had a 2:0 silent carrier genotype (defined in section 1.2.2.4) and two individuals had uninformative results for *SMN1*, exon 7, where the individual could either have a 1:1 or 2:0 genotype. No individual from N/N^B families had an

unequivocal silent carrier genotype, although four individuals from N/N^{B} families had uninformative results for *SMN1*, exon 7, where the individual could either have a 1:1 or 2:0 genotype.

A potential silent carrier rate of 3.3% (4/122) seen in N/N^{B} individuals compares well with the rate of 4% reported in the French-Canadian population (McAndrew *et al.*, 1997) although the true silent carrier rate is likely to be underestimated in the black SA population due to the multiple copies of *SMN1*, exon 7, complicating the determination of phase.

Two variants, c.885+83T>G and c.885+667delAT in exon 8 of the *SMN1* gene have been described to be associated with multiple *SMN1* copies on a single chromosome in combination with a *SMN1*, exon 7 deletion on the other chromosome in African American and Ashkenazi Jewish population groups. It was suggested that these two variants could be useful in refining the silent carrier risk in individuals who have multiple copies of *SMN1*, exon 7 (Luo *et al.*, 2014). An honours project was performed in 2016, supervised by the researcher, to investigate the association of the c.885+83T>G and c.885+667delAT variants to silent carrier haplotypes in the black SA population. Black silent carrier individuals, who had two copies of *SMN1*, exon 7 on a single chromosome (2:0) were identified through family haplotype analysis. The c.885+83T>G and c.885+667delAT variants were observed in 60% (3/5) of known silent carrier individuals (2:0) compared to 0% (0/7) of individuals known to have two copies of *SMN1*, exon 7 copy, one copy on each of their two chromosomes (1:1), suggesting that the two variants are associated with duplicated *SMN1*, exon 7 alleles.

Uninformative M_1/M_1^{B} families were investigated to identify potential silent carriers, who had two copies of *SMN1*, exon 7 but where the phase of these *SMN1*, exon 7 copies could not be accurately determined (2:0 or 1:1). The c.885+83T>G and c.885+667delAT variants were identified in 37.5% (3/8) of these individuals who therefore have an increased risk of being silent carriers. These two variants could be useful in refining the carrier risk of black SA individuals who have two copies of *SMN1*, exon 7 (Khan *et al.*, 2016).

Table 4.2 summarises the results of MLPA analysis across all seven patient and subject groups (Group A-G).

Table 4.2: Summary of main MLPA findings

Patient/Subject category	Main findings
N/N^B n=122	- High variability when compared to N/N^W - No carriers identified (0% vs previous rate of 1/50) - Gene conversion between *SMN2*, exon 8 and *SMN1*, exon 8 - **Multiple copies of *SMN1*, exon 7 (observed in 50.8%):** • Similar rates in African-American & sub-Saharan populations • Could mask presence of silent carriers (2:0. 3:0. 4:0, 5:0 etc.) • No common non-pathogenic haplotypes (60% of families were uninformative; 35 unique haplotypes)
N/N^W n=30	- Majority have two copies of *SMN1*, exon 7 - Similar rates in European & North-American populations
M_1/M_1^B n=75	- Larger deletions extending into the rest of the SMN1 gene and neighbouring genes - No common pathogenic haplotypes (48% of families were uninformative; 17 unique pathogenic haplotypes)
M_1/M_1^W n=30	- Gene conversion between *SMN1*, exon 7 and *SMN2*, exon 7 result in smaller deletions
M_2/M_2^B n=50	- Homozygous deletions of *SMN2*, exon 7 forms part of general variation in black SA population (audit: 12.4%, N/N^B: 27%, U/U^B: 13.9%) - Gene conversion from *SMN2* to *SMN1* result in deletions of *SMN2*, exon 7 & multiple copies of *SMN1*, exon 7 - Larger deletions extending into the rest of the SMN1 gene and neighbouring genes
M_2/M_2^W n=8	- Homozygous deletions of *SMN1*, exon 7 not as common as seen in individuals of black or mixed ancestry - Gene conversion from *SMN2* to *SMN1* result in deletions of *SMN2*, exon 7 & multiple copies of *SMN1*, exon 7
U/U^B n=72	- No novel pathogenic CNVs identified - Multiple *NAIP*, exon 5 copies observed - could have protective effect - Rate of heterozygous deletions of *SMN1*, exon 7 much lower than previously reported (8.3% vs 69.5%) - **Multiple *SMN1*, exon 7 copies (observed in 38.9%):** • Could mask pathogenic CNVs and true heterozygous rate • May be partial, disconnected, non-functional copies

4.4 RNA analysis

This section will discuss the results from section 3.5. Good quality RNA was successfully extracted from blood collected in Tempus Blood RNA tubes stored at a variety of conditions. The Tempus Blood RNA tubes show promise for RNA-based tests in a diagnostic setting, especially since many samples are sent from locations far away from the test laboratory and may take multiple days to arrive which may influence the quality of the RNA.

4.5 Provisional real-time PCR results

This section will discuss the results from section 3.6. A qPCR assay showed a corresponding increase in *SMN1* fluorescent signal to *SMN1*, exon 7 copy number, although there seems to be some overlap in results of different copy numbers and therefore this assay needs further optimisation.

The current diagnostic homozygous *SMN1*, exon 7 deletion assay (described in section 1.2.2.1) is much cheaper than MLPA and qPCR (table 3.10 in section 3.6), indicating that it is still a cost effective first-line testing strategy. The advantage of both qPCR and MLPA over the diagnostic assay is the determination of copy number of *SMN1* and *SMN2*, exon 7 which could be used to determine carrier status and identify patients with multiple copies of *SMN2*, exon 7 for possible clinical trials. The only advantage of MLPA over qPCR is the inclusion of multiple regions across the rest of the SMN1 and SMN2 genes and neighbouring genes and may therefore be useful to give some indication of the extent of deletions/duplications but it is not useful in identifying additional pathogenic CNV patterns and it is very complicated to analyse.

The qPCR assay, once fully optimised, has the potential to detect *SMN1*, exon 7 copy numbers, making it possible to detect heterozygous *SMN1*, exon 7 deletions in carriers of SMA and the 8.3% of U/U^{B} individuals. This would be a clear advantage over the current diagnostic assay, only able to detect the presence or absence of homozygous *SMN1* and *SMN2*, exon 7 deletions and could potentially be implemented as an alternative to the current diagnostic assay and MLPA in the future.

4.6 Implications for diagnostic testing

4.6.1 The clinical utility of MLPA as a diagnostic test

Heterozygous *SMN1*, exon 7 deletions were identified in only 8.3% of U/U^{B} individuals. Multiple copies of *SMN1*, exon 7 could mask heterozygous *SMN1*, exon 7 deletions which may lead to the underestimation of the true heterozygous *SMN1*, exon 7 deletion rate in U/U^{B} individuals.

Furthermore, no additional pathogenic CNVs were identified in U/U^{B} individuals using MLPA (section 4.2.4.1). Once again, multiple copies of *SMN1*, exon 7 may potentially be masking unidentified pathogenic rearrangements of the SMN region. The CNV profiles of U/U^{B} individuals who do not in fact have SMA may also mask true pathogenic CNV patterns.

Previously, heterozygous *SMN1*, exon 7 deletion testing was offered to U/U^{B} individuals, who previously tested negative on the diagnostic assay but the high cost, low pickup rate and the presence of multiple copies confounding results, suggest that MLPA is not an appropriate standard diagnostic procedure in the black SA population.

4.6.2 The clinical utility of MLPA as a test to determine SMA carrier status

The silent carrier genotype (2:0) has been reported to be much more common in African-American individuals (27%) when compared to white individuals (3.6%). As a result, the carrier detection rate in African-American individuals is lower at 70% versus 91% in other population groups (Sugarman *et al.*, 2012).

No N/N^{B} carriers of the heterozygous *SMN1*, exon 7 deletion were identified in this study. Multiple copies of *SMN1*, exon 7 and unidentified CNVs could potentially mask the presence of silent (2:0; 3:0, 4:0, 5:0) carriers and confound the true carrier rate in the individuals of African ancestry.

With the advent of new sequencing technologies, pan-ethnic population-based expanded carrier screening has been gaining momentum internationally. As an example, Israel has implemented a genetic screening program including carrier testing for SMA for couples of reproductive age (Zlotogora et al., 2016).

The American College of Medical Genetics (ACMG) supports the inclusion of SMA into expanded carrier screening tests (Prior and Professional Practice and Guidelines Committee, 2008). The American College of Obstetricians and Gynecologists (ACOG) does not recommend inclusion of SMA into expanded carrier screening tests but does recommend carrier testing for SMA in family members of patients affected with SMA or SMA-like diseases. Furthermore, ACOG comments on a lack of pilot programs and standard laboratory assays to detect SMA (ACOG, 2009).

Caution is advised against population screening in the black SA population, due to the presence of multiple copies of *SMN1*, exon 7 which could confound the actual carrier risk. MLPA may be useful in detecting the carrier risk in members of M_1/M_1^B families but it is highly recommended that MLPA results of samples referred for prenatal and carrier testing should be analysed within a family context in order to identify the phase of multiple *SMN1* genes in all SA populations.

In conclusion, the carrier frequency of SMA cannot be determined with accuracy in the black SA population until the disease mechanism of SMA has been completely uncovered and well understood.

4.6.3 Recommendations for future diagnostic testing

When considering the cost of the homozygous *SMN1*, exon 7 deletion assay, qPCR and MLPA, the homozygous *SMN1*, exon 7 deletion assay is still a reliable first line test for patients presenting with clinical features suggestive of SMA. Once optimised, the qPCR assay could potentially replace this diagnostic assay and MLPA as a quantitative first-line assay.

Patients affected with SMA have previously been shown to have significantly lower FL-SMN1 transcript levels in blood derived samples when compared to negative controls, even when patients were separated by clinical type (Crawford et al., 2012; Tiziano et al., 2010; Zheleznyakova et al., 2011). Patients affected with SMA, types I-III were shown to have a mean FL-SMN transcript value of 38.3±16.3 – 58.4±19.7mol/ng compared to negative controls who were shown to have a mean FL-SMN transcript value of 134.1±65.1mol/ng. SMN protein levels have been shown to be decreased in patients affected with SMA but has not been shown to differ

significantly between different clinical types (with a mean value of 10.1±4.8 – 14.7±10pg/10^6 cells for the different types of SMA when compared to a mean value of 25.2±19.6pg/10^6 cells for negative controls) (Crawford et al., 2012). Furthermore FL-SMN2 has been shown to increase with SMA type but not to correlate with *SMN2* copy number (Crawford et al., 2012; Tiziano et al., 2010).

A future MSc (Med) project (2017) will be focussed on investigating the practical application and clinical utility of RNA expression studies using qRT-PCR as a potential second-line diagnostic test (as described in section 2.3.5.3). qRT-PCR may be able to truly quantitate the expression of full-length SMN transcripts which may be a more accurate indication of the amount of SMN protein produced in U/U^B individuals even in the presence of multiple *SMN1*, exon 7 copies. Figure 4.7 shows a proposed diagnostic approach going forward.

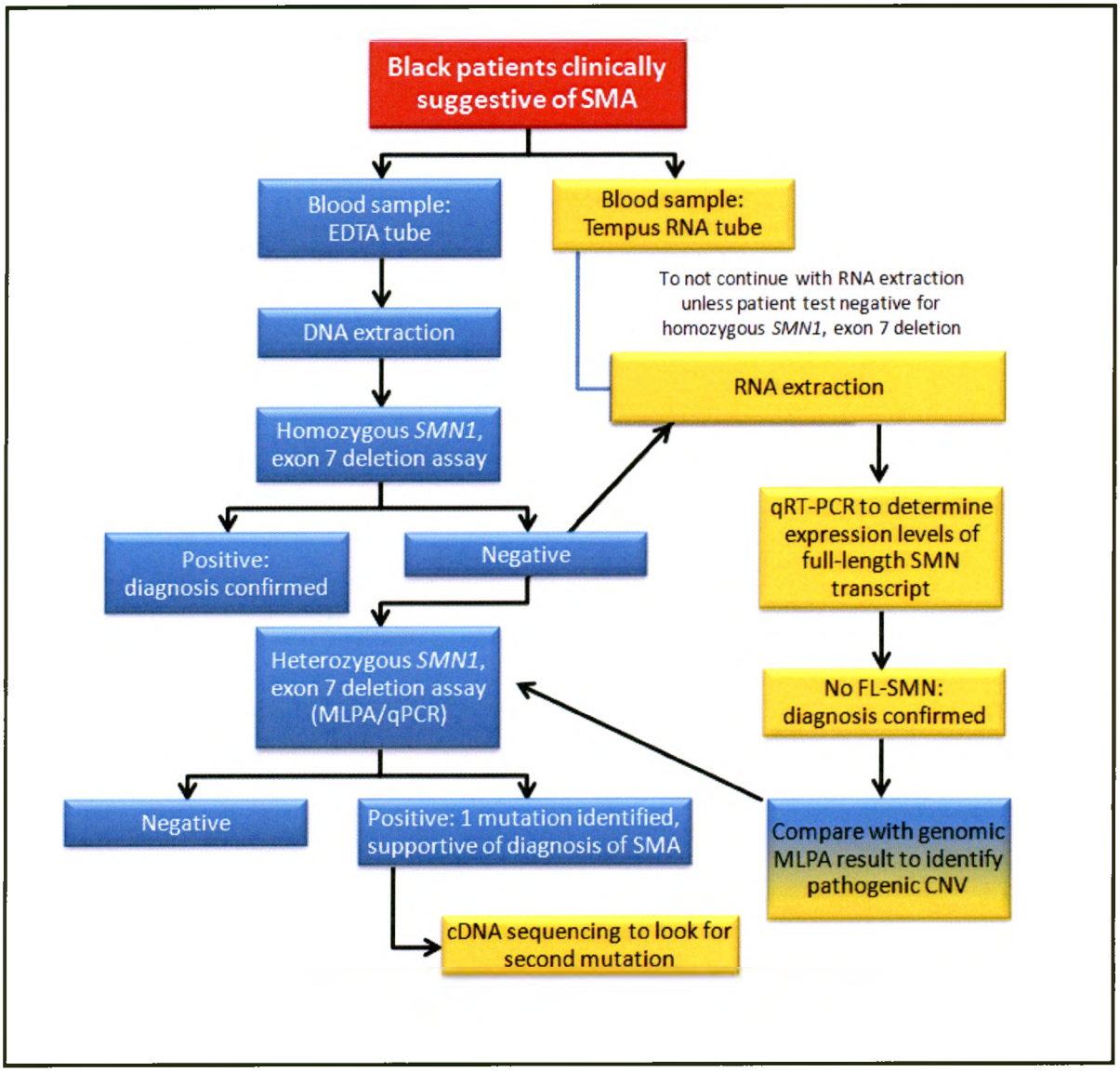


Figure 4.7: Proposed diagnostic strategy for future diagnostic testing for SMA in black patients in the Division
Blue blocks represent DNA-based assays and analysis while yellow blocks represent RNA-based assays and analysis

4.7 Challenges and limitations of the study

4.7.1 Hypervariable SMN region

The SMN region is extremely complex containing multiple pseudogenes (Selig *et al.*, 1995) and repetitive sequences (Bürglen *et al.*, 1996) within a large inverted duplication. Due to this complexity, there is limited understanding of the exact order and location of genes in the SMN region. This is complicated even further since the

location and order of multiple copies of *SMN1*, exon 7 in the black SA population is not understood.

CNV trends observed in the various subject groups tested on MLPA (Groups A-G), suggest that large rearrangements form part of the general variation within the black SA population. SMA was suggested to be a heterogeneous disorder as early as 1971 when it was shown that different types of SMA were observed within a single family (Ghetti *et al.*, 1971). It is possible that the structure of the *SMN* region could differ among different population groups, among different patients affected with SMA and even within the same family, potentially influencing the risk to disease and the severity of the disease.

It is well established that African populations have a higher level of genetic diversity than any other population, due to a combination of genetic drift, multiple migration events, drastic changes in population size and local selection forces (Conrad *et al.*, 2006; Pickrell *et al.*, 2014; Tishkoff and Kidd, 2004; Tishkoff and Williams, 2002). A local group of researchers who investigated CNV in SA populations found that haplotype block lengths are significantly smaller in African populations when compared to non-African populations. These regions seem to coincide with recombination hotspots (Chimusa *et al.*, 2015). More frequent recombination events will lead to a decrease in linkage disequilibrium and consequently smaller haplotype blocks will form. Recombination hotspots have also been shown to be enriched with SNPs (Choudhury *et al.*, 2014).

Perhaps the high variability of the SMN region in the black SA population could be due to frequent recombination events in the SMN region. This hypothesis is supported by a new mutation rate of 3.3% in this study. Frequent recombination events causing smaller haplotype blocks could also explain the lack of linkage disequilibrium of the c.453A>G (centromeric-*GTF2H2)* and c.706C>G (telomeric-*GTF2H2)* variants in the black SA population.

The African Human Genome Variation project (AHGVP) has reported that existing gene arrays and SNP panels will miss common haplotypes within African populations, due to ascertainment bias (a focus on variations not common in African populations)

and that we need to take these differences into account when designing future African projects (Gurdasani *et al.*, 2015).

We need to identify and comprehend natural CNVs in the general SA population in order to fully understand disease mechanisms overlaying these variations, specifically in the SMN region. The Southern African Human Genome Project (SAHGP) shows some promise in creating a better understanding of the baseline CNVs in the general black SA population (Pepper, 2011) although it is unlikely to provide detailed information on the architecture of the SMN region.

4.7.2 High SMN sequence homology

The high sequence homology of the *SMN1* and *SMN2* genes, with only a 5 nucleotide difference between the two genes and the highly variable CNVs of these genes make molecular diagnosis extremely challenging. This limitation complicates and restricts the design of primers and probes in this region and limits the choice of laboratory techniques. In example, primers and probes have to be designed based on the single base pair difference at cDNA position 840 in order to distinguish between the *SMN1* and *SMN2,* exon 7. This same challenge applies to other exons and genes in the *SMN* region and their pseudogenes.

Furthermore, frequent gene conversion events, multiple copies of *SMN1*, exon 7 and potential unidentified CNVs of the SMN region in the black SA population could interfere with the binding of primers/probes used in quantitative techniques, resulting in false negative and/or false positive results.

Since the cDNA sequences of *SMN1* and *SMN2* are almost identical; it will not be possible to determine whether any mutations identified are based in the *SMN1* or the *SMN2* gene, using conventional Sanger sequencing in future research projects.

4.7.3 Limitations of the methodology

The P021-A2 MLPA kit was mainly designed to distinguish between exons 7 and 8 of the *SMN1* and *SMN2* gene, but cannot distinguish between exons 1, 4, and 6 of the *SMN1* and *SMN2* genes. This means that a combined result is observed for these

probes. This makes it difficult to assign multiple copies of specific exons to either *SMN1* or *SMN2.* Similarly, the *NAIP/NAIPΨ,* exon 13 probe was designed to detect the combined copy number of the *NAIP* gene and its centromeric pseudogene, *NAIPΨ* and the *GTF2H2*, exon 5 was designed to detect the combined copy number of telomeric-*GTF2H2* and centromeric-*GTF2H2*.

In addition the copy numbers of *SMN1/2* exons 1, 4, 6 and 8 do not correlate fully with each other or with the *SMN1,* exons 7 an 8 copy numbers suggesting that CNVs of the SMN1 gene do not consist of complete gene copies and that the exons may be non-contiguous. This is further complicated by gene conversion events between *SMN1* and *SMN2*. Result interpretation is therefore incredibly difficult and it is not possible to construct accurate CNV patterns using MLPA.

Any single nucleotide change or rearrangement overlapping with the MLPA or qPCR probe-binding region could result in a lack of amplification of the specific probe, distorting the results and appearing as deletions.

The inclusion of MLPA probes specific for the neighbouring genes *GTF2H2, SERF1B* and *RAD17* could only suggest the extent of CNVs but did not assist with the determination of novel pathogenic CNVs or the order of genes or multiple copies of these genes. Similarly, the c.453A>G and c.706C>G variants were not useful in predicting SMA disease or carrier status in the black or white SA population and the inclusion of these two probes in the P021-A2 kit needs to be questioned.

MRC Holland is considering the removal of probes for *RAD17, SERF1B, GTF2H2* and *NAIP* and the inclusion of additional probes for *SMN1/2* probes, exons 2a, 2b, 3 and 5 in their development of a new version of the P021-A2 SMA MLPA kit and we were asked for our recommendations (personal communication with Dr Mirjam Ketema).

On the basis of this study, the researchers recommended that MRC Holland should include at least one probe specific to the *NAIP* and *GTF2H2* genes in order to provide information on the extent of large deletions for diagnostic purposes but agreed with the removal of other genes located in the SMN region. The researchers advised that analysis of the combined *SMN1/2* exons proved to be very difficult and it would be beneficial if MRC Holland could add guidelines in the SMA product description document to assist the user to analyse copy numbers of these probes.

Furthermore, the researchers advised them to include probes specific to the c.885+83T>G and c.885+667delAT variants in *SMN1* which have been described to be associated with silent carrier genotypes in African American and Ashkenazi Jewish population groups (Luo *et al.*, 2014) and shows some promise in the black SA population. MRC Holland has released the provisional P460 MLPA kit which can detect these two variants in addition to CNVs of *SMN1* and *SMN2*, exon 7.

The researchers will recommend that MRC Holland should add a cautionary statement to the product sheets for these MLPA kits when determining carrier status in individuals of African ancestry. Furthermore, a cautionary statement should be included on the complex architecture of the SMN region and that the exact orientation of genes and multiple copies observed in individuals are not known.

4.7.4 Lack of clinical information

The majority of referrals to the Division for diagnostic SMA testing lack proper clinical information on the request forms and consequently no clinical subtypes could be assigned accurately. The majority of international studies performed to correlate disease severity with the copy number of the *SMN2, NAIP, GTF2H2* and *SERF1A*, rely on the clinical subtype of the patients.

It is also possible that some of the patients referred for SMA testing, do not in fact present with SMA but with similar neurological diseases, although this is unlikely since U/U^{B} individuals were all identified in collaboration with a paediatric neurologist centre who has a special interest in SMA.

5 CONCLUSION

5.1 Conclusions from MLPA analysis

This is the first report summarising CNV patterns of the SMN region in African individuals with known homozygous *SMN1* and *SMN2*, exon 7 deletions (M_1/M_1^B, M_2/M_2^B) and patients who are clinically suggestive of SMA (U/U^B). This is also the first report of CNVs patterns of the SMN region in the general black SA population.

5.1.1 N/N^B individuals

A very high variability of the SMN region was observed in the black SA population when compared to the white SA population. More specifically, multiple copies of *SMN1*, exon 7 were observed in 50.8% of N/N^B individuals and could complicate analysis by masking silent carriers and obstructing haplotype construction. No common non-pathogenic haplotypes were identified in the black SA population. N/N^B individuals have a significantly higher frequency of multiple copies of *NAIP*, exon 5 when compared to U/U^B individuals suggesting a potential protective effect. *NAIP* could have a complex disease modifying effect but the role of *NAIP* in the SMA disease mechanism is poorly understood and needs further investigation.

5.1.2 M_1/M_1^B individuals

Homozygous deletions of *SMN1*, exon 7 seem to more often form part of large deletions including the rest of the SMN1 gene and neighbouring genes in M_1/M_1^B individuals, whereas homozygous deletions of *SMN1*, exon 7 seem to be caused more often by gene conversion from *SMN1*, exon 7 to *SMN2*, exon 7 in M_1/M_1^W individuals. A different genetic profile in M_1/M_1^B and U/U^B individuals have been established by this study and could support previous reports of black SA patients with SMA having a different clinical phenotype from worldwide reports (Moosa and Dawood, 1990). No common pathogenic haplotypes were identified in the black SA population.

5.1.3 M_2/M_2^B individuals

Homozygous *SMN2*, exon 7 deletions were present at a high frequency in the black SA population, suggesting that this CNV forms part of the general variation.

5.1.4 U/U^B individuals

MLPA identified heterozygous deletions of *SMN1*, exon 7 in 8.3% of U/U^B individuals, which contrasts previous SA reports of 69.5%. The previous rate may have been overestimated due to inaccuracies of the previously used dosage assay. MLPA analysis did not identify additional pathogenic rearrangements in U/U^B individuals.

Multiple copies of *SMN1*, exon 7 were observed in 38.9% of U/U^B individuals and could complicate analysis by masking the true rate of heterozygous *SMN1*, exon 7 deletions and possibly the presence of other pathogenic CNVs in U/U^B individuals. It is hypothesised that multiple *SMN1*, exon 7 copies may consist of partial, non-contiguous gene copies. The exact location, order and functionality of these multiple gene copies are not known and therefore do not provide information on their role in the SMA disease mechanism. Expression studies, using qRT-PCR, could potentially provide information about the functional effect of these multiple gene copies.

It has been previously hypothesised that complex population-specific rearrangements of the SMN region could account for the remaining 49% of U/U^B individuals who test negative for the homozygous deletion of *SMN1*, exon 7 (Labrum *et al.*, 2007). This hypothesis is supported by this study, even though the underlying disease mechanism could not be determined.

5.2 Conclusions from RNA extraction and qPCR

The Tempus Blood RNA tubes and extraction kit was successfully validated for future RNA-based studies. The *SMN1*, exon 7 qPCR assay holds promise as an alternative cost-effective and less laborious quantitative technique to MLPA.

5.3 The clinical significance of this study

The P021-A2 MLPA kit did not provide any further clues on the location, order, completeness and functionality of genes in the SMN region. The high cost and

complexity of MLPA does not make it a viable alternative to the current diagnostic assay. Due to the low frequency of heterozygous *SMN1*, exon 7 deletions, and the fact that no additional pathogenic CNVs were identified in U/U^{B} individuals, MLPA is also not an appropriate second-line diagnostic test for SMA.

The hypervariabilty of the SMN region need to be taken into account when counselling black families with SMA. Even in families where the common mutation is identified, multiple gene copies and novel deletion/duplication/conversion events could still complicate carrier testing for the potential partners and prenatal testing. Carrier and prenatal testing can be offered to families who have an immediate family member who is affected with SMA and if sufficient family data is available to interpret the phase of the potential multiple copies of the *SMN1* region.

5.4 The hypervariable SMN region

There is a limited understanding of the genomic architecture of the SMN region due to the high complexity and homology of the region. Perhaps it is time to return to basics and reinvestigate the structure of the SMN region before investigating the disease mechanism.

5.5 Future research

5.5.1 RNA expression

A future MSc (Med) project (2017), to be supervised by the researcher, will be focussed on RNA expression studies to test for the presence and dosage of FL-SMN transcripts in U/U^{B} individuals. The genomic MLPA results of affected individuals with little or no expression of FL-SMN transcripts identified by qRT-PCR can be investigated to try and identify corresponding genomic pathogenic CNV patterns. If multiple copies of the *SMN1* gene are present on MLPA but there is no corresponding FL-SMN transcript, it could be indicative of partial non-functional *SMN1* copies.

Once optimised, the qRT-PCR assay could be offered as a second-line test to patients who test negative on the homozygous deletion of *SMN1*, exon 7 assay.

5.5.2 Transcriptome analysis

Seguencing analysis of the cDNA of U/U^{B} individuals found to have heterozygous *SMN1*, exon 7 deletions, could be performed to potentially identify the second pathogenic mutation in these patients, although there is little supporting evidence for the presence of additional common pathogenic mutations since no clear haplotype could be identified in U/U^{B} individuals with heterozygous deletions of *SMN1*, exon 7.

Furthermore, sequencing of the SMN1 gene was previously performed on random unrelated black individuals in the Division, in order to look for additional pathogenic mutations in the black SA population. Only 38.5% (5/13) of individuals could be fully sequenced, most probably due to the interference of large CNVs of the SMN region and no clear additional pathogenic mutations were identified (Labrum *et al.*, 2007).

Mutations affecting mRNA splicing has been estimated to be the causastive factor of 15-60% of genetic diseases. Next-generation sequencing of the full transcriptome has been shown to detect mutations causative of SMA phenocopies, previously missed by other NGS sequencing strategies. This could be a possible avenue for future research (Kernohan et al., 2017).

5.5.3 Next generation and third generation sequencing

A custom Next-generation sequencing panel could be designed to investigate SMA phenocopies and other neuromuscular disorders with overlapping clinical symptoms in U/U^{B} individuals, although this might be challenging due to the complex nature of the SMN region (Illumina, 2016). Patients, in whom diagnoses of an alternative neuromuscular disease have been excluded (by including on an NGS neuromuscular panel as discussed in section 4.1.1), can then be further investigated for SMA. It is possible that the CNV profiles of U/U^{B} individuals who do not in fact have SMA could mask pathogenic CNV profiles of patients who are affected with SMA.

CNV testing using NGS sequencing is becoming increasingly popular but analysis has proven difficult in regions of high homology such as the SMN region. Feng *et al.* (2017) developed a method called paralogous gene copy number analysis by ratio and sum which is reported to accurately distinguish between one, two and three or more *SMN1* copies (Feng et al., 2017).

Third generation nanopore sequencing technology (Kasianowicz *et al.*, 1996) could be investigated to try and further define the structure of the SMN region. This technology is currently limited to 20 kb reads, which may still be too small to detect the full 500 kb sequence of the SMN region.

Expanded short tandem repeats of up to 3kb have been successfully sequenced in diseases such as spinocerebellar ataxia (Doi et al., 2014). This technology could show some promise for determining the architecture of the SMN region using longer sequencing reads.

5.6 Concluding statements

Multiple copies of *SMN1*, exon 7 were observed as evidence of the incredible hypervariability of the SMN region in the black SA population. These multiple copies confound diagnostic and carrier testing and could potentially consist of partial, non-contiguous copies. Future studies investigating the expression of these multiple gene copies may provide information on their functional effect. No additional pathogenic CNV patterns were identified in U/U^{B} individuals. This study emphasises the lack of understanding of the architecture of the SMN region and the composition of CNVs in the black SA population and needs to be taken into account when counselling and performing diagnostic, carrier and prenatal testing in the black SA population.

Further, new therapies aimed at increasing *SMN2* expression may be problematic in view of the complex architecture of the SMN region and the high number of apparent *SMN2* deletions in the black SA population.

APPENDICES

List of appendices

Appendix A: Summary of probes in P021-A2 MLPA kit

Length (nt)	SALSA MLPA probe	Chromosomal position	
		Reference	Target
64-70-76-82	Q-fragments: DNA quantity; only visible with less than 100 ng sample DNA		
88-92-96	D-fragments: Low signal of 88 or 96 nt fragment indicates incomplete denaturation		
100	X-fragment: Specific for the X chromosome		
105	Y-fragment: Specific for the Y chromosome		
140	Reference probe 01061-0L0727	12q15	
148	Reference probe 01254-L00815	5p13	
157	Reference probe 01112-L00549	3q12	
166	Reference probe 01448-L00932	17p11	
175	Reference probe 00808-L00638	18q21	
185 ¬	**GTF2H2 probe** 01256-L00972		Exon 11 in NM_001515.3 Detects C>G transition
193	Reference probe 01115-L00005	7q21	
202 ¬	**RAD17 probe** 01257-L00184		Exon 14 in NM_002873.1
210	Reference probe 01220-L00689	10p15	
218 ¬	**GTF2H2 probe** 01813-L00818		Exon 8 in NM_001515.3 Detects G>A transition
229	Reference probe 01120-L00060	11q13	
238 ¬	**NAIP probe** 01259-L00811		Exon 5 in NM_004536.2
247	Reference probe 00816-L00334	21q11	
255	Reference probe 00807-L00325	18q23	
270	**SMN1 probe** 01260-L00966		Exon 7 probe - indicates # of SMN1 copies
276 ⌀	**SMN2 probe** 01260-L00967		Exon 7 probe - indicates # of SMN2 copies
285	Reference probe 00824-L00970	3q25	
294	**SMN1 probe** 01812-L01373		Exon 8 – usually indicates # of SMN1 copies
300	**SMN2 probe** 01812-L01372		Exon 8 – usually indicates # of SMN2 copies
310	Reference probe 00871-L00461	13q34	
319	Reference probe 01042-L00791	8q24	
328 ¬	**GTF2H2 probe** 01262-L00971		Exon 5 in NM_001515.3 Detects both GTF2H2 variants
337	Reference probe 00812-L00330	21q21	
346 ¬	**NAIP probe** 01263-L00812		Exon 13 in NM_004536.2
355	Reference probe 00965-L00552	2p13	
364	**SMN1/2 probe** 01814-L00807		Exon 8 Detects # of SMN1 + SMN2 copies
373	Reference probe 01046-L00624	8p11	
382	**SMN1/2 probe** 01265-L00808		Exon 1 Detects # of SMN1 + SMN2 copies
391	Reference probe 01160-L00716	3p25	
400	**SMN1/2 probe** 01816-L00809		Exon 4 Detects # of SMN1 + SMN2 copies
409	Reference probe 00963-L09340	2p16	
419	**SMN1/2 probe** 01815-L00810		Exon 6 Detects # of SMN1 + SMN2 copies
427	Reference probe 01108-L00679	8q23	
433	Reference probe 01057-L00630	17q12	
445	Reference probe 00802-L00320	13q34	
454 ¬	**SERF1B probe** 01269-L00813		Exon 1/Upstream in NM_022978.2
463	Reference probe 00846-L00377	12q24	

From P021-A2 SMA product description (description version 31; 13-07-2013) (http://www.mlpa.com/WebForms/WebFormProductDetails.aspx?).

Appendix B: MLPA worksheet and step-by-step procedure

MLPA worksheet – one tube procedure

Adapted from MLPA Protocol One-tube MDP-v005 (MRC Holland, 2016)

PROBE KIT:_____

DATE:_____ ☐ Bench decontaminated PERFORMED BY:_____

DAY 1:

1.) Hybridisation protocol

- Dilute DNA in TE Buffer (40ng, but can use 20-250ng)
- Allow DNA to resuspend for 45 min
- Add 5µl diluted DNA to 0.5ml PCR tube.
- Denature DNA. **Denaturation protocol:** 98°C – 5 minutes, 25°C – hold
- Prepare Hybe mix as explained below. Vortex reagents briefly before use.

Hybe Master Mix

Reagent	Batch number	1x volume	x mix
SALSA probe mix (black cap)		1.5µl	µl
MLPA buffer (yellow cap)		1.5µl	µl
TOTAL	-	**3µl**	**3µl/reaction**
Mix thoroughly and centrifuge. Extra hybe mix can be frozen and reused.			

- Remove tubes from PCR machine & add 3µl Hybe mix to each tube containing denatured DNA
- Mix by pipetting
- Start **Hybridisation protocol:**
  95°C – 1 minute; 60°C – 18 hours (12-26 hours); 54°C – hold

---

DAY 2:

2.) Ligation protocol

- Prepare a Ligase-65 mastermix
- Vortex Buffer A and B (MRC Holland) before use

Ligase-65 Master Mix

Reagent	Batch number	1x volume	x mix
Ligase-65 buffer A (transparent cap)		3µl	µl
Ligase-65 buffer B (white cap)		3µl	µl
ddH_2O	-	25µl	µl
Mix thoroughly and centrifuge.			
Ligase-65 (green cap)		1µl	µl
TOTAL	-	-	**32µl/reaction**
Mix by pipetting and centrifuge. Extra Ligase-65 mix can be frozen and reused.			

- While the hybridised samples are at 54°C, add 32µl of the Ligase-65 master mix (as described above) to each reaction, mixing by pipetting up and down. Each reaction will consist of total volume of 40µl.
- Start **Ligation protocol** at 54°C – 15 minutes; 98°C – 5 minutes; 20°C – hold
- Remove tubes from PCR machine
- Can store ligated products at 4°C

3.) PCR protocol

- Vortex & centrifuge the SALSA PCR primer mix before use
- Warm the polymerase for 10s in your hand to reduce viscosity
- Prepare the polymerase mix as explained below and store on ice until use

Polymerase Mix:

Reagent	Batch number	1x volume	x mix
SALSA PCR primer mix(brown tube)*		2µl	µl
ddH_2O	-	7.5µl	µl
SALSA polymerase (orange cap)		0.5µl	µl
TOTAL	-	-	**10µl/reaction**
Mix by pipetting and centrifuge. Extra Polymerase mix can be frozen and reused.			

* = FAM labeled primers, One tube PCR primer mix has the MRC Holland logo on the label

- At room temperature add 10µl of Polymerase mix to each tube of ligated product to make up a final volume of 50µl per reaction
- Mix by pipetting & centrifuge
- Start **PCR protocol:**

95°C – 30 seconds; 60°C – 30 seconds; 72°C – 1 minute **x35 cycles**

72°C – 20 minutes, 4°C – hold

- PCR products can be stored for one week at 4°C. For longer periods, storage between 25°C and -15°C is recommended. As the the fluorescent dyes used are light sensitive, the PCR products should be stored in a black bag or aluminium foil.
- Load 1μl of the PCR product with 0.5μl Liz-500 marker and 8.5μl Hi-Di Formamide (Applied Biosystems, Foster City, CA, USA) into a 96-well plate
- Perform fragment analysis on ABI 3130*xl* genetic analyzer

Appendix C: SMA qPCR worksheet: relative quantification of *SMN1* and *SMN2*, exon 7 genomic copy number

qPCR SMN1, SMN2, CFTR PCR worksheet

Run name:	Bench decontaminated? Yes/No	Date:

qPCR Universal PCR Master Mix		Input values	Output values			
	[Stock]	(M)	[Final]	[Final] (M)	1x Volume (µl)	Mastermix (µl)
TaqMan Universal MM	2X		1X		5	105
Forward Primer: SMN-ex7-F / CFTR-F	100 pmole/µl	100 µM	0.9 pmole/µl	900 nM (0.9 µM)	1	21
Reverse Primer: SMN-ex7-R/ CFTR-R	100 pmole/µl	100 µM	0.9 pmole/µl	900 nM (0.9 µM)	1	21
TaqMan Probe: SMN1-Ex7 / SMN2-Ex7 / CFTR	10 pmole/µl	10 µm	0.25 pmole/µl	250 nM (0.25µM)	1	21
DNA	X ng/µl		25-100 ng/µl		1	n/a
H2O					1	21
TOTAL					10	189

9 µl per reaction

Add 1 µl DNA/reaction

How to dilute 100 pmole/µl primer to 900 nM:	10	C1
Make 1/10 dil of 100 pmole/µl primer stock	0.9	C2
900 nM = 900 000 pmole/1 000 000 µl	9	V1 to calculate
900 nM = 0.9 pmole/µl	100	V2
To add V2 value of primer into this amount of ddH2O:	91	H20

How to prepare a dilution of 10 µM for each of the SMN1,	10000	C1

SMN2 and CFTR probes to 250 nM, separately:		
10 µM = 10 000 nM	250	C2
	2.5	V1 to calculate
	100	V2
To add V2 value of probe into this amount of ddH2O:	97.5	H20

Appendix D: Reverse Transcription of RNA worksheet

Reverse Transcription of RNA using Promega ImProm-II system

- Extract RNA using the Tempus RNA tubes and Tempus spin RNA isolation kit (see https://www.thermofisher.com/order/catalog/product/4380204 for the manual)

Table 1: RNA Extraction using Tempus Spin RNA Isolation kit

Name	Code	[RNA] (ng/μl)	260/280 (>2.1)	260/230 (~2)	Conditions/ comment

Check RNA on 0.8 % gel (1μl)

- Not necessary to thaw reagents on ice
- Thaw RNA and enzymes on ice

Table 2: RNA dilution for reverse transcription

Reagent	Lot #	Volume / rxn (μl)
Primer ($Oligo(dT)_{15}$)		1
RNA (~430 ng)		4
TOTAL	-	**5μl**

RNA volume will vary depending on [RNA]. Add 2μl for +ve ctrl.

- Set up positive ctrl (supplied) & blank (no RNA added) with every reaction
- 70 °C heating block for 5 min (double stranded RNA becomes single stranded)
- Place on ice for at least 5 min
- Spin and keep on ice until transcription mix is added

Table 3: Reverse transcription reaction

Reagent	Lot #	Volume (µl)	Master mix x___ (µl)
Nuclease-free water		3.7	
ImProm-II 5x reaction buffer		4	
$MgCl_2$		4.8	
dNTP mix		1	
Recombinant RNasin Ribonuclease Inhibitor		0.5	
ImProm-II Reverse Transcriptase		1	
RNA dilution mix	-	5	-
TOTAL	-	**20 µl**	**15 µl / rxn**

Set up a blank (no reverse transcriptase) **separately** for each reaction

Perform reverse transcription reaction in PCR machine:

25°C – 5 min, 42°C - 60 min, 70°C – 15 min, 4°C – Hold (Store in fridge)

Table 4: Reverse Transcription Reaction:

Sample	Name	Code	Comments
1	Positive ctrl		
2	Blank (no RNA)		
3	Blank (no reverse transcriptase)		
4			
5			
6			
7			
8			
9			
10			

Check on 0.8% agarose gel. Adjust number of rows as necessary.

Appendix E: Ethics clearance certificate

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M130950

NAME: (Principal Investigator)	Ms Elana Vorster and Prof John Rodda
DEPARTMENT:	Pathology, Division of Human Genetics
PROJECT TITLE:	Determining the Molecular Basis of Spinal Muscular Atrophy in the South African Black Population
DATE CONSIDERED:	27/05/2013
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Prof Amanda Krause & Ms Fahmida Essop
APPROVED BY:	Professor PE Cleaton-Jones, Chairperson HREC (Medical)

DATE OF APPROVAL: 21/06/2013

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and ONE COPY returned to the Secretary at Room 10004, 10th floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to resubmit the application to the Committee. I agree to submit a yearly progress report.

Elana Vorster
Principal Investigator Signature

07/01/2014
M130950 Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix F: Statistical test output

1.) Section 3.1.2: Testing for significance in results generated from the SMA audit using the Chi-Square test of Independence: *SMN1*

Chi-Square test result - different genders

Expected Values

Gender	P	N	Total
F	169.785	369.215	539
M	208.215	452.785	661
Total	378	822	1200

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
1200	2	2	1

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*	*Odds Ratio*
Pearson's	0.602937	0.43746	3.841459	no	0.022415	1.10171
Max likeli	0.602289	0.437706	3.841459	no	0.022403	1.10171

Chi-Square test result - different ethnic groups - *SMN1* deletions

Expected Values

Ethnicity	P	N	Total
B	314.9936	676.0064	991
C	5.085669	10.91433	16
I	15.89271	34.10729	50
W	61.02802	130.972	192
Total	397	852	1249

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
1249	4	2	3

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	0.283233	0.96315	7.814728	no	0.015059
Max likeli	0.276371	0.964409	7.814728	no	0.014875

2.) Section 3.1.3: Testing for significance in results generated from the SMA audit using the Chi-Square test of Independence: *SMN2*

Chi-Square test result - SMN2 deletions among different poulation groups

Expected Values

Ethnicity	P	N	Total
B	108.7006	882.2994	991
MA	1.755004	14.245	16
I	5.484388	44.51561	50
W	21.06005	170.94	192
Total	137	1112	1249

Chi-Square Test

SUMMARY	Alpha	0.05	
Count	*Rows*	*Cols*	*df*
1249	4	2	3

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	13.34835	0.003941	7.814728	yes	0.103379
Max likeli	15.7241	0.001292	7.814728	yes	0.112202

Chi-square test result – SMN2 deletions between black (B) and white (W) patients

Ethnicity	P	N
B	123	868
W	9	183

Expected Values

Ethnicity	P	N	Total
B	110.5765	880.4235	991
W	21.4235	170.5765	192
Total	132	1051	1183

Chi-Square Test

SUMMARY	Alpha	0.05	
Count	*Rows*	*Cols*	*df*
1183	2	2	1

CHI-SQUARE				*sig*	*Cramer V*	*Odds Ratio*
	chi-sq	*p-value*	*x-crit*			
Pearson's	9.68034	0.001863	3.841459	yes	0.090459	2.881336
Max likeli	11.6423	0.000645	3.841459	yes	0.099204	2.881336

Chi-square test result – SMN2 deletions between black (B) and mixed ancestry (MA) patients

Ethnicity	P	N
B	123	868
MA	3	13

Expected Values

Ethnicity	P	N	Total
B	123.998	867.002	991
MA	2.001986	13.99801	16
Total	126	881	1007

Chi-Square Test

SUMMARY	Alpha	0.05	
Count	*Rows*	*Cols*	*df*
1007	2	2	1

CHI-SQUARE				*sig*	*Cramer V*	*Odds Ratio*
	chi-sq	*p-value*	*x-crit*			
Pearson's	0.577858	0.447153	3.841459	no	0.023955	0.614055
Max likeli	0.512919	0.473878	3.841459	no	0.022569	0.614055

Chi-Square test result - SMN2 deletions between black (B) & Indian (I) patients

Ethnicity	P	N
B	123	868
I	2	48

Expected Values

Ethnicity	P	N	Total
B	118.9962	872.0038	991
I	6.003842	43.99616	50
Total	125	916	1041

Chi-Square Test

SUMMARY	Alpha	0.05	
Count	*Rows*	*Cols*	*df*
1041	2	2	1

CHI-SQUARE				*sig*	*Cramer V*	*Odds Ratio*
	chi-sq	*p-value*	*x-crit*			
Pearson's	3.18755	0.074201	3.841459	no	0.055335	3.400922
Max likeli	4.116109	0.042472	3.841459	yes	0.062881	3.400922

Chi-Square test result - SMN2 deletions between white (W) & mixed ancestry (MA) patients

Ethnicity	P	N
MA	3	13
W	9	183

Expected Values

Ethnicity	P	N	Total
MA	0.923077	15.07692	16
W	11.07692	180.9231	192
Total	12	196	208

Chi-Square Test

SUMMARY	Alpha	0.05	
Count	*Rows*	*Cols*	*df*
208	2	2	1

CHI-SQUARE				*sig*	*Cramer V*	*Odds Ratio*
	chi-sq	*p-value*	*x-crit*			
Pearson's	5.372449	0.020457	3.841459	yes	0.160714	4.692308
Max likeli	3.658405	0.055787	3.841459	no	0.132622	4.692308

Chi-Square test result - SMN2 deletions between white (W) & Indian (I) patients

Ethnicity	P	N
I	2	48
W	9	183

Expected Values

Ethnicity	P	N	Total
I	2.272727	47.72727	50
W	8.727273	183.2727	192
Total	11	231	242

Chi-Square Test

SUMMARY	Alpha	0.05	
Count	*Rows*	*Cols*	*df*
242	2	2	1

CHI-SQUARE				*sig*	*Cramer V*	*Odds Ratio*
	chi-sq	*p-value*	*x-crit*			
Pearson's	0.043214	0.835322	3.841459	no	0.013363	0.847222
Max likeli	0.044518	0.832893	3.841459	no	0.013563	0.847222

Chi-Square test result - SMN2 deletions between mixed ancestry (MA) & Indian (I) patients

Ethnicity	P	N
MA	3	13
I	2	48

Expected Values

Ethnicity	P	N	Total
MA	1.212121	14.78788	16
I	3.787879	46.21212	50
Total	5	61	66

Chi-Square Test

SUMMARY	Alpha	0.05	
Count	*Rows*	*Cols*	*df*
66	2	2	1

CHI-SQUARE				*sig*	*Cramer V*	*Odds Ratio*
	chi-sq	*p-value*	*x-crit*			
Pearson's	3.766328	0.052294	3.841459	no	0.238884	5.538462
Max likeli	3.176539	0.074703	3.841459	no	0.219384	5.538462

3.) Section 3.2.1.1: Testing for significance in results between *SMN1*, exon 7 copy number among various African populations using the Chi-Square test of Independence

Testing for difference between African American & Sub-Saharan populations

Expected Values

Mali	Nigeria	Kenya	African-Ai	Total
3	1.338645	1.338645	11.32271	14
295	64.44622	64.44622	545.1076	674
330	54.21514	54.21514	458.5697	567
Total	120	120	1015	1255

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
1255	3	3	4

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*		*Cramer V*
Pearson's	5.99662504	0.1994	9.487729	no	0.048878
Max likeliho	6.015782	0.197973	9.487729	no	0.048956

Comparisons of *SMN1*, exon 7 copy number across different African populations

Comparison: Black SA & Mali

Expected Values

SMN1 cop	Black SA	Mali	Total
1	0.488	2.512	3
2	57.74667	297.2533	355
>2 (3-5)	63.76533	328.2347	392
Total	122	628	750

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
750	3	2	2

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	0.746179	0.688604	5.991465	no	0.031542

Comparison: Black SA & Nigeria

Expected Values

SMN1 cop	Black SA	Nigeria	Total
1	1.008264	0.991736	2
2	66.54545	65.45455	132
>2 (3-5)	54.44628	53.55372	108
Total	122	120	242

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
242	3	2	2

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	5.445122	0.065706	5.991465	no	0.150002

Comparison: Black SA & Kenya

>2 (3-5)	62	46

Expected Values

SMN1 cop	Black SA	Kenya	Total
1	0.504132	0.495868	1
2	67.04959	65.95041	133
>2 (3-5)	54.44628	53.55372	108
Total	122	120	242

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
242	3	2	2

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	4.624834	0.099022	5.991465	no	0.138242

Comparison: Black SA & African American

1	0	11
2	60	529
>2 (3-5)	62	475

Expected Values

SMN1 cop	Black SA	African-Ai	Total
1	1.180299	9.819701	11
2	63.19965	525.8004	589
>2 (3-5)	57.62005	479.3799	537
Total	122	1015	1137

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
1137	3	2	2

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	1.876586	0.391295	5.991465	no	0.040626

4.) Section 3.2.1.1: Testing for significance in results between *SMN1*, exon 7 copy number among various Caucasian populations using the Chi-Square test of Independence

Testing for difference between white American & European populations

Expected Values

American	Germany	France	Sweden	Total
28	2.882027	12.78385	10.33413	26
935	133.46	591.9905	478.5495	1204
65	3.657957	16.22565	13.11639	33
Total	140	621	502	1263

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
1263	3	3	4

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	0.811775	0.936863	9.487729	no	0.017927
Max likeli	0.772702	0.94207	9.487729	no	0.01749

Comparisons of *SMN1*, exon 7 copy number across different Caucasian populations

Comparison: White SA & American

Expected Values

	White SA	American	Total
1	0.793951	27.20605	28
2	27.33459	936.6654	964
>2 (3-5)	1.871456	64.12854	66
Total	30	1028	1058

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
1058	3	2	2

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	1.339191	0.511916	5.991465	no	0.035578

Comparison: White SA & Germany

Expected Values

	White SA	Germany	Total
1	0.705882	3.294118	4
2	28.41176	132.5882	161
>2 (3-5)	0.882353	4.117647	5
Total	30	140	170

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
170	3	2	2

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	0.890979	0.640511	5.991465	no	0.072395

Comparison: White SA & France

1	0	13
2	29	593
>2 (3-5)	1	15

Expected Values

	White SA	France	Total
1	0.599078	12.40092	13
2	28.66359	593.3364	622
>2 (3-5)	0.737327	15.26267	16
Total	30	621	651

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
651	3	2	2

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	0.730256	0.694108	5.991465	no	0.033492

Comparison: White SA & Sweden

2	29	479
>2 (3-5)	1	14

Expected Values

	White SA	Sweden	Total
1	0.507519	8.492481	9
2	28.64662	479.3534	508
>2 (3-5)	0.845865	14.15414	15
Total	30	502	532

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
532	3	2	2

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	0.572234	0.751175	5.991465	no	0.032797

5.) Sections 3.2.1.3 & 3.2.4.3: Comparison of homozygous *SMN2*, exon 7 deletion rates between the audit & this study

Comparison of black homozygous SMN2, exon 7 deletions between audit and this study

SMN1 cop	Audit: Black SA	Study: Black SA
P	123	33
N	868	89
Total	991	122

Chi-Square Test

SMN1 cop	Audit: Bla	Study: Bla	Total
P	138.9003	17.09973	156
N	852.0997	104.9003	957
Total	991	122	1113

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
1113	2	2	1

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*	*Odds Ratio*
Pearson's	19.31188	1.11E-05	3.841459	yes	0.131724	0.382174
Max likelihood	16.32158	5.35E-05	3.841459	yes	0.121097	0.382174

Comparison of white homozygous SMN2, exon 7 deletions between audit and this study

SMN1 cop	Audit: White SA	Study: White SA
P	9	2
N	183	28
Total	192	30

Expected Values

Chi-Square Test

SMN1 cop	Audit: Wh	Study: Wh	Total
P	9.513514	1.486486	11
N	182.4865	28.51351	211
Total	192	30	222

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
222	2	2	1

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*	*Odds Ratio*
Pearson's	0.215807	0.642254	3.841459	no	0.031179	0.688525
Max likelihood	0.198879	0.655627	3.841459	no	0.029931	0.688525

Comparison of black homozygous *SMN2*, exon 7 deletion between audit and patients clinically suggestive of SMA

SMN1 copy number	Audit: Black SA	Suggestive
P	123	10
N	868	62
Total	991	72

Expected Values

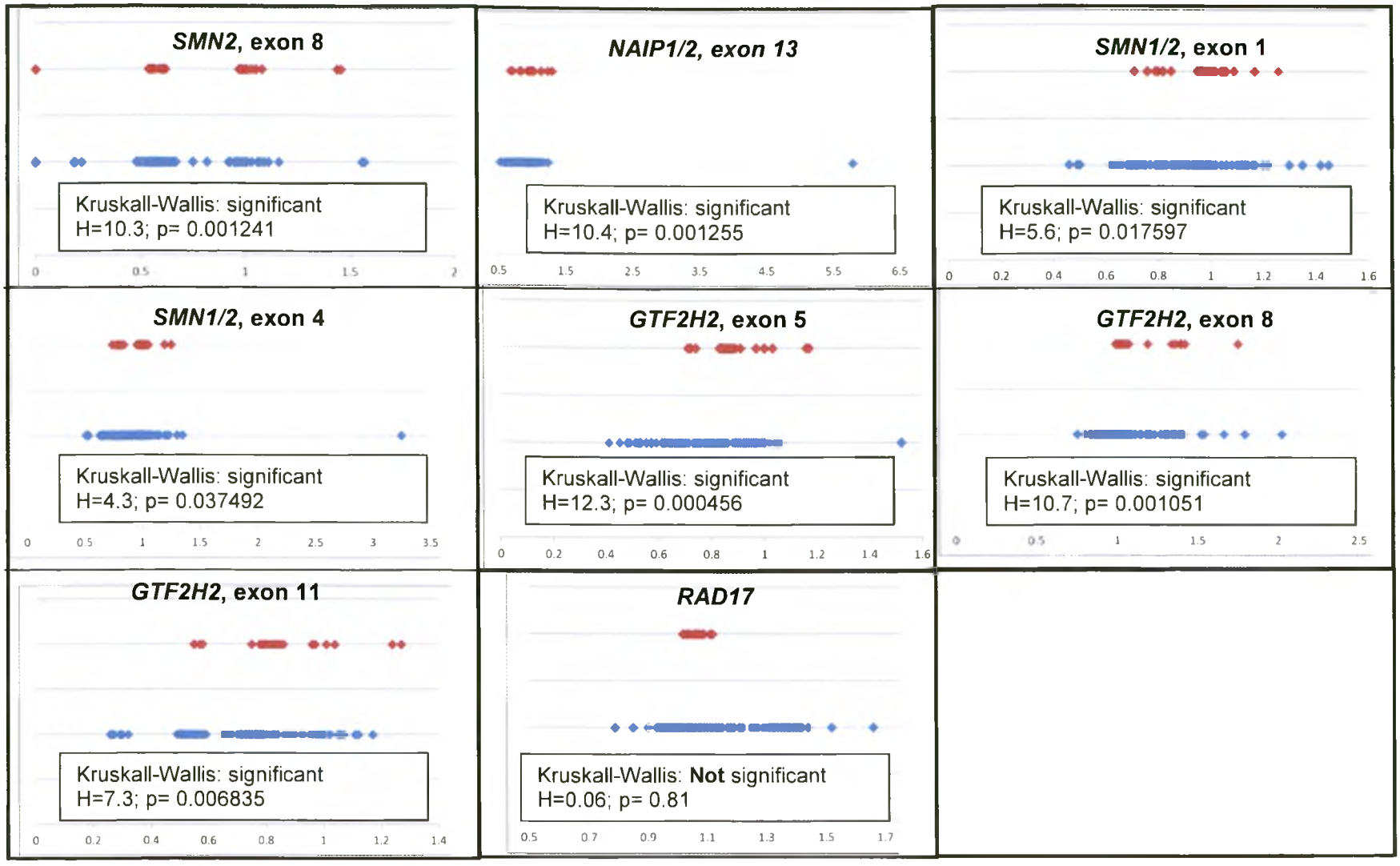
Chi-Square Test

SMN1 cop	Audit: Bla	Suggestiv	Total
P	123.9915	9.008467	133
N	867.0085	62.99153	930
Total	991	72	1063

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
1063	2	2	1

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*	*Odds Ratio*
Pearson's	0.133805	0.714519	3.841459	no	0.011219	0.878571
Max likeli	0.130112	0.718316	3.841459	no	0.011063	0.878571



6.) Section 3.2.1.6: Significant differences in variance between N/N^B and N/N^W individuals across various MLPA probes

7.) Section 3.2.4.4: Chi-squared and Hardy-Weinberg tests to determine significant differences of the rs200357275 & rs201102513 variants

rs201102513

Patient gr	G	A	Total
BT	148.8372	1.162791	150
WT	59.53488	0.465116	60
BC	99.22481	0.775194	100
WC	15.87597	0.124031	16
BN	242.1085	1.891473	244
WN	59.53488	0.465116	60
BS	142.8837	1.116279	144
		6	774

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
774	7	2	6

CHI-SQUARE	*chi-sq*	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	25.155	0.00032	12.59159	yes	0.180278

Patient gr	G	A
BT	144	6
WT	60	0
BC	100	0
WC	16	0
BN	244	0
WN	60	0
BS	144	0

rs2000357275

	rs200357275 (GTF2H2-11)	
	G	C
BT	120	30
WT	42	18
BC	19	81
WC	5	11
BN	59	185
WN	55	5
BS	24	120

Expected Values

	G	C	Total
BT	62.7907	87.2093	150
WT	25.11628	34.88372	60
BC	41.86047	58.13953	100
WC	6.697674	9.302326	16
BN	102.1395	141.8605	244
WN	25.11628	34.88372	60
BS	60.27907	83.72093	144
Total	324	450	774

Chi-Square Test

SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
774	7	2	6

CHI-SQUARE	*chi-sq*	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	261.439	1.47E-53	12.59159	yes	0.581185
Max likeli	277.7259	4.82E-57	12.59159	yes	0.599015

Hardy-Weinberg equilibrium

BS group	GG	GC	CC	Total
observed	5	14	53	72
expected	2	20	20	42
(O-E)^2/E	4.5	1.8	54.45	60.75

p	0.166667	Alpha	0.05
q	0.833333	df	1
p2	0.027778	x-crit	3.84
q2	0.694444	Test value	60.75
2pq	0.277778		

Since 60.75 > 3.84, we reject the null hypothesis. The sample is not in HW equilibrium

8.) Section 3.4.2.2: MANOVA test to compare non-deletion negative chromosomal CNV patterns of parents of positive SMA families and negative chromosomal CNV patterns of negative controls

Multiple ANOVA

ANOVA: Single Factor [SMN1-7]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	0.02467	1	0.02467	0.12616	0.72311	9.07273	0.06312	-0.0092
Within Gr	18.3816	94	0.19555					
Total	18.4063	95	0.19375					

ANOVA: Single Factor [SMN1-8]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	0.21228	1	0.21228	0.84477	0.36039	9.07273	0.16333	-0.0016
Within Gr	23.6211	94	0.25129					
Total	23.8333	95	0.25088					

ANOVA: Single Factor [SMN2-7]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	0.02149	1	0.02149	0.0653	0.79887	9.07273	0.04541	-0.0098
Within Gr	30.9368	94	0.32912					
Total	30.9583	95	0.32588					

ANOVA: Single Factor [SMN2-8]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	0.46327	1	0.46327	1.46248	0.22957	9.07273	0.2149	0.00479
Within Gr	29.7763	94	0.31677					
Total	30.2396	95	0.31831					

ANOVA: Single Factor [RAD17-17]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	0	1	0	#DIV/0!	#DIV/0!	9.07273	#DIV/0!	#DIV/0!
Within Gr	0	94	0		no difference in copy number at all			
Total	0	95	0					

ANOVA: Single Factor [SERF1B-up]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	0	1	0	0	1	9.07273	0	-0.0105
Within Gr	2	94	0.02128					
Total	2	95	0.02105					

ANOVA: Single Factor [SMN1-1]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	0.00175	1	0.00175	0.02828	0.86682	9.07273	0.02988	-0.0102
Within Gr	5.83158	94	0.06204					
Total	5.83333	95	0.0614					

ANOVA: Single Factor [SMN1-4]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	0.01096	1	0.01096	0.26111	0.61056	9.07273	0.09081	-0.0078
Within Gr	3.94737	94	0.04199					
Total	3.95833	95	0.04167					

ANOVA: Single Factor [SMN1-6]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	0.01096	1	0.01096	0.52928	0.46872	9.07273	0.12928	-0.0049
Within Gr	1.94737	94	0.02072					
Total	1.95833	95	0.02061					

ANOVA: Single Factor [SMN1-8]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	0.01096	1	0.01096	0.26111	0.61056	9.07273	0.09081	-0.0078
Within Gr	3.94737	94	0.04199					
Total	3.95833	95	0.04167					

ANOVA: Single Factor [NAIP-13]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	0.03958	1	0.03958	1.2613	0.26427	9.07273	0.19958	0.00271
Within Gr	2.95	94	0.03138					
Total	2.98958	95	0.03147					

ANOVA: Single Factor [NAIP-5]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	1.91053	1	1.91053	6.75416	0.01086	9.07273	0.46183	0.05655
Within Gr	26.5895	94	0.28287					
Total	28.5	95	0.3					

ANOVA: Single Factor [GTF2H2-5]

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between	0.01853	1	0.01853	0.26918	0.6051	9.07273	0.0922	-0.0077
Within Gr	6.47105	94	0.06884					
Total	6.48958	95	0.06831					

9.) Section 4.2.2.1: Comparing the homozygous *SMN1*, exon 7, 8 and *NAIP*, exon 5 deletion results of the current study with data from previous SA studies

Testing if there is a difference in results between previous SA study and this study: Extent of homozygous *SMN1*, exon 7 del in white patients

	White patients	
Deletion	**Previous SA study**	**This study**
SMN1, exon 7 deletion	5	7
SMN1, exon 7 and 8 deletions	8	11
SMN1, exon 7 and *NAIP* deletions	0	1
SMN1, exon 7 and 8 and *NAIP* deletions	10	11
Total	23	30

Expected Values

Deletion	Previous S	This study	Total
SMN1, exon 7 deletion	5.207547	6.792453	12
SMN1, exon 7 and 8 deletion	8.245283	10.75472	19
SMN1, exon 7 and NAIP delet	0.433962	0.566038	1
SMN1, exon 7 and 8 and NAIP	9.113208	11.88679	21
Total	23	30	53

Chi-Square Test

SUMMARY	Alpha	0.05	
Count	*Rows*	*Cols*	*df*
53	4	2	3

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	0.946621	0.814165	7.814728	no	0.133644

Testing if there is a difference in results between previous SA study and this study: Extent of homozygous *SMN1*, exon 7 del in black patients

	Black patients	
Deletion	**Previous SA study**	**This study**
SMN1, exon 7 deletion	18	29
SMN1, exon 7 and 8 deletions	15	17
SMN1, exon 7 and *NAIP* deletions	4	8
SMN1, exon 7 and 8 and *NAIP* deletions	10	21
Total	47	75

Expected Values

Deletion	Previous S	This study	Total
SMN1, exon 7 deletion	18.10656	28.89344	47
SMN1, exon 7 and 8 deletion	12.32787	19.67213	32
SMN1, exon 7 and NAIP delet	4.622951	7.377049	12
SMN1, exon 7 and 8 and NAIP	11.94262	19.05738	31
Total	47	75	122

Chi-Square Test

SUMMARY	Alpha	0.05	
Count	*Rows*	*Cols*	*df*
122	4	2	3

CHI-SQUARE

	chi-sq	*p-value*	*x-crit*	*sig*	*Cramer V*
Pearson's	1.593747	0.660809	7.814728	no	0.114296

Appendix G: Haplotypes constructed from pedigrees of N/N^{B} families

Number of copies: 0 | Normal amount expected: 1 for single genes, 2 for combined genes | 3 | 4

Family	Origin of chromo-some	Family chromosome	Copy Number for SMN Region (MLPA)															Haplotype number
			RAD17	SERF1B	SMN2 Ex 7	SMN2 Ex 8	SMN1 Ex 7	SMN1 Ex 8	SMN1/2 Ex 1	SMN1/2 Ex 4	SMN1/2 Ex 6	SMN1/2 Ex 8	NAIP1 Ex5	NAIP1/2 Ex 13	GTF2H2 Ex 5	GTF2H2-8	GTF2H2-11	
4	Mother	B	1	1	1	0	1	1	2	2	2	2	1	2	2	1	1	1
25	Mother	A	1	1	0	0	1	1	2	2	2	2	1	2	2	1	1	2
25	Father	D	1	1	0	0	1	1	2	2	2	2	1	2	2	1	1	2
8	Mother	A	1	1	0	0	1	1	2	2	2	2	1	2	2	1	1	2
1	Mother	A	1	1	0	0	1	1	2	2	2	2	1	2	2	1	1	2
6	Mother	B	1	1	0	0	1	1	2	2	2	2	1	2	2	1	1	2
5	Father	C	1	1	0	0	1	1	2	2	2	2	1	2	2	1	0	2
7	Mother	B	1	1	0	0	1	1	2	2	2	2	1	2	2	1	0	2
25	Mother	B	1	1	0	0	1	1	0	0	0	0	1	2	2	1	1	3
17	Mother	B	1	0	0	0	1	1	0	0	0	0	2	2	2	1	1	4
40	Mother	B	1	1	1	0	1	1	0	0	2	0	3	0	2	1	1	5
63	Father	D	1	1	0	0	1	1	0	2	2	2	0	2	0	1	1	6
17	Father	C	1	1	0	0	1	1	2	2	2	2	2	2	2	1	1	7
19	Father	C	1	1	0	0	1	1	2	2	2	2	2	2	2	1	1	7
26	Mother	A	1	1	0	0	1	1	2	2	2	2	2	2	2	1	1	7
17	Mother	A	1	2	0	0	1	1	2	2	2	2	2	2	2	1	1	8
2	Father	D	1	1	1	2	1	0	2	2	2	2	1	2	2	1	1	9
63	Mother	A	1	1	0	2	1	1	2	2	2	2	2	2	2	1	1	10
25	Father	C	1	1	2	1	2	2	2	2	2	2	1	4	2	1	1	11
56	Mother	B	1	1	1	0	3	2	2	2	2	2	1	2	2	1	1	12
7	Father	D	1	1	0	0	2	2	2	2	2	2	1	2	2	1	1	13
8	Mother	B	1	1	0	0	2	2	2	2	2	2	1	2	2	1	1	13

Family	Family member	Family chromosome	Copy Number for SMN Region (MLPA)															Haplotype
			RAD17	SERF1B	SMN2 Ex 7	SMN2 Ex 8	SMN1 Ex 7	SMN1 Ex 8	SMN1/2 Ex 1	SMN1/2 Ex 4	SMN1/2 Ex 6	SMN1/2 Ex 8	NAIP1 Ex5	NAIP1/2 Ex 13	GTF2H2 Ex 5	GTF2H2-8	GTF2H2-11	
1	Father	D	1	1	0	0	2	2	2	2	2	2	1	2	2	1	1	13
19	Father	D	1	1	0	0	2	2	2	2	2	2	1	2	2	1	1	13
26	Father	C	1	1	0	0	2	2	2	2	2	2	1	2	2	1	1	13
26	Father	D	1	1	0	0	2	2	2	2	2	2	2	2	2	1	1	14
1	Mother	B	1	1	1	1	2	2	2	2	2	2	1	2	2	1	1	15
12	Father	D	1	1	1	0	2	2	2	2	2	2	2	2	2	1	1	16
52	Mother	A	1	1	1	0	2	2	2	2	2	2	2	2	2	1	1	16
2	Mother	B	1	1	0	1	2	2	2	2	2	2	1	2	2	1	0	17
1	Father	C	1	1	1	1	2	1	2	2	2	2	1	2	2	1	1	18
4	Father	D	1	1	0	1	1	1	4	4	2	2	1	2	2	1	1	19
63	Mother	B	1	1	0	1	1	2	2	2	2	2	1	2	2	1	1	20
52	Father	C	1	1	1	0	1	2	2	2	2	2	2	2	2	1	1	21
7	Mother	A	1	1	2	0	1	2	2	2	2	2	1	2	0	1	0	22
17	Father	D	1	1	2	1	1	1	2	2	2	2	1	2	2	1	1	23
40	Father	C	1	1	2	1	1	1	2	2	2	2	1	2	2	1	1	23
63	Father	C	1	1	2	1	1	1	2	2	2	4	1	2	2	1	1	24
26	Mother	B	1	1	1	2	2	1	2	2	2	2	1	2	2	1	1	25
12	Mother	A	1	1	1	1	2	1	2	2	2	2	2	2	2	1	0	26
7	Father	C	1	1	2	1	1	1	2	2	2	2	1	2	2	1	1	27
58	Father	D	1	1	1	1	1	1	2	2	2	2	1	2	0	1	1	28
4	Mother	A	1	1	1	1	1	1	2	2	2	2	2	2	2	1	1	29
5	Mother	B	1	1	1	1	1	1	2	2	2	2	2	2	2	1	0	29
58	Mother	A	1	1	1	1	1	1	2	2	2	2	2	2	0	1	0	30
40	Father	D	1	1	1	1	1	1	2	2	2	2	1	2	0	1	0	31
19	Mother	A	1	1	1	1	1	1	2	2	2	2	0	2	2	1	0	32
19	Mother	B	1	1	1	0	1	1	2	2	2	2	1	2	2	1	0	33
9	Father	D	1	1	1	1	1	1	2	2	2	2	2	2	2	1	0	34
58	Mother	B	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35

Family	Family member	Family chromosome	Copy Number for SMN Region (MLPA)															Haplotype
			RAD17	SERF1B	SMN2 Ex 7	SMN2 Ex 8	SMN1 Ex 7	SMN1 Ex 8	SMN1/2 Ex 1	SMN1/2 Ex 4	SMN1/2 Ex 6	SMN1/2 Ex 8	NAIP1 Ex5	NAIP1/2 Ex 13	GTF2H2 Ex 5	GTF2H2-8	GTF2H2-11	
58	Father	C	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
56	Father	D	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
2	Mother	A	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
2	Father	C	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
4	Father	C	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
5	Mother	A	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
5	Father	D	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
6	Mother	A	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
6	Father	C	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
6	Father	D	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
10	Mother	A	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
10	Mother	B	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
10	Father	C	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
10	Father	D	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
56	Father	C	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
56	Mother	A	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
9	Mother	A	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
9	Mother	B	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
9	Father	C	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
40	Mother	A	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
12	Mother	B	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
12	Father	C	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
52	Mother	B	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
52	Father	D	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
8	Father	C	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	35
8	Father	D	1	1	1	1	1	1	2	2	2	2\	1	2	2	1	1	35

Appendix H: Pathogenic haplotypes constructed from pedigrees of M_1/M_1^B families

Number of copies: 0 | Normal amount expected: 1 for single genes, 2 for combined genes | 3 | 4

Family	Family chromosome	Copy Number for SMN Region (MLPA)															
		RAD17	SERF1B	SMN2 Ex7	SMN2 Ex8	SMN1 Ex7	SMN1 Ex8	SMN1/2 Ex1	SMN1/2 Ex4	SMN1/2 Ex6	SMN1/2 Ex8	NAIP1 Ex5	NAIP1/2 Ex13	GTF2H2 Ex5	GTF2H2-8	GTF2H2-11	Haplo-type
2	A	1	1	1	1	0	1	2	2	2	2	1	2	2	1	0	1
16	A	1	1	1	1	0	1	2	2	2	2	1	2	2	1	0	1
7	A	1	1	1	1	0	1	2	2	2	2	1	2	2	1	1	1
17	A	1	1	1	0	0	1	2	2	2	2	1	2	2	1	1	2
13	A	1	1	1	0	0	1	2	2	2	2	1	2	2	1	1	2
21	A	1	1	1	0	0	1	0	0	0	0	1	2	2	1	0	3
21	B	1	1	1	0	0	1	2	2	2	2	0	0	0	1	0	4
2	B	1	1	3	1	0	1	2	2	2	2	1	2	0	1	0	5
23	A	1	1	1	1	0	0	2	2	2	2	1	2	2	1	0	6
1	A	1	1	1	1	0	0	2	2	2	2	1	2	2	1	0	6
14	A	1	1	1	1	0	0	2	2	2	2	1	2	2	1	0	6
7	B	1	1	1	1	0	0	2	2	2	2	0	2	2	0	1	7
20	B	1	1	1	1	0	0	2	2	2	2	0	0	0	1	0	8
1	B	1	1	1	2	0	0	2	2	2	2	1	2	2	1	0	9
23	B	1	1	1	1	0	0	0	0	0	0	1	2	0	1	0	10
20	A	1	1	1	1	0	0	0	0	0	0	1	2	2	1	0	11
17	B	1	1	1	1	0	0	0	0	0	0	0	0	2	1	0	12
13	B	1	0	1	1	0	0	0	0	0	0	0	0	2	1	0	13
14	B	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0	14
16	B	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	15
24	A	1	1	0	0	0	0	0	0	0	0	1	2	2	1	0	16
24	B	1	1	2	2	0	0	2	2	2	2	1	2	2	1	1	17

Appendix I: Non-deletion haplotypes of carrier parents of families positive for SMA

Number of copies: 0 | Normal amount expected: 1 for single genes, 2 for combined genes | 3 | 4

Family Number	Family member	Chromosome	Chr ID	Copy Number for SMN Region (MLPA)														
				SMN 1-7	SMN 1-8	SMN 2-7	SMN 2-8	RAD17 -17	SERF 1B	SMN 1-1	SMN 1-4	SMN 1-6	SMN 1-8	NAIP -13	NAI P-5	GTF2H 2-11	GTF2H 2-8	GTF2H 2-5
1	Mother	Chr 1	A	1	2	1	0	1	1	2	2	2	2	2	0	C	G	2
	Father	Chr 1	C	2	3	1	0	1	1	2	2	2	2	2	1	C	G	2
2	Mother	Chr 1	A	2	1	1	1	1	1	2	2	2	2	2	2	C?	G/G	2
	Father	Chr 2	D	1	0	2	2	1	1	2	2	2	2	2	1	C	G	2
7	Mother	Chr 2	B	1	2	1	0	1	1	2	2	2	2	2	1	C	G	2
	Father	Chr 2	D	1	1	1	0	1	1	2	2	2	2	2	0	G	G	2
13	Mother	Chr 2	B	1	2	1	0	1	1	2	2	2	2	2	0	G	G	2
	Father	Chr 2	D	1	1	1	1	1	1	2	2	2	2	2	0	C	G	2
14	Mother	Chr 2	B	1	1	1	1	1	1	2	2	2	2	2	2	C	G	2
	Father	Chr 2	D	1	1	1	1	1	1	2	2	2	2	0	0	G/del?	G	0
16	Mother	Chr 2	B	1	0	1	1	1	1	2	2	2	2	2	1	C	G	2
	Father	Chr 2	D	1	1	1	0	1	1	2	2	2	2	2	0	C	A/del ?	0
17	Mother	Chr 2	B	2	2	0	0	1	1	2	2	2	2	2	1	C	G	2
20	Mother	Chr 1	A	1	1	0	0	1	1	2	2	2	2	2	1	C	G/G	2
	Father	Chr 1	C	1	1	1	1	1	1	2	2	2	2	2	1	G	G	2
21	Mother	Chr 2	B	1	1	0	0	1	1	2	2	2	2	2	2	C	G	2
	Father	Chr 2	D	1	1	1	1	1	1	0	2	2	2	2	1	G	G	2
23	Mother	Chr 1	A	2	2	0	0	1	1	2	2	2	2	2	1	C	G/G?	2
	Father	Chr 1	C	2	2	0	0	1	1	2	2	2	2	2	1	C	G	2
24	Mother	Chr 1	A	1	1	1	1	1	1	2	2	2	2	2	1	C	G	2

Appendix J: Comparison of *SMN1*, exon 7 copy number results between previously used diagnostic dosage assay and MLPA

Patient Code	Heterozygous *SMN1*, exon 7 deletion (1 copy)		Comments
	Dosage assay	MLPA	
SMA5	Positive	Negative (3 copies)	Discrepant
SMA7	Positive	Negative (3 copies)	Discrepant
SMA12	Negative	Negative (4 copies)	Confirmed
SMA80	Positive	Negative (2 copies)	Discrepant
SMA105	Negative	Negative (4 copies)	Confirmed
SMA110	Positive	Negative (2 copies)	Discrepant
SMA149	Negative	Negative (2 copies)	Confirmed

Appendix K: Cost analysis of the current homozygous deletion of *SMN1*, exon 7 diagnostic assay, qPCR and MLPA

Cost analysis: current homozygous deletion of *SMN1*, exon 7 diagnostic assay

REAGENTS

Name of reagent	Price	Volume	Number of reactions	Price per reaction (R)	Price for 10 pts + 5 ctrls (R)	Final price/patient (R)
Amplitaq kit (buffer, enzyme, Taq)	12740.35		6000 (0.5U/sample)	2.12	31.85	3.19
dNTPs	1825.24	20ml of mixed dNTP	8000	0.23	3.42	0.34
Primer-F	300	300 000ul	300 000	0.00	0.02	0.00
Primer-R	300		300 000	0.00	0.02	0.00
NLA-IV kit (buffer, enzyme)	3288.9	1ml	1000	3.29	49.33	4.93
Gel electrophoresis (agarose, TBE, ficoll, marker)				2.00	30.00	3.00
TOTAL COST OF REAGENTS/PATIENT				7.64	114.64	11.46
TOTAL COST OF LABOUR/PATIENT						77.77
TOTAL COST OF TEST/PATIENT						89.23

LABOUR

Measurement taken	Calculations	Comments
Time spent on 1 assay (4.5 hours over 2 days)	270min	Including setup, analysis and reporting
% of annual work hours (115200 min)	0.2%	Assuming 40 hours/week > 160h/month > 1920h/annum > 115200 min/annum
Number of tests done annually	153	
Number of batches (10 patients) done annually	15.3	Number of tests done annually/number of patients in a batch (10)
% time spent on SMA testing/annum	3.2%	Number of batches * % of annual work hours
Annual salary of a medical scientist (C3, NHLS)	371816.08	
Salary spent on a single SMA MLPA assay	11898.11456	
Salary spent/patient	77.77	

Cost analysis: qPCR assay

REAGENTS

Name of reagent	Price	Volume	Number of patients (reactions/patient)	Price per reaction (R)	Price for 10 pts + 6 ctrls (R)	Final price/patient (R)
TAQMAN UNIVERSAL MMIX II	1 049.00	5000 (9ul/pt)	555 (triplicate)	1.89	30.24	3.02
SMN1 probe	4306	50 ul (6nmole)	115 batches (in triplicate)	2.33	37.25	3.72
SMN2 probe	4306	50 ul (6nmole)	115 batches (in triplicate)	2.33	37.25	3.72
CFTR probe	4306	50 ul (6nmole)	115 batches (in triplicate)	2.33	37.25	3.72
SMN-Ex7-qPCR-F primer	168.95	1000ul (10 pmol/ul)	333.33	0.51	8.11	0.81
SMN-Ex7-qPCR-R primer	136.25	1340ul (10 pmol/ul)	446.67	0.31	4.89	0.49
CFTR-qPCR_F primer	114.45	1600ul (10 pmol/ul)	533.33	0.21	3.43	0.34
CFTR-qPCR_R primer	119.9	1330ul (10 pmol/ul)	443.33	0.27	4.33	0.43
TOTAL COST OF REAGENTS/PATIENT				10.16	162.59	16.26
TOTAL COST OF LABOUR/PATIENT						77.77
TOTAL COST OF TEST/PATIENT						94.03

LABOUR

Measurement taken	Calculatio	Comments
Time spent on 1 assay (3.5 hours in 1 day)	210min	Including setup, analysis and reporting
% of annual work hours (115200 min)	0.2%	Assuming 40 hours/week > 160h/month > 1920h/annum > 115200 min/annum
Number of tests done annually	153	
Number of batches (10 patients) done annually	15.3	Number of tests done annually/number of patients in a batch (10)
% time spent on SMA testing/annum	3.2%	Number of batches * % of annual work hours
Annual salary of a medical scientist (C3, NHLS)	371816.1	
Salary spent on a single SMA MLPA assay	11898.11	
Salary spent/patient	77.77	

Cost analysis: MLPA assay

REAGENTS

Name of reagent	Price	Volume	Number of reactions	Price per reaction (R)	Price for 10 pts + 6 ctrls (R)	Final price/patient (R)
P021-A2 probe mix	13 392.00		90	148.80	2380.80	238.08
Reagent mix (EK1)	3625.4		90	40.28	644.52	64.45
HIDI Formamide	380	25000ul	2777.78	0.14	2.19	0.22
GS-LIZ500 standard	4465		800	5.58	89.30	8.93
Sequencing buffer	996	25ml	75 runs	13.28	13.28	1.33
Sequencing Polymer	6257.12	7ml	388.9 runs	16.10	16.10	1.61
TOTAL COST OF REAGENTS/PATIENT				224.18	3146.18	314.62
TOTAL COST OF LABOUR/PATIENT						148.24
TOTAL COST OF TEST/PATIENT						462.86

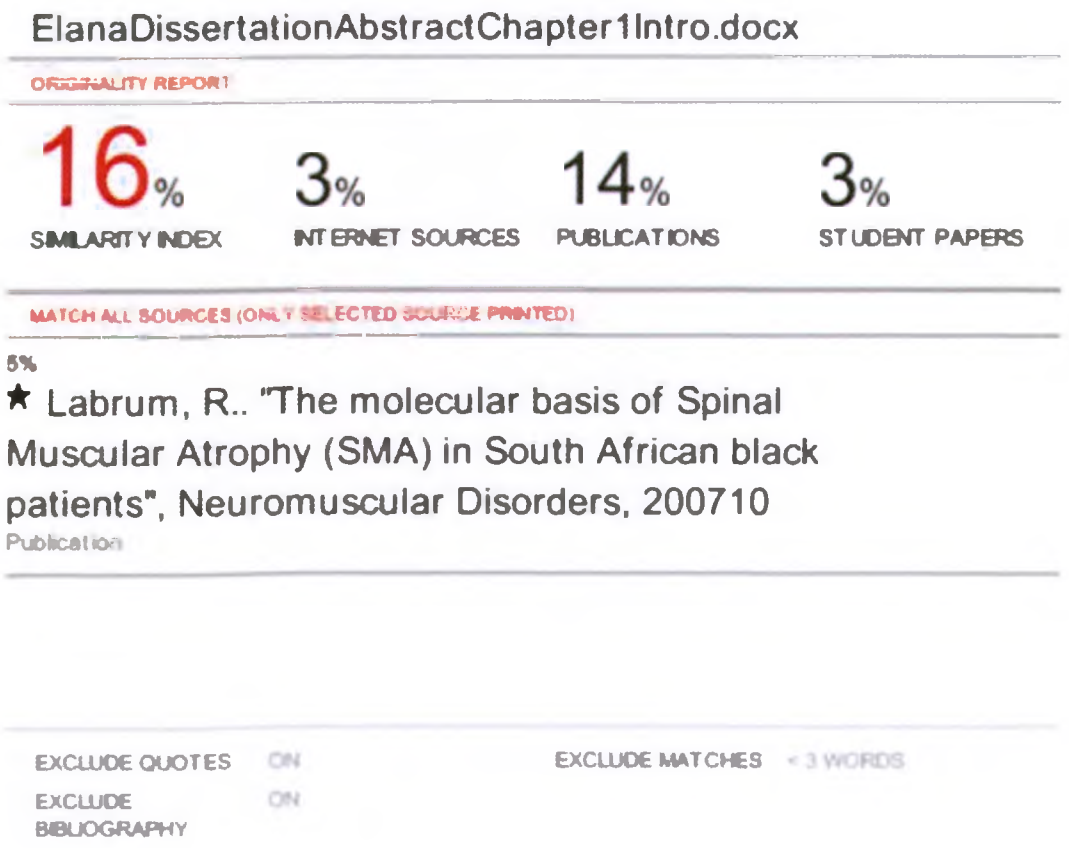
LABOUR

Measurement taken	Calculations	Comments
Time spent on 1 assay (7 hours over 2/3 days)	420min	Including setup, analysis and reporting
% of annual work hours (115200 min)	0.4%	Assuming 40 hours/week > 160h/month > 1920h/annum > 115200 min/annum
Number of tests done annually	153	
Number of batches (10 patients) done annually	15.3	Number of tests done annually/number of patients in a batch (10)
% time spent on SMA testing/annum	6.1%	Number of batches * % of annual work hours
Annual salary of a medical scientist (C3, NHLS)	371816.08	
Salary spent on a single SMA MLPA assay	22680.78	
Salary spent/patient	148.24	

Appendix L: Turnitin plagiarism reports

This dissertation was submitted to WITS Turnitin in four parts: Chapter 1: Abstract and Introduction, Chapter 2: Subjects and methods, Chapter 3: Results and Chapter 4 and 5: Discussion and Conclusion. These reports are quite lengthy and only the title pages were therefore included in the appendices. The full turn-it-in reports are available from the CD accompanying this dissertation.

Chapter 1: Abstract and Introduction



ElanaDissertationAbstractChapter1Intro.docx

ORIGINALITY REPORT

16%	3%	14%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

5%

★ Labrum, R.. "The molecular basis of Spinal Muscular Atrophy (SMA) in South African black patients", Neuromuscular Disorders, 200710

Publication

EXCLUDE QUOTES	ON	EXCLUDE MATCHES	< 3 WORDS
EXCLUDE BIBLIOGRAPHY	ON		

Chapter 2: Subjects and methods

ElanaDissertationChapter2SubjectsMethods.docx

ORIGINALITY REPORT

11%	3%	10%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Ivon Cuscó. "Implementation of SMA carrier testing in genetic laboratories: comparison of two methods for quantifying theSMN1 gene", Human Mutation, 12/2002 Publication	1%
2	Kubo, Yuji, Hisahide Nishio, and Kayoko Saito. "A new method for SMN1 and hybrid SMN gene analysis in spinal muscular atrophy using long-range PCR followed by sequencing", Journal of Human Genetics, 2015. Publication	1%
3	C. Rochette. "SMN gene duplication and the emergence of the SMN2 gene occurred in distinct hominids: SMN2 is unique to Homo sapiens", Human Genetics, 03/29/2001 Publication	1%
4	Krause, Amanda, Claire Mitchell, Fahmida Essop, Susan Tager, James Temlett, Giovanni Stevanin, Christopher Ross, Dobrila Rudnicki, and Russell Margolis. "Junctophilin 3 (JPH3) expansion mutations causing Huntington	<1%

Chapter 3: Results

ElanaDissertationChapter3Results.docx

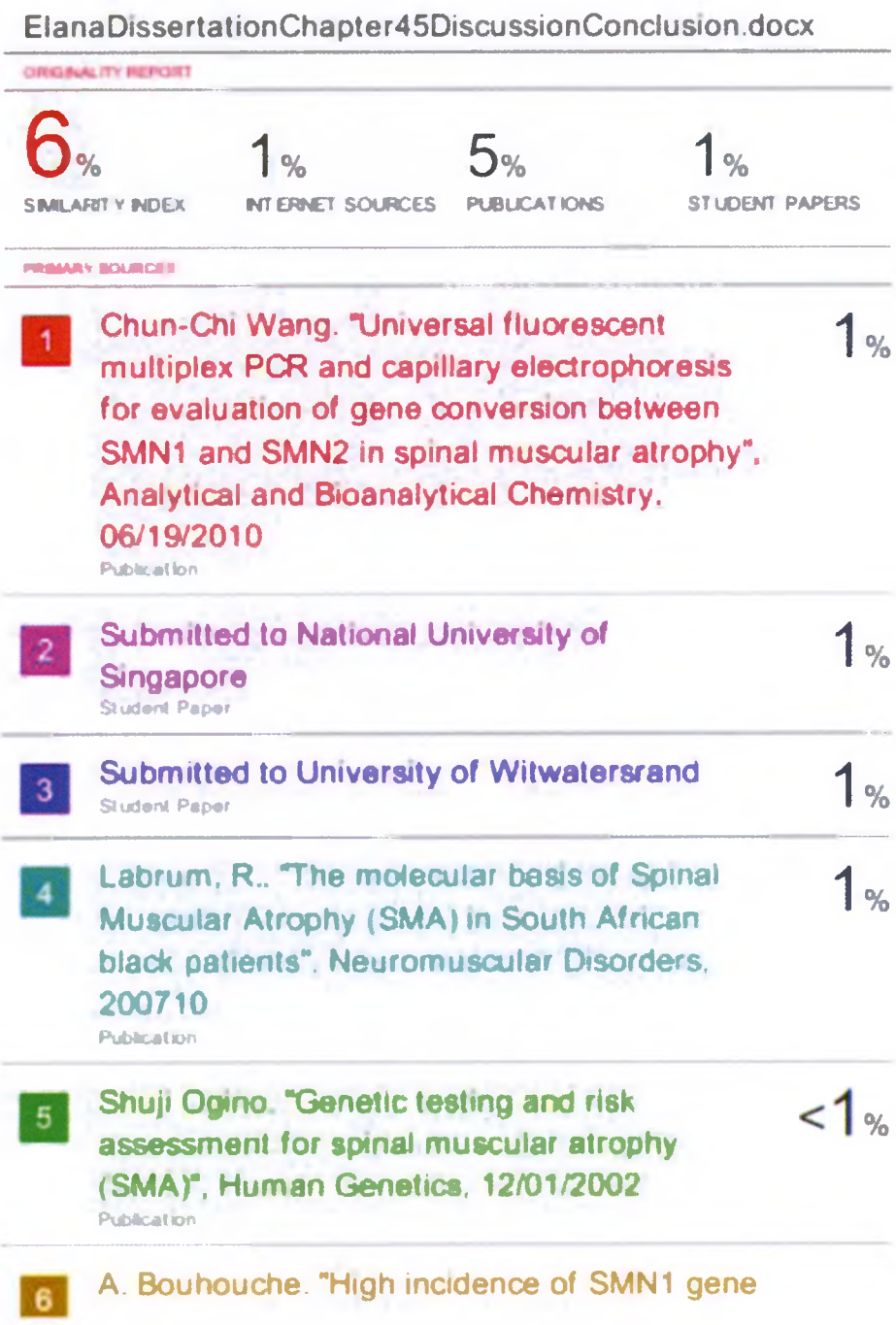
ORIGINALITY REPORT

6%	0%	5%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Abstracts—APASL 2013", Hepatology International, 2013. Publication	1%
2	Submitted to University of Witwatersrand Student Paper	<1%
3	Submitted to National University of Singapore Student Paper	<1%
4	Labrum, R.. "The molecular basis of Spinal Muscular Atrophy (SMA) in South African black patients", Neuromuscular Disorders, 200710 Publication	<1%
5	Kubo, Yuji, Hisahide Nishio, and Kayoko Saito. "A new method for SMN1 and hybrid SMN gene analysis in spinal muscular atrophy using long-range PCR followed by sequencing", Journal of Human Genetics, 2015. Publication	<1%

Chapter 4 and 5: Discussion and Conclusion



ElanaDissertationChapter45DiscussionConclusion.docx

ORIGINALITY REPORT

6%	1%	5%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Chun-Chi Wang. "Universal fluorescent multiplex PCR and capillary electrophoresis for evaluation of gene conversion between SMN1 and SMN2 in spinal muscular atrophy", Analytical and Bioanalytical Chemistry, 06/19/2010 Publication	1%
2	Submitted to National University of Singapore Student Paper	1%
3	Submitted to University of Witwatersrand Student Paper	1%
4	Labrum, R.. "The molecular basis of Spinal Muscular Atrophy (SMA) in South African black patients", Neuromuscular Disorders, 200710 Publication	1%
5	Shuji Ogino. "Genetic testing and risk assessment for spinal muscular atrophy (SMA)", Human Genetics, 12/01/2002 Publication	<1%
6	A. Bouhouche. "High incidence of SMN1 gene	

REFERENCES

Abbs, S., Tuffery-Giraud, S., Bakker, E., Ferlini, A., Sejersen, T., Mueller, C.R., 2010. Best practice guidelines on molecular diagnostics in Duchenne/Becker muscular dystrophies. Neuromuscul. Disord. NMD 20, 422–427. doi:10.1016/j.nmd.2010.04.005

Ankala, A., da Silva, C., Gualandi, F., Ferlini, A., Bean, L.J.H., Collins, C., Tanner, A.K., Hegde, M.R., 2015. A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield. Ann. Neurol. 77, 206–214. doi:10.1002/ana.24303

Arkblad, E., Tulinius, M., Kroksmark, A.-K., Henricsson, M., Darin, N., 2009. A population-based study of genotypic and phenotypic variability in children with spinal muscular atrophy. Acta Paediatr. Oslo Nor. 1992 98, 865–872. doi:10.1111/j.1651-2227.2008.01201.x

Banyer, J.L., Goldwurm, S., Cullen, L., van der Griend, B., Zournazi, A., Smit, D.J., Powell, L.W., Jazwinska, E.C., 1998. The spinal muscular atrophy gene region at 5q13.1 has a paralogous chromosomal region at 6p21.3. Mamm. Genome Off. J. Int. Mamm. Genome Soc. 9, 235–239

Barker, K., 2004. At the Bench: A Laboratory Navigator, Updated Edition, 2nd edition. ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Brkušanin, M., Kosać, A., Jovanović, V., Pešović, J., Brajušković, G., Dimitrijević, N., Todorović, S., Romac, S., Milić Rašić, V., Savić-Pavićević, D., 2015. Joint effect of the SMN2 and SERF1A genes on childhood-onset types of spinal muscular atrophy in Serbian patients. J. Hum. Genet. 60, 723–728. doi:10.1038/jhg.2015.104

Bühler, D., Raker, V., Lührmann, R., Fischer, U., 1999. Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy. Hum. Mol. Genet. 8, 2351–2357.

Burghes, A.H.M., Beattie, C.E., 2009. Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? Nat. Rev. Neurosci. 10, 597–609.

Bürglen, L., Lefebvre, S., Clermont, O., Burlet, P., Viollet, L., Cruaud, C., Munnich, A., Melki, J., 1996. Structure and organization of the human survival motor neurone (SMN) gene. Genomics 32, 479–482. doi:10.1006/geno.1996.0147

Butchbach, M., Burghes, A., 2004. Perspectives on models of spinal muscular atrophy for drug discovery. Drug Discov Today Dis Models. 1:22, 151-156

Campbell, L., Potter, A., Ignatius, J., Dubowitz, V., Davies, K., 1997. Genomic variation and gene conversion in spinal muscular atrophy: implications for disease process and clinical phenotype. Am. J. Hum. Genet. 61, 40–50. doi:10.1086/513886

Carter, T.A., Bönnemann, C.G., Wang, C.H., Obici, S., Parano, E., De Fatima Bonaldo, M., Ross, B.M., Penchaszadeh, G.K., Mackenzie, A., Soares, M.B., Kunkel, L.M., Gilliam, T.C., 1997. A multicopy transcription-repair gene, BTF2p44, maps to the SMA region and demonstrates SMA associated deletions. Hum. Mol. Genet. 6, 229–236.

Chen, J.-M., Férec, C., Cooper, D.N., 2010. Gene Conversion in Human Genetic Disease. Genes 1, 550–563. doi:10.3390/genes1030550

Chimusa, E.R., Meintjies, A., Tchanga, M., Mulder, N., Seoighe, C., Seioghe, C., Soodyall, H., Ramesar, R., 2015. A genomic portrait of haplotype diversity and signatures of selection in indigenous southern African populations. PLoS Genet. 11, e1005052

Choudhury, A., Hazelhurst, S., Meintjes, A., Achinike-Oduaran, O., Aron, S., Gamieldien, J., Jalali Sefid Dashti, M., Mulder, N., Tiffin, N., Ramsay, M., 2014. Population-specific common SNPs reflect demographic histories and highlight regions of genomic plasticity with functional relevance. BMC Genomics 15, 437. doi:10.1186/1471-2164-15-437

Conrad, D.F., Jakobsson, M., Coop, G., Wen, X., Wall, J.D., Rosenberg, N.A., Pritchard, J.K., 2006. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat. Genet. 38, 1251–1260. doi:10.1038/ng1911

Corcia, P., Ingre, C., Blasco, H., Press, R., Praline, J., Antar, C., Veyrat-Durebex, C., Guettard, Y.-O., Camu, W., Andersen, P.M., Vourc'h, P., Andres, C.R., 2012. Homozygous SMN2 deletion is a protective factor in the Swedish ALS population. Eur. J. Hum. Genet. EJHG 20, 588–591. doi:10.1038/ejhg.2011.255

Crawford, T.O., Paushkin, S.V., Kobayashi, D.T., Forrest, S.J., Joyce, C.L., Finkel, R.S., Kaufmann, P., Swoboda, K.J., Tiziano, D., Lomastro, R., Li, R.H., Trachtenberg, F.L., Plasterer, T., Chen, K.S., Pilot Study of Biomarkers for Spinal Muscular Atrophy Trial Group, 2012. Evaluation of SMN protein, transcript, and copy number in the biomarkers for spinal muscular atrophy (BforSMA) clinical study. PloS One 7, e33572. doi:10.1371/journal.pone.0033572

Cure SMA (previously known as "Families of SMA"), accessed 1 August 2016, <http://www.curesma.org>.

de Souza Godinho, F.M., Bock, H., Gheno, T.C., Saraiva-Pereira, M.L., 2012. Molecular Analysis of Spinal Muscular Atrophy: A genotyping protocol based on TaqMan(®) real-time PCR. Genet. Mol. Biol. 35, 955–959.

den Dunnen, J.T., Dalgleish, R., Maglott, D.R., Hart, R.K., Greenblatt, M.S., McGowan-Jordan, J., Roux, A.F., Smith, T., Antonarakis, S.E., Taschner, P.E., 2016, HGVS recommendations for the description of sequence variants: 2016 update. Hum.Mutat. 37, 564-569, accessed 1 August 2016, <http://varnomen.hgvs.org>.

Doi, K., Monjo, T., Hoang, P.H., Yoshimura, J., Yurino, H., Mitsui, J., Ishiura, H., Takahashi, Y., Ichikawa, Y., Goto, J., Tsuji, S., Morishita, S., 2014. Rapid detection of expanded short tandem repeats in personal genomics using hybrid sequencing. Bioinforma. Oxf. Engl. 30, 815–822. doi:10.1093/bioinformatics/btt647

Duale, N., Brunborg, G., Rønningen, K.S., Briese, T., Aarem, J., Aas, K.K., Magnus, P., Stoltenberg, C., Susser, E., Lipkin, W.I., 2012. Human blood RNA stabilization in samples collected and transported for a large biobank. BMC Res. Notes 5, 510. doi:10.1186/1756-0500-5-510

Ensembl Genome Browser, Ensembl accession codes: ENSG00000172062, ENSG00000205571, ENSG00000001626, ENSG00000145736, accessed 11 March 2016, <http://www.ensembl.org>.

Falsone, S.F., Meyer, N.H., Schrank, E., Leitinger, G., Pham, C.L.L., Fodero-Tavoletti, M.T., Holmberg, M., Dulle, M., Scicluna, B., Gesslbauer, B., Rückert, H., Wagner, G.E., Merle, D.A., Nollen, E.A., Kungl, A.J., Hill, A.F., Cappai, R., Zangger, K., 2012, SERF Protein is a Direct Modifier of Amyloid Fiber Assembly. Cell Rep. 2, 358-371.

Feldkötter, M., Schwarzer, V., Wirth, R., Wienker, T.F., Wirth, B., 2002. Quantitative analyses of SMN1 and SMN2 based on real-time lightCycler PCR: fast and highly reliable carrier testing and prediction of severity of spinal muscular atrophy. Am. J. Hum. Genet. 70, 358–368. doi:10.1086/338627

Feng, Y., Ge, X., Meng, L., Scull, J., Li, J., Tian, X., Zhang, T., Jin, W., Cheng, H., Wang, X., Tokita, M., Liu, P., Mei, H., Wang, Y., Li, F., Schmitt, E.S., Zhang, W.V., Muzny, D., Wen, S., Chen, Z., Yang, Y., Beaudet, A.L., Liu, X., Eng, C.M., Xia, F., Wong, L.-J., Zhang, J., 2017. The next generation of population-based spinal muscular atrophy carrier screening: comprehensive pan-ethnic SMN1 copy-number and sequence variant analysis by massively parallel sequencing. Genet. Med. Off. J. Am. Coll. Med. Genet. doi:10.1038/gim.2016.215

Finkel, R.S., Crawford, T.O., Swoboda, K.J., Kaufmann, P., Juhasz, P., Li, X., Guo, Y., Li, R.H., Trachtenberg, F., Forrest, S.J., Kobayashi, D.T., Chen, K.S., Joyce, C.L., Plasterer, T., Pilot Study of Biomarkers for Spinal Muscular Atrophy Trial Group, 2012. Candidate proteins, metabolites and transcripts in the Biomarkers for Spinal Muscular Atrophy (BforSMA) clinical study. PloS One 7, e35462. doi:10.1371/journal.pone.0035462

Francis, M.J., Nesbit, M.A., Theodosiou, A.M., Rodrigues, N.R., Campbell, L., Christodoulou, Z., Qureshi, S.J., Porteous, D.J., Brookes, A.J., Davies, K.E., 1995. Mapping of retrotransposon sequences in the unstable region surrounding the spinal muscular atrophy locus in 5q13. Genomics 27, 366–369. doi:10.1006/geno.1995.1059

Fried, K., Emery, A.E., 1971. Spinal muscular atrophy type II. A separate genetic and clinical entity from type I (Werdnig-Hoffmann disease) and type 3 (Kugelberg-Welander disease). Clin. Genet. 2, 203–209.

Ghetti, B., Amati, A., Turra, M.V., Pacini, A., Del Vecchio, M., Guazzi, G.C., 1971. Werdnig-Hoffmann-Wohlfart-Kugelberg-Welander disease. Nosological unity and clinical variability in intrafamilial cases. Acta Genet. Med. Gemellol. (Roma) 20, 43–58.

Glascock, J., Lenz, M., Hobby, K., Jarecki, J., 2017. Cure SMA and our patient community celebrate the first approved drug for SMA. Gene Ther. doi:10.1038/gt.2017.39

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., Ritchie, G.R.S., Xue, Y., Asimit, J., Nsubuga, R.N., Young, E.H.,

Pomilla, C., Kivinen, K., Rockett, K., Kamali, A., Doumatey, A.P., Asiki, G., Seeley, J., Sisay-Joof, F., Jallow, M., Tollman, S., Mekonnen, E., Ekong, R., Oljira, T., Bradman, N., Bojang, K., Ramsay, M., Adeyemo, A., Bekele, E., Motala, A., Norris, S.A., Pirie, F., Kaleebu, P., Kwiatkowski, D., Tyler-Smith, C., Rotimi, C., Zeggini, E., Sandhu, M.S., 2015. The African Genome Variation Project shapes medical genetics in Africa. Nature 517, 327–332. doi:10.1038/nature13997

Hahnen, E., Schönling, J., Rudnik-Schöneborn, S., Zerres, K., Wirth, B., 1996. Hybrid survival motor neuron genes in patients with autosomal recessive spinal muscular atrophy: new insights into molecular mechanisms responsible for the disease. Am. J. Hum. Genet. 59, 1057–1065.

Hasanzad, M., Golkar, Z., Kariminejad, R., Hadavi, V., Almadani, N., Afroozan, F., Salahshurifar, I., Shafeghati, Y., Kahrizi, K., Najmabadi, H., 2009. Deletions in the survival motor neuron gene in Iranian patients with spinal muscular atrophy. Ann. Acad. Med. Singapore 38, 139–141.

He, J., Zhang, Q.-J., Lin, Q.-F., Chen, Y.-F., Lin, X.-Z., Lin, M.-T., Murong, S.-X., Wang, N., Chen, W.-J., 2013. Molecular analysis of SMN1, SMN2, NAIP, GTF2H2, and H4F5 genes in 157 Chinese patients with spinal muscular atrophy. Gene 518, 325–329. doi:10.1016/j.gene.2012.12.109

Hendrickson, B.C., Donohoe, C., Akmaev, V.R., Sugarman, E.A., Labrousse, P., Boguslavskiy, L., Flynn, K., Rohlfs, E.M., Walker, A., Allitto, B., Sears, C., Scholl, T., 2009. Differences in SMN1 allele frequencies among ethnic groups within North America. J. Med. Genet. 46, 641–644. doi:10.1136/jmg.2009.066969

Human Genome Mutation Database (HGMD), accessed 7 October 2016, <http://www.hgmd.cf.ac.uk/ac/gene.php?gene=SMN1>.

Hsieh-Li, H.M., Chang, J.G., Jong, Y.J., Wu, M.H., Wang, N.M., Tsai, C.H., Li, H., 2000. A mouse model for spinal muscular atrophy. Nat. Genet. 24, 66–70.

Illumina Trusight Sequening panels, accessed 18 February 2017, <https://www.illumina.com/products/trusight-panels>

Jedrzejowska, M., Borkowska, J., Zimowski, J., Kostera-Pruszczyk, A., Milewski, M., Jurek, M., Sielska, D., Kostyk, E., Nyka, W., Zaremba, J., Hausmanowa-Petrusewicz, I., 2008. Unaffected patients with a homozygous absence of the SMN1 gene. Eur. J. Hum. Genet. EJHG 16, 930–934. doi:10.1038/ejhg.2008.41

Jedrzejowska, M., Milewski, M., Zimowski, J., Borkowska, J., Kostera-Pruszczyk, A., Sielska, D., Jurek, M., Hausmanowa-Petrusewicz, I., 2009. Phenotype modifiers of spinal muscular atrophy: the number of SMN2 gene copies, deletion in the NAIP gene and probably gender influence the course of the disease. Acta Biochim. Pol. 56, 103–108.

Kasianowicz, J.J., Brandin, E., Branton, D., Deamer, D.W., 1996. Characterization of individual polynucleotide molecules using a membrane channel. Proc. Natl. Acad. Sci. U. S. A. 93, 13770–13773.

Kernohan, K.D., Frésard, L., Zappala, Z., Hartley, T., Smith, K.S., Wagner, J., Xu, H., McBride, A., Bourque, P.R., Consortium, C.R.C., Bennett, S.A.L., Dyment, D.A., Boycott, K.M., Montgomery, S.B., Warman Chardon, J., 2017. Whole-transcriptome sequencing in blood provides a diagnosis of spinal muscular atrophy with progressive myoclonic epilepsy. Hum. Mutat. 38, 611–614. doi:10.1002/humu.23211

Khan, H., Vorster, E., Krause, A., Essop, F., 2017. Investigating the molecular aetiology of silent carriers of spinal muscular atrophy in the black South African population (Poster). Honours project presentations. Johannesburg, South Africa, 13 October 2017. NHLS & WITS.

Kugelberg, E., Welander, L., 1956. Heredofamilial juvenile muscular atrophy simulating muscular dystrophy. AMA Arch. Neurol. Psychiatry 75, 500–509.

Labrum, R., Rodda, J., Krause, A., 2007. The molecular basis of spinal muscular atrophy (SMA) in South African black patients. Neuromuscul. Disord. NMD 17, 684–692. doi:10.1016/j.nmd.2007.05.005

Lee, T.-M., Kim, S.-W., Lee, K.-S., Jin, H.-S., Koo, S.K., Jo, I., Kang, S., Jung, S.-C., 2004. Quantitative analysis of SMN1 gene and estimation of SMN1 deletion carrier frequency in Korean population based on real-time PCR. J. Korean Med. Sci. 19, 870–873. doi:10.3346/jkms.2004.19.6.870

Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., 1995. Identification and characterization of a spinal muscular atrophy-determining gene. Cell 80, 155–165.

Lorson, C.L., Hahnen, E., Androphy, E.J., Wirth, B., 1999. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc. Natl. Acad. Sci. U. S. A. 96, 6307–6311.

Luo, M., Liu, L., Peter, I., Zhu, J., Scott, S.A., Zhao, G., Eversley, C., Kornreich, R., Desnick, R.J., Edelmann, L., 2014. An Ashkenazi Jewish SMN1 haplotype specific to duplication alleles improves pan-ethnic carrier screening for spinal muscular atrophy. Genet. Med. Off. J. Am. Coll. Med. Genet. 16, 149–156.

Lupski, J.R., Stankiewicz, P., 2005. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet. 1, e49. doi:10.1371/journal.pgen.0010049

MacLeod, M.J., Taylor, J.E., Lunt, P.W., Mathew, C.G., Robb, S.A., 1999. Prenatal onset spinal muscular atrophy. Eur. J. Paediatr. Neurol. EJPN Off. J. Eur. Paediatr. Neurol. Soc. 3, 65–72. doi:10.1053/ejpn.1999.0184

Maranda, B., Fan, L., Soucy, J.-F., Simard, L., Mitchell, G.A., 2012. Spinal muscular atrophy: clinical validation of a single-tube multiplex real time PCR assay for determination of SMN1 and SMN2 copy numbers. Clin. Biochem. 45, 88–91. doi:10.1016/j.clinbiochem.2011.10.019

Markowitz, J.A., Singh, P., Darras, B.T., 2012. Spinal muscular atrophy: a clinical and research update. Pediatr. Neurol. 46, 1–12. doi:10.1016/j.pediatrneurol.2011.09.001

McAndrew, P.E., Parsons, D.W., Simard, L.R., Rochette, C., Ray, P.N., Mendell, J.R., Prior, T.W., Burghes, A.H., 1997. Identification of proximal spinal muscular atrophy carriers and patients by analysis of SMNT and SMNC gene copy number. Am. J. Hum. Genet. 60, 1411–1422. doi:10.1086/515465

Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16, 1215.

Miskovic, M., Lalic, T., Radivojevic, D., Cirkovic, S., Vlahovic, G., Zamurovic, D., Guc-Scekic, M., 2011. Lower incidence of deletions in the survival of motor neuron gene and the neuronal apoptosis inhibitory protein gene in children with spinal muscular atrophy from Serbia. Tohoku J. Exp. Med. 225, 153–159.

Monani, U.R., Lorson, C.L., Parsons, D.W., Prior, T.W., Androphy, E.J., Burghes, A.H., McPherson, J.D., 1999. A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. Hum. Mol. Genet. 8, 1177–1183.

Moosa, A., Dawood, A., 1990. Spinal muscular atrophy in African children. Neuropediatrics 21, 27–31. doi:10.1055/s-2008-1071453

Mrad, R., Dorboz, I., Ben Jemaa, L., Maazoul, F., Trabelsi, M., Chaabouni, M., Mlaiki, B., Miladi, N., Hentati, F., Chaabouni, H., 2006. Molecular analysis of the SMN1 and NAIP genes in 60 Tunisian spinal muscular atrophy patients. Tunis. Médicale 84, 465–469.

MRC Holland, Product Description P021-A2-A316; Probe sequences P021 A2-0511, A2-0613, A2-0415, A2-0316, accessed 10 October 2016, <https://www.mlpa.com/WebForms>.

National Center for Biotechnology information (NCBI), Gene ID: 6606, 6607, 2966, 4671, 5884, 728492, accessed 1 August 2016, <http://www.ncbi.nlm.nih.gov>.

Ndiaye, O., Sall, G., Sylla, A., Diouf, S., Diagne, I., Kuakuvi, N., 2002. [Progressive spinal amyotrophy type I or Werdnig-Hoffman disease. Apropos of 5 cases in Dakar (Senegal)]. Bull. Société Pathol. Exot. 1990 95, 81–82.

Nigro, V., Savarese, M., 2016. Next-generation sequencing approaches for the diagnosis of skeletal muscle disorders. Curr. Opin. Neurol. 29, 621–627. doi:10.1097/WCO.0000000000000371

Noguchi, Y., Onishi, A., Nakamachi, Y., Hayashi, N., Harahap, N.I.F., Rochmah, M.A., Shima, A., Yanagisawa, S., Morisada, N., Nakagawa, T., Iijima, K., Kasagi, S., Saegusa, J., Kawano, S., Shinohara, M., Tairaku, S., Saito, T., Kubo, Y., Saito, K., Nishio, H., 2016. Telomeric Region of the Spinal Muscular Atrophy Locus Is Susceptible to Structural Variations. Pediatr. Neurol. 58, 83–89.

Online Mendelian Inheritance in Man (OMIM), OMIM accession numbers: 253300, 253550, 253400 & 271150, 601748, 600355, 603139, accessed 1 August 2016, <http://www.omim.org>.

Pearn, J.H., Carter, C.O., Wilson, J., 1973. The genetic identity of acute infantile spinal muscular atrophy. Brain J. Neurol. 96, 463–470.

Pearn, J.H., Hudgson, P., Walton, J.N., 1978. A clinical and genetic study of spinal muscular atrophy of adult onset: the autosomal recessive form as a discrete disease entity. Brain J. Neurol. 101, 591–606.

Peeters, K., Chamova, T., Jordanova, A., 2014. Clinical and genetic diversity of SMN1-negative proximal spinal muscular atrophies. Brain J. Neurol. 137, 2879–2896. doi:10.1093/brain/awu169

Pelleboer, R.A., Maaswinkel-Mooy, P.D., Gelderen, H.H., 1989. Werdnig-Hoffmann disease in a Nigerian infant. Trop. Geogr. Med. 41, 282–284.

Pepper, M.S., 2011. Launch of the Southern African Human Genome Programme. South Afr. Med. J. Suid-Afr. Tydskr. Vir Geneeskd. 101, 287–288.

Pfaffl, M.W., 2004, 'Quantification strategies in real-time PCR', in S.A. Bustin (ed.), *A-Z of quantitative PCR*, pp. 87 – 112, La Jolla, CA, USA.

Pickrell, J.K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., Reich, D., 2014. Ancient west Eurasian ancestry in southern and eastern Africa. Proc. Natl. Acad. Sci. U. S. A. 111, 2632–2637. doi:10.1073/pnas.1313787111

Prior, T.W., Krainer, A.R., Hua, Y., Swoboda, K.J., Snyder, P.C., Bridgeman, S.J., Burghes, A.H.M., Kissel, J.T., 2009. A positive modifier of spinal muscular atrophy in the SMN2 gene. Am. J. Hum. Genet. 85, 408–413. doi:10.1016/j.ajhg.2009.08.002

Prior, T.W., Professional Practice and Guidelines Committee, 2008. Carrier screening for spinal muscular atrophy. Genet. Med. Off. J. Am. Coll. Med. Genet. 10, 840–842. doi:10.1097/GIM.0b013e318188d069

Prior, T.W., Swoboda, K.J., Scott, H.D., Hejmanowski, A.Q., 2004. Homozygous SMN1 deletions in unaffected family members and modification of the phenotype by SMN2. Am. J. Med. Genet. A. 130A, 307–310. doi:10.1002/ajmg.a.30251

Qu, Y., Ge, X., Bai, J., Wang, L., Cao, Y., Lu, Y., Jin, Y., Wang, H., Song, F., 2015. Association of copy numbers of survival motor neuron gene 2 and neuronal apoptosis inhibitory protein gene with the natural history in a Chinese spinal muscular atrophy cohort. J. Child Neurol. 30, 429–436. doi:10.1177/0883073814553271

Real Statistics Using Excel, accessed 1 August 2016, < http://www.real-statistics.com>.

Roy, N., Mahadevan, M.S., McLean, M., Shutler, G., Yaraghi, Z., Farahani, R., Baird, S., Besner-Johnston, A., Lefebvre, C., Kang, X., 1995. The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. Cell 80, 167–178.

Samilchuk, E., D'Souza, B., Bastaki, L., al-Awadi, S., 1996. Deletion analysis of the SMN and NAIP genes in Kuwaiti patients with spinal muscular atrophy. Hum. Genet. 98, 524–527.

Sangaré, M., Hendrickson, B., Sango, H.A., Chen, K., Nofziger, J., Amara, A., Dutra, A., Schindler, A.B., Guindo, A., Traoré, M., Harmison, G., Pak, E., Yaro, F.N., Bricceno, K., Grunseich, C., Chen, G., Boehm, M., Zukosky, K., Bocoum, N., Meilleur, K.G., Daou, F., Bagayogo, K., Coulibaly, Y.I., Diakité, M., Fay, M.P., Lee, H.-S., Saad, A., Gribaa, M., Singleton, A.B., Maiga, Y., Auh, S., Landouré, G., Fairhurst, R.M., Burnett, B.G., Scholl, T., Fischbeck, K.H., 2014. Genetics of low spinal muscular atrophy carrier frequency in sub-Saharan Africa. Ann. Neurol. 75, 525–532. doi:10.1002/ana.24114

Scharf, J.M., Endrizzi, M.G., Wetter, A., Huang, S., Thompson, T.G., Zerres, K., Dietrich, W.F., Wirth, B., Kunkel, L.M., 1998. Identification of a candidate modifying gene for spinal muscular atrophy by comparative genomics. Nat. Genet. 20, 83–86. doi:10.1038/1753

Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., Pals, G., 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. 30, e57.

Schwartz, M., Sørensen, N., Hansen, F.J., Hertz, J.M., Nørby, S., Tranebjaerg, L., Skovby, F., 1997. Quantification, by solid-phase minisequencing, of the telomeric and centromeric copies of the survival motor neuron gene in families with spinal muscular atrophy. Hum. Mol. Genet. 6, 99–104.

Selig, S., Bruno, S., Scharf, J.M., Wang, C.H., Vitale, E., Gilliam, T.C., Kunkel, L.M., 1995. Expressed cadherin pseudogenes are localized to the critical region of the spinal muscular atrophy gene. Proc. Natl. Acad. Sci. U. S. A. 92, 3702–3706.

Shawky, R.M., Abd el-Aleem, K., Rifaat, M.M., Moustafa, A., 2001. Molecular diagnosis of spinal muscular atrophy in Egyptians. East. Mediterr. Health J. Rev. Santé Méditerranée Orient. Al-Majallah Al-Ṣiḥḥīyah Li-Sharq Al-Mutawassiṭ 7, 229–237.

Singh, N.N., Lee, B.M., Singh, R.N., 2015. Splicing regulation in spinal muscular atrophy by an RNA structure formed by long-distance interactions. Ann. N. Y. Acad. Sci. 1341, 176–187.

Stevens, G., Yawitch, T., Rodda, J., Verhaart, S., Krause, A., 1999. Different molecular basis for spinal muscular atrophy in South African black patients. Am. J. Med. Genet. 86, 420–426.

Stratigopoulos, G., Lanzano, P., Deng, L., Guo, J., Kaufmann, P., Darras, B., Finkel, R., Tawil, R., McDermott, M.P., Martens, W., Devivo, D.C., Chung, W.K., 2010. Association of plastin 3 expression with disease severity in spinal muscular atrophy only in postpubertal females. Arch. Neurol. 67, 1252–1256. doi:10.1001/archneurol.2010.239

Sugarman, E.A., Nagan, N., Zhu, H., Akmaev, V.R., Zhou, Z., Rohlfs, E.M., Flynn, K., Hendrickson, B.C., Scholl, T., Sirko-Osadsa, D.A., Allitto, B.A., 2012. Pan-ethnic carrier screening and prenatal diagnosis for spinal muscular atrophy: clinical laboratory analysis of >72,400 specimens. Eur. J. Hum. Genet. EJHG 20, 27–32. doi:10.1038/ejhg.2011.134

Talbot, K., Rodrigues, N.R., Ignatius, J., Muntoni, F., Davies, K.E., 1997. Gene conversion at the SMN locus in autosomal recessive spinal muscular atrophy does not predict a mild phenotype. Neuromuscul. Disord. NMD 7, 198–201.

Tazir, M., Geronimi, C., 1990. Chronic childhood spinal muscular atrophies in Algeria. A genetic study. J. Neurol. Sci. 96, 89–101.

The American Congress on Obstetricians and Gynecologists, ACOG Committee Opinion 432: Spinal Muscular Atrophy, accessed 7 February 2017, <http://www.acog.org>

Theodorou, L., Nicolaou, P., Koutsou, P., Georghiou, A., Anastasiadou, V., Tanteles, G., Kyriakides, T., Zamba-Papanicolaou, E., Christodoulou, K., 2015. Genetic findings of Cypriot spinal muscular atrophy patients. Neurol. Sci. Off. J. Ital. Neurol. Soc. Ital. Soc. Clin. Neurophysiol. 36, 1829–1834. doi:10.1007/s10072-015-2263-5

Thermo Fisher, Tempus RNA tubes and Tempus spin RNA isolation kit protocol, accessed 1 August 2016, <https://www.thermofisher.com/order/catalog/product/4380204>

Thomas, N.H., Dubowitz, V., 1994. The natural history of type I (severe) spinal muscular atrophy. Neuromuscul. Disord. NMD 4, 497–502.

Tishkoff, S.A., Kidd, K.K., 2004. Implications of biogeography of human populations for "race" and medicine. Nat. Genet. 36, S21-27. doi:10.1038/ng1438

Tishkoff, S.A., Williams, S.M., 2002. Genetic analysis of African populations: human evolution and complex disease. Nat. Rev. Genet. 3, 611–621. doi:10.1038/nrg865

Tiziano, F.D., Pinto, A.M., Fiori, S., Lomastro, R., Messina, S., Bruno, C., Pini, A., Pane, M., D'Amico, A., Ghezzo, A., Bertini, E., Mercuri, E., Neri, G., Brahe, C., 2010. SMN transcript levels in leukocytes of SMA patients determined by absolute real-time PCR. Eur. J. Hum. Genet. EJHG 18, 52–58. doi:10.1038/ejhg.2009.116

van der Steege, G., Grootscholten, P.M., Cobben, J.M., Zappata, S., Scheffer, H., den Dunnen, J.T., van Ommen, G.J., Brahe, C., Buys, C.H., 1996. Apparent gene conversions involving the SMN gene in the region of the spinal muscular atrophy locus on chromosome 5. Am. J. Hum. Genet. 59, 834–838.

Vorster, E., Essop, F., Krause, A., 2011. MLPA as a diagnostic tool to detect spinal muscular atrophy (SMA) in South African populations (Poster). In: SASHG (South African Society of Human Genetics), *2010 Joint International Conference of the African and Southern African Societies of Human Genetics (SASHG)*. Cape Town, South Africa, 6-9 March 2011. NHLS & WITS: Johannesburg.

Wang, C.-C., Jong, Y.-J., Chang, J.-G., Chen, Y.-L., Wu, S.-M., 2010. Universal fluorescent multiplex PCR and capillary electrophoresis for evaluation of gene conversion between SMN1 and SMN2 in spinal muscular atrophy. Anal. Bioanal. Chem. 397, 2375–2383. Watihayati, M.S., Fatemeh, H., Marini, M., Atif, A.B., Zahiruddin, W.M., Sasongko, T.H., Tang, T.H., Zabidi-Hussin, Z. a. M.H., Nishio, H., Zilfalil, B.A., 2009. Combination of SMN2 copy number and NAIP deletion predicts disease severity in spinal muscular atrophy. Brain Dev. 31, 42–45.

Watihayati, M.S., Fatemeh, H., Marini, M., Atif, A.B., Zahiruddin, W.M., Sasongko, T.H., Tang, T.H., Zabidi-Hussin, Z. a. M.H., Nishio, H., Zilfalil, B.A., 2009. Combination of SMN2 copy number and NAIP deletion predicts disease severity in spinal muscular atrophy. Brain Dev. 31, 42–45. doi:10.1016/j.braindev.2008.08.012

Wilmshurst, J.M., Reynolds, L., Van Toorn, R., Leisegang, F., Henderson, H.E., 2002. Spinal muscular atrophy in black South Africans: concordance with the universal SMN1 genotype. Clin. Genet. 62, 165–168.

Wirth, B., 2000. An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA). Hum. Mutat. 15, 228–237. doi:10.1002/(SICI)1098-1004(200003)15:3<228::AID-HUMU3>3.0.CO;2-9

Wirth, B., Herz, M., Wetter, A., Moskau, S., Hahnen, E., Rudnik-Schöneborn, S., Wienker, T., Zerres, K., 1999. Quantitative analysis of survival motor neuron copies: identification of subtle SMN1 mutations in patients with spinal muscular atrophy, genotype-phenotype correlation, and implications for genetic counseling. Am. J. Hum. Genet. 64, 1340–1356. doi:10.1086/302369

Wirth, B., Schmidt, T., Hahnen, E., Rudnik-Schöneborn, S., Krawczak, M., Müller-Myhsok, B., Schönling, J., Zerres, K., 1997. De Novo Rearrangements Found in 2% of Index Patients with Spinal Muscular Atrophy: Mutational Mechanisms, Parental Origin, Mutation Rate, and Implications for Genetic Counseling. Am. J. Hum. Genet. 61, 1102–1111. doi:10.1086/301608

Yang, C.-W., Chen, C.-L., Chou, W.-C., Lin, H.-C., Jong, Y.-J., Tsai, L.-K., Chuang, C.-Y., 2016. An Integrative Transcriptomic Analysis for Identifying Novel Target Genes Corresponding to Severity Spectrum in Spinal Muscular Atrophy. PloS One 11, e0157426. doi:10.1371/journal.pone.0157426

Yu-Jin, Q., Juan, D., Er-zhen, L., Jin-li, B., Yu-wei, J., Hong, W., Fang, S., 2012. Subtle mutations in the SMN1 gene in Chinese patients with SMA: p.Arg288Met mutation causing SMN1 transcript exclusion of exon7. BMC Med. Genet. 13, 86. doi:10.1186/1471-2350-13-86

Zheleznyakova, G.Y., Kiselev, A.V., Vakharlovsky, V.G., Rask-Andersen, M., Chavan, R., Egorova, A.A., Schiöth, H.B., Baranov, V.S., 2011. Genetic and expression studies of SMN2 gene in Russian patients with spinal muscular atrophy type II and III. BMC Med. Genet. 12, 96. doi:10.1186/1471-2350-12-96

Zlotogora, J., Grotto, I., Kaliner, E., Gamzu, R., 2016. The Israeli national population program of genetic carrier screening for reproductive purposes. Genet. Med. Off. J. Am. Coll. Med. Genet. 18, 203–206. doi:10.1038/gim.2015.55