

https://dspace.org/

Investigating the utility of different methods to detect SMN DNA copy number and RNA expression in black South African patients with spinal muscular atrophy

Tabane, Kgomotso Odirile Peggy

2021
https://hdl.handle.net/10539/33668

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

Investigating the utility of different methods to detect *SMN* DNA copy number and RNA expression in black South African patients with spinal muscular atrophy

Kgomotso Odirile Peggy Tabane

0609571G

Supervisor: Prof Amanda Krause

Co-Supervisor: Mrs Elana Vorster





Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, South Africa

Johannesburg, 2021

DECLARATION

I, Kgomotso Odirile Tabane, declare that this dissertation is my own work. It is being submitted for the degree of Masters of Science in Medicine in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.



_____

(Signature of candidate)

__29th__day of _October__2021__in_Johannesburg

ACKNOWLEDGMENTS

I would like to thank my supervisors Prof Amanda Krause and Mrs Elana Vorster for their continuous support through this study. They have been encouraging and patient throughout. They have guided and assisted me with the academic and administrative parts of my Master's. I would also like to acknowledge my manager Ms Fahmida Essop who has been supportive and understanding during my studies. She has been kind and has allowed me to take study leave when I have needed. I would also like to thank the molecular diagnostic team; they have been encouraging and have cheered me on during the most challenging times. I would like to thank Alukhethi Singo and Marija Kvas from Whitehead Scientific for their technical assistance with qPCR and RT-qPCR. Thank you, Cassandra Soo, for assisting me with the QuantStudio instruments. Finally, I want to thank my family for their utmost support. Thank you to my mother and father Ruth and Walter Tabane for encouraging me to pursue this Master's degree. Thank you to my brother Obakeng and sister Omolemo Tabane. Special thank you to my sister, Omolemo, who has been my support during this Master's study.

This dissertation is dedicated to my mother Ruth and father Walter Tabane, ke a leboga.

Funders:

National Health Laboratory Service Research Trust (NHLSRT)

Faculty of Health Science Research Committee (FRC)

PRESENTATIONS ARISING FROM THIS RESEARCH PROJECT

Southern African Society for Human Genetics (SASHG), 03 Aug 2019 - 06 Aug 2019
Location: Century City Conference Centre, Cape Town, South Africa.

Title: Determining RNA expression levels of the SMN protein in black South African patients clinically affected with spinal muscular atrophy who tested negative for the homozygous deletion of SMN1, exon 7.

ABSTRACT

Spinal muscular atrophy (SMA) is a common neuromuscular disorder occurring as frequently as albinism in the black South African (SA) population. SMA is a neurodegenerative disease characterised by motor neuron loss in the spinal cord causing muscle weakness and atrophy. SMA is caused predominantly by mutations in the survival motor neuron 1 gene (*SMN1*). A homozygous deletion of *SMN1*, exon 7 is the main cause of SMA in ~95% of patients worldwide but only occurs in 51% of black SA patients. Mutations within *SMN2,* a gene copy, are not thought to cause SMA directly, but to modify disease severity. Black SA individuals have been shown to have copy number variations of the *SMN1* and *SMN2* genes which could potentially mask pathogenic mutations. The aim of the study was to determine the clinical utility of different DNA and RNA methods in testing clinically suggestive SMA patients without an *SMN1* deletion. Three new methods, AmplideX® PCR/CE SMN1/2, qPCR and RT-qPCR were optimised and validated as part of this study. Ninety-two subjects were tested using these methods and results were compared to those obtained with Multiplex Ligation-Probe Dependent Amplification (MLPA). Comparative analysis of the three methods indicated that they were all adequate for diagnostic and carrier testing of SMA subjects. The RT-qPCR showed that DNA copy number does not necessarily correlate directly with RNA copy number. Interpreting expression of the *SMN*Δ7 transcript must be done in the context of the *SMN1* and *SMN2* DNA and RNA copy numbers. DNA copy number detection by qPCR was the most affordable method and AmplideX® PCR/CE SMN1/2 was the only method able to detect gene conversions, although the functional significance of these is uncertain. Diagnosis in subjects with clinical features suggestive of SMA could not be confirmed with RT-qPCR. We recommend further clinical assessment of these patients and testing using newer technologies. Further research into the molecular mechanism underlying SMA in patients with African ancestry is required.

TABLE OF CONTENTS

DECLARATION ii
ACKNOWLEDGMENTS iii
PRESENTATIONS ARISING FROM THIS RESEARCH PROJECT iv
ABSTRACT v
TABLE OF CONTENTS vi
LIST OF FIGURES ix
LIST OF TABLES x
LIST OF ABBREVIATIONS AND SYMBOLS xi
1. INTRODUCTION 1
1.1 Clinical epidemiology 1
1.1. Clinical symptoms of SMA 2
1.2. The *SMN* genes 3
1.2.1. *SMN* gene copies 4
1.2.2. Gene conversions and the evolution of the SMN region 5
1.3. SMN transcripts 7
1.4. SMN Protein 9
1.5. SMA disease modifiers 10
1.5.1. *SMN2* copy number as a modifier of SMA disease. 11
1.5.2. DNA methylation 12
1.5.3. Proteins influencing SMA expression. 13
1.6. Therapy for SMA - Nusinersen (Spinraza) treatment 13
1.7. SMA research in Africa 14
1.8. Diagnostic testing at the Division of Human Genetics, Johannesburg 16
1.9. Aim 18
1.10. Objectives 18
2. SUBJECTS AND METHODS 19
2.1 Subjects and controls 19
2.2 Methods 22
2.2.1 Genomic DNA Copy Number Detection 22
2.2.1.1 DNA Extraction 24
2.2.1.2 DNA yield, purity, integrity determination and routine diagnostic testing . 24
2.2.1.3 Homozygous *SMN1/SMN2* deletion method using RFLP. 24
2.2.1.4 Multiplex ligation-dependent probe amplification (MLPA) 25
2.2.1.5 AmplideX® PCR/CE SMN1/2 copy number detection. 28
2.2.1.6 Real-time PCR (qPCR) 30
2.2.2 *SMN1/SMN2* RNA expression analysis 35
2.2.2.1 RNA extraction 36

2.2.2.2 Reverse transcription of RNA to cDNA 36
2.2.2.3 RT-qPCR experiment design 37
2.2.3 Comparative analysis of DNA and transcript levels of *SMN1* and *SMN2* in subjects and controls 38
2.2.4 Ethics 40
2.2.5 Summary of methods 40
3 RESULTS 41
3.1 DNA copy number detection: MLPA, AmplideX and qPCR 41
3.1.1 DNA copy number detection by MLPA 42
3.1.2 DNA copy number detection by AmplideX 46
3.1.3 Gene conversion and DNA copy number 48
3.1.4 DNA copy number detection by qPCR 49
3.1.5 Comparison of MLPA, AmplideX kit and qPCR methods in the detection of *SMN1* DNA copy numbers 53
3.1.6 Summary of comparison of *SMN1* and *SMN2* DNA copy number 56
3.1.7 Costing of methods for determination of SMN DNA copy number 61
3.2 Gene expression: RNA copy number detection 63
3.2.1 RNA expression results of M1/M1 subjects 64
3.2.2 RNA expression results of M2/M2 subjects 66
3.2.3 RNA expression results of N/M1 subjects: DNA *SMN1* copy number versus FL-SMN1 expression 67
3.2.4 RNA expression results of N/N subjects 69
3.2.5 RNA expression results of N/N subjects: DNA copy number of *SMN2* versus FL-*SMN2* transcript 71
3.3 Summary of results 79
4 DISCUSSION 81
4.1 DNA copy number methods – Comparison and their clinical utility 81
4.1.1 Comparison of MLPA DNA copy numbers 82
4.1.2 *SMN1* gene detection by MLPA vs AmplideX vs qPCR 84
4.1.3 Possible reasons for *SMN1* copy number discrepancies. 85
4.1.4 *SMN1* gene copy number in U/U subjects 86
4.1.5 Comparative analysis of N/N subjects - *SMN1* copy number in black population compared to other populations. 87
4.1.6 SMN2 gene detection by MLPA vs AmplideX vs qPCR 89
4.1.7 Comparative analysis of N/N subjects - SMN2 copy number in black population compared 91
4.1.8 Gene conversion and DNA copy number 92
4.2. RNA expression 94
4.2.1 Comparison of *SMN1* DNA copy number versus FL-*SMN1* transcript 95
4.2.2 DNA and RNA analysis of M1/M1 subjects 95
4.2.3 DNA and RNA analysis of M2/M2 subjects 96

4.2.4 DNA and RNA analysis of N/M1 subjects 96
4.2.5 DNA and RNA analysis of N/N subjects 97
4.2.6 DNA and RNA analysis of U/U subjects 99
4.2.7 Comparison of *SMN1*/*SMN2* DNA and RNA results 99
4.3 Clinical utility of SMN1/SMN1 DNA and RNA testing 102
4.4 Future studies 103
4.4.1 SMA modifiers of disease 103
4.4.2 Beyond the copy number 104
4.4.3 Long-read SMN gene sequencing 105
4.5 Advantages and limitations of the study 105
4.5.1 Advantages and limitations of MLPA 105
4.5.2 Advantages and limitations of AmplideX 106
4.5.3 Advantages and limitations of qPCR 106
4.5.4 Advantages and limitations of RT-qPCR 106
5 CONCLUSION 108
REFERENCES 113
APPENDICES 119

LIST OF FIGURES

Figure 1.1. Genetic map of the spinal muscular atrophy locus 4
Figure 1.2. The *SMN* genes. 4
Figure 1.3. Model of alleles present in the normal population and in patients affected with SMA.. 6
Figure 1.4. Six possible ways of gene conversion between *SMN1* and *SMN2* in exons 7 and 8..... 7
Figure 1.5. Difference between *SMN1* and *SMN2* transcription..... 8
Figure 1.6. Modifiers of SMA disease..... 11
Figure 1.7. Epigenetic changes can modify expression. 12
Figure 1.8. Pedigree of a family affected with SMA. 17
Figure 2.1. Strategy for testing of different methods..... 21
Figure 2.2. Flow diagram showing DNA copy number detection. 23
Figure 2.3. Homozygous deletion *SMN1/SMN2* method. 25
Figure 2.4. Steps in the MLPA test process. 26
Figure 2.5. Results of MLPA analysis using the SALSA MLPA probe-mix P021..... 27
Figure 2.6. Peak profile of a sample with 3 copies of SMN1, exon 7 and 1 copy of SMN2 .. 29
Figure 2.7. Example of *SMN1/2* AmplideX Excel report of 30
Figure 2.8. Quantitative PCR by use of TaqMan™ probes..... 31
Figure 2.9. Annotation of *SMN* primers 32
Figure 2.10. Annotation of *CFTR* primers..... 33
Figure 2.11. An example of a qPCR amplification plot visualised on the QuantStudio 3..... 34
Figure 2.12. Example of a gene copy number plot visualised on the QuantStudio 3 34
Figure 2.13. Summary of RNA extraction protocol using Tempus™ Blood RNA Tube 36
Figure 2.14. Summary of reverse transcription using the ImProm-II™ Reverse Transcription System (Promega, Madison, Wisconsin, USA)..... 37
Figure 2.15. Primer and probe sequences of FL-SMN1 and FL-SMN2..... 38
Figure 2.16. Predicted comparative analysis of DNA *SMN1* and *SMN2* copy number versus RNA transcription. 39
Figure 3.1. MLPA results obtained for the *SMN* region. 45
Figure 3.2. N/N, M1/M1 and M2/M2 AmplideX results. 46
Figure 3.3. AmplideX results in subjects with varying gene copy numbers. 48
Figure 3.4. Amplification plot of qPCR. 50
Figure 3.5. Copy number ratio of *SMN1* and *SMN2* in 20 subjects. 51
Figure 3.6. Comparison of *SMN1* copy number results across three methods. 54
Figure 3.7. Comparison of *SMN2* copy number results across three methods. 55
Figure 3.8. *SMN2* copy number compared to genotype. 55
Figure 3.9. Comparison of DNA copy number between MLPA, AmplideX and qPCR. 56
Figure 3.10. Comparison of MLPA, AmplideX and qPCR results in a single subject. 60
Figure 3.11. DNA copy number versus RNA expression in an M1/M1 subject: SMA1988. .. 64
Figure 3.12. An M1/M1 subject with zero copies of *SMN1* on DNA copy number detection and zero copies of FL-SMN1 transcript.. 65
Figure 3.13. DNA copy number versus RNA expression in an M1/M1 subject: SMAR7. 66
Figure 3.14. DNA copy number versus RNA expression in an M2/M2 subject: SMAR2. 67
Figure 3.15. *SMN1* DNA copy number versus RNA expression in N/M1 subjects 68
Figure 3.16. *SMN2* DNA copy number versus RNA expression in N/M1 subjects 69
Figure 3.17. DNA copy number versus gene expression results of N/N subjects. 70
Figure 3.18. *SMN2* DNA copy number versus RNA expression results of N/N subjects. 72
Figure 3.19. DNA copy number versus gene expression results of N/N subjects with matching copies of *SMN1* and *SMN2*. 73
Figure 3.20. DNA copy number versus gene expression results of N/N subjects with variable copies of *SMN1* and *SMN2*. 75

Figure 3.21. DNA copy number versus RNA expression results of N/N subjects with three of more copies of *SMN1*. 76
Figure 3.22. DNA copy number versus RNA expression results of U/U subjects. 77
Figure 4.1. U/U Subject with *SMN* copy number discrepancy. 85
Figure 4.2. Diagnostic algorithm of SMA (Mercuri et al., 2018). 87
Figure 4.3. Comparative analysis of genomic *SMN1* and *SMN2* copy numbers versus RNA transcription. 101

LIST OF TABLES

Table 1.1. Clinical sub-types of spinal muscular atrophy (SMA). Adapted from Butchbach, 2016. 3
Table 2.1. Groups of subjects and controls tested, together with numbers of individuals in each group. 20
Table 2.2. The relationship between DQ values and *SMN1* copy number on MLPA analysis ("MLPA General Protocol MDP-v007.pdf," n.d.) 28
Table 2.3. Expected peak sizes on Applied Biosystems® Genetic Analysers (3500 series) AmplideX PCR/CE SMN1/SMN2 kit protocol guide. 29
Table 2.4. Default *SMN1* and *SMN2* DNA copy number bins vs normalised ratio. 30
Table 2.5. Default *SMN1* Hybrid and *SMN2* Hybrid DNA copy number bins vs normalised ratio 30
Table 2.6. Primer and probe sequences of the *SMN1* and *SMN2* qPCR method 32
Table 2.7. Description of RT-qPCR transcripts 35
Table 2.8. Primer and probe sequences of the *SMN* expression RT-qPCR relative quantification method designed by Vorster et al (2020) 37
Table 3.1. Ethnicity and genotype of subjects tested for DNA copy number. 42
Table 3.2. Comparison of *SMN1*, exon 7 and exon 8 copy number in M1/M1 subjects 43
Table 3.3. Percentage of hybrid *SMN2*-to-*SMN1* copies detected using AmplideX. 49
Table 3.4. Comparison of *SMN1* copy number in different subject groups using different methods. 53
Table 3.5. Summary of DNA copy number methods 59
Table 3.6. Costing results for MLPA, AmplideX and qPCR 61
Table 4.1. Comparison of *SMN1* copy numbers on MLPA/AmplideX/qPCR 85
Table 4.2. Frequency of N/N *SMN1* Copy Number in Different Geographical Regions 88
Table 4.3. Comparison of *SMN2* copy numbers between MLPA/AmplideX/qPCR 90
Table 4.4. Frequency of N/N SMN2 Copy Number in Different Geographical Regions 92

LIST OF ABBREVIATIONS AND SYMBOLS

A	Adenine
ABI	Applied biosystems
bp	Base pairs
C	Cytosine
cm	Centimetre
CNS	Central nervous system
CNV	Copy number variation
C_t	Cycle threshold
Ctrl	Control
ddH_2O	Deionised, distilled water
ddNTPs	Deoxyribonucleic triphosphates
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
DQ	Dosage quotient
EC	Endogenous control
EDTA	Ethylene-diamine-tetra-acetate
EtBr	Ethidium bromide
F	Forward primer
FL-SMN	Full length survival motor neuron
g	Grams
G	Guanine
kb	Kilobases
kDa	Kilo Dalton
L	Litre
Mg^{2+}	Magnesium ion
$MgCl_2$	Magnesium chloride
min	Minutes
ml	Millilitre
MLPA	Multiplex ligation-dependent probe amplification
mm	Millimetre
mM	Millimolar
mRNA	Messenger RNA
NaCl	Sodium Chloride
ng	Nanograms
NGS	Next generation sequencing
NHLS	National Health Laboratory Service
nm	Nanometre
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerase chain reaction
PLS3	Plastin 3
pmol	Picomoles
qPCR	Quantitative/real-time polymerase chain reaction
RT-qPCR	Reverse transcription quantitative PCR
RFLP	Restriction fragment length polymorphism

RNA	Ribonucleic acid
rpm	Revolutions per minute
RQ	Relative quantification
SDS	Sodium Dodecyl Sulphate
SMA	Spinal muscular atrophy
SMN	Survival of Motor Neuron protein
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
UV	Ultraviolet
WES	Whole exome sequencing
WGS	Whole genome sequencing

1. INTRODUCTION

Spinal muscular atrophy is a common neuromuscular disorder occurring as frequently as albinism in the black South African (SA) population. Spinal muscular atrophy (SMA) is a neurodegenerative disease characterised by motor neuron loss in the spinal cord resulting in muscle weakness and atrophy. Five clinical sub-types of SMA have been identified according to the disease manifestations and age of onset (Butchbach, 2016).

Mutations in the survival motor neuron 1 gene (*SMN1*) are the cause of SMA (Lefebvre et al., 1995). A homozygous deletion of *SMN1*, exon 7, is the main cause of SMA in ~95% of patients worldwide, but it only occurs in 51% of black SA patients (Labrum et al., 2007). Mutations within *SMN2*, a gene copy almost identical in sequence to *SMN1*, are not thought to be associated with SMA directly, but modify disease severity (Butchbach, 2016). The high homology of the *SMN1* and *SMN2* genes complicates analysis.

Previous studies performed at the Division of Human Genetics, National Health Laboratory Services (NHLS) and the University of Witwatersrand (henceforth referred to as “the Division”), have shown that subjects of African ancestry may have a hypervariable *SMN* region. Altogether 50.8% of these subjects had multiple copies of *SMN1* compared to 3.5% across various Caucasian populations (Labrum et al., 2007; Stevens et al., 1999). This has made diagnostic and carrier testing in African populations challenging, as large copy number variations (CNVs) of the *SMN* region could interfere with diagnostic tests and may play a role in the SMA disease mechanism. It is possible that some of the multiple *SMN1* gene copies may not be completely functional (Vorster et al., 2020).

1.1 Clinical epidemiology

Spinal muscular atrophy is an autosomal recessive disorder, caused by mutation of the *SMN1* gene in most cases (Lefebvre et al., 1997). A homozygous deletion of exon 7, in *SMN1* causes approximately 95% of SMA cases (Melki et al., 1994). The incidence of SMA worldwide is 1 in 3900–16,000 live births (Belter et al., 2018; Verhaart et al., 2017). The carrier frequency of SMA is 1 in 25-50 in most populations (Ben-Shachar et al., 2011) . Labrum et al (2007) estimated the incidence of SMA in

SA to be 1 in 3574 in the black population and 1 in 1945 in the white population. The carrier frequency in SA was estimated to be 1 in 50 in the black SA population and 1 in 23 in the white SA population (Labrum et al., 2007). The incidence of SMA in SA appears to be higher than in other countries, which could be due to the unique ethnic background of SA populations. The incidence of SMA is slightly higher than that of albinism (1 in 3900) in the SA black population (Kromberg and Jenkins, 1982) and almost as high as that of cystic fibrosis (1 in 2500) in the SA white population (Goldman et al., 2001; Kromberg and Jenkins, 1982).

1.1. Clinical symptoms of SMA

SMA is an early-onset neurodegenerative disease characterised by α-motor neuron loss in the anterior horn of the spinal cord causing muscle weakness and atrophy. The weakness is located more proximal than distal in most SMA subtypes (Farrar and Kiernan, 2015) . Individuals suspected of having spinal muscular atrophy present with a history of motor difficulties, especially a loss of motor skills, proximal muscle weakness, hypotonia, areflexia/hyporeflexia, tongue fasciculations and/or evidence of motor neuron disease on physical examination (Prior and Finanger, 1993).

SMA is a heterogeneous disease which ranges from severe to mild adult-onset phenotypes. It is classified into five clinical sub-types ranging in severity and age of onset; however, the clinical sub-types are more continuous than distinct (Farrar and Kiernan, 2015). Table 1.1 illustrates these subtypes of SMA; type 0 SMA has a prenatal onset with a survival rate of less than six months and is the most severe form. These patients present with very severe hypotonia and early respiratory failure and they are unable to reach any developmental milestones (Butchbach, 2016). Patients with type I SMA, also known as Werdnig-Hoffman disease, have an age of onset before six months, are unable to sit or walk and live for approximately two years. The patients usually present with proximal muscle weakness, hypotonia, and mild contractures; affected infants have problems sucking or swallowing leading to growth failure and recurrent aspiration (Prior and Finanger, 1993). Life expectancy has improved due to respiratory and nutritive care (Corsello et al., 2021).

Patients with type II SMA or Dubowitz disease, are usually diagnosed by 6-18 months and survive into adulthood; however, they have trouble walking and sitting. Type II SMA tends to manifest as progressive proximal leg weakness that is greater than

weakness in the arms (Kolb and Kissel, 2015). Patients may experience finger trembling, general flaccidity, and scoliosis. A study done by Zerres et al in Germany and Poland showed that the SMA II patients were alive at age 25 years (Zerres et al., 1997).

Patients with type III SMA also known as Kugelberg-Welander disease have an age of onset of 18 months or older. These patients walk with difficulty but usually have a normal life expectancy (Kolb and Kissel, 2015).

Patients with SMA type IV have an adult age of onset, have the mildest form of the disease and usually have a normal life expectancy (Kolb and Kissel, 2015). Type IV patients are sometimes misdiagnosed with amyotrophic lateral sclerosis (ALS) due to the late onset of disease and similar clinical features (Kolb and Kissel, 2015).

Table 1.1. Clinical sub-types of spinal muscular atrophy (SMA). Adapted from Butchbach, 2016

Type	Age of onset	Requires Respiratory support at birth	Able to sit	Able to stand	Able to walk	Life expectancy	Predicted SMN2 copy number
0	Prenatal	Yes	No	No	No	<6 months	1
I	<6 months	No	No	No	No	<2 years	2
II	6–18months	No	Yes	No	No	10–40years	3
III	>18months	No	Yes	Yes	Assisted	Adult	3–4
IV	>5 years	No	Yes	Yes	Yes	Adult	>4

1.2. The *SMN* genes

The *SMN1* and *SMN2* genes lie within a 500 kilobase (kb) inverted duplication of chromosome 5q13. There are five protein-coding genes (*SERF1*, *SMN1*, *SMN2*, *NAIP* and *GTF2H2*) within the chromosome 5q13 region as illustrated in Figure 1.1. The *SMN1* gene was first identified and characterised by Lefebvre et al in 1995. Through genetic mapping Lefebvre et al (1995) demonstrated that patients presenting with SMA had large scale deletions on chromosome 5q13. The researchers further showed that deletion of the *SMN1* gene was responsible for SMA. The *SMN1* gene was found in the telomeric region of the chromosome 5q13 repeat.

SMN1 and *SMN2* share 99% nucleotide identity and the critical difference between the two genes is a C>T transition in exon 7 (*SMN2* c.850C>T) that affects the splicing of the genes (Monani et al., 1999a).

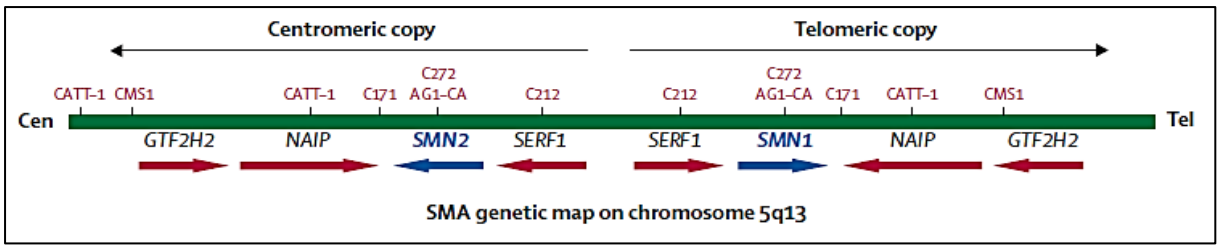


Figure 1.1. Genetic map of the spinal muscular atrophy locus (Lunn and Wang, 2008). *SMN* genes are shown in blue with surrounding genes in black. The genes are in an inverted duplication of chromosome 5q13. The red and blue arrows indicate the direction of transcription of the genes.

1.2.1. *SMN* gene copies

The telomeric copy, *SMN1*, produces the full length SMN protein (FL-SMN) and is the main gene involved in SMA pathogenicity (Zhang et al., 2003). The *SMN* genes are approximately 34kb in length, comprised of 10 exons and nine introns (Lefebvre et al., 1997; Wirth, 2000). More recent research has demonstrated that exon 2 is comprised of two separate exons. To avoid confusion it has now been renamed 2a and 2b (Bürglen et al., 1996). Exon 6b was recently discovered by Seo et al (2016). *SMN1* produces approximately 80% and *SMN2* about 20% of FL-SMN protein. The centromeric copy; *SMN2*, is only distinguishable from *SMN1* by five nucleotides with a critical difference in exon 7 (c.840C>T) that disrupts an exonic splice enhancer (Monani et al., 1999b, 1999a). Figure 1.2 illustrates the difference between the *SMN1* and *SMN2* genes.

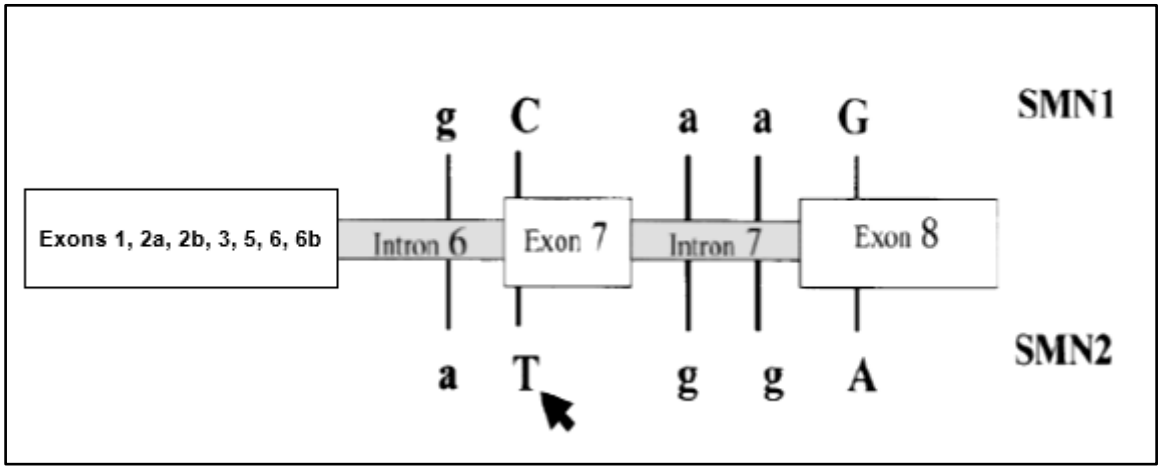


Figure 1.2. The *SMN* genes. *SMN1* and *SMN2* are 99% similar with the critical difference in exon 7 c.840C>T (indicated by the arrow) that results in differential splicing (Wirth, 2000). The *SMN1* gene can be distinguished from the *SMN2* gene by one nucleotide difference in intron 6, one in exon 7, two in intron 7 and one in exon 8. These differences are used to distinguish between *SMN1* and *SMN2* in the design of diagnostic methods (Wirth, 2000).

1.2.2. Gene conversions and the evolution of the SMN region

The SMN region is a complex region prone to gene rearrangement, therefore the sequence of genes in this region is poorly understood. Gene conversion may arise during meiosis, where non-Mendelian segregation of alleles occurs resulting in two copies of the same DNA sequence transferred to one gamete and “zero” DNA sequence to the other gamete. A variety of copies of the DNA sequence will occur therefore resulting in copy number variations (CNV). CNVs vary between individuals and may be disease causing, benign or may modify disease. The frequency of *SMN1* copy number variants differs between ethnic groups. Hendrickson et al (2009) showed that approximately 90% of Caucasians have two copies of *SMN1* compared with approximately 50% of African Americans. Approximately 46% of African Americans had three *SMN1* copies compared to 6% of Caucasians. A similar study by Sangare et al (2014), on sub-Saharan African populations found high copy numbers of *SMN1* and low copy number of *SMN2*. The high copy numbers may be due to a conversion of *SMN2* to *SMN1* (Hendrickson et al., 2009).

The high homology between *SMN1* and *SMN2* and its adjacent genes suggests a gene duplication event occurred. The evolutionary history of the *SMN* genes was studied by Rochette et al (2001) to determine when the duplication of the *SMN* region occurred. They studied the variation of the region in different human controls compared to other non-human primates and mice. They showed that mice and other primates have only one copy of the *SMN* gene and that the duplication of the *SMN* region is a recent event that emerged after separation of human and other primate lineages.

They further illustrated that other primates had *SMN1* gene sequences like those occurring in human *SMN1*. However, the *SMN2* gene was not found in the primates they studied and *SMN2* was found to be unique to humans. They concluded that the *SMN* duplication occurred at least five million years ago prior to human-chimpanzee divergence but the SMN2 gene emerged after the separation of the two lineages. They further suggest that gene conversions events between *SMN1* and *SMN2* are an ongoing process (Rochette et al., 2001).

Figure 1.3 illustrates the different conversions of *SMN1* to *SMN2* and how different alleles affect the SMA phenotype. There is evidence of gene conversion of *SMN1* to

SMN2, and as well as *SMN2* to *SMN1*. Gene conversion of *SMN1* to *SMN2* has been associated with less severe SMA phenotypes (Burghes, 1997). Individuals with greater than three copies of *SMN1*, were more likely to have fewer copies of *SMN2* further supporting the hypothesis of gene conversions of *SMN2* to *SMN1* (Ogino et al., 2003).

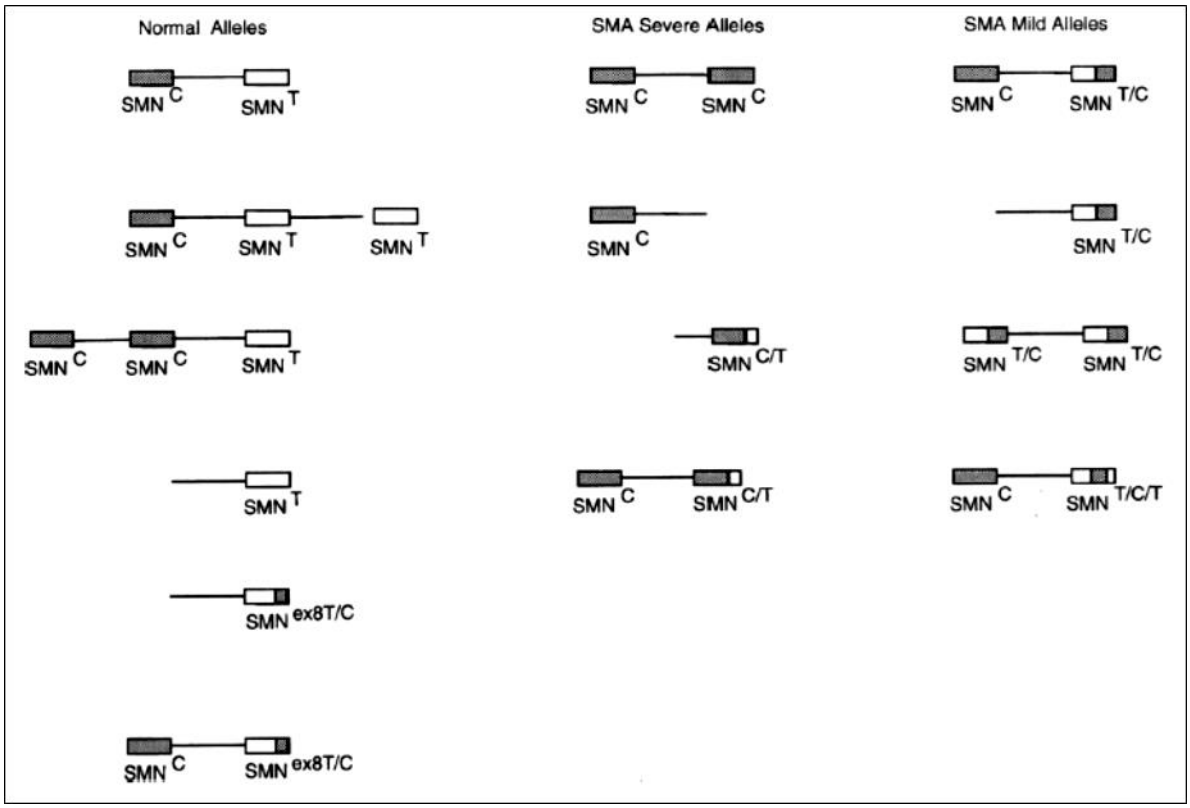


Figure 1.3. Model of alleles present in the normal population and in patients affected with SMA. The severity of SMA is predicted to be associated with the *SMN1*, exon 7 deletion, and the conversion of *SMN1* to SMN2. Alleles with gene conversions are predicted to cause milder SMA. The C allele denotes the normal (wildtype) and the T denotes the mutant allele. Each box represents the number of copies.

Six possible points of gene conversions between *SMN1* and *SMN2* have been proposed and were tested by Wang et al (2010) and illustrated in Figure 1.4. They designed an alternative method to multiplex ligation probe-dependent amplification (MLPA) that was able to detect gene conversions between *SMN1* and *SMN2*, specifically in exons 7 and 8. Identification of gene conversions was helpful for SMA genotyping and diagnosis as gene conversions have been shown to cause a milder phenotype (Ogino et al., 2004).

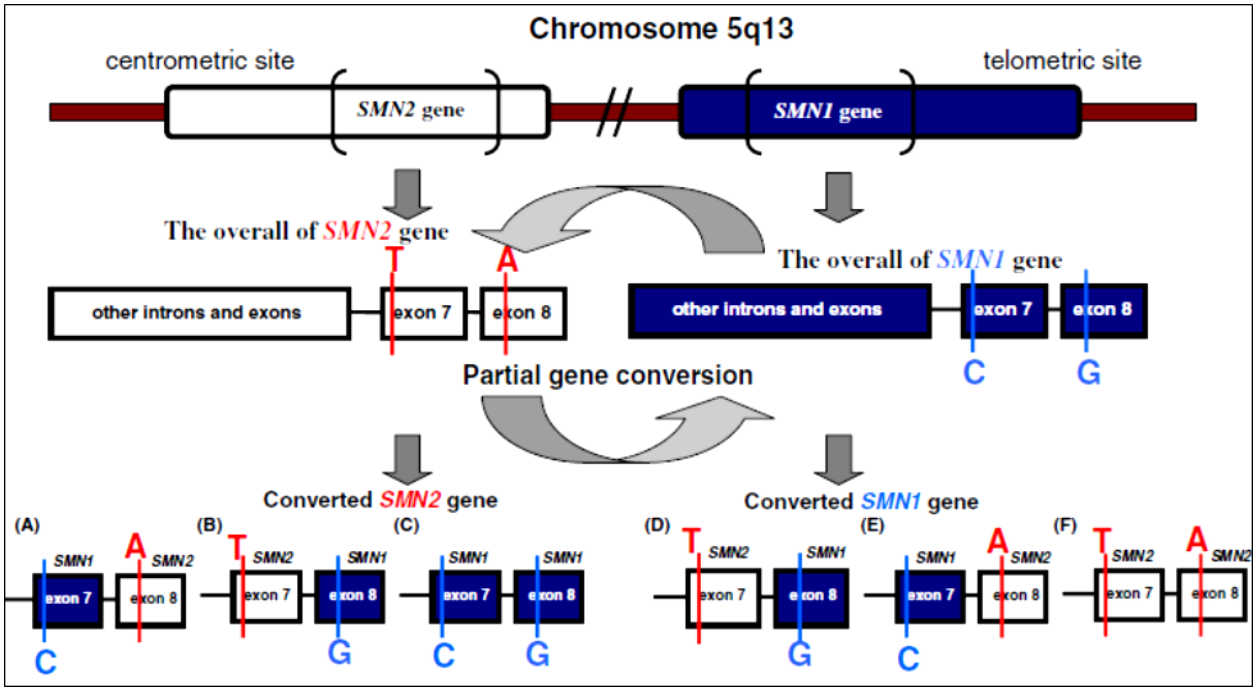


Figure 1.4. Six possible ways of gene conversion between *SMN1* and *SMN2* in exons 7 and 8. The *SMN1* gene, shown in dark blue can convert to *SMN2* (white box) and vice versa. A-C indicate gene conversions of *SMN1* to *SMN2* and D-F indicate gene conversions of *SMN2* to *SMN1.* Gene conversions of SMN genes are complex and only six possible ways are know (Wang et al., 2010).

1.3. SMN transcripts

Transcription in eukaryotic organisms, including humans, is the process of copying genetic information from DNA into the complementary RNA strand. The RNA polymerase enzyme, together with transcription factors, binds to the promoter region of the DNA strand, to create a complementary precursor messenger RNA (pre-mRNA) strand. The pre-mRNA is then spliced to form a mRNA. Splicing is the removal of introns from the pre-mRNA. This is carried out by splicing factors, enhancers, and silencers. A variety of transcripts can be generated from a single pre-mRNA due to alternative splicing (Singh et al., 2012). The *SMN* genes are transcribed into various mRNA transcripts due to variants in the *SMN* gene or regulatory elements. Only 20% of full-length (FL) transcripts are transcribed from *SMN2*; the majority of SMN2 transcripts are truncated as depicted in Figure 1.5 (He et al., 2013; Zhang et al., 2003). The C to T change in exon 7 in *SMN2* reduces the recognition of exon 7 during splicing, resulting in the exclusion of exon 7 in *SMN2* transcripts as illustrated in Figure 1.5. The altered transcripts lacking exon 7 (SMNΔ7) are highly unstable and degrade rapidly (Moulard et al., 1998).

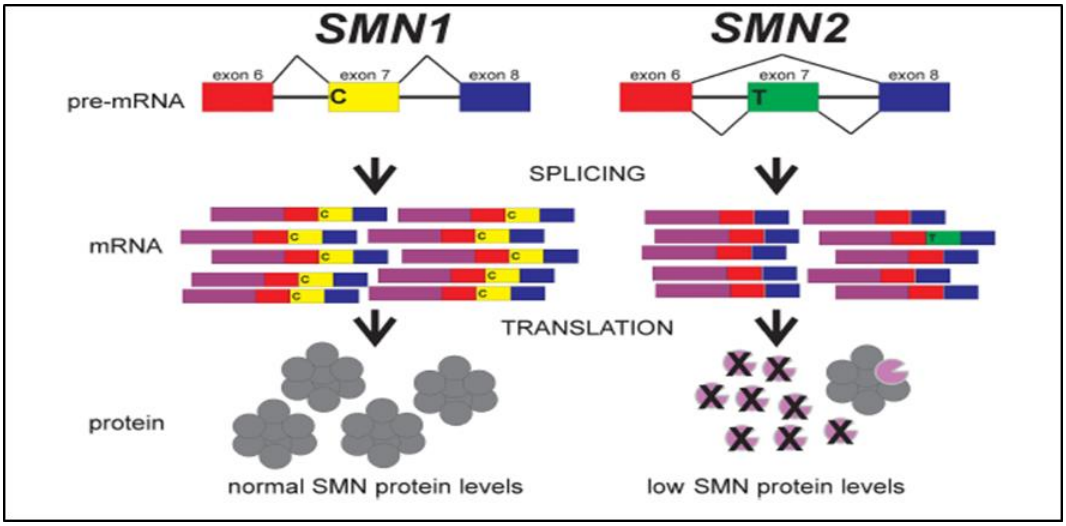


Figure 1.5. Difference between *SMN1* and *SMN2* transcription (Butchbach, 2016). *SMN1* is transcribed into mRNA including exon 7 and is translated into a functional SMN protein. Most of the *SMN2* transcripts result in truncated mRNA strands lacking exon 7 (SMNΔ7); these are translated into an unstable SMN protein which is easily degraded.

There are four possible transcripts that can be generated from the SMN genes: FL-*SMN*, *SMN*Δ*7*, axonal *SMN* (a-*SMN*) and *SMN6B*. The FL-*SMN* transcript (produced by both *SMN1* and *SMN2)* is the main transcript required by most cells. The *SMN*Δ*7* transcript does not contain exon 7 and is produced from alternative splicing of the SMN2 gene or by a mutant *SMN1* gene in patients affected with SMA. The a-SMN and SMN6B are alternative transcripts produced from the *SMN1* and *SMN2* genes, respectively (Seo et al., 2016).

The FL-SMN transcript is made up of 80% *SMN1* transcripts and approximately 20% of SMN2 transcripts are transcribed with the included exon 7. The FL-SMN transcript is the most critical transcript as it produces the functional SMN protein. The SMNΔ7 transcript lacking exon 7 is the predominant transcript derived from *SMN2* in all tissues except testis. SMNΔ7 is translated into an unstable and easily degraded SMN protein.

The a-*SMN* plays a role in mammalian brain development; the a-SMN transcript is preferentially transcribed from the *SMN1* gene (Setola et al., 2007). They identified a-SMN in the human spinal cord, specifically in motor neuron axons in rats, mice, and humans. The protein was observed to be expressed early in embryonic development and was downregulated in adult cells. The a-SMN protein promotes axon growth, stimulates cell mobility, regulates expression of chemokines (CCL2 and CCL7) and insulin-like growth factor-1 (IGF1) (Locatelli et al., 2012). The a-SMN transcript was

not included in this study as it is present in early embryonic development and its association with SMA is not well understood.

The recently identified SMN6B transcript was detected using a mouse model which harbours the *SMN2* human gene and a hybrid *SMN1/SMN2* at the same locus (Seo et al., 2016). Seo et al (2016) sequenced the full transcript and identified a novel SMN transcript with a longer exon 6, the exon contained a portion of intron 6 and was therefore termed exon 6B with the transcript named SMN6B. They went on to determine possible expression of SMN6B in human tissues and identified SMN6B by reverse transcribed qPCR (RT-qPCR) in all human tissues. This transcript was expressed highest in brain and lowest in skeletal muscles, but its role is poorly understood and therefore it was not included as part of the present study (Seo et al., 2016).

Other transcripts of SMN have been identified, indicating the diverse splicing involved in pre-mRNA of the *SMN* gene and possibly other mechanisms of disease pathogenesis. Singh et al (2012) used a multiple exon detection method (MESDA) to identify isoforms of SMN. They identified new splice sites and demonstrated that different transcripts were formed from regulation in the promoter region and splice sites.

1.4. SMN Protein

The SMN protein has numerous roles in all cell types and is required for survival of all mammals. The protein is formed from translation of SMN mRNA into protein; the different transcripts of SMN produce different isoforms of protein with the FL-SMN being the most stable and therefore most significant. The survival motor neuron protein is a molecular chaperone and is essential for the assembly of ribonucleoprotein (RNP) complexes. The RNPs are formed when mRNA and non-coding RNAs (nc-RNA) associate with RNA binding proteins to form complexes that are involved in RNA processing and post-transcriptional regulation (Cooper et al., 2009).

The SMN protein is involved in numerous biological functions such as gene regulation and expression. The protein accumulates in nuclear bodies named gems and cajal bodies (Pellizzoni, 2007). The SMN protein then forms an SMN complex by interacting with Gemin 2-8 proteins and an unrip protein (Pellizzoni, 2007). The SMN complex functions in the assembly of small nuclear ribonucleoproteins (snRNPs) which are

essential in expression of all protein-coding genes (Pillai et al., 2003). Another function of the complex is in the signal recognition particle (SRP) biogenesis which targets polypeptides into the endoplasmic reticulum (Piazzon et al., 2013). The SMN protein is involved in telomerase biogenesis, selenoprotein translation, 3′ end processing of histone mRNAs, pre-mRNA splicing, interaction with transcription factors and enhancers, RNA trafficking, translation regulation of protein arginine methyl transferase 4 (PRMT4), selenoprotein synthesis and Stress granule (SG) formation (Seo et al., 2016; Singh and Singh, 2018). The numerous biological functions of the complex shows the importance of the SMN protein and therefore a mutation in the *SMN* gene or any other proteins involved in its regulation may be involved in SMA pathogenesis.

The *SMN1* gene, exon 7 critical in SMA disease will be referred to as *SMN1 g*oing forward. Further the *SMN2* gene, exon 7 will be referred to as *SMN2* for simplicity. When referring to another specific exon other than 7 in the *SMN* genes that exon will be clearly stated.

1.5. SMA disease modifiers

Disease modifiers are genetic or environmental factors that either cause a milder or more severe form of a disease (Genin et al., 2008). There are several modifiers of SMA, *SMN2* copy numbers, regulatory proteins and DNA methylation which all have the potential to be used for treatment of SMA. Possible modifiers of SMA disease are indicated in Figure 1.6, not all modifiers were included in this study, and further research may indicate other modifiers of SMA.



Figure 1.6. Modifiers of SMA disease

1.5.1. *SMN2* copy number as a modifier of SMA disease.

The number of *SMN2* copies is the main modifier of SMA disease - more copies result in a milder form of SMA. Asymptomatic individuals with a homozygous deletion of *SMN1* have been observed with increased copies of *SMN2* (Jedrzejowska et al., 2008). Patients with more copies of the *SMN2* gene are more likely to have milder symptoms and a later onset of disease (Maretina et al., 2018). However, there have been patients with more than two copies of *SMN2* with severe SMA or patients with only two copies having a milder phenotype - this highlights the possibility of other disease modifiers (Maretina et al., 2018). The number of *SMN2* gene copies in the genome varies from 0 to 8, with an inverse relationship between *SMN2* copy number and severity of SMA disease (Butchbach, 2016). The number of *SMN2* gene copies is significantly increased in black and mixed ancestry populations (12.4% and 18.8% respectively) (Vorster et al., 2020).

1.5.2. DNA methylation

Epigenetics refers to heritable changes in chromosomes that modify gene expression without altering the DNA sequence (Berger et al., 2009). Epigenetic mechanisms include DNA methylation, imprinting and X chromosome inactivation (Berger et al., 2009). DNA methylation is the addition of a methyl group preferentially to cytosine residues within CpG dinucleotides (cytosine followed by guanine) which are concentrated in CpG islands, (see Figure 1.7). Methylation of gene promoters leads to transcription inactivation and aberrant methylation patterns are associated with various diseases including neurodevelopmental and neurogenerative disorders (Maretina et al., 2018).

Hauke et al (2009) showed that *SMN2* gene expression levels are reduced by DNA methylation. There are four CpG islands surrounding the translational site of *SMN2* and the methylation patterns in patients affected with SMA were studied. Significant methylation was demonstrated in SMA type II compared to SMA type III patients (Hauke et al., 2009). Further analysis by whole genome methylation analysis of SMA patients revealed different methylation patterns between SMA types involving other regulatory genes *SLC23A2*, *NCOR2* and *DYNC1H1* genes (Maretina et al., 2019, 2018). DNA methylation was shown to be a modifier of SMA disease and may be important in designing new therapies.

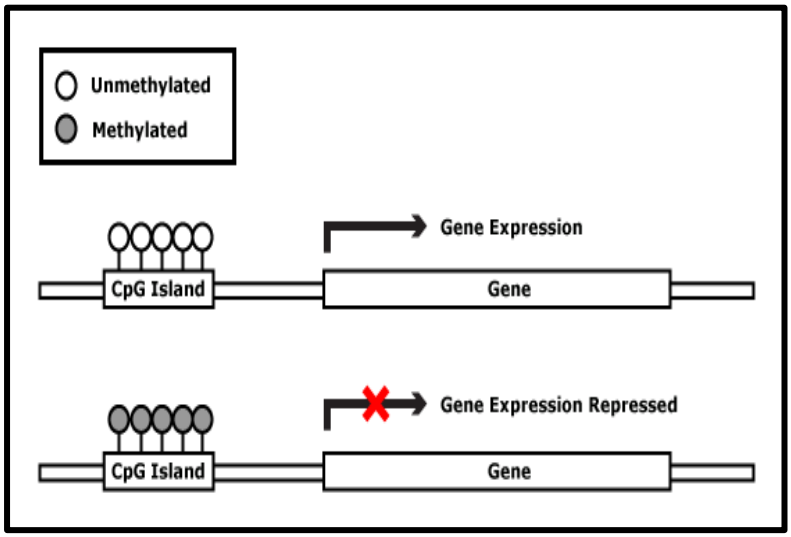


Figure 1.7. Epigenetic changes can modify expression. A methyl group is added to CpG islands and gene expression is repressed (Yousefi et al., 2013).

1.5.3. Proteins influencing SMA expression.

Proteins that modify the actin-binding and cytoskeleton regulation have been studied to determine their role in SMA disease. The Plastin 3 (*PLS3*) expression level was increased in unaffected individuals compared to affected patients (Maretina et al., 2018). Other regulatory proteins have been implicated in SMA pathogenesis, these include phosphatase and tensin homolog (PTEN), Zinc finger protein (ZPR1) and Prolactin (Maretina et al., 2018). The SMN protein is degraded by the ubiquitin/proteome pathway. Treatment of SMA mice models with the proteasome inhibitor, bortezomib, has shown reduced degradation of SMN protein and improved motor function (Kwon et al., 2011).

1.6. Therapy for SMA - Nusinersen (Spinraza) treatment

Once a diagnosis of SMA is confirmed the family of the patient is referred for genetic counselling. Appropriate specialised clinicians can refer patients to clinical trials or treatments available. Most treatments of SMA are supportive but do not cure SMA.

Nusinersen was the first drug therapy approved for the treatment of SMA in December 2016 by the US Food and Drug Administration (FDA), for paediatric and adult patients with SMA. Nusinersen is an antisense oligonucleotide that modifies SMN*2* pre-mRNA to include exon 7 during splicing therefore increasing FL-SMN production (Hua et al., 2010). A clinical study of infants diagnosed with SMA was carried out by Finkel et al (2017). The patients treated with Nusinersen displayed motor milestones responses, survived longer and milestones appeared to improve with time, when compared to patients without Nusinersen treatment (Finkel et al., 2017). Furthermore, patients with later onset SMA treated with Nusinersen also displayed improved milestones (Finkel et al., 2017; Zuluaga-Sanchez et al., 2019). Treatment outcomes varied in different countries, which was likely due to dissimilar supportive care for affected patients (Zuluaga-Sanchez et al., 2019). No adverse effects were observed in most clinical trials (Finkel et al., 2017).

Nusinersen is administered via lumbar puncture starting with an initial dose, followed by four doses every four months thereafter. This treatment costs R7,562,402.70 for the initial dose, followed by R3,782,006.72 annually (Zuluaga-Sanchez et al., 2019). Treatment for SMA is therefore extremely expensive and is not affordable for the

majority of South Africans. South African patients with a confirmed diagnosis of SMA may be included in a clinical trial with a recommendation from a clinician. All patients included in the clinical trials had a confirmed diagnosis of SMA and at least one copy of the *SMN2* gene.

At present molecular diagnosis of SMA is not achieved in approximately 49% of black South African patients presenting with clinical symptoms of SMA. These patients would not be eligible for Nusinersen treatment without a confirmed diagnosis of SMA due to a deletion or mutations in the *SMN1* gene.

1.7. SMA research in Africa

Research in Africa on SMA has been sparse and most studies reported have been done in South Africa. The lack of research undertaken may be due to the high likelihood of early death of patients with SMA in infancy since patients may not have received care particularly in poorly resourced countries. Initial studies focused on clinical features and muscle biopsies of patients suggestive of SMA (Kiepiela et al., 1988; Moosa and Dawood, 2008). Kiepiela *et al* (1988) found no significant difference in T cells and B cells from patients clinically diagnosed with SMA, and their study indicated that the SMN protein does not play a role in immune regulation.

In 1990, research by Moosa and Dawood on 45 patients showed similar clinical characteristics in African patients compared to those of European ancestry, with an additional feature of facial weakness in African children (Moosa and Dawood, 1990). They also noted a lack of positive family history of SMA in African families, possibly because of a reluctance of affected families to provide a family history, due to cultural reasons or rarity of the disease.

Stevens *et al* (1999) aimed to identify the molecular basis of SMA in black patients. Their research was based in Johannesburg and included black patients ascertained through the Division. The *SMN* genes and the *NAIP* gene were analysed in samples from patients confirmed to have SMA on muscle biopsies. They found a homozygous deletion of *SMN1* in 65.5% of their patients, which is far less than found in other studies. They were the first group to suggest a possibility of a different mechanism of SMA disease in black patients.

Wilmshurst et al (2002) did further research in Cape Town, South Africa, with strict inclusion criteria for the patients with SMA. They found a homozygous deletion of *SMN1* in 100% of their black patients. Wilmshurst *et al* (2002) excluded patients with facial weakness and had a smaller sample size of 12 compared to 29 samples in Johannesburg (Stevens et al, 1999). The difference in the two studies could be due to the sample size, the inclusion criteria, or the different geographical location.

Another study was carried out by Labrum *et al* (2007), to further understand the molecular basis of SMA in black SA patients. They found the carrier frequency to be higher than previously thought, at 1 in 50 and 1 in 23 in the black and white populations, respectively. Only 51% of black patients clinically affected with SMA had a homozygous deletion of *SMN1* compared with 95% of the patients in the white population. The diagnostic yield of Labrum *et al* (2007) was less than Stevens et al (1999) most likely due to sample size which was 116 in the former study and 29 in the latter study. Both studies further support the hypothesis that there may be a different molecular basis for black SA patients with an SMA phenotype.

A study performed in Mali to determine the carrier frequency of SMA in Africans was done on 628 Malians, 120 Nigerians, and 120 Kenyans (Sangare, et al., 2014). The results showed that the carrier frequency of SMA in Malians was lower, at 1 in 209 compared to 1 in 30-50 in Europeans (Labrum et al., 2007; Ogino et al., 2004; Sangare, et al., 2014). The carrier frequency was calculated by testing healthy individuals to determine if they had one copy of *SMN1*. Sangare et al (2014) did not account for silent carriers. Their low carrier frequency could have been masked by silent carriers. They further observed the presence of multiple copies of *SMN1*, compared to European and Asian populations. Africans were more likely to have three or more copies of *SMN1* (Sangaré et al., 2014). They also investigated *SMN2* copies and found them to be deleted more frequently in the sub-Saharan Africans compared to other populations (Sangaré et al., 2014). Multiple studies have shown that people of African ancestry are more likely to have three or more copies of *SMN1* (Hendrickson et al., 2009).

Furthermore, Sangare et al (2014) aimed to determine the cause of the *SMN* frequency difference in sub-Saharan Africans, and they found hybrid copies of *SMN1/2* in 14% in Malians with three or more copies of *SMN1*. They observed the same pattern

in other African populations they studied. They hypothesised that gene conversion may be a cause of multiple copies of *SMN1*. However, gene conversion does not fully explain the multiple copies of *SMN1* (Sangaré et al., 2014).

1.8. Diagnostic testing at the Division of Human Genetics, Johannesburg

The Division of Human Genetics performs diagnostic, carrier, and prenatal testing for SMA. Diagnostic testing is performed by detecting the absence or presence of a homozygous deletion of *SMN1*. The absence/presence is detecting by amplification of the *SMN1* and *SMN2* gene, followed by digestion with a restriction enzyme that recognises the critical difference between the two genes. The absence/presence of the genes is then visualised by agarose gel electrophoresis. This method does not quantify the genes and only detects exon 7 of both genes. This method cannot determine carrier status, extent of the *SMN1* or *SMN2* deletion or gene conversions.

A recent study performed by Vorster et al (2020) aimed to further understand the molecular basis of SMA in black patients by studying copy number variations in black SA patients. Patients were tested by Multiplex Ligation-dependent Probe Amplification (MLPA, MRC Holland, Amsterdam, Netherlands), which determines the copy numbers of the *SMN1* and *SMN2* genes, as well as the adjacent genes. The MLPA method was also assessed as a second-tier test to the absence/presence method.

The advantage of the study was that adjacent genes in the *SMN* region were included in the kit, therefore, the extent of the deletion could be determined. Gene conversions could be predicted by analysing the extent of the deletion. Only 50.8% of black patients clinically suggestive of having SMA had a homozygous deletion of *SMN1* and 49% of these patients thus remained without a diagnosis. The results presented by Vorster et al (2020) were consistent with those of previous studies by Labrum et al (2007) and Stevens et al (1999). The multiple copies observed in black populations make it difficult to assess carrier frequencies as some individuals may be silent carriers. It was concluded that MLPA is not an appropriate second tier diagnostic test in the black SA population as no additional diagnoses could be unequivocally confirmed (Vorster et al., 2020). It was hypothesized that some of these multiple gene copies may not be completely functional.

However, MLPA may be useful when testing families with informative pedigrees. Pedigrees from families together with copy number may be used to ascertain carrier status, (see Figure 1.8).

A silent carrier is an individual who has two or more copies of the *SMN1* gene on one chromosome and a deletion on the other chromosome. Silent carriers can pass on the chromosome with the deletion and therefore have children who are affected with SMA. The Figure below shows how a silent carrier may be identified when copy number detection of *SMN1*, is carried out on families with an affected individual or child.

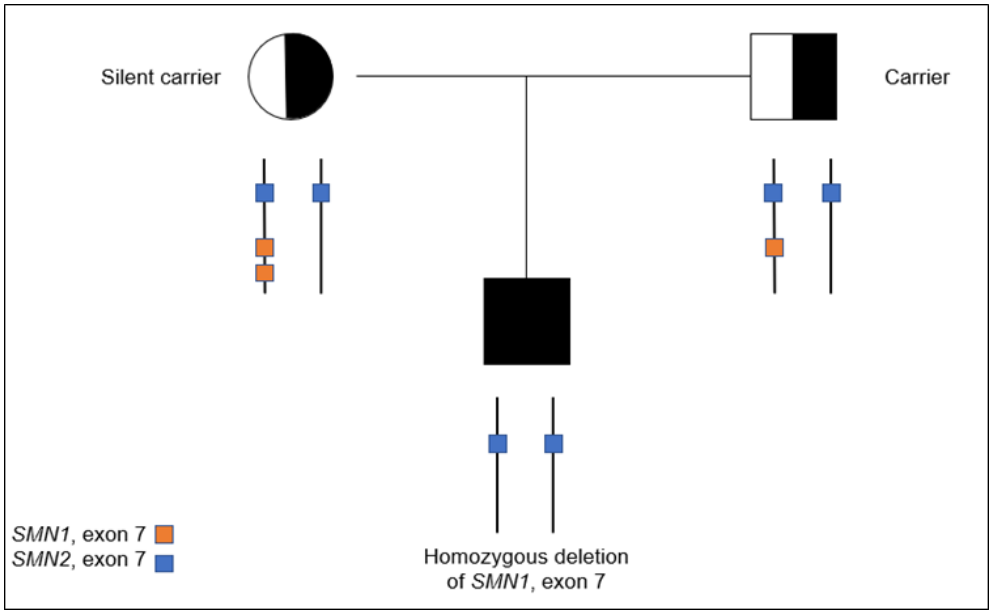


Figure 1.8. Pedigree of a family affected with SMA. The mother has two copies of *SMN1*, however both copies are found on one chromosome. The mother is therefore a silent carrier of SMA. The father has one copy of *SMN1* and is a confirmed carrier of SMA. Their child has zero copies of *SMN1* and two copies of *SMN2*.

The above studies conducted in Africa indicate the need to increase understanding of SMA on the continent and in South Africa specifically. Approximately 49% of black patients with clinical features suggestive of SMA test negative for the homozygous deletion of *SMN1*. These cases could be due to mutations in the *SMN1* gene not detected by routine diagnostic methods, or perhaps other genes affecting expression of *SMN1*. Furthermore, these patients could have a different neuromuscular disease. Nevertheless, further research needs to be performed to elucidate the disease in the many black patients with clinical features of SMA.

1.9. Aim

The aim of the study was to investigate the utility of different methods used to detect *SMN* gene copy number and measure RNA expression in black South African patients with spinal muscular atrophy, and in those with clinical features suggestive of SMA.

1.10. Objectives

- **Objective 1:**

To optimise and validate the AmplideX® PCR/CE *SMN1/2* Kit (Asuragen, Austin, Texas, USA) as an alternative method to determine DNA copy number of SMN genes in patients and controls.

- **Objective 2:**

To optimise and validate qPCR as a method to determine DNA copy number of *SMN1* and *SMN2* in patients and controls.

- **Objective 3:**

To compare the qPCR and AmplideX® PCR/CE *SMN1/2* Kit (Asuragen, Austin, Texas, USA) methods to the already validated/verified MLPA kit (MRC Holland, Amsterdam, Netherlands), to determine the most appropriate approach for DNA copy number detection in black SA SMA patients.

- **Objective 4:**

To optimise and validate reverse transcription of RNA and compare the SMN expression levels with DNA copy number, to identify potential pathogenic copy number variations, and to compare the level of transcripts in patients and controls.

2. SUBJECTS AND METHODS

This chapter will outline the process used to select subjects and control samples and describe the materials and methods used to achieve the aim of the study. The SMA database previously created from MLPA (MRC Holland, Amsterdam, Netherlands) analysis in the Division was used to select samples with various *SMN1* and *SMN2* copy numbers for validation of qPCR and the AmplideX® PCR/CE SMN1/2 Kit (Asuragen, Austin, Texas, USA), a new kit designed by Asuragen (introduced in 2019). Fifty samples were used from the previous study by Vorster et al (2020) and 42 new samples were collected for this study.

Samples of patients with clinical features suggestive of SMA were used for validation of RT-qPCR and comparison of DNA copy number and RNA expression levels of the *SMN1* and *SMN2* genes.

2.1 Subjects and controls

- ***SMN1* exon 7 homozygous deletion controls (M1/M1)**

Patients with a confirmed diagnosis of SMA, homozygous deletion of *SMN1* (genotype M1M1: Mutation 1: deletion of *SMN1*/ Mutation 1: deletion of *SMN1*), results were obtained from previous diagnostic testing. These samples were used as positive controls for validation of qPCR, AmplideX® PCR/CE SMN1/2 Kit (Asuragen, Austin, Texas, USA) and RT-qPCR.

- ***SMN2* exon 7 homozygous deletion controls (M2/M2)**

Samples with a homozygous deletion of *SMN2* (genotype M2M2: Mutation 2: deletion of *SMN2*/ Mutation 2: deletion of *SMN2*), results were obtained from previous diagnostic testing. These samples were used as positive *SMN2* controls for validation of qPCR, AmplideX® PCR/CE SMN1/2 Kit (Asuragen, Austin, Texas, USA) and RT-qPCR.

- **SMA carriers (N/M1)**

Samples with a heterozygous deletion of *SMN1* on MLPA (genotype N/M1: Negative/ Mutation 1: deletion of *SMN1*) were used to determine if RT-qPCR could identify carriers and if so, what range of SMN transcript expression were represented in likely carriers. Table 2.1 illustrates how samples were grouped according to DNA analysis and which controls were used for RT-qPCR.

- **Negative controls (N/N)**

Fresh blood samples were collected from 10 unrelated and unaffected random black SA subjects with no family history of SMA. The CNV status of these subjects was determined by MLPA analysis, and they were included as negative controls in the RT-qPCR method. These samples were used to validate the qPCR method and AmplideX® PCR/CE *SMN1/2 Kit* (Asuragen, Austin, Texas, USA).

- **Patient cohort-Non-deletion (U/U)**

This group forms the focus of this study. Patients of African ancestry who had clinical features suggestive of SMA but who tested negative for a *SMN1* homozygous deletion on DNA studies were identified in collaboration with Prof John Rodda, Head of the Paediatric Neurology Division, Chris Hani Baragwanath Academic Hospital, Johannesburg. The subjects were assumed to have two unidentified mutations (U/U: Unidentified mutation/Unidentified mutation).

Table 2.1. Groups of subjects and controls tested, together with numbers of individuals in each group.

Genotype	Definition	MLPA, AmplideX, qPCR	RT-qPCR	Grand Total
M1/M1	Homozygous deletion of *SMN1, exon 7*: **Affected with SMA**	14	4	**14**
M2/M2	Homozygous deletion of *SMN2,* exon 7	14	1	**14**
N/N	Two or more copies of *SMN1,* exon 7	19	10	**19**
N/M1	One copy of *SMN,1* exon 7: True **Carrier of SMA**	15	2	**15**
U/U	Unidentified mutations, clinical features suggestive of SMA	30	2	**30**
Total tested		92	19	**92**

This table represents the five different subject groups and the number of subjects tested by each method.

*Note: All samples were tested for DNA copy number but not all were tested for RNA expression analysis since RNA was not available from all of these subjects. The total number of samples tested was 92. Each method will be explained fully under methods section.

Previously extracted DNA was stored in the Division and used as part of this study. New blood samples were collected in EDTA and Tempus™ Blood RNA Tubes (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA) from additional black subjects and their parents when available, for DNA and RNA extraction respectively. The strategy for this study is demonstrated by means of a flow diagram in Figure 2.1.

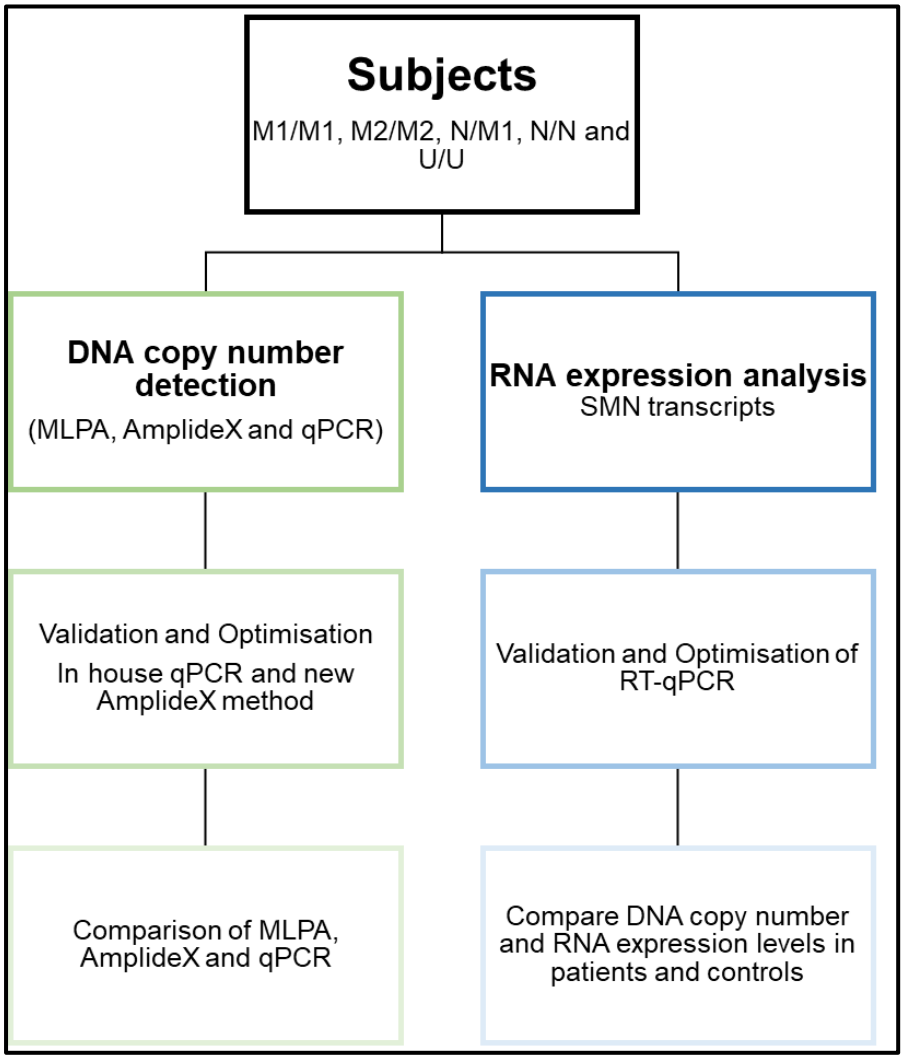


Figure 2.1. Strategy for testing of different methods

2.2 Methods

2.2.1 Genomic DNA Copy Number Detection

The first objective was to determine the utility of alternative methods to detect DNA copy numbers of *SMN1* and *SMN2*. These alternative methods, AmplideX® PCR/CE SMN1/2 Kit (Asuragen, Austin, Texas, USA) and qPCR were first optimised and validated according to NHLS standards. These methods were assessed for cost efficiency, labour intensity, and ability to predict accurate copy number changes. These methods were further assessed to determine if they could provide further insight and diagnosis in patients with suspected but unidentified mutations. To assess the utility of the qPCR and *AmplideX® PCR/CE SMN1/2* kit, both methods were compared with the routine diagnostic test, homozygous deletion of *SMN1* method and MLPA which is used mostly for carrier testing. Figure 2.2 illustrates the procedure followed to determine the diagnostic utility of alternative methods for DNA copy number detection in black SA patients.

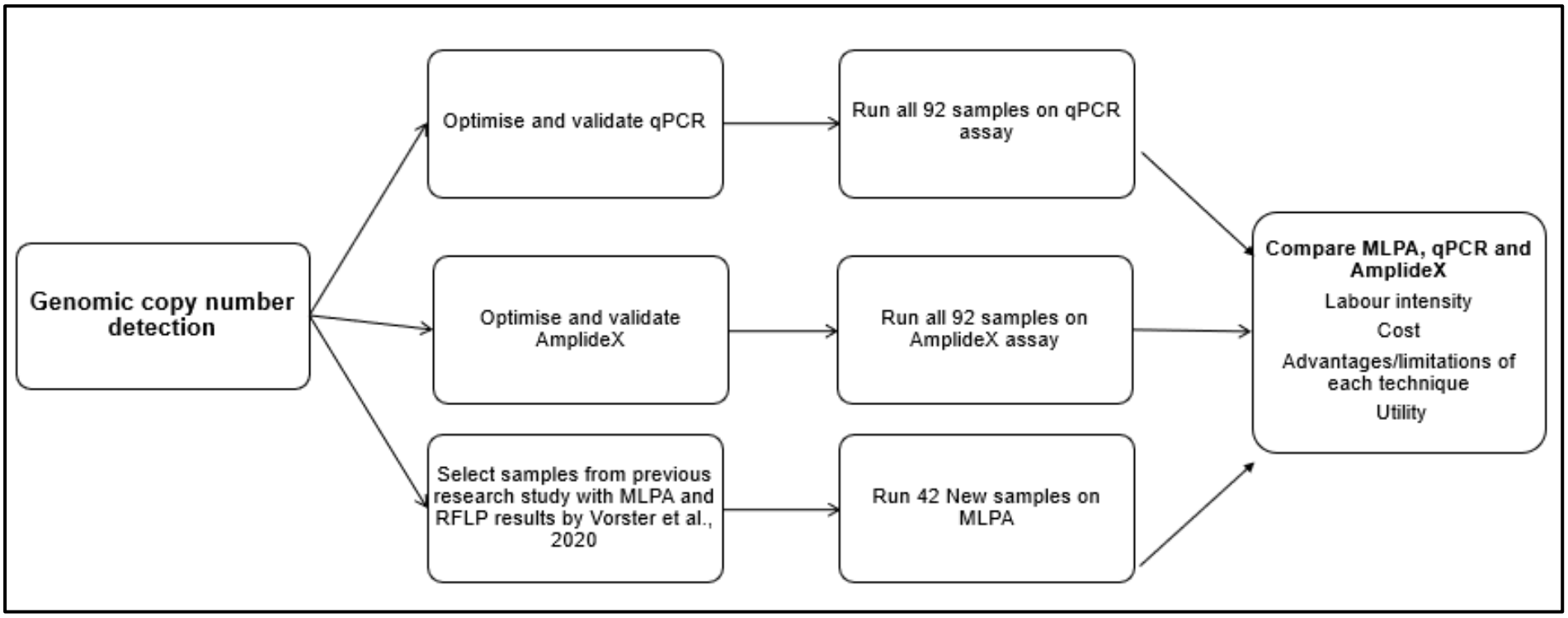


Figure 2.2. Flow diagram showing DNA copy number detection. The *SMN1* and *SMN2* copy numbers were detected by MLPA, qPCR and the AmplideX kit to determine the utility of the methods.

2.2.1.1 DNA Extraction

All the copy number investigations were carried out on DNA extracted from blood samples collected in EDTA tubes. Samples collected before December 2018 were extracted using the salting out DNA extraction method (Miller et al., 1988). Samples collected from January 2019 were extracted using the Flexi Gene DNA kit (QIAGEN Venlo, Netherlands), and the 400µl and 3ml whole blood extraction protocol (Qiagen, 2014). All samples were extracted according to diagnostic protocols of the Division of Human Genetics, see appendix two for DNA extraction worksheet.

2.2.1.2 DNA yield, purity, integrity determination and routine diagnostic testing

DNA yield was measured on the NanoDrop® ND-1000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) with absorbance at 260nm. DNA purity was determined from the A260/A280 ratio (expected 1.7 to 1.9) and A260/A230 ratio (expected 1.8 to 2.0). Samples with impurities were dialysed using Millipore® Filter Membranes (Merk, Darmstadt, Germany). DNA integrity was checked by electrophoresis through a 0.8% agarose gel, and by visualisation under UV light.

2.2.1.3 Homozygous *SMN1/SMN2* deletion method using RFLP.

The method utilises a restriction fragment length polymorphism (RFLP) and is available as part of the routine diagnostic testing procedure in the Division. It was used to determine the absence or presence of *SMN1* or *SMN2* genes. The method utilises the critical one base pair difference in exon 7 of *SMN1* and *SMN2* which can be identified by digesting the PCR product with *Hinf*I endonuclease enzyme. Currently the Division performs RFLP analysis on all samples referred for diagnostic testing and MLPA is used only for carrier testing or where accurate determination of copy number is relevant. The limitations of the method are its inability to test for carriers, copy number of *SMN1* or *SMN2* and to determine the extent of the SMN deletion. The results for the subjects tested were interpreted in comparison with the known control samples, as illustrated in Figure 2.3. All the samples that were referred in for routine diagnostic testing were first assessed on RFLP. Samples without the homozygous

SMN1 deletion, but where the patient had clinical features suggestive of SMA, were then labelled as (U/U).

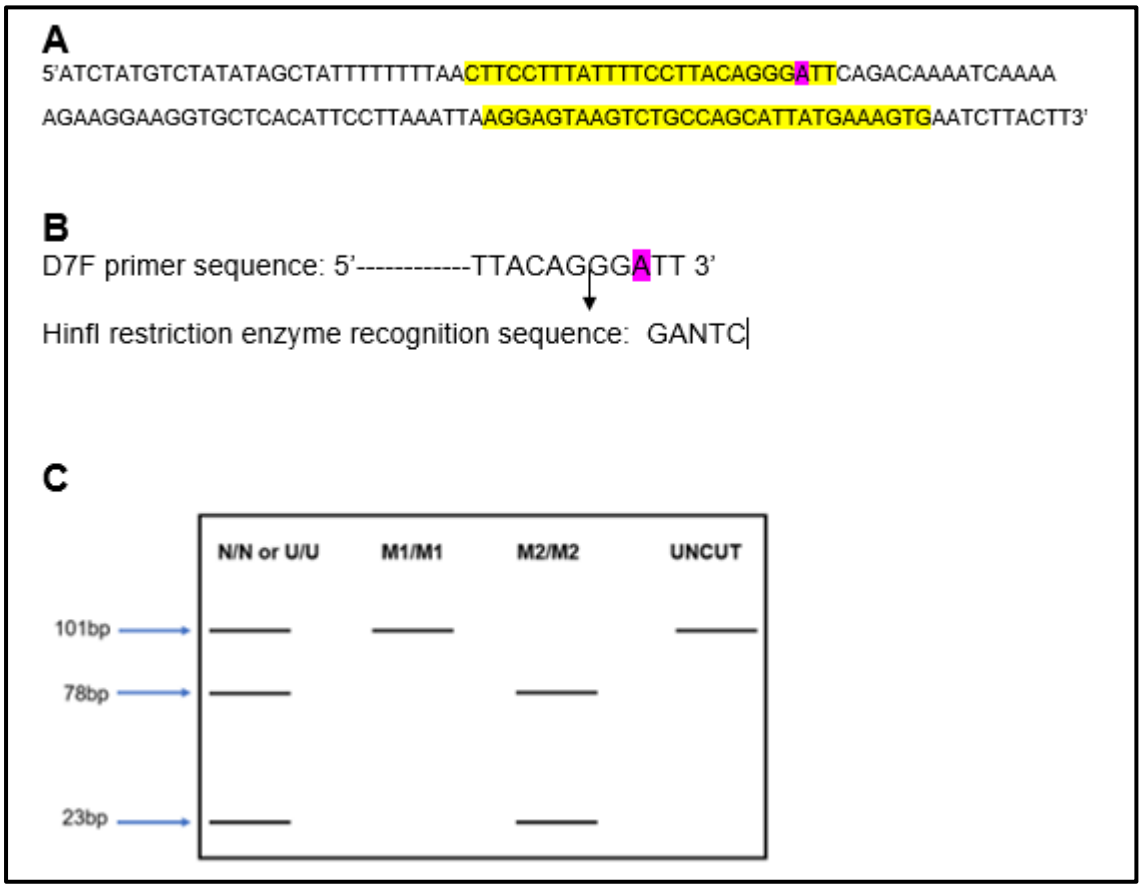


Figure 2.3. Homozygous deletion *SMN1/SMN2* method. **A**, the yellow highlighted region indicates the primer pair mapping to the genomic reference sequence of the *SMN1* and *SMN2* genes; the forward primer has an A introduced instead of T to create a restriction enzyme site. **B**, The *HinfI* recognition site is illustrated. **C**, A 101bp PCR product is amplified and digested with *HinfI* which produces three digest products. N/N or U/U represents subjects who have at least one copy of *SMN1* and *SMN2*, with three bands of 101bp, 78bp and 23bp. M1/M1 represents samples with a homozygous *SMN1* deletion (101bp band only) and M2/M2 represents samples with a homozygous *SMN2* deletion (78bp and 23bp bands only).

2.2.1.4 Multiplex ligation-dependent probe amplification (MLPA)

The MLPA method is a semi-quantitative PCR capable of detecting copy number changes of up to 60 probes in a single reaction. MLPA consists of four basic steps, denaturation, hybridisation, ligation, and amplification (see Figure 2.4 for a summary of the method and appendix 5 for the step-by-step procedure).

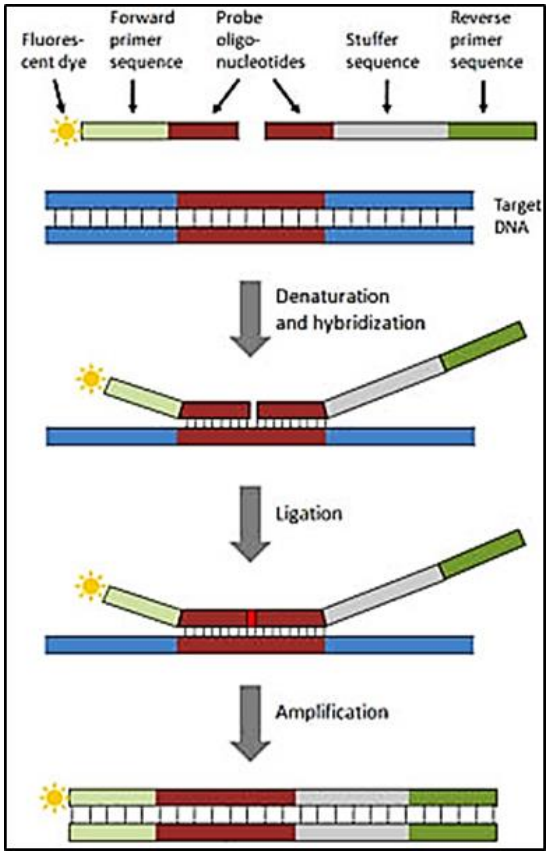


Figure 2.4. Steps in the MLPA test process. Each MLPA kit has up to 60 probes, which are fluorescently labelled with a universal primer sequence and an oligonucleotide specific to the target DNA sequence. The reverse probe oligonucleotide includes a region specific to the target DNA, a stuffer sequence, and a universal reverse primer sequence. Target DNA is denatured and hybridised to disease specific probes, the probes are ligated and amplified by PCR. The PCR products are separated by size using capillary electrophoresis on the 3130*xl* or the 3500*xl* Genetic Analyzer (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA). MLPA analysis was carried out using GeneMapper™ Software (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA) and Coffalyser.Net™ (MRC Holland Amsterdam, Netherlands).(BQUB13-Ecarmona, 2013)

The P021 probe mix

The SALSA MLPA probe-mix P021 (MRC Holland Amsterdam, Netherlands) was used to determine copy number changes of the *SMN1* and *SMN2* genes. This is the kit that will be referred to when mentioning the MLPA method. The P021 probe mix has 37 MLPA probes with amplification products between 140 to 463 nucleotides (nt). Furthermore, the kit contains 9 control probes for DNA quality, denaturation, ligation, and X and Y controls to confirm sex. The kit also contains probes for *SMN1*, exons 7 and 8, *SMN2*, exons 7 and 8, and neighbouring genes *GTF2H2*, *RAD17*, *NAIP*, *SERF1B* and reference regions. MLPA was able to detect homozygous and heterozygous *SMN1* deletions and to determine up to six copies of *SMN1* in black SA

subjects (Vorster et al., 2020). DNA samples were diluted to 40ng and MLPA was carried out according to the one tube protocol described by MRC Holland.

P021 MLPA data analysis

DNA fragments were separated by capillary electrophoresis followed by initial analysis using GeneMapper™ Software (ABI Foster City, California, United States). A GeneScan™ 500 LIZ™ dye size standard (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA) was used to label peaks and determine run quality. All DNA fragments were assessed to confirm proper labelling/binning of sample peaks. Amplified products were further analysed using Coffalyser.Net™ (MRC Holland Amsterdam, Netherlands), with quality checks carried out according to the MRC one tube protocol recommendations. The negative and positive controls were analysed to ensure reproducibility. The *SMN1* and *SMN2* copy numbers were recorded in an Excel© (Microsoft, Redmond, Washington, United States) database and genotypes were assigned for further comparison with the AmplideX kit and DNA qPCR method. Figure 2.5 illustrates a negative subject with no deletions and a subject with a homozygous *SMN1* exon 7and exon 8 deletions.

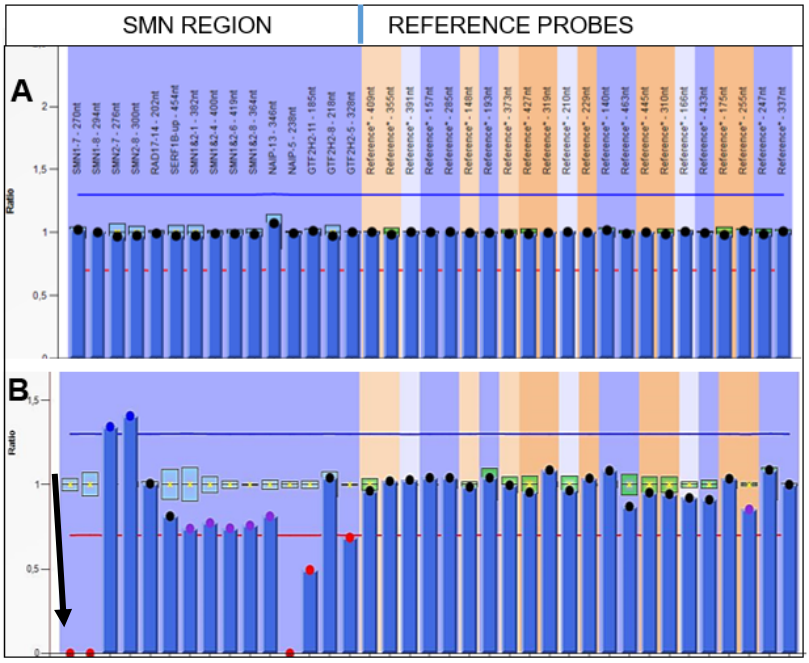


Figure 2.5. Results of MLPA analysis using the SALSA MLPA probe-mix P021. **A**, sample with normal *SMN1* and *SMN2* copy number as well as of adjacent genes. **B**, sample with homozygous *SMN1*, exons 7 and 8 (indicated by the black arrow) deletion and three copies of *SMN2*, exons 7 and 8, suggesting a gene conversion from *SMN1* to *SMN2*. Some variability in other probes is also observed, but these probes are not located in the critical *SMN1*, exon 7 region.

Accurate interpretation of MLPA results depends on the selection of appropriate controls. Three negative controls (with 2 copies of *SMN1* and 2 copies of *SMN2*) were included in each run. A heterozygous *SMN1* deletion (copy number = 1) and a homozygous *SMN1* deletion (copy number = 0) sample were also included as positive controls. The copy number ratio is calculated as a dosage quotient (DQ), which is the final probe height ratio as compared to the internal reference probes of that sample, known as intra-normalisation, and external reference probes (negative controls) for the whole run, known as inter-normalisation. The expected DQ ranges and related copy numbers is indicated in table 2.2.

Table 2.2. The relationship between DQ values and *SMN1* copy number on MLPA analysis ("MLPA General Protocol MDP-v007.pdf," n.d.)

Copy number status	**Copy Number**	**DQ Peak Height**
Homozygous deletion	0 copies	DQ = 0
Heterozygous deletion	1 copy	0.40<DQ<0.65
Identical to reference samples	2 copies	0.80<DQ<1.20
Duplication (phase of copies unknown)	3 copies	1.30 <DQ< 1.65
Triplication (phase of copies unknown)	4 copies	1.75 <DQ< 2.15
Ambiguous result	Variable	All other values

2.2.1.5 AmplideX® PCR/CE SMN1/2 copy number detection.

The AmplideX® PCR/CE SMN1/2 Kit (Asuragen, Austin, Texas, USA) was used to determine *SMN1* and *SMN2* copy numbers in a single reaction. The kit will be referred to as AmplideX for convenience. The kit could serve as an alternative to RFLP and MLPA and was therefore assessed to determine its utility in a diagnostic setting as well as its validity, as it was a new kit. The *AmplideX® PCR/CE SMN1/2* kit contains, 2X PCR mix with dNTPs, polymerase, and buffers, *SMN1/SMN2* and an endogenous control (EC) primers, diluent, calibrator, and control sample. The kit generates copy numbers for both *SMN1* and *SMN2* reported as 0, 1, 2, 3, or ≥4 copies. All subjects were run with a calibrator and a control. All PCR products were loaded with ROX 1000 (Asuragen, Austin, United States) and were analysed by capillary electrophoresis. DNA fragment analysis was done on the GeneMapper™ Software (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA) followed by AmplideX macro-based software for quantification of *SMN* copy numbers. The size standard was checked prior to analyses and accurate labelling of DNA peaks, see appendix 6 for step-by-step procedure. The *SMN* copy numbers were calculated based on the calibrator included in each run. An example of a peak profile is illustrated

in Figure 2.6 and the expected peak sizes are indicated in table 2.3 specific for the 3500xl Genetic Analyzer (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA). *SMN* copy numbers were recorded for comparative studies.

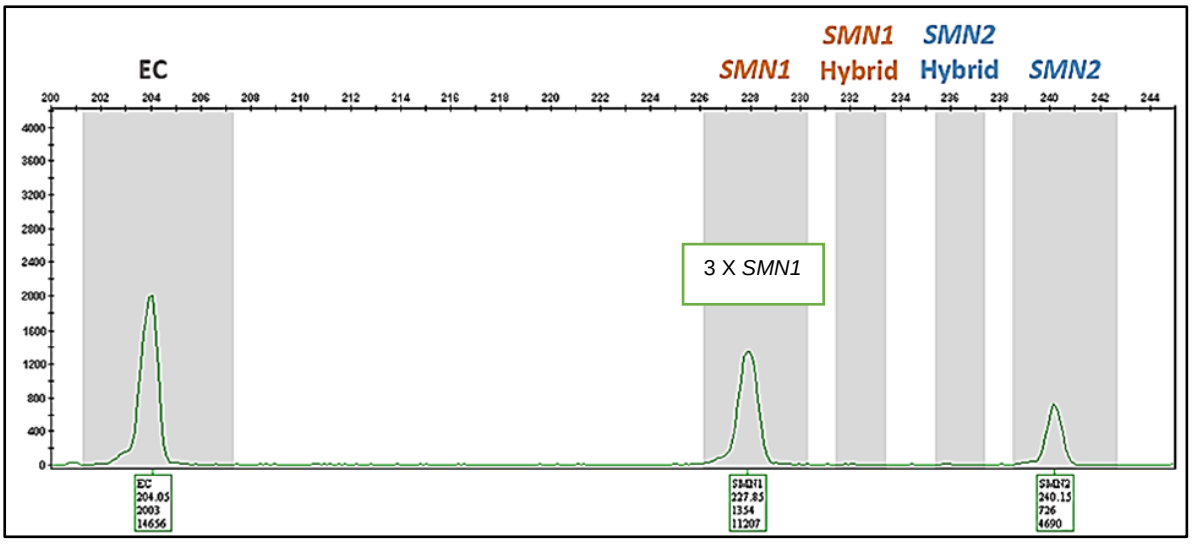


Figure 2.6. Peak profile of a sample with 3 copies of SMN1, exon 7 and 1 copy of SMN2, exon 7 detected by GeneMapper™ Software (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA). The endogenous control (EC) confirms amplification in each reaction.

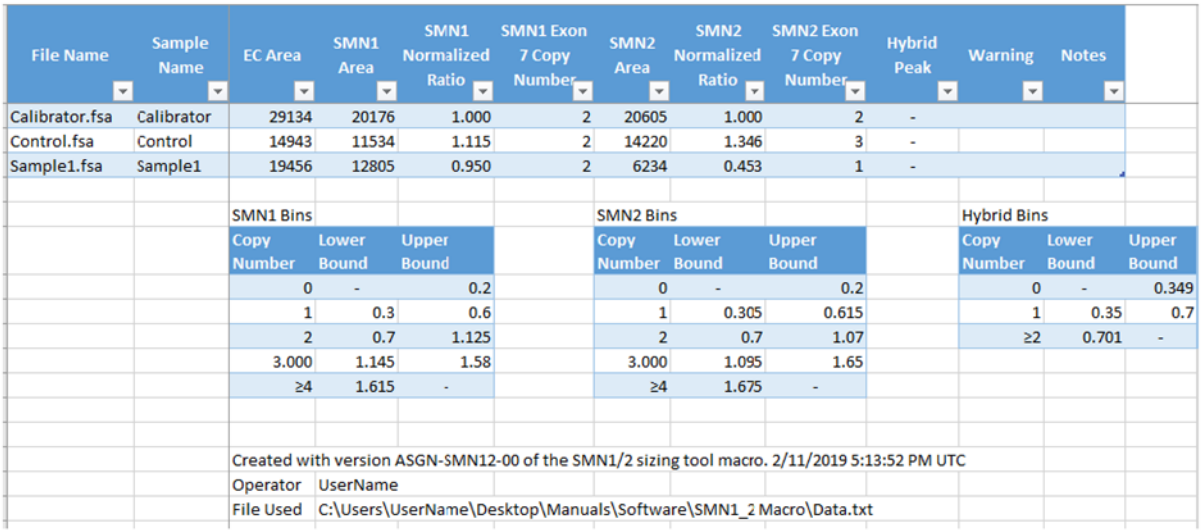
Table 2.3. Expected peak sizes on Applied Biosystems® Genetic Analysers (3500 series) AmplideX PCR/CE SMN1/SMN2 kit protocol guide.

Peak	**Expected Peak Sizes (bp) (3500 Series, 3730 Series)**
EC*	201.30–207.30
SMN1	226.10–230.30
SMN1 Hybrid	231.40–233.40
SMN2 Hybrid	235.40–237.40
SMN2	238.50–242.70

*Endogenous control (EC)

AmplideX® PCR/CE SMN1/2 copy number calculation.

An Excel based macro specifically designed for the analysis of AmplideX data converts peak areas into corresponding *SMN1* and *SMN2* copy numbers. A report was generated in Excel, and an example of an output file is illustrated in figure 2.7. Reference ranges for *SMN* copy numbers were taken from the manufacturer's guidelines as seen in table 2.4 and table 2.5.



File Name	Sample Name	EC Area	SMN1 Area	SMN1 Normalized Ratio	SMN1 Exon 7 Copy Number	SMN2 Area	SMN2 Normalized Ratio	SMN2 Exon 7 Copy Number	Hybrid Peak	Warning	Notes
Calibrator.fsa	Calibrator	29134	20176	1.000	2	20605	1.000	2	-		
Control.fsa	Control	14943	11534	1.115	2	14220	1.346	3	-		
Sample1.fsa	Sample1	19456	12805	0.950	2	6234	0.453	1	-		

SMN1 Bins

Copy Number	Lower Bound	Upper Bound
0	-	0.2
1	0.3	0.6
2	0.7	1.125
3.000	1.145	1.58
≥4	1.615	-

SMN2 Bins

Copy Number	Lower Bound	Upper Bound
0	-	0.2
1	0.305	0.615
2	0.7	1.07
3.000	1.095	1.65
≥4	1.675	-

Hybrid Bins

Copy Number	Lower Bound	Upper Bound
0	-	0.349
1	0.35	0.7
≥2	0.701	-

Created with version ASGN-SMN12-00 of the SMN1/2 sizing tool macro. 2/11/2019 5:13:52 PM UTC

Operator UserName

File Used C:\Users\UserName\Desktop\Manuals\Software\SMN1_2 Macro\Data.txt

Figure 2.7. Example of *SMN1/2* AmplideX Excel report of

Table 2.4. Default *SMN1* and *SMN2* DNA copy number bins vs normalised ratio.

Copy Number Bin	Normalised Ratio	
	SMN1	***SMN2***
0	≤ 0.200	≤ 0.200
1	0.300–0.600	0.305–0.615
2	0.700–1.125	0.700–1.070
3	1.170–1.570	1.095–1.650
≥ 4	≥ 1.615	≥ 1.675

Table 2.5. Default *SMN1* Hybrid and *SMN2* Hybrid DNA copy number bins vs normalised ratio

Hybrid Copy Number Bin	Normalised Ratio
0	< 0.300
1	0.300–0.700
≥ 2	> 0.700

2.2.1.6 Real-time PCR (qPCR)

Quantitative polymerase chain reaction (qPCR) or real time PCR is a method to quantify the amount of PCR product, as it is being amplified (Schefe et al., 2006). There are two qPCR detection methods: TaqMan™ assay (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA) which is sequence specific and the SYBR green method which is based on incorporation of generic non-sequence-specific double-stranded DNA-binding dye. The four stages of qPCR are: linear ground, early exponential, linear exponential phase (log) and plateau phases. The fluorescent signal is calculated during the last phase (Schefe et al., 2006). The method is used for multiple applications such as, genotyping, copy number detection, gene expression etc. In the present study the method was used with TaqMan™ probes

for copy number detection of *SMN1* and *SMN2*. An illustration of TaqMan™ qPCR steps is shown in Figure 2.8, starting with denaturation, annealing of primers and probe, followed by extension of primers and cleavage of fluorophore (dye) and repetition of cycles.

The method was selected for evaluation as an alternative method for diagnostic, carrier, and prenatal testing of SMA. Ten samples were selected for optimisation and validation of qPCR, those selected included three negative controls (N/N), three positive controls (M1/M1), two samples heterozygous for *SMN1* (N/M1) and two samples with a homozygous deletion for *SMN2* (M2/M2). Once qPCR was validated other samples were tested to determine the utility of this method.

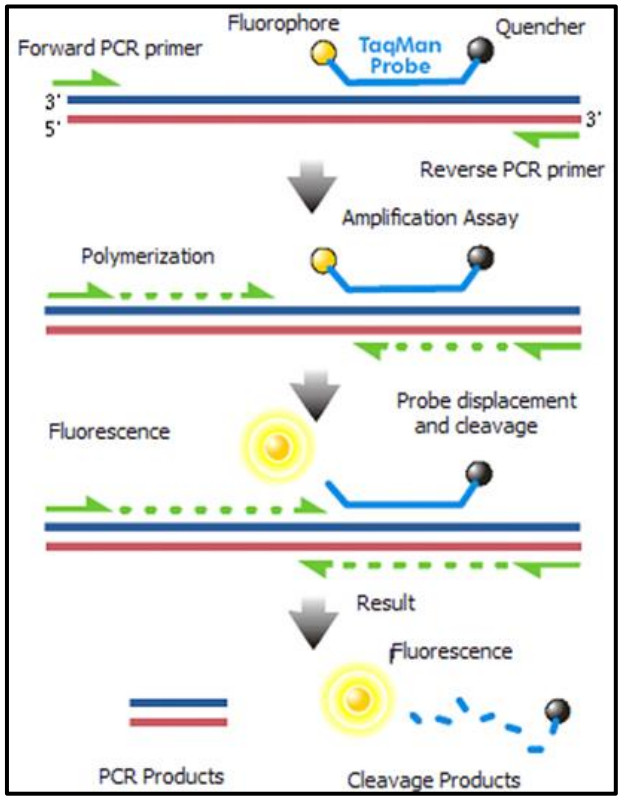


Figure 2.8. Quantitative PCR by use of TaqMan™ probes (Anhuf et al., 2003). Each qPCR method contains a forward and a reverse primer for the target DNA sequence. A TaqMan™ probe has a fluorophore (FAM, VIC, NED etc) on the 5' end which emits light upon excitation. The quencher is located on the 3' end of the TaqMan™ probe and absorbs the fluorescence of the fluorophore. The DNA is denatured into single strands followed by the annealing of primers and TaqMan™ probes to target sequences. This is followed by extension of the primers by DNA polymerase. Once

extension reaches the bound TaqMan™ probe the fluorophore is cleaved and as the fluorophore and quencher are not in close proximity, light is released and is detected by the qPCR machine. The target sequence is amplified, and the fluorescence emitted from the fluorophore is quantified to determine copy number.

SMN1 gene copy qPCR

All DNA samples were diluted to 50 ng/µl. Three negative (N/N) controls were included with each run on the QuantStudio 3 (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA). Primers and TaqMan™ probes were used for *SMN1*, *SMN2* and the *CFTR* gene, which was used as an internal control, as illustrated in Table 2.6. The primers and probes were previously designed as part of a Master's project (Vorster et al., 2020). The primers for *SMN1* and *SMN2* were the same, however the probes for the two genes were different, as illustrated in table 2.6 and Figure 2.9. Furthermore, the CFTR gene was used as an internal control and primers were designed as indicated in table 2.6 and Figure 2.10. See appendix seven for step-by-step procedure.

Table 2.6. Primer and probe sequences of the *SMN1* and *SMN2* qPCR method

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

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

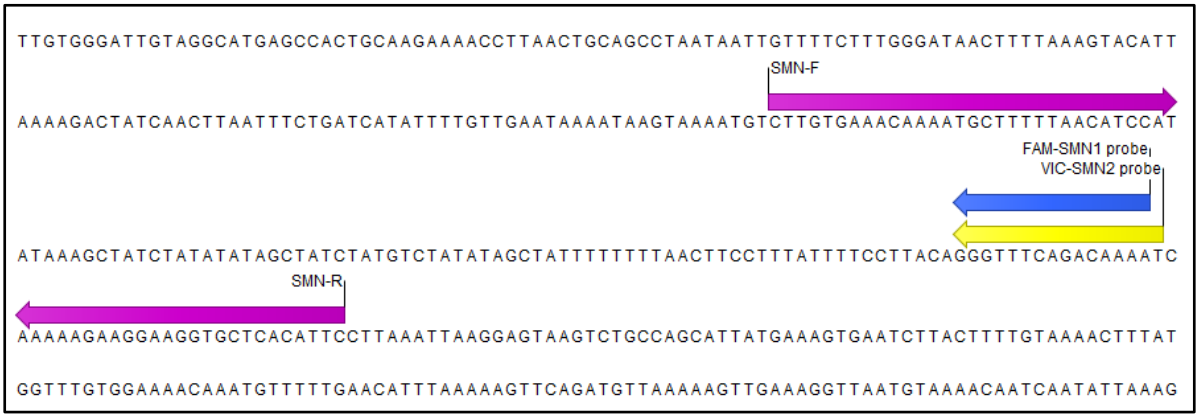


Figure 2.9. Annotation of *SMN* primers, (*SMN1*- NG_008691.1 and *SMN2*- NG_008728.1). A single primer pair SMN-F and SMN-R illustrated in purple was used to amplify both *SMN1* and *SMN2*, exon 7 template DNA. Each *SMN* copy had its own specific probe, blue (FAM) for *SMN1* and yellow (VIC) for the *SMN2* probe. Primer and probe sequences are detailed in table 2.6.

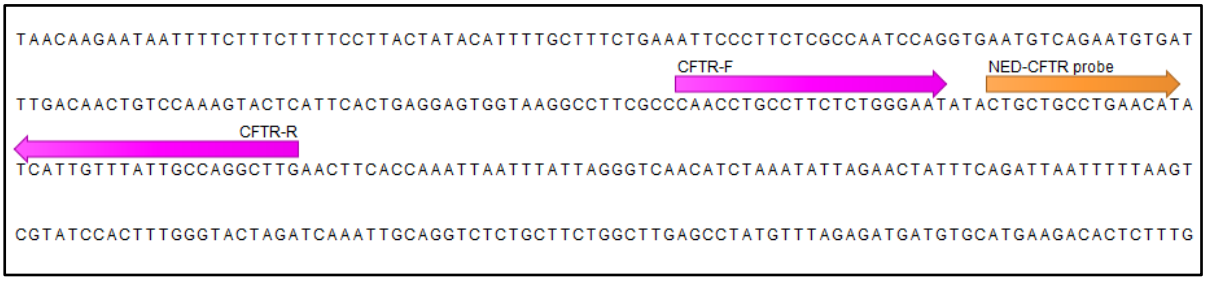


Figure 2.10. Annotation of *CFTR* primers, (NG_016465.4). A single primer pair CFTR-F and CFTR-R in pink and the *CFTR* probe in orange. Primer and probe sequences detailed in table 2.6.

SMN1 and SMN2 qPCR data analysis

The QuantStudio design and analysis software v1.4.1 (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA) was used to create the experiment and import run file onto the QuantStudio 3 (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA). The software, freely available for download from the ThemoFisher website, was designed specifically for the QuantStudio qPCR instruments to create a run, sample worksheet, PCR conditions and analysis of experiment results. Each experiment was set up in triplicate. Amplification data were imported into the QuantStudio design and analysis software v1.4.1 (Applied biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA), and standard curve analysis was used for copy number detection. The amplification plot was analysed to ensure adequate amplification followed by assessment of quality control checks as recommended by the manufacturer.

The cycle threshold (Ct) is defined as the number cycles needed for the fluorescent signal to exceed the background level. The Ct was calculated after PCR amplification, outliers were omitted from analysis and results were reviewed. The copy numbers were determined using relative quantification calculations for relative standard curve and comparative Ct experiments as indicated in Figure 2.11 and Figure 2.12. The *SMN* copy number was measured by determining the coefficient of variation (CV). The CV for each sample was calculated by dividing the standard deviation (SD) of three negative controls in a given subject by the average *SMN* copy number for that subject (Gómez-Curet et al., 2007; Passon et al., 2009).

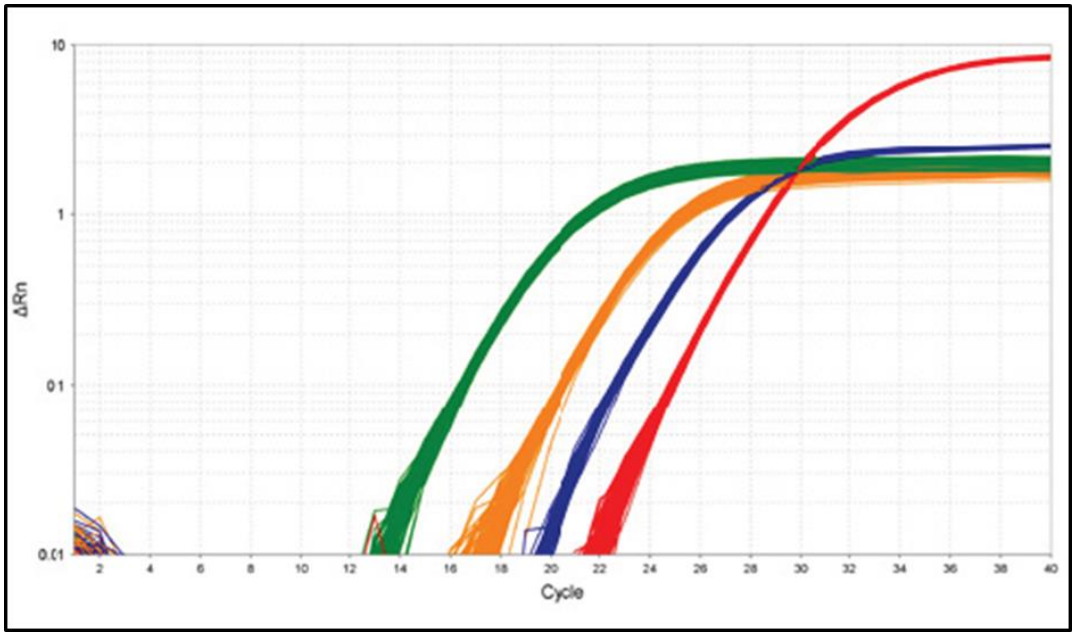


Figure 2.11. An example of a qPCR amplification plot visualised on the QuantStudio 3. Each colour represents a different concentration of DNA used.

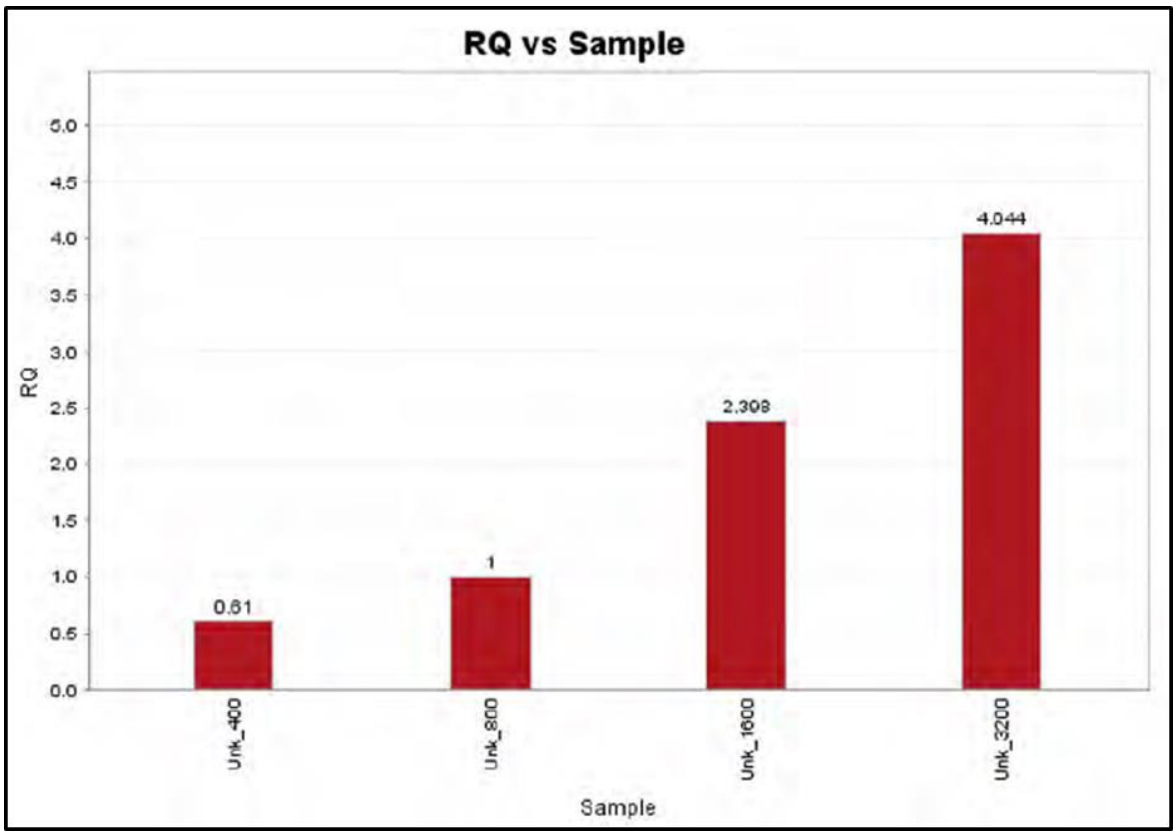


Figure 2.12. Example of a gene copy number plot visualised on the QuantStudio 3. The plot was used to evaluate the level of different samples relative to the reference sample. The fold change is indicated as numbers on top of each sample. Four samples were analysed and the relative quantification (RQ) to the internal control is indicated above each sample.

The qPCR method was more informative than the RFLP, as it determines the DNA copy number of *SMN1.* It therefore detects heterozygous carriers of SMA and is less expensive than MLPA. The qPCR method detected the homozygous *SMN1* deletion, heterozygous deletions of *SMN1* (carriers) and multiple copies of *SMN1* and *SMN2* (Anhuf et al., 2003).

2.2.2 *SMN1/SMN2* RNA expression analysis

Gene expression studies were undertaken to provide an alternative molecular method for testing patients who present with clinical features of SMA but have tested negative for a homozygous *SMN1* deletion using DNA analysis. Reverse transcription quantitative PCR (RT-qPCR) was used to determine mRNA levels of the full length SMN1 (FL-SMN1), full length SMN2 (FL-SMN2) as well as the SMN mRNA lacking exon 7 of both *SMN1* and *SMN2* which was labelled as SMN∆7 (illustrated in table 2.9). The design of primers and probes is outlined below.

Blood was collected in specialised Tempus™ Blood RNA Tube (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA). RNA extraction was performed using the Tempus™ Spin RNA Isolation Kit (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA). See appendix three for RNA extraction procedure and worksheet.

Reverse transcription of RNA into cDNA was performed using the ImProm-II™ Reverse Transcription System (Promega, Madison, Wisconsin, USA). The cDNA was amplified using TaqMan™ probes and RNA expression levels were determined with the comparative Ct experiments with an endogenous control Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) included in each experiment.

Table 2.7. Description of RT-qPCR transcripts

Transcript abbreviation	**Explanation**
FL-SMN1	Full length SMN1
FL-SMN2	Full length SMN2
SMN∆7	SMN transcript lacking exon 7
GAPDH	Endogenous control

2.2.2.1 RNA extraction

Blood was collected in Tempus™ Blood RNA Tubes (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA), so that RNA was stabilised, and RNA expression patterns preserved. Tempus RNA tubes were stored at -70°C. RNA extraction was done was using a Tempus™ Spin RNA Isolation Kit (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA) according to the manufacturer's guidelines. A summary of RNA extraction performed is shown in Figure 2.13 (Duale et al., 2012). RNA yield was measured on the NanoDrop® ND-1000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) with an absorbance at 260nm. RNA integrity was detected on 0.8% agarose gel electrophoresis and visualised by UV. See appendix three four RNA extraction procedure.

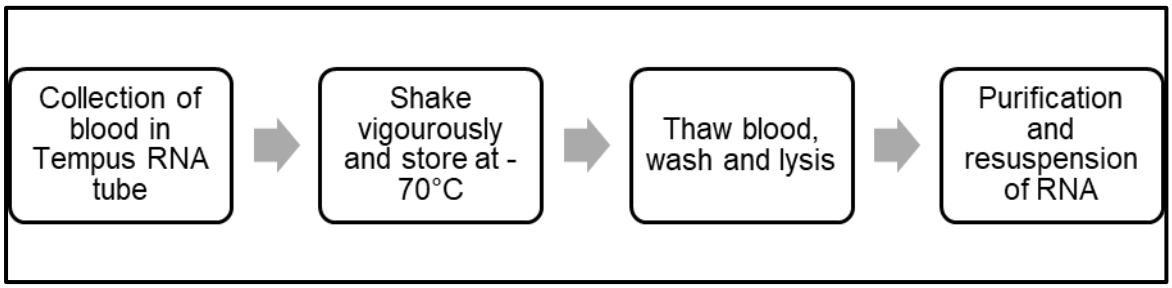


Figure 2.13. Summary of RNA extraction protocol using Tempus™ Blood RNA Tube (Applied biosystems, Thermo Fisher scientific, Waltham, Massachusetts, USA).

2.2.2.2 Reverse transcription of RNA to cDNA

Reverse transcription from freshly collected RNA was performed using the ImProm-II™ Reverse Transcription System (Promega, Madison, Wisconsin, USA), Figure 2.14 (Duale et al., 2012). An endogenous control, *GAPDH*, was used to normalise SMN expression levels using relative quantification. See appendix four for the step-by-step reverse transcription procedure. A worksheet was created for each test and was tabulated using appendix nine.

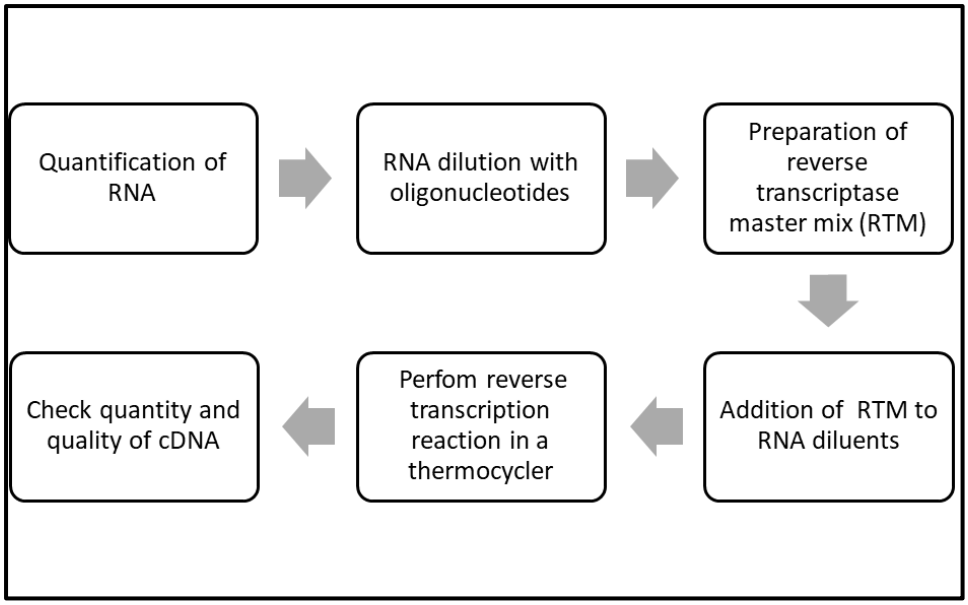


Figure 2.14. Summary of reverse transcription using the ImProm-II™ Reverse Transcription System (Promega, Madison, Wisconsin, USA). Each reaction included a positive control, negative no transcriptase enzyme, no template RNA control and a negative no template RNA control.

2.2.2.3 RT-qPCR experiment design

Table 2.8 and Figure 2.15 illustrates the transcript primers used and their probes. The experiment was carried out in quadruplicate with three controls, each control had two DNA copies of *SMN1* and two DNA copies of *SMN2*. The *GAPDH* endogenous control was included in each experiment to measure the relative gene expression. Analysis of results was carried out on the QuantStudio design and analysis software (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, USA).

Table 2.8. Primer and probe sequences of the *SMN* expression RT-qPCR relative quantification method designed by Vorster et al (2020)

Transcript	Primers	Primer Sequence (5'-3')	Probe	Probe Sequence (5'-3')
FL-SMN1	FL-SMN-F	TACATGAGTGGCTATCATACTGGCTA	FL-SMN1	NED-TATGGGTTT**C**AGACAAA
	FL-SMN-R	AATGTGAGCACCTTCCTTCTTTTT		
FL-SMN2	Same as FL-SMN1		FL-SMN2	VIC-ATATGGGTTT**T**AGACAAAA
SMNΔ2	SMNΔ7-F	CTGATGCTTTGGGAAGTATGTTAATT	Ex7del-SMN	NED-CATGGTACATGAGTGGCTA
	SMNΔ7-R	CCAGCATTTCTCCCATATAATAGCCAGTA		
GAPDH	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.**

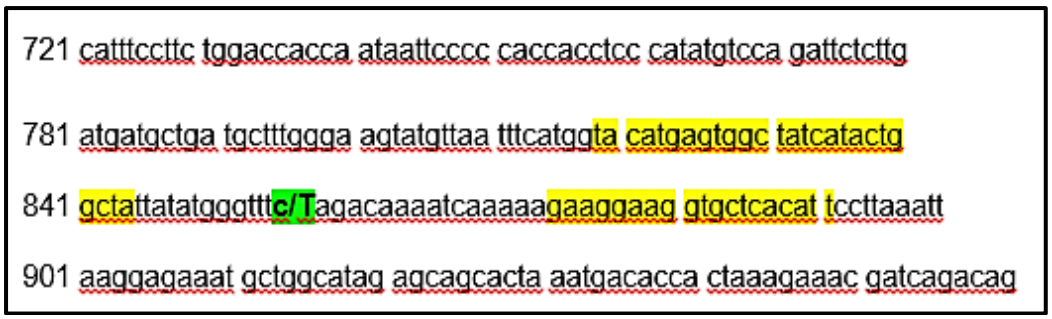


Figure 2.15. Primer and probe sequences of FL-SMN1 (NM_000344.4) and FL-SMN2 (NM_017411.4). The FL-SMN forward and reverse primers are indicated in yellow. The same primers were used to amplify both SMN1 and SMN2 full length transcripts. The probes used to distinguish FL-SMN1 and FL-SMN2 transcripts had a critical nucleotide difference indicated in green.

2.2.3 Comparative analysis of DNA and transcript levels of *SMN1* and *SMN2* in subjects and controls

Comparison of expression levels of the SMN transcripts with the DNA copy number was performed to identify potential pathogenic copy number variations in patients with clinical features suggestive of SMA, who have tested negative for a homozygous deletion of *SMN1*. Figure 2.15 illustrates an approach to compare *SMN1* and *SMN2* gene copy number with RNA expression levels in this study. Please note that this Figure may not necessarily be an accurate representation of the situation in black South African subjects, but it presents a guideline to assist with comparative analysis of DNA copy number and RNA expression levels.

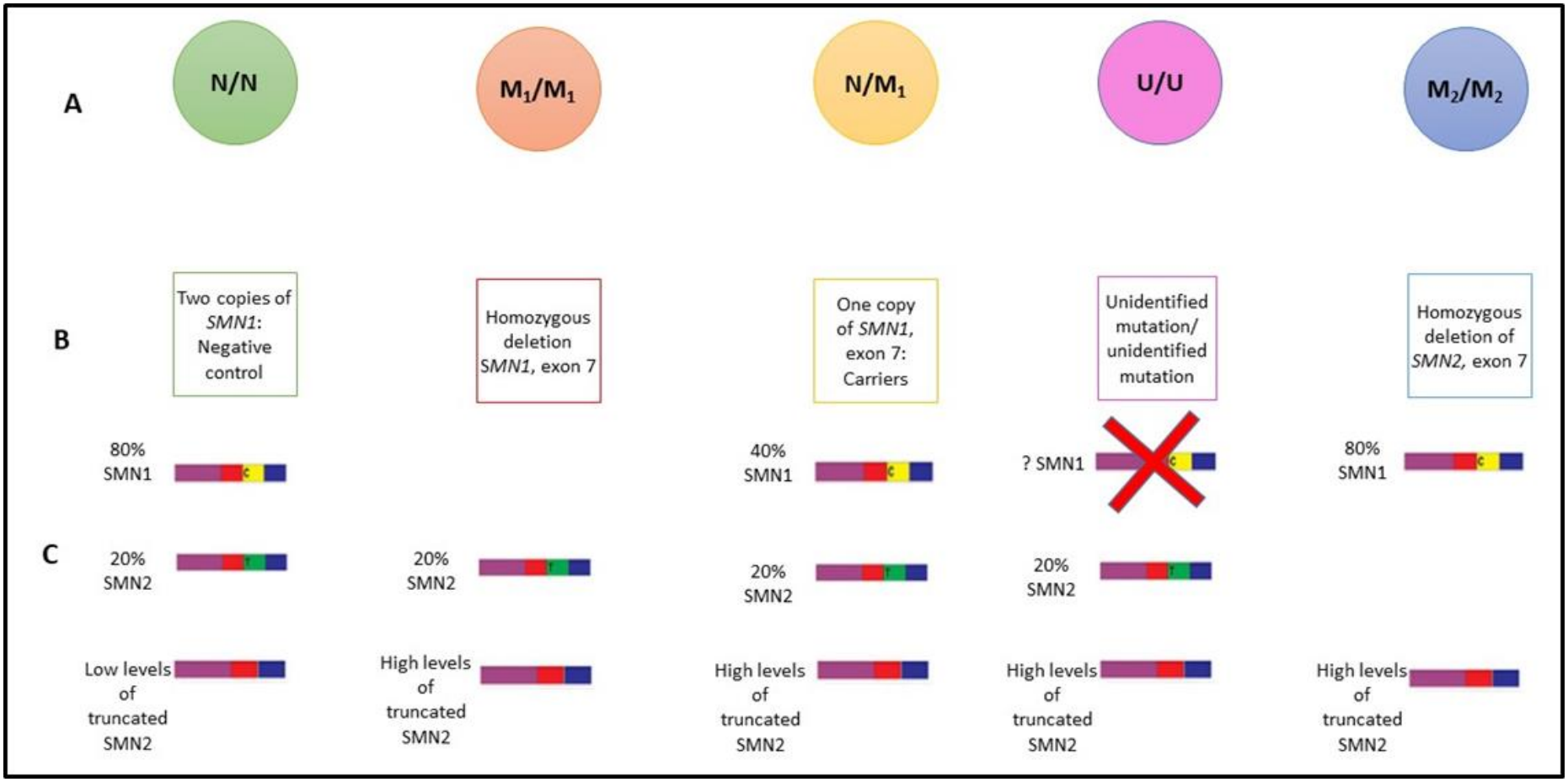


Figure 2.16. Predicted comparative analysis of DNA *SMN1* and *SMN2* copy number versus RNA transcription. **A** indicates the genotypes that may be found from DNA analysis, N indicates a normal SMN1 copy, M1 is a deletion of *SMN1*, exon 7 and U indicates an unidentified mutation. **B** indicates the genotype of subjects and controls. **C** indicates the potential *SMN1* and *SMN2* RNA expression predicted to occur according to each genotype. The RNA expression level of *SMN1* in subjects with M1/U and U/U genotypes is unknown. Three possible transcripts were measured by RT-qPCR - FL-SMN1, FL-SMN2 or truncated SMNΔ7. The percentages of each transcript are estimated according to general S*MN1* and *SMN2* copy number, these percentages may differ in the SA population as they have hypervariable *SMN1* and *SMN2* CNVs.

2.2.4 Ethics

An ethics application was approved unconditionally by the University of the Witwatersrand Committee for Research on Human Subjects (Medical) (Ethics clearance number: M180498) see appendix 1 for certificate.

2.2.5 Summary of methods

Three methods for quantification of *SMN1* and *SMN2* DNA copy numbers were evaluated for their clinical utility in diagnostic and carrier testing of SMA. The RFLP methods detects the presence or absence of both *SMN1* and *SMN2* DNA copies. Subjects previously tested by RFLP who had a homozygous deletion of either *SMN1* or *SMN2* were used as M1/M1 and M2/M2 controls, respectively. All subjects' *SMN* copy numbers were quantified by MLPA, AmplideX and qPCR. The AmplideX kit is a new commercial kit for testing SMA. The kit was validated as part of this research study. Another method qPCR was used to detect *SMN1* and *SMN2* copies and compared with the already utilised P021 Probe mix for MLPA and the new AmplideX kit. RNA expression of 19 subjects who were also tested by MLPA, AmplideX and qPCR was performed. SMN transcripts were quantified to determine the relationship between DNA copy number and RNA copy number as well as to possibly diagnosis subjects with SMA who do not have a typical homozygous deletion of *SMN1*. These suggestive SMA subjects were hypothesised to have no *SMN1* transcripts on RNA expression analysis.

3 RESULTS

This chapter outlines results obtained from DNA copy number analysis (section 3.1) as well as RNA expression studies (section 3.2). Three methods designed to detect DNA copy number were compared – MLPA analysis, an in house designed qPCR-based method and a commercial kit AmplideX® PCR/CE SMN1/2 kit (Asuragen, Austin, United States). The qPCR method and the *AmplideX® PCR/CE SMN1/2* kit have not been previously tested in the Division and were optimised and validated for copy number quantification of *SMN1* and *SMN2*. Both methods were compared with the P021 SALSA MLPA kit (MRC Holland Amsterdam, Netherlands), which is the recommended method for diagnostic testing of SMA and is the current method used for diagnostic testing in the Division when copy number determination is required. The purpose of the comparison was firstly to select the method with the best utility in SA populations. Further, the methods were evaluated to assess whether they would aid in understanding the complex mechanism of SMA disease, specifically in the black SA population, who appear to have a complex *SMN1/SMN2* architecture which complicates diagnostic testing. The AmplideX kit (Asuragen, Austin, United States) and P021 SALSA MLPA kit (MRC Holland Amsterdam, Netherlands) kits will be referred to as AmplideX and MLPA going forward for simplicity.

This chapter will further present data from the validation and optimisation of a RT-qPCR method designed to determine the gene expression of three different SMN transcripts. Since the analysis of DNA copy number methods has previously been shown to be complicated by the complex nature of the *SMN* region, especially in African populations, a gene expression method was investigated as an alternative or second line method. Transcripts levels will be compared with DNA results to determine the relationship between DNA copy number (DNA) and gene expression (RNA).

3.1 DNA copy number detection: MLPA, AmplideX and qPCR

A total of 92 subjects were tested using MLPA, qPCR and AmplideX methods. There were 71 samples from black subjects (71/92, 77.2%) and 21 samples from white subjects (21/92, 22.8%) (as shown in Table 3.1). The subjects were categorised in five different genotype groups - M1/M1, M2/M2, N/N, N/M1 and U/U. Subjects were

labelled according to Table 2.1 in the “Subjects and Methods” section. Because of the concern that the diversity of copy number in black SA populations could complicate optimisation and analysis, previously tested white subjects were also used to validate and optimise the new methods: qPCR and AmplideX.

Table 3.1. Ethnicity and genotype of subjects tested for DNA copy number.

Ethnicity	Genotype*	Number of subjects tested
Black SA	**Total black SA subjects**	**71**
	M1/M1	12
	M2/M2	10
	N/M1	9
	N/N	10
	U/U	30
White SA	**Total white SA subjects**	**21**
	M1/M1	1
	M2/M2	4
	N/M1	7
	N/N	9
Total		**92**

* Genotypes defined in Section 2.1.1 – 2.1.5

3.1.1 DNA copy number detection by MLPA

Fifty subjects of various genotypes were previously tested on MLPA as part of a Master’s project and a further 42 subjects were included in this project as part of validation and optimisation of the new methods (Vorster et al., 2020). Figure 3.1 illustrates examples of typical results obtained using MLPA analysis. The P021-A1 kit has 37 probes, 11 of which are specific to the *SMN* region as defined in section 2.2.1.4. MLPA analysis focussed on *SMN1,* exons 7 and 8 as well as *SMN2,* exons 7 and 8. A deletion or duplication of additional probes in the region could provide details about the possible extent of deletions or duplications as well as the presence of gene conversions. For the purposes of the study only exon 7 of *SMN1* and *SMN2* was analysed and quantified and will therefore be referred as *SMN1* or *SMN2*.

M1/M1 subjects - MLPA DNA copy number detection

A total of 13 M1/M1 subjects were included in the study (12 black, 1 white). These subjects are affected with SMA and were used as positive controls. All 13 had a homozygous deletion of *SMN1,* exon 7 (Table 3.2). However only 4/13 (31%) had a homozygous deletion of *SMN1,* exon 7 and exon 8. Deletion of both exons in the *SMN1* gene illustrate the extent of the deletion as well as the complexity of the *SMN*

region. A subject with a homozygous deletion (zero copies of *SMN1*) is discussed in Figure 3.1.D.

Table 3.2. Comparison of *SMN1*, exon 7 and exon 8 copy number in M1/M1 subjects

Disease Code	Genotype	MLPA *SMN1*, exon 7	MLPA *SMN1*, exon 8
SMA677*	M1/M1	0	0
SMA762	M1/M1	0	2
SMA737*	M1/M1	0	0
SMA995	M1/M1	0	2
SMA1515	M1/M1	0	1
SMA1838*	M1/M1	0	0
SMA1949	M1/M1	0	≥3
SMA1958	M1/M1	0	2
SMA1978	M1/M1	0	2
SMA1988	M1/M1	0	2
SMAR7	M1/M1	0	≥3
SMAR15	M1/M1	0	1
SMAR18*	M1/M1	0	0
Total	13	13	13
Percentage of deletions extending to exon 8 = zero			4/13 (31%)
Percentage of exon 8 greater than one			9/13 (69%)

*Indicates the four subjects with zero copies of SMN1, exons 7 and 8.

M2/M2 subjects - MLPA DNA copy number detection

Fourteen subjects (10 black, 4 white) with a homozygous deletion of *SMN2* were tested, these subjects were not affected with SMA. Subjects with the M2/M2 genotype often had more than two copies of the *SMN1* gene, possibly due to gene conversion from *SMN2* to *SMN1.*

N/M1 subjects - MLPA DNA copy number

Sixteen subjects (9 black, 7 white) with an N/M1 genotype were included in this study. All subjects were previously confirmed to be carriers of one copy of the *SMN1*, exon 7 deletion (M1 mutation, 1:0 genotype). Figure 3.1 C illustrates an N/M1 subject who had one copy of *SMN1*, exon 7 and is therefore a true carrier of SMA. In total 12/16 (75%) of the N/M1 subjects tested had one copy of *SMN1* and 4/16 (25%) had two copies of *SMN1*. All four of the subjects with two copies of SMN1 (2:0 genotype) were black and were previously confirmed silent carriers by use of a family pedigree (see section 1.8 and Figure 1.10 for a definition and illustration of a silent carrier, respectively).

N/N subjects - MLPA DNA copy number detection

A total of 19 N/N subjects (10 black, 9 white) were included in this study. Figure 3.1 A shows a N/N subject with two copies of *SMN1* and *SMN2*.

Multiple copy numbers of *SMN1* were observed in 4/10 (40%) of black and in 0/9 (0%) white N/N individuals. All subjects with three or more copies were given a single value of ≥3 to standardise copy number across the three methods as MLPA does not accurately quantify copy numbers of more than three. Figure 3.1B shows one such subject with four copies of the *SMN1* gene and two copies of the *SMN2* gene. Other exons (1, 4, 6 and 8) in the *SMN1/SMN2* genes were increased, however the probes binding to this region are not able to distinguish *SMN1* and *SMN2* and were thus not analysed further in this study. Both subjects A and B have no family history of SMA and do not present with any features of SMA, therefore illustrating the variation that can be seen in normal black subjects.

U/U subjects - DNA copy number detection using MLPA

Thirty black subjects with the U/U were tested and did not differ visually/statistically from the N/N individuals and appeared similarly to N/N individuals, as illustrated in Figure 3.1 B.

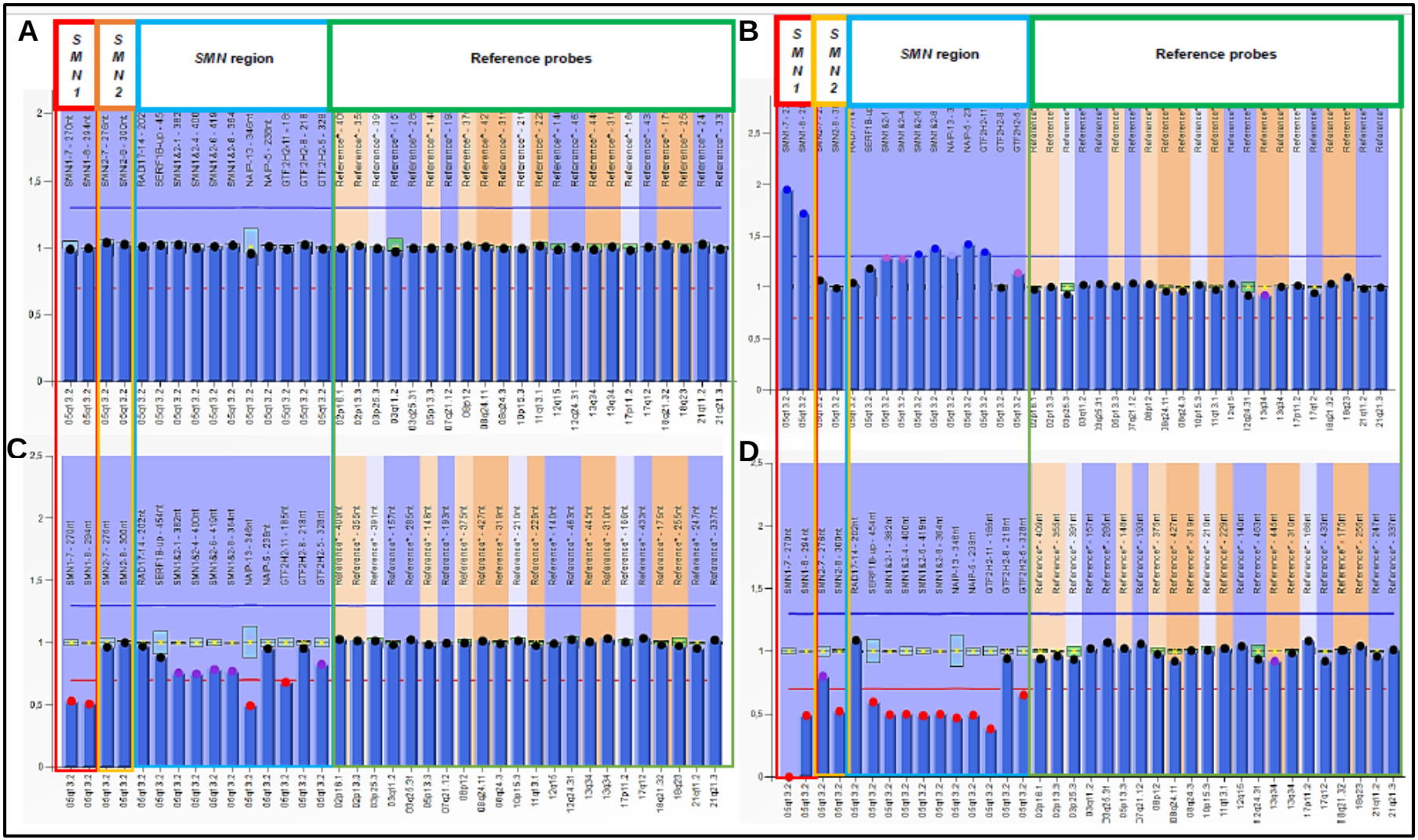


Figure 3.1. MLPA results obtained for the *SMN* region.: **A**. White subject with an N/N genotype with two copies of *SMN1* and two copies of *SMN2*. **B**. Black subject with an N/N genotype with four copies of *SMN1* and 2 copies of *SMN2*. **C**. Subject with an N/M1 genotype with one copy of *SMN1* and two copies of *SMN2*. **D**. Subject affected with SMA with a M1/M1 genotype with zero copies of *SMN1* and two copies of *SMN2* exon.

3.1.2 DNA copy number detection by AmplideX

The AmplideX kit is a commercial kit that detects copy number of exon 7 (not exon 8) of the *SMN1* and *SMN2* genes. The kit was first optimised and validated, followed by testing on the groups of subjects as described in Sections 2.1. The total setup could be done within a day, from dilution to amplification and separation of fragments by capillary electrophoresis. Analysis of data generated by the AmplideX kit was first done using the GeneMapper™ Software to accurately determine the correct size of PCR products and to evaluate the quality of the data. An Excel-based macro designed by Asuragen AmplideX was used to calculate the actual copy numbers. Subjects who had AmplideX results which conflicted with MLPA results, were reloaded on the 3500xl Genetic Analyzer and if results were still conflicted, the test was repeated.

Figure 3.2 illustrates examples of typical results produced by the AmplideX kit. Subject A is a white N/N subject with two copies of both the *SMN1* and *SMN2* genes and is unaffected with SMA. Figure 3.2 B and C indicate subjects with a homozygous deletion of *SMN1* and a homozygous deletion of *SMN2,* respectively.

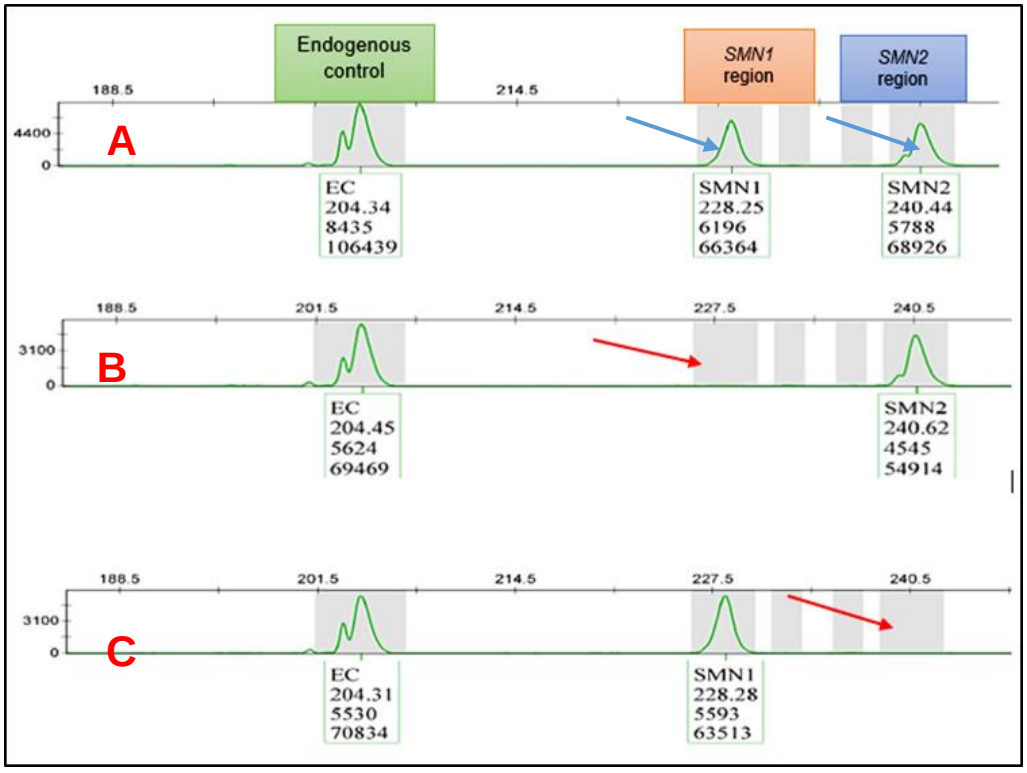


Figure 3.2. N/N, M1/M1 and M2/M2 AmplideX results. **A**. An N/N subject with two copies of both *SMN1* and *SMN2*. **B**. A M1/M1 subject with zero copies (red arrow) of *SMN1* and two copies of *SMN2*. **C**. A M2/M2 subject with two copies of *SMN1* and zero copies (red arrow) of *SMN2*.

All 92 subjects were tested using the AmplideX kit and results compared with results obtained by MLPA. Only two subjects' results did not match the MLPA results - one was a subject with a U/U genotype who had two copies of *SMN1* on MLPA but had only one copy of *SMN1* on the AmplideX kit. The second subject had an N/M1 genotype with one copy of *SMN2* on MLPA and two copies of *SMN2* on the AmplideX kit. All other results obtained on AmplideX correlated with the MLPA results. The AmplideX kit was more sensitive to DNA quality and quantity compared to MLPA.

M1/M1 subjects - AmplideX DNA copy number

Thirteen subjects were tested using the AmplideX kit. All subjects were confirmed to be homozygous for an *SMN1* deletion. The M1/M1 results correlated completely with the MLPA results therefore confirming the AmplideX kit's ability to accurately detect affected SMA subjects with no *SMN1* copies.

M2/M2 subjects - AmplideX DNA copy number

Fourteen M2/M2 subjects were tested using the AmplideX kit and all subjects were confirmed to have zero copies of *SMN2*. These subjects were not affected with SMA but were used to confirm *SMN2* copy numbers which plays a role in understanding the SMA disease and identifying patients eligible for SMA therapy.

N/M1 subjects - AmplideX DNA copy number

This group of subjects were previously identified to be carriers of the *SMN1* deletion on MLPA and pedigree analysis. They were either true carriers (one copy of *SMN1*, 1:0 genotype) or had two or more *SMN1* copies on a single chromosome/in cis (silent carriers, 2:0 genotype) as illustrated in Figure 3.3 A. The AmplideX kit was able to detect true carriers of SMA (N/M1, 1:0 genotype) with one copy of *SMN1* accurately (N/M1) as shown in Figure 3.3.

N/N subjects - AmplideX DNA copy number

To further understand the baseline or normal *SMN* copy number in the general population, 19 N/N subjects tested on MLPA were also included in AmplideX testing. No discrepancies were found with *SMN1* and *SMN2* copies between the AmplideX and MLPA methods. Similarly, to MLPA some N/N subjects had more than two copies of the SMN1 gene as shown in Figure 3.3 B. Comparative analysis between the methods will be reported in section 3.1.7

U/U subjects - AmplideX DNA copy number

This subject group was tested to try to find an alternative method to confirm their clinical diagnosis of SMA or to further understand the SMA mechanism in the black SA population. There were 30 U/U subjects tested on AmplideX, where the subjects' *SMN1* and *SMN2* MLPA copy numbers were previously quantified by Vorster et al (2020) and were included in this study. Only one subject had a discrepant *SMN1* copy number on AmplideX when compared to MLPA. This subject had two *SMN1* copies on MLPA and had one *SMN1* copy on AmplideX. This was the only subject who could potentially be investigated further.

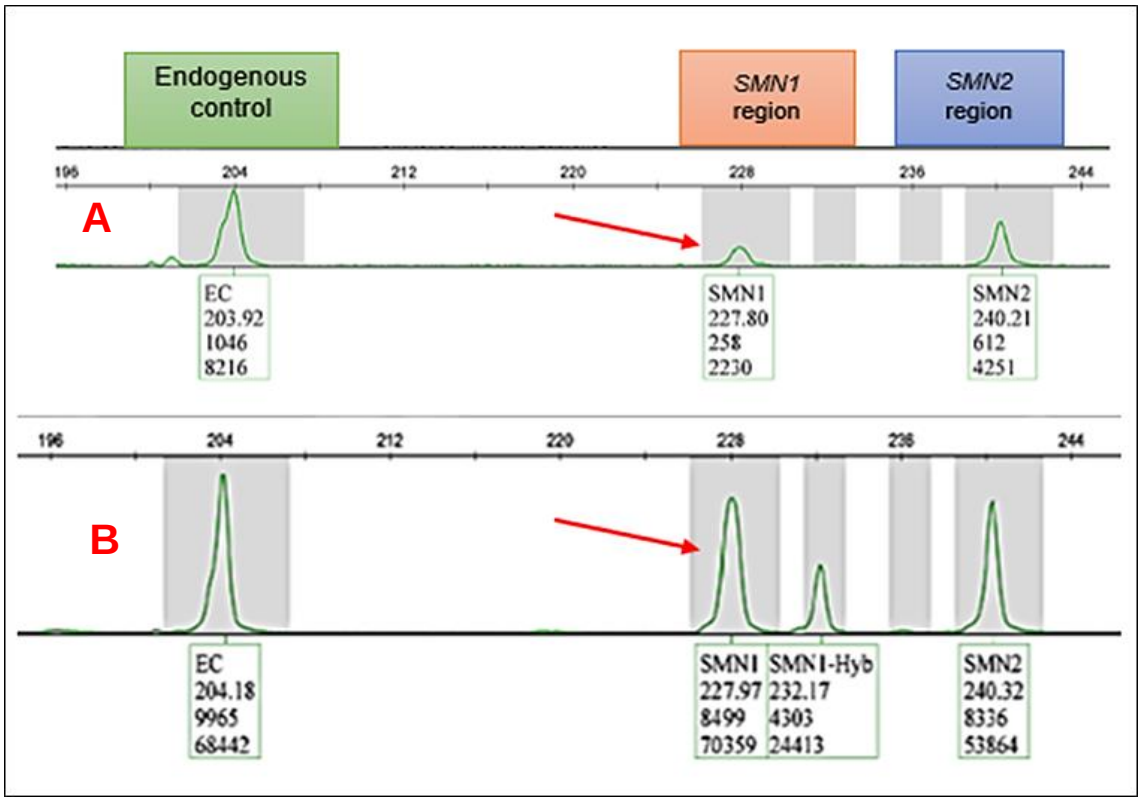


Figure 3.3. AmplideX results in subjects with varying gene copy numbers. **A**. N/M1 subject with one DNA copy of the *SMN1* gene (arrow) and two copies of the *SMN2* gene. **B** N/N subject with four copies of *SMN1* (arrow) calculated with the AmplideX Macro (with one being a hybrid copy) and two copies of *SMN2*.

3.1.3 Gene conversion and DNA copy number

There are at least six possible ways in which gene conversions can occur between *SMN1* and *SMN2*. These possible mechanisms have been explained in Figure 1.4, section 1.3.2. The AmplideX kit is the only method of the four discussed offering detection of gene conversion of *SMN2* into *SMN1* and vice versa. Gene conversion copies are labelled "hybrids". Of the subjects tested 7% (6/92) had a gene conversion

of *SMN2* to *SMN1* (hybrids were only observed in black subjects). Five of the subjects found to have gene conversion events had two *SMN1* copies with one copy being a hybrid, and one subject was found to have four copies of *SMN1* of which some were hybrid copies, as indicated in Table 3.3. In this section the copy numbers were unbundled to illustrate the subject with ≥4 copies. All hybrid copies detected were from *SMN2*-to-*SMN1* gene conversions, as expected in the black SA population who tends to have multiple copies of *SMN1*; *SMN1*-to-*SMN2* gene conversions were not detected in this study.

Table 3.3. Percentage of hybrid ***SMN2***-to-***SMN1*** copies detected using AmplideX.

Copy number	**Number of subjects (Percentage of subjects)**
0	13 (14.13%)
1	13 (14,13%)
2	39 (42,39%)
2 (Hybrid)	5 (5,43%)
3	14 (15,22%)
≥4	7 (6,52%)
≥4 (Hybrid)	1 (1,09%)
Total	92 (100%)

*The red circles represent hybrids SMN2-to-SMN1 copies observed. Five subjects with two SMN1 DNA copies had a one SMN1 hybrid copy. One subject with ≥4 SMN1 DNA copies had one SMN1 hybrid copy.

The hybrid copies were only observed in black subjects and more commonly in those with two or more *SMN1* copies in this study. Analysis of MLPA probes, to possibly verify the gene conversions in the six subjects, did not prove helpful. None of these subjects had any visible gene conversions. Gene conversion accounted for a small portion of *SMN1* copies detected (7% of all subjects).

3.1.4 DNA copy number detection by qPCR

A qPCR method was developed and validated using known controls to determine the utility of this method compared to MLPA and the AmplideX kit. Copy number detection using qPCR could be carried out in a day with dilution, PCR setup, amplification, and detection on the QuantStudio 3, Applied biosystems followed by analysis on the QuantStudio software. All samples were diluted to 50ng/ul and setup in triplicate. Validated qPCR results are shown in Figure 3.4. An example of an amplification plot is illustrated in Figure 3.4 A. The *CFTR* amplification curves were more likely to cluster together, most likely because they were all expected to be present at two copies. However, the *SMN1* and *SMN2* amplification curves did not cluster together as

expected as subjects had variable copy numbers. Outlier peaks/samples were omitted from analysis as recommended for qPCR analysis. These peaks were likely caused by pipetting error and did not affect the results as all samples were run in triplicate. The preliminary copy number was determined using the RQ value (Relative quantification = 2-ΔΔCt), calculated by comparing the results of *SMN1* and *SMN2* to results generated by the *CFTR* housekeeping gene as well as comparing with negative controls (two copies of both *SMN1* and *SMN2*).

All qPCR reactions were multiplexed and included the *SMN* primers and *SMN1* and *SMN2* specific TaqMan™ probes as well as *CFTR* primers and *CFTR* specific TaqMan™ probes. Each reaction also included a calibrator (internal control) which was three known negative controls with two copies of both *SMN1* and *SMN2*.

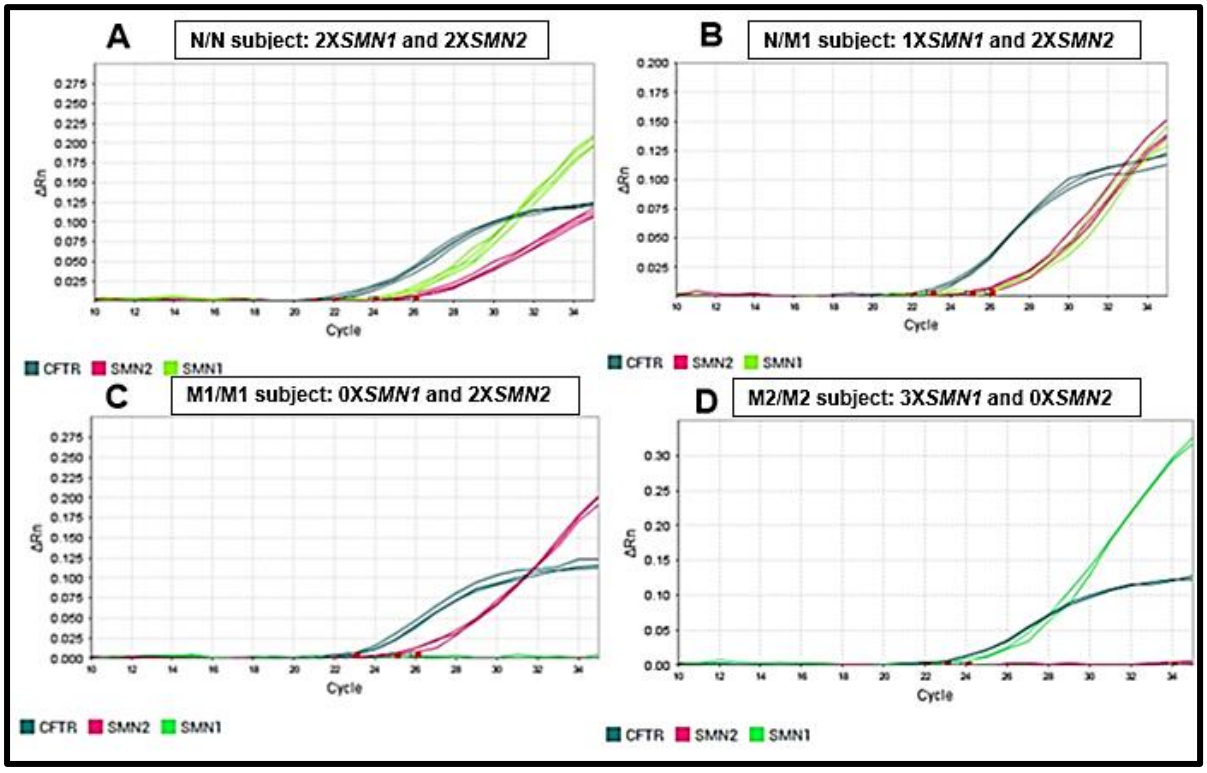


Figure 3.4. Amplification plot of qPCR. The normalised fluorescent value is indicated on the Y axis and the PCR cycle number on the X axis. The *CFTR* gene was labelled in grey and the *SMN1* and *SMN2* genes in green and pink, respectively, **A**. The amplification curves of three N/N controls, these are all subjects with two copies of the *SMN1* and *SMN2* genes. **B**. An amplification curve of a N/M1 subject with a heterozygous deletion of *SMN1* and two copies of *SMN2*. As can be seen when compared with figure A, the amplification curve of *SMN1* is much lower. **C**. A M1/M1 subject with zero copies of *SMN1* (no amplification). **D**. an M2/M2 subject with zero copies of *SMN2* (no amplification).

M1/M1 subjects - qPCR DNA copy number detection

The same subjects tested on MLPA and AmplideX were also tested on qPCR. These subjects also had a homozygous deletion of *SMN1*. A homozygous deletion of either *SMN1* or *SMN2* will not yield any amplification as illustrated in amplification plot Figures 3.5 (SMAR7). Figure 3.5 shows copy number of *SMN* and *CFTR* genes before quantification of absolute copy numbers. All M1/M1 results correlated with both MLPA and AmplideX and thus demonstrating qPCR as a viable alternative method for SMA testing.

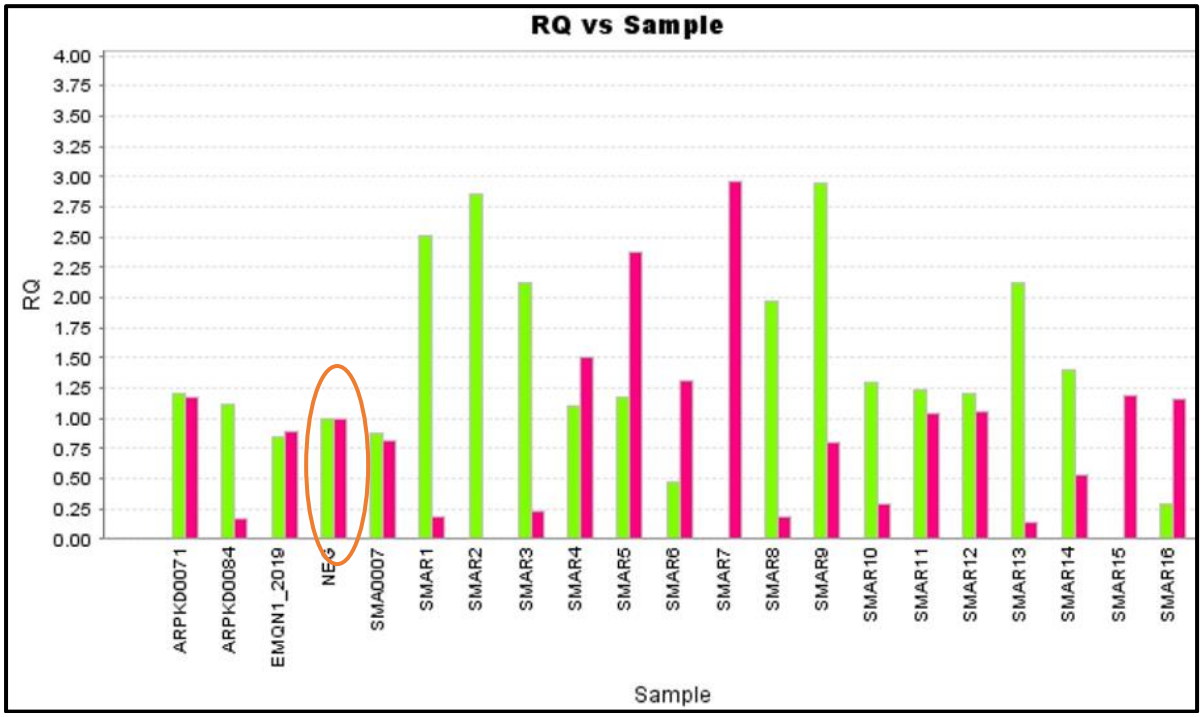


Figure 3.5. Copy number ratio of *SMN1* and *SMN2* in 20 subjects. Green represents *SMN1* and pink represents *SMN2*. Each subject was labelled with a unique identifier displayed on the x-axis; the relative quotient ratio is displayed on the y-axis. The RQ values varied for different subjects, subject ARPKD71 and SMA0007 both had two copies of *SMN1* and *SMN2*. Subject SMAR13 had no amplification of the *SMN2* gene therefore zero copies and subject SMAR7 had zero copes of *SMN1*. The negative controls, circled in orange were normalised to a RQ value of one.

M2/M2 subjects qPCR DNA copy number

Fourteen M2/M2 subjects were tested using qPCR. The quantification of the *SMN2* copies was more challenging than *SMN1*. Similar to MLPA the *SMN2* copy numbers depended on each subject's *SMN1* copy number. However, the use of three internal controls with each reaction resolved this problem. Three internal controls were able to normalise the *SMN1* and *SMN2* ratios. Reliable comparative analysis depends on the quality and choice of control subjects. Figure 3.5 illustrates a typical example of gene

copy number ratios obtained with each experiment. The samples labelled NEG (circled in orange) represent three internal or negative controls (two copies of *SMN1* and *SMN2*) all given the same name “NEG” so they could be normalised to 1. The negative controls were then used to calculate the copy numbers of other subjects in the experiment. As seen in Figure 3.5 the relative quantification ratios seemed inconsistent, the results were exported using Excel and absolute copy numbers were calculated using the calculations described in Section 2.2.1.6. This calculation used to determine copy number was similar to the ones used for MLPA and AmplideX. All M2/M2 subjects had a homozygous deletion of S*MN2* therefore correlating with both MLPA and AmplideX. A typical result of an M2/M2 subject is indicated in Figure 3.4 D and Figure 3.5 SMAR7.

N/M1 subjects - qPCR DNA copy number

All subjects 13/13 (100%) tested on qPCR had the same *SMN1 c*opy numbers as on MLPA and AmplideX. Two of the subjects had no results due to poor and low DNA quality. The qPCR method was as accurate at predicting *SMN1* carriers as MLPA albeit more sensitive to DNA quality/quantity. A typical example of a result of an N/M1 subject is shown in Figure 3.4 B and Figure 3.5 SMAR6 and SMAR16. Both subjects had a decreased *SMN1* amplification compared to negative controls.

N/N subjects - qPCR DNA copy number

All 19 N/N subjects were tested using the qPCR method had results similar to MLPA and AmplideX. The majority - 79% (15/19) of N/N subjects had two copies of *SMN1* and 21% (4/19) had three or more copies of *SMN1*, suggesting a gene conversion from *SMN2* to *SMN1*. There were no discrepancies between qPCR and the other two methods. All subjects with ≥3 were black SA N/N individuals.

U/U subjects - qPCR DNA copy number

Thirty of the U/U subjects were tested using qPCR in the hope that this method would offer a diagnosis for these clinically suggestive SMA subjects. All subjects had the same results as MLPA and AmplideX except one subject who had two copies on MLPA but had one copy of *SMN1* on qPCR and AmplideX. This subject was repeated to exclude a possibility of run artefacts, but their result remained the same.

3.1.5 Comparison of MLPA, AmplideX kit and qPCR methods in the detection of *SMN1* DNA copy numbers

Table 3.4 summarises the comparison of *SMN1* copy numbers obtained using the different methods. This was done to validate the two new methods: the in-house developed qPCR method and the AmplideX commercial kit as alternative methods to MLPA as a diagnostic method as well as to possibly improve the understanding of SMA disease.

Overall, there were two subjects with results discrepant from MLPA results as mentioned in Sections 3.1.1 and 3.1.2. Two subjects failed amplification for qPCR and where therefore noted as "no results". Therefore, a total of 90 subjects were compared between the three methods and 96% (86/90) of the *SMN1* DNA copy number results correlated between the three methods.

Table 3.4. Comparison of *SMN1* copy number in different subject groups using different methods.

Subject group	Copy nr - MLPA	Copy nr-AmplideX	Copy nr - qPCR	Consolidated *SMN1* copy nr
M1/M1	13	13	13	**0**
M2/M2	2	2	3	**2**
	12	12	11	**≥3**
N/M1	12	12	10	**1**
	4	4	4	**2**
	0	0	2	**NR**
N/N	15	15	15	**2**
	4	4	4	**≥3**
U/U	0	1	1	**1**
	24	23	22	**2**
	6	6	6	**≥3**
	0	0	1	**NR**
Total	**92**	**92**	**92**	**-**

*The columns circled in red represent subjects that did not correlate with MLPA.

Comparison of SMN1 copy number results across three methods.

Figure 3.6 summarises the comparison of *SMN1* DNA copy numbers across the five genotypes. All M1/M1 subjects had zero copies of *SMN1* with no discrepancy across the three methods. Subjects with the M2/M2 genotype either had two or three copies of the *SMN1* gene. A discrepancy was observed with the qPCR results where one subject had two copies of *SMN1* on qPCR but had three copies on both MLPA and AmplideX. This subject was marked with a red star in Figure 3.6. Results of the N/M1

and N/N genotypes correlated between all three methods with no discrepancies. Another subject as mentioned above was a U/U subject who had two copies of *SMN1* on MLPA and one copy on both AmplideX and qPCR.

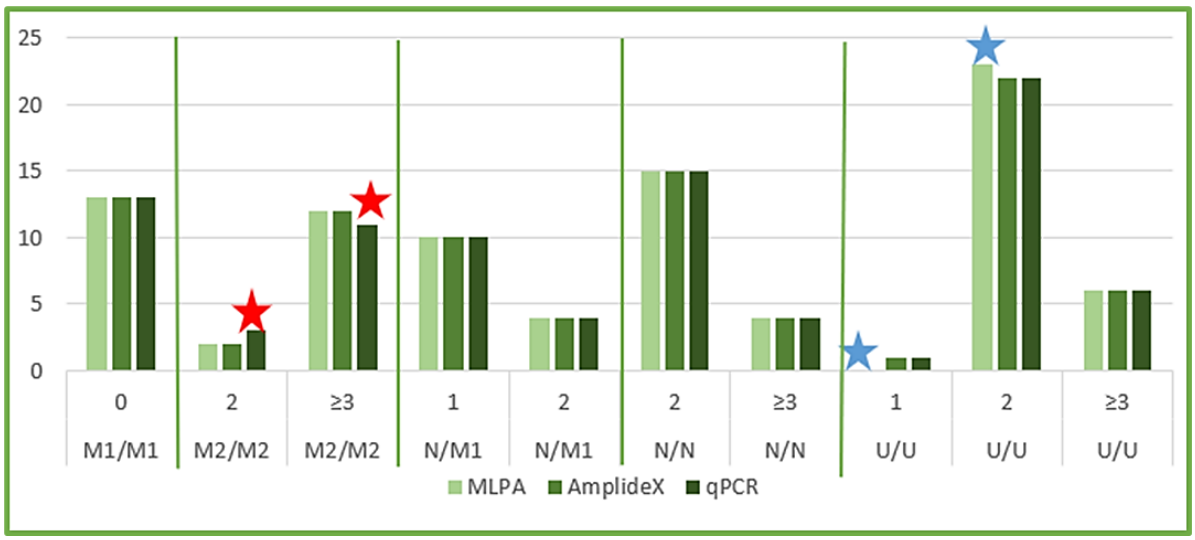


Figure 3.6. Comparison of *SMN1* copy number results across three methods. Five genotypes M1/M1, M2/M2, N/M1, N/N and U/U were compared. Each bar represents the method used for analysis; each coloured star represents a subject that did not have the same *SMN1* DNA copy number.

Comparison of MLPA, AmplideX kit and qPCR methods in the detection of SMN2 DNA copy numbers

The DNA copy number of *SMN2* using the MLPA, qPCR and AmplideX methods were compared to validate the two new methods for diagnostic use and to further investigate the SMA mechanism. All subjects with a homozygous deletion of *SMN1* (M1/M1) had at least two copies of *SMN2*, illustrated in Figure 3.7. Furthermore, there were no conflicts between the three different methods in M2/M2 subjects who had zero copies of *SMN2*. However, there was some variation between the methods in some subjects, i.e., one subject with an N/M1 genotype had two copies of *SMN2* on MLPA but had only one copy of *SMN2* on both qPCR and AmplideX. Methods with conflicting copy number results were repeated to exclude errors during setup. High copy numbers of *SMN2* were less frequent compared to *SMN1*, only seven subjects had ≥3 copies of *SMN2* compared to 22 subjects with ≥3 of *SMN1*. Approximately 50% of subjects had two copies of *SMN2*, 7.6% had ≥3 copies of *SMN2*. Only one subject with a U/U genotype had zero copies of *SMN2*.

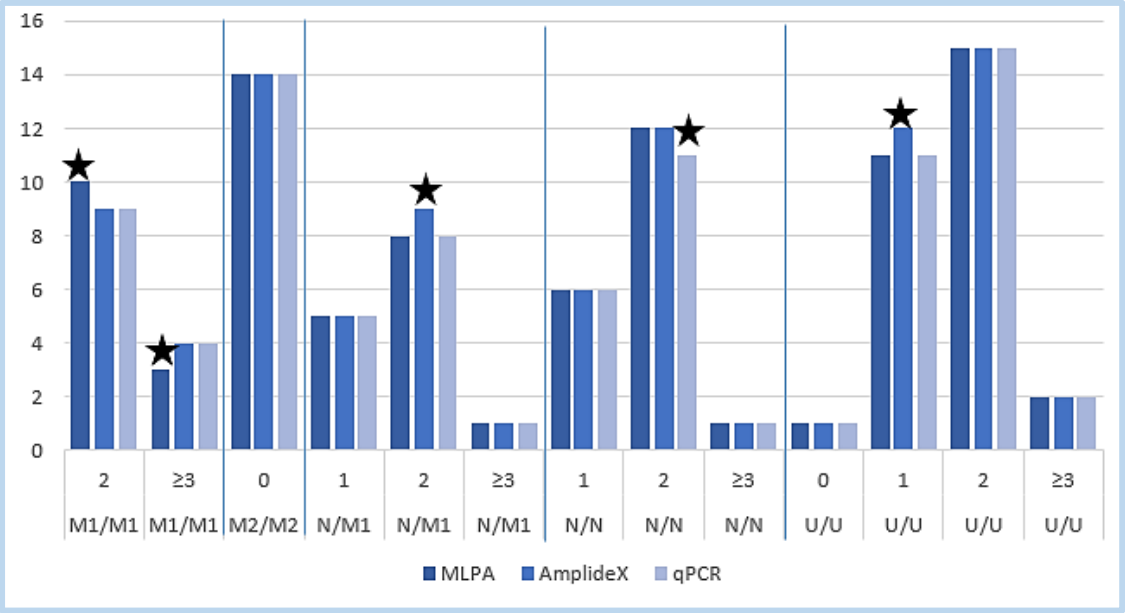


Figure 3.7. Comparison of *SMN2* copy number results across three methods. Five genotypes M1/M1, M2/M2, N/M1, N/N and U/U were compared. Each bar represents the method used for analysis. The stars show subjects with different *SMN2* copies on MLPA compared to AmplideX and qPCR.

Comparison between genotype and SMN2 copy number.

A comparison between genotype and copy number of *SMN2* was further investigated to determine any correlation between *SMN2* copy number and SMA disease. From the Figure 3.8 below, it could be seen that most subjects across all subject groups had two copies of *SMN2* and few had ≥3 copies.

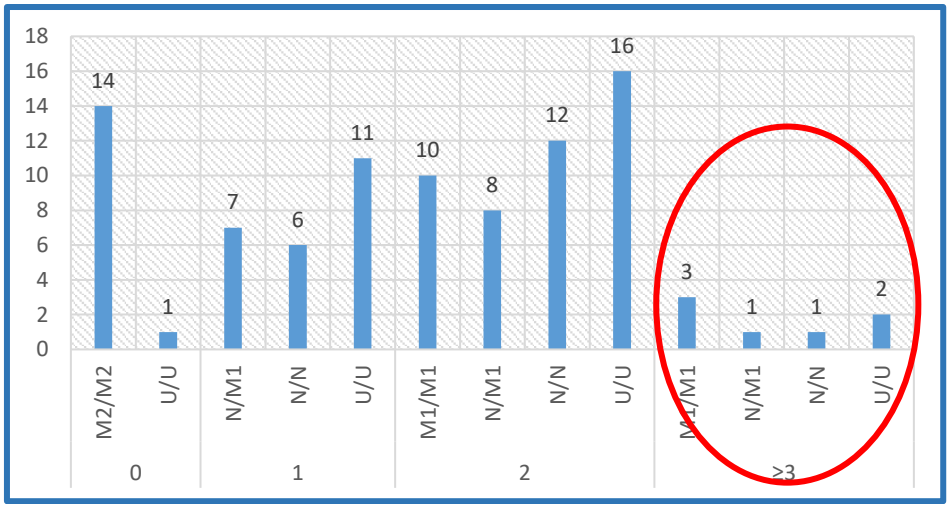


Figure 3.8. *SMN2* copy number compared to genotype. *SMN2* DNA copy numbers were counted from zero to ≥3. The y-axis represents the number of subjects tested with the given genotype versus the *SMN2* DNA copy number detected (x-axis). As expected, all M2/M2 had zero copies of *SMN2* and the least common *SMN2* copy number across all genotypes was ≥3 circled in red.

3.1.6 Summary of comparison of *SMN1* and *SMN2* DNA copy number

The *SMN1* and *SMN2* DNA copy numbers, resulting from the three methods, were evaluated to determine the accuracy of the methods to predict clinical utility. Furthermore, the cost, laboratory workflow, limitations and advantages of each method was investigated. Figure 3.9 compares the results of MLPA, AmplideX and qPCR side by side. The biggest difference between qPCR and the other two methods is higher amplification failure. A higher concentration of DNA was needed for qPCR, compared to MLPA, and it was an in-house designed system compared to the kit based MLPA and AmplideX methods. Most subjects had two copies of *SMN1* or *SMN2*. Multiple copies (≥3 copies) of *SMN1* (23.9%) and a corresponding decrease in *SMN2* copy number (one copy: 23.9%) were frequently observed in black patients, possibly due to gene conversions as illustrated in Figure 3.9.

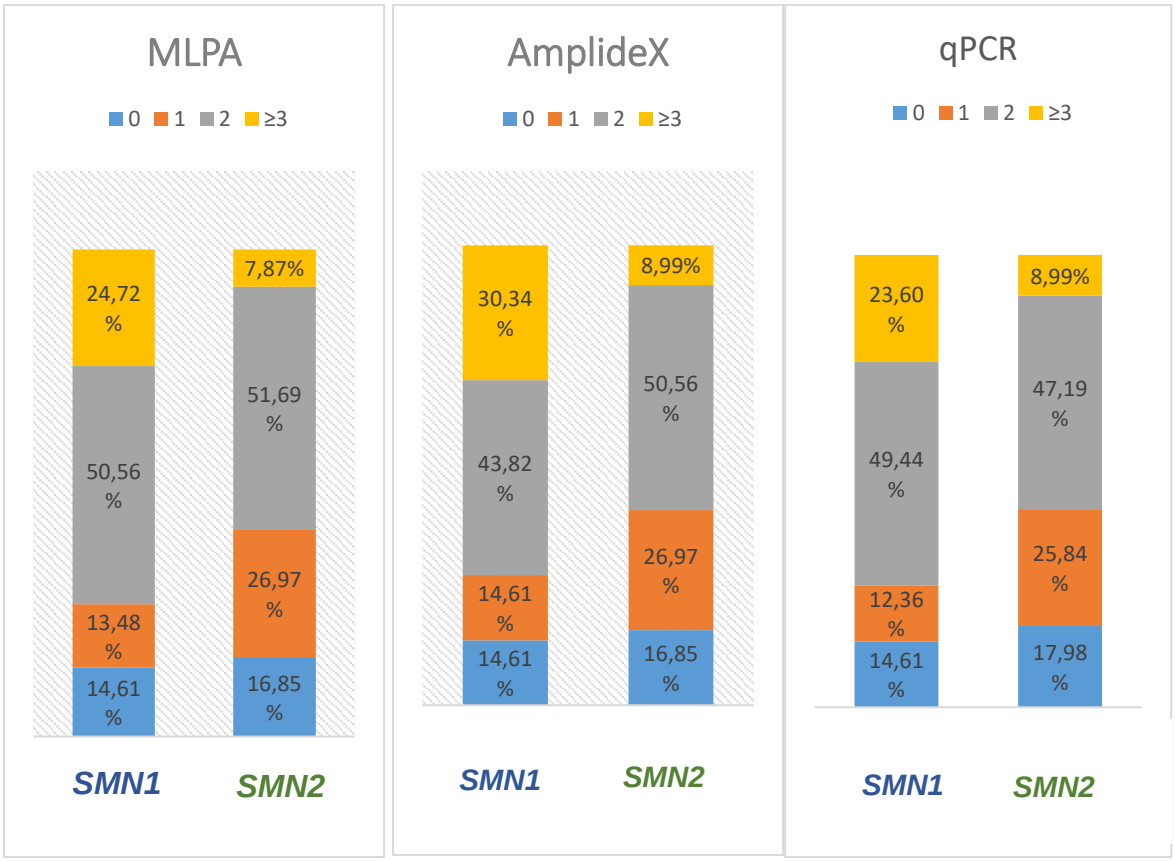


Figure 3.9. Comparison of DNA copy number between MLPA, AmplideX and qPCR. The bars indicate the percentage of patients who had certain copy numbers. Different colours indicate different copy numbers. The most common copy number was 2 (grey area) for both *SMN1* and *SMN2*. The least common copy number was ≥3 on *SMN2* (orange area).

Summary of M1/M1 DNA copy number results

All 13 subjects with known M1/M1 genotypes had homozygous deletions of *SMN1* on MLPA, qPCR and AmplideX, therefore confirming the accuracy of MLPA, qPCR and AmplideX in correctly diagnosing typical SMA patients across all ethnic groups. A total of 77% (10/13) subjects had 2 copies of *SMN2* and 23% (3/13) of subjects had ≥3 copies of *SMN2* suggestive of gene conversions from *SMN1* to *SMN2*. None of the subjects had homozygous deletions of both *SMN1* and *SMN2*.

Summary of M2/M2 DNA copy number results

All 14 subjects with the M2/M2 genotype had zero copies of *SMN2* on all three methods. The majority of M2/M2's had three or more copies of *SMN1* (12/14 or 86% combined black and white), suggesting gene conversions from *SMN2* to *SMN1*. There was one subject, who had one copy of *SMN1*, but this was only observed on qPCR which differed from MLPA and AmplideX. The qPCR method was repeated but still had a result of one copy of *SMN1*.

Summary of N/M1 DNA copy number results

There was a total of 16 subjects (9 black and 7 white) with the N/M1 genotype. A total of 55.5% (5/9) of black subjects had one copy of *SMN1* compared to 100% (7/7) of white subjects. A total of 44.5% (4/9) of black subjects had 2 copies of *SMN1* indicating that these subjects are silent carriers with a 2:0 genotype, with both copies located on the same chromosome (see section 1.8 for a definition of silent carriers). No N/M1 subjects had three or more copies of *SMN1* (multiple copies would all be located on the same chromosome in silent carriers). This group also had one or two copies of *SMN2*. Table 3.5 summarises all the results. Except for two samples which failed qPCR, all three methods produced the same results for this group. Thus, there were no conflicts between MLPA. qPCR and AmplideX in the N/M1 group.

Summary of N/N DNA copy number results

Altogether 19 subjects had a N/N genotype, 60% (6/10) of black subjects had two copies of *SMN1* and 40% (4/10) had three or more copies of *SMN1*, compared to 100% (9/9) of white subjects with only two copies of *SMN1*. See Table 3.5 below for an outline. There was no deviation among the three methods for *SMN1* copy number

detection. In total 40% (4/10) of black subjects with an N/N genotype had a homozygous deletion of *SMN2*, which was not observed in white subjects, suggesting a possible gene conversion from *SMN2* to *SMN1* resulting in multiple copies of *SMN1* in black subjects. Only one subject had three or more copies of *SMN2* (with two copies of SMN1), and most N/N subjects had two copies of both *SMN1* and *SMN2*.

Summary of U/U copy number results

None of the subjects with the U/U (All 29 black subjects) genotype had a homozygous *SMN1* deletion on any of the three methods. One subject who had two copies of *SMN1* on MLPA, had one copy of *SMN1* on the qPCR method and AmplideX kit. MLPA was repeated for this subject and was found to still have two copies of *SMN1*. The possible reasons for this discrepancy are explained in Section 4.1 in the Discussion. Most subjects (76%, 22/29) with the U/U genotype had 2 copies of *SMN1*, with 21% (6/29) having three or more copies of *SMN1*, appearing similar in CNV pattern to the N/N subject group.

Approximately 38% (11/29) of U/U subjects had one *SMN2* copy, 52% (15/29) had two *SMN2* copies and 7% (2/29) had three or more copies of *SMN2*. Overall, four subjects with the U/U genotype had different copy numbers of either *SMN1* or *SMN2* when comparing all three methods. The presence of one copy of *SMN1* would suggest a more likely diagnosis of SMA. The qPCR method had three failures compared with zero with the MLPA and AmplideX kits.

Table 3.5. Summary of DNA copy number methods

Ethnicity	Genotype	Genomic Copy number	SMN1 MLPA	SMN2 MLPA	SMN1 amplidex	SMN2 amplidex	SMN1 qPCR	SMN2 qPCR	Total	Total
Black	M1/M1	0	12	0	12	0	12	0	12 (100% SMN1 del)	70
		1	0	0	0	0	0	0		
		2	0	9	0	8	0	8		
		≥3	0	3	0	4	0	4		
	M2/M2	0	0	10	0	10	0	10	10 (100% SMN2 del)	
		1	0	0	0	0	1	0		
		≥3	10	0	10	0	9	0		
	N/M1	1	5	4	5	4	5	4	9 (55,5 % SMN1 del)	
		2	4	5	4	5	4	5		
	N/N	1	0	4	0	4	0	5	10 (50.0-60.0 % SMN1 del)	
		2	6	5	6	5	6	4		
		≥3	4	1	4	1	4	1		
	U/U	0	0	1	0	1	0	1	29 (0% hom SMN1 del)	
		1	0	11	1	12	1	11		
		2	23	15	22	14	22	15		
		≥3	6	2	6	2	6	2		
White	M1/M1	0	1	0	1	0	1	0	1 (100% SMN1 del)	19
		2	0	1	0	1	0	1		
	M2/M2	0	0	4	0	4	0	4	4 (100% SMN2 del)	
		2	2	0	2	0	2	0		
		≥3	2	0	2	0	2	0		
	N/M1	1	5	1	5	1	5	1	5 (100% SMN1 del)	
		2	0	3	0	3	0	3		
		≥3	0	1	0	1	0	1		
	N/N	1	0	2	0	2	0	2	9 (0% SMN1 del)	
		2	9	7	9	7	9	7		
Total			89	89	89	89	89	89	89	89

*Three subjects that failed amplification on qPCR were removed from comparative analysis, therefore only 89 subjects were summarised in the table.

Comparison of DNA methods in a subject with a gene conversion

One black (N/N) subject with four copies of *SMN1* and two copies of *SMN2* was used to illustrate gene conversion as shown in Figure 3.10. For this section, the absolute copy numbers obtained will be used. The subject had four copies of the *SMN1* gene for both exons 7 and 8 on MLPA, there was no apparent gene conversion on MLPA as there were also two copies of *SMN2* for both exons 7 and 8 indicated in Figure 3.10a. However, hybrid *SMN1* copies (*SMN2*-to-*SMN1*) were detected on AmplideX even though the overall copy number for AmplideX was four for *SMN1* and two for *SMN2*, consistent with the MLPA results as indicated in Figure 3.10b. The copy number on qPCR was five for *SMN1* and two for *SMN2*, which was higher compared to MLPA and AmplideX. One of the limitations of both MLPA and AmplideX is the inability to accurately quantify three or more copies. The result for this subject was consistent between MLPA, AmplideX and qPCR as it was summarised as ≥3 copies of *SMN1* and two copies of *SMN2*.

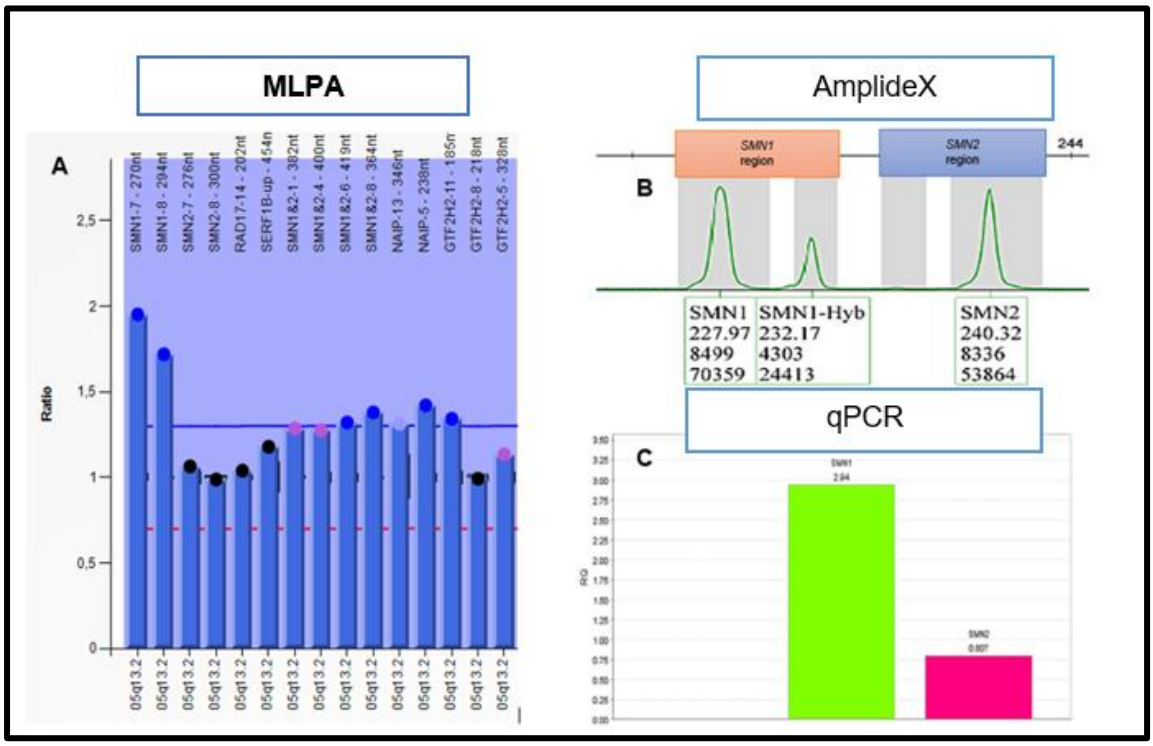


Figure 3.10. Comparison of MLPA, AmplideX and qPCR results in a single subject. **A**. Ratio chart of *SMN* region on chromosome five for the P021 MLPA kit, subject has four copies of *SMN1* (ratio equal to 2) and two copies of *SMN2* (ratio equal to 1). **B**. AmplideX result - a hybrid copy was detected in the *SMN1* region, the overall copy number was four for *SMN1* and two for *SMN2*. **C**. RQ value of the subject on qPCR, green bar represents *SMN1* and pink *SMN2*. The calculated *SMN1* copies were five and two for *SMN2*.

3.1.7 Costing of methods for determination of SMN DNA copy number

The utility of the three methods was further assessed by calculating the cost of each method, using the NHLS diagnostic costing model. The model considers the reagents and consumables used to perform a test, the salary cost of every staff member responsible for the test, the hands-on time to perform the test, and the equipment used, including its maintenance and service. Other factors included in the calculations were EQA's performed, error rate of the method and VAT. Table 3.6 summarises the cost of each method. MLPA is the most expensive method, with the longest hands-on time and setup and analysis at 7.7 hours compared to 3.5 hours for both qPCR and AmplideX. The cost of MLPA was R744.05 per subject, and most of that expense was due to the probe mix and use of the thermocycler and ABI Genetic Analyzer. See appendix 10 for complete costing model.

Table 3.6. Costing results for MLPA, AmplideX and qPCR

SMN1/SMN2 detection Method	Hands on time (Minutes)	Costing per sample
MRC P021 Probe mix kit MLPA	120	744,05
AmplideX kit	60	627,62
In house qPCR	60	165,06

The AmplideX kit had a much lower hands-on time, however the cost of the kit, equipment and equipment maintenance was similar to that of MLPA. The AmplideX kit was optimised to work at half reactions and validated and calculated using the half reaction quantity. The AmplideX also had shorter analysis time, therefore further reducing hands-on time per test.

The qPCR method was the most affordable method at R199,93. The TaqMan™ master mix was significantly cheaper than both the MLPA and AmplideX kits. Furthermore, a centrifuge and real-time thermocycler are the only large instruments used. The hands-on time was shorter than MLPA, and analysis was similar to that performed for AmplideX. The triplicate setup and repeat testing for qPCR was also included.

The number of controls was included when calculating reactions per reagent. All three methods used reference controls (endogenous controls), negative controls and positive controls with each run. Three negative controls were included in each MLPA

and qPCR run. The AmplideX kit came with a calibrator which served the same purpose as a negative control. A positive control was included in each run to determine the reproducibility of the method.

Overall, the qPCR method had the shortest setup time, analysis time, number of reagents, equipment and software used, resulting in it being the most affordable method. All three methods accurately determined a homozygous deletion of *SMN1* and therefore provide a diagnosis for SMA and determine true carriers and silent carriers when a complete pedigree was provided. MLPA was most informative as multiples probes were used in the *SMN* region, but AmplideX could detect hybrid *SMN1/SMN2* copies.

3.2 Gene expression: RNA copy number detection

This section will focus on the RNA expression study of 19 subjects. The RT-qPCR method was optimised as an alternative/second line diagnostic test on four M1/M1 and 10 N/N subjects, one M2/M2, two N/M1, followed by testing of two U/U subjects. Furthermore, a comparison of DNA copy number and RNA expression was done. The expression levels of three different transcripts were analysed: FL-SMN1, FL-SMN2 and (SMN∆7 (see section 2.2.2 for an explanation of the different transcripts). *GAPDH* was used as an endogenous control together with three negative controls (calibrators) in each reaction. The negative controls had previously been found to have two copies of both *SMN1* and *SMN2* on DNA copy number methods.

The most important transcript in SMA was FL-SMN1 as it is the transcript used for protein synthesis as it includes the critical exon 7 (see section 1.3.1 for an explanation of the *SMN* region). FL-SMN1 transcripts produce approximately 80% of SMN protein. Only 20% of full-length transcripts is produced from transcription of the *SMN2* gene (FL-SMN2). Most SMN2 transcripts lack exon 7 (SMNΔ7) and are unstable and easily degraded. In M1/M1 subjects, SMNΔ7 transcripts may be produced from mutated/deleted/converted *SMN1* genes. The SMN∆7 transcript therefore represents a combination of SMN1 and SMN2 transcripts lacking exon 7.

The aim of the RNA expression study was to determine the relationship between *SMN1* and *SMN2* DNA copy number and RNA expression level. This was done to determine whether patients presenting with SMA who appear to have one or more *SMN1* copies on DNA analysis may have non-functional *SMN1* copies which may hamper expression of RNA. RNA expression may also potentially assist with determining the carrier status of subjects with multiple *SMN1* gene copies who may be silent carriers (2:0, 3:0, 4:0 etc.). Furthermore, understanding the RNA expression of the *SMN* region may aid in understanding the mechanism of SMA disease.

In this section the same subject groups investigated in the DNA part of the study were used for analysis in the RNA study (see section 2.1 for a breakdown of subject categories).

3.2.1 RNA expression results of M1/M1 subjects

Four subjects with SMA confirmed on DNA copy number detection methods were analysed. These subjects all had zero copies of *SMN1* on DNA copy number detection methods.

Figures 3.11-14 show examples of the compared DNA copy number and RNA expression in an M1/M1 subject. SMA1988 is an affected subject aged 14 years, who is wheelchair bound. The subject was confirmed to have SMA using the RFLP diagnostic method. The subject had zero copies of *SMN1* on DNA copy number detection and zero expression of the FL-SMN1 transcript and is therefore confirmed to be affected with SMA. The subject had three copies of *SMN2* on DNA methods and RNA expression of FL-SMN2 equivalent to four *SMN2* gene copies, suggesting a gene conversion from *SMN1* to *SMN2.* The subject also had RNA expression equivalent to two gene copies of SMNΔ7, which confirms the presence of non-functional SMN transcripts as expected in M1/M1 subjects.

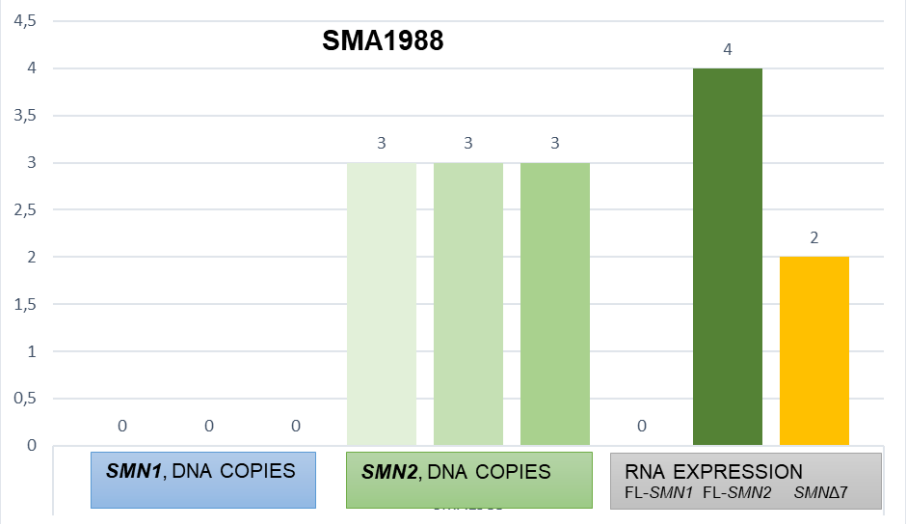


Figure 3.11. DNA copy number versus RNA expression in an M1/M1 subject: SMA1988. The first three bars (zero copies), indicate zero *SMN1* gene copies detected on MLPA, AmplideX and qPCR. The second group of bars (three copies) indicate three *SMN2* gene copies detected on MLPA, AmplideX and qPCR. The third group of bars indicate RNA expression equivalent to zero, four and two gene copies detected by RT-qPCR. This subject is confirmed to be affected with SMA from DNA and RNA detection.

Subject SMAR15 represents a six-month-old child with zero copies of *SMN1* on DNA copy number detection and zero expression of FL-SMN1 and is therefore confirmed to be affected with SMA. The subject has two copies of the *SMN2* gene, RNA expression of FL-SMN1 transcripts equivalent to three *SMN2* copies (Figure 3.12).

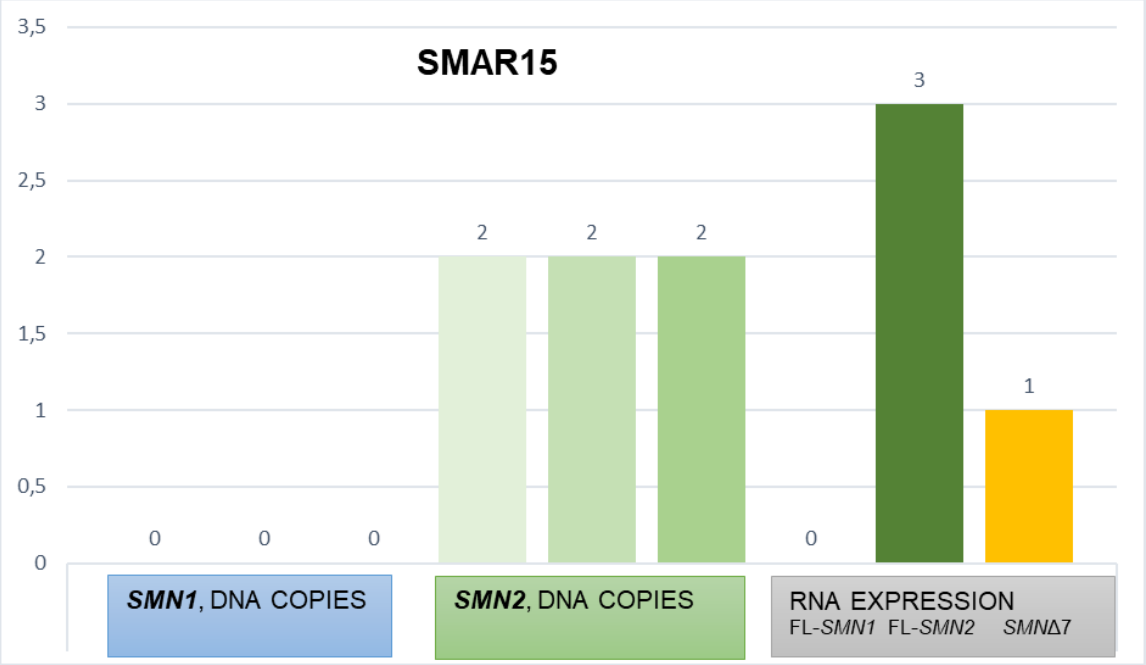


Figure 3.12. An M1/M1 subject with zero copies of *SMN1* on DNA copy number detection and zero copies of FL-SMN1 transcript. The first three bars (zero copies), indicate *SMN1* gene copy number detected on MLPA, AmplideX and qPCR. The second group of bars (two copies) indicate the *SMN2* gene copy numbers detected on MLPA, AmplideX and qPCR. The third group of bars indicate RNA expression levels detected by RT-qPCR. The subject had two DNA copies of *SMN2*, RNA expression levels equivalent to three FL-*SMN2* transcripts and one *SMN*Δ7 transcript.

Subject SMAR7 represents a four-year-old who has been confirmed to be affected with SMA. The subject had zero copies of *SMN1* on DNA copy number detection and zero expression of FL-SMN1 and is therefore confirmed to be affected with SMA. The subject had four copies of *SMN2* on DNA studies and RNA expression of FL-*SMN2* equivalent to six *SMN2* gene copies and seven copies of SMNΔ7, all suggestive of a milder version of SMA (see section 1.6.1, for an explanation on how multiple copies of *SMN2* can decrease the severity of SMA). The DNA copy numbers and RNA expression results for SMAR7 are illustrated in Figure 3.13. This subject was the only M1/M1 individual with more SMNΔ7 expression than FL-SMN2 expression.

.

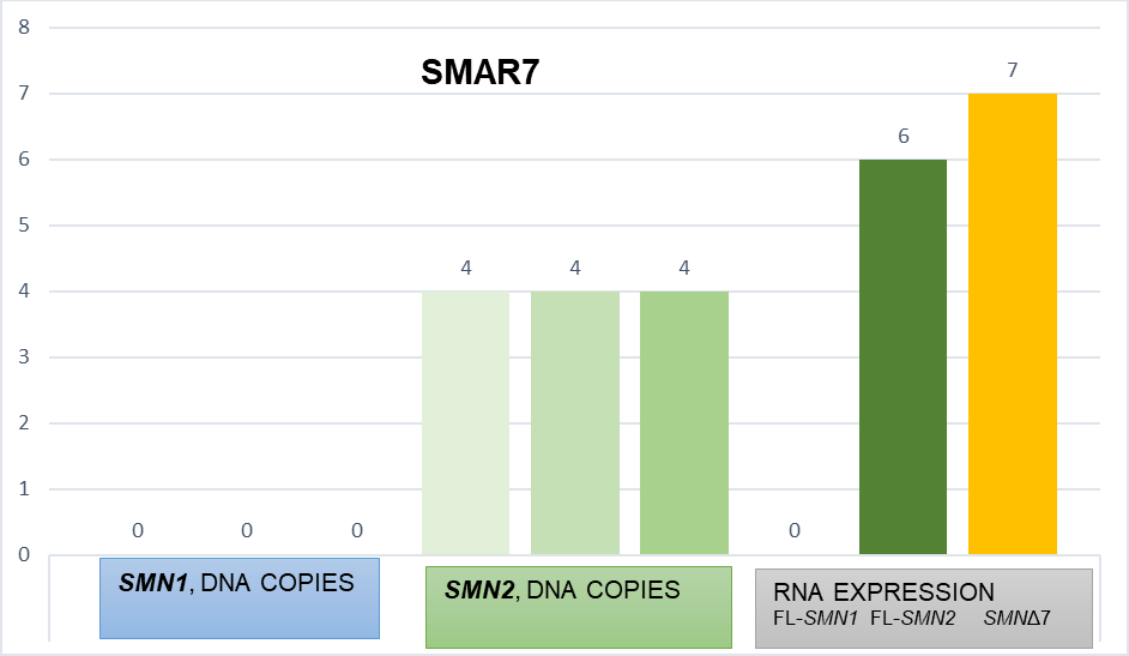


Figure 3.13. DNA copy number versus RNA expression in an M1/M1 subject: SMAR7. An M1/M1 subject with zero copies of *SMN1* on DNA copy number detection and zero FL-SMN1 transcript. SMAR7 had four DNA copies of *SMN2*, RNA expression of FL-SMN2 equivalent to six *SMN2* gene copies and SMNΔ7 expression equivalent to seven gene copies.

3.2.2 RNA expression results of M2/M2 subjects

Only one M2/M2 subject (SMAR2; homozygous deletion of *SMN2* gene) was available for RNA expression analysis with as illustrated in Figure 3.14. This subject had three copies of *SMN1* on all three DNA methods and had FL-SMN1 expression equivalent to four copies of *SMN1*. The *SMN2* DNA copy number and FL-SMN2 expression were zero as expected in M2/M2 subjects. The subject did not have a SMNΔ7 transcript, which is mainly expressed from *SMN2*, as expected. This finding supported an M2/M2 DNA result.

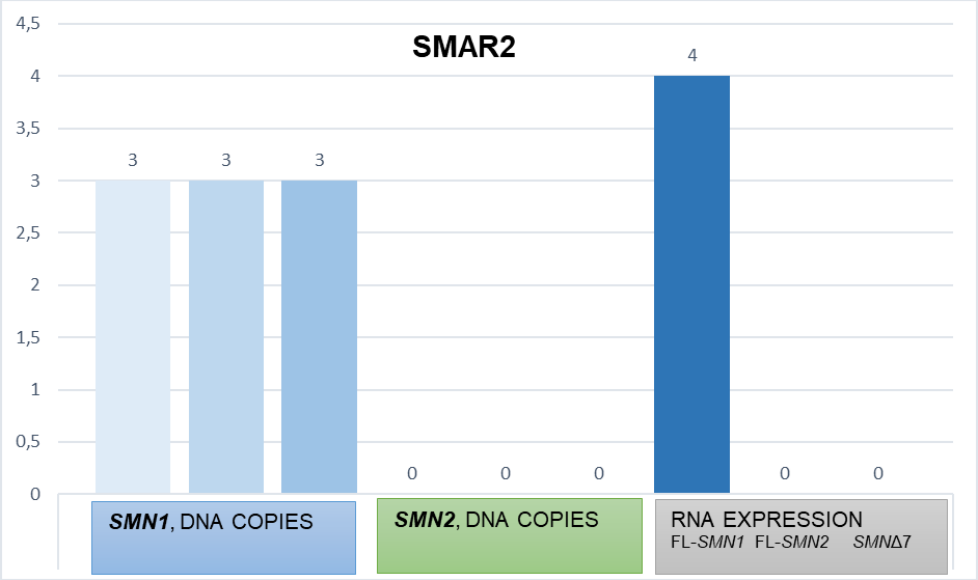


Figure 3.14. DNA copy number versus RNA expression in an M2/M2 subject: SMAR2. The first blue three bars (three copies), indicate *SMN1* gene copy number detected on MLPA, AmplideX and qPCR. The second group of bars (zero copies) indicate the *SMN2* gene copy number detected on MLPA, AmplideX and qPCR. The third group of bars (four, zero and zero) indicate RNA expression levels detected by RT-qPCR. The subject had three DNA copies of *SMN1* and RNA expression level equivalent to four FL-SMN1 transcript copies. The subject had zero DNA copies of *SMN2* and zero RNA expression of FL-SMN2 and SMNΔ7.

3.2.3 RNA expression results of N/M1 subjects: DNA *SMN1* copy number versus FL-SMN1 expression

Two N/M1 subjects were included in the RNA expression study. The two subjects had children confirmed to be affected with SMA and were therefore obligate carriers of SMA. They were also previously confirmed to be carriers of SMA on DNA methods (1:0 genotype). Figure 3.15 illustrates the DNA copy numbers and RNA expression results of these two subjects. Both subjects had one copy of *SMN1* and FL-SMN1 expression equivalent to one copy of *SMN1* which is consistent with being an SMA carrier. There was no discrepancy between DNA *SMN1* copy number and FL-SMN1 expression in the N/M1 group.

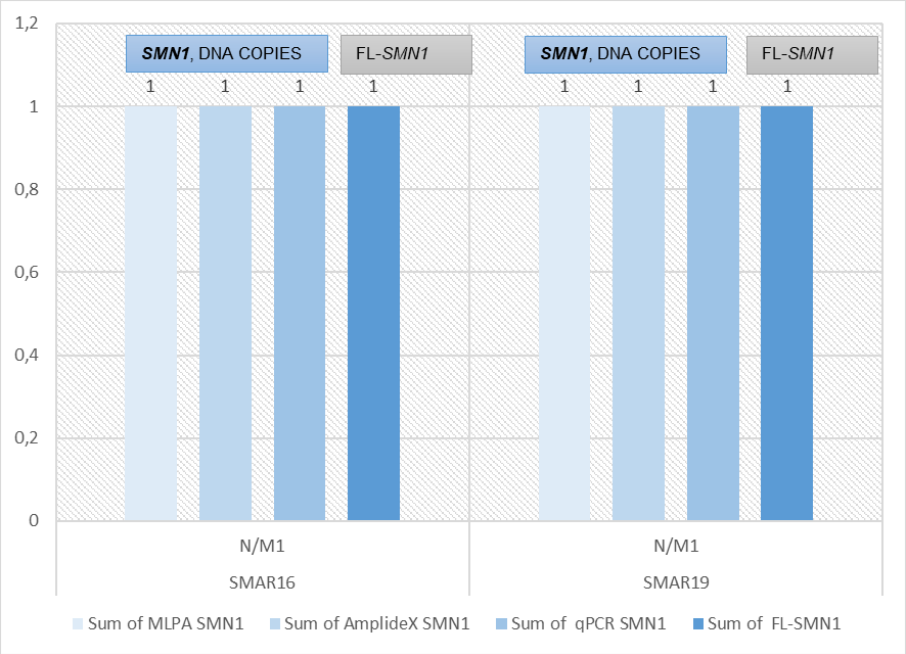


Figure 3.15. *SMN1* DNA copy number versus RNA expression in N/M1 subjects: SMAR16 and SMAR19. Both SMAR16 and SMAR19 had one DNA copy of *SMN1* and FL-SMN1 expression equivalent to one copy of *SMN1*.

The subject SMAR16 had two DNA copies of *SMN2*, and FL-*SMN2* expression levels equivalent to four *SMN2* gene copies and *SMN*Δ7 expression equivalent to one gene copy. These results are depicted in Figure 3.16. The FL-SMN2 transcript expression was twice as high as expected from the *SMN2* copy numbers. However, SMAR16 had decreased levels of SMNΔ7. The second subject SMAR19, had one copy of the *SMN2* gene and FL-SMN2 expression equivalent to one *SMN2* gene copy as illustrated in Figure 3.16. The subject had SMNΔ7 expression equivalent to two gene copies and therefore had an increased amount of this transcript than expected from the DNA *SMN2* copy number and FL-SMN2 transcript expression levels. The two N/M1 mothers will be compared to their children in section 3.3.8.

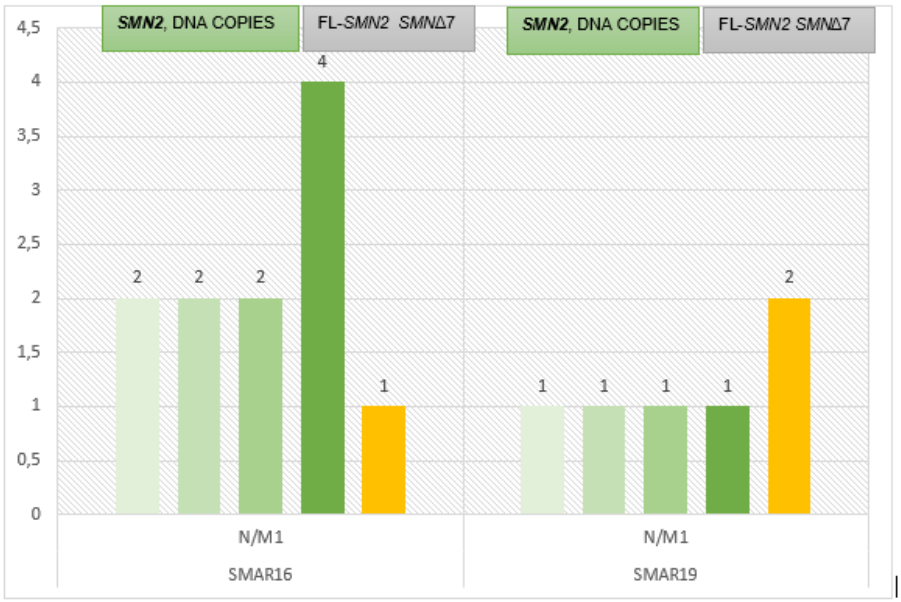


Figure 3.16. *SMN2* DNA copy number versus RNA expression in N/M1 subjects: SMAR16 and SMAR19. The green bar graphs indicate the DNA *SMN2* copies (MLPA, AmplideX and qPCR) and FL-SMN2. The orange bar graph indicates SMNΔ7 transcript expression. Subject SMAR16 had two DNA copies of *SMN2*, FL-SMN2 expression levels equivalent to four copies of *SMN2* gene copies and SMNΔ7 expression equivalent to one gene copy. Subject SMAR19, had one DNA copy of the *SMN2*, FL-*SMN2* expression equivalent to one *SMN2* gene copy and SMNΔ7 expression equivalent to two gene copies.

3.2.4 RNA expression results of N/N subjects

Ten N/N adult subjects with no symptoms or family history of SMA were included in the study. Nine out of ten were black and were tested to determine the normal variation of FL-SMN1, FL-SMN2 and SMN*Δ7* expression levels of black SA subjects. Figure 3.17 illustrates all N/N subjects' DNA *SMN1* copy number versus FL-SMN1 transcript results. Fifty percent (5/10) of the N/N subjects tested had consistent *SMN1* copy number and expression of FL-SMN1 transcripts. Forty percent (4/10) of the subjects had increased FL-SMN1 transcript expression when compared to the corresponding DNA *SMN1* copy number. Only one subject (SMAR9; 10% (1/10)) had FL-SMN1 expression which was lower than expected from the DNA *SMN1* copy number. This subject (SMAR9) had four, copies of *SMN1* on MLPA, four or more copies including an *SMN1* hybrid copy on AmplideX and five copies on qPCR. The DNA copy number for this subject were not consistent across the DNA/genomic methods. This same subject had FL-SMN1 expression suggestive of three copies. These discrepant results shows that three or more copies may not be accurately distinguished using either DNA or RNA based methods and should be interpreted with caution.

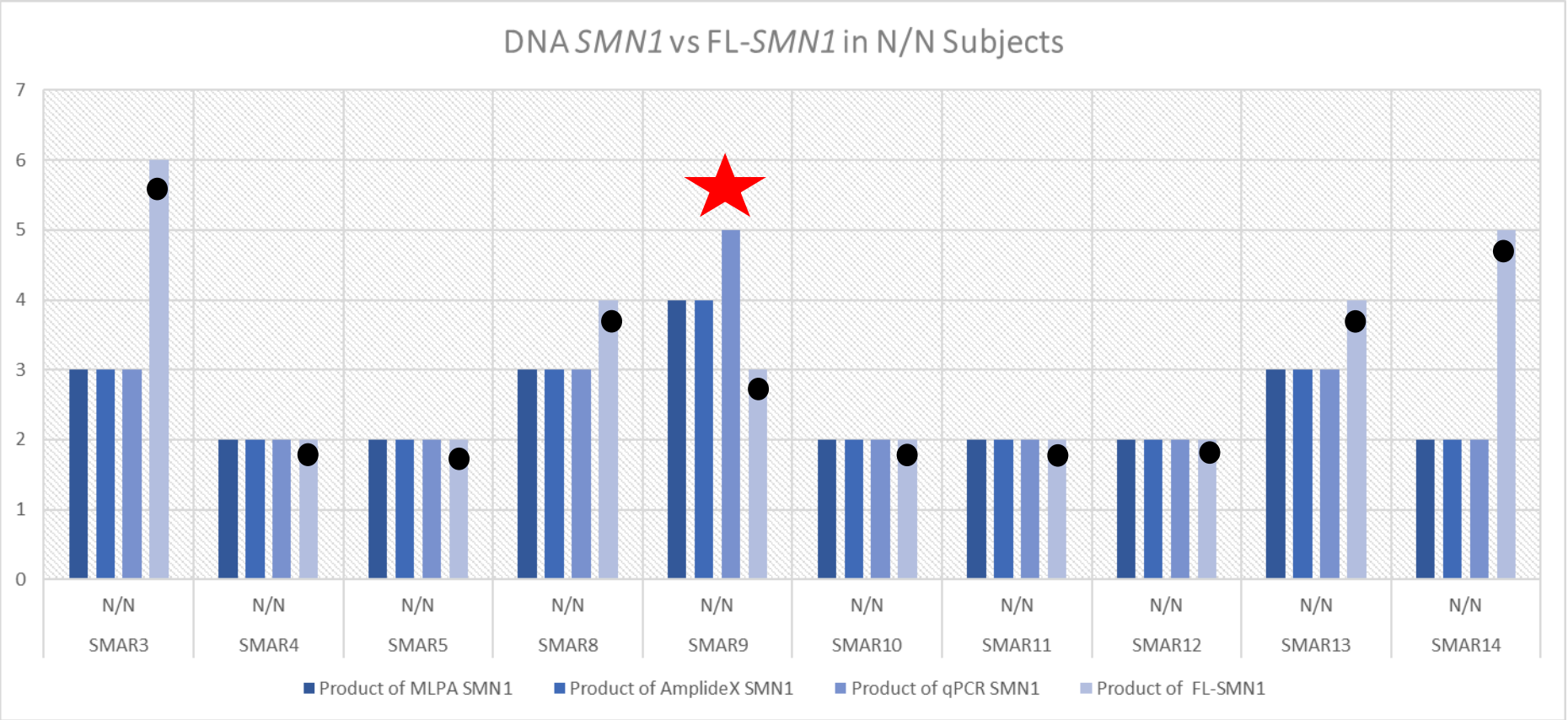


Figure 3.17. DNA copy number versus gene expression results of N/N subjects. Three DNA methods (MLPA, AmplideX and qPCR) were compared to the RT-qPCR expression method. The bar graphs with the black circles indicate FL-SMN1 transcripts. Five out of 10 subjects (SMAR4, SMAR5, SMAR10, SMAR11 and SMAR12) subjects had the same copy numbers on both RNA and DNA analysis. Four out of 10 subjects (SMAR3 SMAR8, SMAR13 and SMAR14) had a FL-SMN1 transcript expression- level which was higher than expected from the DNA *SMN1* copy number. Only one subject, SMAR9 (red star) had decreased FL-SMN1 expression when compared to the DNA *SMN1* copy number and inconsistent copy numbers on AmplideX vs the other methods.

3.2.5 RNA expression results of N/N subjects: DNA copy number of *SMN2* versus FL-*SMN2* transcript

The ten N/N subjects' *SMN2* DNA copy number and RNA expression levels were analysed separately. Figure 3.18 illustrates DNA copy number vs for *SMN2* RNA expression. Only four out of ten N/N subjects had consistent copy number and expression level on DNA and RNA based methods. A surprising difference seen in this group was observed in five of the subjects. These five subjects (SMAR3, SMAR8, SMAR9, SMAR10 and SMAR14) had at least one *SMN2* copy on DNA analysis, however they had zero expression of the FL-SMN2 transcript on RNA analysis. This result suggest that these subjects' *SMN2* DNA copies were not transcribed at all. Figure 3.18 illustrates the association of the DNA copy number with RNA expression in these five subjects. Two subjects (SMAR3 and SMAR4) had increased SMNΔ7 expression compared to their corresponding *SMN2* DNA copy number and FL-SMN2 expression. The *SMN2* DNA copy number and related expression levels appeared to be more variable than that of *SMN1.* Comparison of the RNA and DNA results of *SMN1* and *SMN2,* will be investigated further in the next sections.

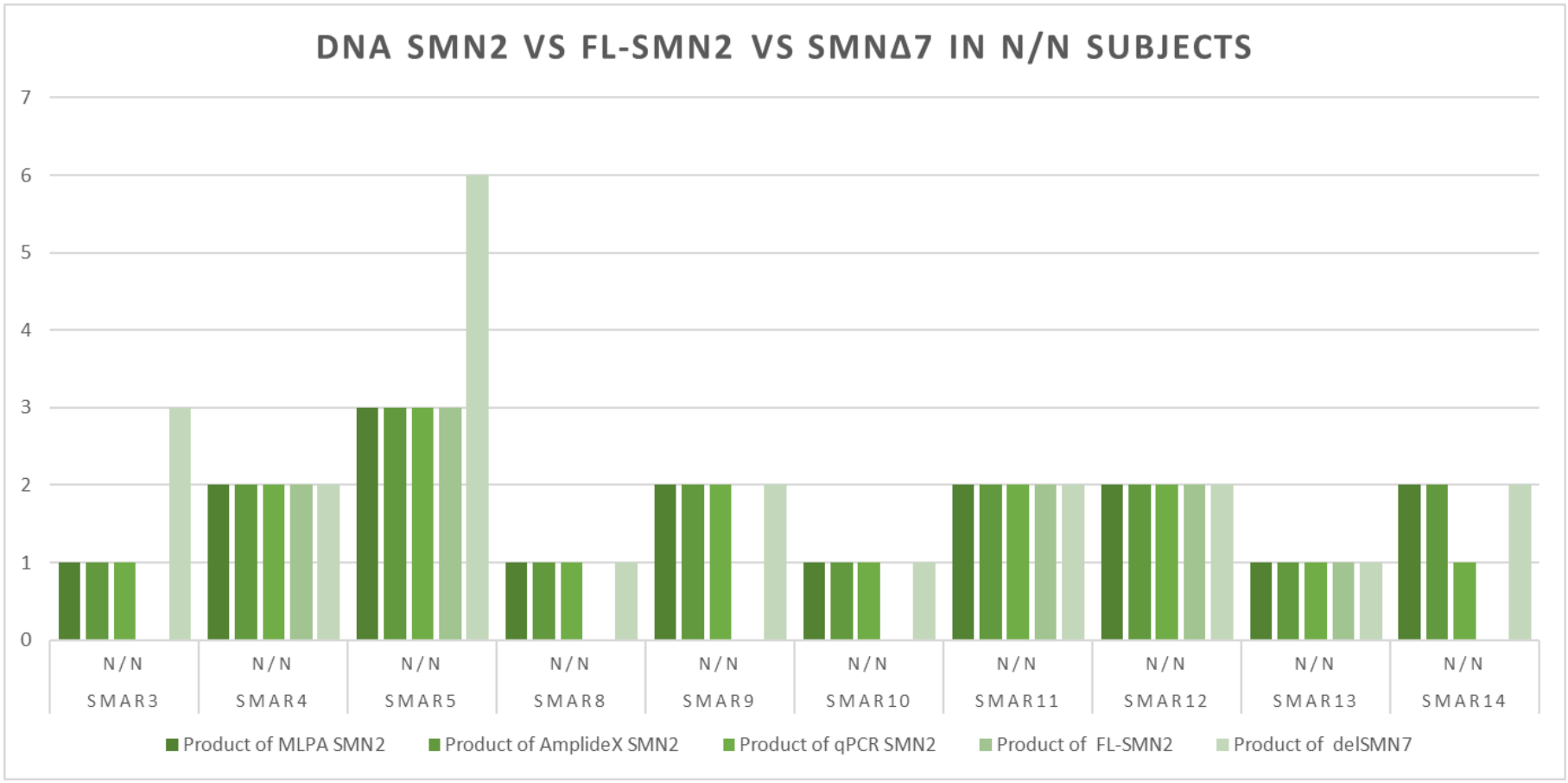


Figure 3.18. *SMN2* DNA copy number versus RNA expression results of N/N subjects. Three DNA methods (MLPA, AmplideX and qPCR) were compared to the RT-qPCR expression method. Four out of 10 subjects (SMAR4, SMAR11, SMAR12 and SMAR13) had the same number of copies of both DNA *SMN2*, as well as equivalent FL-SMN2 and SMNΔ7 expression. Five out of 10 subjects (SMAR3, SMAR8, SMAR9, SMAR10 and SMAR14) had zero FL-SMN2 expression. Two subjects (SMAR3 and SMAR5) had an increased expression of SMNΔ7 transcripts than expected from their corresponding *SMN2* DNA copy number or FL-SMN2 expression.

RNA results of N/N subjects with two DNA copies of SMN1 and SMN2

This section aims to illustrate the variability of both the DNA copy number and expression levels of the *SMN1* and *SMN2* genes in the N/N group. A summary of three N/N subjects with matching copies of both the DNA copy number and expreression levels of both the *SMN1* and *SMN2* genes is shown in Figure 3.19. The DNA copy number of the three subjects is the same as the RNA results (all at two copies).

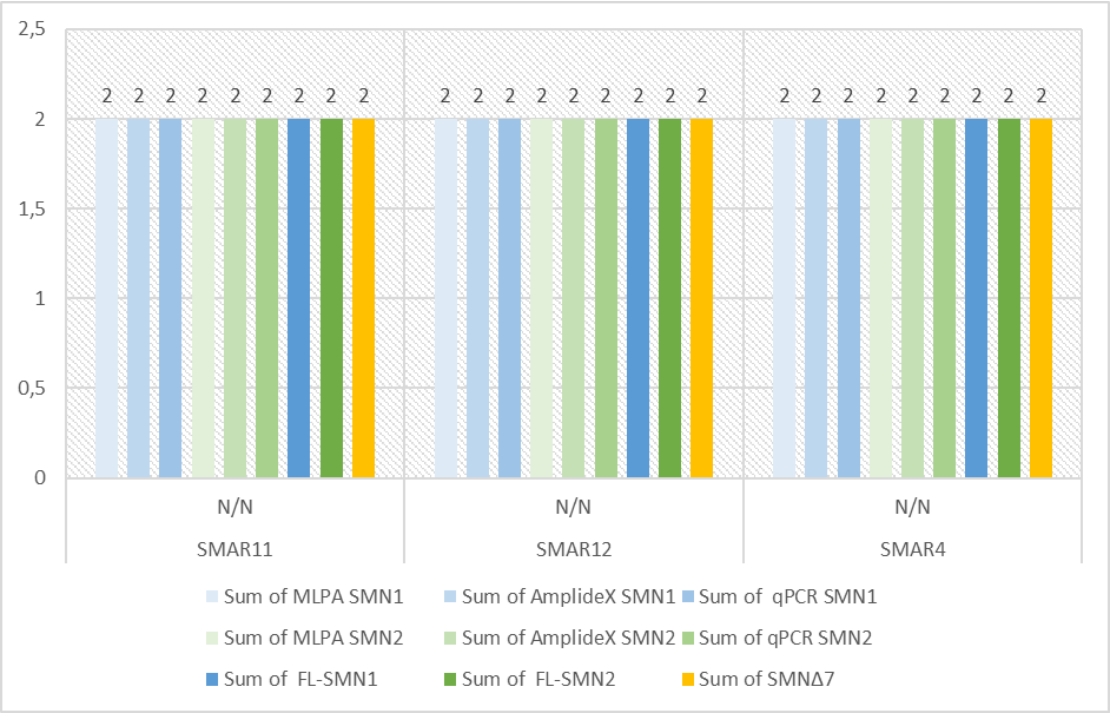


Figure 3.19. DNA copy number versus gene expression results of N/N subjects with matching copies of *SMN1* and *SMN2*. The blue bar graphs represent *SMN1* DNA copy number and FL-SMN1 expression, the green bars represent *SMN2* DNA copy number and FL-SMN2 expression. The orange bar graph represents the SMNΔ7 expression. All three subjects had consistent results on DNA and RNA assays.

***RNA results of N/N subjects with two copies of SMN1 and variable SMN2* expression**

Two subjects, SMAR5 and SMAR10, had two *SMN1* DNA copies and FL-SMN1 expression equivalent to two copies of *SMN1.* Subject SMAR14 had two *SMN1* DNA copies and FL-SMN1 expression equivalent to five *SMN1* DNA copies shown in Figure 3.20. The DNA copy numbers, and transcript expression of these subjects did not correspond as expected. SMAR10 had one copy of *SMN2*, zero expression of FL-SMN2 and SMNΔ7 expression equivalent to one gene copy. These results suggest

that this subject does not have a functional *SMN2* copy. However, the origin of the SMNΔ7 transcript remains uncertain.

Another subject SMAR14, had two DNA copies of *SMN1* and FL-SMN1 expression equivalent to five *SMN1* DNA copies, shown in Figure 3.20. The RNA expression was double what was expected for two *SMN1* DNA copies. This subject had two copies of the *SMN2* gene on MLPA and AmplideX but one copy on qPCR, whereas the transcript levels of this subject did not correlate, with zero FL-SMN2 expression and SMNΔ7 RNA expression equivalent to two gene copies. These results suggest that some of the *SMN2* copies may be *SMN1* copies, incorrectly labelled as *SMN2* copies. This could explain the high FL-SMN1 transcript level observed in this subject. The qPCR and RT-qPCR were repeated for SMAR14, and the same results were obtained. This subject had no obvious gene conversions on MLPA or AmplideX to explain the discrepant results.

The last subject in this section, SMAR5, had two DNA copies of *SMN1* which correlated with the FL-SMN1 expression. The subject had three DNA copies of *SMN2* which correlated with the corresponding FL-SMN2 expression. However, SMNΔ7 expression was very high and was equivalent to six (this gene copies (this phenomenon will be further discussed in section 3.3.9). The SMNΔ7 RT-qPCR was repeated, and the results remained the same. These three N/N subjects did not have the expected two copies of *SMN1* and *SMN2* and did not have the expected results on gene expression. These results further illustrate the variability of *SMN* genes not only at DNA level but also at the RNA level.

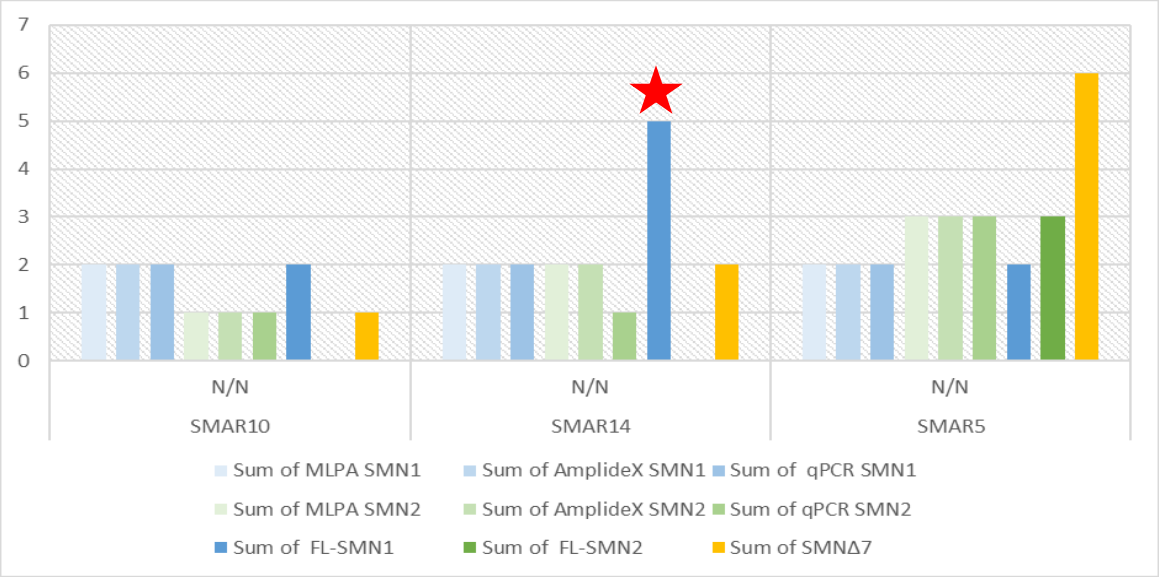


Figure 3.20. DNA copy number versus gene expression results of N/N subjects with variable copies of *SMN1* and *SMN2*. The blue bar graphs represent *SMN1* DNA copy number and FL-SMN1 expression, the green bars represent *SMN2* DNA copy number and FL-SMN2 expression. The orange bar graph represents SMNΔ7 expression. Subject SMAR10, had two DNA copies of *SMN1* which correlated with FL-*SMN1* expression: DNA copy of *SMN2* and zero of the FL-*SMN2*. SMAR14 had two DNA copies of *SMN1* and five of FL-*SMN1*, they had two DNA copies *SMN2* and zero FL-*SMN2* (Discrepancy with the qPCR result). SMAR5 had two DNA copies of *SMN1* and two of FL-*SMN1*. They had three DNA copies of *SMN2*, three of FL-*SMN2* and six of SMNΔ7.

RNA results of N/N subjects with three or more copies of SMN1

The four N/N subjects depicted in Figure 3.21 had three or more DNA copies of *SMN1*. Subject SMAR3 had three DNA copies of *SMN1* and FL-SMN1 expression equivalent to six copies of *SMN1*. The subject had one DNA copy of *SMN2* and zero FL-SMN2 expression, suggesting a possible gene conversion from *SMN2* to *SMN1,* resulting in multiple *SMN1* copies. However, this subject did not have a typical gene conversion on AmplideX testing.

Subject SMAR8 had three DNA copies of *SMN1* and FL-SMN1 expression equivalent to four *SMN1* DNA copies, which is similar but does not correlate. Further, this subject had one DNA copy of *SMN2*, zero FL-SMN2 expression and SMNΔ7 expression equivalent to one gene copy, which may once again indicate a gene conversion from *SMN2* to *SMN1,* resulting in multiple *SMN1* copies.

Another subject, SMAR9 had four DNA copies of *SMN1* on MLPA and AmplideX and five copies of *SMN1* on qPCR and FL-SMN1 expression equivalent to three *SMN1* DNA copy, illustrated in Figure 3.21. This subject had a gene conversion of *SMN2* to *SMN1* on AmplideX analysis, explaining the mechanism by which some of the multiple

SMN1 copies were formed in this subject. The subject also had two DNA copies of *SMN2*, zero expression of FL-SMN2 and SMNΔ7 expression equivalent to two gene copies, further supporting the likelihood of a gene conversion from *SMN2* to *SMN1*. The results of this subject further illustrate the complexity of the *SMN* region.

The last subject in this group SMAR13 had three DNA copies of *SMN1* and FL-SMN1 expression equivalent to four *SMN1* DNA copies (Figure 3.21). SMAR13 had one DNA copy of *SMN2*, FL-SMN2 expression equivalent to one *SMN2* DNA copy and SMNΔ7 expression equivalent to one gene copy. Three copies of *SMN1* in conjunction with one copy of *SMN2* may once again indicate a gene conversion from *SMN2* to *SMN1,* resulting in multiple *SMN1* copies.

All four subjects depicted below were normal adult controls with no known history of SMA, whose DNA results did not fully correlate with RNA results.

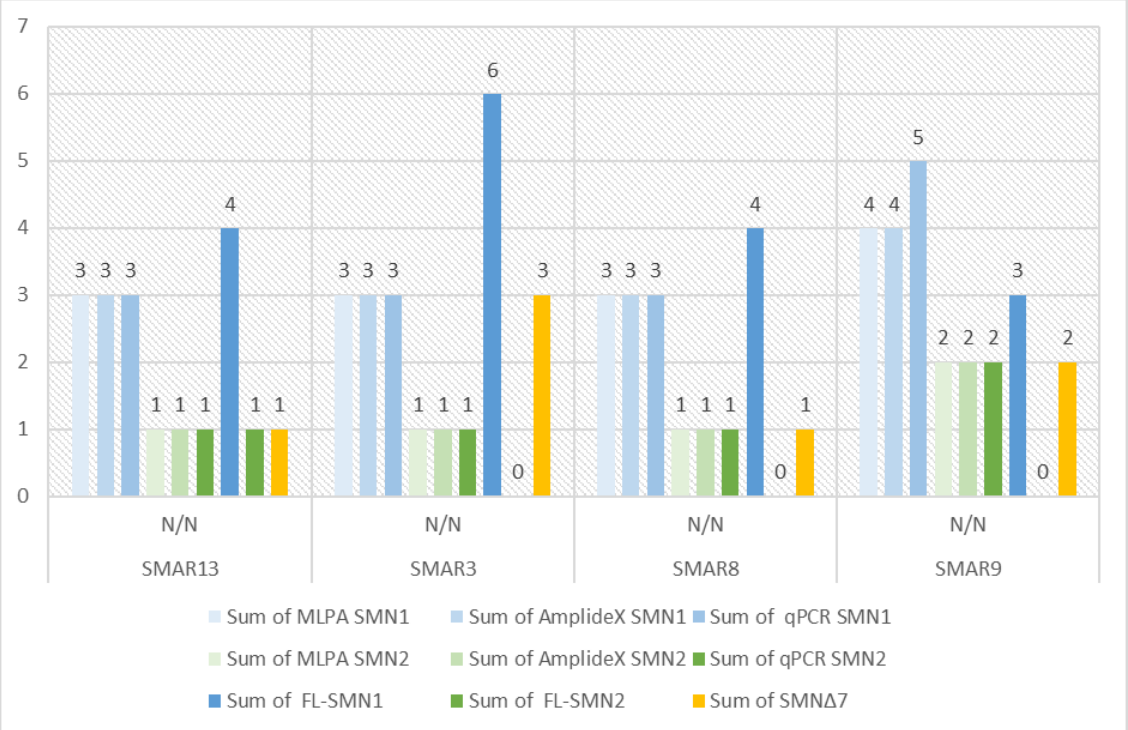


Figure 3.21. DNA copy number versus RNA expression results of N/N subjects with three of more copies of *SMN1***.** All three DNA methods (MLPA, AmplideX and qPCR) were compared to RT-qPCR. The DNA copies of *SMN1* are indicated in blue and FL-SMN1 transcripts in dark blue. DNA copies of *SMN2* are indicated in green and FL-SMN2 in dark green and SMNΔ7 transcripts in orange. All four subjects had three or more copies of *SMN1* and FL-SMN1 expression equivalent to three or more *SMN1* DNA copies.

RNA expression of U/U subjects

Two subjects (SMAR1 and SMAR17) with clinical features of SMA and negative for a homozygous *SMN1* deletion on DNA copy number detection methods were included for RNA expression analysis (Figure 3.22). These subjects had at least one DNA copy of *SMN1* according to the RFLP diagnostic method. The subjects were tested for RNA expression to determine if they had FL-SMN1 transcripts. Absence of FL-SMN1 transcripts would confirm the clinical diagnosis of SMA in the U/U subjects.

SMAR1 had three DNA copies of *SMN1* and correlating FL-SMN1 expression equivalent to three *SMN1* DNA copies (shown in Figure 3.22). The subject also had one DNA copy of *SMN2*, zero expression of FL-SMN2 and gene expression of SMNΔ7 equivalent to one gene copy. SMAR17 had two DNA copies of *SMN1* and correlating FL-SMN1 expression equivalent to two *SMN1* DNA copies. The subject had two DNA copies of *SMN2*, correlating with FL-SMN2 expression equivalent to two *SMN2* DNA copies and SMNΔ7 expression equivalent to three gene copies. Both SMAR1 and SMAR17 had RNA expression of the FL-SMN1 and FL-SMN2 transcripts; therefore, the clinical diagnosis of SMA was not confirmed.

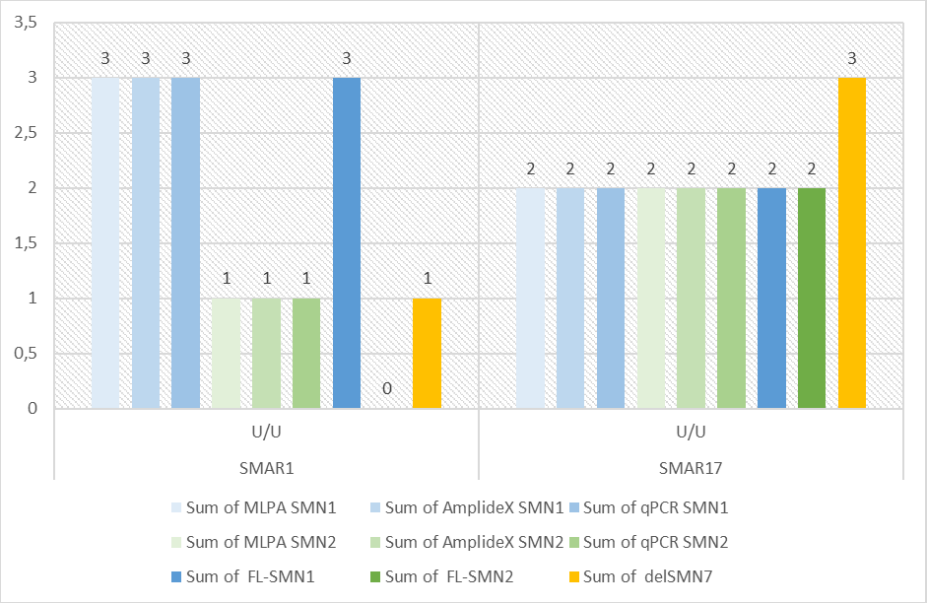


Figure 3.22. DNA copy number versus RNA expression results of U/U subjects. SMAR1 had three DNA copies of *SMN1* (blue) and FL-SMN1 expression equivalent to three *SMN1* DNA copies (dark blue). Both these patients also had one DNA copy of *SMN2* (green), zero expression of FL-SMN2 and zero expression of the SMNΔ7 (orange) transcript. SMAR17's DNA copies correlated with its RNA expression.

Comparison of RNA expression in families affected with SMA.

Two families, who had family members, who were confirmed to be affected with SMA on RFLP, were included for RNA expression testing. The aim was to determine whether the *SMN* transcript levels reflect what occurs at the DNA level.

Family One consists of a mother who is a confirmed true carrier (1:0 genotype) of *SMN1* on DNA methods and her son, who is affected with SMA and was found to have zero DNA copies of *SMN1* (homozygous deletion,0:0 genotype). The DNA and RNA expression levels in the mother and her son correlated. The RT-qPCR was able to accurately determine the *SMN1* carrier status of the mother. Both mother and child had two DNA copies of *SMN2.* The FL-SMN2 expression was increased in both mother and son. The SMA severity in the affected child was not provided and could therefore not be correlated with the *SMN2* copy number.

Family Two also consisted of a mother who is a confirmed true carrier (1:0 genotype) of an *SMN1* and her affected daughter, who is affected with SMA and was found to have zero DNA copies of *SMN1* (homozygous deletion, 0:0 genotype) confirmed by DNA copy number detection methods. The mother had DNA copy number and RNA expression level equivalent to one copy of both *SMN1 and SMN2*. The affected daughter had zero DNA copies of *SMN1* and zero expression of FL-SMN1, as expected. The daughter had two DNA copies of *SMN2* and FL-SMN2 expression level equivalent to four *SMN2* DNA copies.

Both affected children had increased expression levels of FL-SMN2 compared to the corresponding DNA copies. *SMN1* DNA results and RNA expression results correlated for all children and mothers, thus accurately confirming the diagnosis of SMA in the children and carrier status of the mothers.

RNA expression results of SMN∆7

The levels of *SMN∆7* was calculated to determine whether multiple copies of the *SMN1* gene correlated with increased SMN∆7 RNA expression. It has been hypothesised that multiple *SMN1* copies (≥3) frequently observed in the black SA population, may not be complete or functional copies.

3.3 Summary of results

This chapter has outlined the results from the three DNA copy number detection methods used. The MLPA method was kit based and is the recommended method for SMA testing, however it is expensive, time consuming and was inadequate for carrier testing of black subjects since silent carriers with multiple *SMN1* copies on a single chromosome will be missed. Two other methods AmplideX (new method) and qPCR (designed in house) were tested to determine their clinical utility and whether they could provide alternatives to MLPA. Both AmplideX and qPCR were optimised and validated before comparative analysis.

Comparative analysis of the three DNA copy methods indicated that they were all adequate for diagnostic and carrier testing of SMA subjects. There was a discrepancy in one subject for the *SMN1* gene copy number. The *SMN2* DNA copies were mostly the same but varied more between the three methods when compared with *SMN1* gene copies.

The AmplideX kit was the only method able to detect gene conversions (hybrid copies). These hybrid copies resulted from gene conversion of *SMN1*-to-*SMN2* and *SMN2*-to-*SMN1.* Gene conversions were observed in 6/89 (7%) of subjects tested. Most of the hybrid copies were seen in subjects with two DNA copies of *SMN1.* Only *SMN2*-to-*SMN1* hybrids were observed. Analysis of other *SMN* exons on MLPA for these subjects did not confirm or disprove gene conversions. Diagnostic testing for gene conversions does not seem to add any value. It is also possible that the AmplideX kit may miss gene conversions/rearrangements unique to black SA individuals.

The qPCR method was designed in-house, it was the least expensive and required the same time as AmplideX. The method, however, had more amplification failures. This was likely due to poor DNA quality and quantity as it requires the most DNA of all of the DNA based methods (set up in triplicates with 50ng in each well). Comparison of these methods indicated that they each had their advantages and limitations.

RNA expression analysis by RT-qPCR was performed to attempt to further understand the mechanism of SMA. This method was designed in house, optimised and results then compared with DNA copy number results. Five subject groups were analysed and comparative analysis of DNA *SMN1* and *SMN2* gene copy numbers and RNA expression levels were performed. The aim of RT-qPCR was to determine if clinically

suggestive SMA patients who have tested negative for deletions of *SMN1* on DNA studies could be found to have non-functional *SMN1* copies on RNA expression studies. These U/U subjects were hypothesised to have zero FL-SMN1 transcripts and increased SMN∆7 transcript levels. However, this was not true as both U/U subjects showed expression of FL-SMN1. SMAR17 had an increased SMN∆7 transcript level, which may indicate the presence of at least one non-functional *SMN1* copy, however this cannot confirm a diagnosis of SMA. Similar to the DNA methods FL-SMN2 expression was much more variable across methods and did not always correlate, as would have been expected, with SMNΔ7 expression.

The most exciting observation made with RNA expression studies was in subjects with unusually high *SMN*∆7 transcript levels (copy numbers of ≥3). These subjects' high SMN∆7 transcript levels could not be explained by *SMN2* copy number alone and may be due to non-functional *SMN1* copies. This would support the hypothesis that some of the multiple *SMN1* gene copies observed in black SA individuals may be incomplete/non-functional copies which may mask the true carrier and clinical status of subjects when using traditional DNA copy number methods. These four subjects were more likely to have variable expression levels of FL-SMN1, FL-SMN2 and SMNΔ7.

RNA expression studies showed the complexity of the *SMN* region and emphasises the complex relationship of DNA copy number versus RNA expression. The following chapter will aim to unpack these complex relationships by comparing these results with the literature.

4 DISCUSSION

This chapter will discuss the results obtained from genomic copy number analysis using the MLPA, AmplideX® PCR/CE *SMN1/2* and qPCR techniques and how they compared to previous studies done on SMA, particularly in black patients. The methods will be compared to determine which offers the best clinical utility for diagnostic and carrier testing of SMA in South African populations.

RNA expression analysis using RT-qPCR will be discussed and compared with the DNA methods to determine whether it resolved some of the limitations seen with DNA studies. This chapter will comment on the advantages and limitations of both the DNA techniques and RNA expression methods by comparing results to articles published on current SMA testing and how testing in South Africa equates to international guidelines.

The first section (4.1) will focus on DNA copy number detection methods, followed by a comparison of *SMN1* results between the three methods (sections 4.1.1 – 4.1.5) and *SMN2* results between the three (4.1.6 – 4.1.7) followed by a discussion on gene conversion in section 4.1.8. RNA expression (RT-qPCR) results will be discussed in section 4.2, the clinical utility of *SMN1/SMN2* testing in section 4.3, potential future studies in section 4.4 and finally the advantages and limitations in section 4.5.

4.1 DNA copy number methods – Comparison and their clinical utility

Spinal muscular atrophy is an autosomal recessive disorder commonly caused by a homozygous deletion of exon 7 in the *SMN1* gene. Accurate detection of the homozygous deletion is essential in confirming diagnosis. The second *SMN* gene copy - *SMN2* modifies SMA disease, the more copies an affected SMA individual has the less severe the disease is.

These methods have been compared to determine their clinical utility for diagnosis of SMA, determining carrier status, ability to accurately quantify copy numbers and their cost effectiveness.

4.1.1 Comparison of MLPA DNA copy numbers

The study done by Vorster et al (2020) investigated 197 subjects to determine the molecular mechanism of SMA disease in the black SA population. The authors compared differences in copy number between black and white subjects and determined carrier frequencies in the two populations and discussed the diagnostic utility of MLPA for SMA testing in SA populations. The study aimed to identify possible pathogenic CNVs in black SMA patients with clinical features suggestive of SMA (U/U), but no common pathogenic CNVs were identified. This study aimed to continue the investigation from Vorster et al (2020), to further understand the molecular mechanism of SMA disease and to determine the clinical utility of different DNA and RNA expression methods.

SMN1 gene detection by MLPA

Approximately 95-98% of SMA affected individuals show a homozygous deletion of *SMN1* worldwide. However only 51% of cases can be confirmed in the black SA population. This study had 92 subjects (71 black and 21 white). All 13 subjects with the M1/M1 genotype had a homozygous deletion of *SMN1*, indicated in table 3.2 (section 3.1.1). Therefore, a diagnosis of SMA could be confirmed in these subjects.

In the M1/M1 group only 4 out of 13 (31%) subjects (12 black and one white) had a homozygous deletion of both exons 7 and 8. The M1/M1 sample size was very small, however the results illustrated the complexity of the *SMN* region and genetic diversity in the black SA population compared to other populations. This difference in extent of *SMN1* deletions has been previously documented in studies from South Africa (Labrum et al., 2007; Stevens et al., 1999; Vorster et al., 2020). Further, the difference in copy numbers between exons in the same gene was proposed to be due to complex rearrangements which would not be detected by MLPA (Vorster et al., 2020). This further highlighted the need for further analysis of the *SMN* region by methods that might accurately determine the arrangement of *SMN1* gene exons as well as multiple gene copies (see section 4.4 for a further discussion on possible future studies of the SMN region).

The MLPA results obtained from this study supported the findings seen by Vorster et al (2020) - the *SMN* region in black subjects appeared to be more complex than white

in subjects (Vorster et al., 2020). In this study, 28% (20/71) of black subjects had three or more copies of *SMN1* compared to 9% (2/21) of white subjects.

In the M2/M2 genotype group (n=14 black subjects), the average *SMN1* copy number was ≥3. Furthermore 40% of N/N subjects who were black had ≥3 *SMN1* copies and 20% of U/U subjects had ≥3 *SMN1* copies. The previous study done by Vorster et al (2020), showed that 50.8% of all black N/N subjects had multiple copies of *SMN1* compared to this study's 40%. The observation could be due to a difference in sample size as the previous study had 122 subjects compared to this study's 10 subjects.

The multiple copies of *SMN1* have also been noted in previous studies performed in African populations - approximately 47% (475/1015) of African Americans had three copies of *SMN1* (Hendrickson et al., 2009). Studies done in different African countries further support the prevalence of multiple copies of *SMN1*, with multiple *SMN1* copies observed in approximately 52.5% of Malians, 37% Nigerians and 38% of Kenyans (Sangare, et al., 2014). All these studies illustrate that individuals of African ancestry have more copies of *SMN1* compared to other populations. The physical arrangement of these gene copies remains uncertain. Further it is not determined whether the exons are contiguous, or the genes are functional (Vorster et al., 2020).

SMN2 gene detection by MLPA

The majority of SMN2 transcripts lack exon 7, it is thus essential to distinguish between the two genes. The range of *SMN2* copies is 0 to three in healthy individuals, provided they have at least one *SMN1* copy. The range of *SMN2* copies is one to four copies in those affected with SMA (Butchbach, 2016). This study showed similar ranges of zero to ≥3 *SMN2* copies in the genotypes M2/M2, N/M1 and N/N (potentially healthy individuals). Affected SMA subjects M1/M1 had 2 to ≥3 *SMN2* copies. M1/M1 subjects with zero or one copy of the *SMN2* gene were not observed in this study and is suspected to be incompatible with life (Butchbach, 2016).

Multiple copies of *SMN2* were less frequently observed: 8.5% (6/71) of black subjects tested had ≥3 copies of *SMN2* compared to 28% seen for *SMN1*. The *SMN2* copy number was similar to those seen by Vorster et al (2020). at 11.5%. Multiple copies of *SMN2* are not as common as *SMN1*.This has been attributed to the evolution of the *SMN* region (Vorster et al., 2020).

The MLPA method not only detected *SMN1*, exon 7 and *SMN2*, exon 7 copy numbers but also the copy numbers of other *SMN1/SMN2* exons and neighbouring genes in this region. The SALSA MLPA P021 probe kit is able to detect copy numbers of the *GTF2H2*, *RAD17*, *NAIP* and *SERF1B* genes. The kit also contained probes for exons 1, 4, 6 and 8 of both *SMN1* and *SMN2*, but could not distinguish between copies derived from the *SMN1* or the *SMN2* gene.

The advantages of MLPA were its ability to detect homozygous deletions of both *SMN1* and *SMN2*, detect true carriers (1:0 genotype) and to determine the extent of deletions by looking at adjacent *SMN1/SMN2* exons and genes. The limitation of MLPA was the hands-on time of 120 minutes, cost of testing compared to the in house RFLP method and inability to detect gene conversion (hybrid) *SMN* copies.

4.1.2 *SMN1* gene detection by MLPA vs AmplideX vs qPCR

Once AmplideX and qPCR were optimised and validated all samples were tested using these methods. The two methods were compared to MLPA as it is the gold standard for copy number detection and specifically for SMA diagnostic and carrier testing (Verhaart et al., 2017). Table 4.1 (Summary of table 3.5) illustrates the minimum, maximum and average *SMN1* copy numbers observed using the three methods. As seen below there was only one discrepancy, seen in the U/U group. The minimum *SMN1* copy number observed in the U/U group was two for MLPA, however AmplideX and qPCR had one copy each. This discrepancy was caused by one subject. Overall, 92 subjects were tested on all three methods of which three failed amplification on qPCR and therefore only 89 subjects could be compared. Only one subject had a discrepant *SMN1* copy number resulting in 88/89 (98.8%) correlation between MLPA, AmplideX and qPCR for detection of *SMN1* copy numbers. When comparing only AmplideX and qPCR there was a 100% correlation. Therefore, all three techniques could accurately diagnose SMA. The discrepant result will be discussed in section 4.1.3.

Table 4.1. Comparison of *SMN1* copy numbers on MLPA/AmplideX/qPCR

Genotype	Minimum *SMN1*			Maximum *SMN1*			Average *SMN1*		
	MLPA	AmplideX	qPCR	MLPA	AmplideX	qPCR	MLPA	AmplideX	qPCR
M1/M1	0	0	0	0	0	0	0	0	0
M2/M2	2	2	0	≥3	≥3	≥3	≥3	≥3	≥3
N/M1	1	1	1	2	2	2	1	1	1
N/N	2	2	2	≥3	≥3	≥3	2	2	2
U/U	**2**	**1**	**1**	≥3	≥3	≥3	2	2	2

*One U/U subject with discrepant result is indicated in bold.

4.1.3 Possible reasons for *SMN1* copy number discrepancies.

The subject with two copies of *SMN1* on MLPA and one copy on both AmplideX and qPCR is shown in Figure 4.1. This subject had clinical features suggestive of SMA and did not have the homozygous *SMN1*, exon 7 deletion on the RFLP diagnostic assay. There are multiple reasons for discrepant copy number results. Firstly, there could be a single nucleotide polymorphism (SNP) in the primer binding regions of the AmplideX and qPCR primers (~20bp). The primer sequences of the AmplideX kit were not available to investigate the overlap. A SNP in the patient may prevent primer binding and may appear as a deletion on quantitative analysis. MLPA probes are quite large (~100bp to 450bp) and could still bind to the region of interest if the potential SNP is not located at the ends of the probe, which could explain why the MLPA probe bound successfully and the AmplideX and qPCR primers did not.

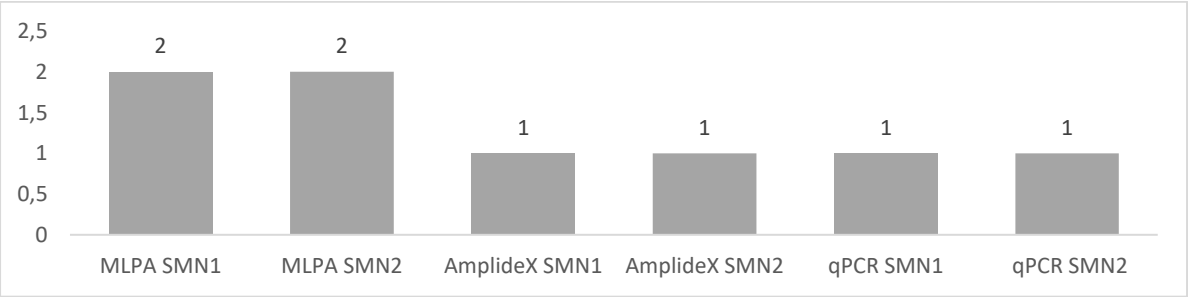


Figure 4.1. U/U Subject with *SMN* copy number discrepancy**.** The subject had two *SMN1* and *SMN2* copies on MLPA. They had only one copy of each on AmplideX and qPCR.

Secondly the subject could have a gene conversion of *SMN2*-to-*SMN1* only detected by MLPA. There are complex gene rearrangements in the *SMN* region, and it is possible that not all would be detected by AmplideX or qPCR. DNA sequencing of the *SMN* region of this subject could possibly explain if this subject only has one *SMN1* copy or a potential gene conversion.

Thirdly this subject could possibly have a deletion of *SMN1* on one chromosome and a point mutation on the second chromosome. This option could be clinically significant. It is estimated that point mutations account for 2% of all SMA cases. All three methods used were only able to detect deletions and duplications and would not have been able to detect point mutations. There are several point mutations described in the *SMN1* gene all causing SMA (Sharifi et al., 2021; Sun et al., 2020; Zapletalová et al., 2007). Continuous analysis of the *SMN1* gene has shown that deletions are not the only cause for SMA. If the subject has a pathogenic point mutation and clinical features suggestive of SMA, it could be supportive of a diagnosis of SMA, regardless of whether the subject has one or two copies of *SMN1*. This would only be determined by further sequencing of the *SMN1* gene.

4.1.4 *SMN1* gene copy number in U/U subjects

There were 30 U/U subjects tested for DNA copy number. These subjects had clinical features suggestive of SMA and all had at least one *SMN1* copy. These subjects were selected because their clinical symptoms were reported to be indistinguishable from those of SMA. There are many reasons why these subjects remain undiagnosed. Figure 4.2 summarises a suggested diagnostic algorithm for SMA testing. Detection of *SMN1* and *SMN2* copy numbers is recommended as the initial starting point. DNA sequencing is the recommended second step for a consanguineous family. Complete family histories in the South African patients are not always available, but high rates of consanguinity do not seem a likely explanation. Therefore, this approach would not be feasible in a SA diagnostic setting. Further clinical evaluation is recommended to determine a differential diagnosis of other neuromuscular disorders is a recommended step in undiagnosed SMA patients. Due to the numerous genes responsible for neuromuscular disorders gene panels or whole exome/genome sequencing is recommended (Mercuri et al., 2018). Testing for the homozygous deletion of *SMN1* followed by NGS using WES might be the approach in patients suspected of having SMA.

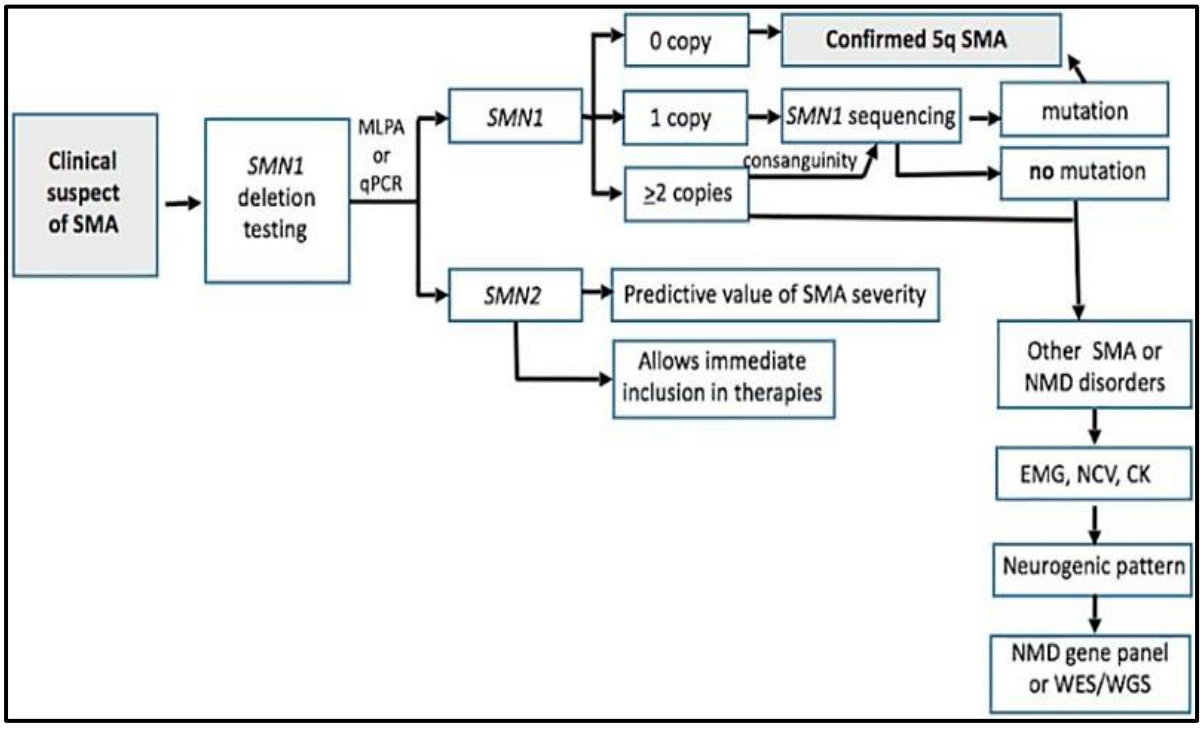


Figure 4.2. Diagnostic algorithm of SMA (Mercuri et al., 2018).

4.1.5 Comparative analysis of N/N subjects - *SMN1* copy number in black population compared to other populations.

The *SMN1* gene copy numbers observed in this study were compared to previous studies and summarised in table 4.2. In this study 60% (n=10) of subjects had two copies and 40% had three or more copies of *SMN1*. Individuals from African populations are more likely to have ≥3 copies than other populations. The sample size in this study (n=10 N/N subjects) was much smaller than other studies, however a similar pattern of multiple copies of *SMN1* could be seen within patients of African ancestry in this study.

Table 4.2. Frequency of N/N *SMN1* Copy Number in Different Geographical Regions

***SMN1* Copy#**	**sub-Saharan Africa**					**North American**		**Europe**			**Asia**	
	(Present study and Vorster, et al ., 2020)		**(Sangare, et al., 2014)**			**(Vijzelaar et al., 2019)**		**(Anhuf et al., 2003 and Corcia et al., 2006)**			**(Lee et al., 2004 and Hasanzad et al., 2009)**	
	South Africa, N/N n=10*	**South Africa 2013-2017, n=122**	**Mali, n=628**	**Nigeria, n=120**	**Kenya, n=120**	**African American, n=1015**	**Caucasian, n=1020**	**Germany, n=140**	**France, n=621**	**Sweden, n=502**	**South Korea, n=326**	**Iran, n=200**
1	0 (0%)	0 (0%)	3 (0.5%)	2 (2%)	1 (1%)	11 (1%)	28 (2.7%)	4 (4%)	13 (2%)	9 (2%)	7 (2%)	10 (5%)
2	6 (60%)	60 (49.2%)	295 (47%)	72 (60%)	73 (61%)	529 (52%)	935 (91%)	132 (94%)	593 (96%)	479 (95%)	254 (78%)	150 (75%)
≥3	4 (40%)	62 (50.8%)	330 (53%)	46 (38%)	46 (38%)	475 (47%)	65 (6.3%)	4 (3%)	15 (2%)	14 (3%)	65 (20%)	33 (17%)

*Present study, n=10 (black SA subjects)

4.1.6 SMN2 gene detection by MLPA vs AmplideX vs qPCR

Mutations in *SMN2,* as explained previously in section 1.3, do not cause SMA but *SMN2* is a known modifier of SMA disease. Table 4.3 summarises the *SMN2* copy numbers observed. Three subjects did not have the same *SMN2* copy number across all three methods. One subject was an M1/M1 subject with two copies of *SMN2* on MLPA and ≥3 copies on AmplideX and qPCR. The diagnosis of SMA still remained accurate, however the discrepant *SMN2* copy number could be important for patients undergoing therapy. The recommendation is that a minimum of 2 *SMN2* gene copies are required for eligibility for therapy, so the patient would theoretically qualify (Deutsch et al., 2021). The predicted impact of therapy could be different.

The second subject was an N/N subject with two copies of *SMN2* on both MLPA and AmplideX and one on qPCR. The *SMN2* copy number in this subject was not critical as the subject did not present with any clinical symptoms and had no family history of SMA. The third subject had two copies of *SMN2* on both MLPA and qPCR but had one copy on AmplideX. This subject had clinical symptoms suggestive of SMA with at least one *SMN1* gene copy. The clinical significance of the *SMN2* copy number in this subject was not critical as their diagnosis remained unknown.

Discrepancies were observed in all three methods. There was no single method that correlated 100% with the other, methods - therefore the discrepant *SMN2* copy numbers could be due to miscalculation of the *SMN2* copy number or the complexity of the *SMN* region. The exact reasons for these discrepancies are uncertain and it is unknown which method correctly detect the copy number most accurately. Since there is so much variability in the *SMN1* copy number in black SA subjects, this could potentially complicate/jeopardise accurate *SMN2* copy number calculations. The best way to report on *SMN2* DNA copy is zero (homozygous deletion), one *SMN2* DNA copy detected or at least two *SMN2* DNA copies present. A high number of homozygous *SMN2* deletions were observed in black subjects. This was not surprising as it has been seen before in populations of African ancestry (Sangare, et al., 2014; Vorster et al., 2020).

Table 4.3. Comparison of *SMN2* copy numbers between MLPA/AmplideX/qPCR

Ethnicity	Genotype	Number of subjects detected per genotype			
		DNA copy number observed	*SMN2* MLPA	*SMN2* AmplideX	*SMN2* qPCR
Black	M1/M1	0	0	0	0
		1	0	0	0
		2	9	8	8
		≥3	3	4	4
	M2/M2	0	10	10	10
		1	0	0	0
		≥3	0	0	0
	N/M1	1	4	4	4
		2	5	5	5
	N/N	1	4	4	5
		2	5	5	4
		≥3	1	1	1
	U/U	0	1	1	1
		1	11	12	11
		2	15	14	15
		≥3	2	2	2
White	M1/M1	0	0	0	0
		2	1	1	1
	M2/M2	0	4	4	4
		2	0	0	0
		≥3	0	0	0
	N/M1	1	1	1	1
		2	3	3	3
		≥3	1	1	1
	N/N	1	2	2	2
		2	7	7	7
Total			89	89	89

*Values in red represent discrepant copy numbers. Three subjects did not have same SMN2 copy number across all three methods.

4.1.7 Comparative analysis of N/N subjects - SMN2 copy number in black population compared

The *SMN2* copies were compared with other populations as shown in Table 4.4. The table illustrates frequencies calculated by Sangare et al 2014, Anhuf et al 2003 and Corcia et al 2006. Frequencies detected by Vorster et al (2020), were also included for comparison. All 122 subjects (black SA population) in that study were from individuals with no family history of SMA and did not present with any symptoms of SMA. The sample size of N/N subjects in this study was much smaller (n=10). Homozygous deletions (zero copies) of *SMN2* appear to be more common in sub-Saharan populations. In this study a homozygous deletion of *SMN2* was not observed in the N/N subjects, possibly due to our small sample size. We compared the Vorster et al (2020) with other populations. The most common copy number in South Africa was two (59% in the Vorster et al (2020). study), which was similar to European populations. Zero copies of *SMN2* were more common in sub-Saharan populations than in Europeans with Nigeria having the highest frequency at 27%. The differences in *SMN2* copy number suggest some sort of gene conversions of *SMN2*-to-*SMN1* in sub-Saharan populations, resulting in multiple *SMN1* copy numbers. Another hypothesis relates to the evolution of *SMN2.* It has been suggested that the duplication of the *SMN* region to form *SMN1* and *SMN2* happened at different times in European and African populations (Rochette et al., 2001). This could mean that further *SMN* duplications and *SMN1/2* gene conversions may be unique in African populations. In addition, the methods used to detect *SMN1* and *SMN2* gene copy numbers have been optimised in European populations. It is unclear how well they reflect true gene copy numbers in African populations. Further testing using long range DNA sequencing may provide answers by looking at the structure and orientation of multiple *SMN1/2* gene copies.

Table 4.4. Frequency of N/N SMN2 Copy Number in Different Geographical Regions

SMN2 Copy#	sub-Saharan Africa					Europe		
	South Africa 2021 n=10*	**South Africa 2013-2017 Cohort n=122**	**Mali, n=5 613**	**Nigeria , n=5 120**	**Kenya, n=5 120**	**Germany, n=5 100**	**France, n=5 621**	**Sweden , n=5 502**
0	0(0%)	15 (12.3%)	150 (24%)	33 (27%)	23 (19%)	9 (9%)	52 (8%)	37 (8%)
1	4 (40%)	20 (16.4%)	276 (45%)	66 (55%)	53 (44%)	38 (38%)	239 (38%)	212 (42%)
2	5(50%)	73 (59.8%)	172 (28%)	21 (18%)	38 (32%)	48 (48%)	321 (52%)	247 (49%
≥3	1 (10%)	14 (11.5%)	15 (3%)	0 (0%)	6 (5%)	5 (5%)	9 (2%)	6 (1%)
Study	**Present study**	(Vorster, et al., 2020)	(Sangare, et al., 2014)			(Anhuf et al., 2003)	(Corcia et al., 2006)	

*Ten black SA subjects with the N/N genotype. The n, represents sample size of the population in each study

4.1.8 Gene conversion and DNA copy number

The AmplideX kit is a new kit designed for quantification of *SMN1/2* gene copy number. The kit has fluorescently labelled primers for *SMN1*, exon 7 and *SMN2*, exon 7 as well as an endogenous control. In addition to detecting *SMN1* and *SMN2* gene dosage the kit has been designed to detect *SMN1*-to-*SMN2* and *SMN2*-to-*SMN1* gene conversions (Asuragen, 2020). Hybrid *SMN1* contains *SMN1*, exon 7 and *SMN2*, intron 7. An *SMN2* hybrid contains *SMN2*, exon 7 and *SMN1*, intron 7. The kit was evaluated as an alternative diagnostic method for SMA testing. The kit was able to diagnose homozygous deletions of *SMN1*, exon 7, determine traditional carriers (1:0 genotype) as well as gene conversions. The kit however did not detect the copy number of adjacent exons or genes in the *SMN* region and therefore, the extent of deletions could not be determined. The extent of *SMN* deletions is not generally required in a diagnostic setting and does not alter results or influence treatment. However, deletion size has been reported to correlate with severity in European populations (Lefebvre et al., 1997; Wirth et al., 2006). The AmplideX kit provided similar results to the MLPA kit. The kit used less DNA (20ng) and was less laborious; however, the cost of testing (R627.62) was similar to that of MLPA (R744.05).

Frequency of SMN gene conversions

There were 6/92 (7%) subjects found to have gene conversions, using the AmplideX kit. All subjects had the *SMN1* hybrid (*SMN1* exon 7 and *SMN2* intron 7) gene. Five of the subjects with hybrid copies had two copies of *SMN1.* A reassessment of MLPA results did not indicate any obvious gene conversions and qPCR primers cannot

detect gene conversions. Only one subject from the N/N group who had ≥3 copies of *SMN1* had a gene conversion. This was surprising as 40% of our N/N subjects had ≥3 copies of *SMN1* (Table 4.2). This suggests that gene conversions may not be the only mechanism of creating multiple *SMN1* copies in the black SA population – there could be potential large rearrangements of the *SMN* region, not detectable with current methods. It is also possible that gene conversions of the *SMN* region in Africans over years of evolution have resulted in multiple *SMN1* copies being part of the normal/general haplotypes seen in African population, and not due to recent gene conversion. Further, the AmplideX kit may only be optimised to detect some gene conversion events and may not detect other events with different breakpoints.

None of the M1/M1 subjects had gene conversions, perhaps supporting Rochette et al. (2001), that duplication of the *SMN* region occurred at different times in the European and African populations. Sangare et al. (2014) found 19% of N/N subjects to have *SMN2*-to-*SMN1* gene conversions (*SMN1* hybrids). They had a larger populations size (246 healthy controls) compared with only 10 N/N black SA subjects in this study. Sangare et al. also concluded that gene conversions events were far too few to explain the multiple copies found in African populations.

Possible role of SMN gene duplications and gene conversion

Gene duplications have been observed not only in the *SMN* region but in other regions in the human genome. Most gene duplications are not pathogenic but are likely to become a source for further gene rearrangements. These complex rearrangements are likely in turn to cause further gene conversions as they may be predisposed to non-homologous recombination events. Gene conversions may interrupt proper translation and result in non-functional proteins (Chen et al., 2007).

The 500kb *SMN* region is thought to have originated from an inverted duplication that gave rise to the *SMN1* and *SMN2* genes. The inverted duplication occurred after divergence of the humans and mouse lineages; therefore, mouse lineages only have one copy of the *SMN* gene. The inverted duplication occurred before divergence of the humans and chimpanzee lineages, but the *SMN2* gene is unique to humans (Rochette et al., 2001). The average *SMN1* copy number is reported to be three and two in healthy individuals in African populations and European populations, respectively (Vijzelaar et al., 2019). Further, Africans were more likely to have more

than three copies of *SMN1* compared to other populations, suggesting that further duplication/rearrangement of the *SMN* region may have taken place in African individuals (Vorster et al., 2020).

The hybrid copies identified in this study may not have provided the full picture of the structure of the *SMN* region in black SA subjects. Multiple full or partial copies of *SMN1* may mask the true carrier and diagnostic status of black SA subjects. Furthermore, it is possible that these multiple *SMN1* copies may not be fully functional gene copies, which may further complicate the accurate diagnosis and carrier detection of SMA. More questions arising from this part of the study include: 1.) Are the multiple *SMN1* copies observed, in-tact copies?

2.) Are *SMN1* copies all transcribed into functional transcripts? This aspect was addressed by investigating the correlation between the genomic *SMN1* copy number and RNA expression levels.

4.2. RNA expression

The *SMN1* and *SMN2* genes are both transcribed with approximately 80% of all SMN protein translated from *SMN1*. This section will discuss the results obtained from RNA expression analysis of the Full length SMN1 transcript (FL-SMN1), Full length SMN2 (FL-SMN2, usually accounts for 20% of all SMN transcripts) and SMN1/SMN2 lacking exon 7 (SMNΔ7). The SMNΔ7 transcript is mostly produce from the SMN2 gene, see Figure 2.16 for explanation of transcripts. These transcripts were assessed to determine whether there is a correlation between DNA copy number and RNA expression. Further, the study aimed to determine the clinical utility of RNA expression as an alternative/extended diagnostic method to DNA analysis. Due to the long duration of optimisation of RT-qPCR and the Covid-19 pandemic, only 19 samples could be collected for RNA studies. The results obtained provided some insights not seen on DNA analysis. Subjects from all five genotypes were included in RNA analysis (M1/M1, M2/M2, N/M1, N/N and U/U).

Baseline RNA expression levels were calculated using three internal controls (negative controls with two copies of *SMN1* and *SMN2* genes). An increased expression of FL-SMN1 transcripts (80%) was expected in N/N controls who had an increased *SMN1* DNA copy number and lower expression of FL-SMN2 transcripts

(20%) were expected due to *SMN2* often being deleted in the black SA population. It was further expected that expression of the SMNΔ7 transcripts will increase with the number of deletions of exon 7 of either the *SMN1* or *SMN2* genes. In this study the FL-SMN1 transcript expression levels correlated with the *SMN1* DNA copy number. In most subjects the FL-SMN2 transcript expression levels correlated with the *SMN2* DNA copy number. It was observed that subjects with increased expression of FL-SMN1 transcripts were more likely to have lower expression of FL-SMN2 transcripts. The expression of SMNΔ7 correlated with the number of deletions of either *SMN1* or *SMN2.*

4.2.1 Comparison of *SMN1* DNA copy number versus FL-*SMN1* transcript

All 19 subjects (M1/M1=4, M2/M2=1, N/M1=2, N/N=10, U/U=2) were analysed and DNA copy number of the *SMN1* gene and RNA expression levels of the SMN transcripts were compared. The FL-SMN1 transcripts' expression levels were labelled from zero to ≥3 copies to represent the RNA expression expected to represent zero to ≥3 DNA copies. All samples were set up in quadruplicates and outliers were omitted from analysis.

4.2.2 DNA and RNA analysis of M1/M1 subjects

Comparative analysis of DNA and RNA showed that all four subjects had zero *SMN1* gene copies and no expression of the FL-SMN1 transcript, as expected. There was no discrepancy between the DNA *SMN1* copy number and RNA expression level and therefore, a diagnosis of SMA was confirmed in these subjects. The ability to confirm diagnosis by RT-qPCR was important as it could be used as an alternative to DNA analysis. This may indicate that none of the *SMN2* gene copies had been converted to *SMN1* and therefore it is likely that subjects had "true" homozygous deletions of *SMN1.* The sample size was very small; however, it was able to prove that indeed RT-qPCR could confirm SMA disease.

All four subjects had at least two copies of the *SMN2* gene on DNA methods and on RT-qPCR. The FL-SMN2 transcript expression level was much higher than expected from the DNA *SMN2* copy numbers for all four subjects (see section 3.3.1). The increased FL-SMN2 expression could be due to the way RNA transcripts levels were calculated. Another possibility could be due to the negative controls used. Thirdly only one endogenous control, *GAPDH* was used, it is recommended to use three

endogenous controls for RNA expression studies (Bas et al., 2004). We used only one GAPDH for SMN transcript analysis as previously used in other studies of SMA copy number and RNA expression (Simard et al., 2007; Sumner et al., 2003). These studies showed a correlation between DNA copy number and RNA expression, however these studies also reported complex copy numbers as seen in this study.

4.2.3 DNA and RNA analysis of M2/M2 subjects

To further show proof of principle one subject with a homozygous *SMN2*, exon 7 (zero copies of *SMN2*) deletion was analysed. Firstly, there was no expression of FL-SMN2 on RT-qPCR, as expected and therefore the DNA and RNA results correlated. More subjects with a homozygous *SMN2* deletion would have to be studied to confirm these results, however this showed that RT-qPCR was a reliable method to detect both homozygous *SMN1* and *SMN2* deletions.

Secondly, the subject had zero copies of the SMNΔ7 transcript. The SMNΔ7 transcripts are produced from *SMN1* or *SMN*2 genes and occur due to a deletion of exon 7 and should mostly be contributed from the *SMN2* gene in healthy subjects. The absence of SMNΔ7 transcript suggests that this subject does not have the *SMN2* gene. MLPA analysis indicated a homozygous deletion of *SMN2*, exon 8 as well, further supporting a possible full gene deletion of the *SMN2* gene. The lack of the SMNΔ7 transcript fits with the fact that this subject has at least three copies of *SMN1,* suggesting a gene conversion from *SMN2* to *SMN1*. This further suggests that this subject had three full length *SMN1* DNA copies, as there were no SMNΔ7 transcripts. The RNA and DNA results of this subject correlate: The subject has at least three DNA copies of *SMN1* and zero DNA copies of *SMN2*.

4.2.4 DNA and RNA analysis of N/M1 subjects

Two mothers who were confirmed carriers of SMA on DNA analysis were included as N/M1 subjects. Both subjects had one copy of *SMN1* on DNA methods. RNA expression by RT-qPCR showed them to have RNA expression suggesting one FL-SMN1 copy. RT-qPCR could accurately detect “true carriers” of SMA (one copy of SMN1, 1:0 genotype). There was no RNA available from silent carrier subjects (2:0 genotype), therefore this haplotype could not be evaluated. Similarly, to DNA methods, RT-qPCR detects only final transcript expression and family pedigrees are still needed

to confirm silent carriers. Informative pedigrees remain essential in predicting silent carriers (see section 1.8).

The first subject had one *SMN1* copy and two *SMN2* copies on DNA methods and expression of the SMNΔ7 transcript representing one *SMN1* DNA copy and expression of FL-SMN2 transcript representing 4 *SMN2* DNA copies on RT-qPCR. The discrepant *SMN2* results between DNA and RNA methods could be due to faulty calculations or potentially due to biological differences of DNA copy number versus expression of the *SMN1* and *SMN2* genes. The SMNΔ7 transcript likely comes from *SMN1* (the subject has a heterozygous deletion of *SMN1*). This is likely the case as this subject is a confirmed carrier of SMA.

The second subject had one *SMN1* copy and one *SMN2* copy on DNA methods - and expression of the SMNΔ7 transcript representing deletions of two *SMN* DNA copies and expression of the FL-SMN2 transcript and two SMNΔ7 transcripts on RT-qPCR. The RNA transcripts and DNA copy numbers correlated for this subject. The two SMNΔ7 copies are likely one from each of the *SMN1* and *SMN2* genes (subject has a heterozygous deletion of both *SMN1* and *SMN2*).

The RT-qPCR results showed that DNA copy number does not necessarily correlate with RNA expression. The interpretation of the expression of SMNΔ7 transcripts must be done in context of the *SMN1* and *SMN2* DNA copy numbers and RNA expression levels of other transcripts Although results of SMNΔ7 were not essential for diagnostic purposes it offered better understanding of the SMN genes.

4.2.5 DNA and RNA analysis of N/N subjects

RNA expression of the N/N subjects further showed the complexity of working with African samples. There were ten subjects with no symptoms of SMA and no family history of SMA. A typical cohort of N/N subjects in other populations would have shown two copies of both *SMN1* and *SMN2* in most individuals. Six of the ten N/N subjects in this study had two *SMN1* copies and only four had two copies of *SMN2* on DNA methods. RNA expression showed only five of the six subjects to have two FL-*SMN1* expression representing two *SMN1* DNA copies.

One subject (SMAR14) had increased FL-SMN1 transcripts compared to their *SMN1* DNA copies. FL-SMN1 expression represented five copies on qRT-PCR compared to

two copies on DNA methods. There could be some upregulation of transcription of the *SMN1* gene in this subject due to factors not tested for by methods in this study or DNA methods were not able to detect all their *SMN1* copies. Surprisingly, this subject had two DNA copies of *SMN2* on MLPA and AmplideX and one on qPCR. The subject had zero expression of the FL-SMN2 transcript. This suggests that the subject had no *SMN2*, exon 7 and that DNA methods may have inaccurately identified some of the *SMN1* copies as *SMN2* copies. It further suggests that some gene conversion event occurred where *SMN2* was converted to *SMN1*. Due to discrepancies seen in two methods (qPCR and RT-qPCR), testing was repeated on AmplideX, qPCR and RT-qPCR. The results remained the same, therefore laboratory error was excluded. Further, the subject had SMNΔ7 RNA expression levels suggestive of two *SMN* DNA copies. This transcript could be produced from either the *SMN1* or *SMN2* genes, but due to the complexity of the *SMN* region in the subject, the origin of SMNΔ7 transcripts could not be deduced. Long range sequencing of the SMN region would be ideal to fully understand the orientation and quantities of the SMN genes in this subject.

Four of the N/N subjects had ≥3 copies of *SMN1* on DNA methods, three of the four subjects had increased expression of FL-SMN1 transcripts, as expected and one had decreased FL-SMN1 transcript (four *SMN1* copies on DNA methods and expression suggestive of three copies on RNA methods). All subjects would still be classified as having ≥3 copies on *SMN1* on both DNA and RNA methods.

There were five N/N subjects with at least one *SMN2* copy on DNA analysis and expression levels of zero of the FL-SMN2 transcript on RT-qPCR. There are multiple reasons for no amplification of FL-SMN2 transcripts. Firstly, RNA is less stable than DNA and the FL-SMN2 transcripts could have degraded. However, all transcripts are expected to be affected if degradation is the cause, which was not the case. A second cause could be due to the calculation of expression levels in this study. Three negative controls (calibrators) were included in each run to determine copy numbers, perhaps calculations were too stringent. Thirdly, transcription of the SMN genes could be regulated differently in different tissues and perhaps there was no transcription of *SMN2* in the blood for these four subjects. The primers used will have to be assessed for any SNPs, that may influenced primer binding and amplification (known SNPs were checked, however rare SNPs may have been unaccounted for).

The SMNΔ7 expression levels correlated with both FL-SMN1 and FL-SMN2 expression levels in 9/10 subjects. Only one subject had increased SMNΔ7 transcripts. This subject (SMAR5) had two copies of *SMN1* on DNA methods as well as FL-SMN1 expression suggestive of two copies on RNA methods. The subject also had three copies of SMN2 on DNA methods and FL-SMN2 expression levels suggestive of three *SMN2* copies on the RNA method. The subject had five SMNΔ7 transcripts, which likely originated from the SMN2 gene, even though the copy numbers do not fully correlate. RNA expression of SMN genes illustrated that DNA copies do not necessarily correlate with RNA expression. RNA expression is tissue specific and is sensitive to environment conditions (e.g., drugs) (Sumner et al., 2006; Tiziano et al., 2013).

4.2.6 DNA and RNA analysis of U/U subjects

Only two subjects who had clinical symptoms of SMA with no *SMN1* deletion were analysed. Both the *SMN1* DNA copies and FL-SMN1 expression correlated in these subjects (two copies). Therefore, these subjects had no deletion of *SMN1*, exon 7 on MLPA, AmplideX, qPCR (DNA methods) and RT-qPCR (RNA method). These subjects will have to be re-evaluated clinically to confirm if they are truly likely to have SMA. These subjects could have rare point mutations in the *SMN1* gene that account for 2% of SMA cases or a complex rearrangement not detectable by current methods may have been the cause (Vorster et al). However, the presence of FL-SMN1 transcript may suggest another genetic diagnosis and they could have mutations in other SMA related genes. A molecular diagnosis in these subjects remains unconfirmed, as these subjects had two *SMN1* DNA copies and expression of the FL-SMN1 transcripts, representative of two DNA copies. Further testing to determine the cause of disease in these two subjects is recommended, by first doing a clinical assessment on the subjects and selecting the most appropriate molecular test.

4.2.7 Comparison of *SMN1*/*SMN2* DNA and RNA results

There is a limited understanding of the correlation and interaction of DNA copies versus RNA expression levels. Figure 2.16 (section 3.2) summarises the expected results. We expected to find expression representative of 80% FL-SMN1 and 20% FL-SMN2 in N/N subjects with two DNA copies of both *SMN1* and *SMN2*. Our results

showed that FL-SMN1 expression correlated with *SMN1* DNA copy number. FL-SMN2 expression did not correlate fully with *SMN2* DNA copy and may be due to the underlying complexity of the *SMN* region. The FL-SMN2 transcript was not expressed in 4/10 subjects who had at least one copy of *SMN2* on DNA analysis. However, high levels of truncated SMNΔ7 were detected in these subjects as expected if *SMN2* genes were present.

All M1/M1 subjects had higher levels of FL-SMN2 transcript expression than what was expected from the number of *SMN2* gene copies on DNA methods. Only one subject had an increase in SMNΔ7. This suggests that FL-SMN2 was upregulated in these subjects to compensate for the missing FL-SMN1 transcript. Figure 4.3 was modified to reflect the results observed in this study. The transcript levels were different in each subject and depended on the specific subject's *SMN1* and *SMN2* DNA copy numbers. Analysis and interpretation proved challenging. It was not possible to fully correlate DNA copies of *SMN1* and *SMN2* to expression of the transcripts

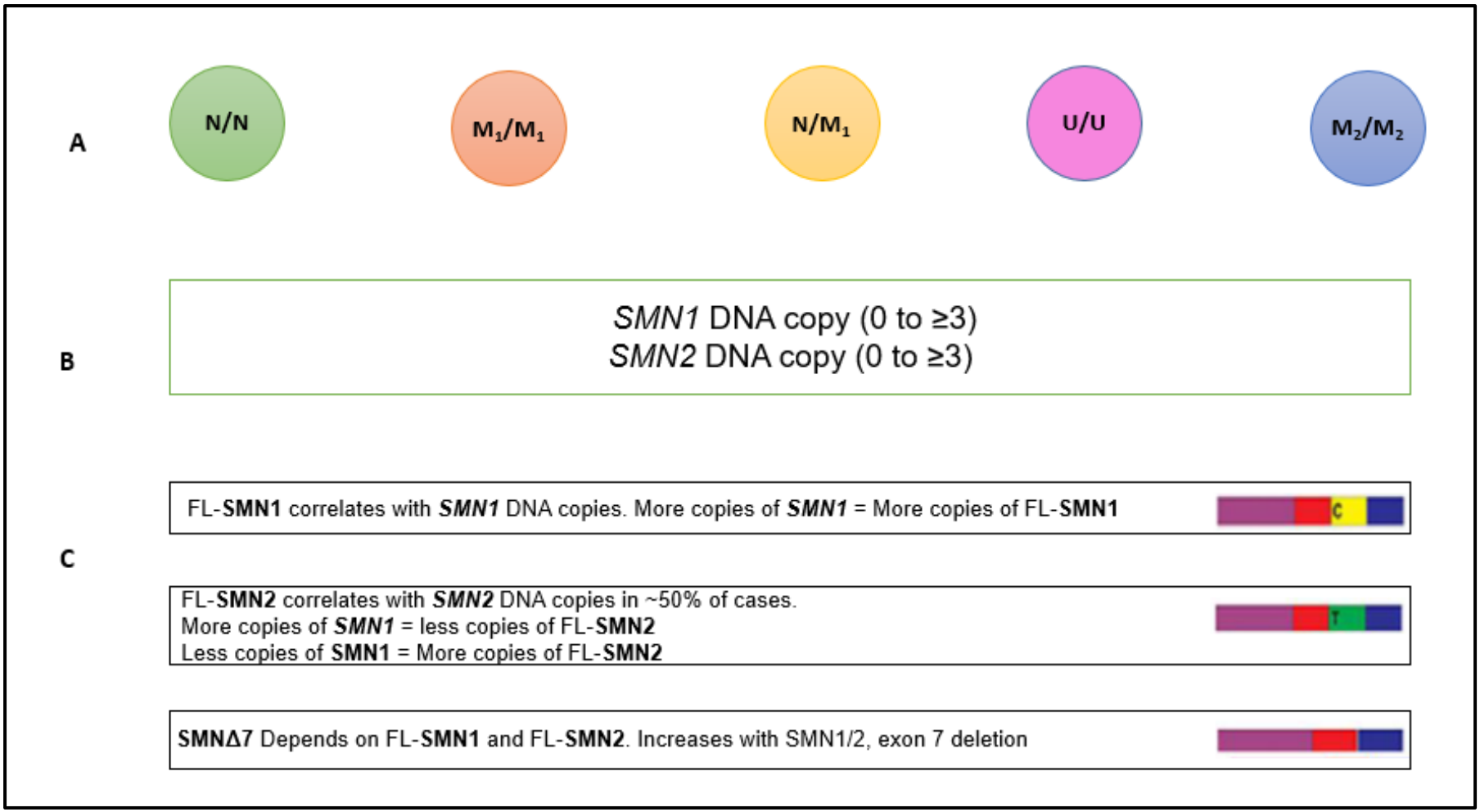


Figure 4.3. Comparative analysis of genomic *SMN1* and *SMN2* copy numbers versus RNA transcription**. A** indicates the genotypes, N indicates a normal *SMN1* copy, M1 is a deletion of *SMN1*, exon 7 and U indicates an unidentified mutation. **B** indicates the number of *SMN1* and *SMN2* DNA copy numbers. **C** indicates the *SMN1* and *SMN2* RNA expression levels that were observed according to *SMN1/2* DNA copy numbers. Three transcripts were analysed FL-SMN1, FL-SMN2 and SMNΔ7. The transcript levels depended on each subject's DNA copy numbers

4.3 Clinical utility of SMN1/SMN1 DNA and RNA testing

Overall, four methods were used to determine the molecular basis of SMA disease. Each method had its own advantages and limitations. The MLPA method is the gold standard for diagnosis and carrier testing of SMA. It is also the most expensive method at R744.05 and the most labour intensive taking approximately 120 minutes hands-on time. MLPA was able to detect *SMN1* and *SMN2* copy numbers accurately and offered more insight into other exons of the *SMN1* and *SMN2* genes as well as other adjacent genes. However, MLPA was not able to specifically detect gene conversions, even though results may suggest gene conversions.

The AmplideX kit was a new kit introduced and evaluated. The kit costs R627.62 per sample and had 60 minutes of hands-on time. Analysis of AmplideX results was easy and straightforward as it only focused on exon 7 of the *SMN1* and *SMN2* genes. Furthermore, the kit was able to detect gene conversions. In this study, the identification of gene conversion did not add clinical value. Therefore, *SMN1/2* AmplideX was a valuable kit but did not offer more than MLPA could at a similar price. It however halved the hands-on time, which could improve turnaround time.

The qPCR method to detect *SMN1* and *SMN2* DNA copy number was the most affordable method at R165.06 per sample with 60 minutes hands-on time. Analysis was straightforward and quick to perform. Although not the gold standard for SMA testing, the qPCR method has been utilised the most for SMA testing in the literature (Feldkotter et al., 2002; Maretina et al., 2018; Scheffer et al., 2001) This further supports the clinical utility of qPCR. In this study we encountered more failures with qPCR than MLPA or AmplideX. If this assay was to be implemented diagnostically, the DNA quality issue would need to be investigated further.

RNA expression analysis using RT-qPCR did not offer a diagnosis to the two subjects tested with clinical features of SMA, but no deletions detected on DNA analysis. These subjects are probably not affected with SMA, and a further clinical assessment is recommended. If SMA is still suspected, further testing of the *SMN1* gene to detect point mutations would be recommended. The RNA analysis has raised several additional questions:

1. Do subjects with ≥3 *SMN1 DNA* copies have increased expression of functional FL-SMN1 transcript?

2. What is the actual gene conversion rate in Africans?

3. Are there additional gene conversion events not detectable by current testing methods?

The cost of RT-qPCR was R294.65 per sample with 60 minutes hands-on time. RNA analysis by RT-qPCR would not be ideal as a diagnostic procedure, due to the challenges in sampling of RNA. Special tubes need to be used for collection of RNA and once collected, samples must be sent immediately to the laboratory. Immediate extraction of the RNA must be done, or tubes must be stored at -70°C. Further, RNA analysis needs to be performed in a larger cohort of *SMN* deletion negative subjects to determine its clinical utility in U/U subjects.

Taking the above into account further clinical assessment is recommended to confirm that the U/U subjects do indeed present with features suggestive of SMA. Once clinical diagnosis is determined, DNA testing for SMA or other related disorders should be performed. The best testing approach is next generation sequencing of *SMN* genes as well as other neuromuscular diseases, either by a targeted panel or WGS or WES.

4.4 Future studies

Current studies seem to be focused on the effects of Nusinersen therapy on SMA patients. Other studies focus on new-born screening, modifiers of SMA disease and non-deletion SMA patients (Loeber et al., 2021). This section will propose the next steps that need to be followed in further understanding SMA disease in African patients and the methods required to confirm the diagnosis in the 49% of African patients with a clinical phenotype suggestive of SMA who remain molecularly undiagnosed.

4.4.1 SMA modifiers of disease

The heterogeneity of the SMN region causes multiple sub-types (see section 1.2, for explanation) of SMA disease. Many modifiers have been demonstrated to play a role: the role of the *SMN2* gene and DNA methylation has been explained in section 1.6. DNA methylation of *SMN2* was evaluated in Chinese children (Cao et al., 2016). The children in their study had a confirmed diagnosis of SMA and compared affected

children with the same *SMN1* and *SMN2* copy numbers. The authors were able to demonstrate that *SMN2* is differentially methylated depending on the variants the children carried resulting in different SMA sub-types. They were able to show that DNA methylation of *SMN2* as well as copy number of *SMN2* can change the severity of disease (Cao et al., 2016). In this study the *SMN2* DNA copy numbers and RNA expression levels were evaluated, however due to limited clinical information the *SMN2* copy number versus SMA phenotype could not be explored. DNA methylation analysis of the *SMN* genes in affected patients, carriers and healthy individuals should be considered in the future.

4.4.2 Beyond the copy number

We have demonstrated the complexity of the *SMN* genes, therefore multiple approaches to testing may offer a diagnosis to non-deletion SMA patients. Approximately 95% of SMA patients have a homozygous deletion of *SMN1*, exon 7. The remaining 5% have a heterozygous deletion of *SMN1* on one chromosome in combination with a point mutation in *SMN1* on the other chromosome. Point mutations were not explored in this study as they account for a small fraction of SMA patients and are difficult to detect due to similarity of the *SMN1* and *SMN2* genes (Butchbach, 2016). Sequencing of the *SMN1* and *SMN2* genes is the recommended approach to confirm a clinical diagnosis of SMA. Previous Sanger sequencing studies done in the division did not find definitive variants that caused SMA in the *SMN1* gene (Labrum et al., 2007). However, it is recommended to test for mutations in the *SMN* genes as well as other genes causing neuromuscular disease. Next generation sequencing can be performed by using a panel with genes known to be associated with non-chromosome 5q spinal muscular atrophy genes shown in appendix 11. A larger neuromuscular panel would have to be designed to include as many candidate genes as possible. Alternatively whole exome sequences may be a better option, specifically in our diverse population. There is a continuous discovery of new genes associated with SMA. African populations may have mutations in other genes not yet detected (Fang et al., 2021). Whole exome or NGS testing by panel may be more appropriate methods of diagnosing non-deletion patients.

4.4.3 Long-read SMN gene sequencing

Long read sequencing is a third-generation sequencing method that overcomes the limitations of next generation sequencing. There are two major long-read technologies, Pacific BioSciences (PacBio) sequencing and the Oxford Nanopore Technologies (ONT) (Logsdon et al., 2020). PacBio can analyse longer sequences (reads) from 10kb to 60kb depending on the PacBio technology (Rhoads and Au, 2015). Conventional Next generation sequencing can analyse DNA in approximately 250 bp fragments. PacBio can therefore sequence fragments more than 100 times longer than the size of most NGS technologies. The ONT uses linear molecules for sequencing and analysis reads from 10kb to 200kb depending on the platform (Logsdon et al., 2020). Another advantage of long-read sequencing is that it does not pause between reads like NGS, therefore making it useful in determining the gene arrangement of the *SMN* genes as well as other adjacent genes (Rhoads and Au, 2015). It can also be used to determine the copy numbers of the *SMN* genes, while determining the sequences of these genes. PacBio is able to simultaneously determine DNA sequences and epigenetic changes (Nakano et al., 2017). One *SMN* gene is approximately 34kb long, the PacBio or ONT will thus be able to sequence one complete gene copy but may struggle with multiple copies and large rearrangements. These technologies may be helpful in further determining the arrangement of *SMN* genes as well as the methylation status of the *SMN* region in the black SA population.

4.5 Advantages and limitations of the study

4.5.1 Advantages and limitations of MLPA

The MLPA method is more expensive than RFLP, it is also relatively more difficult to analyse, and more labour intensive compared to RFLP. Although the method is more informative, most of the patients who come for diagnostic testing of SMA need only the presence or absence of *SMN1* determined. The RFLP is sufficient for diagnostic testing albeit only for patients of European ancestry, in which 95% are confirmed using this method. However, diagnosis in 49% patients of African ancestry is not achieved using either method (Vorster et al ., 2020). Firstly, an affordable method with the ability to accurately detect DNA copy number of *SMN1* and *SMN2* was desired to combine the function of both RFLP method and the MLPA. Secondly, another method was required to try to detect the molecular cause of SMA in the remaining 49% of patients

of African ancestry. These issues prompted this researcher to find alternative and affordable methods for diagnostic, carrier, and prenatal testing of SMA.

4.5.2 Advantages and limitations of AmplideX

The AmplideX kit is a quantitative method yielding similar results to MLPA and qPCR and is more informative than the RFLP. However, the kit is similar to qPCR in that the extent of the SMN deletion cannot be determined. The AmplideX procedure is far less laborious than the RFLP or MLPA. Also, low DNA concentration and volume were needed for the method and interpretation of results was fairly quick and easy. The AmplideX was the only method able to detect gene conversion, however gene conversions would not be reported. The kit was also as costly as MLPA and was therefore not a suitable candidate to replace both RFLP and MLPA.

4.5.3 Advantages and limitations of qPCR

The method is a quantitative method, like MLPA, which detects DNA copy numbers of *SMN1* and *SMN2*. The method has been cited multiple times in the literature as an alternative to MLPA (Zarkov et al., 2015). The method requires a single reaction to be performed, unlike MLPA which is a two-day procedure. qPCR was used to analyse only the two *SMN* genes and specifically exon 7. None of the other exons or adjacent genes in the *SMN* region can be analysed, which is a limiting factor of the method. DNA concentrations and dilution of qPCR must be precise, therefore pipetting errors have to be minimal. This method was the most affordable, less labour intensive than MLPA and relatively easy to analyse. It also required less equipment than RFLP, MLPA and AmplideX.

4.5.4 Advantages and limitations of RT-qPCR

Determining the absence of the FL-SMN1 transcript may support the clinical diagnosis of subjects who have clinical features suggestive of SMA but tested negative on DNA analysis. RT-qPCR has an advantage as transcript levels of *SMN* may be determined without knowing the molecular mechanisms of disease, while still offering a diagnosis to these affected subjects and their families. One of the main challenges with this method is the collection of blood in an EDTA (for DNA extraction) and a Tempus tube (for RNA extraction) from diagnostic centres. The doctor would need the specialised Tempus tube and the sample should reach the lab as soon as possible to prevent degradation of the RNA. Furthermore, RT-qPCR depends on the accurate quantitation

and dilution of cDNA; also, RNA was extracted from whole blood which may not completely reflect expression of *SMN* in muscle tissue where it is most essential. RNA expression may not accurately determine carriers but can offer a diagnosis for affected subjects. Our samples sizes for RNA expression analysis were very small (n=19) and we only had two U/U subjects. Therefore, diagnosis in non-deletion subjects was not confirmed. A larger samples size, with more U/U subjects who have strong suspicion of SMA would need to be tested for RNA expression. However, RT-qPCR is a promising method for alternative/additional testing of SMA and it has been validated in a diagnostic testing.

5 CONCLUSION

Diagnostic testing of black SA patients in the Division of Human Genetics detects *SMN1* deletions in 51% of patients with a suggestive phenotype. The remaining 49% of black SA patients do not have a diagnosis.

The Division currently uses RFLP analysis to detect a homozygous deletion of *SMN1* and confirm the diagnosis of SMA. This method can only detect the absence or presence of the *SMN1* and *SMN2* genes. It does not detect copy numbers and therefore cannot detect heterozygous deletions in patients or carriers of SMA. *SMN1* and *SMN2* gene copy numbers are detected by MLPA in the Division. The MLPA method is the gold standard for SMA analysis since it can detect homozygous deletions, copy numbers and is able to determine carrier status. MLPA is not used as a first line test in the Division as it far more expensive than RFLP. Two other methods for DNA copy number detection were validated and compared to MLPA to determine their clinical utility and to further investigate the SMA disease mechanism.

The *AmplideX® PCR/CE SMN1/2* kit is a new quantitative kit that had never been used in the Division before. The kit can determine homozygous *SMN1 and SMN2* deletions, *SMN1 and SMN2* copy numbers as well as SMA carrier status. Additionally, the kit was able to detect some gene conversions of *SMN1* to *SMN2* and vice versa. In this study *AmplideX® PCR/CE SMN1/2* results correlated with MLPA results, with only one subject who had a discrepancy in *SMN1* copy number. Three discrepancies were observed in *SMN2* copy numbers. In this study 6% of subjects had gene conversions and these were observed in N/N and U/U subjects, supporting previous findings that gene conversion is common in black SA subjects and form part of normal variation and does not necessarily add much diagnostic value. The kit had less hands-on time but was not as expensive as MLPA. Further, the kit was only able to detect exon 7 of *SMN1* and *SMN2* genes and none of the adjacent exons or genes. The clinical utility of the AmplideX kit is the same as MLPA, it can confirm the diagnoses of SMA and determine carriers.

qRT-PCR was designed in-house and validated similarly to the *AmplideX® PCR/CE SMN1/2* kit. DNA copy numbers were compared between MLPA and AmplideX. The method was able to detect homozygous *SMN1* and *SMN2* deletions, *SMN1* and *SMN2*

copy numbers and carriers of SMA. The method was the most affordable and had a hands-on time similar to the *AmplideX® PCR/CE SMN1/2* kit. qPCR is the recommended method of testing of SMA in SA diagnostic setting. It could replace both RFLP as a diagnostic method and MLPA as a method for CNV detection in a single reaction. Currently patients with a heterozygous deletion of *SMN1* are not detected by the current RFLP diagnostic assay in the Division. The diagnostic yield may be increased by using qPCR. Currently quantitative analysis of the *SMN* genes is only done when *SMN1* and *SMN2* copy numbers are requested in a known SMA patient for possible therapy or to determine carrier status in families affected by SMA. Testing for SMA families (diagnosis, *SMN2* copy number for therapy and carrier status) could be performed using one affordable and quick method. This will ensure that the Division follows the same testing approaches as recommended overseas.

Additional testing using the *AmplideX® PCR/CE SMN1/2* kit and qPCR did not identify additional SMA patients from the U/U subject group.

The second part of this study consisted of optimisation and validation of RT-qPCR for detection of FL-SMN1, FL-SMN2 and SMNΔ7 transcripts. Nineteen subjects were tested on RT-qPCR and results were compared with DNA methods to investigate the correlation of DNA copy number with RNA expression.

RNA expression results showed that subjects with homozygous deletions of *SMN1* also had zero expression of FL-SMN1 transcripts and subjects with homozygous deletions of *SMN2* had zero expression of FL-SMN2 transcripts, as expected. RNA expression analysis detects zero copies of both the FL-SMN1 and FL-SMN2 transcripts. Two subjects with the N/M1 genotype previously shown to have one copy of *SMN1* had FL-SMN1 expression levels suggestive of one copy, therefore, carrier status could be determined by RT-qPCR. Most DNA copy numbers of *SMN1* correlated with FL-SMN1 expression levels. However, this was not the case for *SMN2* DNA copy and FL-SMN2 expression levels, which were discrepant in several cases.

It seems that the *SMN2* copy numbers increased with decreasing *SMN1* copies and vice versa, likely due to gene conversion. Similarly, FL-SMN1 expression increased with decreased FL-SMN2 expression and vice versa. The SMNΔ7 correlated with FL-SMN1 and FL-SMN2 transcripts. Deletions of the *SMN2* gene increased transcripts of

SMNΔ7. Analysis of the SMNΔ7 transcript showed interesting variation but was not useful in SMA diagnosis.

This study provided some insights into the correlation of DNA copy number versus RNA expression of the *SMN* genes, but a study with a larger sample size is recommended, more specifically on U/U subjects to determine if additional SMA patients could be identified using this method. Thus far, this does not seem to be likely. Two blood tubes could be taken at the same time, one for DNA analysis and one for RNA analysis and results could be compared. Gene expression studies may be useful in cases where multiple *SMN1* and *SMN2* copies are detected in U/U subjects. Zero expression of FL-SMN1 could confirm SMA and indicate incorrectly assigned *SMN1* copies by DNA methods. The *SMN* region is complex, and it is challenging to distinguish between *SMN1* and *SMN2*, exon 7 using the single base pair difference between these two genes – it is possible that current DNA methods may incorrectly assign copies to be *SMN1* or *SMN2,* may also miss or mis-diagnose some gene conversion or rearrangement events.

A transcriptome analysis of patients with clinical features suggestive of SMA could be beneficial to determine the RNA profile of affected patients and perhaps identify candidate genes or pathways.

This study illustrated the complexity of the *SMN* region and various factors that could contribute to the pathogenesis of SMA. One single approach to SMA testing might not be sufficient in reaching a molecular diagnosis in the black SA population in patients with suggestive features. Accurate clinical information is critical to guide the appropriate approach followed by quantitative detection of both the *SMN1* and *SMN2* genes. If a diagnosis of SMA is not confirmed, further clinical assessment is recommended followed by DNA sequencing. Next generation sequencing is the most appropriate method to look for point mutations in the *SMN1* gene as well as other genes that may be associated with SMA. Perhaps WGS or WES is the best approach for sequencing in African patients as there remain a large group of undiagnosed patients in whom the molecular diagnosis is uncertain. Once candidate genes are found, targeted NGS sequencing panels may be designed specifically for SA population groups.

A homozygous deletion of *SMN1* has been observed in healthy individuals further adding to the complexity of SMA (Jedrzejowska et al., 2008). These individuals were shown to have more than three copies of the *SMN2* gene. This shows the importance of *SMN2* as a modifier of SMA disease. DNA methylation studies on *SMN* genes have also shown that methylation of *SMN2* changes the severity of SMA. Not all methods for understanding the molecular basis of SMA were explored in this study, further highlighting the importance of continued research.

Long range DNA sequencing to determine how gene rearrangements occur and how they play a role in SMA disease should be explored. Long-range sequencing may show if the multiple *SMN1* copies observed in N/N and U/U subjects are full gene copies or not. It may also reveal complex rearrangements common to African populations that may further explain the evolution of the *SMN* region.

The aim of this study was to determine the clinical utility of DNA and RNA methods in testing clinically suggestive SMA patients without the *SMN1* deletion. This study illustrated the advantages and limitations of each method and has shown qPCR to be the most affordable diagnostic method for detecting *SMN1* and *SMN2* copy numbers in a SA diagnostic setting. This study was unable to offer any additional diagnosis to other suggestive SMA patients. Further clinical assessment of these patients is recommended using NGS analysis or either targeted, WES or WGS with the *SMN* genes included. RT-qPCR shows promise as a second line diagnostic test for U/U subjects, but this method needs to be further assessed since no U/U subjects in this study were found to have zero expression of FL-SMN1 transcripts.

Further research of the molecular mechanism of SMA in the black SA population is needed. New-born screening for SMA is starting to be offered widely in developed countries (Loeber et al., 2021). However, this depends on comprehensively understanding the true incidence of SMA disease as well as its molecular mechanism. Currently only 51% of black SA patients who come for SMA testing are found to have a homozygous *SMN1*, exon 7 deletion. The remaining 49% could be due to point mutations, modifier genes, other SMA related genes, an incorrect diagnosis, or complex rearrangements in the *SMN* region, not yet discovered by current methods. New-born screening in SA individuals can only be pursued once the molecular mechanism of SMA disease is better understood in this population. This study was

able to answer some of the objectives as well as yield more questions to resolve. The study of SMA remains challenging and hopefully future research can explain and unravel the complexity of SMA disease, so that patients with a clinical SMA diagnosis can receive a molecular diagnosis. This is required to ultimately facilitate therapy and to provide opportunities for risk assessment and family cascade screening.

REFERENCES

Anhuf, D., Eggermann, T., Rudnik-Schöneborn, S., Zerres, K., 2003. Determination of SMN1 and SMN2 copy number using TaqMan™ technology. Hum Mut 22, 74–78. https://doi.org/10.1002/humu.10221

Asuragen, 2020. Whitepaper: Interpreting Variants with the AmplideX® PCR/CE SMN1/2 Plus and SMA Plus Kits. VH Bio. URL https://www.vhbio.com/interpreting-variants-pcr-ce-smn1-2-sma/ (accessed 4.26.21).

Bas, A., Forsberg, G., Hammarström, S., Hammarström, M.-L., 2004. Utility of the Housekeeping Genes 18S rRNA, β-Actin and Glyceraldehyde-3-Phosphate-Dehydrogenase for Normalization in Real-Time Quantitative Reverse Transcriptase-Polymerase Chain Reaction Analysis of Gene Expression in Human T Lymphocytes. Scand J Immunol 59, 566–573.

Belter, L., Cook, S.F., Crawford, T.O., Jarecki, J., Jones, C.C., Kissel, J.T., Schroth, M., Hobby, K., 2018. An overview of the Cure SMA membership database: Highlights of key demographic and clinical characteristics of SMA members. J Neuromuscul Dis 5, 167–176. https://doi.org/10.3233/JND-170292

Ben-Shachar, S., Orr-Urtreger, A., Bardugo, E., Shomrat, R., Yaron, Y., 2011. Large-scale population screening for spinal muscular atrophy: Clinical implications. Genet Med 13, 110–114.

Berger, S.L., Kouzarides, T., Shiekhattar, R., Shilatifard, A., 2009. An operational definition of epigenetics. Genes Dev. 23, 781–783. BQUB13-Ecarmona, 2013. Català: Procés MLPA.

Burghes, A.H., 1997. When is a deletion not a deletion? When it is converted. Am. J. Hum. Genet. 61, 9–15.

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.

Butchbach, M.E.R., 2016. Copy Number Variations in the Survival Motor Neuron Genes: Implications for Spinal Muscular Atrophy and Other Neurodegenerative Diseases. Front Mol Biosci 3, 7.

Cao, Y., Qu, Y., He, S., Li, Y., Bai, J.-L., Jin, Y., Wang, H., Song, F., 2016. Association between SMN2 methylation and disease severity in Chinese children with spinal muscular atrophy. J Zhejiang Univ Sci B 17, 76–82.

Chen, J.-M., Cooper, D.N., Chuzhanova, N., Férec, C., Patrinos, G.P., 2007. Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet 8, 762–775.

Cooper, T.A., Wan, L., Dreyfuss, G., 2009. RNA and disease. Cell 136, 777–793.

Corsello, A., Scatigno, L., Pascuzzi, M.C., Calcaterra, V., Dilillo, D., Vizzuso, S., Pelizzo, G., Zoia, E., Mandelli, A., Govoni, A., Bosetti, A., Francavilla, R., Indrio, F., Fabiano, V., Zuccotti, G.V., Verduci, E., 2021. Nutritional, Gastrointestinal and Endo-Metabolic Challenges in the Management of Children with Spinal Muscular Atrophy Type 1. Nutrients 13, 2400.

Deutsch, L., Osredkar, D., Plavec, J., Stres, B., 2021. Spinal Muscular Atrophy after Nusinersen Therapy: Improved Physiology in Pediatric Patients with No Significant Change in Urine, Serum, and Liquor 1H-NMR Metabolomes in Comparison to an Age-Matched, Healthy Cohort. Metabolites 11.

Duale, N., Brunborg, G., RÅ¸nningen, 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 Research Notes 5, 510–510.

Fang, Y.-L., Li, N., Zhi, X.-F., Zheng, J., Liu, Y., Pu, L.-J., Gu, C.-Y., Shu, J.-B., Cai, C.-Q., 2021. Discovery of specific mutations in spinal muscular atrophy patients by next-generation sequencing. Neurol Sci 42, 1827–1833.

Farrar, M.A., Kiernan, M.C., 2015. The Genetics of Spinal Muscular Atrophy: Progress and Challenges. Neurotherapeutics 12, 290–302.

Feldkotter, 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.

Finkel, R.S., Mercuri, E., Darras, B.T., Connolly, A.M., Kuntz, N.L., Kirschner, J., Chiriboga, C.A., Saito, K., Servais, L., Tizzano, E., Topaloglu, H., Tulinius, M., Montes, J., Glanzman, A.M., Bishop, K., Zhong, Z.J., Gheuens, S., Bennett, C.F., Schneider, E., Farwell, W., De Vivo, D.C., ENDEAR Study Group, 2017. Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. N. Engl. J. Med. 377, 1723–1732.

Genin, E., Feingold, J., Clerget-Darpoux, F., 2008. Identifying modifier genes of monogenic disease: strategies and difficulties. Hum Genet 124, 357–368.

Goldman, A., Labrum, R., Claustres, M., Desgeorges, M., Guittard, C., Wallace, A., Ramsay, M., 2001. The molecular basis of cystic fibrosis in South Africa. Clin Genet 59, 37–41.

Gómez-Curet, I., Robinson, K.G., Funanage, V.L., Crawford, T.O., Scavina, M., Wang, W., 2007. Robust quantification of the SMN gene copy number by real-time TaqMan PCR. Neurogenetics 8, 271–278.

Hauke, J., Riessland, M., Lunke, S., Eyüpoglu, I.Y., Blümcke, I., El-Osta, A., Wirth, B., Hahnen, E., 2009. Survival motor neuron gene 2 silencing by DNA methylation correlates with spinal muscular atrophy disease severity and can be bypassed by histone deacetylase inhibition. Hum Mol Genet 18, 304–317.

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.

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.

Hua, Y., Sahashi, K., Hung, G., Rigo, F., Passini, M.A., Bennett, C.F., Krainer, A.R., 2010. Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. Genes Dev. 24, 1634–1644.

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. 16, 930–934.

Kiepiela, P., Dawood, A.A., Moosa, A., Coovadia, H.M., Coward, P., 1988. Evaluation of immunoregulatory cells in Duchenne muscular dystrophy and spinal muscular atrophy among African and Indian patients. J. Neurol. Sci. 84, 247–255.

Kolb, S.J., Kissel, J.T., 2015. Spinal Muscular Atrophy. Neurologic clinics 33, 831–846.

Kromberg, J.G., Jenkins, T., 1982. Prevalence of albinism in the South African negro. S. Afr. Med. J. 61, 383–386.

Kwon, D.Y., Motley, W.W., Fischbeck, K.H., Burnett, B.G., 2011. Increasing expression and decreasing degradation of SMN ameliorate the spinal muscular atrophy phenotype in mice. Hum. Mol. Genet. 20, 3667–3677.

Labrum, R., Rodda, J., Krause, A., 2007. The molecular basis of Spinal Muscular Atrophy (SMA) in South African black patients. Neuromuscul Disord 17, 684–692.

Lefebvre, S., BĂĽrglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Paslier, D.L., Frezal, J., Cohen, D., Weissenbach, J., Munnich, A., Melki, J., 1995. Identification and characterization of a spinal muscular atrophy-determining gene. Cell 80.

Lefebvre, S., Burlet, P., Liu, Q., Bertrandy, S., Clermont, O., Munnich, A., Dreyfuss, G., Melki, J., 1997. Correlation between severity and SMN protein level in spinal muscular atrophy. Nat Genet 16, 265–269.

Locatelli, D., Terao, M., Fratelli, M., Zanetti, A., Kurosaki, M., Lupi, M., Barzago, M.M., Uggetti, A., Capra, S., D'Errico, P., Battaglia, G.S., Garattini, E., 2012. Human axonal survival of motor neuron (a-SMN) protein stimulates axon growth, cell motility, C-C motif ligand 2 (CCL2), and insulin-like growth factor-1 (IGF1) production. J. Biol. Chem. 287, 25782–25794.

Loeber, J.G., Platis, D., Zetterström, R.H., Almashanu, S., et al., 2021. Neonatal Screening in Europe Revisited: An ISNS Perspective on the Current State and Developments Since 2010. Int J Neonatal Screen 7.

Logsdon, G.A., Vollger, M.R., Eichler, E.E., 2020. Long-read human genome sequencing and its applications. Nat Rev Genet 21, 597–614.

Lunn, M.R., Wang, C.H., 2008. Spinal muscular atrophy. Lancet 371, 2120–2133.

Maretina, M., Egorova, A., Baranov, V., Kiselev, A., 2019. DYNC1H1 gene methylation correlates with severity of spinal muscular atrophy. Ann. Hum. Genet. 83, 73–81.

Maretina, M.A., Zheleznyakova, G.Y., Lanko, K.M., Egorova, A.A., Baranov, V.S., Kiselev, A.V., 2018. Molecular Factors Involved in Spinal Muscular Atrophy Pathways as Possible Disease-modifying Candidates. Curr. Genomics 19, 339–355.

Melki, J., Lefebvre, S., Burglen, L., Burlet, P., Clermont, O., Millasseau, P., Reboullet, S., Benichou, B., Zeviani, M., Paslier, D.L., Et, A., 1994. De novo and inherited deletions of the 5q13 region in spinal muscular atrophies. Science 264, 1474–1477.

Mercuri, E., Finkel, R.S., Muntoni, F., Wirth, B., Montes, J., et al, SMA Care Group, 2018. Diagnosis and management of spinal muscular atrophy: Part 1: Recommendations for diagnosis, rehabilitation, orthopedic and nutritional care. Neuromuscul Disord 28, 103–115.

Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research 16, 1215.

MLPA General Protocol MDP-v007.pdf, n.d.

Monani, U.R., Lorson, C.L., Parsons, D.W., Prior, T.W., Androphy, E.J., Burghes, A.H., McPherson, J.D., 1999a. A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. Hum Mol Genet 8, 1177–1183.

Monani, U.R., McPherson, J.D., Burghes, A.H., 1999b. Promoter analysis of the human centromeric and telomeric survival motor neuron genes (SMNC and SMNT). Biochim. Biophys. Acta 1445, 330–336.

Moosa, A., Dawood, A., 2008. Spinal Muscular Atrophy in African Children. Neuropediatrics 21, 27–31.

Moulard, B., Salachas, F., Chassande, B., Briolotti, V., Meininger, V., Malafosse, A., Camu, W., 1998. Association between centromeric deletions of the SMN gene and sporadic adult-onset lower motor neuron disease. Ann Neurol 43, 640–644.

Nakano, K., Shiroma, A., Shimoji, M., Tamotsu, H., Ashimine, N., Ohki, S., Shinzato, M., Minami, M., Nakanishi, T., Teruya, K., Satou, K., Hirano, T., 2017. Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. Hum Cell 30, 149–161.

Ogino, S., Gao, S., Leonard, D.G.B., Paessler, M., Wilson, R.B., 2003. Inverse correlation between SMN1 and SMN2 copy numbers: evidence for gene conversion from SMN2 to SMN1. Eur J Hum Genet 11, 275–277.

Ogino, S., Wilson, R.B., Gold, B., 2004. New insights on the evolution of the SMN1 and SMN2 region: simulation and meta-analysis for allele and haplotype frequency calculations. Eur J Hum Genet 12.

Passon, N., Pozzo, F., Molinis, C., Bregant, E., Gellera, C., Damante, G., Lonigro, R.I., 2009. A simple multiplex real-time PCR methodology for the SMN1 gene copy number quantification. Genet Test Mol Biomarkers 13, 37–42.

Pellizzoni, L., 2007. Chaperoning ribonucleoprotein biogenesis in health and disease. EMBO Rep 8, 340–345.

Piazzon, N., Schlotter, F., Lefebvre, S., Dodré, M., Méreau, A., Soret, J., Besse, A., Barkats, M., Bordonné, R., Branlant, C., Massenet, S., 2013. Implication of the SMN complex in the biogenesis and steady state level of the Signal Recognition Particle. Nucleic Acids Res 41, 1255–1272.

Pillai, R.S., Grimmler, M., Meister, G., Will, C.L., Lührmann, R., Fischer, U., Schümperli, D., 2003. Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. Genes Dev. 17, 2321–2333.

Prior, T.W., Finanger, E., 1993. Spinal Muscular Atrophy, in: Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J., Stephens, K., Amemiya, A. (Eds.), GeneReviews®. University of Washington, Seattle, Seattle (WA).

Qiagen, 2014. FlexiGene DNA Handbook - (EN) - QIAGEN [WWW Document]. URL https://www.qiagen.com/at/resources/resourcedetail?id=6b147421-7846-411e-9993-bb01563b807e&lang=en (accessed 4.28.20).

Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics, SI: Metagenomics of Marine Environments 13, 278–289.

Rochette, C., Gilbert, N., Simard, L., 2001. SMN gene duplication and the emergence of the SMN2 gene occurred in distinct hominids: SMN2 is unique to Homo sapiens. Hum Genet 108, 255–266.

Sangare, M., Hendrickson, B., Sango, H.A., Chen, K., et al, 2014. Genetics of low spinal muscular atrophy carrier frequency in sub-Saharan Africa. Annals of Neurology 75, 525–532.

Schefe, J.H., Lehmann, K.E., Buschmann, I.R., Unger, T., Funke-Kaiser, H., 2006. Quantitative real-time RT-PCR data analysis: current concepts and the novel “gene expression’s CTdifference” formula. J Mol Med 84, 901–910.

Scheffer, H., Cobben, J.M., Matthijs, G., Wirth, B., 2001. Best practice guidelines for molecular analysis in spinal muscular atrophy. Eur J Hum Genet 9, 484–491.

Seo, J., Singh, N.N., Ottesen, E.W., Lee, B.M., Singh, R.N., 2016. A novel human-specific splice isoform alters the critical C-terminus of Survival Motor Neuron protein. Sci Rep 6, 30778.

Setola, V., Terao, M., Locatelli, D., Bassanini, S., Garattini, E., Battaglia, G., 2007. Axonal-SMN (a-SMN), a protein isoform of the survival motor neuron gene, is specifically involved in axonogenesis. Proc. Natl. Acad. Sci. U.S.A. 104, 1959–1964.

Sharifi, Z., Taheri, M., Fallah, M.-S., Abiri, M., Golnabi, F., Bagherian, H., Zeinali, R., Farahzadi, H., Alborji, M., Tehrani, P.G., Amini, M., Asnavandi, S., Hashemi, M., Forouzesh, F., Zeinali, S., 2021. Comprehensive Mutation Analysis and Report of 12 Novel Mutations in a Cohort of Patients with Spinal Muscular Atrophy in Iran. J Mol Neurosci.

Simard, L.R., Bélanger, M.-C., Morissette, S., Wride, M., Prior, T.W., Swoboda, K.J., 2007. Preclinical validation of a multiplex real-time assay to quantify SMN mRNA in patients with SMA. Neurology 68, 451–456.

Singh, N.N., Seo, J., Rahn, S.J., Singh, R.N., 2012. A multi-exon-skipping detection assay reveals surprising diversity of splice isoforms of spinal muscular atrophy genes. PLoS ONE 7, e49595.

Singh, R.N., Singh, N.N., 2018. Mechanism of Splicing Regulation of Spinal Muscular Atrophy Genes. Adv Neurobiol 20, 31–61.

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.

Sumner, C.J., Huynh, T.N., Markowitz, J.A., Perhac, J.S., Hill, B., Coovert, D.D., Schussler, K., Chen, X., Jarecki, J., Burghes, A.H.M., Taylor, J.P., Fischbeck, K.H., 2003. Valproic acid increases SMN levels in spinal muscular atrophy patient cells. Ann Neurol 54, 647–654.

Sumner, C.J., Kolb, S.J., Harmison, G.G., Jeffries, N.O., Schadt, K., Finkel, R.S., Dreyfuss, G., Fischbeck, K.H., 2006. SMN mRNA and protein levels in peripheral blood: biomarkers for SMA clinical trials. Neurology 66, 1067–1073.

Sun, Y., Kong, X., Zhao, Z., Zhao, X., 2020. Mutation analysis of 419 family and prenatal diagnosis of 339 cases of spinal muscular atrophy in China. BMC Med Genet 21, 133.

Tiziano, F.D., Melki, J., Simard, L.R., 2013. Solving the puzzle of spinal muscular atrophy: what are the missing pieces? Am. J. Med. Genet. A 161A, 2836–2845.

Verhaart, I.E.C., Robertson, A., Wilson, I.J., Aartsma-Rus, A., Cameron, S., Jones, C.C., Cook, S.F., LochmÄĽller, H., 2017. Prevalence, incidence and carrier frequency of linked spinal muscular atrophy“ a literature review. Orphanet Journal of Rare Diseases 12, 124.

Vijzelaar, R., Snetselaar, R., Clausen, M., Mason, A.G., Rinsma, M., Zegers, M., Molleman, N., Boschloo, R., Yilmaz, R., Kuilboer, R., Lens, S., Sulchan, S., Schouten, J., 2019. The frequency of SMN gene variants lacking exon 7 and 8 is highly population dependent. PLoS One 14.

Vorster, E., Essop, F.B., Rodda, J.L., Krause, A., 2020. Spinal Muscular Atrophy in the Black South African Population: A Matter of Rearrangement? Front Genet 11, 54.

Wang, C.C., Chang, J.G., Chen, Y.L., Jong, Y.J., Wu, S.M., 2010. Multi-exon genotyping of SMN gene in spinal muscular atrophy by universal fluorescent PCR and capillary electrophoresis. Electrophoresis 31.

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.

Wirth, B., Brichta, L., Schrank, B., Lochmüller, H., Blick, S., Baasner, A., Heller, R., 2006. Mildly affected patients with spinal muscular atrophy are partially protected by an increased SMN2 copy number. Hum Genet 119, 422–428.

Yousefi, P., Huen, K., Schall, R.A., Decker, A., Elboudwarej, E., Quach, H., Barcellos, L., Holland, N., 2013. Considerations for normalization of DNA methylation data by Illumina 450K BeadChip assay in population studies. Epigenetics 8, 1141–1152.

Zapletalová, E., Hedvicáková, P., Kozák, L., Vondrácek, P., Gaillyová, R., Maríková, T., Kalina, Z., Jüttnerová, V., Fajkus, J., Fajkusová, L., 2007. Analysis of point mutations in the SMN1 gene in SMA patients bearing a single SMN1 copy. Neuromuscul Disord 17, 476–481.

Zarkov, M., Stojadinovic, A., Sekulic, S., Barjaktarovic, I., Peric, S., Kekovic, G., Draskovic, B., Stevic, Z., 2015. Association between the SMN2 gene copy number and clinical characteristics of patients with spinal muscular atrophy with homozygous deletion of exon 7 of the SMN1 gene. Vojnosanit Pregl 72, 859–863.

Zerres, K., Rudnik-Schöneborn, S., Forrest, E., Lusakowska, A., Borkowska, J., Hausmanowa-Petrusewicz, I., 1997. A collaborative study on the natural history of childhood and juvenile onset proximal spinal muscular atrophy (type II and III SMA): 569 patients. J. Neurol. Sci. 146, 67–72.

Zhang, H.L., Pan, F., Hong, D., Shenoy, S.M., Singer, R.H., Bassell, G.J., 2003. Active transport of the survival motor neuron protein and the role of exon-7 in cytoplasmic localization. J Neurosci 23, 6627–6637.

Zuluaga-Sanchez, S., Teynor, M., Knight, C., Thompson, R., Lundqvist, T., Ekelund, M., Forsmark, A., Vickers, A.D., Lloyd, A., 2019. Cost Effectiveness of Nusinersen in the Treatment of Patients with Infantile-Onset and Later-Onset Spinal Muscular Atrophy in Sweden. PharmacoEconomics.

APPENDICES

Appendix 1

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

R14/49 Miss Kgomotso Odirile Peggy Tabane et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M180498

NAME: (Principal Investigator)	Miss Kgomotso Odirile Peggy Tabane et al
DEPARTMENT:	Human Genetics - School of Pathology Chris Hani Baragwanath Academic Hospital National Health Laboratory Service
PROJECT TITLE:	Determining the RNA expression levels of SMN protein in the black South African non-SMN1 deletion patients clinically affected with spinal muscular atrophy
DATE CONSIDERED:	Ad hoc
DECISION:	Approved unconditionally
CONDITIONS:	A Sub-study under Primary Study M130950 Ms E. Vorster & Prof J. Rodda
SUPERVISOR:	Prof A. Krause and Mrs E. Vorster
APPROVED BY:	Professor CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	04/06/2018

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 Research Office Secretary on the Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193. University of the Witwatersrand. 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**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in April and will therefore be due in the month of April each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

Date 12/06/2018

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix 2

DNA EXTRACTION WORKSHEET

DATE:	EXTRACTED BY:
SOP: GPL4270	
BENCH DECONTAMINATED:	

Salting out DNA extraction ☐ Ref: Miller SA, Dykes DD and Polesky HF (1988). Nucleic Acids Research 16 (3): 1215

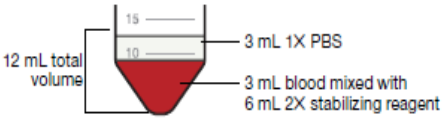
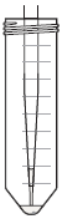
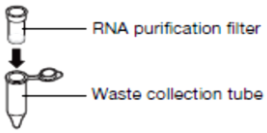
Flexigene DNA Extraction ☐ Lot Number: _____

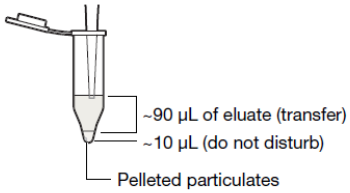
Phenol Cholorform DNA Extraction ☐ Lot Number: _____

LANE	DISEASE CODE	NAME	SURNAME	DATE RECEIVED	[DNA] [ng/ul]	RATIO 260/280	RATIO 260/230	1xTE VOL	COMMENT

Appendix 3

Isolation RNA from whole blood using Tempus kit. Date:

Step	Description	Tick
1	Thaw Tempus RNA whole blood tube (18ºC to 25ºC)	
2	Pour whole blood in 50 mL tube	
3	Pipet 3 mL of 1X PBS into 50 mL tube, **bring total volume to 12 mL** 	
4	Vortex vigorously at maximum speed for **30 sec** " Vortexing for less may cause clogging purification consumable	
5	Centrifuge the tube **4ºC** at **2500 rpm** for **30 min**	
6	Carefully pour off supernatant **Note:** RNA pellet is transparent and invisible	
7	Leave tube inverted on absorbent paper for **1 to 2 min**	
8	Blot the remaining drops of liquid off the rim of the tube with clean absorbent paper	
9	Pipet **400µl** of RNA purification resuspension solution into tube, then vortex briefly to resuspend the RNA pellet 	
10	Label RNA purification filter, then insert the filter into the waste collection tube 	
11	Pre-wet the filtration membrane by pipetting **100µL** RNA purification **Wash solution 1** into purification filter	
12	Pipet **400µL** of the resuspended RNA into the purification filter then centrifuge for **30 sec** at **14 000 rpm**	
13	Remove the purification filter, discard the liquid waste collected in the waste tube, then re-insert the purification filter into the waste tube	
14	Pipet **500µL** RNA purification **wash solution 1** into the purification filter and centrifuge for **30 sec** at **14 000 rpm**. Discard liquid waste	

15	Pipet **500 µL** RNA purification **wash solution 2** into the purification filter and centrifuge for **30 sec** at **14 000 rpm**. Discard liquid waste	
16	Remove the purification filter, discard the liquid waste collected in the waste tube, then re-insert the purification filter into the waste tube	
17	Repeat step **15** and **16**	
18	Centrifuge for **30 sec** at **14 000 rpm** to dry the sample	
19	Transfer the purification filter to a **NEW** labelled collection tube	
20	Pipet **100µL** nucleic acid purification elution solution into purification filter, close the cap and incubate for **2 min** at **70ºC**	
21	Centrifuge for **30 sec** at **14 000 rpm**	
22	Pipet the collected RNA eluate back into the purification filter, then centrifuge for **2 min** at **14 000 rpm**	
23	Discard the purification filter, then transfer approximately **90 µL** of the RNA eluate to a new labelled collection tube  ***NB!*** *When transferring the RNA eluate, carefully pipet the liquid out of the collection tube starting from the top of the liquid to ensure that the pelleted particulates are not disturbed*	
24	Check **2 µL** of RNA on 1% agarose gel and spec. Store RNA at -20ºC or -80ºC for long term storage	

Appendix 4

Reverse Transcription of RNA using Promega ImProm-II system

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

Table 1: RNA Extraction using Tempus Spin RNA Isolation kit

RNA EXTRACTION WORKSHEET								
DATE:				EXTRACTED BY:				
BENCH DECONTAMINATED:								
Tempus RNA Extraction				Lot Number: _____				
LANE	**DISEASE CODE**	**NAME**	**SURNAME**	**DATE RECEIVED**	**[RNA] [ng/ul]**	**RATIO 260/280**	**RATIO 260/230**	**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.

Table 3: Positive control

Reagent	**Lot #**	**Volume / rxn (µl)**
Primer ($Oligo(dT)_{15}$)		1
Positive control RNA (~**1µg**)		2
Nuclease-Free water		2
TOTAL	-	**5µl**

Table 3: Negative control

Reagent	**Lot #**	**Volume / rxn (µl)**
Primer ($Oligo(dT)_{15}$)		1
Nuclease-Free water		4
TOTAL	-	**5µl**

- Set up positive ctrl (supplied) and two negative controls (no-template and no-reverse transcriptase)
-70 °C heating block for 5 min (double stranded RNA becomes single stranded)

- 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:

LANE	DISEASE CODE	NAME	SURNAME	DATE RECEIVED	[cDNA] [ng/ul]	RATIO 260/280	RATIO 260/230	COMMENT
		Positive control						
		Negative control	No template					
		Negative control	No polymerase					

Check on 0.8% agarose gel

Appendix 5

General protocol for MLPA – One tube procedure

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. Vortex reagents briefly before use

Hybe 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
- 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 before use

Ligase-65 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
Mix by pipetting and centrifuge. Extra Ligase-65 mix can be frozen and reused.			
Hybridised products	**-**	8μl	-
TOTAL	**-**	**40μl**	**32μl/reaction**

- ❖ While the samples are at 54°C, add 32μl of the Ligase-65 mix to each reaction, mixing by pipetting up and down.
- ❖ 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 (Differs between the one tube & two tube protocols)

- ❖ Vortex & centrifuge the SALSA PCR primer mix before use
- ❖ Warm the polymerase for 10s in your hand to reduce viscosity
- ❖ Prepare the polymerase mix 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
Mix by pipetting and centrifuge. Extra Polymerase mix can be frozen and reused.			
MLPA ligated products	-	40μl	-
TOTAL	-	**50μl**	**10μl/reaction**

* = 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
- ❖ 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 PCR products with 0.5μl Liz-500 marker and 8.5μl Hi-dye
- ❖ Run products on ABI 3130*xl* genetic analyzer

Name of run: _____ **Date:** _____

Appendix 6

SMN1/SMN2 Dx Worksheet

Asuragen

PCR Master Mix:	Lot	Vol / Rxn (µL)	MasterMix Volumes	Total Rxns:	13
2X PCR mix		3,75	48,75		
SMN1/2 HEX Primer mix		2,75	35,75		
Sample, SMN1 calibrator, SMN control or NTC		*2,0*	*add separately*		
	Total Volume (µL)	8,50	84,50		

Mix aliquot 8µL into individual wells: **Note handling of enzyme and vigorous master mix vortex.**

Add 2µL (20-80ng) of the appropriate sample or control to each well.

Recommended 20ng DNA Input

1 HOLD	25 CYCLES	1 HOLD	1 HOLD
94°C / 2 min	94°C / 30 s 53°C / 30 s 72°C /30 s	72°C / 5 min	4°C forever

Following the PCR run, begin capillary electrophoresis setup or alternatively, store PCR product at -15 to -30 °C until future use. Protect PCR product from light.

CE MasterMix	Lot	Vol / Rxn (µL)	MasterMix Volumes	Total Rxns:	13
Hi-Di Formamide		11,0	143,0		
ROX Ladder		2,0	26,0		
PCR product		*2,0*	*add separately*		
		15,0	169,0		

Mix aliquot 13µL into individual wells:

Add 2µL of PCR product, and mix / centrifuge to remove all bubbles.

Denature plate in pre-warmed thermalcycler and prepare CE for injection. **Consult Asuragen technical services for CE injection support.

1 HOLD	1 HOLD
95°C / 2 min	place on ice

Appendix 7

Copy number detection qPCR

qPCR SMN1, SMN2, CFTR PCR worksheet		Run name:		Bench decontaminated:		Date:	
qPCR Universal PCR Master Mix		Input values	Output values				
	[Stock]	**(M)**	**[Final]**	**[Final] (M)**	**1x Volume (µl)**	**Mastermix (µl)**	X MM
TaqMan™ Universal MM	2X		1X		7		
Forward Primer: SMN1-ex7-F	100 pmole/µl	100 µM	0.9 pmole/µl	900 nM (0.9 µM)	1		
Reverse Primer: SMN1-ex7-R	100 pmole/µl	100 µM	0.9 pmole/µl	900 nM (0.9 µM)	1		
Forward Primer: SMN2-ex7-F	100 pmole/µl	100 µM	0.9 pmole/µl	900 nM (0.9 µM)	1		
Reverse Primer: SMN2-ex7-R	100 pmole/µl	100 µM	0.9 pmole/µl	900 nM (0.9 µM)	1		
Forward Primer: CFTR-F 900nM	100 pmole/µl	100 µM	0.9pmole/µl	900 nM (0.9 µM)	1		
Reverse Primer: CFTR-R 900nM	100 pmole/µl	100 µM	0.9pmole/µl	900 nM (0.9 µM)	1		
TaqMan Probe: SMN1-Ex7	10 µm	10 µm	0.25 pmole/µl	250 nM (0.25µM)	1		
TaqMan™ Probe: CFTR	5 µm	5 µm	0.25 pmole/µl	250 nM (0.25µM)	1		
TaqMan™ Probe: SMN2-Ex7	10 µm	10 µm	0.25 pmole/µl	250 nM (0.25µM)	1		
DNA	X ng/µl		25-100 ng/µl		1		
H2O					1		
TOTAL					18		
						17,0	

Appendix 8

RNA expression RT-qPCR

FL-SMN1, FL-SMN2, SMNΔ2, GAPDH	**Run name:**		**Date:**

qPCR Universal PCR Master Mix		Input values	Output values	**Bench decontaminated:**			X MM
	[Stock]	**(M)**	**[Final]**	**[Final] (M)**	**1x Volume (µl)**	**Mastermix (µl)**	
TaqMan™ Universal MM	2X		1X		5		
Forward Primer:	100 pmole/µl	100 µM	0.9 pmole/µl	900 nM (0.9 µM)	1		
Reverse Primer:	100 pmole/µl	100 µM	0.9 pmole/µl	900 nM (0.9 µM)	1		
TaqMan™ Probe:	10 pmole/µl	10 µm	0.25 pmole/µl	250 nM (0.25µM)	1		
DNA	X ng/µl		25-100 ng/µl		1		
H2O					1		
TOTAL					10		
						9 µl per reaction	
						Add 1 µl cDNA/reaction	

Appendix 9

Patient worksheet
SPINAL MUSCULAR ATROPHY

Method: MLA/AmplideX/qPCR/RT-qPCR **Run #**

Date set up:

Nr	DATE RECEIVED	SMA CODE	ALTERNATIVE CODES	DNA BOX	SURNAME	NAME	DNA CONC (ng/ul)	DILUTION (__ng/ul)		Method:		COMMENT
								DNA	TE BUFFER	*SMN1* Ex7	*SMN2* Ex7	
1												
2												
3												
4												
5												
6												
7												
8												
9					**Neg control**							
10					**Neg control**							
11					**Neg control**							
12					**Blank (NTC)**							

Extend accordingly

Appendix 10

Costing MLPA (reagents, consumables, labour and maintenance)									
Scenario	**Asset cost p/a**	**Staff cost p/a**	**Reagent cost p/a**	**Total costs p/a**	**Outcome: No of Tests p/a**	**Cost p/Outcome**	**Overheads (20%)**	**VAT (15%)**	**Final Cost p/Outcome**
Base Case	47546,56	19648,46	72408,89	139603,91	200,00	698,02	139,60	97,72	935,35
Error Rate	1,00	1,10	1,10		200,00				
Adjusted Cost p/a	47546,56	21613,31	79649,77	148809,64	200,00	744,05	148,81	104,17	997,02
Adjusted Cost p/test	237,73	108,07	398,25	**744,05**					
Costing Results AmplideX® PCR/CE SMN1/2 (reagents, consumables, labour and maintenance)									
Scenario	**Asset cost p/a**	**Staff cost p/a**	**Reagent cost p/a**	**Total costs p/a**	**Outcome: No of Tests p/a**	**Cost p/Outcome**	**Overheads (20%)**	**VAT (15%)**	**Final Cost p/Outcome**
Base Case	47546,56	12363,28	58525,34	118435,19	200,00	592,18	118,44	82,90	793,52
Error Rate	1,00	1,10	1,10		200,00				
Adjusted Cost p/a	47546,56	13599,61	64377,88	125524,05	200,00	627,62	125,52	87,87	841,01
Adjusted Cost p/test	237,73	68,00	321,89	**627,62**					
Costing qPCR (reagents, consumables, labour and maintenance)									
Scenario	**Asset cost p/a**	**Staff cost p/a**	**Reagent cost p/a**	**Total costs p/a**	**Outcome: No of Tests p/a**	**Cost p/Outcome**	**Overheads (20%)**	**VAT (14%)**	**Final Cost p/Outcome**
Base Case	9554,27	12363,28	8962,18	30879,72	200,00	154,40	30,88	21,62	206,89
Error Rate	1,00	1,10	1,10		200,00				
Adjusted Cost p/a	9554,27	13599,61	9858,40	33012,27	200,00	165,06	33,01	23,11	221,18
Adjusted Cost p/test	47,77	68,00	49,29	165,06					

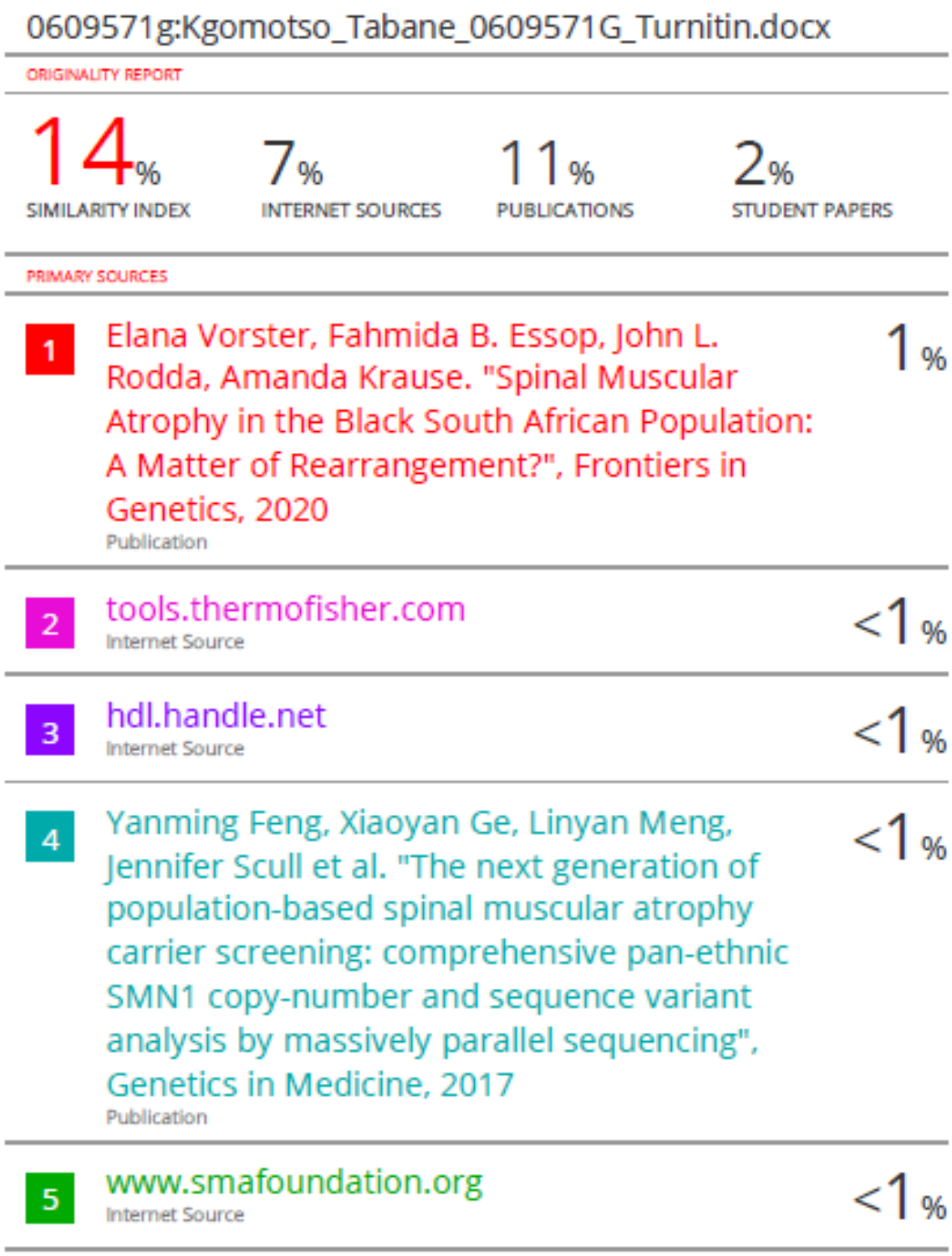
Appendix 11

Known Spinal muscular atrophy genes.

Group	Name	OMIM	Gene	Locus	Mode of inheritance
SMA	Spinal muscular atrophy (SMA)	253300 253550 253400 271150	SMN1	5q13.2	Autosomal recessive
XLSMA	X-linked spinal muscular atrophy type 1 (SMAX1)	313200	NR3C4	Xq12	X-linked recessive
	X-linked spinal muscular atrophy type 2 (SMAX2)	301830	UBA1	Xp11.23	X-linked recessive
	X-linked spinal muscular atrophy type 3 (SMAX3)	300489	ATP7A	Xq21.1	X-linked recessive
DSMA	Distal spinal muscular atrophy type 1 (DSMA1)	604320	IGHMBP2	11q13.3	Autosomal recessive
	Distal spinal muscular atrophy type 2 (DSMA2)	605726	SIGMAR1	19p13.3	Autosomal recessive
	Distal spinal muscular atrophy type 4 (DSMA4)	611067	PLEKHG5	1p36.31	Autosomal recessive
	Distal spinal muscular atrophy type 5 (DSMA5)	614881	DNAJB2	2q35	Autosomal recessive
	Distal spinal muscular atrophy type VA (DSMAVA)	600794	GARS	7p14.3	Autosomal dominant
	Distal spinal muscular atrophy type VB (DSMAVB)	614751	REEP1	2p11	Autosomal dominant
	Distal spinal muscular atrophy with calf predominance	615575	FBXO38	5q32	Autosomal dominant
	Distal spinal muscular atrophy with vocal cord paralysis	158580	SLC5A7	2q12.3	Autosomal dominant
	Congenital distal spinal muscular atrophy	600175	TRPV4	12q24.11	Autosomal dominant
	Scapuloperoneal spinal muscular atrophy (SPSMA)	181405	TRPV4	12q24.11	Autosomal dominant
	Autosomal dominant distal spinal muscular atrophy	158590	HSPB8	12q24.23	Autosomal dominant
	Finkel-type proximal spinal muscular atrophy (SMA-FK)	182980	VAPB	20q13.32	Autosomal dominant
	Jokela-type spinal muscular atrophy (SMA-J)	615048	CHCHD10	22q11.2–q13.2	Autosomal dominant
	Spinal muscular atrophy with lower extremity predominance 1 (SMALED1)	158600	DYNC1H1	14q32	Autosomal dominant
	Spinal muscular atrophy with lower extremity predominance 2 (SMALED2)	615290	BICD2	9q22.31	Autosomal dominant
	Spinal muscular atrophy with progressive myoclonic epilepsy (SMA-PME)	159950	ASAH1	8p22	Autosomal recessive
PCH	Spinal muscular atrophy with pontocerebellar hypoplasia (SMA-PCH)	607596	VRK1	14q32	Autosomal dominant

Appendix 12

Turnit in report



0609571g:Kgomotso_Tabane_0609571G_Turnitin.docx

ORIGINALITY REPORT

14%	7%	11%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Elana Vorster, Fahmida B. Essop, John L. Rodda, Amanda Krause. "Spinal Muscular Atrophy in the Black South African Population: A Matter of Rearrangement?", Frontiers in Genetics, 2020 Publication	1%
2	tools.thermofisher.com Internet Source	<1%
3	hdl.handle.net Internet Source	<1%
4	Yanming Feng, Xiaoyan Ge, Linyan Meng, Jennifer Scull et al. "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", Genetics in Medicine, 2017 Publication	<1%
5	www.smafoundation.org Internet Source	<1%