

**ANALYSIS OF TRANSCRIPTOME AND SPLICING  
CHANGES IN ZEBRAFISH MODEL FOR  
SPINAL MUSCULAR ATROPHY**

**HIMANSHU VYAS**

**NATIONAL UNIVERSITY OF SINGAPORE**

**2016**

**ANALYSIS OF TRANSCRIPTOME AND SPLICING  
CHANGES IN ZEBRAFISH MODEL FOR  
SPINAL MUSCULAR ATROPHY**

**HIMANSHU VYAS**

(M. Tech. Biotechnology,

Indian Institute of Technology Guwahati)

**A THESIS SUBMITTED FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY**

**DEPARTMENT OF BIOLOGICAL SCIENCES**

**NATIONAL UNIVERSITY OF SINGAPORE**

**2016**

## **Declaration**

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have dully acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been used for any degree in any university previously.

---

Himanshu Vyas

January 2016

## **ACKNOWLEDGEMENT**

I would like to thank my supervisor, Assoc. Prof. Christoph Winkler, for his valuable guidance and supervision during the entire project. He has been patient with all the mistakes and confusions that accompanied me throughout my work. He has been a constant source of motivation and a driving force behind the success of this work.

I would also like to thank all the members of Winkler lab, both former and present, for their support in last 4 years. I would like to specially thank Anita for all her help and support during RNAseq work. I am also very grateful to Manish, Husvinee, Angela, Sudha, Shijie, Zoltan, Tingsheng for all the friendly conversations and being there whenever I needed to discuss science. I want to thank Shermaine and Viktor for their help with the CRISPR work. I also very thankful to Mr. Balan and entire fish facility team for providing zebrafish embryos.

I am obliged to Shermaine and Kelvin for providing me with motor neuron transgenic line and Schwann cell calcium sensor line, respectively.

I owe earnest thankfulness to Dr. Candida Vaz for introducing me to the Bioinformatics and teaching me all the software. I also thank our collaborator Dr. S. Mathavan (GIS) for all his help with the RNAseq, Dr. Vivek Tanavde and Dr. Brian Parker (BII) for their help with Bioinformatic analysis of my data. I am grateful to our collaborator Dr. Sibylle Jablonka (University of Wuerzburg) for providing with IGHMBP2 MO and IGHMBP2 antibody for DSMA1 project.

I am grateful to all my friends in Singapore, especially Sarath, Manish, Digant and Dhiraj, who have been there with me since starting, be it holiday trips, night-outs or scientific and political discussions.

Last but not least, I want to thank my wife and my parents for their unlimited love and support without which this was not possible. I wish to dedicate this work to them and my late grandfather, who has been my idol for as long as I can remember.

Thank you.

## Table of contents

<b>ACKNOWLEDGEMENT</b> .....	<b>i</b>
<b>PUBLICATIONS</b> .....	<b>vi</b>
<b>CONFERENCE CONTRIBUTIONS</b> .....	<b>vii</b>
<b>SUMMARY</b> .....	<b>viii</b>
<b>LIST OF FIGURES</b> .....	<b>x</b>
<b>LIST OF TABLES</b> .....	<b>xi</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>xii</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1 Spinal Muscular Atrophy (SMA) .....	1
1.1.1 Loss of Survival of Motor Neuron 1 ( <i>SMN1</i> ) causes Spinal Muscular Atrophy..	1
1.1.2 Role of SMN in assembly of U snRNP complexes .....	4
1.1.3 The role of SMN in motor axons .....	8
1.1.4 Non-cell autonomous effects in SMA .....	10
1.2 The role of Schwann cells for maintenance of motor neurons .....	12
1.3 Distal Spinal Muscular Atrophy Type 1 (DSMA1) .....	15
1.4 RNA sequencing (RNAseq) .....	16
1.5 CRISPR/Cas system .....	18
1.6 Aim of the project .....	20
<b>2. MATERIAL AND METHODS</b> .....	<b>21</b>
2.1 Zebrafish strains and handling .....	21
2.2 Antisense morpholino oligonucleotide (MO) knockdown .....	22
2.3 Whole-mount immunostaining .....	22
2.4 Whole zebrafish embryo dissociation .....	23
2.5 FACS .....	25
2.6 RNAseq library preparation .....	25
2.6.1 Total RNA extraction .....	26
2.6.2 Double stranded cDNA synthesis .....	26
2.6.3 Covaris shearing .....	28
2.6.4 Qubit assay .....	28
2.6.5 cDNA library generation .....	29
2.6.6 Quantitative PCR .....	30
2.7 RNAseq data analysis .....	31
2.8 Whole-mount in situ hybridisation .....	32
2.8.1 Cloning of riboprobes for WISH .....	32
2.8.2 Probe synthesis .....	33
2.8.3 In situ hybridization .....	34
2.9 DNA Sequencing .....	36
2.10 Agarose gel electrophoresis .....	36

2.11 RNA extraction .....	37
2.11.1 DNase I digestion .....	38
2.11.2 RNeasy clean-up .....	38
2.12 First strand cDNA synthesis for RT-PCR .....	39
2.13 Alternative splicing assay by semiquantitative RT-PCR .....	39
2.14 General PCR .....	40
2.15 Western Blot .....	40
2.16 CRISPR/Cas9 .....	43
2.16.1 CRISPR target site design and gRNA preparation .....	43
2.16.2 Cas9 mRNA preparation .....	44
2.16.3 Microinjection .....	45
2.16.4 Genomic DNA (gDNA) extraction .....	46
2.16.5 <i>srsf6b</i> mutant screening .....	46
<b>3. RESULTS .....</b>	<b>47</b>
3.1 A knockdown of <i>Smn</i> leads to motoaxonal defects in zebrafish embryos .....	47
3.2 Generation of a zebrafish model for DSMA1 .....	48
3.2.1 An IGHMBP2 knockdown leads to motoaxonal defects similar to SMA .....	48
3.3 HB9:eGFP/mCherry embryo dissociation and FAC sorting of motor neurons .....	52
3.3.1 Optimization of embryo dissociation conditions and FAC sorting of motor neurons .....	52
3.3.2 Total RNA extraction from sorted motor neurons .....	56
3.4 Preparation of cDNA libraries from <i>Smn</i> -deficient motor neurons .....	57
3.4.1 Synthesis of double stranded cDNA from total RNA derived from <i>Smn</i> -deficient motor neurons .....	58
3.4.2 Preparation of cDNA libraries .....	59
3.4.3 RNA sequencing of <i>Smn</i> -deficient motor neurons .....	61
3.4.4 RNA sequencing of FAC sorted DSMA1 motor neurons .....	62
3.5 Transcriptome analysis identifies differentially expressed genes in <i>Smn</i> deficient motor neurons .....	64
3.5.1 Mapping of RNAseq data to zebrafish genome .....	65
3.5.2 Differential gene expression analysis of <i>Smn</i> deficient motor neurons identifies deregulated transcripts .....	66
3.5.3 Gene Ontology (GO) enrichment using Partek .....	68
3.5.4 Identification of differentially expressed genes in DSMA1 MN .....	68
3.6 <i>Smn</i> knockdown causes alternative splicing changes in motor neurons .....	72
<b>Related neuronal diseases .....</b>	<b>76</b>
3.7 RT-PCR validation of alternative splice events identified in SMA motor neurons .....	77
3.8 Transcriptome analysis of FAC sorted Schwann cells after <i>Smn</i> knockdown .....	79
3.8.1 Dissociation of <i>Smn</i> deficient HB9:D3cpv/SC embryos and FAC sorting .....	80
3.8.2 cDNA library generation and RNA sequencing of SMA Schwann cells .....	81
3.8.3 Differential gene expression analysis of Schwann cells .....	83
3.9 <i>Smn</i> knockdown leads to alternative splicing in Schwann cells .....	86
3.9.1 Identification of alternative splice events using MATS .....	86
3.9.2 RT-PCR validation of alternative splice events identified in SMA Schwann cells .....	87

3.10 srsf6b splicing factor as a novel candidate target of Smn .....	88
3.10.1 Expression pattern of srsf6b .....	89
3.10.2 CRISPR/Cas mediated knock-out of <i>srsf6b</i> .....	90
<b>4. DISCUSSION .....</b>	<b>95</b>
4.1 Method development for RNAseq of FAC sorted motor neurons and Schwann cells from zebrafish embryos.....	96
4.2 Smn deficiency affects gene expression in motor neurons and Schwann cells .....	99
4.3 Deficiency of Smn and Ighmbp2 affects genes related to axonogenesis .....	102
4.4 Identification of alternatively spliced transcripts in SMA motor neurons and Schwann cells.....	104
4.4.1 Smn deficiency affects pre-synaptic proteins in motor neurons .....	106
4.4.2 Smn deficiency leads to alternative splicing of genes involved in axonal transport.....	107
4.4.3 <i>srsf6b</i> as a novel candidate target of Smn .....	110
4.5 Future experiments .....	113
4.6 Conclusion .....	114
4.7 Transcriptome analysis of RNAseq data based on GRCz10 genome assembly.....	115
<b>APPENDIX .....</b>	<b>119</b>
<b>REFERENCES .....</b>	<b>148</b>

## **PUBLICATIONS**

See, K., Yadav, P., Giegerich, M., Cheong, P. S., Graf, M., Vyas, H., Lee, S. G., Mathavan, S., Fischer, U., Sendtner, M., Winkler, C. (2014). "SMN deficiency alters Nrnx2 expression and splicing in zebrafish and mouse models of spinal muscular atrophy." *Hum Mol Genet.* 23(7):1754-70.

Krieger, F., Dombert, B., Vyas, H., Clemens, S., Elflein, N., Alzheimer, M., Rachor, J., Fischer, U., Winkler, C. and Jablonka, S. (2016). "Ighmbp2 deficiency corresponds to differentiation defects in motoneurons." Manuscript in preparation

## CONFERENCE CONTRIBUTIONS

9<sup>th</sup> European Zebrafish Meeting, June 28– July 2, 2015, Oslo, Norway:

Oral: RNA sequencing of FACS-sorted motor neurons and Schwann cells reveals pre-mRNA splicing changes in a zebrafish model for Spinal Muscular Atrophy; Vyas, H., Vaz, C., Parker, B., Tay, S., See, K., Tanavde, V., Mathavan, S., Winkler, C.

18<sup>th</sup> Biological Science Graduate Congress, January 5-7, 2014, Kuala Lumpur, Malaysia:

Oral: Zebrafish models for SMA and SMARD1: Elucidation of the pathomechanism by RNAseq; Vyas, H., See, K., Krieger, F., Jablonka S., Winkler. C.

Singapore Neuroscience Association symposium 2013 on Neurons and Glia: Advances in Health and Disease, April 26, 2013, Singapore, Singapore:

Poster: Characterization of Neurexin2a in a zebrafish model for Spinal Muscular Atrophy; See, K., Koh, A., Vyas, H., Neo, Y. S., Mathavan, S., Winkler, C.

**(Awarded best poster award)**

## SUMMARY

Spinal muscular atrophy (SMA) is a progressive neurodegenerative disease characterized by the loss of lower  $\alpha$ -motor neurons in spinal cord leading to progressive muscle weakness, paralysis and in severe cases death. SMA is caused by reduced levels of ubiquitously expressed Survival motor neuron (SMN) protein. It remains unclear how a SMN deficiency leads to an apparently motor neuron specific phenotype and whether other cell types are also involved in SMA pathology. One hypothesis is based on SMN's critical role in assembly of spliceosomal U snRNPs involved in pre-mRNA splicing, while another hypothesis proposes motor-neuron specific function of SMN. However, the genes downstream of SMN are still largely unknown. Apart from this, involvement of other cells in SMA pathology has also been debated. Distal spinal muscular atrophy type 1 (DSMA1), like SMA, is a neurodegenerative disease affecting motor neurons. It is caused by mutations in the Immunoglobulin  $\mu$ -binding protein 2 (*IGHMBP2*) gene leading to deficiency of this ubiquitously expressed protein. IGHMBP2 is also reported to be involved in RNA metabolism similar to SMA. However, molecular mechanisms leading to motor neuron degeneration remains elusive.

Here in this study, I have optimised a protocol for performing RNA sequencing on FAC sorted motor neurons and Schwann cells under *Smn* deficient conditions. Using this protocol, the transcriptomes of motor neurons and Schwann cells under disease conditions were analysed. Defects in axonogenesis related genes were identified to be deregulated in motor neurons leading to motoaxonal defects in SMA and DSMA1 diseases. Transcriptome analysis identified down-regulation of genes important for functioning of motor neurons and Schwann cells. Differentially expressed genes, however, were different in both cell types indicating that *Smn*

deficiency leads to cell-type specific mRNA level changes. RNAseq data analysis revealed alternative splicing in both cell types. The alternative splicing data strongly suggests that Smn deficiency leads to splicing abnormalities in motor neurons and Schwann cells, which primarily affect genes important for normal physiology including synapse formation and axonal transport. I report on the identification of the splicing factor *srsf6b* in zebrafish as a possible novel downstream target of Smn and propose an 'amplifier mechanism' of splicing deficiencies to explain motor neuron vulnerability in SMA. Taken together, this study, for the first time, identifies splicing abnormalities in zebrafish motor neurons and Schwann cells under Smn deficiency that could be implicated in SMA pathology.

## LIST OF FIGURES

**Figure 1.** Molecular basis of SMA

**Figure 2.** Role of SMN in assembly of uridine-rich small nuclear ribonucleoproteins (UsnRNPs)

**Figure 3.** Motoaxonal defects in zebrafish *smn* morphants

**Figure 4.** Motoaxon outgrowth defects in *lghmbp2* deficient zebrafish embryos

**Figure 5.** Hb9-mCherry positive cells after embryo dissociation.

**Figure 6.** FAC sorting of eGFP positive motor neuron cells

**Figure 7.** Total RNA extracted from FAC sorted motor neurons

**Figure 8.** Assessment of quality for extracted RNA, amplified cDNA and final library during SMA MN library preparation using Agilent 2100bioanalyser

**Figure 9.** Assessment of quality for extracted RNA, amplified cDNA and final library during DSMA1 motor neuron library preparation using Agilent bioanalyser

**Figure 10.** Summary of transcriptomic changes identified in SMA motor neurons by RNAseq analysis

**Figure 11.** Gene Ontology (GO) analysis of differentially expressed genes between *Smn* deficient and control motor neurons

**Figure 12.** Alternative splice events in *Smn*-deficient motor neurons identified by MATS software

**Figure 13.** RT-PCR validation of alternative splicing events in SMA MN identified by RNAseq

**Figure 14.** FAC sorting of YFP positive Schwann cells.

**Figure 15.** Assessment of quality for extracted RNA, amplified cDNA and final library during SMA Schwann cell library preparation using Agilent bioanalyser

**Figure 16.** Gene Ontology (GO) analysis of differentially expressed genes between *Smn*-deficient and control Schwann cells

**Figure 17.** RT-PCR experiments validated alternative splicing changes in SMA Schwann cells identified by RNAseq

**Figure 18.** *Smn* deficiency leads to alternative splicing of *srsf6b* in motor neurons and Schwann cells

**Figure 19.** CRISPR/Cas mediated *srsf6b* knock-out

**Figure 20.** *srsf6b* knock-out leads to motoaxonal defects

**Figure 21.** 'Amplifier mechanism' to explain vulnerability of motor neurons to splicing defects in SMA

**Figure 22.** Gene Ontology (GO) analysis of differentially expressed genes between *Smn* deficient and control motor neurons.

## LIST OF TABLES

**Table 1:** List of used Morpholinos (MO)

**Table 2:** List of antibodies used for immunostaining

**Table 3:** List of primers used

**Table 4:** Preparation of SDS-PAGE gels

**Table 5:** List of antibodies used for Western blot

**Table 6:** *srsf6b* CRISPR target sites

**Table 7:** Number of cells obtained after embryo dissociation with 2 hrs of papain incubation

**Table 8:** List of samples used for SMA MN RNAseq

**Table 9:** List of samples used for DSMA1 MN RNAseq

**Table 10:** Illumina HiSeq sequencing results for SMA motor neuron libraries

**Table 11:** SMA MN RNAseq reads mapped to zebrafish reference genome Zv9 using TopHat

**Table 12:** DSMA1 MN RNAseq results

**Table 13:** List of differentially expressed genes that overlap between DSMA1 MN and SMA MN

**Table 14:** List of potential candidate genes with alternative splicing in motor neurons under *Smn* deficiency

**Table 15:** List of samples used for SMA SC RNAseq

**Table 16:** SMA SC RNAseq results

**Table 17:** SMA SC RNAseq reads mapped to zebrafish reference genome Zv9 using TopHat

**Table 18.** List of deregulated genes in SMA MN related to calcium ion binding and developmental process

**Table 19.** Comparison of mapping percentages between Zv9 and GRCz10 mapped data

## LIST OF ABBREVIATIONS

<b>AS</b>	Alternatively spliced
<b>bp</b>	basepair
<b>BSA</b>	Bovine serum albumin
<b>CaP</b>	Caudal Primary
<b>EtOH</b>	Ethanol
<b>FACS</b>	Fluorescence activated cell sorting
<b>g</b>	G-force or Relative centrifugal force (RCF)
<b>hb</b>	hindbrain
<b>hrs</b>	hours
<b>hpf</b>	hours post fertilization
<b>M</b>	Molarity
<b>mb</b>	midbrain
<b>mins</b>	minutes
<b>MN</b>	Motor neuron
<b>MO</b>	Morpholino
<b>MQ</b>	Milli-Q water
<b>n</b>	number of samples
<b>NBT</b>	Nitro blue tetrazolium
<b>NMJ</b>	Neuromuscular Junction
<b>PBS</b>	Phosphate buffered saline
<b>PBST</b>	PBS+ Tween-20
<b>r.t.</b>	room temperature
<b>rpm</b>	revolutions per minute
<b>RT</b>	Reverse Transcription
<b>SC</b>	Schwann cell
<b>sc</b>	spinal cord
<b>sec</b>	seconds
<b>SMN</b>	Survival of Motor Neuron
<b>SSC</b>	sodium chloride/sodium citrate
<b>SSCT</b>	SSC+ Tween-20
<b>TE</b>	Tris-EDTA buffer
<b>tel</b>	telencephalon
<b>V</b>	volt
<b>WT</b>	wild-type

# 1. INTRODUCTION

## 1.1 Spinal Muscular Atrophy (SMA)

### 1.1.1 Loss of Survival of Motor Neuron 1 (*SMN1*) causes Spinal Muscular Atrophy

Spinal Muscular Atrophy (SMA), an autosomal recessive neurodegenerative disease, is characterized by the loss of lower  $\alpha$ -motor neurons in the spinal cord leading to progressive muscle weakness, paralysis and in severe cases death. It is one of the most common genetic causes of infant mortality (Wirth 2000). This disease is categorized into five subtypes on the basis of time of onset and severity. The most severe form is Type I (Werdnig-Hoffmann disease) with an onset before 6 months and an average lifespan of about 8 months. Type II and Type III (Kugelberg-Welander disease) are less severe forms of SMA with onset between 6-18 months and after 18 months, respectively. Type IV is the mildest of all forms with first symptoms of disease occurring after 30 years (Pearn et al. 1978). Type 0 is the 5th SMA subtype which has prenatal onset and the infant dies after birth.

The SMA-causing gene, Survival Motor Neuron (*SMN*), is ubiquitously expressed and was identified to be located on chromosome 5q13 (Brzustowicz et al. 1990). Lefebvre et al later reported that the *SMN* gene is present in two copies in humans: *SMN1* (telomeric copy) and *SMN2* (centromeric copy), thought to have arisen by intrachromosomal genetic duplication (Lefebvre et al. 1995) (Fig. 1). *SMN1* and *SMN2* genes have a very high level of homology including intronic and promoter sequences. The *SMN2* gene differs functionally from *SMN1* gene due to a single nucleotide transition, C to T, at position +6 in exon 7. This transition is thought to cause disruption of an Exonic Splicing Enhancer (ESE) (Cartegni and Krainer 2002,

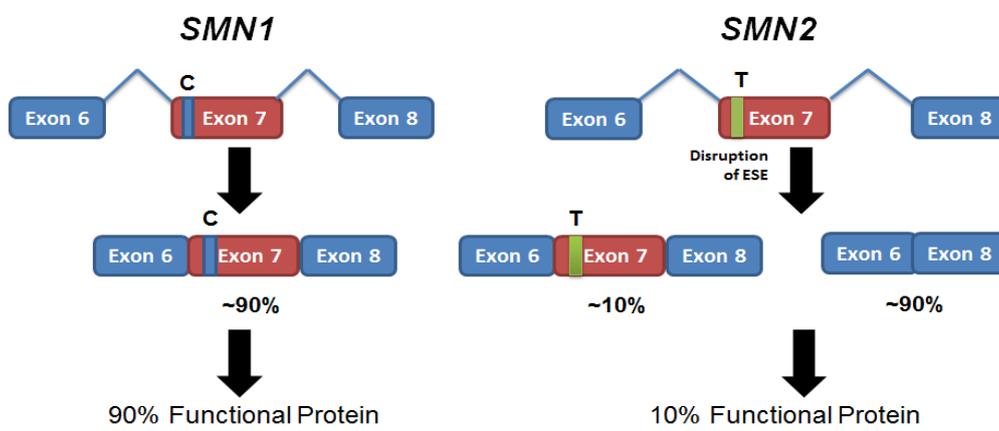
Cartegni et al. 2006) or creation of an Exonic Splicing Silencer (ESS) (Kashmina and Manley 2003) leading to increased skipping of exon 7. This results in generation of about 10% of full length protein while the remaining 90% is truncated protein (SMN $\Delta$ 7), which is unstable and gets degraded (Pellizzoni et al. 1999, Monani 2005, Burnett et al. 2009).

Most of the childhood-onset patients have recessively-inherited mutations in *SMN1*, with more than 95% of the cases attributed to deletions in exon 7 of the *SMN1* gene (Lefebvre et al. 1995). Since *SMN2* produces low levels of functional SMN protein, it cannot fully compensate the homozygous loss of *SMN1*. However, it has been found that *SMN2* modulates the severity of the disease due to its variable copy number in the human genome (Lefebvre et al. 1997). Thus, the copy number of *SMN2* and severity of SMA are inversely correlated. Yet to date, the exact molecular mechanisms by which the reduction of ubiquitously expressed SMN leads to a motor neuron specific phenotype in SMA remains unclear.

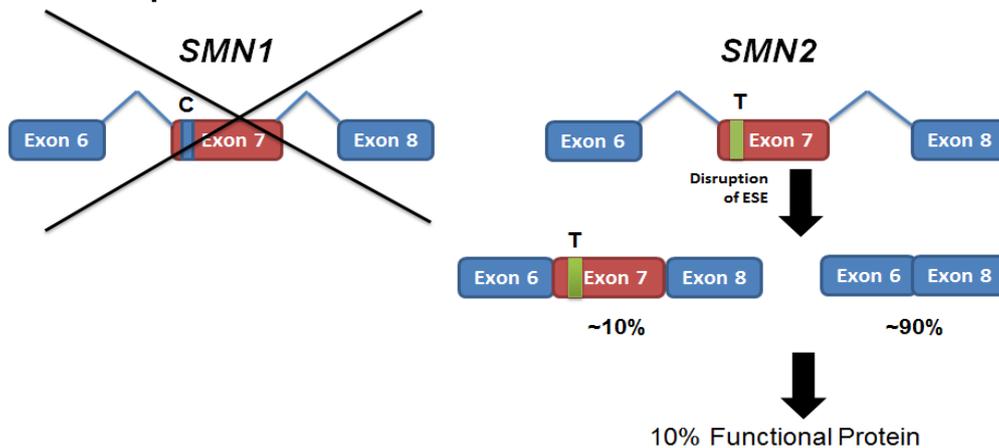
SMN is a ubiquitously expressed protein and in motor neurons it is found to be localized in both cell bodies as well as in axonal growth cones and neurites (Bechade et al. 1999, Dombert et al. 2014). There has been a highly controversial debate for more than a decade over how reduction of this ubiquitously expressed protein leads to a seemingly cell-type specific defect, but the reason is still largely unknown. Currently, there are three hypotheses prevailing among the scientific community to explain the role of SMN mutations in the pathomechanism of SMA. One hypothesis claims that the loss of SMN's ubiquitous function, i.e. assembly of U snRNP spliceosomes, causes an alteration in the splicing of a motor neuron-specific gene (or genes) (Burghes and Beattie 2009). The second hypothesis proposes a motor neuron-specific role of SMN, which is regulation of axonal mRNA transport

(Burghes and Beattie 2009). Disruption of this function, due to loss SMN, has been proposed to cause motor neuron degeneration and ultimately SMA. The final hypothesis in this debate is fairly recent and is based on the idea that there could be non-cell autonomous contributions from non-motor neurons causing the SMA phenotype (Hamilton and Gillingwater 2013, Shababi et al. 2013, Hua et al. 2015).

### A Normal individual



### B SMA patient



**Figure 1. Molecular basis of SMA.** (A) In healthy individuals, the *SMN1* gene produces approx. 90% full length transcript while *SMN2*, due to a base transition, produces about 10% full length and 90% truncated transcript (without Exon 7). Overall sufficient functional protein is produced for survival and maintenance of motor neurons. (B) In SMA patients, due to homozygous mutations, *SMN1* becomes non-functional. *SMN2*, however, remains functional but produces only bare amounts of functional protein which is not enough for survival and maintenance of motor neurons.

### 1.1.2 Role of SMN in assembly of U snRNP complexes

SMN has a well-established role in the assembly of Uridine-rich small nuclear ribonucleoproteins (U snRNPs) (Fischer et al. 1997, Meister et al. 2001, Pellizzoni et al. 2002), which are important for the formation of the spliceosome involved in pre-mRNA splicing (Pellizzoni et al. 1998). There are five types of U snRNPs- U1, U2, U4, U5, and U6 snRNP, that play major roles in splicing (Sperling 2016). Each of these snRNPs consists of one U snRNA, seven Sm proteins and a set of specific proteins (Will and Luhrmann 2001, Matera and Wang 2014). During U snRNP biogenesis, U snRNA molecules, following their transcription by RNA polymerase II, are exported out of the nucleus with the help of nuclear-cap-binding complex proteins (CBC), phosphorylated adaptors for RNA export (PHAX), exportin (XPO1) and Ras-related nuclear protein GTP (RAN) (Fig. 2A) (Izaurralde et al. 1995, Ohno et al. 2000). pICln binds to the spliceosomal Sm proteins (D1, D2, E, F G) forming a 6S complex in the cytoplasm and delivers them to the Protein Arginine Methyltransferase 5 (PRMT5) complex (Fig. 2B) (Chari et al. 2008, Grimm et al. 2013, Neuenkirchen et al. 2015). Sequestration of Sm proteins by pICln prevents their random association with U snRNAs (Chari et al. 2008). The PRMT5 complex later methylates the Sm proteins and recruits them to the SMN complex (Chari et al. 2008). SMN complex consists of SMN oligomer, Gemin2-7 and unr interacting protein (unrip) (Gubitza et al. 2004). This complex acts as a scaffold for formation of heptameric ring of Sm proteins (B/B', D1, D2, D3, E, F & G) onto U snRNAs to form U snRNPs (Fig. 2C) (Meister et al. 2001, Pellizzoni et al. 2002, Zhang et al. 2011). The SMN complex has also been shown to play an important role in preventing non-specific binding of the Sm core (Kroiss et al. 2008). Pellizzoni et al. showed that in the absence of SMN, Sm core proteins bind non-specifically to rRNA and tRNA *in vitro* (Pellizzoni et al. 2002). After the assembly of U snRNPs,

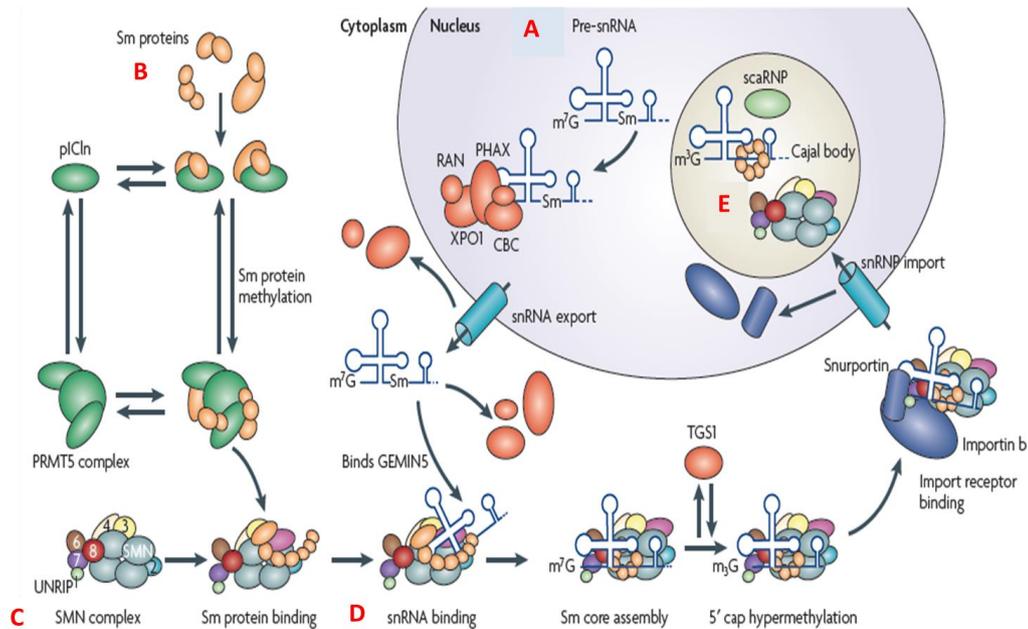
the 5' 7-methylguanosine ( $m^7G$ ) cap of U snRNP hyper-methylates to trimethylguanosine ( $m_3G$ ) with the help of the trimethylguanosine synthetase I enzyme (Fig. 2D) (Mouaikel et al. 2002). This acts as a signal for nuclear import of the U snRNPs via binding to snurportin1 (SPN) and importin  $\beta$  (IP- $\beta$ ) (Palacios et al. 1997, Huber et al. 1998). *In vitro* studies have shown that there is coupled import of SMN and snRNPs into the nucleus (Narayanan et al. 2004). Once inside the nucleus, SMN also helps to target U snRNPs to Cajal bodies for further maturation to form spliceosomes for pre-mRNA splicing (Fig. 2E) (Sleeman and Lamond 1999, Narayanan et al. 2002, Narayanan et al. 2004).

In most of the SMA cases, there is a homozygous loss of *SMN1*, but there are other missense mutations identified in *SMN1*, which have been linked to SMA (Wirth 2000). A common feature of majority of these mutations in *SMN1* is that they are present in the C-terminus of the protein (Exons 6-8) (Seng et al. 2015). Seng et al. showed that this region, which is partially absent in *SMN $\Delta$ 7* transcripts, is very important for snRNP assembly (Seng et al. 2015). Various animal models including mouse, zebrafish and *Drosophila*, have been developed to study SMA (Burghes and Beattie 2009, Bebee et al. 2012, Edens et al. 2015). All of these animals have one copy of SMN gene equivalent to human *SMN1*, loss of which has been shown to be lethal. The time point of lethality, however, is modulated by the levels of maternal SMN (Burghes and Beattie 2009). To model a SMA phenotype in mouse, varying copies of human *SMN2* have been expressed in a *Smn* null mice background. With low copy numbers (two) of *SMN2*, mice develop a severe SMA phenotype while a high copy number (8-16) results in mice with normal phenotype (Hsieh-Li et al. 2000, Monani et al. 2000). By adding human *SMN2* lacking exon 7 (*SMN $\Delta$ 7*) to a *Smn*<sup>-/-</sup>;*SMN2*<sup>+/+</sup> background, this generated  $\Delta 7$

SMA mice or type II SMA mice, which are the most widely used mouse models (Le et al. 2005).

There is a correlation between the levels of snRNP assembly and SMA disease severity. The levels of snRNP assembly activity was found to be strongly reduced in severe SMA mice ( $SMN2^{+/+}; Smn^{-/-}$ ), mildly reduced in mild SMA carrier mice ( $SMN2^{+/+}; Smn^{+/-}$ ) but restored to normal levels in phenotypically rescued high copy SMN2 mice ( $SMN2^{+/+}; SMN2(566)^{+/+}; Smn^{-/-}$ ) (Gabanella et al. 2007). Other proteins involved in snRNP biogenesis such as Gemins 2, 6, and 8, were also found to be significantly decreased in severe SMA mice (Gabanella et al. 2007). Introduction of purified U snRNPs has been shown to rescue the defective axonal phenotype in *Smn* knockdown conditions in zebrafish as well as the developmental arrest in *Xenopus* embryos (Winkler et al. 2005). Thus, this suggests that there is a connection between snRNP assembly and the SMA phenotype. Loss of ubiquitous SMN has been reported to result in tissue specific defects in the repertoire of snRNAs and pre-mRNA splicing, which suggests that splicing is affected to varying extents in many different tissues and is not restricted to motor neurons (Zhang et al. 2008). This suggests that at low levels of SMN, certain transcripts may be more susceptible to aberrant splicing and down-regulated to a greater degree in motor neurons than in other tissues, which leads to the defects observed in the motor neurons (Zhang et al. 2008). The eukaryotic genes consist of two types of introns which are spliced by two classes of spliceosomal snRNPs. Majority of the introns are removed by U2-dependent or Major spliceosome comprised of U1, U2, U4, U5 and U6 snRNPs, while a small percentage of introns (0.5%), however, are spliced by U12-dependent or Minor spliceosome formed by U11, U12, U4atac, U6atac and U5 snRNPs (Lotti et al. 2012, Turunen et al. 2013). SMN deficiency alters the snRNP

profile of tissues in nonuniform manner by only decreasing levels of the minor spliceosome related snRNPs (Gabanella et al. 2007, Zhang et al. 2008). This has led to the hypothesis that SMN deficiency affects the splicing of genes with U12 introns. Consistent with this, type I SMA patients derived lymphoblasts show impaired minor tri-snRNP assembly (Boulisfane et al. 2011). In another report, Lotti et al. reported that a SMN deficiency perturbs the splicing and decreases the expression of U12 intron-containing genes in *Drosophila*. They also identified a U12 intron-containing gene *stasimon*, required for neurotransmitter release in motor neurons of *Drosophila*, as SMN-target gene in this species, which gets aberrantly spliced in SMN deficient conditions (Lotti et al. 2012). One of the former Ph.D. students in our lab, Kelvin See, also showed alternative splicing of *neurexin 2aa* (*nrxn2aa*) isoforms upon *Smn* knockdown in zebrafish embryos (See et al. 2014). Neurexins are cell adhesion molecules present at synapses that interact with neuroligins for forming functional synapses (Dean et al. 2003). These studies provided first evidence that splicing of important neuronal genes is being affected by a deficiency in SMN, which in turn affects the functioning of neurons and might ultimately lead to a SMA phenotype.



**Figure 2. Role of SMN in assembly of uridine-rich small nuclear ribonucleoproteins (U snRNPs).** (A) pre-U snRNAs after transcription are exported out of the nucleus following binding with nuclear-cap-binding complex (CBC) proteins, export adaptor proteins (PHAX), exportin (XPO1) and Ras-related nuclear protein GTP (RAN). (B) In the cytoplasm, Sm proteins bind to pICln and then to PRMT5 complex, which methylate these Sm proteins. The Sm proteins later dissociate from the pICln-PRMT5 complex and bind to SMN complex. (C) SMN complex assembles the Sm core proteins (B/D1/D2/D3/E/F/G) via RG domain interaction between SMN and Sm proteins. (D) The SMN complex then transfers this assembled Sm core onto the U snRNAs to form U snRNPs. The 5' cap of U snRNP gets hypermethylated and binds to snurportin1 (SNP) and importin β (IP-β) to facilitate the intake of the entire complex into the nucleus. (E) Following re-entry into the nucleus, the U snRNPs are targeted to Cajal bodies to form spliceosomes for pre-mRNA splicing. (Modified from Burghes and Beattie 2009)

### 1.1.3 The role of SMN in motor axons

Another hypothesis proposes that SMN is critical for regulation of axonal mRNA transport, and disruption of this leads to SMA (Burghes and Beattie 2009). SMN has been reported to be involved in motor neuron-specific functions apart from the canonical snRNP assembly process. Rossoll et al. reported expression of SMN at the leading edge of neurite outgrowth and growth cones *in vitro* (Rossoll et al. 2002). A year later the same group reported that SMN forms a complex with hnRNP-R, which translocates β-actin mRNA along axons to the growth cones in culture. Any deficiency in SMN

affects this transport of  $\beta$ -actin mRNA to the growth cones (Rossoll et al. 2003). Jablonka et al. later reported that the reduced levels of  $\beta$ -actin mRNA in growth cones affect the translocation of voltage-gated calcium channels (VGCC) to membrane clusters in SMA mouse motor neurons *in vitro* (Jablonka et al. 2007). The VGCC are important for the influx of  $\text{Ca}^{2+}$  ions into motor neurons during synaptic activity. This defect in VGCC clustering correlates to the reduced frequency of local  $\text{Ca}^{2+}$  transients (Jablonka et al. 2007). SMN has also been found to interact with several other RNA binding proteins (RBPs), which are involved in the transport, stability and localized translation of mRNAs in neurons. Several of these proteins, including hnRNP-U, hnRNP-Q (Rossoll et al. 2002), and FMRP (Piazzon et al. 2008), but also EWS (Young et al. 2003), Nucleolin (Lefebvre et al. 2002) and Unrip (Carissimi et al. 2005) have previously been identified as components of KIF5-associated RNA transport granules (Kanai et al. 2004). KIFs are kinesin superfamily proteins that act as molecular motors for both anterograde and retrograde transport of mRNAs selectively to dendrites (Nakagawa et al. 2000, Setou et al. 2000, Kanai et al. 2004). Recently, Dombert et al. showed localization of SMN and hnRNP-R in close proximity in axons and axon terminals *in vitro* and *in vivo* (Dombert et al. 2014). This report provided evidence in support of the hypothesis that SMN is involved in the axonal translocation of hnRNP-R and hnRNP-R bound RNA particles.

However, it remains to be investigated which of the two hypotheses, U snRNP assembly defects and axonal mRNA dysregulation, is primarily affected by low SMN levels and whether there is a possible link between these two hypotheses. It could be possible that reduced level of SMN causes aberrant splicing of transcripts found specifically in motor neurons that

negatively affect axon outgrowth and guidance, contributing to the pathomechanism of SMA.

#### **1.1.4 Non-cell autonomous effects in SMA**

Traditionally, SMN deficiency was thought to exclusively cause lower motor neuron degeneration, leading to denervation and atrophy of skeletal muscles. However, there have been several recent reports that challenge this theory of SMA being solely a motor neuron disease, and instead claim that SMA is a multi-system disorder. This has led to the idea that non-cell autonomous mechanisms could play important roles in manifestation of this disease. The first significant piece of evidence came from Park et al., where they showed that depletion of SMN specifically in motor neurons lead to a dystrophic phenotype but not a full blown SMA phenotype (Park et al. 2010). This shows that systemic loss rather than motor neuron-specific loss of SMN is responsible for generating a SMA disease phenotype.

Hamilton and Gillingwater have reviewed the function of SMN in a variety of cell types other than motor neurons (Hamilton and Gillingwater 2013). Among the various tissues that have been studied with regard to SMA, the most obvious is muscle. In fetuses with severe SMA, skeletal muscle myotubes have been shown to be smaller, thus indicating a delay in muscle growth and maturation (Martinez-Hernandez et al. 2009). This report also suggested that skeletal muscles and motor neurons undergo different pathogenetic processes in SMA during development. Using severe SMA mouse models, Murray et al. showed that the extent of pre-synaptic motor nerve terminal loss does not correlate with shrinkage of the post-synaptic motor endplate suggesting that loss of muscle cells is independent to neuronal degeneration (Murray et al. 2008). This further contributes to the idea of SMN is important for survival of muscles.

Apart from muscles, other organs like heart, vasculature, liver, bones and others have also been studied in SMA animal models. The cardio-vascular system has been suggested to be one of several pathogenic targets in SMA. In severe SMA patients as well as severe SMA mouse models, arrhythmia and/or cardiomyopathy has been observed with high frequency (Bevan et al. 2010, Heier et al. 2010). In severe SMA mouse models, these phenotypes are observed well before the onset of any neuromuscular abnormalities (Heier et al. 2010). SMA mouse models treated with the Histone deacetylase inhibitor (HDACi) Trichostatin A showed vascular necrosis raising the possibility that vascular dysfunction is a part of SMA pathology (Narver et al. 2008). Development of an osteoporosis-like phenotype has been reported in mild SMA mice (Shanmugarajan et al. 2009, Shanmugarajan et al. 2010). This phenotype, unlikely to be caused by muscle atrophy, suggests a possible role for SMN in bone development. More recently, Hua et al. reported deficiencies in liver-related processes during SMA pathogenesis (Hua et al. 2011). They identified the reduction of a hepatic protein, IGF-binding protein, acid labile subunit (IGFALS), which stabilizes IGF-1 protein, in SMA mice. IGF-1 is known to have roles in the peripheral nervous system (PNS), postnatal growth, cardiac development and function, which are all affected in SMA (Sullivan et al. 2008). Interestingly, reduced gene expression of *Igfals* precedes the neuromuscular phenotype in SMA mice (Hua et al. 2011).

Neuronal populations other than motor neurons have also been studied in SMA animal models. In the spinal cord of the SMN $\Delta$ 7 mouse model, a significant reduction in glutamatergic central synapses, positioned closely to the soma and proximal dendrites of motor neurons, was observed (Ling et al. 2010, Mentis et al. 2011). This was due to the loss of proprioceptive sensory neurons that form synapses with motor neurons. A recent study in a

*Drosophila* SMA model also showed that a motor neuron dysfunction is not primary to disease pathology but rather a downstream effect of neural circuit dysfunction (Imlach et al. 2012). They reported that a SMN deficiency primarily affects the function of sensory neurons and interneurons, which can be completely rescued by expressing SMN in these neurons (Imlach et al. 2012). Another recent study describing gastrointestinal symptoms in SMA mice provides interesting evidence that the enteric nervous system is involved in the pathology of SMA (Gombash et al. 2015). A role of SMN in Schwann cells has also been investigated by Hunter et al. These authors reported SMN-dependent intrinsic defects in Schwann cells of SMA mice, such as myelination defects, delayed maturation of axo-glial interactions and abnormalities of the extracellular matrix in the PNS (Hunter et al. 2014). Taken together, all these studies provide support to the idea that SMA is not exclusively a motor neuron disease but could also affect other tissues, which could directly or indirectly lead to SMA in a non-cell autonomous fashion.

## **1.2 The role of Schwann cells for maintenance of motor neurons**

The nervous system is basically composed of two cell types, neurons and glial cells. Glial cells have many indispensable functions during the development and function of the nervous system as well as during disease conditions. Schwann cells are a major type of glial cells in the PNS, where they form layers of myelin sheaths wrapping around the peripheral nerve axons. There are two types of mature Schwann cells, those that are myelinating (forming a myelin sheath around large diameter axons) and those that are non-myelinating (positioned around smaller diameter axons). Myelinated nerves have long myelinated segments interspersed with nodes of Ranvier where the myelin sheath terminates and the electrical impulses are

conducted along the axons in a saltatory fashion. This increases the impulse conduction significantly as compared to non-myelinated axons (Hartline and Colman 2007, Nave 2010). Myelination of axons has another advantage in reduction of energy consumption during restoration of ion gradients (Nave 2010). Traditionally, Schwann cells have been believed to be important for motor neuron survival and support, but recent studies have highlighted their crucial roles in formation, function, plasticity and elimination of synapses (Eroglu and Barres 2010).

Perisynaptic Schwann cells (PSCs) or terminal Schwann cells are an important sub-type of non-myelinating Schwann cells found at neuromuscular junctions (NMJ) where they form a tripartite synapse between nerve, muscle and PSC (Araque et al. 1999, Jessen 2004, Armati and Mathey 2013). Although, PSCs are dispensable for the early nerve-muscle contact, they play a critical role in the maintenance of NMJs, synaptic growth and maturation (Ko and Robitaille 2015). The underlying molecular mechanism, however, is not well understood. PSCs induce the formation of NMJs by secreting TGF- $\beta$ 1 that upregulates neuronal agrin (Feng and Ko 2008). Neuronal agrin is important for clustering of postsynaptic acetylcholine receptors (AChR) (Sanes and Lichtman 2001). This increased clustering of AChR on muscle cells is indicative of the synaptogenesis effect of TGF-  $\beta$ 1 (Feng and Ko 2008). PSCs can detect synaptic activity, decode the transmission and react differentially to produce an adapted modulation (Ko and Robitaille 2015). They increase intracellular  $Ca^{2+}$  levels in response to nerve evoked transmitter release (Jahromi et al. 1992), talk-back to the nerve and muscle cells by releasing gliotransmitters and modulate the synaptic communications (Araque et al. 2014). PSCs also play an important role in activity-dependent synapse elimination during development where they engulf and phagocytose

axons that fail to form functional synapses (Eroglu and Barres 2010, Smith et al. 2013).

Schwann cells can also communicate with axons via secreted vesicles (Lopez-Verrilli and Court 2012). Schwann cells have been reported to transfer polyribosomes to axons after axonal damage and during regeneration (Court et al. 2008, Court et al. 2011, Lopez-Verrilli et al. 2013). These transported ribosomes, secreted in the form of microvesicles (Lopez-Verrilli and Court 2012), suggest that Schwann cells support local protein synthesis in axons during nerve injury and regeneration. Another study reported that Schwann cell-derived exosomes enhance axonal regeneration after nerve injury (Lopez-Verrilli et al. 2013). Schwann cell-derived exosomes containing pathogenic prions have been proposed as means to spread these proteins from CNS to PNS (Fevrier et al. 2004). Taken together, these studies highlight a new emerging concept in the field of neuron-glia interaction whereby axons receive transcripts and/or protein synthesis machineries from Schwann cells.

Given the importance of neuron-glia interactions in the functioning of motor neurons, it is conceivable to postulate an involvement of Schwann cells in the SMA pathomechanism. Consistently, SMN-dependent intrinsic Schwann cell defects in SMA mouse models have recently been reported (Hunter et al. 2014). Similarly, a former Ph.D. student in our lab, Kelvin See, had reported that *Smn* deficiency leads to Schwann cell defects which are independent of motor neuron activity in a zebrafish model of SMA (Kelvin See Ph.D. thesis; manuscript in preparation). Interestingly, the motor axonal defects in zebrafish caused by *Smn* deficiency can be partially rescued by Schwann cell-specific expression of *Smn* in transgenic zebrafish (Shermaine Tay, FYP thesis).

Further investigation needs to be done to explain the role of SMN in Schwann cells and its association to motor neuron degeneration.

### **1.3 Distal Spinal Muscular Atrophy Type 1 (DSMA1)**

DSMA1, like SMA, is an autosomal recessive neurodegenerative disease likewise affecting the  $\alpha$ -motor neurons in the spinal cord. The disease normally manifests in the first two years after birth with a characteristic distal muscle atrophy and paralysis of the diaphragm (Grohmann et al. 2001, Diers et al. 2005, Guenther et al. 2007, Guenther et al. 2009). Although it is known that some of these pathological events correspond to motor neuron loss in the spinal cord (Grohmann et al. 2001) little is known about the molecular mechanisms leading to this motor neuron/axon loss. This disease is caused by mutations in the gene encoding Immunoglobulin  $\mu$ -binding protein 2 (IGHMBP2) located on chromosome 11 (Grohmann et al. 2001).

IGHMBP2 is an ATP-dependent helicase, which belongs to the SF1 helicase superfamily. The precise cellular function of *IGHMBP2* remains elusive, although it has been implicated in transcription and pre-mRNA processing (Shieh et al. 1995, Molnar et al. 1997, Zhang et al. 1999). More recently, IGHMBP2 has been reported to associate with ribosomes and tRNAs suggesting that it is functionally linked to translation (de Planell-Saguer et al. 2009, Guenther et al. 2009). The majority of disease causing mutations identified in DSMA1 patients is found in the helicase domain of IGHMBP2 (Grohmann et al. 2001, Guenther et al. 2007, Guenther et al. 2009). Thus, loss of helicase activity could be the primary defect of mutated IGHMBP2. In order to study the DSMA1 disease and characterise the function of IGHMBP2, the *neuromuscular degeneration (nmd)* mutant mouse has been

used (Cox et al. 1998, Grohmann et al. 2004). The *nmd* mice carry a point mutation in intron 4 of *Ighmbp2* that leads to alternative splicing and production of a truncated transcript (Cox et al. 1998). Mutant mice initially show progressive paralysis in hindlimbs with myopathic changes in the diaphragm at later stages (Cox et al. 1998, Grohmann et al. 2004). Grohmann et al showed that *Ighmbp2* protein is found in the nucleus as well as cell bodies, axons and growth cones, but its expression is low in the nucleus when compared to cell bodies or axons (Grohmann et al. 2004). Unlike in the case of SMA where motor neuron cell death is observed only late when paralysis is already clearly detectable, in DSMA1 the motor neuron death is prominent already during early stages of the disease followed by endplate degeneration (Grohmann et al. 2004, Krieger et al. 2013). This hints at a role of IGHMBP2 in motor neuron survival. However, it is still remains to be investigated why this disease primarily affects motor neurons.

#### **1.4 RNA sequencing (RNAseq)**

RNAseq is a high-throughput sequencing method which can be used to map and quantify the complete transcriptome. Briefly, RNA samples (total RNA or mRNA) is used for generating libraries of cDNA fragments with attached adaptors. The cDNA libraries are then sequenced in a high-throughput manner from one end (single-end sequencing) or both ends (paired-end sequencing). The reads obtained after sequencing are between 30-400 bp in length depending on the type of sequencing platform used. The sequence reads are later mapped to a reference genome or could be assembled together without genomic sequence to generate a transcription map which provides the information on gene expression levels as well as the structure of the gene.

RNAseq has gained considerable attention due to its various advantages over the existing microarray technology (Wang et al. 2009). First and foremost, RNAseq is not dependent on the knowledge of genomic sequence. Thus, it can be used for non-model organisms with no genome sequences available. Second, RNAseq has a very low background signal and no upper limit for quantification. Also, it is highly sensitive for genes with very high and low expression levels, which usually are missed with DNA microarrays. Another advantage is its high reproducibility even with very low RNA input samples. RNAseq has single-base resolution which makes it highly efficient in detecting sequence variations like SNPs and importantly splice isoforms.

RNAseq has become a very attractive tool in zebrafish research. It was recently used for improving the zebrafish reference genome annotation by adjusting intron/exon boundaries of known genes, confirm their expression and improve the coverage of 3' UTRs (Collins et al. 2012). With the availability of a fully annotated zebrafish reference genome as well as pipelines for analysis of zebrafish data, RNAseq has been used for identification of splice variants and novel transcripts, SNP discovery and quantification of transcript levels in zebrafish (Qian et al. 2014). Rosel et al. employed RNAseq to study alternative splicing regulation in zebrafish where they identified a new role for U1C (a snRNP protein) as a splicing regulator (Rosel et al. 2011). RNAseq has also been used in zebrafish developmental biology to understand the changes in transcriptomes during different developmental stages as well as retinogenesis (Vesterlund et al. 2011, Uribe et al. 2012). Taken together, these studies provide strong arguments for the preferred use of RNAseq over DNA microarrays in transcriptomic studies in zebrafish.

## 1.5 CRISPR/Cas system

The clustered regularly interspersed short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system has recently been established as a tool for targeted genome editing in zebrafish and other models with high efficiency. CRISPR/Cas is a type of bacterial defence mechanism that provides adaptive immunity against viruses (Al-Attar et al. 2011, Bhaya et al. 2011, Wiedenheft et al. 2012). This system is composed of CRISPR arrays with gene-targeting sequences or spacers and Cas genes that encode a family of endonucleases (Al-Attar et al. 2011, Bhaya et al. 2011, Wiedenheft et al. 2012). Briefly, the CRISPR RNA (crRNA) binds to trans-activating crRNA (tracrRNA) and direct sequence-specific degradation of viral genomes using Cas proteins (Brouns et al. 2008, Jinek et al. 2012). A recent *in vitro* study has reported that Cas9 endonucleases can be programmed to cleave specific sites in the genome using a single guide RNA (sgRNA) consisting of fused crRNA and a trans-activating crRNA (tracrRNA) (Jinek et al. 2012). This paper introduced CRISPR/Cas system as a tool for precision targeting similar to Zinc finger nucleases (ZFNs) and Transcription activator-like nucleases (TALENs). However, unlike ZFN or TALEN, the CRISPR/Cas system has a low cost and is easy to design for nearly any genomic sequence, and provides very high efficiency of single gene mutations or multiple gene mutations at the same time (Jao et al. 2013, Wang et al. 2013).

In a span of only three years, since Jinek et al published their results (Jinek et al. 2012), the CRISPR/Cas system has become an extensively used tool to modify target genes with great success in human cells, mice, rat, zebrafish, bacteria, fruit flies, nematodes and crops (Pennisi 2013). The CRISPR/Cas system has been recently used for generating single mutations and even multiple mutations in zebrafish, *Drosophila*, human and mouse cells (Chang

et al. 2013, Cong et al. 2013, Hwang et al. 2013, Mali et al. 2013, Wang et al. 2013, Bassett and Liu 2014, Ota et al. 2014). The underlying idea, similar to ZFNs and TALENs, is to generate site-specific double-strand breaks in the genome and allow for erroneous non-homologous end joining (NHEJ) mechanism to repair and generate insertions/deletions (indels) at the targeted site (Durai et al. 2005, Miller et al. 2011). Apart from knock-outs, knock-ins of exogenous sequence is also possible with the CRISPR system (Cong et al. 2013, Wang et al. 2013). Interestingly, the CRISPR/Cas system is not limited to insertions/deletions but can also be used to modulate endogenous gene expression (Larson et al. 2013, Qi et al. 2013, Kabadi et al. 2014). Recently, a new RNA-guided endonuclease, Cpf1, similar to Cas9 was identified that can be used to target specific DNA sequences (Zetsche et al. 2015). Unlike Cas9, Cpf1 creates a staggered double-strand cut distal to a 5' T-rich PAM sequence (Zetsche et al. 2015).

In 2013, Hwang et al. were the first to show the use of this system in zebrafish research (Hwang et al. 2013). After this, the CRISPR/Cas system has completely revolutionized the genome editing field in zebrafish. Multiplex genome engineering is possible in zebrafish by injecting a single Cas9 mRNA along with one or multiple sgRNAs (Ota et al. 2014, Shah et al. 2015). Recently, Jao et al. reported biallelic gene disruption in zebrafish by using zebrafish codon-optimised Cas9 protein (Jao et al. 2013). The observed mutation rates were between 75-99% suggesting complete knock-out of the genes in the F0 generation. This is very advantageous for doing phenotypic analysis in injected fish without the need to wait for homozygosity in the F2 generation. Thus, CRISPR/Cas system has proven to be a simple and robust tool for doing genome engineering in zebrafish for functional analysis of newly identified genes.

## 1.6 Aim of the project

The ultimate aim of this project is to study the changes in the transcriptomes of motor neurons and Schwann cells under SMN deficient conditions, and identify alternative spliced transcripts potentially involved in SMA pathology. The working hypothesis in this study is that SMN deficiency leads to splicing changes in genes critical for motor neurons and Schwann cells, which lead to motor neuron degeneration and finally SMA. These splicing changes could be detected as alternatively spliced transcripts by RNAseq data analysis.

To achieve this goal, the first main aim was to establish a protocol based on Fluorescence activated cell sorting (FACS) in order to isolate pure motor neuron and Schwann cell populations from transgenic zebrafish embryos. In our lab, we already had established motor neuron (HB9:eGFP and HB9:mCherry) and Schwann cell (HB9:D3cpv/SC) transgenic reporter lines that were used in this study. The optimized protocol was then applied to sort these specific cell types under *Smn* deficient as well as DSMA1 conditions.

The second aim was to establish a protocol for generating libraries from *Smn* and *Ighmbp2* deficient FAC sorted motor neurons and Schwann cells and perform RNAseq. The RNAseq data were then used for bioinformatics analysis to identify gene expression changes as well as alternative spliced transcripts in response to *Smn* and *Ighmbp2* knockdown. These alternative spliced transcripts were then screened to identify potential candidate genes, expressed exclusively in motor neurons or neurons in general, that can be implicated in SMA. Finally, a CRISPR/Cas based functional study of the most promising candidate identified by RNAseq was performed.

## **2. MATERIAL AND METHODS**

### **2.1 Zebrafish strains and handling**

All wild-type fish used in this study were obtained from the fish facility of the Department of Biological Sciences, NUS. All fish were kept under controlled conditions (14 hrs light/10 hrs dark, 28°C constant water temperature and salt concentration) and were fed three times a day with brine shrimps. To obtain freshly fertilized zebrafish eggs, one male was separated from one female by a plastic separator in a small mating tank overnight and put together the next morning. After spawning, embryos were collected in a small sieve and gently washed under running water tap before being manipulated during the course of experiments. All injected or chemically treated embryos were kept in Danieau's solution in a 28°C incubator. To prevent pigment formation 1X PTU (0.15 g / 100 ml water) was added to the fish medium at 10 hpf which was replaced every 24 hrs until the embryos reached the desired stage.

#### **Transgenic zebrafish lines**

Motor neuron reporter lines, HB9:eGFP and HB9:mCherry, expressing fluorescent reporter proteins under control of the motor neuron-specific promoter HB9, were used in this study. The HB9:eGFP line was generated by Shermaine Tay (Shermaine Tay, FYP thesis), and the HB9:mCherry line was generated by Kelvin See (Kelvin See, Ph.D. thesis). The used Schwann cell calcium sensor line, HB9:D3cpv/SC, was also generated by Kelvin See. This line drives expression of calcium sensor protein, D3cpv, in Schwann cell precursors, immature Schwann cells and dorsal neural crest cells.

## 2.2 Antisense morpholino oligonucleotide (MO) knockdown

All working solutions for antisense Morpholino Oligonucleotides (MOs; Genetools, USA) used for injections were prepared fresh by diluting the stock MO solutions in appropriate amounts of sterile MQ H<sub>2</sub>O. MO solutions were incubated at 65°C for 10 mins and cooled on ice before loading into glass capillaries (Harvard Apparatus), which had been pulled into sharp tips using a needle puller (Narishige, Japan). Using a Femtojet Microinjector (Eppendorf), MOs were injected into the yolk region just below the cytoplasm into one to two cell stage embryos. Standard Control MO which targets the human  $\beta$ -globin gene and has no target in the zebrafish genome was used as control.

**Table 1. List of used Morpholinos (MO)**

Name	Sequence (5'- 3')	Concentration (mg/ml)
SMN MO	CGACATCTTCTGCACCATTGGC	3.1
Control MO	CCTCTTACCTCAGTTACAATTTATA	3.1
IGHMBP2 MO	CAACCTCCATAGCTGTTAAGTTGGG	6.25

## 2.3 Whole-mount immunostaining

At 31 hpf, embryos were manually dechorionated and fixed in 4% PFA/PBST at 4°C overnight. Fixed embryos were washed 4 x 5 mins at r.t. with PBST before storing them in MeOH at -20°C prior to use. Embryos were rehydrated in 50% MeOH/H<sub>2</sub>O for 5 minutes, in MQ H<sub>2</sub>O for 5 mins and 1 hour at r.t. The embryos were then blocked for 1 hour at r.t. in PBDT solution (1% DMSO, 1% BSA, 0.5% Triton X-100, 2.5% sheep serum, 1X PBS, 0.5% Tween-20). Embryos were then incubated with primary antibody (Table 2) diluted in PBDT at 4°C overnight. Afterwards, embryos were washed in 0.1% Triton X-100 in PBST at r.t. 4 x 1 hour each and incubated with the appropriate biotin-coupled secondary antibody (Vectastain Elite ABC kit, Vector laboratories, USA) at

4°C overnight. Then, embryos were washed again 4 x 1 hour each in 0.1% Triton X-100 in PBST. A Vectastain ABC solution was prepared by mixing a drop each of Solution A and B (Vectastain kit) in 2.5 ml PBS and incubated for 30 min at r.t. After the washes, the embryos were incubated for 1 hour at r.t in Vectastain ABC solution followed by 4 x 30 min washes in 0.1 Triton X-100. The embryos after wash were pre-incubated for 30 min in diaminobenzidine solution (DAB, 1 tablet in 5 ml H<sub>2</sub>O) followed by staining in Urea/H<sub>2</sub>O<sub>2</sub>/DAB solution (1 tablet DAB and 1 tablet Urea/H<sub>2</sub>O<sub>2</sub> in 5 ml H<sub>2</sub>O). The staining was continued until a strong signal was observed and then stopped by removing the staining solution and washing with PBST 3 x 5 min. For storage, embryos were kept in PBST at 4°C.

**Table 2. List of antibodies used for immunostaining**

Antibody	Species	Working solution	Company
Anti-znp1	mouse	1:100	ZIRC
Anti-HuC/HuD	mouse	1:1000	Sigma
Anti-acetylated Tubulin	mouse	1:1000	Molecular probes
Biotinylated anti-mouse IgG	horse	1:1000	Vector laboratories

## 2.4 Whole zebrafish embryo dissociation

The papain dissociation system kit (Worthington Labs, cat no. LK003150) was used for dissociation of zebrafish embryos.

### Preparation of solutions

Papain enzyme solution was prepared by adding 5 ml of Earle's Balanced Salt Solution (EBSS) to the Papain vial (Vial 2) to prepare a solution with a final concentration of 20 units/ml. The solution was incubated in a 37°C water bath for 10 mins for papain to dissolve properly. 500 µl EBSS was added to the DNase I vial (Vial 3) to prepare DNase I solution (2000 units/ml). 250 µl of reconstituted DNase I solution was added to papain to get a Papain/DNase I

solution. 32 ml of EBSS was added to the Albumin-Ovomucoid protease inhibitor vial (Vial 4) and mixed properly. The Stop solution was prepared by mixing 150  $\mu$ l DNase I, 2.7 ml of EBSS and 300  $\mu$ l of the Albumin-Ovomucoid inhibitor solution.

### **Embryo dissociation**

100 zebrafish embryos at 28 hpf were dechorionated and transferred to an Eppendorf tube and excess fish medium was pipetted out. 1 ml of cold Ringer's solution (116 mM NaCl, 2.9 mM KCl, 5 mM HEPES, 1.8 mM CaCl<sub>2</sub>, pH7.2) was added to the Eppendorf tube and embryos were deyolked by pipetting them several times with a 1000  $\mu$ l pipette tip. The deyolked embryos were pelleted down by centrifugation at 300 g for 5 mins at 4°C and the supernatant was discarded. 500  $\mu$ l of Papain-DNase I solution was added to the embryos and incubated on a thermomixer at 30°C for 45 mins and 1200 rpm shaking. The tubes were removed after every 15 min and the embryos were triturated by pipetting thoroughly with a 200  $\mu$ l pipette tip to obtain single cell suspensions. The single cells were pelleted down by centrifugation at 300 g for 5 min at 4°C and the supernatant was discarded. 300  $\mu$ l of the Stop solution was added to the tubes and incubated at r.t. for 5 mins on a rotator shaker to terminate papain digestion. The cells were pelleted down by centrifugation at 300 g, 5 mins, 4°C and re-suspended in Leibovitz's L-15 medium with 2% Fetal bovine serum (FBS). Cells were filtered through the cell strainer cap of a FACS tube and kept on ice until sorting was done. For samples with more than 100 embryos, the volume of the used 2% FBS/L-15 medium was varied.

## **2.5 FACS**

DAPI (1:1000) was added to the cell samples 5 mins before starting of the FACS procedure. Cell sorting was performed in a pre-cooled BD FACS Aria II SORP cell sorter. First, a light scatter plot (side scatter, SSC-A vs. forward scatter, FSC-A) was used to gate homogenous cell populations and to exclude debris and cell aggregates according to size and granularity. The gated cells were then further analysed to exclude doublets, before cell viability was assessed. Only cells with no or low DAPI fluorescence (355 nm laser, 450/50 band filter) were considered as viable single cells. Finally, GFP, mCherry and YFP positive cells were sorted out at a laser excitation of 488 nm (530/30 filter), 561 nm (585/25 filter) and 488 nm (540/25 filter), respectively. Cells were directly sorted into 400 µl TRIzol<sup>®</sup> in a 1.5 ml low-binding reaction tube, whereby multiple sorting rounds of the same sample type were combined into one collection tube. However, not more than 80,000 cells were collected per tube to ensure a 3:1 volume ratio of TRIzol<sup>®</sup> to aqueous component, or 5:1 phenol to chloroform during the RNA isolation step. After collection, the tubes were vigorously vortexed for 2 min and stored at -80°C until further processing. For collecting more than 80,000 cells in a single tube, collection was done into 500 µl of 20% FBS/L-15 medium. The cell suspension was kept on ice. The cells were pelleted by centrifugation for 5 mins at 2000 g at 4°C, the supernatant was carefully removed, leaving behind 50 µl of liquid. The cells were finally re-suspended in 400 µl TRIzol<sup>®</sup> for RNA isolation.

## **2.6 RNAseq library preparation**

Preparation of libraries and RNA sequencing was done in collaboration with Dr. Sinnakaruppan Mathavan at the Genome Institute of Singapore.

### **2.6.1 Total RNA extraction**

RNA was isolated using a phenol-chloroform (TRIzol<sup>®</sup>; Invitrogen) protocol with a column-based purification (PureLink<sup>™</sup> RNA Micro Kit; Invitrogen). Frozen cell samples were thawed and lysed by vortexing, pipetting and using a pestle-homogenizer on ice. Afterwards, the volume was adjusted to 1 ml with fresh TRIzol<sup>®</sup> and the samples were incubated at r.t. for 5 mins. 0.2 ml of chloroform was added and the tubes were vigorously shaken for 30 sec, before incubating them for another 5 mins at r.t. The tubes were centrifuged at  $\geq 12,000$  g for 15 mins at 4°C. The upper aqueous phase was recovered and an equal volume of 100% EtOH was added and mixed by inverting. RNA samples were then further purified using the PureLink Kit following the manufacturer's instructions. RNA was eluted with 14  $\mu$ l RNase free water. 1.2  $\mu$ l of each RNA sample was used for quality assessment and the rest was stored at -80°C before proceeding with the RNAseq library preparation. An on-column DNA digestion was omitted to avoid RNA loss. An Agilent 2100 Bioanalyzer and RNA 6000 Pico Kit (Agilent) was used for RNA quality and quantity assessment. RNA pico chips were prepared following manufacturer's instructions. RNA aliquots were heated to 70°C for 2 mins before loading 1  $\mu$ l onto a pico chip. For the following steps, RNA samples with a RNA integrity number (RIN) of 8.0 or higher were used.

### **2.6.2 Double stranded cDNA synthesis**

First and second strand cDNA synthesis was done with 1-10 ng total RNA using the SMARTer<sup>®</sup> Ultra<sup>™</sup> Low Input RNA for Illumina<sup>®</sup> Sequencing Kit, following the manufacturer's instructions. A positive RNA control with the same amount was included. For first strand cDNA synthesis, RNA samples and control RNA were diluted with 1  $\mu$ l Reaction buffer (19  $\mu$ l of Dilution buffer and 1  $\mu$ l of RNase inhibitor) and Nuclease-free water in a 96-well PCR plate

to get a final volume of 10  $\mu$ l. These samples were then incubated with 1  $\mu$ l of 3' SMART CDS Primer IIA at 72°C for 3 mins. The following components were added to the RNA-Primer mix and mixed thoroughly by pipetting: 4  $\mu$ l of 5X First-strand buffer, 0.5  $\mu$ l of DTT, 1  $\mu$ l of dNTP mix (20 mM), 1  $\mu$ l of SMARTer IIA oligonucleotide, 0.5  $\mu$ l of RNase inhibitor and 2  $\mu$ l of SMARTScribe reverse transcriptase. The samples were incubated for 90 mins at 42°C and the reaction was terminated by incubation at 72°C for 10 mins. After this, a clean-up was done using SPRI AMPure XP beads (Beckman Coulter). For this, 36  $\mu$ l of the beads were added to the reaction mix, and pipetted up and down at least 10 times and incubated for 8 mins at r.t. Samples were then placed on a magnetic separation device and left for more than 5 mins until the solution was clear. The solution was removed using a pipette without disturbing the beads to which the first-strand cDNA was bound. The plate was sealed, briefly spun down and placed again onto the magnetic stand for 2 more mins and the supernatant was carefully discarded.

For second strand synthesis, a long distance PCR was performed using the Advantage 2 PCR kit (Clontech). The following reaction components were added to the DNA-bound beads and mixed thoroughly by pipetting: 5  $\mu$ l of Advantage 2 PCR buffer, 2  $\mu$ l of dNTP mix (10 mM), 2  $\mu$ l of IS PCR primer, 2  $\mu$ l of 50X Advantage 2 polymerase mix and 39  $\mu$ l of Nuclease-free water. The PCR was performed with the following conditions: 1 min 95°C, 15 sec 95°C, 30 sec 65°C, 6 min 68°C, 10 min 72°C. 13 cycles were used for 10 ng total RNA input, 14 cycles for 5 ng and 16 cycles for 1 ng. PCR products were purified using 90  $\mu$ l of fresh SPRI AMPure XP beads. After adding the beads, the solution was mixed thoroughly by pipetting 10 times and incubated for 5 mins to separate the beads. The clear solution was discarded. The beads

were washed twice with 200 µl of 80% EtOH and then air dried. 15 µl of Purification buffer was used for eluting the cDNA. The plate was incubated for 2 mins at r.t. and mixed by pipetting 10 times and placed on the magnetic stand for 1 min. 13 µl of purified cDNA in the clear supernatant was pipetted out carefully without any bead contamination, transferred to DNA lo-bind Eppendorf tubes and stored at -20°C.

cDNA aliquots were analysed using the Agilent 2100 Bioanalyzer and High Sensitivity (HS) DNA Chip. HS DNA chips were prepared by following the manufacturer's instructions and 1 µl of each cDNA sample was loaded onto the chip for QC. Successful cDNA synthesis and amplification should yield no product in the negative control (see Appendix 1), and a distinct peak spanning 400 bp to 9,000 bp, peaked at ~2,000 bp for the positive control RNA sample (see Appendix 1). Contaminated samples will have a broader peak, and an abnormally high yield. Only samples with confirmed quality were used for subsequent shearing and library amplification. The libraries were stored at -20°C until RNAseq submission.

### **2.6.3 Covaris shearing**

The Covaris AFA system was used to produce short cDNA fragments of 200-500 bp. 80 µl purification buffer was added to cDNA samples and the mix was transferred to a Covaris tube. Shearing was done with the following parameters: duty 10%, intensity 5, burst cycle 200, time 5 mins and sweeping mode frequency. The sheared cDNA samples were transferred to new DNA lo-bind tubes.

### **2.6.4 Qubit assay**

A Qubit<sup>®</sup> dsDNA HS Assay Kit was used to determine the DNA concentration in a Qubit<sup>®</sup> 2.0 Fluorometer. Qubit working solutions were prepared by diluting

the Qubit dsDNA HS Reagent 1:200 in Qubit dsDNA HS Buffer in an Eppendorf tube. 1 µl of each cDNA sample was diluted in 199 µl of Qubit working solution and mixed. For each run, two standards were prepared. 10 µl of each standard was diluted with 190 µl of Qubit working solution and mixed. The solutions were incubated in the dark for 2 mins at r.t. Finally, the concentrations were measured on a fluorometer.

### **2.6.5 cDNA library generation**

The cDNA libraries were generated using the NEBNext<sup>®</sup> UltraTM DNA Library Prep Kit for Illumina<sup>®</sup> and NEBNext<sup>®</sup> Multiplex Oligos for Illumina<sup>®</sup> (Index primer set 1). For this, 55.5 µl of sheared cDNA was mixed with 3 µl of End prep enzyme mix and 6.5 µl of End repair reaction buffer, and incubated for 30 mins at 20°C followed by 30 mins at 65°C. The following components were added to the End prep reaction mix and mixed well by pipetting: 15 µl of Blunt/TA ligase master mix, 2.5 µl of NEBNext adaptor and 1 µl of Ligation enhancer. For cDNA inputs of less than 100 ng, 10 times diluted adaptor was used. The samples were incubated at 20°C for 20 mins. 3 µl of USER enzyme was added to all samples before incubating for 20 mins at 37°C. A clean-up of adaptor-ligated cDNA without size selection was performed with AMPure XP beads. 86.5 µl of beads were added to the samples, mixed by pipetting and incubated for 5 mins at r.t. The sample plate was placed on the magnetic stand for 5 mins to allow the beads to separate and the clear supernatant was carefully discarded. The beads were washed thrice with 200 µl of 80% EtOH and air dried. The DNA was eluted from the beads by adding 28 µl of 0.1X TE, mixed well and the sample plate was placed on a magnetic stand for separating the beads from cDNA. 23 µl of clear supernatant containing adaptor ligated DNA was carefully removed.

The DNA was then PCR amplified by adding 25 µl of NEBNext high fidelity 2X PCR master mix, 1 µl of Universal primer and 1 µl of the corresponding Index primer (primers 1-12). PCR was done with the following conditions: 98°C for 30 sec, 98°C for 10 sec, 65°C for 30 sec, 72°C for 30 sec, 72°C for 5 min. The used cycle numbers were according to the input amount of sheared cDNA based on Qubit readings: 9 cycles for 100-300 ng, 12 cycles for 35-60 ng, 15 cycles for 20-35 ng, and 18 cycles for 5-20 ng of cDNA. The amplification product was cleaned up using AMPure XP beads. Libraries were kept at -20°C. A quality assessment was done using an Agilent 2100 Bioanalyzer as described previously and qPCR was performed for quantifying the amount of each library.

#### **2.6.6 Quantitative PCR**

For qPCR analysis, serial dilutions of the library samples were used with the Light Cycler<sup>®</sup> 480 SYBR Green I Master from Roche together with qPCR Primer 1.1 (5'AATGATACGGCGACCACCGAGAT 3') and qPCR Primer 2.1 (5'CAAGCAGAAGACGGCATACGA 3') from 1st BASE Oligos on a Roche Light Cycler<sup>®</sup>. Two criteria defined a successful qPCR assay: the R<sup>2</sup> (coefficient of determination) values of the linear regression equation were bigger than 0.980 and CP (crossing point cycle; or Ct cycle threshold) values of the replicates were similar, which also reflected in evenly spaced amplification curves. After calculation of the molarities, the libraries were multiplexed and sequenced on an Illumina HiSeq platform.

## **2.7 RNAseq data analysis**

### **RNAseq data processing and alignment**

Raw RNAseq data obtained from sequencing were cleaned by removing adaptor sequences (5'-5Phos/GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT C/deoxyU/A CAC TCT TTC CCT ACA CGA CGC TCT TCC GAT C\*T-3') using the CutAdapt 1.9.1 tool (See Appendix 2A for commands). The adaptor sequences from the 5' and 3' ends were removed to yield actual RNA sequences. The trimmed sequence reads were then mapped to the zebrafish reference genome, Zv9, using TopHat 2.0.12 tool (See Appendix 2B for commands). The TopHat output (bam files) contained sequence alignment data for all reads, which were used for further downstream analysis.

### **Quantification and differential gene expression analysis**

Differential gene expression analysis was done to detect genes, which are expressed differentially among control and test samples. The analysis was done using Partek Genomics Suite 6.6 at the Bioinformatics Institute Singapore. Bam files obtained from the TopHat program were uploaded into the Partek software that uses the sequence alignment data to quantify expression levels of all genes/transcripts. The expression levels of genes/transcripts are denoted in terms of Reads Per Kilobase of exon model per Million mapped reads (RPKM). Analysis of variance (ANOVA) test was used to obtain the differential expression analysis. Finally, a list of genes with significant differential expression (Fold change >2 and p-value <0.05) was generated.

### **Alternative splicing events analysis**

Multivariate Analysis of Transcript Splicing (MATS; version 3.0.9) was applied to the aligned RNAseq data (bam files) to identify alternatively spliced (AS)

transcripts. The alternatively spliced events were categorised into: skipped exon (SE), mutually exclusive exon (MXE), retained intron (RI), alternative 3' splice site (A3SS) and alternative 5' splice site (A5SS). Events with a false discovery rate (FDR) less than 0.1 were filtered out as most significant AS events.

### **Gene ontology (GO) enrichment analysis**

GO enrichment analysis was performed using the list of genes with statistically significant differential expression or significant AS events. Partek Genomics Suite 6.6 was used for the analysis. The program applies Fischer's exact test to quantify the enrichment scores and p-value for each GO term identified. Terms with p-value less than 0.05 (Enrichment score >3) were filtered out as most significant.

## **2.8 Whole-mount in situ hybridisation**

### **2.8.1 Cloning of riboprobes for WISH**

*srsf6b* riboprobes (sense and antisense) were generated from linearized plasmids by *in vitro* transcription. For plasmid cloning, primers were designed to amplify a 593bp region spanning exon 4 to exon 6 in the *srsf6b* cDNA (transcript ENSDART0000002318). Phusion DNA polymerase (Finnzymes) was used to amplify the target region from cDNA generated from wild-type embryos at 31 hpf. The PCR product generated by phusion polymerase had blunt ends, which was suitable for cloning into the pJET 1.2/blunt Cloning vector (Thermo scientific). This pJET vector contains multiple cloning sites and a T7 promoter site for *in vitro* transcription. For cloning, 1-4 µl of PCR product was mixed with 5 µl of 2X Reaction buffer, 0.5 µl of pJET vector (50

ng/μl), 1 μl of T4 DNA ligase (5 U/μl) and topped up to 10 μl with MQ H<sub>2</sub>O. The ligation mix was incubated for 15 min at r.t., before transformation into E. Coli DH5α chemically competent cells. Colony PCR and sequencing were performed to select positive colonies with proper orientation of the riboprobe sequence. Forward integration of the PCR product was used for sense probe preparation, while the reverse orientation was used for antisense probes.

## **2.8.2 Probe synthesis**

### **Linearisation of pJET/Insert plasmid DNA**

XbaI (Fermentas) restriction enzyme was used for linearizing the pJET vectors containing the inserts. The XbaI cutting site was at the end of the insert sequence. The digestion reaction was incubated overnight at 37°C and checked on a 1% agarose gel to confirm complete linearization. The linearised product was gel extracted and purified to remove digestion reaction components before starting the *in vitro* transcription.

### ***In vitro* transcription**

For *in vitro* transcription, 1 μg of linearised DNA template was mixed with 4 μl of 5X transcription buffer, 2 μl of 10X DIG-labeled NTP mix, 0.5 μl of Ribolock RNase inhibitor, 1 μl of T7 RNA polymerase and topped up to 20 μl with MQ H<sub>2</sub>O. The reaction mix was incubated at 37°C for 2 hrs and 40 mins. 1 μl of DNase I was added to the reaction mix and incubated for 20 min at 37°C to remove the DNA template. A RNA clean-up was performed using RNeasy mini kit and the RNA was eluted in 50 μl. The eluted RNA was mixed with 150 μl MQ H<sub>2</sub>O, 20 μl 3M Sodium acetate (pH 5.2) and 500 μl 100% EtOH for overnight precipitation at -80°C. Following this, the samples were centrifuged at full speed for 45 mins at 4°C and the supernatant was discarded. The pellet was washed with 500 μl 80% EtOH and centrifuged at full speed for 30 min at

4°C. The RNA pellet was air dried to remove excess EtOH and re-suspended in 25 µl MQ H<sub>2</sub>O. To check for the purity and correct size of the generated riboprobe, 1 µl was mixed with 1 µl RNA loading dye and 8 µl H<sub>2</sub>O and denatured at 80°C for 10 mins. After denaturation, the solution was visualised on a 1% agarose gel. 1 µl of riboprobe was also used for checking the concentration on a Nanodrop. The remaining 23 µl riboprobe was mixed with 77 µl HybMix to generate a stock riboprobe solution, which was stored at -20°C.

### **2.8.3 In situ hybridization**

#### **Preparation of embryos**

Embryos at 31 hpf were fixed in 4% paraformaldehyde/phosphate buffered saline containing 0.01% Tween-20 (PFA/PBST) overnight at 4°C. On the next day, the embryos were washed in 1X PBST 4 x 5 mins, 1 x 5 mins in 100% MeOH and finally stored in 100% MeOH at -20°C until use. The MeOH stored embryos were rehydrated in a stepwise gradient of 75%, 50% and 25% MeOH/PBST for 5 mins each at r.t., followed by 2 x 5 mins washes with 1X PBST at r.t.

#### **Proteinase K treatment**

The rehydrated embryos were incubated in freshly diluted proteinase K (1:8000 dilution in 1X PBST) for 5 mins at r.t. The treatment was stopped by two washes with glycine (2 µg/ml in PBST) followed by re-fixation in 4% PFA/PBST for 20 min at r.t. The fixed embryos were washed with 1X PBST, 5 x 5 mins at r.t.

## **Hybridisation**

Embryos were pre-hybridised in 500 µl of HybMix in a water bath at 65°C for 1 hour. The riboprobe stocks were diluted 1:50 or 1:100 in HybMix to prepare working solutions. Diluted riboprobes were denatured at 80°C for 10 mins and immediately chilled on ice prior to use. After pre-hybridisation, the HybMix was removed and embryos were incubated overnight at 65°C in the heat-denatured riboprobe solution. On the next day, the riboprobes were collected and stored at -20°C for future use. The embryos were washed twice in SSCT Buffer I (50% formamide/ 2X SSC Tween) for 30 mins, once in SSCT Buffer II (2X SSC Tween) for 30 mins and twice in SSCT Buffer III (0.2X SSC Tween) for 30 mins. All washes were performed at 65°C. Afterwards, embryos were washed twice in 1X PBST for 1 min.

## **Antibody incubation**

Embryos were then incubated in 500 µl block solution (5% sheep serum/PBST) for 1 hour at r.t., followed by incubation in 500 µl alkaline phosphatase coupled Anti-DIG antibody (1:2000) for 2 hrs at r.t. Embryos were then washed 6 times in 1X PBST for 20 min at r.t to remove any unbound antibody. One wash was performed overnight at 4°C.

## **NBT/BCIP staining**

Embryos were incubated twice in pre-staining buffer (0.1 M NaCl, 0.05 M MgCl<sub>2</sub>, 0.1 M Tris-HCl pH 9.5 and 0.1% Tween) for 5 mins at r.t. NBT staining buffer (0.1 M NaCl, 0.1 M Tris-HCl pH 9.5, 0.1% Tween and 2% NBT) was added to the embryos and the samples were incubated in the dark at r.t until the colour had developed. The staining time for *srsf6b* riboprobe was 4-5 hrs. The staining was stopped by three washes of 1X PBST for 5 mins and the embryos were finally stored at 4°C in 4% PFA/PBST.

## **2.9 DNA Sequencing**

All sequencing reactions were performed using the BigDye® Terminator v3.1 cycle sequencing kit on a PCR thermal cycler. A typical setup of the PCR sequencing reaction includes 2 µl of BigDye ready reaction premix, 1 µl of 5x BigDye sequencing buffer, 1 µl of primer, 100 ng of DNA template and MQ H<sub>2</sub>O to give a final volume of 10 µl. The thermal cycler conditions used were as follows: 96°C for 2 mins, followed by 35 cycles of 96°C at 15 sec, 50°C to 60°C (depending on the T<sub>m</sub> of primers) for 15 sec and 62°C for 4 mins. The products were then kept at 4°C before precipitation. To precipitate the PCR products, samples were transferred to fresh 1.5 ml tubes. 10 µl MQ H<sub>2</sub>O, 2 µl of 3M Sodium acetate pH 5.2, and 50 µl of 100% EtOH were added to the solution and mixed well by pipetting. The samples were then centrifuged at full speed for 45 mins at 4°C in a cooling centrifuge (Sorvall Legend, Thermoelectron). The supernatant was carefully removed and 500 µl of pre-chilled 70% EtOH was used for washing the DNA pellet by centrifuging at full speed for 30 mins at 4°C. The supernatant was carefully removed and the DNA pellet was air dried at 50°C in a heat block for 5-10 mins. The precipitated PCR products were kept at -20°C prior to sending for sequencing. Sequencing electrophoresis reactions were performed by the DNA Sequencing Laboratory (DSL) staff on an Applied Biosystems 3130xl Genetic Analyser (ABI) and sequences were analysed using GENTle 1.9.4.

## **2.10 Agarose gel electrophoresis**

Agarose gels were prepared freshly by dissolving agarose powder (1st Base) in 1X TAE buffer to obtain the appropriate concentration, typically 1% to 3%. The SYBR® safe DNA gel stain solution was added to the dissolved solution at 1:15,000 dilution, and cast into gel casting moulds. Electrophoresis was

carried out at 90-110V constant voltage in 1X TAE running buffer using a gel electrophoresis chamber (BioRad) and PowerPac™ basic power supply unit (BioRad). Gel documentation was performed using the GeneSnap software (SynGene) and a G:BOX gel documentation system (SynGene). ImageJ was used for quantification of gel band intensities on background corrected images. The housekeeping genes *gapdh* or *β-actin* were used for normalization to ensure equal loading amounts and accurate comparisons of ratios or intensities.

## **2.11 RNA extraction**

20-30 embryos per sample at the appropriate stage were taken for RNA extraction. The embryos were homogenised on ice in 350 µl of Buffer RLT with β-mercaptoethanol (1:100) using pestles. The lysate was centrifuged for 3 mins at full speed (14,000 g) using a table top centrifuge (Sorvall legend micro 21, Thermoelectron) and the supernatant was transferred to a fresh 1.5 ml Eppendorf tube. 1 volume of 70% EtOH was added and mixed by pipetting. The mixture, up to 700 µl, was then transferred to a fresh RNeasy spin column and centrifuged for 15 sec at 10,000 rpm, before discarding the flow through. 700 µl of Buffer RW1 was added to the spin column, centrifuged at 10,000 rpm for 15 sec before discarding the flow through. 500 µl of Buffer RPE was used to wash the spin column by centrifugation at 10,000 rpm for 15 sec. One more wash with 500 µl Buffer RPE was done by centrifugation at 10,000 rpm for 2 mins. The RNeasy spin column was transferred to a fresh 2 ml collection tube and centrifuged at full speed for 1 min to remove any residual liquid. The RNeasy spin column was then transferred to a fresh 1.5 ml Eppendorf collection tube and 30 µl of RNase free H<sub>2</sub>O was added into the

centre of the column for elution by centrifugation at 10,000rpm for 1 mins. The elution was repeated with another 30 µl of RNase free H<sub>2</sub>O.

### **2.11.1 DNase I digestion**

All samples used for semiquantitative RT-PCR were subjected to an additional step of DNase I digestion to eliminate traces of genomic DNA contamination prior to use. 60 µl of RNA samples were incubated with 2 µl of DNase I enzyme (Fermentas) for 60 mins at 37°C in a covered heat block (Eppendorf Thermostat). Samples were then subjected to a subsequent round of RNA clean-up using the RNeasy mini kit.

### **2.11.2 RNeasy clean-up**

Clean-up of RNA was performed after DNase I digestion to purify the RNA before further experiments. The RNA samples were adjusted to a volume of 100 µl with RNase free H<sub>2</sub>O, before addition of 350 µl of Buffer RLT and mixed well. 250 µl of 100% EtOH was added to the solution, mixed well and transferred to an RNeasy mini spin column. The flow through was discarded and the spin column was washed with two successive rounds of 500 µl of Buffer RPE by centrifugation at 10,000 rpm for 15 sec and 2 mins, respectively. One more round of 1 min centrifugation at 10,000 rpm was done on an empty column to remove residual liquid. The RNeasy spin column was then transferred to a fresh 1.5 ml tube and 30 µl of RNase free H<sub>2</sub>O was added to the centre of the column and centrifuged at 10,000 rpm for 1 min. The elution was repeated with 30 µl of RNase free H<sub>2</sub>O. The eluted RNA samples were stored at -80°C.

## **2.12 First strand cDNA synthesis for RT-PCR**

Reverse transcription (RT) of the extracted RNA was used to generate cDNA using the Revertaid first strand cDNA synthesis kit (Fermentas). Total RNA in the range of 1 ng to 1 µg was used as input. RNA samples were mixed with 1 µl of oligo dT primer and MQ H<sub>2</sub>O to a volume of 12 µl in PCR tubes. The tubes were incubated at 65°C for 5 mins in a PCR machine (Veriti Thermal Cycler, Applied Biosystems), spun down and kept on ice. The following reagents were added to the tubes: 4 µl of 5X reaction buffer, 1 µl of RiboLock RNase Inhibitor (20 U/µl), 2 µl of 10 mM dNTP mix and 1 µl of RevertAid M-MuLV reverse transcriptase enzyme (200 u/µl) and incubated at 42°C for 60 mins, followed by 70°C for 5 mins and cooled at 4°C. Minus RT (-RT) controls were generated for each batch of RNA samples by using the same setup, except for the omission of the reverse transcriptase enzyme. Samples were stored at -80°C until further use.

## **2.13 Alternative splicing assay by semiquantitative RT-PCR**

Semiquantitative RT-PCR was applied for validation of the alternative splicing events identified by bioinformatics analysis. Primers were designed to bind to exons flanking the alternatively spliced exon (See Table 3). Since the PCR products obtained differed in size by 33-100 bp, 2% gels were used for proper separation.

## 2.14 General PCR

**Table 3. List of primers used**

zfSrsf6b_Ex4_Fwd	GTTTCGCTCCCACTCTGATATG	riboprobe for <i>srsf6b</i> (593 bp)
zfSrsf6b_Ex6_Rev	GAGAAGCAGATCTAGAACGGGAA C	
pJET1.2 Forward Sequencing Primer	CGACTCACTATAGGGAGAGCGGC	colony screening
pJET1.2 Reverse Sequencing Primer	AAGAACATCGATTTTCCATGGCAG	
zfSrsf6b_Ex1_F	GGTGTACATCGGCAAGCTG	alternative splicing at exon 2 of <i>srsf6b</i>
zfSrsf6b_Ex4_R	CTCATATCAGAGTGGGAGCGA	
zfSnx9b_Ex5_F	GTGTTTGACAATAATGCCTCTGG G	alternative splicing at exon 6 of <i>snx9b</i>
zfSnx9b_Ex7/8_R	CTTCCCCGCTATAAACAGACAAC	
zfZgc:91909_Ex1_F	GGAGGAAACACATACTGCTCTTC	alternative splicing at exon 3 of <i>zgc:91909</i>
zfZgc:91909_Ex4_R	CTTGTAAGTGGCCTGAATGAGAAA C	
zfSnx27a_Ex10_F	CTTTGAATACGTGCGGGGAG	alternative splicing at exon 12 of <i>snx27a</i>
zfSnx27a_Ex13_R	TCTGCGTGTTATGGTGGTCT	
CRISPR RNAamp F	CATTATGGTGAAAGTTGGAAC	gBlock amplification
CRISPR RNAamp R	AAAAGCACCGACTCGGTGCCAC	
zfSrsf6b_5'UP_Crispr	AGTGTTTAGTATCGCAGGAAGTG	<i>srsf6b</i> mutant screening
zfSrsf6b_Intron1_Crispr	TGTACACAAACAATCCGCCATT	

## 2.15 Western Blot

### Sample preparation

70-80 dechorionated embryos at 31 hpf were transferred to an Eppendorf tube and excess fish medium was added. 1 ml of deholking buffer (55 mM NaCl, 1.8 mM KCl, 1.25 mM NaHCO<sub>3</sub>, 4 mM PMSF) was added and embryos were deholked by pipetting them up and down with 200 µl pipette tips. The embryos were shaken at 1000 rpm for 5 min at r.t. on a thermomixer followed by centrifugation at 300 g, 30 sec at r.t., and the supernatant was discarded. The deholked embryos were washed twice with 1 ml wash buffer (110 mM NaCl, 3.5 mM KCl, 2.7 mM CaCl<sub>2</sub>, 10 mM Tris/Cl pH 8.5) by incubating on a

thermomixer at r.t. 1000 rpm for 2 mins and centrifugation at 300 g, 30 sec. The embryos were lysed by adding RIPA buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate, 4 mM PMSF; 1  $\mu$ l per embryo) and pipetting with 200  $\mu$ l tips on ice. An equal volume of 2X SDS loading buffer (125 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 0.2% bromophenol blue and 10% 2-mercaptoethanol) was added to the tube and mixed by pipetting before heating at 95°C for 5 mins. The samples were frozen at -20°C until further use.

### SDS-PAGE

SDS denaturing gels were used for analysing the protein samples. A 10% resolving gel and a 5% stacking gel were prepared (Table 4). The frozen protein samples were thawed, heated at 95°C for 5 mins and centrifuged at full speed for 3 mins. 25  $\mu$ l of each sample and 5  $\mu$ l of a PageRuler pre-stained protein ladder (Fermentas) was loaded onto the gel and run for 60 mins at a constant voltage of 200V in electrode buffer (25 mM Tris, 190 mM Glycine, 0.1% SDS with pH 8.3).

**Table 4. Preparation of SDS-PAGE gels**

	<b>Resolving gel (10%)</b>	<b>Stacking gel (5%)</b>
<b>H<sub>2</sub>O</b>	4.1 ml	5.7 ml
<b>30% Acrylamide/Bis solution</b>	3.3 ml	1.7 ml
<b>Buffer</b>	2.5 ml of 1.5 M Tris-HCl pH 8.8	2.5 ml of 0.5 M Tris-HCl pH 6.8
<b>10% SDS</b>	100 $\mu$ l	100 $\mu$ l
<b>10% APS</b>	100 $\mu$ l	100 $\mu$ l
<b>TEMED</b>	10 $\mu$ l	10 $\mu$ l
	10 ml	10 ml

## Electrophoretic transfer and immuno blot

After gel separation, the stacking gel was removed and the remaining gel piece was set up for electrophoretic transfer. A PVDF membrane was used for transfer after activating it with 100% MeOH. The transfer was done using a Mini trans-blot electrophoretic transfer cell (BioRad) at a constant current of 200 mA for 1 hour. A transfer buffer (25 mM Tris, 190 mM Glycine, 0.1% SDS, 10% MeOH with pH 8.3) was used during the transfer and cooled down with an ice pad placed into the chamber. After transfer, the PVDF membrane was washed thrice with 1X TBST (50 mM Tris, 150 mM NaCl, 0.1% Tween with pH 7.4) for 10 mins on a rotator shaker followed by blocking in 5% milk in 1X TBST for 1 hour at r.t. The membrane was then incubated with primary antibody diluted in 5% milk/1X TBST and incubated overnight at 4°C. After this, the primary antibodies were recovered and the membrane was washed 3 times in 1X TBST for 10 mins. This was followed by incubation with secondary antibody (diluted in 5% milk/1X TBST) for 1.5 hrs at r.t. The membrane was washed 3 times in 1X TBST for 10 mins before visualisation of protein samples. For visualisation, the SuperSignal™ west femto maximum sensitivity substrate was used.

**Table 5. List of antibodies used for Western blot**

Antibody	Species	Working solution	Company
Anti-human IGHMBP2	Goat	1:50	Santa cruz
Anti-PCNA	Mouse	1:500	Abcam
Anti-goat HRP	Donkey	1:20,000	Abcam
Anti-mouse HRP	Goat	1:5,000	Abcam

## 2.16 CRISPR/Cas9

### 2.16.1 CRISPR target site design and gRNA preparation

CRISPR target sites in *srsf6* were identified using the CRISPRscan program (<http://www.crisprscan.org/>) (Moreno-Mateos et al. 2015). For this, the gene sequence of *srsf6b* (ENSDART00000002318) was submitted and the software predicted all possible target sites with 19 or 20 nt, for generation of guide RNAs (gRNAs). Exon 1 was targeted to introduce an early mutation as this increased the likelihood to completely disrupt the protein coding sequence. Target sites with a score of >70 and no off-targets were selected. Score values of >70 indicate highly efficient gRNA sites (Moreno-Mateos et al. 2015). Possible off-targets were also identified by performing BLAST searches in zebrafish genome. Two target sites were selected, one each in exon 1 and intron 1 (Table 6). The selected target sites were used to design standard gBlocks® (see Appendix 9) with upstream T7 promoter sequence and ordered at IDT (Singapore).

**Table 6. *srsf6b* CRISPR target sites**

Name	Target site
srsf6b_E1	GGCTTGCCGATGTACACCCG
srsf6b_I1	GGTGGACGGGAACGCGCGGTG

The gBlocks were first amplified using Takara PrimeStar MAX DNA polymerase and CRISPR RNAamp primer pairs (Table 3). 4-10 ng of gBlock was mixed with 25 µl of 2X PrimeStar enzyme mix and 1 µl each of forward and reverse primers. The final volume was adjusted to 50 µl with H<sub>2</sub>O, mixed and run with the following PCR conditions for 25 cycles: 98°C for 10 sec, 55°C for 5 sec and 72°C for 5 sec. The amplified gBlock DNAs were gel

extracted and sequenced with CRISPR RNAamp primers to check for any point mutation.

The gRNAs were transcribed from the respective gBlock using a MEGAscript<sup>TM</sup> T7 kit (Invitrogen). For setting up a reaction mix, 400 ng of gBlock was mixed 2  $\mu$ l of T7 10X Reaction buffer, 2  $\mu$ l each of T7 ATP, GTP, CTP and ATP solution, and 2  $\mu$ l of T7 Enzyme mix. The total volume of the reaction mix was adjusted to 20  $\mu$ l with H<sub>2</sub>O and mixed well. The reaction mix was incubated at 37°C for 5 hrs, after which 1  $\mu$ l of TURBO DNase was added, for digestion of gBlock DNA, and incubated for at least 45 min at 37°C. A Sodium acetate precipitation was used for clean-up and purification of the gRNAs. 115  $\mu$ l of fresh MQ H<sub>2</sub>O, 15  $\mu$ l of 3M Sodium acetate (pH 5.2) and 300  $\mu$ l of 100% EtOH was added to the reaction mix, mixed thoroughly and incubated overnight at -80°C. Then, the precipitated gRNAs were centrifuged at full speed for 45 mins at 4°C. The supernatant was discarded, and the pellet washed with 80% EtOH and centrifuged again at full speed for 30 mins at 4°C. The supernatant was carefully discarded and the RNA pellet air-dried before re-suspending it in 20  $\mu$ l of fresh MQ H<sub>2</sub>O. The quality and purity of the gRNAs was assessed by running an aliquot on a 1% agarose gel. Samples with single bands were selected for injection. The Nanodrop readings were recorded for concentration measurement. gRNAs were stored at -80°C.

### **2.16.2 Cas9 mRNA preparation**

The Cas9 mRNA was transcribed from the pCS2-nCas9n plasmid (Addgene plasmid # 47929) (Jao et al. 2013). This plasmid expressed a zebrafish codon-optimised Cas9 protein, having nuclear localisation signal sequence, under SP6 promoter (Jao et al. 2013). The plasmid was first digested with NotI (Fermentas). 5  $\mu$ g of pCS2-nCas9n plasmid was linearised with 3  $\mu$ l of

NotI enzyme (30 units) in a reaction volume of 100  $\mu$ l. The digestion mix was incubated 37°C overnight. Then, the entire digestion reaction was run on a gel. The linearised plasmid band was gel extracted and stored at -20°C.

For Cas9 mRNA preparation, a SP6 mMessage mMachine kit (Life Technologies) was used. A 20  $\mu$ l transcription reaction was set up with at least 1  $\mu$ g of linearised pCS2-nCas9n plasmid DNA, 10  $\mu$ l of 2X NTP/CAP, 2  $\mu$ l of 10X Reaction buffer and 2  $\mu$ l of SP6 Enzyme mix. The reaction mix was incubated at 37°C for 2 hrs followed by addition of 1  $\mu$ l of TURBO DNase and incubation at 37°C for another 30 mins. The Cas9 mRNA was purified using Sodium acetate precipitation. 115  $\mu$ l of fresh MQ H<sub>2</sub>O, 15  $\mu$ l of 3M Sodium acetate (pH 5.2) and 300  $\mu$ l of 100% EtOH was added to the reaction mix and mixed properly before incubating at -80°C overnight. Then, the solution was centrifuged at full speed for 45 mins at 4°C, and the supernatant was carefully discarded. The RNA pellet was washed with 80% EtOH and centrifuged again at full speed for 30 mins at 4°C. Finally, the supernatant was removed completely and the RNA pellet was air dried before re-suspending it in 20  $\mu$ l of fresh MQ H<sub>2</sub>O. The quality and purity of Cas9 mRNA was checked on a gel. The concentration was measured on a Nanodrop and the Cas9 mRNA was stored at -80°C.

### **2.16.3 Microinjection**

100 ng/ $\mu$ l of each gRNA and 300 ng/ $\mu$ l of Cas9 mRNA were mixed and injected into one-cell stage zebrafish embryos. The CRISPR mix was injected directly into the cytoplasm of the embryos. The injected embryos were raised at 28°C until the experimental stage was reached.

#### **2.16.4 Genomic DNA (gDNA) extraction**

For gDNA extraction, each injected embryo was lysed in 30 µl of 50 mM NaOH at 95°C for 5 mins. The solution was vortexed and spun down, repeating these steps three times. After the third round of 95°C incubation, 2 µl of Tris/HCl pH 8.2 was added for neutralization. The sample was mixed properly, vortexed briefly, spun down and incubated on ice for 3 mins. After this, the gDNA samples were stored at -20°C until further use.

For gDNA preparation from immune-stained embryo heads, a DNA lysis buffer (10 mM Tris/HCl pH 8.2, 50 mM KCl, 0.3% Tween20, 0.3% Nonidet 40; freshly added Proteinase K (18.6 mg/ml) at 1:50 dilution) was used. The stained embryos were decapitated in 100% Glycerol and heads were used for gDNA extraction. Before starting, the embryo heads were washed thoroughly with H<sub>2</sub>O for 2-3 times to completely remove the glycerol. 20 µl of DNA lysis buffer was used for each head sample. The samples were incubated at 55°C for 20 mins, vortexed thoroughly, spun down and incubated at 55°C overnight. On the following day, the gDNA samples were vortexed thoroughly again, spun down and incubated at 95°C for 15 mins for proteinase K inactivation. The samples were then stored at -20°C until further use.

#### **2.16.5 *srsf6b* mutant screening**

To screen CRISPR injected embryos for introduced mutations in *srsf6b*, gDNA samples were PCR amplified using primers flanking the two CRISPR target sites in the *srsf6b* gene (Table 3). 24 CRISPR injected embryos per batch of injection were screened. Wild-type embryo gDNA was used as negative control.

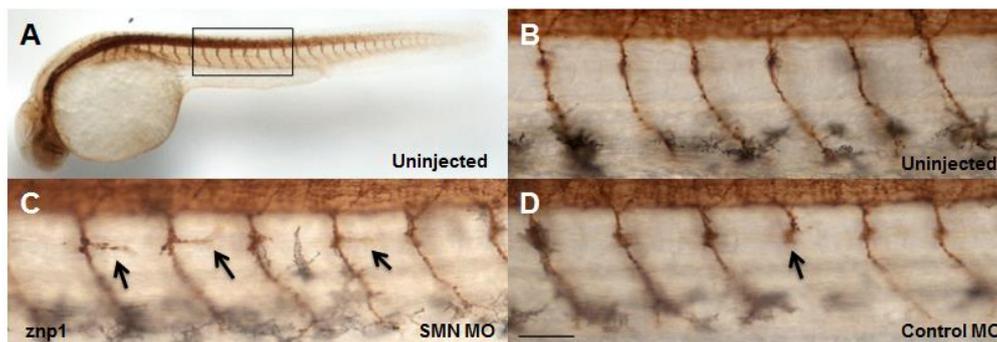
### 3. RESULTS

#### 3.1 A knockdown of *Smn* leads to motoaxonal defects in zebrafish embryos

Zebrafish is a well-established model organism for SMA and has been extensively used in study of the pathomechanism of this disease (McWhorter et al. 2003, Winkler et al. 2005, See et al. 2014). Earlier reports have shown that MO based *Smn* knockdown in zebrafish embryos leads to motoaxonal defects in caudal primary motor neurons (CaP) in the form of branching and truncation of axons (McWhorter et al. 2003, Winkler et al. 2005, See et al. 2014). Zebrafish CaP motor neurons have a very stereotypic growth pattern with the axons extending ventrally from the ventral root at around 18 hpf to the ventral edge of the axial muscles forming NMJs (Myers et al. 1986). By about 31 hpf all CaP motor neurons in the anterior spinal cord are fully developed with NMJs formed at the ventral muscles. These CaP motoaxons can be visualised by immunostaining with  $\alpha$ -znp1 antibody that labels synaptotagmin 2 on motoaxons.

To begin with, I repeated the *Smn* knockdown using MO and performed the immunostaining on *smn* morphants in order to replicate the motoaxonal defects shown in earlier reports. The SMN MO used was designed against the translational start site of the *smn* gene (Table 1) and was used at a concentration of 3.1 mg/ml as described previously (McWhorter et al. 2003, Winkler et al. 2005, See et al. 2014). To assess the specificity of knockdown of the *smn* gene, a standard control MO (Control MO; Table 1) was used. The Control MO targets an intron in the human  $\beta$ -globin gene and has no target in the zebrafish genome, thus it acts as negative control. I analysed a total of 120 motoaxons in 10 uninjected HB9-mCherry embryos (12 axons per

embryo) and the same axon number in 10 control embryos (Fig. 3B). In both control situations, I observed 3.3% (4 defective motoaxons out of 120) of motoaxons exhibiting axon outgrowth deficiencies such as ectopic branching or axon truncations. Consistent with previous reports (McWhorter et al. 2003, Winkler et al. 2005), it was found that upon *Smn* knockdown the number of defective axons was significantly increased to 12.5% (15 defective motoaxons out of 120) (Fig. 3C). This suggests that the motoaxonal defects seen in the *smn* morphants was due to the deficiency of *Smn* protein rather than a non-specific effect. These defects have been shown to be partially rescued by overexpressing full length *SMN* mRNA (See et al. 2014, Hao le et al. 2015), which further confirms that the defects were specific to *Smn* knockdown.



**Figure 3. Motoaxonal defects in zebrafish *smn* morphants.** Brightfield images of *znp1* immunostained HB9-mCherry embryos at 31 hpf. (A) Lateral view of HB9-mCherry embryo. Box indicates region above yolk extension where motoaxon morphology was assessed. (B) Higher magnification view of uninjected HB9-mCherry embryo. (C) Representative embryo after knockdown of *Smn* showing severe axon bifurcations (arrows). (D) Motoaxons in embryo injected with standard control Morpholino (MO). Scale bar = 50  $\mu$ m.

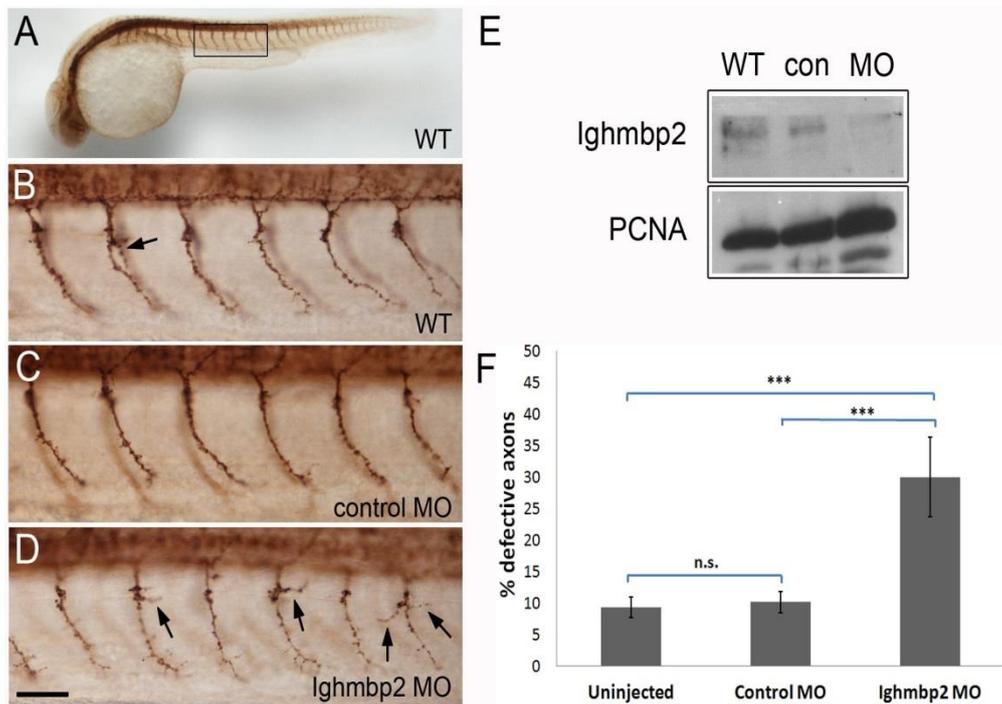
## 3.2 Generation of a zebrafish model for DSMA1

### 3.2.1 An IGHMBP2 knockdown leads to motoaxonal defects similar to SMA

Mouse studies have shown that mutations in IGHMBP2 lead to motoaxonal degeneration (Grohmann et al. 2004). Given the advantage of zebrafish for

live imaging of motor neurons, I wanted to know what happens to motoaxons under IGHMBP2 deficient conditions in zebrafish embryos. Antisense morpholino oligos were designed against the translation start site of *Ighmbp2a* (Table 1) and injected into zebrafish embryos at the 1-2 cell stage at 6.25 mg/ml concentration. Motor neuron morphology was assessed at 31 hpf using immunostaining with  $\alpha$ -znp1. In control embryos at 31 hpf, the axons of CaP motor neurons were fully extended and have reached their targets in the ventral myotome (Fig. 4A-C). Axon outgrowth was usually uninterrupted without obvious branching or truncation of axons, with a few rare exceptions (Fig. 4B,F; arrow). A total of 720 motoaxons were analysed in 60 uninjected wild-type embryos (12 axons per embryo) and the same axon number in 60 embryos injected with a Standard control morpholino (Fig. 4C). In both control situations, it was observed that 9.3% and 10.1%, respectively, of motoaxons exhibited axon outgrowth deficiencies such as ectopic branching or axon truncations. In contrast, the number of defective axons was significantly increased to 30.0% in *ighmbp2* morphants (Fig. 4D,F; 720 axons analysed in 60 embryos in six independent injection experiments). *In situ* hybridisation experiment with *myoD* riboprobe, that stains early muscle progenitors, revealed no muscle development defects in *ighmbp2* morphants (data not shown). The knock-down of endogenous *Ighmbp2* protein was confirmed by Western blot analysis (Fig. 4E). In mouse model for DSMA1, motor axon degeneration occurs very early and precedes fiber atrophy (Krieger et al. 2013). Reduced axon elongation and pathfinding has been observed in *Ighmbp2* deficient cultured motor neurons (Krieger et al, unpublished data). Taken together, the results, obtained in zebrafish after *ighmbp2* knockdown, were similar to what was observed in mouse. This opens the possibility to use the motor neuron FAC sorting protocols in the zebrafish model also for IGHMBP2 deficient embryos, to address the

pathomechanism underlying DSMA1. Also the motoaxonal defects (axon branching and truncation) observed in zebrafish *ighmbp2* morphants were similar to that seen in *smn* morphants (Fig. 3C) as well as in *nrxn2a* morphants (See et al. 2014). This opens up a new question: Why does a deficiency in three different proteins, all having independent yet mutually exclusive functions, manifest in a similar cell-type specific phenotypic defect. Comparison of differentially expressed genes in motor neurons of *smn* morphants as well as *ighmbp2* morphants using RNAseq would help to shed some light on this.



**Figure 4. Motoaxon outgrowth defects in Ighmbp2 deficient zebrafish embryos.** (A) Lateral view of zebrafish embryo at 31 hpf after immunostaining with znp1. Box indicates region above yolk extension where motoaxon morphology was assessed. 12 motoaxons (6 on either side of the body axis) were analysed per embryo. (B) Higher magnification view of uninjected wild-type (WT) embryo. Note bifurcation in one axon (arrow). (C) Motoaxons in embryo injected with Standard control Morpholino (MO). (D) Representative embryo after knock-down of Ighmbp2 showing severe axon bifurcations (arrows). (E) Western blot analysis of endogenous Ighmbp2 protein levels in uninjected (WT), control Morpholino injected (con; 6.25 mg/ml) and Ighmbp2 Morpholino (MO; 6.25 mg/ml) injected embryos. The equivalent of approximately 10 embryos was loaded per lane. PCNA (proliferating cell nuclear antigen) was analysed on the same filter as loading control. (E) Quantitative analysis of motoaxon defects. Percentage of defective axons in ten randomly picked embryos derived from six independent injection experiments (total number of axons analysed per sample: 720). Significant increase in frequency of motoaxon outgrowth defects (bifurcations, truncations) in Ighmbp2 deficient embryos ( $p = 0.0096$  for uninjected versus Ighmbp2 MO;  $p = 0.0123$  for control MO versus Ighmbp2 MO).

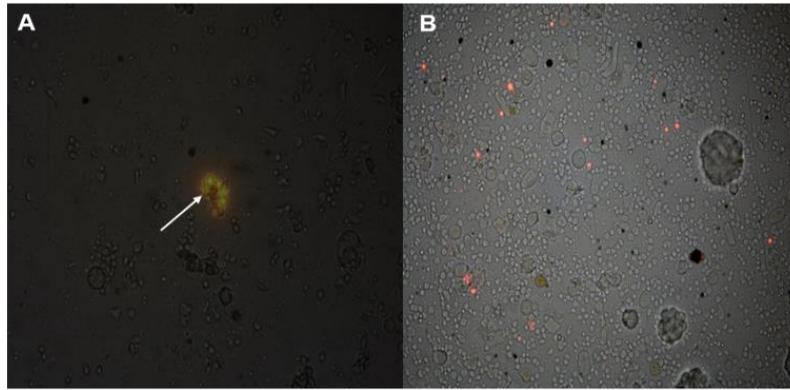
### **3.3 HB9:eGFP/mCherry embryo dissociation and FAC sorting of motor neurons**

#### **3.3.1 Optimization of embryo dissociation conditions and FAC sorting of motor neurons**

With the SMN MO knockdown working well in my hands, I next used whole embryo dissociation to obtain single cell suspensions. For establishing an optimised protocol, HB9-mCherry embryos were used for generating single cell suspensions that were subjected to FAC sorting. A 'Papain Dissociation System protocol' (Worthington Labs), similar to one reported by Cerda et al., was adopted for dissociating whole zebrafish embryos (Cerda et al. 2009).

Initially, I started dissociating the embryos according to the manufacturer's instructions with some modifications. 100-300 dechorionated HB9:mCherry embryos at 28 hpf were deyolked in Ringer's solution and incubated with 500 µl of Papain/DNase I solution at 30°C followed by mechanical dissociation. Cells were recovered in L-15 media containing 1% FBS and observed under the Fluorescence microscope. The critical step in this protocol was to optimize the papain incubation condition to obtain a maximum yield of live mCherry positive cells (motor neurons). Initially, 30 min papain incubation on a thermomixer at 700 rpm (as described in Cerda et al. 2008) was used, but these conditions were too mild to yield single mCherry positive cells. When observed under the fluorescence microscope, cell clumps and small tissue fragments were seen indicating incomplete dissociation (Fig. 5A).

Next, the incubation time was increased to 2 hrs with 1400 rpm shaking. These conditions yielded 12,500-23,400 single mCherry positive cells per 100 dissociated embryos (Fig. 5B) and the yield was 0.6-0.7% (Table 7).



**Figure 5. Hb9-mCherry positive cells after embryo dissociation.** (A) Fluorescence image after papain dissociation for 30 min. mCherry positive tissue fragments (arrow) were visible indicating incomplete dissociation. (B) Single mCherry positive cells obtained from 120 embryos after 2 hrs of papain incubation.

**Table 7. Number of cells obtained after embryo dissociation with 2 hrs of papain incubation**

Batch	No. of embryo	Total no. of cells obtained (500 µl)	Total mCherry positive cells obtained (500 µl)	% Yield	Total cells per 100 embryo	Total mCherry positive cells per 100 embryo
1	120	1.9 million	15,000	0.76	1.6 million	12,500
2	105	3.2 million	24,588	0.76	3.0 million	23,416
3	318	9.4million	58,300	0.62	2.9 million	18,333

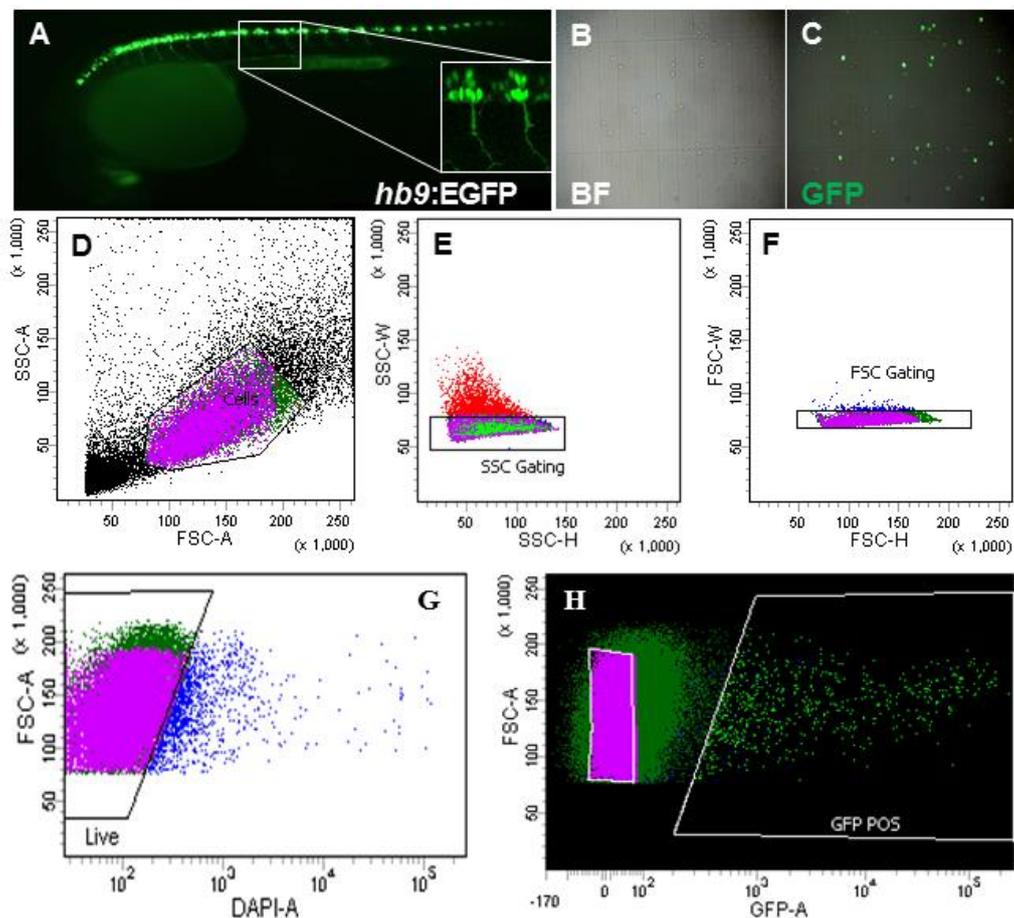
Next, it was necessary to assess the condition of these single cells, i.e. whether they were still alive after being dissociated from whole embryos. For this, an initial round of FAC sorting was employed. The FACS instrument can quantify the exact percentage of live and dead cells in the sample. HB9:eGFP were used for FACS optimisation and all subsequent experiments.

The sorting was done using a BD FACSAria Cell sorter available at the Biopolis Shared Facility. The single cell suspensions after dissociation were stained with DAPI for labeling dead cells and introduced into the FACS

machine where they were sorted using different parameters. Firstly, the total sample was analysed to separate a homogeneous cell population from cell aggregates and cell debris on the basis of size (Forward Scattering, FSC) and granularity (Side Scattering, SSC) (Fig. 6D). 60% of cells were gated at this step while rest was discarded as debris or cell aggregates. This population was then further scrutinized to obtain single cells and eliminating doublets using SSC and FSC gating (Fig. 6E,F). More than 90% of cells were selected for the next step. These defined single cells were subjected to live/dead screening by analysing DAPI fluorescence. DAPI does not penetrate cell membranes well and therefore live cells do not take it up efficiently. In contrast, DAPI enters dead cells readily, thus labeling them (Fig. 6G). More than 99% of the gated cells were found to be alive. Finally, after removing DAPI positive dead cells, live single cells were screened for specific fluorescence. Cells with high fluorescence intensity were then sorted from the rest of the weak/non-fluorescent cells (Fig. 6H). The sorted cells were collected in L-15 medium containing 20% FBS. Wild-type embryos as well as non-transgenic sibling embryos were used to optimize the parameters for FAC sorting.

Using these reference settings with minor adjustments, I was successful in identifying and collecting motor neuron cell populations from dissociated cells obtained from whole embryos after 2 hrs of Papain incubation. Unfortunately, the yield of mCherry positive cells was low in the initial attempts (approx. 0.3% of total live cells). This indicated that the selected embryo dissociation conditions were too harsh, thus destroying the cells. To improve the number of sorted cells, the dissociation conditions were made milder by reducing papain incubation time to 45 min and shaking to 1200 rpm. The cell survival increased and the yield of cells also increased from 0.3% to 1% yielding a

significantly higher number of mCherry positive cells. Also with reduction in papain incubation time, total duration of the whole procedure was reduced to less than 2 hrs. This is important as a shorter procedure ensures that the transcriptome profile of sorted cells does not change much and RNAseq results are therefore more accurate. On average, 35,000 GFP positive motor neurons were sorted out from 290 HB9:eGFP embryos. The sorted GFP positive motor neurons were collected in 20% FBS/L-15 medium, pelleted down, re-suspended in TRIZOL and stored at -80°C.



**Figure 6. FAC sorting of eGFP positive motor neuron cells.** (A) HB9:eGFP embryo at 30 hpf. Brightfield (B) and fluorescent image (C) of sorted GFP positive cells. (D) Total cells were presented in a FSC-A vs. SSC-A plot to screen for homogenous cell populations (enclosed region) without cell debris and cell aggregates (everything except enclosed region). (E and F) The cell population was then checked again on SSC and FSC plots to remove any doublets and for selection of only single-cell populations (SSC and FSC gating). (G) These populations were checked for live cells (enclosed region; DAPI negative) among the entire single cell population. Cells on right side of

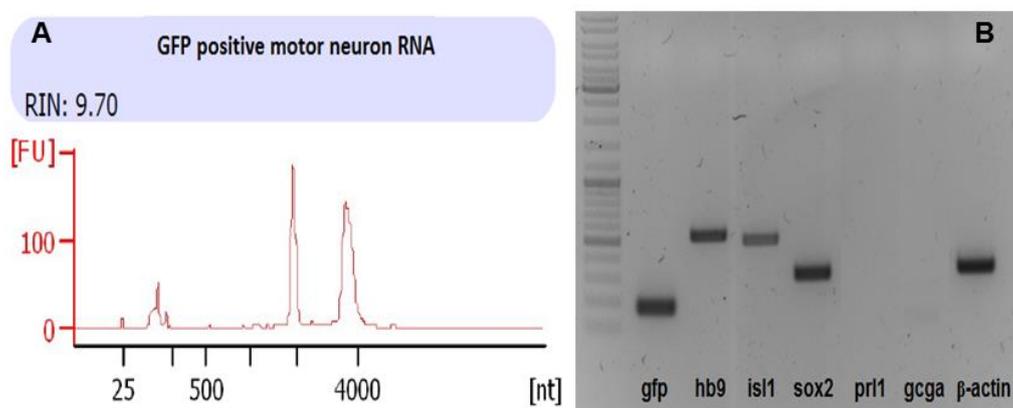
plot were dead (DAPI positive). (H) Finally live single cells were sorted on the basis of eGFP fluorescence. The GFP POS gated population represents live eGFP positive cells that were finally collected in L-15 media containing 0.5% FBS.

### 3.3.2 Total RNA extraction from sorted motor neurons

The cell samples frozen in TRIZOL were used for extracting total RNA. For obtaining RNA samples with optimal quantity and high quality from the FAC sorted motor neuron cells, a phenol-chloroform based extraction method followed by column purification was used. RNA extraction from very low cell numbers (less than 5,000) usually gave low quality RNA. To improve quality and quantity of RNA, the TRIZOL frozen cells were pooled together to achieve more than 15,000 cells. With the optimised RNA extraction protocol, 28 ng of total RNA was purified from 35,000 GFP positive motor neurons. The quality of the purified RNA samples was assessed using RNA 6000 Pico kit and Agilent 2100 bioanalyser. The Agilent bioanalyser gives an RNA Integrity Number (RIN) as a read out for RNA quality. A RIN value of 1 represents degraded RNA while value of 10 indicates an intact RNA. All of the RNA samples generated had RIN values >9 which was suitable for the downstream RNA library preparation. Fig. 7A shows an Agilent bioanalyser profile for one of the GFP positive motor neuron RNA samples with a RIN of 9.7. RNA samples were stored at -80°C until the start of the library preparation.

To confirm the identity of the sorted cells, a RT-PCR was performed on RNA obtained from sorted HB9:eGFP positive motor neurons. Primers were designed for the following specific lineage markers: *hb9*, *gfp*, *isl1*, *sox2* (for motor neurons); *prl1* (pituitary gland); and *gcga* (pancreas). The RT-PCR results showed strong expression of *hb9* and *isl1*. A strong *gfp* band indicated

presence of motor neurons in the sorted cells (Fig. 7B). *Sox2* is expressed in motor neuron progenitors until they exit the cell cycle, switch off *sox2*, turn on *hb9* expression and differentiate into mature motor neurons (Briscoe and Novitch 2008). A considerable *sox2* band can be explained by the presence of HB9:eGFP positive motor neurons with residual *sox2* transcripts (Fig. 7B). Absence of bands for *prl1* and *gcga* indicates absence of pituitary and pancreatic cell types in the sorted cell population, respectively (Fig. 7B). Taken together, the RT-PCR results suggest an enrichment of motor neurons in the sorted cells after FACS.



**Figure 7. Total RNA extracted from FAC sorted motor neurons.** (A) Agilent bioanalyser profile for a GFP positive motor neuron RNA sample with a RIN of 9.7. (B) RT-PCR gel image showing strong expression of motor neuron markers (*hb9*, *isl1*, *sox2* and *gfp*) but no expression of non-motor neuron lineage markers (*prl1* and *gcga*) suggesting enrichment of motor neuron populations after FAC sorting.  $\beta$ -actin was used as loading control.

### 3.4 Preparation of cDNA libraries from Smn-deficient motor neurons

After having optimised the protocol for embryo dissociation, FAC sorting and RNA extraction for control embryos, the next step was to use these conditions for sorting out GFP positive motor neurons from Smn deficient zebrafish embryos (hereafter named 'SMA MN' for simplicity). SMN has a well-

established role in the assembly of U snRNPs and is thus important for formation of spliceosome (Pellizzoni et al. 2002, Gubitz et al. 2004). It has been proposed that alternative or aberrant splicing of transcripts occurs in motor neurons under SMN deficient conditions (Zhang et al. 2008). To identify alternatively or aberrantly spliced transcripts, RNAseq was performed on RNA from motor neurons under *Smn* deficient conditions. HB9:eGFP transgenic embryos were injected with SMN MO and Control MO. Embryos were dissociated at 28 hpf and subjected to FAC sorting to obtain motor neurons. Total RNA was extracted from the sorted motor neurons and high quality RNA samples (RIN >9.5) were then used for preparation of cDNA libraries using the SMARTer Ultra Low Input RNA for Illumina sequencing-HV (Clontech Laboratories) and NEBNext Ultra Library Prep kit for Illumina (New England BioLabs).

#### **3.4.1 Synthesis of double stranded cDNA from total RNA derived from *Smn*-deficient motor neurons**

The first step in library preparation was synthesis of cDNA from total RNA. For cDNA synthesis, the SMARTer Ultra Low Input RNA for Illumina sequencing-HV system from Clontech Laboratories was used. This kit has been previously used for preparation of high quality cDNA samples for RNAseq from very low input RNA (Head et al. 2014, Shanker et al. 2015) as well as single neurons (Qiu et al. 2012). For RNAseq library preparation, 1-10 ng of RNA samples with high quality (RIN >9.5), as assessed from Agilent bioanalyzer, were used (Fig. 8A,B). A total of three sets of the motor neuron sample from *smn* morphants (SMA MN) and control (Control MN) embryos were used for RNAseq library generation. Table 8 lists the details of all samples that were used for preparing SMA motor neuron RNAseq libraries.

RNA samples were subjected to cDNA synthesis followed by amplification using long-distance PCR. The PCR cycles were increased from the recommended 12 to 13 cycles for Sample 1 (both SMA MN and Control MN), and to 16 PCR cycles for Sample set 2 and 3. Positive and negative control reactions were also set up to check whether the kit was working properly as well as to exclude any contamination in the reaction mixture. Next, the amplified double stranded cDNA samples were assessed for quality using HS DNA kit and Agilent bioanalyser (Fig. 8C,D). Finally, the cDNA samples were sheared into 200-500 bp fragments using Covaris AFA system which uses a focused ultrasonic acoustic energy. The fragmented cDNA samples were then used for library preparation.

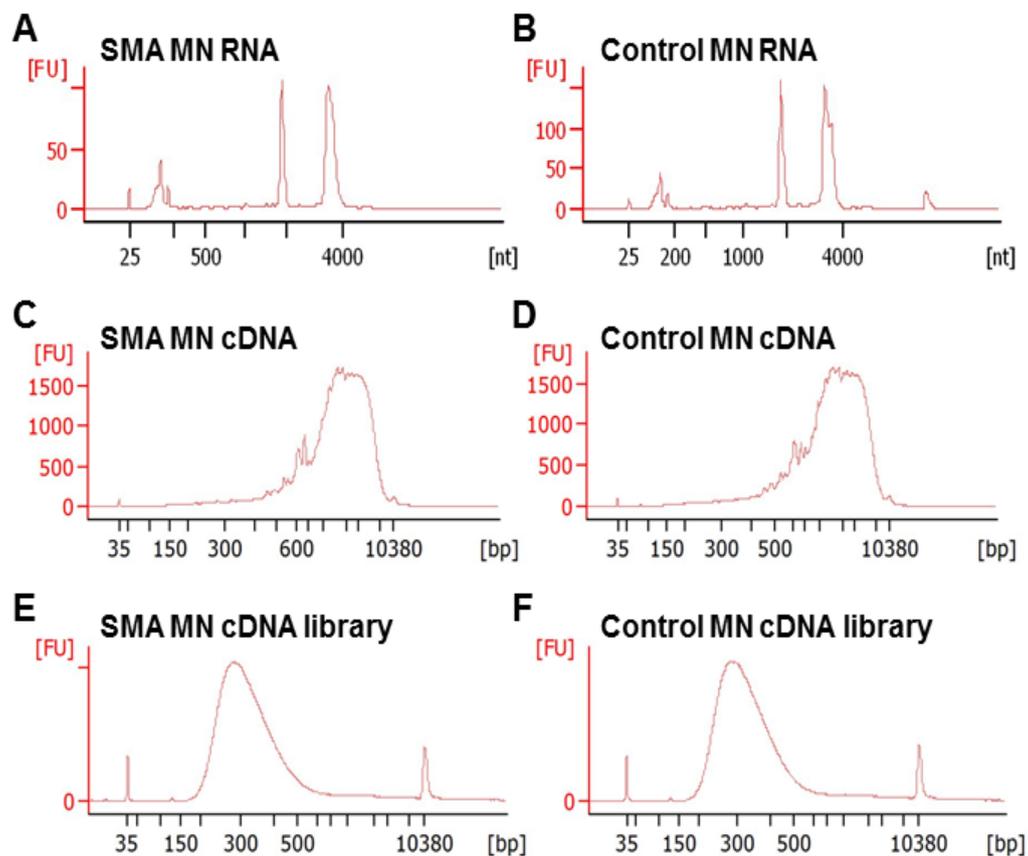
**Table 8. List of samples used for SMA MN RNAseq**

Samples		No. of embryos	Total cells	RNA yield (ng)	RIN	RNA input for SMARTer kit (ng)	DNA input for NEBNext kit (ng)
Control MN	Sample 1	254	36000	27.4	9.5	6.81	4.8
	Sample 2	405	37370	29.7	9.7	2.7	145.41
	Sample 3	206	31350	27.6	9.5	2.7	162.61
SMA MN	Sample 1	259	18000	10.2	9.5	7.65	5.15
	Sample 2	378	15600	11.4	9.6	2.7	137.64
	Sample 3	191	15200	19.9	9.5	2.7	157.06

### 3.4.2 Preparation of cDNA libraries

NEBNext Ultra Library Prep kit for Illumina (New England BioLabs) was used for preparing libraries from fragmented cDNA samples. cDNA samples were subjected to blunt-end repair and dA-tailing followed by ligation of adaptors which contain annealing sites for index primers and the universal primer. Index primers 6 and 12 were used for multiplexing two libraries while primers

2, 6, 7 and 12 were used for multiplexing four libraries in one lane. The number of PCR cycles for amplification step was optimised according to different input cDNA amounts. Finally, the libraries were cleaned up using the AMPure beads followed by assessment of quality of each library using HS DNA kit and Agilent bioanalyser (Fig. 8E,F). Successful cDNA library samples were multiplexed and submitted for RNAs sequencing



**Figure 8. Assessment of quality for extracted RNA, amplified cDNA and final library during SMA MN library preparation using Agilent 2100 bioanalyser.** Each step in the protocol was checked by running 1  $\mu$ l sample on a bioanalyser. The Y-axis represents the fluorescence intensity detected (FU). The X-axis represents fragment size of RNA or DNA in nucleotides (nt) or base pair (bp), respectively. Representative profiles for one SMA MN sample and one Control MN sample are shown. (A,B) Electropherograms of input RNA for SMA and control MN sample used for RNAseq library preparation. RIN value for both samples was 9.5. (C,D) Electropherograms of amplified cDNA obtained after first strand cDNA synthesis and amplification during SMARTer protocol. (E,F) Electropherograms of cDNA libraries after completion of NEBNext library prep protocol, with lengths of all the fragmented cDNA between 200-500 bp.

### **3.4.3 RNA sequencing of Smn-deficient motor neurons**

All individual libraries generated were multiplexed into one lane to achieve 100-200 million sequencing reads and were sequenced using the paired-end sequencing method on an Illumina HiSeq platform. Libraries were quantified using qPCR and multiplexed before submitting for sequencing at the Genome Institute of Singapore (GIS). Performing qPCR was necessary for achieving accurate normalisation and clustering, which was not possible using the Qubit assay or bioanalyser (Vuyisich et al. 2014). A total of three sets of motor neuron libraries were generated from Smn deficient and control embryos. Ideally, a single lane in Illumina HiSeq platform can generate more than 400 million paired-end reads. The number of samples/libraries to be multiplexed was dependent on how much depth/number of reads was required for each library. For identifying the changes in gene expression profile, low depth (approx. 30-50 million reads) is sufficient. In my case where the aim was to identify alternative splice transcripts or for identification of novel genes, very deep sequencing was required (more than 100 million reads). The Sample 1 library for both SMA MN and Control MN was run on a single lane on the Illumina HiSeq platform, while library sets for Samples 2 and 3 were multiplexed and run on one lane. The sequencing was performed using the paired-end method, where each cDNA fragment was sequenced from both directions (forward and reverse), with read lengths of 75-100 bp. The data generated using this method was suitable for identification of alternatively spliced transcripts expressed at very low levels. Typically, 40 million mapped reads provides reliable measurement of a single transcript per cell (Mortazavi et al. 2008). The RNAseq was performed with a Next Generation Sequencing platform at the Genome Institute of Singapore in collaboration with Dr. S. Mathavan.

### 3.4.4 RNA sequencing of FAC sorted DSMA1 motor neurons

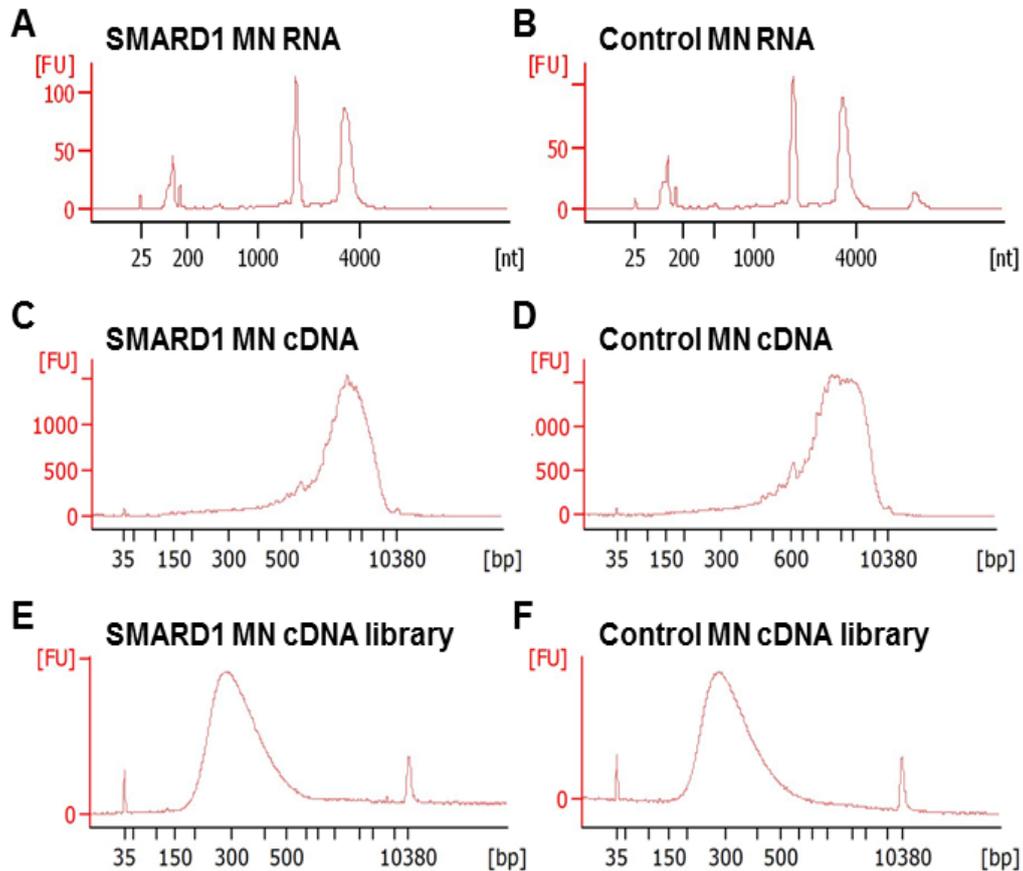
An *ighmbp2* knockdown leads to motoaxonal defects in zebrafish embryos (Fig. 4D), which were similar to those observed after *smn* knockdown (Fig. 3C) (McWhorter et al. 2003, Winkler et al. 2005, See et al. 2014) and *nrxn2a* knockdown (See et al. 2014). In order to understand what leads to such defects, I generated RNAseq data for FAC sorted motor neurons from IGHMBP2 deficient and control embryos. Using the optimised protocols for embryo dissociation, FAC sorting and RNA extraction protocols, high quality total RNA (RIN >9) was extracted from FAC sorted motor neurons under IGHMBP2 deficient and control conditions.

**Table 9. List of samples used for DSMA1 MN RNAseq**

Sample	No. of embryos	Total cells	RNA yield (ng)	RIN	RNA input for SMARTer kit (ng)	DNA input for NEBNext kit (ng)
<b>DSMA1 MN Sample 1</b>	179	15,720	33.5	9.4	2.7	207.01
<b>Control MN Sample 1</b>	142	21,760	33.3	9.3	2.7	311.91

Only one set of motor neuron library was generated from IGHMBP2 MO injected (DSMA1 MN) and control (Control MN) embryos. Details for the samples used for RNAseq is listed are Table 3. In this case, 2.7 ng of input RNA (Fig. 9A,B) was used as input with the SMARTer kit for preparing cDNA and later it was amplified using LD-PCR with 16 cycles. The resulting amplified cDNA was checked for quality using Agilent bioanalyser (Fig. 9C,D) and subjected to Covaris shearing. The sheared cDNA samples were then quantified using Qubit HS DNA assay and used for preparing libraries using NEBNext kit with the PCR cycles adjusted according to the sheared cDNA input (Table 9). Index Primers 7 and 12 were used for barcoding the libraries.

The libraries were assessed for quality using an Agilent bioanalyser (Fig. 9E,F) and multiplexed together on the basis of qPCR results. The libraries were finally run on Illumina HiSeq platform using paired-end sequencing method with 100 bp read length.



**Figure 9. Assessment of quality for extracted RNA, amplified cDNA and final library during DSMA1 motor neuron library preparation using Agilent bioanalyser.** Each step in the protocol was checked by running 1 $\mu$ l sample on bioanalyser. The Y-axis represents the fluorescence intensity detected (FU). The X-axis represents the fragment size of RNA or DNA in nucleotides (nt) or base pair (bp), respectively. (A,B) Electropherograms of input RNA for DSMA1 and control MN samples used for RNAseq library preparation with RIN values of 9.4 and 9.3, respectively. (C,D) Electropherograms of amplified cDNA obtained after first strand cDNA synthesis and amplification during SMARTer protocol. (E,F) Electropherograms of cDNA libraries after completion of NEBNext library prep protocol, with lengths of all the fragmented cDNA between 200-500 bp.

### 3.5 Transcriptome analysis identifies differentially expressed genes in Smn deficient motor neurons

After the completion of sequencing run, the multiplexed libraries were demultiplexed to sort the reads according to their barcodes (index primers) and finally the data for individual libraries were downloaded from the GIS server in fastq formats. Sample 1 motor neuron library set which was sequenced alone produced more than 200 million reads each (Table 10). Sample 2 and 3 library set were multiplexed together thus yielding approx. 100 million reads (Table 10) except for Sample 3 Control MN library (78 million reads). This difference in the no. of reads obtained could be due to improper pooling of the four libraries (Sample 2 and 3 library set) before submitting for the sequencing run. However, the overall quality of sequencing, as assessed by FastQC, was very high in that the average base Phred quality score (Q) was around 35 (Table 10). As a point of reference, a value of 30 for Q represents base call accuracy of 99.9%. The % of  $\geq$ Q30 bases (percentage of base calls with greater than 99.9 % accuracy) was higher than 90% for all the libraries (Table 10).

**Table 10. Illumina HiSeq sequencing results for SMA motor neuron libraries**

Sample	Yield (Mbases)	# Reads	% of $\geq$ Q30 Bases	Mean Quality Score
SMA MN Sample 1	14,676	216,714,968	90.10	35.20
Control MN Sample 1	14,784	218,480,980	90.27	35.29
SMA MN Sample 2	10,164	108,355,474	90.12	35.02
Control MN Sample 2	10,014	106,870,888	90.33	35.1
SMA MN Sample 3	9,984	106,732,056	90.65	35.21
Control MN Sample 3	7,411	78,892,948	90.64	35.2

The raw RNAseq data (fastq files) was then used for bioinformatics analysis, which was done in collaboration with Dr. Candida Vaz, Dr. Vivek Tanavde and Dr. Brian Parker at the Bioinformatics Institute Singapore (BII).

### 3.5.1 Mapping of RNAseq data to zebrafish genome

The raw fastq files contain sequences of all the reads obtained for a particular library. These read sequences were comprised of actual cDNA sequence plus the adaptor sequence which was added during the library preparation step and thus needs to be removed before mapping them to the zebrafish reference genome. For trimming off the adaptor sequence, CutAdapt tool was used (See Appendix 2a). This tool identifies any particular sequence in the reads and removes it. It can be used to remove poly-A tails, primer, or in case of RNAseq data, adaptor sequences. The trimmed read sequences were then mapped to zebrafish reference genome, Zv9, using the TopHat version 2.0.12 (see Appendix 2b).

**Table 11. SMA MN RNAseq reads mapped to zebrafish reference genome Zv9 using TopHat**

	Control MN			SMA MN		
	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3
<b>Total Reads</b>	218,480,980	106,870,888	78,892,948	216,714,968	108,355,474	106,732,056
<b>Reads Mapped</b>	83.2%	72.6%	73.7%	81.9%	71.4%	73.5%
<b>Total Paired Reads</b>	75,375,208	29,919,473	22,498,973	74,202,859	29,763,979	30,416,870
<b>Concordant Pairs Mapped</b>	74.6%	58.5%	59.1%	74.1%	57.3%	59.6%

The total no. of reads per sample were between 78,892,948 and 218,480,980 (Table 11); between 71.4% and 83.2% of the reads were uniquely mapped to the zebrafish reference genome. The no. of aligned pairs per sample were between 22,498,973 and 75,375,208 (Table 11), out of which concordant pair

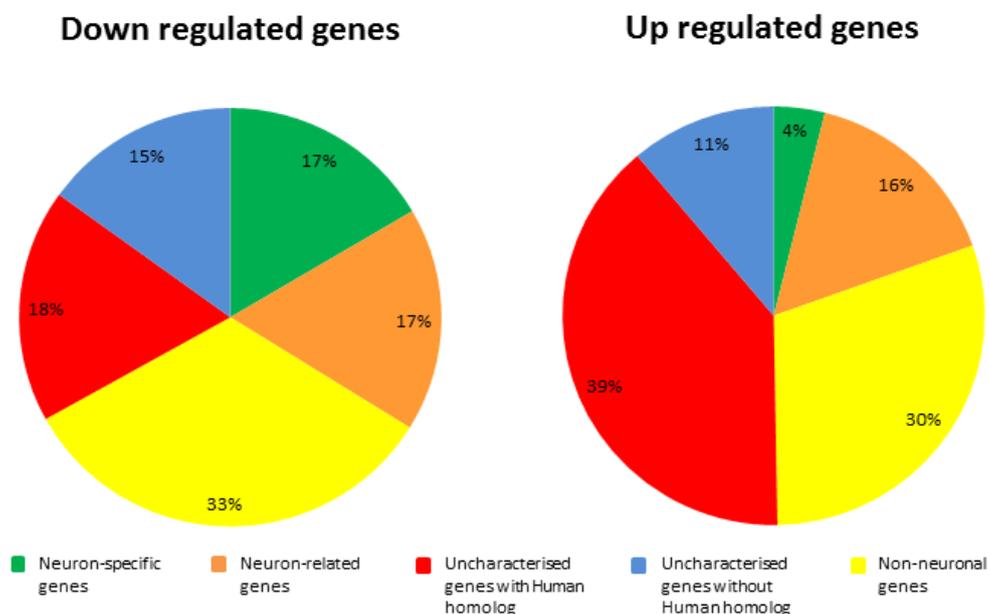
reads were between 57.3% and 74.6% uniquely mapped to the zebrafish reference genome (Table 11).

### **3.5.2 Differential gene expression analysis of Smn deficient motor neurons identifies deregulated transcripts**

After running the TopHat tool for mapping of raw RNAseq data, bam files were generated as an output which contains sequence alignment data for all the mapped reads. This was the main file for all the downstream RNAseq data analysis. The bam files for all six libraries were used for differential gene expression analysis using Partek Genomics Suite 6.6 (Partek Inc.). This is one of the most commonly used software for NGS and Microarray data analysis. All the bam files were imported into the software and the mRNA quantification was done to identify all the transcripts that were differentially expressed across the samples. The normalisation method used by Partek Genomics Suite was Reads Per Kilobase of exon model per Million mapped reads or RPKM, which represents the expression of particular gene normalised to gene length and the total no. of mapped reads obtained in the RNAseq experiment. These RPKM values can then be compared across different experimental conditions to determine the changes in regulation for each gene.

A total of 13,923 expressed genes were identified in motor neurons at 28 hpf out of which 312 genes were found to be differentially expressed in SMA MN (Fold change >2 and p-value <0.05). Out of 312 differentially expressed genes, 133 were down-regulated in SMA MN as compared to Control MN while 179 were up-regulated. A list of all differentially expressed genes is provided in Appendix 3. I categorised the differentially expressed genes based on their available expression data in the ZFIN database. The genes were categorised into five groups (Fig. 10): (1) Neuron-specific genes,

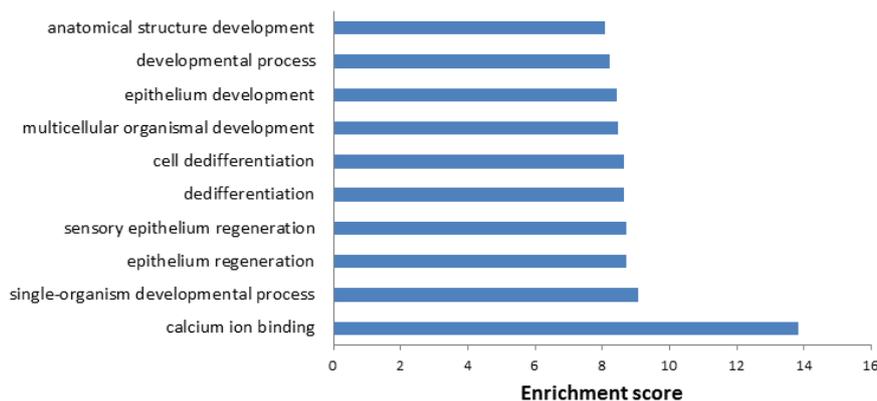
expressed exclusively in zebrafish neurons; (2) Neuron-related genes, which not specific to neurons but also expressed in other tissues of zebrafish; (3) Non-neuronal genes, which are not reported to be expressed in neurons but found in other non-neuronal tissues of zebrafish; (4) Uncharacterised genes with known human homologs but with no known zebrafish function; (5) Uncharacterised genes without human homolog thus representing novel zebrafish-specific genes. Interestingly, 33% of the down-regulated and 50% of the up-regulated genes were found to be uncharacterised in zebrafish, among which a majority of transcripts have known human homologs (Fig 10). In case of down-regulated genes, 34% have a neuronal expression while the rest (33%) has no reported neuronal expression. On the other hand, only 20% of the up-regulated genes have neuronal expression and 30% were without reported neuronal expression.



**Figure 10. Summary of transcriptomic changes identified in SMA motor neurons by RNAseq analysis.** Pie-charts showing the genes identified in down-regulated and up-regulated gene lists, classified by the expression data from the ZFIN database.

### 3.5.3 Gene Ontology (GO) enrichment using Partek

To determine the biological relevance of the differentially expressed genes under Smn knockdown conditions, Gene set analysis feature in Partek genomics suite was used. Gene Ontology (GO) analysis revealed a list of GO terms that were over represented in the list of differentially expressed genes. This would consequently suggest a particular biological process or function that was being affected because of Smn knockdown in motor neurons. A total of 166 GO terms were found to be over represented with a p-value cut-off of 0.05. The top ten most enriched terms are shown in Fig. 11. Many of these terms were related to developmental processes suggesting that a Smn deficiency in motor neurons affects genes related to neuronal differentiation.



**Figure 11. Gene Ontology (GO) analysis of differentially expressed genes between Smn deficient and control motor neurons.** GO analysis was performed on differentially expressed genes with at least a 2-fold change and p-values less than or equal to 0.05. Top 10 GO terms enriched are represented in a bar graph with the Y-axis showing the enriched GO terms and the X-axis showing the respective enrichment score values.

### 3.5.4 Identification of differentially expressed genes in DSMA1 MN

The RNAseq data for DSMA1 MN and Control MN libraries were analysed in a similar fashion as mentioned above for SMA MN. In this case, only one set each of DSMA1 MN and Control MN libraries was generated and sequenced. 95,716,496 and 103,384,698 reads were obtained from sequencing of Control

MN and DSMA1 MN libraries, respectively, with high sequencing quality (Table 12). The Mean Phred quality score (Q) was 35 for both the samples with more than 90% of base calling at 99.9% accuracy. After the adaptor trimming of the raw RNAseq data using the CutAdapt tool, the trimmed reads were mapped to Zv9 zebrafish reference genome using TopHat. 73.1% and 68.4% of total reads in Control MN and DSMA1 MN data respectively, were mapped uniquely to the zebrafish reference genome (Table 12).

**Table 12. DSMA1 MN RNAseq results**

	<b>Control MN</b>	<b>SAMRD1 MN</b>
	<b>Sample 1</b>	<b>Sample 1</b>
<b>Total Reads</b>	9,57,16,496	10,33,84,698
<b>Reads Mapped</b>	73.1%	68.4%
<b>Total Paired Reads</b>	2,62,94,281	2,73,35,743
<b>Concordant Pairs Mapped</b>	39.1%	56%

Using Partek Genomics Suite, 13,487 expressed genes were identified out of which 1,666 were differentially expressed in DSMA1 MN (Fold change >2). In this case, p-values were not obtained as the dataset was n=1, so only a fold change cut-off was used. Out of 1,666 differentially expressed genes, 740 were down-regulated while 926 were up-regulated.

In order to analyse genes that were differentially regulated in both the SMA MN and DSMA1 MN, I compared the two differential gene expression lists. 93 genes were found to be overlapping among the two lists out of which 21 genes had opposite regulation pattern in both lists (Table 13A), 19 genes were down-regulated in both lists (Table 13B), and 53 were up-regulated in both lists (Table 13C). GO enrichment analysis was performed to determine the biological relevance of these overlapping genes. A total of 197 functional terms were identified to be significantly over represented (p-value <0.05) among the overlapping genes. Some of the top terms were related to

regulation of cellular component movement, dedifferentiation, regulation of axon guidance and axonogenesis etc. This suggests that genes involved in axonogenesis or axon guidance were commonly affected both in SMA or DSMA1 zebrafish larvae.

**Table 13. List of differentially expressed genes that overlap between DSMA1 MN and SMA MN**

<b>A. Genes with opposite regulation trend in both lists</b>			<b>B. Genes down-regulated in both lists</b>		
Gene names	Fold Change		Gene names	Fold Change	
	SMARD1 MN	SMA MN		SMARD1 MN	SMA MN
<i>urp2</i>	-10.15	2.75	<i>gnb3a</i>	-31.97	-4.55
<i>zgc:162324</i>	-4.31	4.66	<i>zp2l2</i>	-8.12	-14.27
<i>si:dkey-169i5.4</i>	-3.55	2.56	<i>si:dkey-11p23.7</i>	-5.07	-11.99
<i>tldr6</i>	-3.33	2.46	<i>zgc:195356</i>	-4.06	-9.99
<i>si:dkey-188i13.10</i>	-3.04	2.91	<i>ttni2a.1</i>	-3.04	-21.09
<i>lmcd1</i>	-2.93	2.09	<i>ahr1b</i>	-2.88	-2.41
<i>zgc:193681</i>	-2.88	2.02	<i>si:ch211-251b21.1</i>	-2.75	-2.34
<i>slc27a6</i>	-2.68	2.08	<i>arl3l2</i>	-2.71	-3.22
<i>plxnb2b</i>	-2.54	3.95	<i>glis3</i>	-2.71	-16.43
<i>hcst</i>	-2.37	2.66	<i>npas4a</i>	-2.56	-2.59
<i>rad21l1</i>	-2.37	12.33	<i>sst1.1</i>	-2.35	-2.37
<i>serpinb1</i>	-2.25	2.10	<i>cdnf</i>	-2.25	-4.57
<i>zgc:195173</i>	-2.16	2.94	<i>rtn4rl2a</i>	-2.24	-5.24
<i>acvr1l</i>	-2.03	2.06	<i>rx2</i>	-2.20	-3.59
<i>si:ch211-14a17.7</i>	-2.03	4.10	<i>vax1</i>	-2.14	-2.21
<i>zgc:172079</i>	2.08	-2.26	<i>smyd1a</i>	-2.08	-3.49
<i>apobec2a</i>	2.08	-2.55	<i>camk2a</i>	-2.03	-4.88
<i>adh8a</i>	2.96	-3.73	<i>ctssa</i>	-2.03	-2.85
<i>nitr3c</i>	3.94	-6.22	<i>zgc:66455</i>	-2.00	-3.60
<i>olfm2b</i>	4.93	-2.67			
<i>ugt2a1</i>	13.83	-4.48			

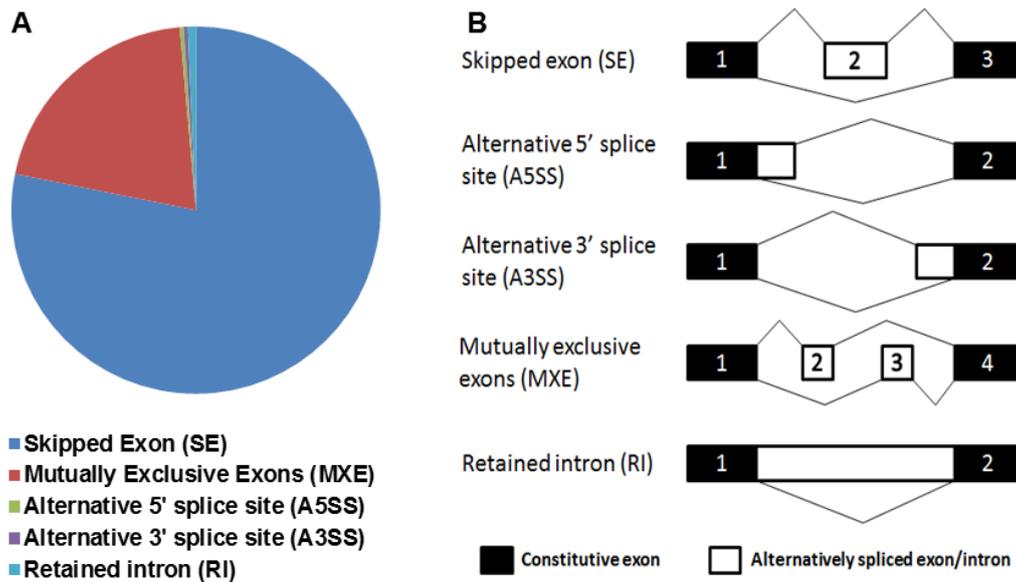
  

<b>C. Genes up-regulated in both lists</b>					
Gene names	Fold Change		Gene names	Fold Change	
	SMARD1 MN	SMA MN		SMARD1 MN	SMA MN
<i>tspan15</i>	2.01	2.80	<i>LOC100536659</i>	3.16	3.35
<i>hsd3b7</i>	2.02	2.03	<i>fbxo32</i>	3.16	3.11
<i>tpa</i>	2.03	2.22	<i>cdh15</i>	3.22	4.42
<i>tnfaip2a</i>	2.03	2.04	<i>myl4</i>	3.25	2.34
<i>si:dkey-261j4.5</i>	2.08	3.10	<i>zgc:113363</i>	3.36	3.21
<i>eva1bb</i>	2.09	4.78	<i>si:ch211-222k6.3</i>	3.38	2.76
<i>ef2ak2</i>	2.13	6.70	<i>zgc:101663</i>	3.40	2.56
<i>abhd6b</i>	2.14	2.36	<i>zp3a.2</i>	3.52	2.42
<i>fam129aa</i>	2.15	3.79	<i>dkk1b</i>	3.55	4.67
<i>rbm24a</i>	2.15	2.42	<i>ggh</i>	3.63	2.61
<i>mmp9</i>	2.21	2.42	<i>lingo3a</i>	3.78	2.82
<i>hmox1a</i>	2.21	3.01	<i>apoeb</i>	3.84	4.38
<i>zgc:113337</i>	2.34	3.05	<i>ccdc106b</i>	3.94	4.27
<i>si:ch211-237l4.6</i>	2.37	2.14	<i>si:dkey-222p3.1</i>	3.94	3.02
<i>slc9a2</i>	2.46	6.25	<i>sepw2b</i>	4.02	2.74
<i>gata1a</i>	2.56	2.09	<i>jam2a</i>	4.59	2.74
<i>kcnc4</i>	2.56	2.45	<i>dmrt2a</i>	4.70	2.60
<i>cyp2k16</i>	2.58	6.57	<i>gnmt</i>	4.86	2.36
<i>efemp2b</i>	2.66	2.84	<i>slc5a2</i>	4.93	3.76
<i>tagln2</i>	2.69	2.45	<i>tmed1a</i>	5.27	2.44
<i>si:ch211-81a5.5</i>	2.71	3.55	<i>chrng</i>	5.62	2.75
<i>csf2rb</i>	2.74	2.21	<i>apoc1l</i>	5.83	6.29
<i>fn1b</i>	2.80	3.74	<i>fbxl22</i>	5.96	2.26
<i>cyp3c1</i>	2.81	2.36	<i>mespba</i>	8.87	19.49
<i>phlda2</i>	2.90	2.24	<i>pcdh8</i>	10.36	3.38
<i>calhm2</i>	2.96	6.74	<i>mgsppb</i>	10.84	39.23
<i>baiap2l1b</i>	3.12	3.03			

### **3.6 Smn knockdown causes alternative splicing changes in motor neurons**

For identification of alternatively spliced transcripts, the Multivariate Analysis of Transcript Splicing (MATS) software was used. The bam files, obtained after TopHat mapping, were used as input for the MATS program. MATS uses, for each exon, the counts of RNAseq reads mapped to exon-exon junctions for estimating the exon inclusion levels in two samples. The statistical model of MATS then calculates a p-value and the values for the false discovery rate (FDR) for each exon. MATS automatically detects and analyses the alternative splicing (AS) events related to all major alternative splicing patterns. It analyses skipped exon (SE), mutually exclusive exons (MXE), retained intron (RI), alternative 3' (A3SS) and 5' (A5SS) splice site events (Fig. 12B). Overall 13,721 AS events were detected by MATS on comparing the SMA MN and control MN RNAseq data. Out of these 13,721 AS events, 284 events were found to be significant based on a FDR cut off of 0.1 (10%). Appendix 6 provides the list of all the significant AS events identified in SMA MN. The majority of these 284 events were SE events (222 events, 78%) (Fig. 12A). Apart from SE, other types of alternative splicing patterns were also detected (Fig. 12A): MXE (58 events, 20%), A3SS (1 event), A5SS (1 event) and RI (2 events). The number of AS events was not directly proportional to the number of genes as some genes had more than one AS event happening at different exons. There were only 248 genes found to be alternatively spliced with 27 of them having more than one AS event.

All of the 248 AS genes were screened for potential candidate genes with the search criteria that they were expressed in motor neurons or any other type of neurons. With the use of PubMed, ZFIN and NCBI databases, I shortlisted 16



**Figure 12. Alternative splice events in *Smn*-deficient motor neurons identified by MATS software.** (A) Total number of significant AS events, identified in *Smn*-deficient motor neurons, represented as pie chart. A FDR cut-off of 0.1 (10%) was used for defining significant events. (B) MATS detects five types of alternative splice events and analyses them. SE event represents inclusion or exclusion of an exon between two constitutive exons. A5SS and A3SS events represent use of alternative 5' or 3' splice site, respectively, during the splicing. MXE event represents inclusion of mutually exclusive exons in between two constitutive exons. RI event represents inclusion of an intron during splicing.

potential candidate genes based on either known biological function in zebrafish neurons or reported function of their human, mouse or other homologs in neurons. All of these shortlisted genes had exon skipping in their respective transcripts. Table 14 lists the shortlisted candidate genes with their respective inclusion values. These values represent level of inclusion/exclusion of a particular exon in the transcripts. Negative values indicate exclusion of the exon from transcripts under *Smn* deficiency, while positive values mean that the exon is normally excluded in controls, but is retained in *smn* morphants. Most of these 16 genes were found to be involved in intracellular transport (9 out of 16) like endocytosis, intracellular trafficking etc. All significant AS events were then analysed for their biological relevance

using the elim method for GO enrichment. A total 100 GO terms were found to be significant with elim value less than or equal to 0.01 (1%). A list of all the significantly enriched GO terms is provided in Appendix 8. Among the top enriched GO terms were Integral to membrane, cytoplasm, catalytic activity, RNA binding, nucleus etc.

**Table 14. List of potential candidate genes with alternative splicing in motor neurons under Smn deficiency**

S. No.	Gene	Gene Description	Skipped exon	Inclusion values*	Consequence	Affected protein domain	Related neuronal diseases
1.	<i>foxp2</i>	Forkhead box P2	4	-0.362	Loss of first 92 aa corresponding to exon 4	Transcription factor, fork head domain	Developmental verbal dyspraxia (Lai et al. 2001)
2.	<i>srsf6b</i>	Serine/arginine-rich splicing factor 6b	2	-0.298	Truncated protein of 37 aa instead of 355 aa	RNA recognition motif domain	Huntington's disease (Fernandez-Nogales et al. 2016)
3.	<i>snx9b</i>	Sorting nexin 9b	6	-0.294	Truncated protein of 188 aa instead of 581 aa	WASP binding domain	
4.	<i>zgc:91909</i>	Ras-related protein rab-7-like	3	-0.294	Truncated protein of 147 aa instead of 204 aa with only first 18 aa conserved	Small-GTP binding domain	Charcot-Marie-Tooth disease (Cogli et al. 2010, Ponomareva et al. 2016)
5.	<i>cadm1a</i>	Cell adhesion molecule 1a	10	-0.27	Loss of 11 aa corresponding to exon 10		Autism spectrum disorder (Zhiling et al. 2008)
6.	<i>rab15</i>	RAB15, member RAS oncogene family	5	-0.24	Loss of 30 aa corresponding to exon 5	Small-GTP binding domain	
7.	<i>zgc:77650</i>	ADP-ribosylation factor 4	2	0.069	Inclusion of exon 2	Small GTP binding domain	
8.	<i>si:dkey-119o24.1</i>	DEAH (Asp-Glu-Ala-His) box polypeptide 36 ATP-dependent RNA helicase (DHX36)	17	0.075	Inclusion of exon 17	Helicase-associated domain	
9.	<i>nrxn1a</i>	Neurexin 1a	11	0.099	Inclusion of exon 11 at Splice site 3 (SS3)	Laminin G domain	Pitt-Hopkins like mental retardation (Zweier et al. 2009, Peippo and Ignatius 2011)
10.	<i>si:ch211-225b11.1</i>	Uncharacterised	7	0.114	Inclusion of exon 7	Sodium:Neurotransmitter symporter	
11.	<i>snx14</i>	Sorting nexin 14	22	0.133	Inclusion of exon 22	Sorting nexin C-terminal domain	Spinocerebellar ataxia (Thomas et al. 2014, Akizu et al. 2015)

<b>S. No.</b>	<b>Gene</b>	<b>Gene Description</b>	<b>Skipped exon</b>	<b>Inclusion values*</b>	<b>Consequence</b>	<b>Affected protein domain</b>	<b>Related neuronal diseases</b>
12.	<i>nav3</i>	Neuron navigator 3	6	0.191	Inclusion of exon 6		
13.	<i>snx27a</i>	Sorting nexin 27a	12	0.251	Inclusion of exon 12		
14.	<i>vcla</i>	Vinculin a	19	0.313	Inclusion of exon 19	Vinculin/alpha-catenin domain	
15.	<i>copz2</i>	Coatamer protein complex, subunit zeta 2	9	0.333	Inclusion of exon 9		
16.	<i>ptenb</i>	Phosphatase and tensin homolog B	4	0.412	Inclusion of exon 4	Dual specificity phosphatase, catalytic domain	Spinal muscular atrophy (Little et al. 2015)

\*Inclusion values represent level of inclusion/exclusion of a particular exon in the transcripts in SMA MN as compared to Control MN. Negative values indicate exclusion of the exon while positive values mean inclusion of exon under Smn knockdown condition.

### 3.7 RT-PCR validation of alternative splice events identified in SMA motor neurons

Semiquantitative RT-PCR was performed to validate the results obtained from alternative splicing analysis on RNAseq data. For this assay, I used cDNA from FAC sorted motor neuron samples. I shortlisted a total of 16 candidate genes with alternative splicing that can possibly be implicated in SMA pathology. Out of these, I successfully validated three genes with alternative splicing in sorted motor neurons.

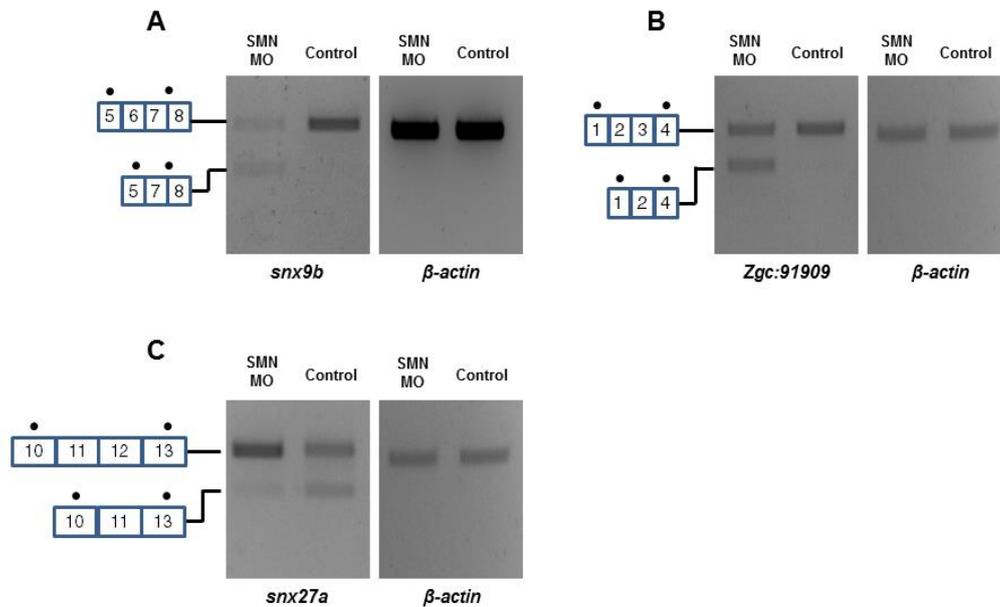
Sorting nexin (SNX) is a family of proteins characterised by the presence of a phox homology (PX) domain and is involved in diverse intracellular functions like endocytosis, protein sorting and endosomal signalling (Cullen 2008). Three genes from this family were found to be alternatively spliced in SMA motor neurons, *snx9b*, *snx14*, *snx27a*. I successfully confirmed the alternative splicing for *snx9b* and *snx27a*. From the alternative splicing analysis, *snx9b* was observed to have increased skipping of exon 6 in SMA MN as compared to controls (Table 14). This skipping leads to creation of a pre-mature stop codon resulting in a truncated protein of 188 aa instead of 581 aa and would possibly reduce overall Snx9b levels. Alternative splicing was confirmed by RT-PCR analysis, where a lower band, corresponding to exon 6 skipped transcripts, was present in SMA MN (Fig. 13A). Also the upper band, representing exon 6 including transcripts, was reduced in intensity as compared to the upper band in case of controls.

In case of *snx27a*, alternative splicing analysis revealed an increased retention of exon 12 in SMA MN as compared to controls (Table 14). This was validated by RT-PCR where a stronger upper band (ex10-11-12-13) in SMA MN was observed as compared to controls, while the lower band (ex10-11-13) was reduced in case on SMA motor neurons (Fig. 13B). This shows that

Smn deficiency leads to increased retention of exon 12 in *snx27a* transcripts. Exon 12 retention in *snx27a* only affected the last two exons in the gene and did not lead to a truncated protein.

*zgc:91909* is an uncharacterised gene in zebrafish that codes for a novel protein similar to human RAB7A, which is a member of the RAB family of proteins involved in regulation of vesicular transport (Pfeffer 1994, Hutagalung and Novick 2011). In case of *zgc:91909*, there was increased skipping of exon 3 in SMA MN leading to frameshift in the coding sequence and translating into non-functional protein (Table 14). Alternative splicing was also confirmed by RT-PCR that revealed the presence of a lower band corresponding to the exon 3 skipped transcripts (Fig. 13C).

Interestingly, the alternative splicing in all three genes did not manifest into mRNA level changes, indicating that Smn deficiency did not affect the overall gene expression (Zhang et al. 2013). Taken together, the above results suggest that Smn deficiency leads to alternative splicing of genes which could be important for normal neuronal physiology.



**Figure 13. RT-PCR validation of alternative splicing events in SMA MN identified by RNAseq.** RT-PCR reactions confirmed exon skipping in SMA MN. *snx9b* and *zgc:91909* showed increased exon skipping (exon 6 and 3, respectively) in *smn* morphants while *snx27a* showed increased exon retention (exon 12) in *smn* morphants as compared to Control MN.  $\beta$ -actin was used as loading control. Spliced isoforms are shown as boxes labeled with the corresponding exon number. Black dots on top of the boxes indicate primer binding sites.

### 3.8 Transcriptome analysis of FAC sorted Schwann cells after

#### Smn knockdown

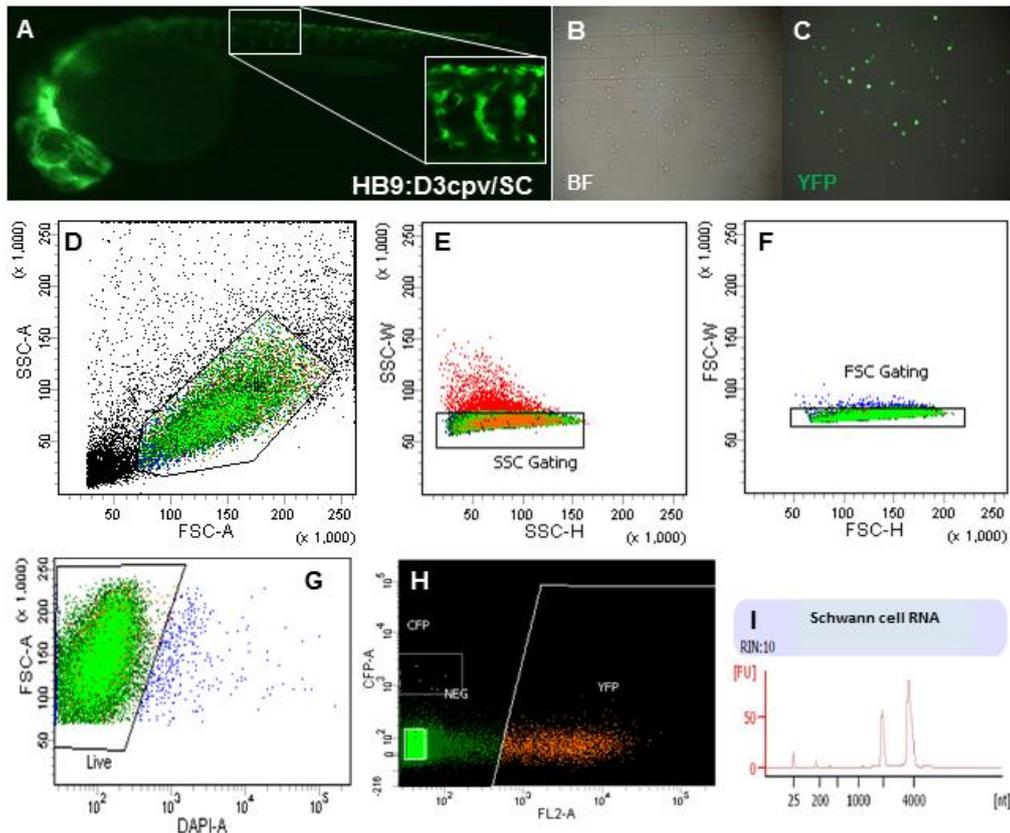
Schwann cells wrap around motor axons and play a crucial role in support and maintenance of neuromuscular junctions (NMJs), which are found to be defective in SMA cases. Gillingwater's group has shown that a SMN deficiency leads to intrinsic Schwann defects in a mouse model of SMA (Hunter et al. 2014). A former graduate student in our lab, Kelvin See, showed that *Smn* knockdown in zebrafish leads to Schwann cells defects, which are independent of motor neuron activity (Kelvin See, PhD Thesis). Shermaine Tay, a graduate student in our lab, also showed that the motoaxonal defects upon *Smn* knockdown can be partially rescued by

transgenic expression of *Smn* in Schwann cells alone (Shermaine Tay, FYP thesis). This opens up the possibility that splicing defects in Schwann cells might contribute to motoaxonal defects. To address this, I performed RNAseq on FAC sorted Schwann cells in the zebrafish SMA model. I used a transgenic Schwann cell calcium sensor line, HB9:D3cpv/SC, generated by Kelvin See, for this experiment. This line expresses calcium sensor D3cpv in Schwann cell precursors as well as immature in Schwann cells (Fig. 14A).

### **3.8.1 Dissociation of *Smn* deficient HB9:D3cpv/SC embryos and FAC sorting**

The optimised protocol for HB9:eGFP embryo dissociation and motor neuron FAC sorting was used for sorting Schwann cells from HB9:D3cpv embryos under *Smn* deficient and control conditions at 28 hpf. Since the age of embryos used was the same as in the case of HB9:eGFP embryos, the dissociation conditions were not changed. The calcium sensor protein, D3cpv, which emits fluorescence in CFP as well as in the YFP channel (by FRET in its Ca<sup>2+</sup> bound form) was used for sorting Schwann cells. Interestingly, almost all of the sorted cells were YFP positive while only a few cells were CFP positive (Fig. 14H). This could be due to the presence of excessive calcium in the L-15 medium and serum. Thus, only YFP emission was used for FAC sorting Schwann cells from dissociated HB9:D3cpv embryos (Fig. 14D-H).

On average, 100 HB9:D3cpv embryos generated close to 108,000 YFP positive cells after FAC sorting. The yield of sorted YFP positive cells was between 4-5%, which was considerably higher than the GFP positive MN cells (0.8-1%). The sorted Schwann cells generated on an average 11 ng of high quality (RIN >8.5) total RNA, as assessed by the Agilent bioanalyser (Fig. 14I).



**Figure 14. FAC sorting of YFP positive Schwann cells.** (A) HB9:D3cpv embryo at 30 hpf. Brightfield (B) and fluorescent (C) image of sorted YFP positive cells under fluorescence microscope. (D) Total cells were represented in a FSC-A vs. SSC-A plot to screen for homogenous cell populations (enclosed region) without cell debris and cell aggregates (everything except enclosed region). (E and F) The selected cell population was then checked again on SSC and FSC plots for removing any doublets and only single cell population was selected (SSC and FSC gating). (G) Cells were checked for live cells (enclosed region; DAPI negative) among the entire single cell population. Cells on right side of plot were dead (DAPI positive). (H) Finally live single cells were sorted on the basis of YFP fluorescence. The YFP gated cells were the live YFP positive cells that were finally collected in L-15 media containing 0.5% FBS. (I) Agilent Bioanalyser profile of total RNA extracted from the sorted YFP positive cells. BF- bright field.

### 3.8.2 cDNA library generation and RNA sequencing of SMA Schwann cells

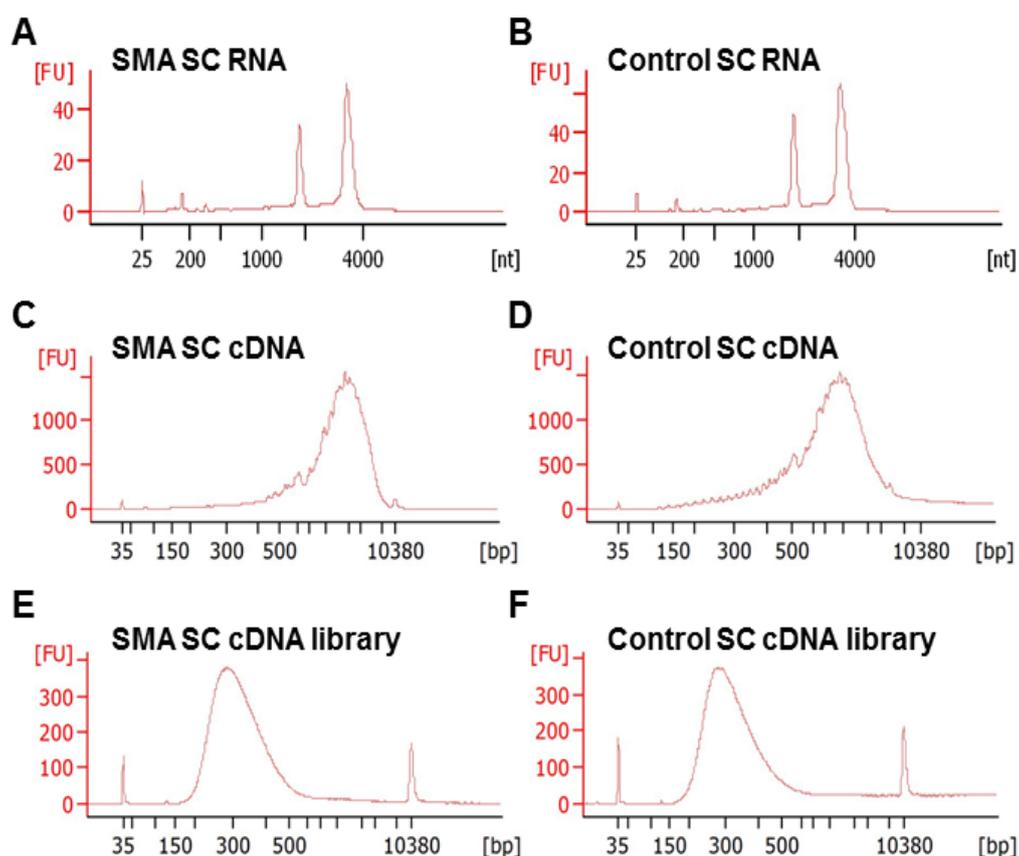
Two sets of high quality RNA samples (RIN >9) isolated from FAC sorted Schwann cells, both from *smn* morphants (SMA SC) and control embryos (Control SC), were selected to generate cDNA libraries for RNAseq (Table 15; Fig. 15A,B). The library preparation protocol was identical to the one used for SMA MN and DSMA1 MN libraries. First, the RNA samples were used for

ds cDNA synthesis with Oligo-dT primer and Moloney Murine Leukemia virus reverse transcriptase followed by LD-PCR amplification. The amplified ds cDNA was purified using magnetic beads and checked on an Agilent bioanalyser using the HS DNA assay kit (Fig. 15C,D). Both sets of samples produced high quality cDNA profiles similar to the cDNA profile for control RNA (see Appendix 1).

**Table 15. List of samples used for SMA SC RNAseq**

Sample		No. of embryos	Total cells	RNA yield (ng)	RIN	RNA input for SMARTer kit (ng)	DNA input for NEBNext kit (ng)
Control SC	Sample 1	48	111,000	12.5	9.4	5	24.4
	Sample 2	149	89,000	14.5	9.7	2.7	201
SMA SC	Sample 1	298	144,000	25.8	9.8	5	67.5
	Sample 2	229	92,000	9.4	9.5	2.7	121

Next, the amplified cDNA samples were subjected to Covaris shearing to generate small fragments of 200-500 bp in size. These fragmented cDNAs were then used to generate libraries using adaptors and barcoded with different Index primers. Finally, the libraries were purified and checked for quality on an Agilent bioanalyser. All libraries produced good profiles with a distinct peak spanning 200-500 bp, peaked at approx. 300 bp (Fig. 15E,F). All cDNA libraries were quantified using qRT-PCR and pooled to give identical amounts of each library. The multiplexed libraries were submitted for sequencing with the Next Generation Sequencing platform at the Genome Institute of Singapore in collaboration with Dr. S. Mathavan. The sequencing was performed on an Illumina HiSeq platform with paired-end sequencing (75-100 bp read length).



**Figure 15. Assessment of quality for extracted RNA, amplified cDNA and final library during SMA Schwann cell library preparation using Agilent bioanalyser.** Each step in the protocol was checked by running 1 $\mu$ l sample on a bioanalyser. The Y-axis represents the fluorescence intensity detected (FU). The X-axis represents the fragment size of RNA or DNA in nucleotides (nt) or base pair (bp), respectively. Representative profiles for one SMA SC and one Control SC sample are shown here. (A,B) Electropherograms of input RNA for SMA and control SC sample used for RNAseq library preparation with RIN values of 9.5 and 9.7, respectively. (C,D) Electropherograms of cDNA obtained after first strand cDNA synthesis and amplification during SMARTer protocol. (E,F) Electropherograms of cDNA libraries after completion of NEBNext library prep protocol, with lengths of all the fragmented cDNA between 200-500 bp.

### 3.8.3 Differential gene expression analysis of Schwann cells

The RNAseq data obtained from sequencing SMA SC libraries was used for bioinformatics analysis in collaboration with Dr. Candida Vaz and Dr. Vivek Tanavde at Bioinformatics Institute Singapore. The RNAseq results for SMA SC libraries are shown in Table 16. The total number of reads obtained was

between 103,973,678 and 204,591,620. The Mean Phred quality score (Q) obtained was above 35 with more than 90% of base calls at 99.9% accuracy.

**Table 16. SMA SC RNAseq results**

<b>Sample</b>	<b>Yield (Mbases)</b>	<b># Reads</b>	<b>% of <math>\geq</math> Q30 Bases</b>	<b>Mean Quality Score</b>
<b>SMA SC Sample 1</b>	17,151	182,588,058	92.25	35.7
<b>Control SC Sample 1</b>	19,237	204,591,620	92.1	35.63
<b>SMA SC Sample 2</b>	9,728	103,973,678	90.14	35.05
<b>Control SC Sample 2</b>	10,767	116,017,986	90.58	35.27

Similar to the SMA MN and DSMA1 MN RNAseq data, adaptor sequences were first removed from all reads obtained and mapped to the zebrafish reference genome, Zv9, using TopHat tool (2.0.12). The total number of reads per sample was between 103,973,678 and 204,591,620 (Table 17); between 71.4% and 83.2% of reads were uniquely mapped to the zebrafish reference genome. The no. of aligned pairs per sample were between 20,631,802 and 57,931,738 (Table 17), out of which concordant pair reads were between 29.3% and 61.3% uniquely mapped to zebrafish reference genome (Table 17).

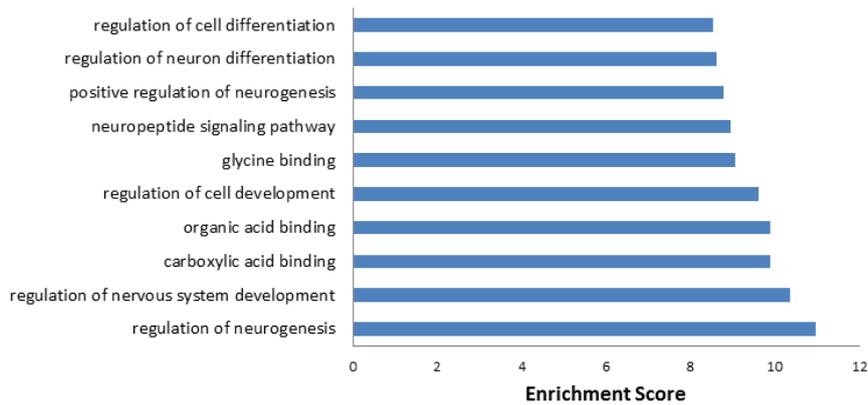
The mapped reads when analysed on Partek for differentially expressed genes. A total of 13,910 genes were identified to be expressed in the sorted Schwann cells at 28 hpf out of which 226 genes were found to be differentially expressed in SMA SC (Fold change  $>2$  and p-value  $<0.05$ ). Out of 226 differentially expressed genes, 129 were down-regulated in SMA SC as compared to Control SC while 97 were up-regulated. A list of all the differentially expressed genes is provided in the Appendix 4. Comparison of

differentially expressed genes in SMA SC and SMA MN revealed only 27 overlapping genes, indicating cell type-specific defects under Smn deficiency (Zhang et al. 2013).

**Table 17. SMA SC RNAseq reads mapped to zebrafish reference genome Zv9 using TopHat**

	Control SC		SMA SC	
	Sample 1	Sample 2	Sample 1	Sample 2
<b>Total Reads</b>	204,591,620	116,017,986	182,588,058	103,973,678
<b>Reads Mapped</b>	76.2%	47.2%	76.2%	72.4%
<b>Total Paired Reads</b>	57,931,738	20,631,802	53,882,303	28,910,040
<b>Concordant Pairs Mapped</b>	29.3%	36.7%	61.3%	58.3%

To determine the biological relevance of these differentially expressed genes in SMA SC, Partek software was used. GO enrichment analysis identified 195 GO terms to be over represented with a p-value cut-off of 0.05. Interestingly, terms related to development and differentiation were the most enriched terms with “regulation of neurogenesis” and “regulation of nervous system development” as the top most enriched term in the list. Apart from this, neuropeptide signalling pathway was also found to be affected in SMA SC. Fig. 16 shows a graphical representation of the top 10 GO terms with their respective enrichment scores.



**Figure 16. Gene Ontology (GO) analysis of differentially expressed genes in Snn deficient and control Schwann cells.** GO analysis was performed on differentially expressed genes with at least a 2-fold change and p-value less than or equal to 0.05. The top 10 GO terms enriched are represented in a bar graph with the Y-axis showing the enriched GO terms and X-axis showing the respective enrichment score values.

### 3.9 Snn knockdown leads to alternative splicing in Schwann cells

#### 3.9.1 Identification of alternative splice events using MATS

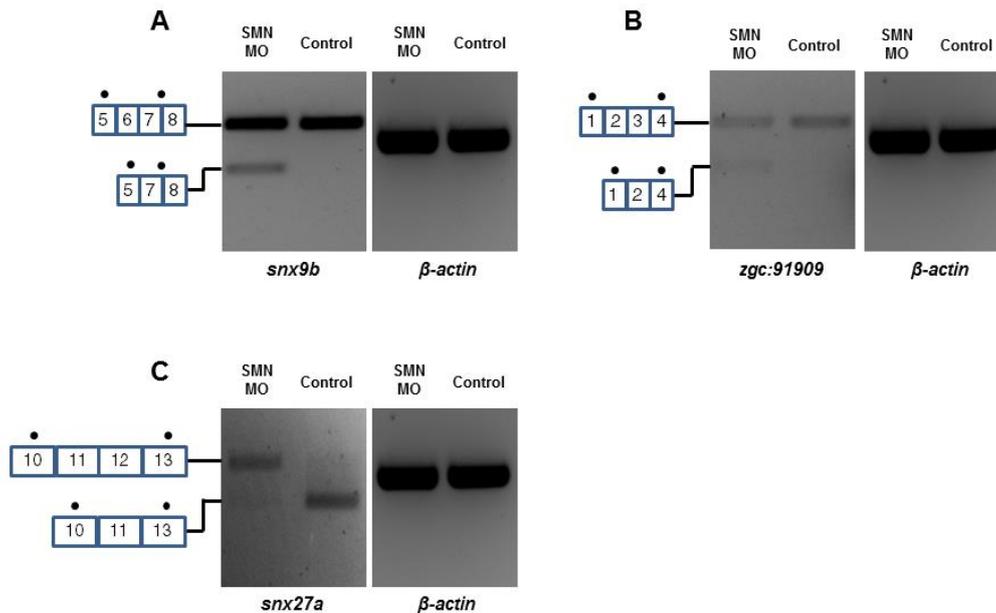
Alternative splicing analysis was also performed on the SMA SC RNAseq data using MATS. The analysis identified a total of 7133 AS events, out of which only 39 AS events were filtered as significant (FDR <0.1). Appendix 7 lists out all the AS events in SMA SC. Unlike in motor neurons, only a few significant AS events were detected in Schwann cells. This could be due to low percentage mapping of the Control Schwann cell RNAseq data. A total 37 genes were found to be alternatively spliced with two of them having more than one AS event. Out of 39 AS events, 38 were skipped exon events while only one mutually exclusive event was detected. Interestingly, 29 AS events were similar to those found in SMA motor neurons while only 10 AS events were unique to Schwann cells. Among these 37 AS genes, five of them were found in the potential candidate gene list (Table 14), although, with different

inclusion values. These genes were *srsf6b*, *snx9b*, *zgc:91909*, *snx27a* and *ptenb* which were validated using RT-PCR.

### **3.9.2 RT-PCR validation of alternative splice events identified in SMA Schwann cells**

Using a semiquantitative RT-PCR assay, I successfully validated the alternative splicing of *snx9b*, *snx27a* and *zgc:91909*. cDNA obtained from FAC sorted Schwann cells was used for the assay. *snx9b*, as per alternative splicing analysis, had exon skipping at exon 6 which was confirmed by RT-PCR. A lower *snx9b* band, corresponding to exon 6 skipped transcripts, was observed in SMA Schwann cells, while control Schwann cells did not show expression of this transcript (Fig. 17A). The same was the case with *zgc:91909*, where alternative splicing analysis showed increased skipping of exon 3 in *zgc:91909* transcripts. RT-PCR result also showed that SMA Schwann cells expressed transcripts with exon 3 skipped along with full length transcripts. Control Schwann cells, on the other hand, showed expression of only full length transcript (ex1-2-3-4) (Fig. 17B).

For *snx27a*, alternative splicing analysis reported increased retention of exon 12 in SMA Schwann cells as compared to controls. This was also confirmed using the RT-PCR assay. The gel showed presence of a strong *snx27a* upper band (ex10-11-12-13) in SMA Schwann cells while the controls appeared to express the exon 12 skipped transcripts (strong lower *snx27a* band) (Fig. 17C). This, unlike other previous cases, showed an almost switch from exon 12 skipped transcripts in controls to exon 12 retained transcripts under *Smn* deficiency. Taken together, these results indicate that *Smn* deficiency leads to an alternative splicing in Schwann cells.



**Figure 17. RT-PCR experiments validated alternative splicing changes in SMA Schwann cells identified by RNAseq.** RT-PCR reactions confirmed the exon skipping events in SMA Schwann cells. *snx9b* and *zgc:91909* showed increased exon skipping (exon 6 and 3, respectively) in *smn* morphants while *snx27a* showed increased exon retention (exon 12) in *smn* morphants as compared to Control SC.  $\beta$ -actin was used as loading control. Spliced isoforms are shown as boxes labeled with the corresponding exon number. Black dots on top of the boxes indicate primer binding sites.

### 3.10 *srsf6b* splicing factor as a novel candidate target of *Smn*

*Smn* deficiency leads to alternative splicing of genes in motor neurons as well as Schwann cells. From alternative splicing analysis of both cell types one potential candidate gene serine/arginine-rich splicing factor 6 (*srsf6b*) garnered particular interest. *Srsf6b* belongs to Serine-arginine (SR) family of proteins characterised by the presence of carboxy-terminal Arginine-serine (RS) domain and at least one RNA binding domain (RRM) preceding it (Zahler et al. 1992). SR proteins are involved in both constitutive (Mayeda et al. 1999) and alternative splicing (Zahler et al. 1993) of pre-mRNA, as well as post-splicing activities like mRNA nuclear export, translation and nonsense mediated decay (Shepard and Hertel 2009). Interestingly, SRSF proteins

have been shown to regulate splicing of *SMN1* and *SMN2* transcripts (Wee et al. 2014).

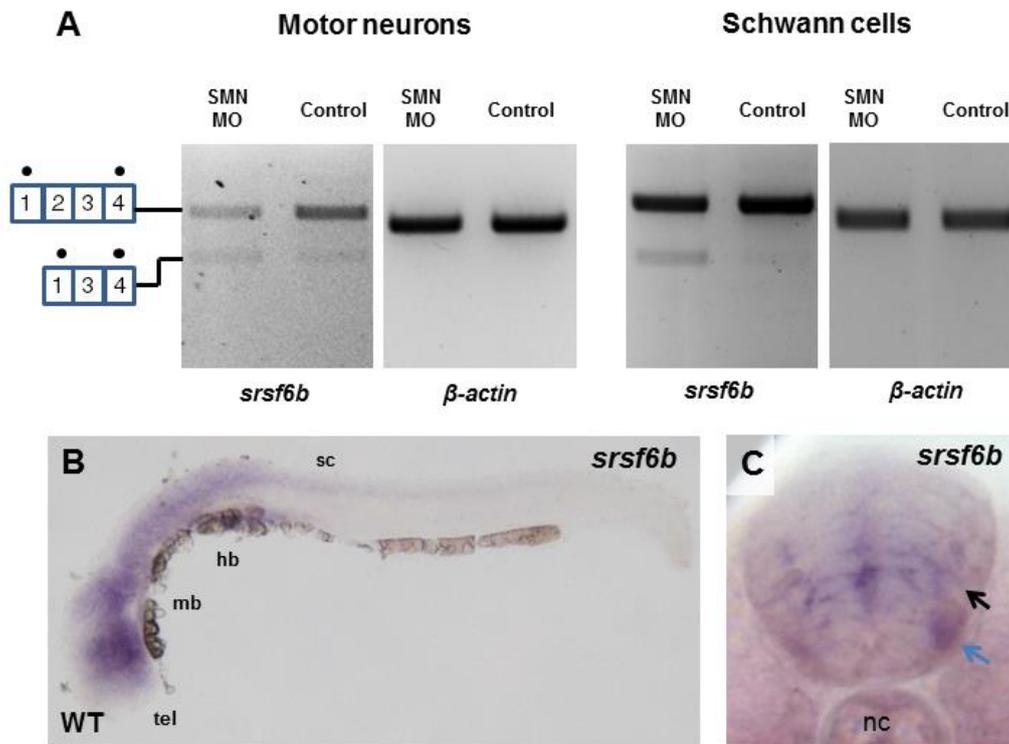
Alternative splicing analysis revealed increased skipping of exon 2 in *srsf6b* transcripts in both motor neurons and Schwann cells under *Smn* deficient conditions. Semiquantitative RT-PCR assay was performed to validate this alternative splicing. RT-PCR performed using primers flanking exon 2 of *srsf6b* showed a weak lower band corresponding to exon 2 skipped transcripts in both SMA motor neurons and Schwann cells (Fig. 18A). In case of controls, the majority of RT-PCR product was corresponding to exon 2 inclusive transcripts (upper band) while the lower band was very weak (Fig. 18A). These results confirm that *Smn* deficiency affects the splicing of *srsf6b* in motor neurons and Schwann cells. Exclusion of exon 2 in *srsf6b* transcripts would lead to creation of pre-mature stop codon generating a 37 aa truncated protein with no RS and RRM domains. This suggests that *Smn* deficiency leads to reduction in functional *Srsf6b* proteins in SMA MN and SC and might lead to enhanced splicing defects in these cell types.

### **3.10.1 Expression pattern of *srsf6b***

Having established the effect of *Smn* on alternative splicing of *srsf6b*, it was important to verify whether *srsf6b* was expressed in motor neurons of zebrafish. According to ZFIN database and literature, there was no reported expression of *srsf6b* in zebrafish. To determine the expression pattern of *srsf6b* whole mount RNA *in situ* hybridization was performed.

It was found that *srsf6b* was expressed in various parts of the central nervous system including telencephalon, midbrain, hindbrain and spinal cord, with highest levels in its anterior part (Fig. 18B). Expression of *srsf6b* in the spinal cord was restricted to its ventral domain where it was expressed in the region

containing motor neurons and V2 interneurons (Fig. 18C). Expression in Schwann cells, however, was not detectable using this method. This suggests a possible role of *srsf6b* in zebrafish nervous system development, and importantly in the development of motor neurons.



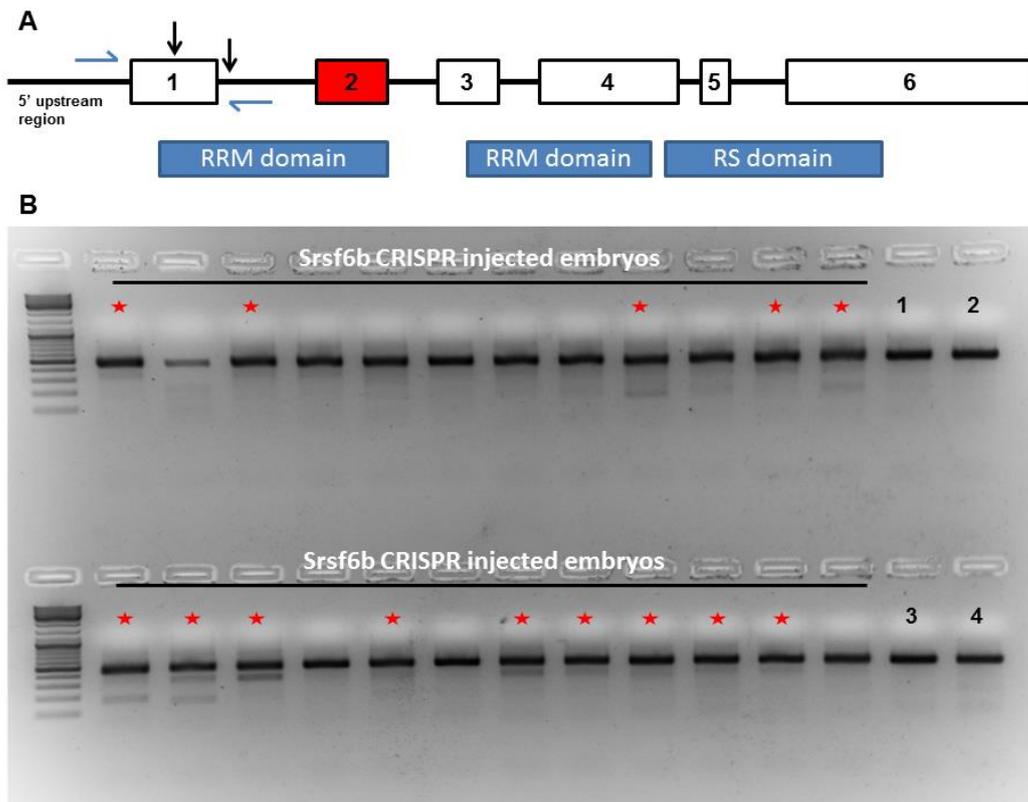
**Figure 18. Smn deficiency leads to alternative splicing of *srsf6b* in motor neurons and Schwann cells.** (A) RT-PCR assay validated the alternative splicing of *srsf6b* in FAC sorted motor neurons and Schwann cells under Smn deficient conditions. Loss of Smn increases skipping of exon 2 in *srsf6b* transcripts in both cell types.  $\beta$ -actin was used as loading control. Spliced isoforms are shown as boxes labeled with the corresponding exon number. Black dots on top of the boxes indicate primer binding sites. (B) Whole mount *in situ* hybridization identified expression of *srsf6b* in telencephalon (tel), midbrain (mb), hindbrain (hb) and spinal cord (sc) at 31 hpf. (C) Transverse sections through trunk of *srsf6b* stained embryo showing expression in the ventral spinal cord labeling regions of primary motor neurons (blue arrow) and V2 interneurons (black arrow). nc- notochord.

### 3.10.2 CRISPR/Cas mediated knock-out of *srsf6b*

Having demonstrated that Smn deficiency causes alternative splicing of *srsf6b* and that *srsf6b* is expressed in the ventral spinal cord, I speculated

that it might be important for motor neuron development. Therefore, I next analysed the effect of a mutation of this gene on motor neurons. For this, I employed a transient CRISPR/Cas based functional study. Two CRISPR target sites were identified in *srsf6b* using CRISPRscan program with high score and no off-targets. The target sites, 93 bp apart, were located in exon 1 and intron 1 of the *srsf6b* gene (Fig. 19A). These target sites were used for preparing gRNAs. 100 ng/μl of each gRNA and 300 ng/μl of Cas9 mRNA were injected into wild-type embryos and surviving injected fish were analysed for indels in *srsf6b* gene sequence. The evaluation of CRISPR efficiency and identification of *srsf6b* potential mutants was done by analysing gDNA from CRISPR injected fish using primers flanking the two target sites (Fig. 19A). The two gRNAs would recognize their respective target sites in the *srsf6b* gene and direct the Cas9 protein for producing double stranded breaks at these sites. These breaks, repaired by erroneous Non-Homologous End Joining (NHEJ) mechanism would generate insertions and/or deletions (Indels), which were identified using the flanking primers. In my case, with double CRISPR, ideally a big deletion corresponding to the distance between the two target sites was expected. The wild-type band was 475 bp in size. Because of the mosaic character of the introduced mutations, a positive mutant embryo was expected to have wild-type band along with smaller mutant band. A total of 46 CRISPR injected embryos were analysed, out of which 26 embryonic gDNAs showed smaller bands of varying sizes. Fig 19B shows a representative result. Wild-type gDNAs were used as negative controls, which gave a single band of 475 bp. Mutant embryos showed mutant bands in the size range of 200-400 bp, with a majority of them having an approx. 400 bp mutant band. Five embryos showed a large deletion in the *srsf6b* gene, with mutant bands of 200 bp. The overall efficiency of F0 potential *srsf6b* mutants was 56.5%. The injected embryos, at 100 ng/μl dose

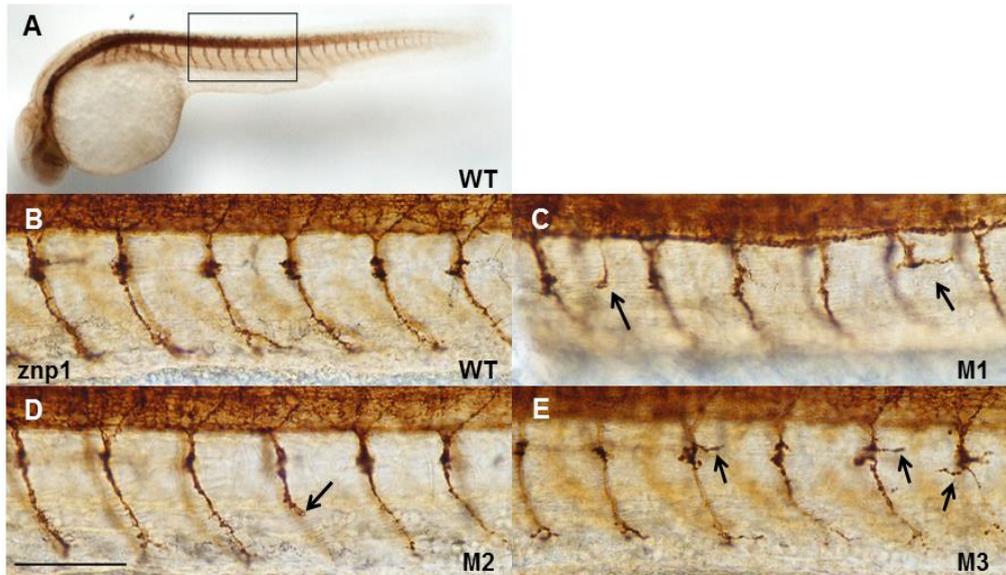
of each gRNA, did not show a high lethality, however, a few embryos developed morphological defects like a curved body curvature and delayed development. Some, but not all, of the defective embryos showed mutations in *srsf6b* suggesting that the loss of Srsf6b leads to morphological defects in embryos.



**Figure 19. CRISPR/Cas mediated *srsf6b* knock-out.** (A) Structure of *srsf6b* gene with RRM and RS domains. Exon 2 marked with red box is skipped in *Smn* deficient motor neurons and Schwann cells. CRISPR target sites are marked with black arrows and screening primers are marked with blue arrows. (B) Representative gel image showing screening of 24 *srsf6b* double CRISPR injected embryos. Mutant embryos with smaller bands are marked with red star. Four wild-type embryos (1, 2, 3 and 4) were used as negative controls.

Next, I performed whole mount immunostaining on *srsf6b* double CRISPR injected embryos to analyse the effect of a mosaic *srsf6b* knock-out on motor axonal development. Injected embryos were raised until 31 hpf and surviving embryos were used for the analysis of motor axonal phenotypes. Embryos

were stained with the  $\alpha$ -znp1 antibody that stains motoaxons and embryos were analysed in similar manner as was done for *smn* and *ighmbp2* morphants. A total of 39 *srsf6b* double CRISPR injected embryos were analysed for motoaxonal defects. Interestingly, 27 out of 39 embryos (69.23%) were found to have two or more axonal defects (referred to as 'defective') in the form of branching or truncation. Fig. 20C-E shows representative images of embryos with motoaxonal defects. gDNA extracted from heads of all of these embryos was separately analysed for mutants. A total of 19 embryos (48.7%) showed mutant bands. However, when the two results were compared, the data was not completely overlapping. Out of the 27 defective embryos, 12 showed a corresponding mutant band while 15 did not show the mutant band. On the other hand, six out of seven embryos that showed a mutant band had only one defect, while one embryo had no defect but still showed mutant band on gel. The absence of mutant bands for the defective embryos could be explained by the fact that only one of the two CRISPRs worked leading to small indels that could not be picked up on the gel, but disrupted the protein coding sequence. Alternatively, due to the mosaic character of the introduced biallelic mutations, this could have affected the trunk but not the head.



**Figure 20. *srsf6b* knock-out leads to motoaxonal defects.** (A) Lateral view of wild-type embryo. Box indicates region above yolk extension where motoaxon morphology was assessed. (B) Higher magnification view of wild-type embryo. (C-E) Representative images of three *srsf6b* double CRISPR injected embryos (M1, M2, M3) showing severe axon defects (branching and truncation). Arrows mark the defective axons. Scale bar= 100  $\mu$ m.

Taken together, these results indicate that *srsf6b* is indeed important for motor neuron differentiation, as a knock-out of *srsf6b* caused motoaxonal defects similar to those seen in *smn* or *ighmbp2* morphants (Fig. 3C and 4D). This further provides evidence that *srsf6b* is a novel downstream target of Smn.

## 4. DISCUSSION

Currently there are three main hypotheses to explain why a reduction of ubiquitous SMN protein leads to degeneration of motor neurons. The first hypothesis is based on SMN's well-established role in U snRNP assembly. According to this, loss of SMN causes reduced U snRNP levels, consequently affecting proper splicing of the motor neuron-specific genes, leading to motor neuron defects (Pellizzoni et al. 1998, Meister et al. 2001, Pellizzoni et al. 2002, Winkler et al. 2005). Apart from this, a role of SMN in motoaxonal mRNA regulation (Rossoll et al. 2002, Rossoll et al. 2003) and finally in the maintenance of a neural circuitry has been proposed (Imlach et al. 2012). A common denominator in all the three hypotheses is that SMN loss directly or indirectly causes motor neuron defects, which ultimately leads to SMA phenotype. Recent reports have identified transcriptomic changes and alternative splicing in motor neurons under SMN deficient conditions (Zhang et al. 2013, Huo et al. 2014, Maeda et al. 2014, Saal et al. 2014, Ng et al. 2015). Microarray analysis in whole zebrafish embryos identified *nrxn2a* as a novel downstream target of Smn which gets alternatively spliced when Smn levels are reduced (See et al. 2014). Schwann cells defects are also reported in the mouse SMA model (Hunter et al. 2014). However, little is known about the transcriptomes of motor neurons or surrounding Schwann cells in the zebrafish SMA model. Here, I report transcriptomic and splicing changes in these cells types after knock-down of Smn.

#### **4.1 Method development for RNAseq of FAC sorted motor neurons and Schwann cells from zebrafish embryos**

In this study, I report a method for transcriptome profiling of FAC sorted motor neurons and Schwann cells from whole zebrafish embryos. As a first step towards this, I optimised protocols for dissociation of whole embryos and isolating motor neurons and Schwann cells using FACS. A papain dissociation system, adopted from Cerda et al., was used for dissociating whole zebrafish embryos (Cerda et al. 2009). Papain, a cysteine protease, has previously been reported to be better than trypsin, collagenase or other proteases for dissociation of rat neural tissues (Huettner and Baughman 1986). It has also been used for dissociating neural cells from rat spinal cords, particularly for creating neuroblast and neural stem cell cultures (Shihabuddin 2008). Cerda et al reported dissociation of zebrafish transgenic embryos using papain to FAC sort interneurons for RNA profiling (Cerda et al. 2009).

I successfully optimised the protocol for dissociation of 28 hpf zebrafish transgenic embryos to obtain live single cell suspensions of motor neurons and Schwann cells suitable for FACS. Since papain is a protease, it was critical to optimise the incubation time for the dissociation so as to prevent cell lysis. 30 mins of papain incubation was used for dissociation of cells from trunks of zebrafish embryos at 27 hpf (Cerda et al. 2009). However, in this study, 45 mins of papain incubation at 30°C was found to be optimal for recovery of live single cells. As per the manufacturer's recommendations, papain incubation should be performed at 37°C. However, 38-39°C has been reported to induce heat shock in zebrafish embryos causing an unfolded protein response (UPR), activation of innate and adaptive immunity and changes in the transcriptome (Shoji and Sato-Maeda 2008, Todd et al. 2008,

Costa et al. 2011, Long et al. 2012). Loss of SMN has recently been reported to activate UPR in motor neurons generated from iPSCs derived from SMA patients (Ng et al. 2015). Therefore, to avoid any non-specific effects of the heat shock on transcriptomes of motor neurons and Schwann cells, 30°C was used as incubation temperature, which did not affect the dissociation of embryos.

The single cells obtained after dissociation were subjected to FACS for sorting of motor neurons and Schwann cells. Wild type embryos as well as non-transgenic siblings were used for setting up the FACS parameters to sort true fluorescent positive cells. On an average 12,000 GFP positive motor neurons (0.8-1%) and 108,000 YFP positive Schwann cells (4-5%) were sorted out from 100 transgenic embryos (HB9:eGFP or HB9:D3cpv/SC). With the optimised FACS settings, the percentage of cells sorted out was consistent throughout the study. The entire optimised protocol, from dissociation to cell sorting, was done in less than 2 hrs. This ensured that any possible transcriptomic changes in the cells were kept to a minimum during the procedure and consequently resulted in more reliable RNAseq results. A recent study discussed the effect of isolation and separation procedures on transcriptomic changes (Richardson et al. 2015). The group used microarray analysis and concluded that the isolation procedure rather than FACS perturbs the gene expression of cells (Richardson et al. 2015). In order to avoid any significant differences among the samples due the dissociation protocol, all the samples were prepared with exactly similar conditions. Also, the transcriptomic or splicing changes that might still occur due the embryo dissociation protocol could be normalised while comparing controls and SMN morphants. For RNA extraction from FAC sorted cells, a phenol-chloroform based extraction method followed by column purification was used. High

quality (RIN >9) total RNA, suitable for RNAseq, was achieved from this approach. The identity of the sorted cells was confirmed by analysing them with a fluorescence microscope and by performing RT-PCR for motor neuron-specific markers like *hb9* and *isl1*. Although this entire protocol was developed for motor neurons and Schwann cells from whole zebrafish embryos, it can be easily be adapted for other cells types and adult tissues.

Using the optimised protocol for dissociation of embryos and FACS, I successfully sorted motor neurons and Schwann cells under *Smn* deficient conditions from HB9:eGFP and HB9:D3cpv/SC embryos, respectively. Motor neurons were also sorted from HB9:eGFP embryos under *Ighmbp2* deficient conditions. For cases with low RNA inputs, RNA amplification kits are usually applied prior to library construction for transcriptomic profiling (Shanker et al. 2015). The RNA yields obtained from sorted motor neurons and Schwann cells was also very low, thus RNA amplification was performed using the SMARTer kit (Clontech) to generate high quality amplified cDNA. Shanker et al. evaluated different RNA amplification kits and reported that libraries generated using Clontech had the highest percentage of unique reads as well as exonic reads compared to other kits (Shanker et al. 2015). High quality cDNA libraries were then generated using NEBNext library prep kit (and submitted for sequencing. The sequencing was performed on an Illumina HiSeq platform with sequencing depth between 100-200 million reads which was suitable for alternative splicing analysis as well as to identify differentially expressed genes. Additionally, sequencing was done using the paired-end method to generate high quality data suited for alternative splicing analysis. The RNAseq data was finally used for bioinformatics analysis to identify alternatively spliced transcripts as well up- and down-regulated genes under disease conditions.

## 4.2 Smn deficiency affects gene expression in motor neurons and Schwann cells

Low levels of SMN leading to reduced snRNP levels has been hypothesised to result in aberrant splicing of pre-mRNAs leading to degradation of wrongly spliced transcripts by non-sense mediated decay (Pellizzoni et al. 1998, Pellizzoni et al. 2002, Gubitz et al. 2004, Behm-Ansmant et al. 2007, Zhang et al. 2008, Wittkopp et al. 2009). Microarrays have been previously used for differential gene expression studies using *smn* morphant zebrafish embryos (See et al. 2014), motor neuron cultures (Anderson et al. 2004) and whole spinal cords from SMA mouse models (Zhang et al. 2008, Baumer et al. 2009, Murray et al. 2010). As mentioned previously, RNAseq is advantageous for transcriptomic profiling because high signal-to-noise ratio, independence from hybridization efficiency, high reproducibility for low RNA inputs and identification of novel RNA transcripts. In this study, RNAseq was performed on FAC sorted motor neurons and Schwann cells under Smn deficient conditions. Mapping of RNAseq data to the zebrafish reference genome, Zv9, identified similar numbers of expressed genes in both cell populations, i.e. 13,923 and 13,907 genes in motor neurons and Schwann cells, respectively. Differential gene expression analysis revealed 312 and 226 genes (Fold change >2, p-value <0.05) to be deregulated in SMA MN (n=3) and SC (n=2). Comparing the differentially expressed genes identified in zebrafish SMA MN with earlier microarray and RNAseq studies in other animal models identified only a few overlapping genes (Huo et al. 2014, Saal et al. 2014, Ng et al. 2015, Doktor et al. 2016). This could be due various reasons like use of different animal models (mouse, cell culture or zebrafish), different time points analysed (asymptomatic, pre-symptomatic, post-symptomatic) or differences in techniques (RNAseq or microarrays). Although both cell populations

surprisingly expressed similar genes overall (~95% overlap between motor neuron and Schwann cell expressed genes), only 27 genes were affected in common (see Appendix 5), indicating that Smn deficiency causes cell type-specific mRNA level changes (Zhang et al. 2013). Out of these 27 overlapping genes, three genes had opposite regulation patterns while 24 genes were regulated similarly in both cell types. Also, the total number of differentially expressed genes was about 2.2% and 1.6% of expressed genes in motor neurons and Schwann cells, respectively. This suggests that a Smn deficiency does not cause widespread transcriptome changes (Zhang et al. 2013).

Low levels of mapping were observed for Control SC samples (29% and 36% mapping of Sample 1 and 2, respectively) to the zebrafish genome. This could be either due to low quality of cDNA libraries or degeneration of libraries during sequencing. Neither of them was plausible reason, as the quality was assessed for libraries prior to sequencing as well sequenced reads, analysed by FastQC reports. Both of the QC results were good. Contamination of NGS data with microbial genome sequences has been reported earlier (Lusk 2014, Strong et al. 2014), which lowers the percentage of mapped reads to reference genome. Since the quality of mapped reads was good, the data was used for downstream analysis.

GO enrichment analysis was performed on the differentially expressed genes from both cell types to determine the biological relevance of these genes. 166 GO terms were found to be significantly over-represented ( $p$ -value $<0.05$ ) among the deregulated genes (both up- and down-regulated) in SMA MN. The most enriched GO term identified was “calcium ion binding”, indicating that Smn deficiency affects genes that encode for calcium ion-interacting proteins. Reduction in SMN has been previously reported to affect regulation

of intraterminal  $Ca^{2+}$  levels (Ruiz et al. 2010), indicating a possible deregulation of calcium ion binding proteins. In a zebrafish SMA model, reduced levels of Smn causes significant reduction of evoked  $Ca^{2+}$  influx into the pre-synaptic axon terminal (See et al. 2014). A total of 26 genes in the list were found to be related to “calcium ion binding” (Table 18). 15 out of these 26 genes were down-regulated, suggesting that loss of Smn leads to deficiencies in these proteins, and consequently, reduced  $Ca^{2+}$  levels in pre-synaptic axon terminals. A recent study on RNAseq of SMA mESC-derived motor neurons, reported down-regulation of transcripts related to neuron development and up-regulation of pluripotency and cell proliferation genes (Maeda et al. 2014). bHLH transcription factors like *ascl1a*, *neurog1*, *olig2*; *fox* genes, *pax6b*, *fezf2* among others that have been reported to be involved in neuronal cell fate specification and commitment were differentially expressed. A total of 75 genes related to developmental processes were found to be deregulated in SMA MN (Table 18). Taken together, this indicates a possible effect of Smn reduction on neuronal development in zebrafish.

**Table 18. List of deregulated genes in SMA MN related to calcium ion binding and developmental process**

Genes	GO term
actn3a, casq1a, casq1b, casq2, dld, jag1a, mmp13a, myl1, mylz3, pvalb2, pvalb3, s100z, tnnc2, zgc:162595, zgc:66455 <i>anxa13, capn2a, cdh15, efemp2b, myl4, myl9a, oc90, pcdh8, pls3, s100v2, scgn</i>	<b>Calcium ion binding</b>
lft1, foxg1d, rtn4rl2a, neurog1, camk2a, tnnt3b, ascl1a, zic5, rx2, olig2, smyd1a, fezf2, tnnc2a.4, sox21b, tspan12, rx1, six3b, murca, dld, npas4a, stac3, apobec2a, pou3f3a, pax6b, fzd5, tbx2b, cx43, grem2b, vax1, dmrt3a, jag1a, hoxa13a, eng1b, apela, sfrp1a, mibp2 <i>kdr1, csad, acvr11, gata1a, parp3, optn, ptpreb, ttpa, smad9, sox7, socs1a, fsta, trim2a, kremen1, arrb1, rbm24a, mmp9, unc45a, foxi1, tagln2, snrkb, dmrt2a, klf17, epor, jam2a, oc90, nfe2, casp8, mink1, hapln4, fn1b, apoeb, dkk1b, wnt1, wnt10b, ap1m3, tectb, mespba, itga2b</i>	<b>Developmental process</b>

Gene names in black represent down-regulated genes while red represents up-regulated genes.

Genes affected in SMA SC showed enrichment of terms related to development and differentiation, indicating Smn deficiency affects similar pathways in Schwann cells and motor neurons. Although the differentially expressed genes between the two cell types were not similar (only 27 overlapping genes), the affected processes appeared to be similar. Neuropeptide signalling pathway was also found to be affected in SMA SC, indicating defects in axon-glia communication. Neurotransmission from nerve terminals leads to increase in intracellular Ca<sup>2+</sup> levels in Schwann cells as well as regulates their gene expression (Jahromi et al. 1992, Georgiou et al. 1994). Taken together, Smn deficiency leads to Ca<sup>2+</sup> deregulation in both, motor neurons and Schwann cells, affecting their activity at NMJs. Apart from this, the RNAseq data from both cell types identified interesting genes that would allow elucidating new pathways involved in neuron-Schwann cell communication.

### **4.3 Deficiency of Smn and Ighmbp2 affects genes related to axonogenesis**

Loss of IGHMBP2 due to mutations in the protein coding gene leads to motor neuron degeneration and ultimately to DSMA1 (Grohmann et al. 2001). Ighmbp2 deficient mouse motor neurons showed defects in axon outgrowth and pathfinding in culture (Krieger et al., unpublished data) similar to the axonal defects observed in zebrafish *ighmbp2* morphants *in vivo*. The axonal defects were significantly increased in *ighmbp2* morphants (30% defective axons) as compared to control or uninjected embryos (10% and 9.3%, respectively). Interestingly, these axonal defects in *ighmbp2* morphants were similar to those observed in *smn* and *nrxn2a* morphant zebrafish embryos (See et al. 2014). It has been reported that axons in *smn* and *nrxn2a*

morphants pause at the choice point for a longer duration than normal and eventually grow ventrally towards the second intermediate target (McWhorter et al. 2003, See et al. 2014). However, they remain truncated and eventually sprout branches at later stages. A similar axonal out-growth pattern can be speculated in case of *lghmbp2* morphants.

In order to understand what causes such defects, RNAseq data were generated from *lghmbp2* deficient FAC sorted motor neurons and control embryos. Differential gene expression analysis identified 1,666 affected genes with fold change >2. Since the data were generated for n=1, p-values could not be determined, hence the large number of differentially regulated genes. Out of the 1,666 differentially expressed genes, 740 were down-regulated and 926 were up-regulated. To analyse the genes commonly regulated in both SMA MN and DSMA1 MN, the differentially expressed gene lists in both cases were compared. A total of 93 genes were found commonly regulated in both disease conditions. Interestingly, GO analysis of these overlapping genes identified terms like “regulation of axonogenesis” and “regulation of axon guidance” among the top most terms. This suggests that reduced levels of *Smn* and *lghmbp2*, directly or indirectly, affect genes that are important for axon out growth and guidance. It could also be speculated that *Nrxn2a* deficiency could also affect similar genes. Taken together, *lghmbp2* and *Smn* deficiency leads to motoaxonal defects in zebrafish models for SMA and DSMA1 due to deregulation of axonogenesis related genes.

#### **4.4 Identification of alternatively spliced transcripts in SMA motor neurons and Schwann cells**

Reduced snRNP levels, in SMN deficiency, not necessarily lead to aberrant splicing of pre-mRNAs, alternative splicing is also possible (Zhang et al. 2008). This could lead to changes in the ratio of the two spliced isoforms, generation of aberrantly spliced isoforms not found in unaffected controls or do not correspond to any known isoform (Zhang et al. 2008). RNAseq, as mentioned earlier, has inherent advantages over traditional microarrays in identification of novel splice isoforms. Microarrays have been used to identify splicing changes in SMA mouse models (Zhang et al. 2008, Baumer et al. 2009) as well as SMN-depleted neuroblastoma cells (Huo et al. 2014). Zhang et al. reported that SMN deficiency affects the snRNPs repertoire of each cell uniquely, thus producing cell type-specific effects on splicing (Zhang et al. 2008). There has been an ongoing debate on whether SMN deficiency has any preferential effect on the minor spliceosome splicing pathway (Gabanella et al. 2007). More recently, RNAseq has been employed to analyse the splicing changes as well as to identify spurious or novel splice isoforms (Doktor et al. 2016). This is majorly due to the inherent advantages of RNAseq over traditional microarray that lack complete coverage of whole transcriptome with high efficiency. A very recent RNAseq study in the SMA mouse model showed increased splicing defects in U12-dependent introns, which leads to dysregulation of genes related to neuronal function (Doktor et al. 2016). Alternative splicing analysis, performed using the MATS program, identified 284 significant AS events (FDR <0.1) in SMA MN while 39 events were identified in SMA SC. Exon skipping is the most prevalent type of alternative splicing in vertebrates and invertebrates and is believed to contribute most to phenotypic complexity (Sugnet et al. 2004, Kim et al.

2008). In case of SMA MN and SMA SC, the majority of AS events were due to exon skipping. AS analysis identified only a small number of splicing changes in SMA SC which could be attributed to the relatively low alignment of Control SC reads (29% and 36% mapping of Sample 1 and 2, respectively) to the zebrafish genome. Interestingly, out of these 39 AS events, 29 were similar to those identified in SMA MN while only 10 were unique to SMA SC. Due to high similarity of the AS events in both cell types, the enriched GO terms were also found to be similar (data not shown). This indicates that *Smn* deficiency leads to splicing defects in similar genes in both cell types.

A total of 284 and 39 splicing events (in 248 and 37 genes, respectively) were identified in SMA MN and SC, respectively. Most of these did not overlap with differentially expressed genes (only five genes in motor neurons and one gene in Schwann cell were common). This finding suggests that most of the alternative splicing events affect the relative abundance of various isoforms but not the overall transcript level (Zhang et al. 2013). These AS genes identified in SMA MN, like differentially expressed gene list, were found to be very different from earlier studies in other SMA animal models with only a few common genes (Zhang et al. 2008, Baumer et al. 2009, Zhang et al. 2013, Huo et al. 2014, Ng et al. 2015, Doktor et al. 2016).

An interesting uncharacterised gene, *si:ch211-225b11.1*, was identified to be alternatively spliced in SMA MN. This gene is a homolog of *Drosophila inebriated (ine)*, which encodes a Na<sup>+</sup>/Cl<sup>-</sup>-dependent neurotransmitter transporter protein Ine. *ine* mutant flies show increased excitability of motor neurons due to defective re-uptake of the neurotransmitter substrate leading to overstimulation of motor neurons (Soehnge et al. 1996, Huang et al. 2002). Alternative splicing in *si:ch211-225b11.1* leads to inclusion of exon 7 in SMA MN which creates a pre-mature stop codon leading to the formation of a

truncated protein of 338 aa instead of 609 aa. This AS destroys the Sodium:neurotransmitter symporter domain and could lead to defects in zebrafish motor neuron excitability, similar to *Drosophila*. This makes *si:ch211-225b11.1* a potential candidate gene that could be implicated in SMA pathology.

#### **4.4.1 Smn deficiency affects pre-synaptic proteins in motor neurons**

Neurexins are cell adhesion molecules found at the pre-synaptic terminals that interact with post-synaptic neuroligins to trigger synapse formation (Dean et al. 2003). They exist in over 1000 isoforms due to alternative splicing causing different binding properties, thus regulating synapse specificity and function (Missler and Sudhof 1998, Rowen et al. 2002, Iijima et al. 2011). Alternative splicing at SS1 and SS3 of *Nrxn2a* in rat neurons has been shown to be caused by depolarisation and  $Ca^{2+}$  levels inside neurons (Rozić-Kotliroff and Zisapel 2007). Activity-dependent alternative splicing at SS4 of *Nrxn1* mediated by SAM68 has also been reported (Iijima et al. 2011). A recent study in the zebrafish SMA model reported alternative splicing of *neurexin 2aa* (*nrxn2aa*) at splice site 3 (SS3) in response to reduced Smn levels (See et al. 2014). Alternative splicing of *nrxn1a* at SS3 was identified in SMA MN leading to inclusion of exon at SS3. SS3 alternative splicing in *nrxn1a*, similar to *nrxn2aa*, could also be dependent on  $Ca^{2+}$  levels and modulated by depolarisation of neurons (See et al. 2014). See et al. reported down-regulation and alternative splicing of *nrxn2aa* in *smn* morphants (See et al. 2014), however, RNAseq data from SMA MN reveal no change in *nrxn2aa* splicing as well as mRNA levels. This could be due to the fact that *nrxn2aa* was shown to be expressed in other neurons apart from motor neurons (See et al. 2014). This suggests that *nrxn2aa* remains unaffected under Smn deficiency in zebrafish motor neurons, but not in other neurons. Alternative

splicing of another cell adhesion molecule coding gene, *cell adhesion molecule 1a (cadm1a)*, was also identified in SMA MN. CADM1 encodes a synaptic cell adhesion molecule SynCAM1 in vertebrates which has a role in assembly of synapses (Biederer et al. 2002, Fogel et al. 2007). Taken together, these provide evidence of a link between the splicing defects and motor axon defects observed under reduced Smn levels. A recent study in the SMA mouse model also reported that splicing and other mRNA level changes affect specific genes that are critical for motor neuron synapses, providing evidence for a link between splicing abnormality and motoaxonal dysfunction (Zhang et al. 2013).

#### **4.4.2 Smn deficiency leads to alternative splicing of genes involved in axonal transport**

Intracellular transport or axonal transport is very important for the maintenance and function of neurons because of their extreme polarity and size. Axonal transport supplies proteins and small organelles to the axons and nerve terminals and at the same time also clears out misfolded proteins from axons to avoid any toxic build up (Perlson et al. 2010). Transport is also critical for intracellular neural transmission and allows neurons to respond to trophic signals and stress (Perlson et al. 2010). Defects in axonal transport have been linked to a variety of neurodegenerative diseases such as Alzheimer, Parkinson, Huntington, ALS and others (Perlson et al. 2010, Millecamps and Julien 2013). In SMA $\Delta$ 7 mice, defective axonal transport of synaptic vesicle 2 and synaptotagmin proteins leading to synaptic transmission defects have been reported (Dale et al. 2011). Motor neurons derived from SMN deficient hESCs as well as from SMA1 patient iPSCs shows abnormal mitochondrial transport as an early pathological change (Xu et al. 2016).

Schwann cells, like motor neurons, polarise their surface into multiple membrane domains by sorting and targeting specific proteins to these domains (Trapp et al. 1981, Trapp and Quarles 1982, Cornbrooks et al. 1983). These proteins after being synthesised in cytoplasm, are sorted using trans-golgi network, transported by carrier vesicles to their target sites using microtubules (Trapp et al. 1995). Vesicular transport has been reported to play a critical role in “glia-to-axon” communication (Lopez-Verrilli and Court 2012). Polyribosomes from Schwann cells are secreted in the form of microvesicles and transported to the adjoining axons during nerve injury and regeneration (Court et al. 2008, Court et al. 2011, Lopez-Verrilli and Court 2012, Lopez-Verrilli et al. 2013). Interestingly, alternative splicing analysis in SMA MN and SC revealed some interesting transport related genes that could be implicated in SMA pathology.

The RAB family of GTPase proteins has been reported to be involved in regulation of vesicular transport (Pfeffer 1994, Hutagalung and Novick 2011). Mouse Rab7 plays an important role in retrograde axonal transport of the neurotrophin Nerve growth factor (NGF) (Hendry et al. 1974, Saxena et al. 2005, Deinhardt et al. 2006). The neurotrophin family of proteins have major roles in CNS and PNS including survival, development, nerve repair and synaptic plasticity (Huang and Reichardt 2001, Reichardt 2006). Thus, the retrograde transport of these proteins from axon tips to soma is very critical for neurons. Components of the retrograde transport routes, such as dynein and dynactin, which target neurotrophins and other survival signals to the soma have been implicated in motor neuron diseases in humans and mice (Hafezparast et al. 2003, Puls et al. 2003). Alternative splicing in the uncharacterised gene *zgc:91909*, encoding a novel protein similar to human RAB7, was identified in both SMA MN and SC. In both cell types, this leads to

generation of a non-functional truncated protein with incorrect amino acid sequence. This suggests reduction in functional protein from *zgc:91909* in both cell types. Rab15, another member of the RAB family, differentially regulates the early steps of endocytic trafficking and directs the membrane traffic from early endosomes to recycling endosomes for a transport back to the membrane (Zuk and Elferink 2000, Strick and Elferink 2005, Hutagalung and Novick 2011). Alternative splicing of *rab15* was identified in SMA MN, leading to exclusion of exon 5 from *rab15* transcripts under Smn deficiency. These results suggest that deregulation of axonal retrograde transport (i.e. by alternative splicing of *zgc:91909* and *rab15*) could affect processes important for SMA pathology.

Alternative splicing analysis also identified genes encoding members of the SNX family of proteins, to be affected by reduced Smn levels in zebrafish. SNX proteins are involved in intracellular processes like endocytosis, protein sorting and endosomal signalling (Cullen 2008). *snx9b* and *snx27a* were found to be alternatively spliced in both SMA MN and SC. *snx14*, however, was only found in SMA MN. Smn deficiency leads to an increase in exon 6 skipping of *snx9b* in both cell types, which was validated using RT-PCR experiments. A human homolog SNX9 is present in the presynaptic compartment of cultured hippocampal neurons, where it influences the synaptic vesicle endocytosis by interaction with dynamin 1 and N-WASP (Shin et al. 2007). SNX9 is also involved in clathrin-mediated endocytosis through its interaction with dynamin and clathrin (Lundmark and Carlsson 2003, Soulet et al. 2005). Interestingly exclusion of exon 6 in *snx9b* leads to creation of a pre-mature stop codon. This generates a truncated Snx9b protein of 188 aa with its WASP and SH3 binding domain (for dynamin interaction) missing. Taken together, a Smn deficiency could lead to Snx9b

protein dysfunction, which is possibly linked to defects in synaptic transmission, making it an interesting candidate in SMA pathology.

Sorting nexin 27 (SNX27) is a brain-enriched PDZ domain protein that has been shown to regulate the trafficking of certain G protein-gated potassium channels (Kir3), which are important for controlling neuronal excitability in the brain (Lunn et al. 2007). It is also important for maintaining glutamate receptors and is required for normal synaptic activity (Wang et al. 2013). Alternative splicing analysis revealed increased inclusion of exon 12 in SMA MN and SC, which was also confirmed by performing RT-PCR. The consequence of this splicing, however, is not expected to be severe as only the last two exons, with no known domains, were affected.

Taken together, these results point to a new mechanism, linking Smn deficiency-induced splicing defects with deregulated intracellular transport inside motor neurons and Schwann cells, which could contribute to SMA pathology.

#### **4.4.3 *srsf6b* as a novel candidate target of Smn**

Alternative splicing analysis of SMA MN and SC identified an interesting candidate, serine/arginine-rich splicing factor 6 (*srsf6b*), which encodes a member of the SR protein family which are important for both constitutive and alternative splicing of pre-mRNAs (Zahler et al. 1993, Mayeda et al. 1999). Studies have reported alternative splicing in splice factor coding genes under SMN deficient conditions that could potentially exacerbate the transcriptomic changes in SMA over time by inducing a second wave of splicing abnormalities (Zhang et al. 2013, Huo et al. 2014, Doktor et al. 2016). Alternative splicing of *Srsf5* and *Srsf10*, two other members of SR protein family, has recently been reported in the SMA mouse model and the splicing

change was found not only in the spinal cord, but also in brain, muscle and liver (Doktor et al. 2016). Skipping of exon 2 in *srsf6b* transcripts was identified by RNAseq in both SMA MN and SC and further validated using RT-PCR. This exon skipping leads to a truncated protein with only 37 aa, which is likely non-functional, suggesting a reduction in overall Srsf6b protein levels under Smn deficient conditions. Consequently, this would affect the normal functioning of Srsf6b protein in these two cell types. The expression pattern, as determined by in situ hybridisation, revealed that *srsf6b* is expressed in the whole brain and spinal cord, with highest levels in its anterior part. In the spinal cord at 31 hpf, *srf6b* expression was restricted to the ventral domain with expression in regions containing motor neurons and V2 interneurons, suggesting a possible role in neuronal differentiation.

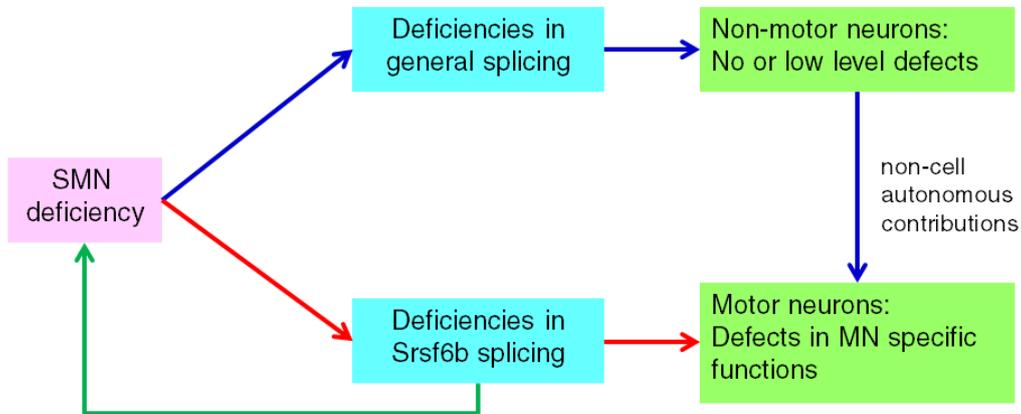
The function of *srsf6b* in motor neuron development was assessed by injecting CRISPRs targeting two sites in *srsf6b* gene, one each in exon 1 and intron 1. The efficiency of this *srsf6b* double CRISPR injection in generating mutations was very high (26 mutant embryos out of 46 injected; 56.52%), as determined by the presence of mutant bands in the injected embryos. This high efficiency allowed analysis of a mosaic phenotype directly in the injected embryos (in an approach called 'transient CRISPR mutagenesis'; (Willems et al. 2015)). Overall, the injected embryos were morphologically normal with low lethality when a 100 ng/μl dose of each gRNA was injected. However, some embryos developed a curved body curvature and delayed growth and several of them showed a *srsf6b* mutation by PCR. Being a splicing factor, Srsf6b is expected to be critical during development, thus explaining the morphological defects observed in some mutant embryos. *znp1* immunostaining was performed to assess the effect of a Srsf6b loss on motoaxons. Interestingly, 69.2% of the embryos showed more than one

motoaxonal defects. Motoaxonal defects observed in injected embryos after *znp1* immunostaining, provided further evidence of *Srsf6b*'s important role in motor neuron differentiation. At 31hpf, *srsf6b* showed expression mostly in the anterior part of spinal cord, suggesting a role for *Srsf6b* in fully developed motor neurons with NMJs already formed (Myers et al. 1986).

SR proteins are involved in pre-mRNA splicing of a wide variety of genes including the alternative splicing of *SMN1* and *SMN2*. SRSF1, SRSF2 and SRSF9 affect the inclusion of exon 7 in *SMN1* (Cartegni and Krainer 2002, Young et al. 2002, Kashmina and Manley 2003, Cartegni et al. 2006). Using HeLa cells and fibroblasts of SMA patients, Wee et al. reported a role of SRSF6 in modulating exon 7 splicing in *SMN* (Wee et al. 2014). However, it is presently unclear whether this is in a positive or negative manner in motor neurons. Based on this, there is a possibility that in motor neurons *Srsf6b* enhances exon 7 skipping in *smn*, suggesting a positive feedback and hence an 'amplification' of splicing defects in motor neurons. However, this needs to be addressed in future experiments.

A reduction of *Smn* leads to alternative splicing of *srsf6b* in motor neurons, resulting in overall lower levels of *Srsf6b* proteins. Also, *srsf6b* was found to be expressed in motor neurons and to be important for motor neuron differentiation. Taken together, I propose the following model to explain increased motor neuronal vulnerability in SMA (Fig. 21). According to this model, *Smn* deficiency leads to a general splicing deficiency, mediated by reduced snRNP levels, in all cells. Reduced levels of *Srsf6b* in SMA MN initiates a second wave of splicing deficiencies affecting particularly motor neuron-specific genes, leading to defects in motor neurons. On the other hand, general splicing deficiencies in non-motor neurons lead to no or low level defects in these cells, but possibly contribute non-cell autonomously to

cause motor neuron defects. Thus, a possible ‘amplifier mechanism’, consisting of deficiencies in general and *Srsf6b* splicing, exists in motor neurons making them especially vulnerable to *Smn* deficiency in SMA.



**Figure 21. ‘Amplifier mechanism’ to explain vulnerability of motor neurons to splicing defects in SMA.** Loss of *Smn* leads to deficiencies in general splicing as well as *srsf6b* splicing in SMA MN, leading to motor neuron-specific defects. In non-motor neurons, low level splicing defects caused due to general splicing deficiency contribute non-cell autonomously to motor neuron defects. *Srsf6b* enhances exon 7 skipping in *smn*. In *Smn* deficient conditions, low level of *Srsf6b* leads to increased *smn* exon 7 inclusion, elevating overall *Smn* protein levels. This consequently improves splicing of *srsf6b* and could possibly ameliorate the downstream splicing defects.

## 4.5 Future experiments

### Generation and characterisation of *srsf6b* mutant

The *znp1* immunostaining results presented in this thesis suggest a role of *Srsf6b* in motor neuron differentiation. However, this experiment was performed by a transient and mosaic *srsf6b* knock-out, and thus needs to be further corroborated with data from stable mutants. RNAseq data identified alternative splicing of *srsf6b* also in SMA SC, thus it would be interesting to test for possible Schwann cell defects in these mutants. By crossing the *srsf6b* mutants with motor neuron and Schwann cell calcium sensor lines, HB9:D3cpv/MN and HB9:D3cpv/SC, it would be interesting to analyse the

excitability and synaptic transmission of both cell types in a *srsf6b* mutant background. The proposed role of Srsf6b as a negative regulator of *smn* exon 7 inclusion can also be validated in the *srsf6b* mutants by analysing the splicing pattern of *smn* as well as determining changes in Smn protein levels. In this respect, I have already generated the gRNAs, targeting two sites in *srsf6b* gene, for injection into wild-type embryos.

### **Functional characterisation of novel genes affected in SMA MN and SC**

A major advantage of RNAseq is that it identifies novel genes and transcripts. RNAseq of SMA MN and SC identified a large number of previously uncharacterised novel genes affected by reduced Smn levels. It would be interesting to characterise these novel genes and analyse them in context of SMA. One such novel gene identified was *si:ch211-225b11.1*. The *Drosophila* homolog of this gene encodes for Na<sup>+</sup>/Cl<sup>-</sup> dependent neurotransmitter transporter protein and controls neuronal excitability. It will be interesting to characterise the expression and function of this gene in the context of motor neuron differentiation.

### **4.6 Conclusion**

In this study, I have optimised a protocol for performing RNAseq on FAC sorted motor neurons and Schwann cells from whole zebrafish embryos, which can easily be adapted for other cell types or tissues. Using this protocol, the transcriptomes of motor neurons and Schwann cells under disease conditions were analysed. Defects in axonogenesis related genes were identified as a common denominator of motoaxonal defects in SMA and DSMA1 diseases. The alternative splicing data presented here strongly suggests that Smn deficiency leads to splicing and other mRNA expression

abnormalities in motor neurons, which primarily affect genes important for normal neuronal physiology including synapse formation and axonal transport. I report identification of the splicing factor *srsf6b* in zebrafish as a possible novel downstream target of Smn and propose an ‘amplifier mechanism’ of splicing deficiencies to explain motor neuron vulnerability in SMA. To the best of my knowledge, this is the first report analysing transcriptome wide splicing changes in the zebrafish SMA model.

#### **4.7 Transcriptome analysis of RNAseq data based on GRCz10 genome assembly**

During the thesis writing, all the bioinformatics analysis done and reported was based on Zv9 zebrafish reference genome assembly. However, after the submission of this thesis, an updated reference genome assembly was released called GRCz10. This updated reference genome has more than 1000 new clone sequences and improvements in order and orientation of assembly sequences. For this, a different set of bioinformatics tools and strategy was employed. Briefly, the raw RNAseq data for both motor neurons and Schwann cells was trimmed to remove adaptor sequences using Trimmomatic software (version 0.36) instead of CutAdapt tool. The commands for running trimmomatic are provided in Appendix 2. The trimmed RNAseq data was then mapped to GRCz10 zebrafish reference genome using Spliced Transcripts Alignment to Reference (STAR) software (version 2.4.1). Previously, TopHat was used for mapping of trimmed data with Zv9 genome assembly. The STAR aligner commands are provided in Appendix 2. The mapping % for all the libraries (Table 19) considerably improved with STAR-GRCz10 as compared to TopHat-Zv9 mapping (Table 11 and 17). The STAR aligned data (bam files) were then analysed using Partek Genomics

suite for quantification, differential gene expression and GO enrichment analysis as mentioned earlier.

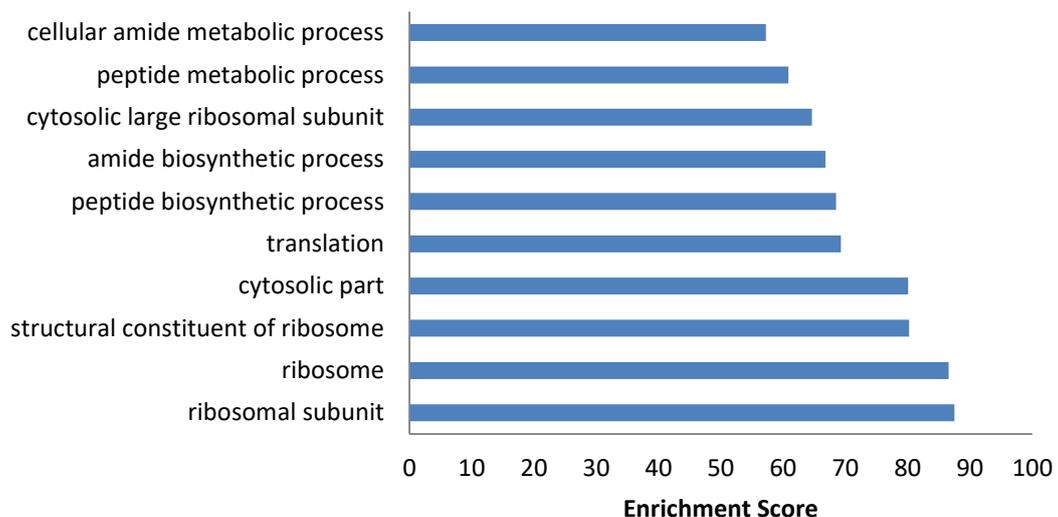
**Table 19. Comparison of mapping percentages between Zv9 and GRCz10 mapped data**

		Mapping % with Zv9 genome assembly	Mapping % with GRCz10 genome assembly
<b>Control MN</b>	Sample 1	74.6%	89.01%
	Sample 2	58.5%	82.22%
	Sample 3	59.1%	82.09%
<b>SMA MN</b>	Sample 1	74.1%	87.1%
	Sample 2	57.3%	80.52%
	Sample 3	59.6%	81.38%
<b>Control SC</b>	Sample 1	29.3%	85.72%
	Sample 2	36.7%	53.56%
<b>SMA SC</b>	Sample 1	61.3%	86.42%
	Sample 2	58.3%	81.66%

#### **(A) SMA MN**

In case of SMA MN (n=3), Partek identified 27,258 genes to be expressed in motor neurons at the 28 hpf stage. These genes were then filtered and low expressing genes (genes with RPKM values less than 1 in at least 4 out of 6 samples) were excluded. After filtering out low expression genes, the samples were analysed for clustering and Sample 1 of both Control MN and SMA MN were found to be clustering together and not with other Control and SMA MN samples. Ideally, all the controls should appear as one cluster and the test samples as another cluster. Clustering of Sample 1 in both control and SMA MN might be due to low level of significant difference among the two samples. Sample 1 from both Control MN and SMA MN was then rejected from the downstream analysis to better statistically significant genes among controls and SMA MN. The filtering reduced the overall genes to about 13,484 genes, out of which 223 genes were found to be significantly differentially expressed in SMA MN at a False Discovery rate (FDR) of 20% (0.2) and a fold change greater than 1.5. Out of the 223 genes, 117 were found to be down-regulated

while 106 were up-regulated in SMA MN. GO enrichment analysis identified 311 significantly enriched terms (p-value <0.05, Enrichment score >3). Most of the top GO terms were related to ribosome and translation indicating deregulation of genes related to them (Figure 22). A closer look at the differentially expressed gene list revealed more than 30 ribosomal protein coding genes, both for the small and large subunits, and all of them were down-regulated under Smn knockdown conditions. This indicates a possible defect in ribosomal assembly and translation in SMA MN. This was also supported by Pathway analysis using online resource David. Ribosome pathway was found to be significantly affected in down-regulated genes while Oxidative phosphorylation was the only enriched pathway among the up-regulated genes.



**Figure 22. Gene Ontology (GO) analysis of differentially expressed genes between SMN deficient and control motor neurons.** GO analysis was performed on differentially expressed genes with at least a 1.5-fold change and FDR less than 0.2 (20%). Top 10 GO terms enriched are represented in a bar graph with the Y-axis showing the enriched GO terms and the X-axis showing the respective enrichment score values.

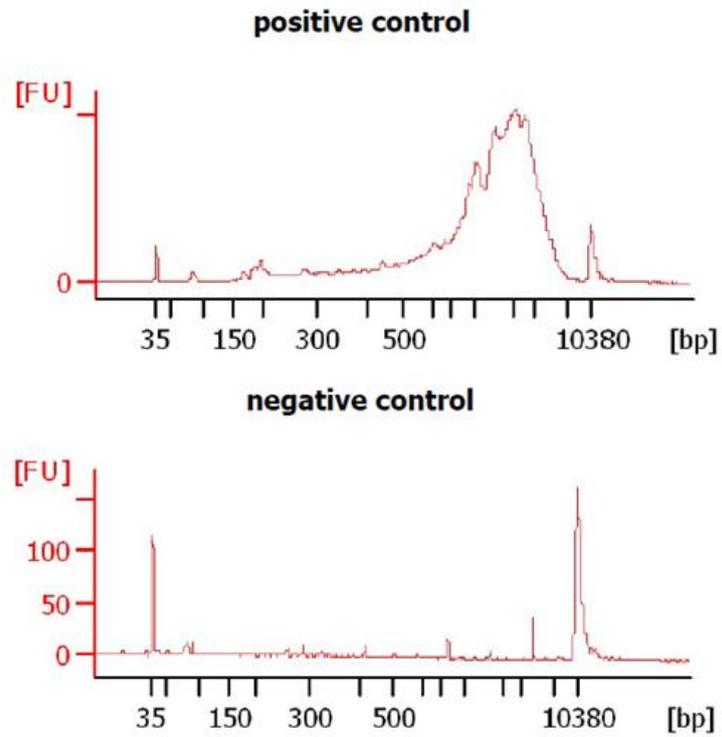
## **(B) SMA SC**

Partek software identified a total of 26,285 genes to be expressed in Schwann cells. Low expression genes (RPKM less than 1 in at least 3 out of 4 samples) were filtered out. 12,666 genes passed the criterion and were used for differential gene expression analysis. Out of 12,666 genes, 338 genes were found to be differentially expressed with fold change greater than 1.5 and p-value less than 0.05 (FDR 0.22). Among these differentially expressed genes, 159 genes were found to be down-regulated while 179 were up-regulated. On comparing the differentially expressed gene list from both cell types, only five genes were found to be commonly affected. *ascl1a*, *med14* and *her3* were down-regulated in both cell types while *itm2ba* was up-regulated and *prdm14* was down-regulated in Schwann cells but up-regulated in motor neurons. This again highlights an interesting observation that the SMN knockdown leads to tissue-specific gene expression changes.

GO analysis of differentially expressed genes in SMA SC identified a total of 271 terms to be significantly enriched (p-value < 0.05, enrichment score >3). Genes involved in developmental process, myosin filament assembly related processes, syntaxin-1 binding etc. were identified to be deregulated under *Smn* knockdown conditions. Alternative splicing analysis using the newly annotated GRCz10 assembly is currently underway for both cell types.

## APPENDIX

### Appendix 1. Amplified double stranded cDNA profiles for Positive control RNA and Negative control samples after SMARTer kit



Positive control cDNA generated using 1.3 ng of Control RNA (SMARTer kit)

## Appendix 2. RNAseq data processing and alignment software commands.

### (A) CutAdapt tool

```
cutadapt -b GATCGGAAGAGCACACGTCTGAACTCCAGTC -g
ACACTCTTTCCCTACACGACGCTCTTCCGATCT input.fastq > output.fastq
2> report.txt
```

Here in the above code “-b” attribute looks for the adaptor sequence at 5’ or 3’ end while “-g” looks for 5’ adaptor

input.fastq - Raw RNAseq reads with the adaptor sequence

output.fastq – Trimmed reads without adaptor sequence

### (B) TopHat mapping tool

```
#!/bin/bash
```

```
#PBS -j oe
```

```
#PBS -l walltime=720:00:00
```

```
#PBS -l nodes=1:ppn=4
```

```
export PATH=/cluster/apps/x86_64/packages/bowtie-2.2.3/bin:$PATH export
PATH=/cluster/apps/x86_64/packages/samtools-0.1.19/bin:$PATH
/cluster/apps/x86_64/packages/tophat-2.0.12/bin/tophat --GTF
/home/vyash/Danio_rerio_NCBI/NCBI/Zv9/Annotation/Genes/genes.gtf --
num-threads 8 --min-anchor 8 --splice-mismatches 1 --min-intron-length 50 --
max-intron-length 500000 --min-isoform-fraction 0.15 --max-multihits 1 --
segment-length 25 --segment-mismatches 2 --min-coverage-intron 50 --max-
coverage-intron 20000 --min-segment-intron 50 --max-segment-intron 500000
--keep-fasta-order --read-mismatches 3 --read-gap-length 3 --read-edit-dist 3
--read-realign-edit-dist 3 --max-insertion-length 3 --max-deletion-length 3 --
mate-inner-dist 200 --mate-std-dev 20 --no-coverage-search --library-type fr-
unstranded --output-dir /home/vyash/HVyas-RZN001-tophatoutput
/home/vyash/Danio_rerio_NCBI/NCBI/Zv9/Sequence/Bowtie2Index/genome
/home/vyash/RZN001-HV/RZN001-R1-trimmed.fastq /home/vyash/RZN001-
HV/RZN001-R2-trimmed.fastq
```

```
whoami
```

```
sleep 5
```

Default parameters were used for running TopHat with only few changes.

- mate-inner-dist : Distance between the mates (concordant paired-end reads) aligning to the genome. It is calculated as Fragment length minus two times the read length.
- mate-std-dev : The standard deviation for the distribution on inner distances between mate pairs

### (C) Trimmomatic tool

```
#!/bin/bash

#PBS -j oe

#PBS -l walltime=720:00:00

#PBS -l nodes=1:ppn=4

java -jar /home/vyash/Trimmomatic-0.36/trimmomatic-0.36.jar PE -trimlog
/home/vyash/RZN001/RZN001-removal-log.txt /home/vyash/RZN001/HS002-
PE-R00137_AC5F75ACXX.RZN001_CTTGTA_L003_R1.fastq
/home/vyash/RZN001/HS002-PE-
R00137_AC5F75ACXX.RZN001_CTTGTA_L003_R2.fastq
/home/vyash/RZN001/RZN001_CTTGTA_L003_forward_paired_long.fastq
/home/vyash/RZN001/RZN001_CTTGTA_L003_forward_unpaired_long.fastq
/home/vyash/RZN001/RZN001_CTTGTA_L003_reverse_paired_long.fastq /
home/vyash/RZN001/RZN001_CTTGTA_L003_reverse_unpaired_long.fastq
ILLUMINA_CLIP:Adaptor-file.fa:2:30:10:8:TRUE LEADING:3 TRAILING:3
SLIDINGWINDOW:4:15 MINLEN:36

whoami

sleep 5
```

A custom made adaptor file, **Adaptor-file.fa**, was generated that contained information on sequences that needs to be removed from all the samples. The adaptor file contained the following sequence information-

```
>PrefixPE/1
```

```
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTAT
GCCGTCTTCTGCTTG
```

```
>PrefixPE/2
```

```
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCG
CCGTATCATT
```

```
>PE1
```

```
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTAT
GCCGTCTTCTGCTTG
```

>PE1\_rc

CAAGCAGAAGACGGCATAACGAGATCGTGATGTGACTGGAGTTCAGACGT  
GTGCTCTTCCGATCT

>PE2

AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCG  
CCGTATCATT

>PE2\_rc

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC  
TTCCGATCT

#### **(D) STAR aligner tool**

```
#!/bin/bash
```

```
#PBS -j oe
```

```
#PBS -l walltime=720:00:00
```

```
#PBS -l nodes=1:ppn=4
```

```
/cluster/apps/x86_64/packages/STAR_2.4.1d/src/bin/Linux_x86_64_static/STAR --runThreadN 4 --genomeDir /home/vyash/STAR-Danio10-new/Danio-  
rerio-genome-dir-new --readFilesIn /home/vyash/RZN001-HV/  
RZN001_CTTGTA_L003_forward_paired_long.fastq /home/vyash/RZN001-  
HV/ RZN001_CTTGTA_L003_reverse_paired_long.fastq --outFilterType  
BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --  
alignSJDBoverhangMin 1 --alignIntronMin 20 --alignIntronMax 1000000 --  
outSAMtype BAM Unsorted SortedByCoordinate --outFileNamePrefix  
/home/vyash/RZN001-star-output/RZN001trimmed
```

```
whoami
```

```
sleep 5
```

### Appendix 3. List of differentially expressed genes in SMA MN

S. No.	Gene symbol	p-value	Fold Change	Regulation
1	<i>zgc:153759</i>	0.01821	30757.60	down
2	<i>tnni2a.1</i>	0.00852	21.09	down
3	<i>glis3</i>	0.01793	16.43	down
4	<i>lft1</i>	0.00284	16.03	down
5	<i>zp2l2</i>	0.00137	14.27	down
6	<i>si:dkey-11p23.7</i>	0.00698	11.99	down
7	<i>foxg1d</i>	0.02047	11.03	down
8	<i>lta</i>	0.00412	10.58	down
9	<i>zgc:195356</i>	0.02966	9.99	down
10	<i>slc38a8</i>	0.00548	8.75	down
11	<i>tssk6</i>	0.01694	8.57	down
12	<i>gsto1</i>	0.00896	7.92	down
13	<i>loxl5a</i>	0.00075	7.61	down
14	<i>hspb2</i>	0.04110	6.73	down
15	<i>crygm2d17</i>	0.01003	6.63	down
16	<i>nitr3c</i>	0.04813	6.22	down
17	<i>mettl11b</i>	0.02312	6.10	down
18	<i>pvalb3</i>	0.03452	5.58	down
19	<i>ifitm5</i>	0.02242	5.29	down
20	<i>rtn4rl2a</i>	0.01377	5.24	down
21	<i>tas2r3</i>	0.01896	5.23	down
22	<i>neurog1</i>	0.00731	4.91	down
23	<i>camk2a</i>	0.00259	4.88	down
24	<i>cpa5</i>	0.01614	4.83	down
25	<i>si:dkey-228a15.1</i>	0.03432	4.72	down
26	<i>s100z</i>	0.02611	4.59	down
27	<i>cdnf</i>	0.00107	4.57	down
28	<i>gnb3a</i>	0.00939	4.55	down
29	<i>lrrc20</i>	0.00450	4.54	down
30	<i>ugt2a1</i>	0.01250	4.48	down
31	<i>pvalb2</i>	0.04406	4.22	down
32	<i>gadd45gb.1</i>	0.02548	4.20	down
33	<i>si:ch211-152f23.5</i>	0.02775	4.16	down
34	<i>hoga1</i>	0.03192	4.16	down
35	<i>zgc:158296</i>	0.03575	4.09	down
36	<i>nme2b.2</i>	0.04330	3.95	down
37	<i>pde6h</i>	0.01088	3.86	down
38	<i>tgm2a</i>	0.03862	3.86	down
39	<i>dhrs7cb</i>	0.02528	3.79	down
40	<i>hsc70</i>	0.03104	3.78	down
41	<i>adh8a</i>	0.00622	3.73	down
42	<i>casq1b</i>	0.02529	3.71	down
43	<i>tnnt3b</i>	0.03209	3.69	down
44	<i>ascl1a</i>	0.00616	3.68	down
45	<i>zic5</i>	0.02122	3.61	down
46	<i>zgc:66455</i>	0.01260	3.60	down
47	<i>rx2</i>	0.04132	3.59	down
48	<i>myl1</i>	0.03618	3.55	down
49	<i>zgc:174931</i>	0.01112	3.54	down
50	<i>olig2</i>	0.00032	3.54	down
51	<i>ckmb</i>	0.02980	3.52	down
52	<i>smyd1a</i>	0.02125	3.49	down
53	<i>mylz3</i>	0.04338	3.47	down
54	<i>cyp1b1</i>	0.01890	3.47	down
55	<i>prom1b</i>	0.04727	3.41	down
56	<i>pygma</i>	0.04600	3.37	down
57	<i>casq1a</i>	0.00373	3.33	down

S. No.	Gene symbol	p-value	Fold Change	Regulation
58	<i>fezf2</i>	0.00227	3.30	down
59	<i>actc1b</i>	0.04051	3.29	down
60	<i>ckma</i>	0.04469	3.23	down
61	<i>arl3l2</i>	0.03485	3.22	down
62	<i>tnni2a.4</i>	0.02600	3.22	down
63	<i>myoz2a</i>	0.03393	3.22	down
64	<i>gsx1</i>	0.00386	3.15	down
65	<i>sox21b</i>	0.03705	3.10	down
66	<i>zgc:110045</i>	0.00155	3.08	down
67	<i>tnnc2</i>	0.04595	3.03	down
68	<i>zgc:173585</i>	0.02855	3.01	down
69	<i>mybpc2b</i>	0.01510	3.00	down
70	<i>rtn2b</i>	0.00580	2.98	down
71	<i>tpma</i>	0.02934	2.97	down
72	<i>vcla</i>	0.01696	2.94	down
73	<i>tspan12</i>	0.02997	2.92	down
74	<i>hhatla</i>	0.01275	2.89	down
75	<i>rx1</i>	0.04731	2.88	down
76	<i>six3b</i>	0.00084	2.88	down
77	<i>murca</i>	0.02183	2.85	down
78	<i>cldn8</i>	0.04820	2.85	down
79	<i>ctssa</i>	0.04103	2.85	down
80	<i>actn3a</i>	0.02073	2.76	down
81	<i>gapdh</i>	0.04016	2.75	down
82	<i>ak1</i>	0.02303	2.69	down
83	<i>olfm2b</i>	0.01121	2.67	down
84	<i>zgc:162356</i>	0.00785	2.67	down
85	<i>dld</i>	0.00401	2.62	down
86	<i>npas4a</i>	0.00492	2.59	down
87	<i>zgc:162595</i>	0.01228	2.58	down
88	<i>stac3</i>	0.03101	2.57	down
89	<i>apobec2a</i>	0.02191	2.55	down
90	<i>nr5a5</i>	0.02680	2.55	down
91	<i>bhlhe22</i>	0.04264	2.54	down
92	<i>aqp4</i>	0.03332	2.50	down
93	<i>si:dkey-23c22.6</i>	0.01270	2.47	down
94	<i>si:ch211-170d8.2</i>	0.00053	2.42	down
95	<i>si:dkey-238o13.4</i>	0.03789	2.42	down
96	<i>hspb6</i>	0.00249	2.42	down
97	<i>ahr1b</i>	0.04587	2.41	down
98	<i>foxd1l</i>	0.00877	2.41	down
99	<i>pou3f3a</i>	0.04802	2.40	down
100	<i>pax6b</i>	0.00110	2.40	down
101	<i>fzd5</i>	0.01483	2.39	down
102	<i>sst1.1</i>	0.04195	2.37	down
103	<i>tbx2b</i>	0.02647	2.36	down
104	<i>mmp13a</i>	0.04201	2.36	down
105	<i>pde6a</i>	0.00030	2.35	down
106	<i>si:ch211-243a20.3</i>	0.01151	2.34	down
107	<i>si:ch211-251b21.1</i>	0.03428	2.34	down
108	<i>cx43</i>	0.00467	2.31	down
109	<i>grem2b</i>	0.02388	2.28	down
110	<i>casq2</i>	0.03795	2.26	down
111	<i>zgc:172079</i>	0.03948	2.26	down
112	<i>znf362a</i>	0.01593	2.26	down
113	<i>rtn4ip1</i>	0.01120	2.25	down
114	<i>vax1</i>	0.00948	2.21	down
115	<i>lbx1b</i>	0.01099	2.21	down
116	<i>dmrt3a</i>	0.00069	2.20	down

S. No.	Gene symbol	p-value	Fold Change	Regulation
117	<i>jag1a</i>	0.02254	2.20	down
118	<i>zgc:158463</i>	0.00086	2.17	down
119	<i>sc:d217</i>	0.03683	2.16	down
120	<i>p2rx5</i>	0.04061	2.16	down
121	<i>hoxa13a</i>	0.03985	2.15	down
122	<i>wfdc1</i>	0.02289	2.14	down
123	<i>eng1b</i>	0.01893	2.12	down
124	<i>med14</i>	0.00745	2.12	down
125	<i>eno3</i>	0.02714	2.10	down
126	<i>dct</i>	0.02848	2.09	down
127	<i>hdhd1</i>	0.00739	2.08	down
128	<i>apela</i>	0.01265	2.07	down
129	<i>sfrp1a</i>	0.00740	2.06	down
130	<i>zc3h10</i>	0.02535	2.04	down
131	<i>mycn</i>	0.02849	2.03	down
132	<i>cmyb</i>	0.00851	2.03	down
133	<i>mibp2</i>	0.00842	2.00	down
134	<i>nipsnap1</i>	0.02462	55.72	up
135	<i>kcnj8</i>	0.00468	52.65	up
136	<i>mespbb</i>	0.02217	39.23	up
137	<i>itga2b</i>	0.00832	19.60	up
138	<i>mespba</i>	0.00336	19.49	up
139	<i>zgc:194733</i>	0.00699	16.58	up
140	<i>rhbg</i>	0.02410	14.45	up
141	<i>rad211</i>	0.00167	12.33	up
142	<i>paqr8</i>	0.02610	10.83	up
143	<i>cfb</i>	0.04887	9.86	up
144	<i>si:ch211-12h2.6</i>	0.01577	9.29	up
145	<i>isg15</i>	0.00941	9.03	up
146	<i>utrnp</i>	0.01844	7.14	up
147	<i>tectb</i>	0.01871	7.02	up
148	<i>calhm2</i>	0.02584	6.74	up
149	<i>EIF2AK2</i>	0.00142	6.70	up
150	<i>cyp2k16</i>	0.04218	6.57	up
151	<i>slc27a1b</i>	0.03850	6.43	up
152	<i>apoc1l</i>	0.00118	6.29	up
153	<i>slc9a2</i>	0.00185	6.25	up
154	<i>ela2l</i>	0.01147	5.71	up
155	<i>ap1m3</i>	0.01836	5.68	up
156	<i>zgc:101699</i>	0.00167	5.58	up
157	<i>wnt10b</i>	0.01059	5.46	up
158	<i>wnt1</i>	0.01209	5.42	up
159	<i>si:ch211-14a17.6</i>	0.03936	5.31	up
160	<i>eva1bb</i>	0.02054	4.78	up
161	<i>mat2al</i>	0.00688	4.78	up
162	<i>trhr2</i>	0.01348	4.75	up
163	<i>erbb3a</i>	0.04388	4.67	up
164	<i>dkk1b</i>	0.01314	4.67	up
165	<i>zgc:162324</i>	0.01894	4.66	up
166	<i>ppil6</i>	0.01128	4.55	up
167	<i>ptgdsa</i>	0.01497	4.50	up
168	<i>zgc:65811</i>	0.02279	4.47	up
169	<i>cdh15</i>	0.02448	4.42	up
170	<i>apoeb</i>	0.00152	4.38	up
171	<i>rpz5</i>	0.00385	4.35	up
172	<i>ccdc106b</i>	0.00646	4.27	up
173	<i>si:ch211-14a17.7</i>	0.02613	4.10	up
174	<i>lrrc48</i>	0.00198	3.97	up
175	<i>plxnb2b</i>	0.01036	3.95	up

S. No.	Gene symbol	p-value	Fold Change	Regulation
176	<i>upb1</i>	0.02984	3.88	up
177	<i>fam129aa</i>	0.04582	3.79	up
178	<i>slc5a2</i>	0.04235	3.76	up
179	<i>hepacama</i>	0.02629	3.74	up
180	<i>fn1b</i>	0.00018	3.74	up
181	<i>ddx43</i>	0.01340	3.67	up
182	<i>si:ch211-81a5.5</i>	0.02457	3.55	up
183	<i>htra3a</i>	0.03111	3.51	up
184	<i>glrba</i>	0.00230	3.51	up
185	<i>pcdh8</i>	0.00462	3.38	up
186	<i>loc100536659</i>	0.02368	3.35	up
187	<i>zgc:153740</i>	0.01157	3.28	up
188	<i>acsl5</i>	0.00984	3.25	up
189	<i>zgc:113363</i>	0.01237	3.21	up
190	<i>nrm</i>	0.02060	3.20	up
191	<i>fbxo32</i>	0.01697	3.11	up
192	<i>si:dkey-261j4.5</i>	0.01274	3.10	up
193	<i>s100v2</i>	0.03128	3.07	up
194	<i>hapln4</i>	0.02525	3.07	up
195	<i>smkr1</i>	0.00547	3.07	up
196	<i>zgc:113337</i>	0.02348	3.05	up
197	<i>kcnk1a</i>	0.01861	3.04	up
198	<i>baiap211b</i>	0.01304	3.03	up
199	<i>si:dkey-222p3.1</i>	0.01833	3.02	up
200	<i>hmox1a</i>	0.00295	3.01	up
201	<i>mink1</i>	0.02653	2.98	up
202	<i>casp8</i>	0.03940	2.97	up
203	<i>zgc:195173</i>	0.02624	2.94	up
204	<i>rpe65b</i>	0.03594	2.94	up
205	<i>si:dkey-188i13.10</i>	0.04310	2.91	up
206	<i>gcca</i>	0.03785	2.88	up
207	<i>ccdc125</i>	0.03596	2.86	up
208	<i>efemp2b</i>	0.01026	2.84	up
209	<i>mgc172218</i>	0.02885	2.83	up
210	<i>nfe2</i>	0.02006	2.83	up
211	<i>lingo3a</i>	0.02723	2.82	up
212	<i>tspan15</i>	0.00502	2.80	up
213	<i>mdm1</i>	0.03314	2.76	up
214	<i>si:dkey-121a9.3</i>	0.04280	2.76	up
215	<i>si:ch211-222k6.3</i>	0.01442	2.76	up
216	<i>oc90</i>	0.01133	2.76	up
217	<i>urp2</i>	0.03162	2.75	up
218	<i>chrng</i>	0.02193	2.75	up
219	<i>jam2a</i>	0.03123	2.74	up
220	<i>epor</i>	0.04414	2.74	up
221	<i>sepw2b</i>	0.04551	2.74	up
222	<i>sesn3</i>	0.02784	2.70	up
223	<i>klf17</i>	0.01081	2.66	up
224	<i>hcst</i>	0.01637	2.66	up
225	<i>acot8</i>	0.01070	2.64	up
226	<i>ggh</i>	0.00480	2.61	up
227	<i>dmrt2a</i>	0.00344	2.60	up
228	<i>zgc:92326</i>	0.02318	2.60	up
229	<i>si:dkeyp-59c12.1</i>	0.04683	2.59	up
230	<i>pls3</i>	0.02885	2.58	up
231	<i>si:dkey-169i5.4</i>	0.00081	2.56	up
232	<i>snrkb</i>	0.01632	2.56	up
233	<i>zgc:101663</i>	0.00888	2.56	up
234	<i>lgi2b</i>	0.03168	2.55	up

S. No.	Gene symbol	p-value	Fold Change	Regulation
235	<i>scgn</i>	0.03943	2.50	up
236	<i>cx35.4</i>	0.01711	2.47	up
237	<i>tdrd6</i>	0.01782	2.46	up
238	<i>kcnc4</i>	0.02887	2.45	up
239	<i>eps8l1</i>	0.02633	2.45	up
240	<i>kcnk1b</i>	0.00410	2.45	up
241	<i>tagln2</i>	0.01354	2.45	up
242	<i>tmed1a</i>	0.01159	2.44	up
243	<i>ppp1r1c</i>	0.00198	2.44	up
244	<i>foxi1</i>	0.00821	2.44	up
245	<i>unc45a</i>	0.01104	2.43	up
246	<i>zp3a.2</i>	0.04623	2.42	up
247	<i>mmp9</i>	0.04253	2.42	up
248	<i>rbm24a</i>	0.02840	2.42	up
249	<i>zgc:153219</i>	0.03311	2.40	up
250	<i>pik3cg</i>	0.02360	2.38	up
251	<i>itm2bb</i>	0.01041	2.38	up
252	<i>arb1</i>	0.01255	2.37	up
253	<i>cyp3c1</i>	0.00206	2.36	up
254	<i>gnmt</i>	0.00323	2.36	up
255	<i>abhd6b</i>	0.02214	2.36	up
256	<i>myl4</i>	0.00946	2.34	up
257	<i>kremen1</i>	0.01810	2.34	up
258	<i>itcha</i>	0.04642	2.33	up
259	<i>trim2a</i>	0.03284	2.32	up
260	<i>fsta</i>	0.01466	2.31	up
261	<i>capn2a</i>	0.01954	2.30	up
262	<i>socs1a</i>	0.01474	2.29	up
263	<i>arhgdig</i>	0.01995	2.28	up
264	<i>zgc:92912</i>	0.03169	2.27	up
265	<i>fbxl22</i>	0.04956	2.26	up
266	<i>abcg4b</i>	0.02109	2.26	up
267	<i>phlda2</i>	0.02738	2.24	up
268	<i>sox7</i>	0.04464	2.24	up
269	<i>smad9</i>	0.01979	2.24	up
270	<i>ldlrp1b</i>	0.02442	2.22	up
271	<i>loc556326</i>	0.03008	2.22	up
272	<i>ttpa</i>	0.00304	2.22	up
273	<i>ptpreb</i>	0.00117	2.21	up
274	<i>csf2rb</i>	0.02916	2.21	up
275	<i>hsd3b1</i>	0.00257	2.21	up
276	<i>slc9a8</i>	0.00092	2.20	up
277	<i>rab36</i>	0.00990	2.20	up
278	<i>myl9a</i>	0.01923	2.19	up
279	<i>zgc:154077</i>	0.00882	2.18	up
280	<i>pxmp2</i>	0.03518	2.16	up
281	<i>rtn3</i>	0.00639	2.16	up
282	<i>timp2b</i>	0.02234	2.14	up
283	<i>si:ch211-237l4.6</i>	0.02600	2.14	up
284	<i>optn</i>	0.03697	2.13	up
285	<i>ttc25</i>	0.02459	2.13	up
286	<i>parp3</i>	0.02832	2.13	up
287	<i>acot11b</i>	0.01756	2.12	up
288	<i>serpinb1</i>	0.00250	2.10	up
289	<i>uck1</i>	0.02434	2.10	up
290	<i>lmcd1</i>	0.00417	2.09	up
291	<i>srsf7b</i>	0.00815	2.09	up
292	<i>prkg1a</i>	0.04795	2.09	up
293	<i>gata1a</i>	0.04864	2.09	up

S. No.	Gene symbol	p-value	Fold Change	Regulation
294	<i>slc27a6</i>	0.02240	2.08	up
295	<i>kcnj2a</i>	0.04862	2.07	up
296	<i>lrrc15</i>	0.02223	2.06	up
297	<i>acvr1l1</i>	0.03358	2.06	up
298	<i>ppp1r14aa</i>	0.03277	2.05	up
299	<i>cntn1b</i>	0.00443	2.05	up
300	<i>arf3a</i>	0.00616	2.05	up
301	<i>csad</i>	0.02771	2.04	up
302	<i>tnfaip2a</i>	0.01266	2.04	up
303	<i>gadd45ga</i>	0.03181	2.04	up
304	<i>anxa13</i>	0.00197	2.03	up
305	<i>skap2</i>	0.02760	2.03	up
306	<i>hsd3b7</i>	0.03565	2.03	up
307	<i>fuom</i>	0.01242	2.03	up
308	<i>kitlgb</i>	0.01167	2.02	up
309	<i>kdr1</i>	0.01777	2.02	up
310	<i>zgc:193681</i>	0.03643	2.02	up
311	<i>scrn2</i>	0.02459	2.01	up
312	<i>ppap2cb</i>	0.02415	2.00	up

#### Appendix 4. List of differentially expressed genes in SMA SC

S. No.	Gene symbol	p-value	Fold Change	Regulation
1	<i>mir137-2</i>	0.00029	679516000	down
2	<i>mir137-3</i>	0.00029	679516000	down
3	<i>mogat2</i>	0.04119	70.99	down
4	<i>mpz</i>	0.01056	60.35	down
5	<i>pfkmb</i>	0.01360	26.03	down
6	<i>nsmfb</i>	0.00378	24.01	down
7	<i>slc30a1b</i>	0.04665	21.47	down
8	<i>sox12</i>	0.03287	20.20	down
9	<i>fgf22</i>	0.00709	14.94	down
10	<i>olfm3a</i>	0.02106	12.29	down
11	<i>scpp8</i>	0.03430	11.48	down
12	<i>ifitm5</i>	0.01914	11.45	down
13	<i>zgc:91944</i>	0.02645	10.57	down
14	<i>cabp5b</i>	0.01111	10.56	down
15	<i>si:ch211-71m22.3</i>	0.01957	9.66	down
16	<i>tfap2d</i>	0.01957	9.66	down
17	<i>fbp2</i>	0.04358	9.26	down
18	<i>cldn10a</i>	0.03314	9.21	down
19	<i>zgc:171579</i>	0.03830	8.56	down
20	<i>her3</i>	0.01431	8.41	down
21	<i>glrba</i>	0.04471	7.60	down
22	<i>si:dkey-32m20.1</i>	0.00375	6.50	down
23	<i>fabp7b</i>	0.00826	6.45	down
24	<i>gnat1</i>	0.03507	6.11	down
25	<i>zgc:136336</i>	0.03117	5.73	down
26	<i>mpp5b</i>	0.04991	5.43	down
27	<i>bcl11aa</i>	0.04772	5.40	down
28	<i>glra1</i>	0.03602	5.04	down
29	<i>and3</i>	0.01010	5.03	down
30	<i>vil1</i>	0.04306	5.03	down
31	<i>serp2</i>	0.03473	5.01	down
32	<i>ptf1a</i>	0.00156	4.99	down
33	<i>chgb</i>	0.03735	4.51	down
34	<i>usp43a</i>	0.04205	4.27	down
35	<i>zgc:114175</i>	0.00474	4.18	down
36	<i>prss35</i>	0.03827	4.15	down
37	<i>dgat1a</i>	0.04250	4.13	down
38	<i>pdyn</i>	0.02786	4.11	down
39	<i>zgc:103438</i>	0.03259	4.11	down
40	<i>si:dkey-228a15.1</i>	0.02110	4.10	down
41	<i>bmp8a</i>	0.01624	4.03	Down
42	<i>itga11a</i>	0.03257	3.99	Down
43	<i>scn12aa</i>	0.04025	3.84	Down
44	<i>saga</i>	0.01180	3.82	down
45	<i>zgc:56231</i>	0.00796	3.69	down
46	<i>ppox</i>	0.02045	3.61	down
47	<i>foxq1b</i>	0.00510	3.57	down
48	<i>figf</i>	0.01039	3.55	down
49	<i>pvalb1</i>	0.02603	3.52	down
50	<i>ccbe1</i>	0.00439	3.47	down
51	<i>cx39.9</i>	0.04389	3.43	down
52	<i>phkg1b</i>	0.00297	3.40	down
53	<i>insb</i>	0.02628	3.37	down
54	<i>tnfsf13b</i>	0.01659	3.33	down
55	<i>cbln1</i>	0.03988	3.28	down
56	<i>grin2bb</i>	0.00223	3.02	down
57	<i>ascl1a</i>	0.00288	3.00	down
58	<i>zgc:158423</i>	0.02038	2.97	down
59	<i>crabp1b</i>	0.01991	2.95	down

S. No.	Gene symbol	p-value	Fold Change	Regulation
60	<i>lrrc20</i>	0.03633	2.91	down
61	<i>ascl1b</i>	0.03650	2.84	down
62	<i>neurod4</i>	0.02242	2.84	down
63	<i>foxf2a</i>	0.01707	2.80	down
64	<i>pvalb2</i>	0.00905	2.79	down
65	<i>onecut1</i>	0.01203	2.77	down
66	<i>zgc:193807</i>	0.02103	2.75	down
67	<i>lgi1b</i>	0.01362	2.75	down
68	<i>nrsn1</i>	0.03929	2.74	down
69	<i>zgc:158296</i>	0.01088	2.72	down
70	<i>jph2</i>	0.02593	2.69	down
71	<i>rbfox1l</i>	0.00401	2.67	down
72	<i>si:ch211-245j22.3</i>	0.03795	2.62	down
73	<i>krt96</i>	0.03989	2.61	down
74	<i>actc1b</i>	0.00361	2.61	down
75	<i>vcam1</i>	0.03253	2.60	down
76	<i>si:ch211-155m12.1</i>	0.01904	2.58	down
77	<i>smyd1a</i>	0.01059	2.55	down
78	<i>tnni2a.4</i>	0.02791	2.54	down
79	<i>rell2</i>	0.02682	2.54	down
80	<i>tlx2</i>	0.00793	2.53	down
81	<i>slc1a3b</i>	0.00436	2.51	down
82	<i>pou4f1</i>	0.01116	2.50	down
83	<i>emx2</i>	0.03232	2.49	down
84	<i>srt1a</i>	0.02704	2.49	down
85	<i>gbx2</i>	0.00917	2.47	down
86	<i>caprin2</i>	0.01767	2.46	down
87	<i>ckma</i>	0.02277	2.45	down
88	<i>alox5a</i>	0.04358	2.45	down
89	<i>crygm2d5</i>	0.01278	2.44	down
90	<i>foxd1</i>	0.03965	2.44	down
91	<i>rem2</i>	0.00820	2.41	down
92	<i>myl1</i>	0.00951	2.39	down
93	<i>foxc1a</i>	0.04491	2.37	down
94	<i>ak1</i>	0.01652	2.35	down
95	<i>fbx17</i>	0.00423	2.34	down
96	<i>ckmb</i>	0.03136	2.34	down
97	<i>gpr27</i>	0.03390	2.33	down
98	<i>cd151</i>	0.00476	2.32	down
99	<i>pdlim5a</i>	0.02377	2.32	down
100	<i>nhlh2</i>	0.03855	2.31	down
101	<i>wnt9a</i>	0.00447	2.28	down
102	<i>nme4</i>	0.02162	2.26	down
103	<i>mylz3</i>	0.01389	2.25	down
104	<i>hey2</i>	0.03236	2.25	down
105	<i>wu:fc66h01</i>	0.03338	2.24	down
106	<i>gadd45gb.1</i>	0.02522	2.24	down
107	<i>gstr</i>	0.01277	2.22	down
108	<i>rbfox1</i>	0.02586	2.22	down
109	<i>snap25a</i>	0.02974	2.22	down
110	<i>kif3cb</i>	0.00905	2.22	down
111	<i>rtn1b</i>	0.00215	2.22	down
112	<i>si:busm1-160c18.6</i>	0.02432	2.21	down
113	<i>syn1</i>	0.03649	2.21	down
114	<i>phyhiplb</i>	0.04577	2.19	down
115	<i>hoxb5b</i>	0.01904	2.13	down
116	<i>tpma</i>	0.04070	2.12	down
117	<i>nfil3-2</i>	0.03679	2.11	down
118	<i>dpf1</i>	0.03994	2.11	down
119	<i>palm1a</i>	0.00563	2.10	down
120	<i>mid1ip1b</i>	0.02322	2.09	down
121	<i>tmod4</i>	0.01139	2.07	down

S. No.	Gene symbol	p-value	Fold Change	Regulation
122	<i>si:dkey-23c22.6</i>	0.03584	2.06	down
123	<i>mylpfb</i>	0.03377	2.05	down
124	<i>crispld1b</i>	0.02771	2.04	down
125	<i>acta1b</i>	0.02828	2.04	down
126	<i>olig4</i>	0.04823	2.04	down
127	<i>bcat1</i>	0.00914	2.02	down
128	<i>sept5b</i>	0.00242	2.02	down
129	<i>tagln3b</i>	0.00952	2.00	down
130	<i>dpp4</i>	0.01460	75.86	up
131	<i>cfhl2</i>	0.02403	54.25	up
132	<i>trpa1b</i>	0.01648	53.73	up
133	<i>lrp2bp</i>	0.00258	49.65	up
134	<i>c8g</i>	0.04862	18.48	up
135	<i>zgc:66449</i>	0.01275	18.41	up
136	<i>otos</i>	0.00548	16.17	up
137	<i>nipsnap1</i>	0.01910	14.57	up
138	<i>zp2.6</i>	0.00837	13.48	up
139	<i>sim1a</i>	0.00828	12.97	up
140	<i>zgc:92162</i>	0.00828	12.97	up
141	<i>ifnphi1</i>	0.00854	12.22	up
142	<i>si:zfos-1425h8.1</i>	0.02596	11.92	up
143	<i>loxl5b</i>	0.02596	11.92	up
144	<i>c8b</i>	0.01176	11.37	up
145	<i>ntf7</i>	0.01078	11.15	up
146	<i>zgc:173837</i>	0.02584	10.04	up
147	<i>lat</i>	0.03698	9.76	up
148	<i>kcnj8</i>	0.01999	9.69	up
149	<i>cyp2aa8</i>	0.03276	9.56	up
150	<i>neu3.5</i>	0.01743	8.70	up
151	<i>loc567472</i>	0.01813	8.64	up
152	<i>apobec2b</i>	0.04255	8.13	up
153	<i>trim35-31</i>	0.02062	8.09	up
154	<i>nr4a2a</i>	0.02062	8.09	up
155	<i>glra2</i>	0.03708	7.53	up
156	<i>ppil6</i>	0.01264	7.18	up
157	<i>fam113</i>	0.03726	7.07	up
158	<i>slc24a2</i>	0.04144	7.04	up
159	<i>nos2b</i>	0.00100	6.83	up
160	<i>enpep</i>	0.03410	6.40	up
161	<i>tyrobp</i>	0.04840	5.95	up
162	<i>ifit8</i>	0.04616	5.51	up
163	<i>cdkn2c</i>	0.01276	5.25	up
164	<i>si:dkeyp-110e4.3</i>	0.03335	4.82	up
165	<i>smkr1</i>	0.00796	4.77	up
166	<i>rspo1</i>	0.00158	4.76	up
167	<i>fhl2a</i>	0.02239	4.63	up
168	<i>pawr</i>	0.01645	4.56	up
169	<i>si:ch211-154a22.8</i>	0.00382	4.56	up
170	<i>sptssb</i>	0.01695	4.46	up
171	<i>ta</i>	0.04735	4.29	up
172	<i>dock5</i>	0.03205	4.07	up
173	<i>adh8a</i>	0.03968	3.99	up
174	<i>prom1b</i>	0.03890	3.98	up
175	<i>loc100500728</i>	0.03004	3.90	up
176	<i>il17c</i>	0.02490	3.84	up
177	<i>grip2a</i>	0.02048	3.64	up
178	<i>nr5a2</i>	0.03364	3.61	up
179	<i>loc567180</i>	0.01882	3.61	up
180	<i>zp2.1</i>	0.00909	3.58	up
181	<i>chadlb</i>	0.03359	3.52	up
182	<i>zgc:153738</i>	0.03557	3.25	up
183	<i>si:ch211-57h10.1</i>	0.04141	3.18	up

S. No.	Gene symbol	p-value	Fold Change	Regulation
184	<i>morn3</i>	0.02287	3.16	up
185	<i>camk2n1b</i>	0.04183	3.09	up
186	<i>ehd2b</i>	0.00582	2.94	up
187	<i>trpv4</i>	0.00460	2.89	up
188	<i>ptgdsb</i>	0.01716	2.81	up
189	<i>col9a1b</i>	0.04219	2.74	up
190	<i>apoc1l</i>	0.03741	2.74	up
191	<i>sesn3</i>	0.04346	2.67	up
192	<i>mdm2</i>	0.01221	2.66	up
193	<i>lrp2a</i>	0.03452	2.65	up
194	<i>runx2b</i>	0.01334	2.55	up
195	<i>zgc:66484</i>	0.04074	2.48	up
196	<i>ankar</i>	0.01753	2.44	up
197	<i>her11</i>	0.01284	2.44	up
198	<i>plod2</i>	0.01620	2.43	up
199	<i>cyp2aa3v1</i>	0.02830	2.41	up
200	<i>klhl21</i>	0.01631	2.39	up
201	<i>he2</i>	0.03155	2.38	up
202	<i>olfml3b</i>	0.02651	2.35	up
203	<i>sid4</i>	0.01681	2.34	up
204	<i>npffr1l2</i>	0.00580	2.34	up
205	<i>zgc:110712</i>	0.02124	2.33	up
206	<i>gtpbp1l</i>	0.01621	2.29	up
207	<i>pkhd1l1</i>	0.03879	2.25	up
208	<i>dpyda.1</i>	0.02222	2.23	up
209	<i>si:dkey-91m11.5</i>	0.00200	2.21	up
210	<i>asb13a.2</i>	0.02059	2.17	up
211	<i>pnocb</i>	0.00809	2.16	up
212	<i>e2f5</i>	0.03417	2.16	up
213	<i>loc794484</i>	0.03453	2.14	up
214	<i>si:ch211-244b2.4</i>	0.03453	2.14	up
215	<i>zgc:136493</i>	0.02698	2.11	up
216	<i>gabra6a</i>	0.02770	2.10	up
217	<i>tlr3</i>	0.02476	2.10	up
218	<i>cacnb2a</i>	0.00774	2.09	up
219	<i>slc4a2a</i>	0.03008	2.06	up
220	<i>tubd1</i>	0.04921	2.05	up
221	<i>stat7</i>	0.04971	2.05	up
222	<i>zgc:113426</i>	0.03633	2.05	up
223	<i>zgc:193681</i>	0.00831	2.04	up
224	<i>camkvb</i>	0.02045	2.02	up
225	<i>zgc:165514</i>	0.04473	2.02	up
226	<i>zgc:175107</i>	0.03396	2.01	up

**Appendix 5. List of differentially expressed genes common to SMA MN and SC**

<b>(A) Genes with opposite regulation trend</b>			
<b>S. No.</b>	<b>Gene names</b>	<b>Fold Change</b>	
		<b>SMA MN</b>	<b>SMA SC</b>
1	<i>adh8a</i>	-3.73	3.99
2	<i>glrba</i>	3.51	-7.60
3	<i>prom1b</i>	-3.41	3.98
<b>(B) Genes down-regulated in both lists</b>			
1	<i>ifitm5</i>	-5.29	-11.45
2	<i>si:dkey-228a15.1</i>	-4.72	-4.10
3	<i>lrrc20</i>	-4.54	-2.91
4	<i>pvalb2</i>	-4.22	-2.79
5	<i>gadd45gb.1</i>	-4.20	-2.24
6	<i>zgc:158296</i>	-4.09	-2.72
7	<i>asc1a</i>	-3.68	-3.00
8	<i>myl1</i>	-3.55	-2.39
9	<i>ckmb</i>	-3.52	-2.34
10	<i>smyd1a</i>	-3.49	-2.55
11	<i>mylz3</i>	-3.47	-2.25
12	<i>actc1b</i>	-3.29	-2.61
13	<i>ckma</i>	-3.23	-2.45
14	<i>tnni2a.4</i>	-3.22	-2.54
15	<i>tpma</i>	-2.97	-2.12
16	<i>ak1</i>	-2.69	-2.35
17	<i>si:dkey-23c22.6</i>	-2.47	-2.06
<b>(C) Genes up-regulated in both lists</b>			
1	<i>nipsnap1</i>	55.72	14.57
2	<i>kcnj8</i>	52.65	9.69
3	<i>apoc1l</i>	6.29	2.74
4	<i>ppil6</i>	4.55	7.18
5	<i>smkr1</i>	3.07	4.77
6	<i>sesn3</i>	2.70	2.67
7	<i>zgc:193681</i>	2.02	2.04

## Appendix 6. List of AS events identified in SMA MN

### (A) Skipped exon AS events

geneSymbol	chr	strand	exonStart_0base	exonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	InclLevel Difference
ptenb	12	-	18482144	18482188	18481820	18482059	18482466	18482511	0	0.412
zgc:136639	2	-	22431832	22431907	22431667	22431739	22431977	22432077	3.69E-17	0.406
ptpreb	17	+	29391604	29391701	29390738	29390810	29393243	29393317	2.16E-132	0.36
zgc:165580	17	+	19332145	19332289	19329309	19329524	19332461	19332594	3.97E-164	0.353
copz2	3	+	24399366	24399399	24399117	24399156	24401054	24401645	0.0010874	0.333
melk	1	-	20709535	20709622	20708388	20708517	20710511	20710610	4.07E-06	0.318
dll4	20	+	28282568	28282677	28281373	28282091	28282930	28283618	1.81E-42	0.314
vcla	13	-	25440846	25441038	25438586	25438790	25445064	25445250	0.0076964	0.313
zgc:77838	21	-	5270280	5270352	5264911	5264996	5295568	5296327	3.11E-19	0.304
si:dkey-57a22.11	9	+	23011522	23011600	23011155	23011331	23011793	23011833	2.15E-48	0.299
mettl20	25	-	29777244	29777375	29774501	29774933	29792980	29793322	3.49E-07	0.292
calcr1a	9	-	43436606	43436648	43433737	43433998	43439509	43439728	1.61E-104	0.276
naa30	17	+	44515679	44515735	44515295	44515419	44520359	44523265	5.73E-37	0.263
snx27a	19	+	9267840	9267889	9266968	9267097	9269978	9270104	9.29E-08	0.251
ppp5c	15	-	2738101	2738167	2737889	2737988	2739221	2739343	4.03E-06	0.249
zgc:162576	17	+	42113203	42113284	42110438	42110601	42119446	42120017	9.86E-29	0.244
tbl1xr1a	11	+	8751012	8751110	8750860	8750924	8755159	8755220	2.03E-203	0.235
myom1a	2	+	30631792	30631981	30631019	30631133	30636188	30636317	5.28E-05	0.226
ccnt2a	9	-	24272951	24273014	24272683	24272729	24274684	24274813	5.38E-16	0.22
prkcbp1l	23	-	9947491	9947569	9947022	9947138	9947734	9947790	1.18E-262	0.215
map7d2a	5	-	25649734	25649815	25645497	25645610	25653003	25653103	3.11E-141	0.195
nav3	4	+	19829258	19829327	19817838	19818022	19834542	19834685	0.0058625	0.191
ptpn2a	16	-	11652840	11652968	11650025	11650181	11659517	11659679	9.84E-05	0.189
gnmt	17	-	49084872	49085000	49084673	49084787	49091454	49091737	9.55E-18	0.188
ccdc93	9	-	39417471	39417570	39417340	39417397	39418862	39418974	7.85E-91	0.187
chd2	18	-	24762723	24762840	24760407	24760560	24764807	24764869	0.0153645	0.186
ash2l	10	-	2802789	2802807	2800288	2800391	2802902	2802978	1.51E-09	0.183
rasl11a	7	-	52775097	52775154	52770813	52771284	52775794	52775945	1.47E-11	0.176
pho	5	+	65383451	65383509	65383269	65383367	65386804	65394595	6.61E-104	0.166
ccn1a	18	+	39853090	39853165	39852454	39852747	39854442	39854552	9.76E-27	0.163
mtfmt	7	+	54213413	54213490	54212737	54212816	54216274	54216652	5.81E-08	0.16
mthfd1b	17	-	49558065	49558119	49557843	49557980	49559738	49559798	2.45E-81	0.159

## Appendix 6A. (contd..)

geneSymbol	chr	strand	exonStart_Obase	exonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	InclLevel Difference
ncalda	16	-	11482548	11482610	11440893	11441293	11491536	11491628	0.0012406	0.157
dzip1	6	+	7182987	7183036	7182596	7182725	7187096	7187245	1.45E-12	0.156
tnikb	24	-	27312868	27313015	27310149	27310243	27313315	27313424	0.0370144	0.154
slc25a46	5	-	57336927	57336985	57330075	57331276	57337290	57337347	3.10E-109	0.153
zgc:63907	16	-	46220791	46220996	46218555	46218700	46226625	46226686	2.95E-10	0.153
lrrc40	6	+	29831369	29831542	29831151	29831294	29833063	29833154	6.14E-18	0.151
tmem141	5	+	63139243	63139310	63138374	63138431	63139545	63139629	0.0028726	0.151
zgc:171429	9	+	8955787	8956225	8954520	8954600	8959667	8959806	0.0560507	0.146
kat7	12	-	6681634	6681685	6679510	6679658	6682216	6682360	2.35E-126	0.145
mknk2a	2	+	22470087	22470161	22469471	22469551	22471300	22471405	9.72E-06	0.138
atp2b3b	23	-	20178807	20178849	20175290	20175455	20181879	20182005	0.0082438	0.138
fkn	5	-	72951785	72951842	72951227	72951428	72952933	72953088	2.99E-51	0.135
etf1b	10	+	21518356	21518496	21517960	21518136	21521366	21521505	8.97E-20	0.135
palm2	10	+	4806732	4806834	4804528	4804550	4811303	4812013	0.0083147	0.135
snx14	20	-	881444	881485	879451	879571	885115	885227	5.88E-26	0.133
ptp4a2b	19	-	37020794	37020903	37020617	37020710	37021182	37021457	2.22E-08	0.132
rasl11a	7	-	52774809	52774892	52770813	52771284	52775794	52775945	1.52E-09	0.13
zgc:172295	1	-	27739970	27740044	27719129	27719267	27762582	27762679	0.002891	0.13
aamp	6	+	59646330	59646421	59644657	59644746	59646520	59646675	0.0627878	0.13
srek1	5	-	55757839	55757914	55757644	55757759	55760608	55760740	5.53E-07	0.127
trim33	8	+	11749390	11749441	11744779	11744932	11749637	11750361	5.87E-15	0.125
zgc:101840	20	-	9624961	9625056	9614622	9614680	9640293	9640457	5.74E-25	0.124
clasp2	19	+	44369688	44369796	44368961	44369003	44372784	44372891	2.66E-37	0.123
rab4b	15	-	8975296	8975377	8969887	8970002	8991519	8991664	4.89E-08	0.123
sec31a	10	-	5113837	5113933	5108976	5109070	5114036	5114237	2.11E-05	0.122
gabrg2	21	+	40871128	40871152	40863601	40863807	40879879	40880456	0.000193	0.122
mpzl1l	10	+	285404	285510	282421	282551	288869	292623	0.0742519	0.122
ergic3	6	+	50358017	50358082	50353079	50353176	50360176	50360313	8.60E-174	0.121
dgcr6	8	-	1825064	1825165	1822365	1822506	1828265	1828426	1.60E-75	0.121
tspan18a	7	+	27929967	27930042	27888875	27889049	27950852	27950948	0.0150772	0.121
ndrq4	25	+	12611061	12611100	12610745	12610797	12613573	12618682	1.15E-27	0.119
zgc:91976	17	+	23590629	23590796	23587289	23587407	23592342	23594895	2.21E-09	0.118
cep57l1	20	-	32185291	32185408	32185095	32185173	32185545	32185667	9.02E-71	0.117
tsr2	8	-	8890775	8890906	8890385	8890476	8894400	8894486	1.31E-36	0.117
nip7	18	-	4483409	4483496	4482156	4482295	4486203	4486341	3.56E-05	0.117
zgc:162431	23	-	30273126	30273189	30270995	30271125	30282950	30283032	0.0007815	0.117

## Appendix 6A. (contd..)

geneSymbol	chr	strand	exonStart_Obase	exonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	InclLevel Difference
immt	14	+	20032423	20032528	20026771	20026922	20038602	20038719	1.18E-30	0.116
osbp19	8	-	17162192	17162231	17158737	17158837	17162832	17162883	1.20E-11	0.116
si:ch211-225b11.1	5	-	22584455	22584580	22576972	22577085	22584800	22584904	1.18E-30	0.114
smc2	1	+	18282294	18282333	18281430	18281553	18282437	18282548	1.47E-92	0.11
fmr1	14	+	21163217	21163334	21163034	21163128	21169097	21169268	0.0043671	0.109
zgc:91976	17	+	23587652	23587806	23587289	23587407	23592342	23594895	1.18E-07	0.107
aldh9a1b	2	+	5713906	5714052	5710136	5710335	5715973	5716103	0.0013932	0.107
rcor2	7	-	26157753	26157876	26157398	26157455	26161314	26161593	2.77E-21	0.105
amph	2	-	31615426	31615480	31611929	31612100	31616123	31616177	0.0010333	0.101
nrxn1a	12	+	26022927	26022954	26014891	26015095	26036723	26036846	5.25E-05	0.099
grin1b	5	-	31242076	31242187	31236439	31238473	31249510	31249656	0.0261435	0.098
scn8ab	6	-	39334919	39335042	39328804	39329089	39338675	39338849	0.0002555	0.096
zgc:112175	23	+	4415628	4415744	4414986	4415088	4415849	4416010	0.0244282	0.096
stk24b	6	+	12454440	12454503	12454240	12454358	12462228	12462377	0.0995308	0.09
kansl3	8	-	53594464	53594542	53592155	53592335	53597188	53597345	8.75E-05	0.089
alcama	10	-	29419431	29419458	29419036	29419151	29422447	29422577	8.50E-25	0.088
tmem161b	5	-	49843488	49843592	49833279	49833366	49853374	49853570	1.54E-12	0.085
nprl3	3	-	55930223	55930293	55922561	55922694	55930801	55931414	1.82E-05	0.085
rhd	13	-	46124367	46124447	46121189	46121263	46133117	46133255	0.0572333	0.085
garnl3	5	-	34999812	34999887	34999616	34999731	34999977	35000082	1.85E-51	0.084
zgc:153606	25	-	22447290	22447432	22447072	22447204	22452846	22453056	0.0001699	0.084
cast	21	-	10003350	10003407	10001841	10001898	10003624	10003675	0.0005927	0.083
rnf111	7	-	31904734	31904815	31902661	31902757	31905198	31905316	0.0867303	0.083
fmr1	14	+	21163217	21163334	21161243	21161392	21169097	21169268	1.94E-09	0.082
si:ch211-140m22.7	9	+	20540761	20540825	20527705	20527808	20540910	20541035	0.0254538	0.082
dpp7	5	+	30802414	30802550	30801194	30801358	30802641	30802723	3.61E-23	0.078
psma6b	15	+	46848401	46848496	46847760	46847957	46849877	46850428	1.18E-10	0.077
fmr1	14	+	21163034	21163128	21161243	21161392	21169097	21169268	0.056565	0.076
mthfd1b	17	-	49543018	49543116	49536826	49537000	49545767	49545895	5.06E-08	0.075
si:dkey-119o24.1	7	+	71175471	71175570	71166214	71166358	71183864	71184038	1.84E-05	0.075
wwp2	25	+	35819422	35819560	35816007	35816129	35824317	35824393	1.41E-12	0.073
capza1a	6	-	48273855	48274062	48266432	48266583	48276501	48276565	0.0440833	0.07
zgc:77650	18	+	7019133	7019214	7017445	7017591	7023235	7023345	0.0005616	0.069
ptbp2b	2	-	18283651	18283711	18282869	18282985	18284397	18284596	0.0931317	0.069
brd8	14	+	7679744	7679870	7679371	7679474	7683711	7683896	3.71E-20	0.068
trove2	2	+	11730499	11730616	11727586	11727703	11731015	11731159	1.35E-08	0.068

## Appendix 6A. (contd..)

geneSymbol	chr	strand	exonStart_0base	exonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	InclLevel Difference
pyroxd2	13	-	40644621	40644694	40641526	40641683	40647591	40647726	5.99E-10	0.067
dbnlb	8	+	54621015	54621087	54615539	54615657	54621232	54621336	0.0024939	0.067
lamb1a	25	-	32548488	32548665	32545558	32545638	32578124	32578313	2.26E-45	0.066
ero1lb	11	-	46600418	46600492	46597626	46597747	46600946	46601029	1.63E-06	0.064
emc10	3	-	30393234	30393337	30392356	30392660	30396121	30396215	8.64E-12	0.063
impad1	2	-	11433991	11434087	11431284	11431446	11434285	11434445	0.0778992	0.061
si:ch211-216123.2	18	+	17421766	17421966	17418528	17418670	17427555	17427831	1.11E-27	0.06
tyh3b	1	-	10971467	10971647	10969706	10969782	10975774	10975911	0.0002286	0.06
dpf3	20	+	28505736	28505778	28502369	28502442	28513091	28513229	9.10E-10	0.058
pnkp	15	+	14013301	14013394	14012999	14013070	14013704	14013801	0.0001462	0.057
cstf3	18	-	44173340	44173436	44168060	44168093	44173587	44173689	0.0816906	0.057
atp6v1h	2	+	30522124	30522178	30514647	30514752	30523194	30523292	0.0986457	-0.052
avl9	12	-	15297581	15297648	15293142	15293615	15303538	15303659	1.80E-07	-0.053
ncor1	5	-	43512439	43512589	43510125	43510269	43514882	43515083	0.0004309	-0.053
cdkl5	11	+	31101742	31101865	31101038	31101144	31104045	31104261	5.34E-11	-0.057
rspry1	18	+	17395235	17395362	17394294	17394407	17396510	17396569	1.22E-11	-0.064
rraga	14	+	17944562	17944685	17944381	17944477	17944845	17944937	0.067898	-0.065
mibp2	18	-	4948122	4948171	4947821	4947999	4949034	4949125	0.0499535	-0.066
dmd	1	-	10018272	10018387	10012089	10012182	10024681	10024925	0.0588136	-0.071
nif3l1	12	+	4565794	4565878	4561085	4561209	4570385	4570705	1.69E-05	-0.073
serpinb1l3	24	+	43923696	43923864	43922269	43922412	43926062	43926518	0.0352831	-0.073
cwc22	9	-	44667761	44667822	44635124	44635202	44681527	44681612	7.96E-05	-0.076
casp2	16	-	19630976	19631126	19629652	19629728	19639128	19639281	0.0004072	-0.082
cxxc1l	8	-	8272562	8272663	8271058	8271186	8272749	8272862	0.0558729	-0.083
si:ch211-160d20.3	18	+	35136401	35136478	35133564	35133708	35138821	35139866	2.05E-25	-0.088
eya1	24	+	13839464	13839530	13838222	13838350	13846246	13846324	0.0061107	-0.088
otub1a	21	+	27898780	27898879	27897197	27897259	27901978	27902063	1.33E-05	-0.089
coq4	5	+	63027390	63027520	63027219	63027322	63027742	63027836	0.0100031	-0.089
tlk1a	9	+	3545614	3545757	3544897	3545177	3565807	3565929	0.0333669	-0.093
nexn	8	-	19052138	19052330	19051650	19051735	19055062	19055434	5.42E-09	-0.098
olfm3a	24	-	29614320	29614476	29613875	29614095	29622573	29622717	0.01364	-0.098
pank4	11	+	16241055	16241237	16240556	16240710	16241544	16241626	1.02E-14	-0.1
slx4ip	13	-	35616366	35616480	35616055	35616169	35616741	35616855	1.21E-06	-0.105
ap3d1	22	-	22538989	22539078	22538818	22538904	22540156	22540346	3.38E-05	-0.106
ppp3ca	21	+	28463499	28463547	28462856	28462954	28468301	28470211	9.40E-25	-0.107
nif3l1	12	+	4563477	4563616	4561085	4561209	4570385	4570705	0.0156311	-0.107

Appendix 6A. (contd..)

geneSymbol	chr	strand	exonStart_0base	exonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	InclLevel Difference
cbsa	1	+	28008207	28008314	28001272	28001531	28008574	28008709	0.0159099	-0.107
mbip	17	+	38470308	38470374	38467344	38467441	38470837	38470990	5.36E-07	-0.108
gabpb2b	16	+	41011708	41011827	41008182	41008347	41011913	41012327	1.83E-08	-0.111
cpt1a	7	+	50270839	50270927	50268413	50268521	50272474	50272670	8.24E-05	-0.111
zgc:101016	5	-	27374389	27374503	27371703	27373094	27377839	27378294	1.08E-08	-0.112
si:dkey-286j15.1	22	+	32718693	32718875	32718382	32718548	32720721	32720849	0.0261914	-0.112
nif3l1	12	+	4563477	4563616	4561085	4561209	4565794	4565878	2.68E-11	-0.115
mus81	7	+	21015253	21015339	21013096	21013177	21015443	21015596	0.0012166	-0.116
LOC794757	22	+	25190435	25190511	25189467	25189603	25191105	25191239	6.24E-09	-0.118
zgc:110366	2	+	6847743	6847868	6845984	6846044	6852620	6853273	1.80E-05	-0.118
pum1	19	+	44690129	44690322	44686844	44686966	44692450	44693856	0.0004337	-0.121
slc25a39	3	-	30036880	30036931	30035816	30035926	30038279	30038339	2.18E-18	-0.122
lrrc40	6	+	29835022	29835100	29834534	29834645	29835516	29835702	1.61E-22	-0.124
cratb	19	+	18402766	18402886	18402354	18402455	18405598	18405721	3.25E-05	-0.125
zgc:55558	3	+	16291739	16291899	16290822	16291001	16298544	16299951	5.59E-157	-0.128
kcnma1a	13	+	16931154	16931183	16927699	16927924	16934670	16934822	0.0051851	-0.133
tfb1m	17	-	49427581	49427733	49416803	49416923	49431522	49431627	1.06E-19	-0.138
ppp2r5d	13	-	3901745	3901799	3901425	3901596	3903467	3903576	7.72E-08	-0.14
eif2d	8	-	38560005	38560180	38559756	38559921	38563886	38564007	5.64E-56	-0.141
psip1	1	+	27252154	27252353	27251720	27251779	27254997	27255094	7.13E-21	-0.144
taf6l	7	+	19212183	19212680	19210414	19210543	19215695	19217567	1.36E-06	-0.146
golga1	8	-	43544500	43544575	43541474	43541605	43548228	43548319	4.06E-07	-0.148
odz4	15	+	10723451	10723670	10656006	10656279	10747466	10747713	0.0137908	-0.148
odc1	17	+	52273121	52273255	52271121	52271256	52275325	52275499	2.19E-20	-0.149
bbs7	14	-	49317790	49317856	49317554	49317694	49317953	49318028	2.57E-35	-0.156
zgc:112023	6	-	8632169	8632306	8631377	8631506	8632393	8632725	7.67E-05	-0.157
wdr41	21	+	7552173	7552222	7545667	7546131	7554156	7554285	3.06E-33	-0.159
slc29a4	3	+	40333432	40333556	40328336	40328593	40340362	40340550	1.91E-20	-0.161
nr5a2	22	-	23082643	23082785	23074668	23075074	23084491	23084610	7.27E-42	-0.162
cdkl5	11	+	31078765	31078825	31077159	31077280	31087892	31088082	1.66E-07	-0.162
fbxo22	7	+	31303318	31303484	31301375	31301528	31304465	31305382	0.0005745	-0.166
pard3	2	+	43415556	43415601	43409047	43409190	43417598	43417829	2.21E-09	-0.168
rad18	6	+	41784237	41784367	41780790	41780910	41799494	41799541	1.06E-15	-0.17
ascc2	5	-	19688086	19688154	19687874	19687988	19689564	19689738	0.0007366	-0.174
csnk1g2a	2	+	22519310	22519459	22513924	22513994	22519548	22519783	1.30E-10	-0.176
elmo1	19	-	35388935	35388971	35382991	35383091	35401913	35402083	7.35E-138	-0.177

## Appendix 6A. (contd..)

geneSymbol	chr	strand	exonStart_0base	exonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	InclLevel Difference
mkl1b	12	+	19810948	19811035	19810682	19810817	19811141	19811202	2.37E-13	-0.18
scn2b	15	+	12936697	12936765	12936410	12936606	12939298	12940842	3.88E-08	-0.184
fam3a	23	+	25433300	25433324	25426861	25426969	25433413	25433537	1.26E-09	-0.185
msrb3	4	-	11971377	11971531	11967052	11967161	11978585	11978673	1.44E-05	-0.186
zmat5	5	+	33714216	33714297	33714048	33714111	33716042	33716151	3.48E-10	-0.189
mlh1	13	+	43257035	43257126	43255510	43255631	43257222	43257321	1.80E-42	-0.191
si:dkey-46l15.1	6	-	50565813	50565888	50549825	50549983	50568161	50568253	8.82E-92	-0.198
cdkl5	11	+	31078765	31078825	31077159	31077280	31081117	31081208	2.92E-07	-0.204
tbpl1	23	+	31839698	31839762	31838881	31838964	31839844	31839948	1.45E-08	-0.205
cyp2p10	20	+	25617677	25617838	25617444	25617594	25620762	25620939	5.90E-06	-0.207
pbx3b	8	-	34764046	34764124	34761214	34761261	34764689	34764855	0.0009128	-0.208
si:ch211-22i13.2	20	-	13926804	13926874	13926522	13926684	13927427	13927613	2.46E-287	-0.217
kdsr	2	-	13034313	13034397	13034114	13034198	13034484	13034676	1.42E-20	-0.228
fgfr2	13	-	46998391	46998658	46995484	46995559	47032754	47033011	4.48E-11	-0.228
rab24	21	+	37698031	37698091	37697890	37697941	37699066	37700268	7.37E-74	-0.23
rab15	20	-	28811960	28812050	28810532	28810598	28812306	28812384	7.15E-09	-0.24
polr3gla	19	-	25036239	25036296	25036034	25036111	25039030	25039099	0.026561	-0.241
zgc:162431	23	-	30253820	30253903	30251819	30252766	30255392	30255509	0.0015118	-0.247
upp1	16	-	18262563	18262675	18262315	18262474	18263094	18263143	0.0024989	-0.248
afap111b	21	-	29178293	29178372	29171593	29171749	29178882	29178975	0.0005114	-0.257
eif4e3	23	-	10833717	10833873	10830270	10832309	10835228	10835295	2.25E-37	-0.258
zgc:112466	21	-	229427	229477	229066	229193	230124	230223	5.87E-21	-0.26
cadm1a	21	+	23658825	23658858	23619570	23619743	23689082	23689211	2.24E-06	-0.27
tbp	13	-	24609972	24610140	24608891	24608986	24610243	24610335	1.54E-173	-0.273
ptpn13	21	+	8854505	8854634	8849905	8850333	8857330	8857424	6.46E-06	-0.283
pmvk	16	-	8152297	8152427	8150792	8152212	8153823	8153976	8.30E-36	-0.289
mthfd1b	17	-	49514486	49514594	49511436	49511589	49515216	49515394	5.33E-28	-0.292
snx9b	20	-	43078665	43078804	43076099	43076181	43084230	43084396	1.27E-124	-0.294
zgc:91909	8	+	53244356	53244483	53240686	53240905	53247444	53247516	9.86E-14	-0.294
zgc:153901	9	+	25243578	25243674	25243254	25243447	25244400	25244589	0.0043491	-0.294
srsf6b	11	+	1489084	1489248	1487469	1487604	1490569	1490691	2.31E-22	-0.298
eya3	19	+	25385908	25386040	25385561	25385671	25386240	25386320	4.72E-38	-0.322
pot1	25	+	28290895	28290974	28277055	28277175	28291396	28291494	4.27E-58	-0.325
wdr41	21	+	7549093	7549209	7545667	7546131	7552173	7552222	4.47E-28	-0.329
si:ch211-284b7.3	2	-	5687295	5687373	5684736	5684802	5690364	5690446	1.93E-122	-0.334
pet112	1	-	23409017	23409131	23397583	23397782	23411570	23411721	4.45E-05	-0.335

geneSymbol	chr	strand	exonStart_Obase	exonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	InclLevel Difference
wdr41	21	+	7549093	7549209	7545667	7546131	7554156	7554285	0.0007194	-0.342
lpar2b	1	-	59943763	59943796	59938281	59938611	59947173	59947427	0.001061	-0.342
foxp2	4	-	5885772	5885950	5854178	5854268	5940846	5940911	0.0440833	-0.362
sepn1	17	+	24521976	24522078	24521347	24521460	24522170	24523355	0	-0.408
zgc:63674	2	+	20670032	20670164	20669812	20669872	20670479	20670567	1.52E-33	-0.452
mogat3b	25	+	16967040	16967109	16965085	16965175	16967208	16967303	2.77E-91	-0.481
osbp16	9	-	1410398	1410491	1402248	1402417	1413039	1413143	9.44E-21	-0.535
rmnd5b	14	+	51492793	51492931	51491250	51491393	51493033	51493134	0	-0.587
ptges3a	23	+	27297788	27297845	27297616	27297706	27299207	27299235	1.98E-178	-0.74
mdm2	4	+	22063129	22063204	22062973	22063045	22064660	22064794	7.70E-308	-0.807

**(B) Mutually exclusive exon AS events**

geneSymbol	chr	strand	1stExon Start_Obase	1stExonEnd	2ndExonStart_Obase	2ndExonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	InclLevel Difference
mdm2	4	+	22064660	22064794	22065101	22065151	22062973	22063045	22065262	22065321	2.44E-117	0.408
rgs12b	1	-	41597618	41597715	41598335	41598403	41596812	41597373	41598765	41598862	0.033852846	0.302
cadm1a	21	+	23639862	23639946	23658825	23658858	23619570	23619743	23689082	23689211	7.36E-20	0.244
methfd1b	17	-	49514486	49514594	49515216	49515394	49511436	49511589	49518894	49518995	5.64E-19	0.235
zgc:63674	2	+	20670479	20670567	20670999	20671185	20669812	20669872	20671811	20671928	9.50E-88	0.221
msi2b	15	-	29831831	29832011	29848475	29848598	29816580	29816735	29876177	29876252	0.000105091	0.197
upb1	8	+	31571366	31571448	31573263	31573306	31569615	31569785	31577052	31577207	0.058373803	0.181
rab15	20	-	28811960	28812050	28812306	28812384	28810532	28810598	28815556	28815617	5.81E-10	0.144
zmp:0000000686	3	+	60198681	60198963	60202056	60202317	60197856	60198150	60205249	60205516	0.033852846	0.144
sulf2b	23	-	15317606	15317660	15319653	15319687	15314865	15317132	15322241	15322365	1.78E-08	0.133
dgcr6	8	-	1822365	1822506	1825064	1825165	1819021	1819406	1828265	1828426	0.002266556	0.121
ncs1a	5	-	34048574	34048663	34050992	34051071	34046162	34046240	34051970	34052109	0.069756709	0.121
rcor2	7	-	26157398	26157455	26157753	26157876	26152572	26152625	26161314	26161593	7.00E-30	0.119
glmna	2	+	10637701	10637837	10639963	10640040	10632969	10633033	10648472	10649203	0.000105091	0.118
tex261	7	-	9510124	9510278	9514629	9514709	9506871	9506974	9518538	9518712	3.41E-32	0.116
reep3a	17	-	19443548	19443685	19444367	19444488	19439856	19440428	19446887	19447001	0.00686345	0.113
map7d2a	5	-	25653003	25653103	25656415	25656579	25645497	25645610	25659434	25659512	1.70E-28	0.11

Appendix 6B. (contd..)

geneSymbol	chr	strand	1stExonStart_0base	1stExonEnd	2ndExonStart_0base	2ndExonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	IncLevelDifference
hipk3b	18	+	27898725	27898849	27909809	27909929	27875746	27877008	27912776	27912863	0.011188039	0.107
melk	1	-	20707035	20707188	20709535	20709622	20704399	20704664	20710511	20710610	7.27E-05	0.106
qkia	17	+	27604540	27604683	27629112	27629229	27579731	27580116	27639990	27640134	2.01E-240	0.096
nom1	7	-	42041432	42041565	42041938	42042060	42040108	42040240	42042250	42042421	0.034834675	0.094
mnat1	13	+	31881360	31881513	31883216	31883320	31871902	31872092	31883775	31883916	0.01769779	0.093
odc1	17	+	52271121	52271256	52273121	52273255	52270230	52270353	52275325	52275499	5.01E-10	0.089
pemt	12	+	1531594	1531706	1562879	1562954	1526401	1526547	1564315	1566132	2.76E-11	0.085
lemd3	4	-	11984405	11984487	11985063	11985242	11984139	11984245	11986366	11986469	0.000413679	0.077
cdon	18	-	42487765	42487879	42489815	42489990	42481794	42482001	42492487	42492771	1.99E-07	0.072
pcf11a	10	-	43569792	43569936	43576703	43576787	43567404	43567564	43576951	43577054	0.058496646	0.071
adarb1a	22	-	12807237	12807264	12819972	12820030	12793681	12794619	12832950	12833138	0.006016418	0.067
abi2a	9	+	14063058	14063217	14066970	14067057	14054379	14054501	14073827	14074001	8.67E-06	0.058
agfg1a	15	-	34463353	34463557	34468451	34468571	34459857	34460017	34470687	34470847	5.31E-11	0.051
hoxc3a	23	+	36178766	36178818	36196567	36196737	36138840	36138951	36196927	36197195	0.002266556	-0.053
nr6a1b	21	+	7743288	7743344	7745771	7745923	7739453	7739681	7748647	7748833	0.002829475	-0.058
adck3	20	+	38528235	38528329	38533074	38533216	38525080	38525162	38536556	38536664	0.005699956	-0.061
dgcr6	8	-	1825064	1825165	1828265	1828426	1822365	1822506	1830518	1830653	0.032432488	-0.065
tlk1a	9	+	3545614	3545757	3565807	3565929	3544897	3545177	3566451	3566523	0.004379908	-0.07
rtn1b	20	+	20763179	20763318	20765748	20765807	20762886	20763094	20767016	20768113	0.005699956	-0.07
aplp2	18	-	42804960	42805134	42813908	42814008	42794004	42794087	42818533	42818662	0.007302814	-0.07
adarb1a	22	-	12807237	12807264	12819972	12820030	12800273	12800345	12832950	12833138	6.53E-18	-0.074
msrb1a	3	-	18624713	18624828	18625439	18625582	18620802	18623176	18627008	18627178	0.058373803	-0.078
xrn1	2	+	16463538	16463652	16464279	16464379	16460365	16460553	16464485	16464643	2.72E-05	-0.08
hagh	3	-	18606922	18607040	18608198	18608263	18604712	18604821	18609706	18609870	8.16E-06	-0.081
adck3	20	+	38528235	38528329	38537960	38538026	38525080	38525162	38538109	38538196	0.000159081	-0.081
rad18	6	+	41784237	41784367	41799494	41799541	41780790	41780910	41817999	41818262	1.30E-08	-0.105
si:ch211-57i17.1	20	+	46995223	46995365	46997610	46997876	46993794	46993911	47005668	47005744	3.36E-10	-0.108
hdac4	9	-	47058912	47059123	47064792	47064905	47053400	47053594	47065264	47065393	1.02E-10	-0.111
nif3l1	12	+	4563477	4563616	4565794	4565878	4561085	4561209	4570385	4570705	1.28E-08	-0.116
upb1	8	+	31557702	31557797	31563756	31563918	31556544	31556632	31569615	31569785	0.022005778	-0.127
bcar3	8	-	15625695	15625824	15666975	15667015	15617316	15617756	15678952	15679281	0.083699072	-0.145
slc52a2	19	-	22820224	22820348	22821972	22822825	22817617	22818186	22822909	22823046	6.53E-05	-0.151
wdr41	21	+	7549093	7549209	7552173	7552222	7545667	7546131	7554156	7554285	0.000345264	-0.156
atp7a	14	+	23430961	23431096	23431856	23432051	23429736	23429891	23433295	23433478	0.044817223	-0.195
nbas	20	-	33111158	33111221	33116126	33116260	33099753	33099876	33123707	33123881	6.53E-05	-0.216

Appendix 6B. (contd..)

geneSymbol	chr	strand	1stExonStart_0base	1stExonEnd	2ndExonStart_0base	2ndExonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	IncLevel Difference
calcr1a	9	-	43436606	43436648	43439509	43439728	43433737	43433998	43439995	43440062	1.62E-213	-0.22
slc50a1	16	+	25544865	25544943	25545374	25545498	25543186	25543397	25550234	25550396	1.60E-05	-0.225
vps39	17	+	50620145	50620287	50622664	50622751	50616275	50616416	50627058	50627132	2.39E-08	-0.226
zgc:162576	17	+	42110438	42110601	42113203	42113284	42108713	42108776	42119446	42120017	0.002266556	-0.276
slc7a7	7	+	20737619	20737723	20739853	20739950	20736689	20736813	20740994	20741144	0.01825428	-0.358
si:ch211-284b7.3	2	-	5690364	5690446	5691753	5691810	5684736	5684802	5694006	5694090	5.20E-257	-0.359

(C) Alternative 3' splice site AS events

geneSymbol	chr	strand	longExonStart_0base	longExonEnd	shortES	shortEE	flankingES	flankingEE	FDR	IncLevel Difference
samd13	11	+	8300032	8300075	8300035	8300075	8299769	8299940	3.84E-16	-0.23

(D) Alternative 5' splice site AS events

geneSymbol	chr	strand	longExonStart_0base	longExonEnd	shortES	shortEE	flankingES	flankingEE	FDR	IncLevel Difference
ptp4a2b	19	-	37020794	37021457	37021182	37021457	37020617	37020710	2.09E-12	0.164

(E) Retained intron AS events

geneSymbol	chr	strand	riExonStart_0base	riExonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	IncLevel Difference
ccnt2a	9	-	24267577	24270178	24267577	24269032	24269127	24270178	3.84E-25	0.166
tmem234	13	-	33540557	33540930	33540557	33540709	33540896	33540930	1.69E-05	-0.333

## Appendix 7. List of AS events identified in SMA SC

### (A) Skipped exon AS events

geneSymbol	chr	strand	exonStart_Obase	exonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	InclLevel Difference
mdm2	4	+	22063129	22063204	22062973	22063045	22064660	22064794	0	-0.775
ptges3a	23	+	27297788	27297845	27297616	27297706	27299207	27299235	0	-0.7
wdr41	21	+	7549093	7549209	7545667	7546131	7552173	7552222	0.003753895	-0.477
mthfd1b	17	-	49514486	49514594	49511436	49511589	49515216	49515394	0	-0.402
mogat3b	25	+	16967040	16967109	16965085	16965175	16967208	16967303	8.63E-07	-0.384
rmnd5b	14	+	51492793	51492931	51491250	51491393	51493033	51493134	0	-0.371
pet112	1	-	23409017	23409131	23397583	23397782	23411570	23411721	0.000212796	-0.314
sepn1	17	+	24521976	24522078	24521347	24521460	24522170	24523355	0	-0.313
znf609	7	-	55793542	55793908	55792628	55792707	55794414	55794739	0.015730428	-0.291
zgc:91909	8	-	53244356	53244483	53240686	53240905	53247444	53247516	6.50E-13	-0.285
odc1	17	+	52273121	52273255	52271121	52271256	52275325	52275499	0	-0.237
si:ch211-22i13.2	20	-	13926804	13926874	13926522	13926684	13927427	13927613	4.13E-10	-0.209
zgc:63674	2	+	20670032	20670164	20669812	20669872	20670479	20670567	0.041763267	-0.207
srsf6b	11	+	1489084	1489248	1487469	1487604	1490569	1490691	9.14E-14	-0.187
si:ch211-160d20.3	18	+	35136401	35136478	35133564	35133708	35138821	35139866	0.001554926	-0.168
snx9b	20	-	43078665	43078804	43076099	43076181	43084230	43084396	0.000644025	-0.159
pum1	19	+	44690129	44690322	44686844	44686966	44692450	44693856	2.58E-07	-0.155
stx6	22	-	16712415	16712478	16712209	16712338	16713637	16713732	5.47E-05	-0.141
nenf	17	+	45563920	45563981	45563361	45563805	45569379	45569483	0.066148894	-0.136
csnk1g2a	2	+	22519310	22519459	22513924	22513994	22519548	22519783	0.004977037	-0.119
prdx5	21	+	26056193	26056325	26053497	26053632	26056437	26056476	0.058653284	-0.113
cds2	5	+	22937693	22937790	22936711	22936836	22941625	22941723	0.015413046	-0.102
fam3a	23	+	25433300	25433324	25426861	25426969	25433413	25433537	0.00090344	-0.099
fmr1	14	+	21163217	21163334	21163034	21163128	21169097	21169268	0.007290535	0.139
ergic3	6	+	50358017	50358082	50353079	50353176	50360176	50360313	1.68E-07	0.148
phf2	11	-	28545121	28545237	28542410	28542589	28546538	28546658	0.01289792	0.167
ccnl1a	18	+	39853090	39853165	39852454	39852747	39854442	39854552	0	0.196
ccdc93	9	-	39417471	39417570	39417340	39417397	39418862	39418974	2.06E-05	0.198
aamp	6	+	59646330	59646421	59644657	59644746	59646520	59646675	0.000212796	0.218
snx15	7	-	21123289	21123325	21123093	21123214	21123781	21123965	0.011685496	0.221
mthfd1b	17	-	49558065	49558119	49557843	49557980	49559738	49559798	0.01289792	0.264
melk	1	-	20709535	20709622	20708388	20708517	20710511	20710610	0.001297067	0.268

## Appendix 7A. (contd..)

geneSymbol	chr	strand	exonStart_0base	exonEnd	upstreamES	upstreamEE	downstreamES	downstreamEE	FDR	InclLevel Difference
ptp4a2b	19	-	37021182	37021457	37020617	37020710	37040965	37041074	0.008983682	0.287
calcr1a	9	-	43436606	43436648	43433737	43433998	43439509	43439728	0.062196472	0.302
prkcbp1l	23	-	9947491	9947569	9947022	9947138	9947734	9947790	2.52E-09	0.318
ptenb	12	-	18482144	18482188	18481820	18482059	18482466	18482511	1.67E-07	0.359
zgc:165580	17	+	19332145	19332289	19329309	19329524	19332461	19332594	0.083978973	0.376
snx27a	19	+	9267840	9267889	9266968	9267097	9269978	9270104	9.16E-05	0.633

**(B) Mutually exclusive exon AS events**

geneSymbol	chr	strand	1stExonStart_0base	1stExon End	2ndExonStart_0base	2ndExon End	upstream ES	upstream mEE	downstream mES	downstream EE	FDR	InclLevel Difference
wdr41	21	+	7549093	7549209	7552173	7552222	7545667	7546131	7554156	7554285	0.02799	-0.409

## Appendix 8. List of significantly enriched GO terms identified from AS analysis of

### SMA MN

GO.ID	Term	Annotated	Significant	Expected	elim
GO:0008150	biological_process	4220	189	347.16	1.2E-28
GO:0005575	cellular_component	3564	153	280.94	5.8E-27
GO:0003674	molecular_function	4154	184	347.36	1.3E-25
GO:0016021	integral to membrane	747	36	58.88	2E-07
GO:0005737	cytoplasm	1227	55	96.72	1.8E-06
GO:0003824	catalytic activity	1737	78	145.25	0.000016
GO:0003723	RNA binding	228	15	19.07	0.000023
GO:0005634	nucleus	954	42	75.2	0.000029
GO:0044464	cell part	2792	122	220.08	0.000035
GO:0008152	metabolic process	2641	117	217.26	0.000095
GO:0008589	regulation of smoothened signaling pathway	6	3	0.49	0.00012
GO:0061351	neural precursor cell proliferation	7	3	0.58	0.00021
GO:0035335	peptidyl-tyrosine dephosphorylation	47	6	3.87	0.00022
GO:0004725	protein tyrosine phosphatase activity	47	6	3.93	0.00022
GO:0006468	protein phosphorylation	249	14	20.48	0.00023
GO:0006886	intracellular protein transport	138	10	11.35	0.00025
GO:0007165	signal transduction	660	31	54.3	0.00026
GO:0005524	ATP binding	486	21	40.64	0.00028
GO:0033743	peptide-methionine (R)-S-oxide reductase activity	2	2	0.17	0.00035
GO:0006352	transcription initiation, DNA-dependent	19	4	1.56	0.00036
GO:0016023	cytoplasmic membrane-bounded vesicle	74	7	5.83	0.00036
GO:0043234	protein complex	622	27	49.03	0.00039
GO:0016192	vesicle-mediated transport	190	15	15.63	0.00041
GO:0006355	regulation of transcription, DNA-dependent	402	18	33.07	0.0005
GO:0004674	protein serine/threonine kinase activity	153	10	12.79	0.00057
GO:0003677	DNA binding	372	17	31.11	0.00058
GO:0044428	nuclear part	298	16	23.49	0.00067
GO:0010951	negative regulation of endopeptidase activity	10	3	0.82	0.00069
GO:0031090	organelle membrane	285	17	22.47	0.00082
GO:0022891	substrate-specific transmembrane transporter activity	184	13	15.39	0.00101
GO:0031167	rRNA methylation	3	2	0.25	0.00102
GO:0030091	protein repair	3	2	0.25	0.00102
GO:0016433	rRNA (adenine) methyltransferase activity	3	2	0.25	0.00103
GO:0019003	GDP binding	25	4	2.09	0.00109
GO:0002009	morphogenesis of an epithelium	88	7	7.24	0.00124
GO:0006897	endocytosis	66	6	5.43	0.00138
GO:0008170	N-methyltransferase activity	27	4	2.26	0.00147
GO:0016788	hydrolase activity, acting on ester bonds	200	16	16.72	0.00155
GO:0009790	embryo development	209	15	17.19	0.00177
GO:0006464	protein modification process	591	33	48.62	0.00188
GO:0007423	sensory organ development	122	8	10.04	0.00197
GO:0008270	zinc ion binding	347	15	29.02	0.0021
GO:0006139	nucleobase-containing compound metabolic process	972	48	79.96	0.00219
GO:0051276	chromosome organization	125	8	10.28	0.00229
GO:0044427	chromosomal part	75	6	5.91	0.0023
GO:0005743	mitochondrial inner membrane	52	5	4.1	0.00239
GO:0006397	mRNA processing	74	6	6.09	0.0025
GO:0048562	embryonic organ morphogenesis	52	5	4.28	0.00272
GO:0003779	actin binding	75	8	6.27	0.00279
GO:0030334	regulation of cell migration	16	3	1.32	0.00298
GO:0045132	meiotic chromosome segregation	5	2	0.41	0.00332
GO:0034453	microtubule anchoring	5	2	0.41	0.00332
GO:0048738	cardiac muscle tissue development	17	3	1.4	0.00356
GO:0019829	cation-transporting ATPase activity	17	3	1.42	0.00358
GO:0005488	binding	2500	101	209.05	0.0037
GO:0035239	tube morphogenesis	56	5	4.61	0.00376
GO:0061371	determination of heart left/right asymmetry	35	4	2.88	0.00388
GO:0051188	cofactor biosynthetic process	43	6	3.54	0.00416

GO.ID	Term	Annotated	Significant	Expected	elim
GO:0022603	regulation of anatomical structure morphogenesis	58	5	4.77	0.00438
GO:0005768	endosome	60	5	4.73	0.00447
GO:0001518	voltage-gated sodium channel complex	6	2	0.47	0.00466
GO:0007420	brain development	84	6	6.91	0.0047
GO:0005525	GTP binding	141	8	11.79	0.00484
GO:0044237	cellular metabolic process	2015	93	165.77	0.00485
GO:0044238	primary metabolic process	1967	89	161.82	0.00489
GO:0045214	sarcomere organization	6	2	0.49	0.00492
GO:0006807	nitrogen compound metabolic process	1054	55	86.71	0.00519
GO:0044425	membrane part	923	44	72.76	0.00521
GO:0044281	small molecule metabolic process	1155	58	95.02	0.00524
GO:0005794	Golgi apparatus	117	7	9.22	0.00524
GO:0050790	regulation of catalytic activity	220	13	18.1	0.0053
GO:0070013	intracellular organelle lumen	260	13	20.49	0.00545
GO:0006396	RNA processing	167	14	13.74	0.00554
GO:0051252	regulation of RNA metabolic process	423	21	34.8	0.00559
GO:0044446	intracellular organelle part	789	39	62.19	0.0058
GO:0048856	anatomical structure development	677	39	55.69	0.00584
GO:0007049	cell cycle	184	11	15.14	0.00586
GO:0000793	condensed chromosome	21	3	1.66	0.0061
GO:0048869	cellular developmental process	389	18	32	0.00642
GO:0071842	cellular component organization at cellular level	588	28	48.37	0.00667
GO:0001946	lymphangiogenesis	7	2	0.58	0.00681
GO:0006744	ubiquinone biosynthetic process	7	2	0.58	0.00681
GO:0016814	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amidines	7	2	0.59	0.00683
GO:0004437	inositol or phosphatidylinositol phosphatase activity	7	2	0.59	0.00683
GO:0032482	Rab protein signal transduction	41	4	3.37	0.00689
GO:0034220	ion transmembrane transport	120	7	9.87	0.00706
GO:0006812	cation transport	121	7	9.95	0.00737
GO:0051015	actin filament binding	22	3	1.84	0.00757
GO:0004888	transmembrane signaling receptor activity	94	6	7.86	0.00814
GO:0000118	histone deacetylase complex	8	2	0.63	0.00849
GO:0048747	muscle fiber development	23	3	1.89	0.00855
GO:0006366	transcription from RNA polymerase II promoter	125	7	10.28	0.00876
GO:0060840	artery development	8	2	0.66	0.00897
GO:0042559	pteridine-containing compound biosynthetic process	8	2	0.66	0.00897
GO:0021903	rostrocaudal neural tube patterning	8	2	0.66	0.00897
GO:0004867	serine-type endopeptidase inhibitor activity	8	2	0.67	0.009
GO:0008289	lipid binding	69	5	5.77	0.00917
GO:0005886	plasma membrane	277	13	21.83	0.00922
GO:0005773	vacuole	46	4	3.63	0.00935
GO:0030695	GTPase regulator activity	97	6	8.11	0.00944

## Appendix 9. *srsf6b* CRISPR gBlocks

### (A) gblock\_srsf6b\_E1

CATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAAGTCAAAAGCCTCCGGTCCGGAG  
GCTTTTGACTTTCTGCTATGGAGGTCAGGTATGATTTAAATGGTCAGTATTGAGCCTCAGGA  
AACAGCTATGACATCAAGCTGACTAGATAATCTAGCTGATCGTGGACCGATCATACGTATA  
ATGCCGTAAGATCACGGGTCGCAGCACAGCTCGCGGTCCAGTAGTGATCGACACTGCTCG  
ATCCGCTCGCACCGCTAGCTAATACGACTCACTATAggctgccgatgtacaccgGTTTTAGAGCT  
AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG  
GTGCTTTT

### (B) gblock\_srsf6b\_I1

CATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAAGTCAAAAGCCTCCGGTCCGGAG  
GCTTTTGACTTTCTGCTATGGAGGTCAGGTATGATTTAAATGGTCAGTATTGAGCCTCAGGA  
AACAGCTATGACATCAAGCTGACTAGATAATCTAGCTGATCGTGGACCGATCATACGTATA  
ATGCCGTAAGATCACGGGTCGCAGCACAGCTCGCGGTCCAGTAGTGATCGACACTGCTCG  
ATCCGCTCGCACCGCTAGCTAATACGACTCACTATAggtggacgggaacgcgcggtGTTTTAGAG  
CTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT  
CGGTGCTTTT

TAATACGACTCACTATA - T7 promoter sequence

ggctgccgatgtacaccg - *srsf6b* exon 1 gRNA target site

ggtggacgggaacgcgcggtg - *srsf6b* intron 1 gRNA target site

## REFERENCES

- Akizu, N., V. Cantagrel, M. S. Zaki, L. Al-Gazali, X. Wang, R. O. Rosti, E. Dikoglu, A. B. Gelot, B. Rosti, K. K. Vaux, E. M. Scott, J. L. Silhavy, J. Schroth, B. Copeland, A. E. Schaffer, P. L. Gordts, J. D. Esko, M. D. Buschman, S. J. Field, G. Napolitano, G. M. Abdel-Salam, R. K. Ozgul, M. S. Sagiroglu, M. Azam, S. Ismail, M. Aglan, L. Selim, I. G. Mahmoud, S. Abdel-Hadi, A. E. Badawy, A. A. Sadek, F. Mojahedi, H. Kayserili, A. Masri, L. Bastaki, S. Temtamy, U. Muller, I. Desguerre, J. L. Casanova, A. Dursun, M. Gunel, S. B. Gabriel, P. de Lonlay and J. G. Gleeson (2015). "Biallelic mutations in SNX14 cause a syndromic form of cerebellar atrophy and lysosome-autophagosome dysfunction." *Nat Genet* **47**(5): 528-534.
- Al-Attar, S., E. R. Westra, J. van der Oost and S. J. Brouns (2011). "Clustered regularly interspaced short palindromic repeats (CRISPRs)-the hallmark of an ingenious antiviral defense mechanism in prokaryotes." *Biological Chemistry* **392**: 277-289.
- Anderson, K. N., D. Baban, P. L. Oliver, A. Potter and K. E. Davies (2004). "Expression profiling in spinal muscular atrophy reveals an RNA binding protein deficit." *Neuromuscul Disord* **14**(11): 711-722.
- Araque, A., G. Carmignoto, P. G. Haydon, S. H. Oliet, R. Robitaille and A. Volterra (2014). "Gliotransmitters travel in time and space." *Neuron* **81**(4): 728-739.
- Araque, A., V. Parpura, R. P. Sanzgiri and P. G. Haydon (1999). "Tripartite synapses: glia, the unacknowledged partner." *Trends in Neuroscience* **22**: 208-215.
- Armati, P. J. and E. K. Mathey (2013). "An update on Schwann cell biology--immunomodulation, neural regulation and other surprises." *J Neurol Sci* **333**(1-2): 68-72.
- Bassett, A. R. and J. L. Liu (2014). "CRISPR/Cas9 and genome editing in Drosophila." *J Genet Genomics* **41**(1): 7-19.
- Baumer, D., S. Lee, G. Nicholson, J. L. Davies, N. J. Parkinson, L. M. Murray, T. H. Gillingwater, O. Ansoorge, K. E. Davies and K. Talbot (2009). "Alternative splicing events are a late feature of pathology in a mouse model of spinal muscular atrophy." *PLoS Genet* **5**(12): e1000773.
- Beebe, T. W., C. E. Dominguez and D. S. Chandler (2012). "Mouse models of SMA: tools for disease characterization and therapeutic development." *Hum Genet* **131**(8): 1277-1293.
- Bechade, C., P. Rostaing, C. Cisterni, R. Kalisch, V. La Bella, B. Pettmann and A. Triller (1999). "Subcellular distribution of survival motor neuron (SMN) protein: possible involvement in nucleocytoplasmic and dendritic transport." *Eur J Neurosci* **11**(1): 293-304.
- Behm-Ansmant, I., I. Kashima, J. Rehwinkel, J. Sauliere, N. Wittkopp and E. Izaurralde (2007). "mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons." *FEBS Lett* **581**(15): 2845-2853.
- Bevan, A. K., K. R. Hutchinson, K. D. Foust, L. Braun, V. L. McGovern, L. Schmelzer, J. G. Ward, J. C. Petruska, P. A. Lucchesi, A. H. M. Burghes and B. K. Kaspar (2010). "Early heart failure in the SMN 7 model of spinal muscular atrophy and correction by postnatal scAAV9-SMN delivery." *Human Molecular Genetics* **19**(20): 3895-3905.

- Bhaya, D., M. Davison and R. Barrangou** (2011). "CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation." *Annu Rev Genet* **45**: 273-297.
- Biederer, T., Y. Sara, M. Mozhayeva, D. Atasoy, X. R. Liu, E. T. Kavalali and T. C. Sudhof** (2002). "SynCAM, a synaptic adhesion molecule that drives synapse assembly." *Science* **297**(5586): 1525-1531.
- Boulisfane, N., M. Choleza, F. Rage, H. Neel, J. Soret and R. Bordonne** (2011). "Impaired minor tri-snRNP assembly generates differential splicing defects of U12-type introns in lymphoblasts derived from a type I SMA patient." *Hum Mol Genet* **20**(4): 641-648.
- Briscoe, J. and B. G. Novitch** (2008). "Regulatory pathways linking progenitor patterning, cell fates and neurogenesis in the ventral neural tube." *Philos Trans R Soc Lond B Biol Sci* **363**(1489): 57-70.
- Brouns, S. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin and J. van der Oost** (2008). "Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes." *Science* **321**: 960-964.
- Brzustowicz, L. M., T. Lehner, L. H. Castilla, G. K. Penchaszadeh, K. C. Wilhelmsen, R. Daniels, K. E. Davies, M. Leppert, F. Ziter, D. Wood, V. Dubowitz, K. Zerres, I. Hausmanowa-Petrusewicz, T. Ott, T. L. Munsat and T. C. Gilliam** (1990). "Genetic mapping of chronic childhood-onset spinal muscular atrophy to chromosome 5q11.2-13.3." *Nature* **344**: 540-541.
- Burghes, A. H. and C. E. Beattie** (2009). "Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick?" *Nat Rev Neurosci* **10**(8): 597-609.
- Burnett, B. G., E. Munoz, A. Tandon, D. Y. Kwon, C. J. Sumner and K. H. Fischbeck** (2009). "Regulation of SMN protein stability." *Mol Cell Biol* **29**(5): 1107-1115.
- Carissimi, C., J. Baccon, M. Straccia, P. Chiarella, A. Maiolica, A. Sawyer, J. Rappsilber and L. Pellizzoni** (2005). "Unrip is a component of SMN complexes active in snRNP assembly." *FEBS Lett* **579**(11): 2348-2354.
- Cartegni, L., M. L. Hastings, J. A. Calarco, E. de Stanchina and A. R. Krainer** (2006). "Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2." *Am J Hum Genet* **78**(1): 63-77.
- Cartegni, L. and A. R. Krainer** (2002). "Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1." *Nat Genet* **30**(4): 377-384.
- Cerda, G. A., M. Hargrave and K. E. Lewis** (2009). "RNA profiling of FAC-sorted neurons from the developing zebrafish spinal cord." *Dev Dyn* **238**(1): 150-161.
- Chang, N., C. Sun, L. Gao, D. Zhu, X. Xu, X. Zhu, J. W. Xiong and J. J. Xi** (2013). "Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos." *Cell Res* **23**(4): 465-472.
- Chari, A., M. M. Golas, M. Klingenhager, N. Neuenkirchen, B. Sander, C. Englbrecht, A. Sickmann, H. Stark and U. Fischer** (2008). "An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs." *Cell* **135**(3): 497-509.

- Cogli, L., C. Progida, R. Lecci, R. Bramato, A. Kruttgen and C. Bucci** (2010). "CMT2B-associated Rab7 mutants inhibit neurite outgrowth." Acta Neuropathol **120**(4): 491-501.
- Collins, J. E., S. White, S. M. Searle and D. L. Stemple** (2012). "Incorporating RNA-seq data into the zebrafish Ensembl genebuild." Genome Res **22**(10): 2067-2078.
- Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang** (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science **339**(6121): 819-823.
- Cornbrooks, C. J., D. J. Carey, J. A. McDonald, R. Timpl and R. P. Bunge** (1983). "In vivo and in vitro observations on laminin production by Schwann cells." Proc Natl Acad Sci U S A **80**(12): 3850-3854.
- Costa, C. Z., S. E. da Rosa and M. M. de Camargo** (2011). "The unfolded protein response: how protein folding became a restrictive aspect for innate immunity and B lymphocytes." Scand J Immunol **73**(5): 436-448.
- Court, F. A., W. T. Hendriks, H. D. MacGillavry, J. Alvarez and J. van Minnen** (2008). "Schwann cell to axon transfer of ribosomes: toward a novel understanding of the role of glia in the nervous system." J Neurosci **28**(43): 11024-11029.
- Court, F. A., R. Midha, B. A. Cisterna, J. Grochmal, A. Shakhbazau, W. T. Hendriks and J. Van Minnen** (2011). "Morphological evidence for a transport of ribosomes from Schwann cells to regenerating axons." Glia **59**(10): 1529-1539.
- Cox, G. A., C. L. Mahaffey and W. N. Frankel** (1998). "Identification of the mouse neuromuscular degeneration gene and mapping of a second site suppressor allele." Neuron **21**(6): 1327-1337.
- Cullen, P. J.** (2008). "Endosomal sorting and signalling: an emerging role for sorting nexins." Nat Rev Mol Cell Biol **9**(7): 574-582.
- Dale, J. M., H. Shen, D. M. Barry, V. B. Garcia, F. F. Rose, Jr., C. L. Lorson and M. L. Garcia** (2011). "The spinal muscular atrophy mouse model, SMADelta7, displays altered axonal transport without global neurofilament alterations." Acta Neuropathol **122**(3): 331-341.
- de Planell-Saguer, M., D. G. Schroeder, M. C. Rodicio, G. A. Cox and Z. Mourelatos** (2009). "Biochemical and genetic evidence for a role of IGHMBP2 in the translational machinery." Hum Mol Genet **18**(12): 2115-2126.
- Dean, C., F. G. Scholl, J. Choih, S. DeMaria, J. Berger, E. Isacoff and P. Scheiffele** (2003). "Neurexin mediates the assembly of presynaptic terminals." Nat Neurosci **6**(7): 708-716.
- Deinhardt, K., S. Salinas, C. Verastegui, R. Watson, D. Worth, S. Hanrahan, C. Bucci and G. Schiavo** (2006). "Rab5 and Rab7 control endocytic sorting along the axonal retrograde transport pathway." Neuron **52**(2): 293-305.
- Diers, A., M. Kaczinski, K. Grohmann, C. Hubner and G. Stoltenburg-Didinger** (2005). "The ultrastructure of peripheral nerve, motor end-plate and skeletal muscle in patients suffering from spinal muscular atrophy with respiratory distress type 1 (SMARD1)." Acta Neuropathol **110**(3): 289-297.

- Doktor, T. K., Y. Hua, H. S. Andersen, S. Broner, Y. H. Liu, A. Wieckowska, M. Dembic, G. H. Bruun, A. R. Krainer and B. S. Andresen** (2016). "RNA-sequencing of a mouse-model of spinal muscular atrophy reveals tissue-wide changes in splicing of U12-dependent introns." Nucleic Acids Res.
- Dombert, B., R. Sivadasan, C. M. Simon, S. Jablonka and M. Sendtner** (2014). "Presynaptic localization of Smn and hnRNP R in axon terminals of embryonic and postnatal mouse motoneurons." PLoS One **9**(10): e110846.
- Durai, S., M. Mani, K. Kandavelou, J. Wu, M. H. Porteus and S. Chandrasegaran** (2005). "Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells." Nucleic Acids Res **33**(18): 5978-5990.
- Edens, B. M., S. Ajroud-Driss, L. Ma and Y. C. Ma** (2015). "Molecular mechanisms and animal models of spinal muscular atrophy." Biochim Biophys Acta **1852**(4): 685-692.
- Eroglu, C. and B. A. Barres** (2010). "Regulation of synaptic connectivity by glia." Nature **468**(7321): 223-231.
- Feng, Z. and C. P. Ko** (2008). "Schwann cells promote synaptogenesis at the neuromuscular junction via transforming growth factor-beta1." J Neurosci **28**(39): 9599-9609.
- Fernandez-Nogales, M., M. Santos-Galindo, I. H. Hernandez, J. R. Cabrera and J. J. Lucas** (2016). "Faulty splicing and cytoskeleton abnormalities in Huntington's disease." Brain Pathol.
- Fevrier, B., D. Vilette, F. Archer, D. Loew, W. Faigle, M. Vidal, H. Laude and G. Raposo** (2004). "Cells release prions in association with exosomes." Proc Natl Acad Sci U S A **101**(26): 9683-9688.
- Fischer, U., Q. Liu and G. Dreyfuss** (1997). "The SMN-SIP1 Complex Has an Essential Role in Spliceosomal snRNP Biogenesis." Cell **90**: 1023-1029.
- Fogel, A. I., M. R. Akins, A. J. Krupp, M. Stagi, V. Stein and T. Biederer** (2007). "SynCAMs organize synapses through heterophilic adhesion." J Neurosci **27**(46): 12516-12530.
- Gabanella, F., M. E. Butchbach, L. Saieva, C. Carissimi, A. H. Burghes and L. Pellizzoni** (2007). "Ribonucleoprotein assembly defects correlate with spinal muscular atrophy severity and preferentially affect a subset of spliceosomal snRNPs." PLoS One **2**(9): e921.
- Georgiou, J., R. Robitaille, W. S. Trimble and M. P. Charlton** (1994). "Synaptic regulation of glial protein expression in vivo." Neuron **12**(2): 443-455.
- Gombash, S. E., C. J. Cowley, J. A. Fitzgerald, C. C. Iyer, D. Fried, V. L. McGovern, K. C. Williams, A. H. Burghes, F. L. Christofi, B. D. Gulbransen and K. D. Foust** (2015). "SMN deficiency disrupts gastrointestinal and enteric nervous system function in mice." Hum Mol Genet **24**(13): 3847-3860.
- Grimm, C., A. Chari, J. P. Pelz, J. Kuper, C. Kisker, K. Diederichs, H. Stark, H. Schindelin and U. Fischer** (2013). "Structural basis of assembly chaperone-mediated snRNP formation." Mol Cell **49**(4): 692-703.
- Grohmann, K., W. Rossoll, I. Kobsar, B. Holtmann, S. Jablonka, C. Wessig, G. Stoltenburg-Didinger, U. Fischer, C. Hubner, R. Martini and M. Sendtner** (2004). "Characterization of Ighmbp2 in motor neurons

and implications for the pathomechanism in a mouse model of human spinal muscular atrophy with respiratory distress type 1 (SMARD1)." Hum Mol Genet **13**(18): 2031-2042.

**Grohmann, K., M. Schuelke, A. Diers, K. Hoffmann, B. Lucke, C. Adams, E. Bertini, H. Leonhardt-Horti, F. Muntoni, R. Ouvrier, A. Pfeufer, R. Rossi, L. Van Maldergem, J. M. Wilmshurst, T. F. Wienker, M. Sendtner, S. Rudnik-Schoneborn, K. Zerres and C. Hubner** (2001). "Mutations in the gene encoding immunoglobulin mu-binding protein 2 cause spinal muscular atrophy with respiratory distress type 1." Nat Genet **29**(1): 75-77.

**Gubitz, A. K., W. Feng and G. Dreyfuss** (2004). "The SMN complex." Exp Cell Res **296**(1): 51-56.

**Guenther, U. P., L. Handoko, B. Laggerbauer, S. Jablonka, A. Chari, M. Alzheimer, J. Ohmer, O. Plottner, N. Gehring, A. Sickmann, K. von Au, M. Schuelke and U. Fischer** (2009). "IGHMBP2 is a ribosome-associated helicase inactive in the neuromuscular disorder distal SMA type 1 (DSMA1)." Hum Mol Genet **18**(7): 1288-1300.

**Guenther, U. P., R. Varon, M. Schlicke, V. Dutrannoy, A. Volk, C. Hubner, K. von Au and M. Schuelke** (2007). "Clinical and mutational profile in spinal muscular atrophy with respiratory distress (SMARD): defining novel phenotypes through hierarchical cluster analysis." Hum Mutat **28**(8): 808-815.

**Hafezparast, M., R. Klocke, C. Ruhrberg, A. Marquardt, A. Ahmad-Annuar, S. Bowen, G. Lalli, A. S. Witherden, H. Hummerich, S. Nicholson, P. J. Morgan, R. Oozageer, J. V. Priestley, S. Averill, V. R. King, S. Ball, J. Peters, T. Toda, A. Yamamoto, Y. Hiraoka, M. Augustin, D. Korthaus, S. Wattler, P. Wabnitz, C. Dickneite, S. Lampel, F. Boehme, G. Peraus, A. Popp, M. Rudelius, J. Schlegel, H. Fuchs, M. Hrabe de Angelis, G. Schiavo, D. T. Shima, A. P. Russ, G. Stumm, J. E. Martin and E. M. Fisher** (2003). "Mutations in dynein link motor neuron degeneration to defects in retrograde transport." Science **300**(5620): 808-812.

**Hamilton, G. and T. H. Gillingwater** (2013). "Spinal muscular atrophy: going beyond the motor neuron." Trends Mol Med **19**(1): 40-50.

**Hao le, T., P. Q. Duy, J. D. Jontes and C. E. Beattie** (2015). "Motoneuron development influences dorsal root ganglia survival and Schwann cell development in a vertebrate model of spinal muscular atrophy." Hum Mol Genet **24**(2): 346-360.

**Hartline, D. K. and D. R. Colman** (2007). "Rapid conduction and the evolution of giant axons and myelinated fibers." Curr Biol **17**(1): R29-35.

**Head, S. R., H. K. Komori, S. A. LaMere, T. Whisenant, F. Van Nieuwerburgh, D. R. Salomon and P. Ordoukhanian** (2014). "Library construction for next-generation sequencing: overviews and challenges." Biotechniques **56**(2): 61-64, 66, 68, passim.

**Heier, C. R., R. Satta, C. Lutz and C. J. DiDonato** (2010). "Arrhythmia and cardiac defects are a feature of spinal muscular atrophy model mice." Hum Mol Genet **19**(20): 3906-3918.

**Hendry, I. A., K. Stockel, H. Thoenen and L. L. Iverson** (1974). "The retrograde axonal transport of Nerve growth factor." Brain Res **68**: 103-121.

- Hsieh-Li, H. M., J. G. Chang, Y. J. Jong, M. H. Wu, N. M. Wang, C. H. Tsai and H. Li** (2000). "A mouse model for spinal muscular atrophy." Nat Genet **24**(1): 66-70.
- Hua, Y., Y. H. Liu, K. Sahashi, F. Rigo, C. F. Bennett and A. R. Krainer** (2015). "Motor neuron cell-nonautonomous rescue of spinal muscular atrophy phenotypes in mild and severe transgenic mouse models." Genes Dev **29**(3): 288-297.
- Hua, Y., K. Sahashi, F. Rigo, G. Hung, G. Horev, C. F. Bennett and A. R. Krainer** (2011). "Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model." Nature **478**(7367): 123-126.
- Huang, E. J. and L. F. Reichardt** (2001). "Neurotrophins: roles in neuronal development and function." Annu Rev Neurosci **24**: 677-736.
- Huang, X., Y. Huang, R. Chinnappan, C. Bocchini, M. C. Gustin and M. Stern** (2002). "The Drosophila inebriated-encoded neurotransmitter/osmolyte transporter: dual roles in the control of neuronal excitability and the osmotic stress response." Genetics **160**(2): 561-569.
- Huber, J., U. Cronshagen, M. Kadokura, C. Marshallsay, T. Wada, M. Sekine and R. Luhrmann** (1998). "Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure." EMBO J **17**(14): 4114-4126.
- Huettner, J. E. and R. W. Baughman** (1986). "Primary Culture of Identified Neurons from the Visual Cortex of Postnatal Rats." The Journal of Neuroscience **6**(10): 3044-3060.
- Hunter, G., A. Aghamaleky Sarvestany, S. L. Roche, R. C. Symes and T. H. Gillingwater** (2014). "SMN-dependent intrinsic defects in Schwann cells in mouse models of spinal muscular atrophy." Hum Mol Genet **23**(9): 2235-2250.
- Huo, Q., M. Kayikci, P. Odermatt, K. Meyer, O. Michels, S. Saxena, J. Ule and D. Schumperli** (2014). "Splicing changes in SMA mouse motoneurons and SMN-depleted neuroblastoma cells: evidence for involvement of splicing regulatory proteins." RNA Biol **11**(11): 1430-1446.
- Hutagalung, A. H. and P. J. Novick** (2011). "Role of Rab GTPases in membrane traffic and cell physiology." Physiol Rev **91**(1): 119-149.
- Hwang, W. Y., Y. Fu, D. Reyon, M. L. Maeder, S. Q. Tsai, J. D. Sander, R. T. Peterson, J. R. Yeh and J. K. Joung** (2013). "Efficient genome editing in zebrafish using a CRISPR-Cas system." Nat Biotechnol **31**(3): 227-229.
- Iijima, T., K. Wu, H. Witte, Y. Hanno-Iijima, T. Glatzer, S. Richard and P. Scheiffele** (2011). "SAM68 regulates neuronal activity-dependent alternative splicing of neurexin-1." Cell **147**(7): 1601-1614.
- Imlach, W. L., E. S. Beck, B. J. Choi, F. Lotti, L. Pellizzoni and B. D. McCabe** (2012). "SMN is required for sensory-motor circuit function in Drosophila." Cell **151**(2): 427-439.
- Izaurralde, E., J. Lewis, C. Gamberi, A. Jarmolowski, C. McGuigan and I. W. Mattaj** (1995). "A cap-binding protein complex mediating U snRNA export." Nature **376**: 709-712.

- Jablonka, S., M. Beck, B. D. Lechner, C. Mayer and M. Sendtner** (2007). "Defective Ca<sup>2+</sup> channel clustering in axon terminals disturbs excitability in motoneurons in spinal muscular atrophy." J Cell Biol **179**(1): 139-149.
- Jahromi, B. S., R. Robitaille and M. P. Charlton** (1992). "Transmitter release increases intracellular calcium in perisynaptic schwann cells in situ." Neuron **8**: 1069-1077.
- Jao, L. E., S. R. Wente and W. B. Chen** (2013). "Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system." Proceedings of the National Academy of Sciences of the United States of America **110**(34): 13904-13909.
- Jessen, K. R.** (2004). "Glial cells." Int J Biochem Cell Biol **36**(10): 1861-1867.
- Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier** (2012). "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." Science **337**: 816-821.
- Kabadi, A. M., D. G. Ousterout, I. B. Hilton and C. A. Gersbach** (2014). "Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector." Nucleic Acids Res **42**(19): e147.
- Kanai, Y., N. Dohmae and N. Hirokawa** (2004). "Kinesin transports RNA: isolation and characterization of an RNA-transporting granule." Neuron **43**(4): 513-525.
- Kashmina, T. and J. L. Manley** (2003). "A negative element in *SMN2* exon 7 inhibits splicing in spinal muscular atrophy." Nat Genet **34**(4): 460-463.
- Kim, E., A. Goren and G. Ast** (2008). "Alternative splicing: current perspectives." Bioessays **30**(1): 38-47.
- Ko, C. P. and R. Robitaille** (2015). "Perisynaptic Schwann Cells at the Neuromuscular Synapse: Adaptable, Multitasking Glial Cells." Cold Spring Harb Perspect Biol **7**(10).
- Krieger, F., N. Elflein, R. Ruiz, J. Guerra, A. L. Serrano, E. Asan, L. Tabares and S. Jablonka** (2013). "Fast motor axon loss in *SMARD1* does not correspond to morphological and functional alterations of the NMJ." Neurobiol Dis **54**: 169-182.
- Kroiss, M., J. Schultz, J. Wiesner, A. Chari, A. Sickmann and U. Fischer** (2008). "Evolution of an RNP assembly system: a minimal SMN complex facilitates formation of UsnRNPs in *Drosophila melanogaster*." Proc Natl Acad Sci U S A **105**(29): 10045-10050.
- Lai, C. S. L., S. E. Fischer, J. A. Hurst, F. Vargha-Khadem and A. P. Monaco** (2001). "A forkhead-domain gene is mutated in a severe speech and language disorder." Nature **413**: 519-523.
- Larson, M. H., L. A. Gilbert, X. Wang, W. A. Lim, J. S. Weissman and L. S. Qi** (2013). "CRISPR interference (CRISPRi) for sequence-specific control of gene expression." Nat Protoc **8**(11): 2180-2196.
- Le, T. T., L. T. Pham, M. E. Butchbach, H. L. Zhang, U. R. Monani, D. D. Covert, T. O. Gavrilina, L. Xing, G. J. Bassell and A. H. Burghes** (2005). "SMN $\Delta$ 7, the major product of the centromeric survival motor neuron (*SMN2*) gene, extends survival in mice with spinal muscular atrophy and associates with full-length SMN." Hum Mol Genet **14**(6): 845-857.

- Lefebvre, S., L. Burglen, S. Reboullet, O. Clermont, P. Burlet, L. Viollet, B. Benichou, C. Cruaud, P. Millasseau, M. Zeviani, D. L. Paslier, J. Frezal, D. Cohen, J. Weissenbach, A. Munnich and J. Melki (1995). "Identification and Characterization of a Spinal Muscular Atrophy-Determining Gene." *Cell* **80**: 155-165.
- Lefebvre, S., P. Burlet, Q. Liu, S. Bertrand, O. Clermont, A. Munnich, G. Dreyfuss and J. Melki (1997). "Correlation between severity and SMN protein level in spinal muscular atrophy." *Nat Genet* **16**: 265-269.
- Lefebvre, S., P. Burlet, L. Viollet, S. Bertrand, C. Huber, C. Belser and A. Munnich (2002). "A novel association of the SMN protein with two major non-ribosomal nucleolar proteins and its implication in spinal muscular atrophy." *Hum Mol Genet* **11**(9): 1017-1027.
- Ling, K. K., M. Y. Lin, B. Zingg, Z. Feng and C. P. Ko (2010). "Synaptic defects in the spinal and neuromuscular circuitry in a mouse model of spinal muscular atrophy." *PLoS One* **5**(11): e15457.
- Little, D., C. F. Valori, C. A. Mutsaers, E. J. Bennett, M. Wyles, B. Sharrack, P. J. Shaw, T. H. Gillingwater, M. Azzouz and K. Ning (2015). "PTEN depletion decreases disease severity and modestly prolongs survival in a mouse model of spinal muscular atrophy." *Mol Ther* **23**(2): 270-277.
- Long, Y., L. Li, Q. Li, X. He and Z. Cui (2012). "Transcriptomic characterization of temperature stress responses in larval zebrafish." *PLoS One* **7**(5): e37209.
- Lopez-Verrilli, M. A. and F. A. Court (2012). "Transfer of vesicles from schwann cells to axons: a novel mechanism of communication in the peripheral nervous system." *Front Physiol* **3**: 205.
- Lopez-Verrilli, M. A., F. Picou and F. A. Court (2013). "Schwann cell-derived exosomes enhance axonal regeneration in the peripheral nervous system." *Glia* **61**(11): 1795-1806.
- Lotti, F., W. L. Imlach, L. Saieva, E. S. Beck, T. Hao le, D. K. Li, W. Jiao, G. Z. Mentis, C. E. Beattie, B. D. McCabe and L. Pellizzoni (2012). "An SMN-dependent U12 splicing event essential for motor circuit function." *Cell* **151**(2): 440-454.
- Lundmark, R. and S. R. Carlsson (2003). "Sorting nexin 9 participates in clathrin-mediated endocytosis through interactions with the core components." *J Biol Chem* **278**(47): 46772-46781.
- Lunn, M. L., R. Nassirpour, C. Arrabit, J. Tan, I. McLeod, C. M. Arias, P. E. Sawchenko, J. R. Yates, 3rd and P. A. Slesinger (2007). "A unique sorting nexin regulates trafficking of potassium channels via a PDZ domain interaction." *Nat Neurosci* **10**(10): 1249-1259.
- Lusk, R. W. (2014). "Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data." *PLoS One* **9**(10): e110808.
- Maeda, M., A. W. Harris, B. F. Kingham, C. J. Lumpkin, L. M. Opdenaker, S. M. McCahan, W. Wang and M. E. Butchbach (2014). "Transcriptome profiling of spinal muscular atrophy motor neurons derived from mouse embryonic stem cells." *PLoS One* **9**(9): e106818.
- Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville and G. M. Church (2013). "RNA-guided human genome engineering via Cas9." *Science* **339**(6121): 823-826.

- Martinez-Hernandez, R., C. Soler-Botija, E. Also, L. Alias, L. Caselles, I. Gich, S. Bernal and E. F. Tizzano** (2009). "The Developmental Pattern of Myotubes in Spinal Muscular Atrophy Indicates Prenatal Delay of Muscle Maturation." Journal of Neuropathology and Experimental Neurology **68**(5): 474-481.
- Matera, A. G. and Z. Wang** (2014). "A day in the life of the spliceosome." Nat Rev Mol Cell Biol **15**(2): 108-121.
- Mayeda, A., G. R. Screaton, S. D. Chandler, X. D. Fu and A. R. Krainer** (1999). "Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements." Molecular and Cellular Biology **19**(3): 1853-1863.
- McWhorter, M. L., U. R. Monani, A. H. Burghes and C. E. Beattie** (2003). "Knockdown of the survival motor neuron (Smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding." J Cell Biol **162**(5): 919-931.
- Meister, G., D. Buhler, R. Pillai, F. Lottspeich and U. Fischer** (2001). "A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs." NATURE CELL BIOLOGY **3**: 945-949.
- Mentis, G. Z., D. Blivis, W. Liu, E. Drobac, M. E. Crowder, L. Kong, F. J. Alvarez, C. J. Sumner and M. J. O'Donovan** (2011). "Early functional impairment of sensory-motor connectivity in a mouse model of spinal muscular atrophy." Neuron **69**(3): 453-467.
- Millecamps, S. and J. P. Julien** (2013). "Axonal transport deficits and neurodegenerative diseases." Nat Rev Neurosci **14**(3): 161-176.
- Miller, J. C., S. Tan, G. Qiao, K. A. Barlow, J. Wang, D. F. Xia, X. Meng, D. E. Paschon, E. Leung, S. J. Hinkley, G. P. Dulay, K. L. Hua, I. Ankoudinova, G. J. Cost, F. D. Urnov, H. S. Zhang, M. C. Holmes, L. Zhang, P. D. Gregory and E. J. Rebar** (2011). "A TALE nuclease architecture for efficient genome editing." Nat Biotechnol **29**(2): 143-148.
- Missler, M. and T. C. Sudhof** (1998). "Neurexins: three genes and 1001 products." Trends Genet **14**(1): 20-26.
- Molnar, G. M., A. Crozat, S. K. Kraeft, Q. P. Dou, L. B. Chen and A. B. Pardee** (1997). "Association of the mammalian helicase MAH with the pre-mRNA splicing complex." Proc Natl Acad Sci U S A **94**: 7831-7836.
- Monani, U. R.** (2005). "Spinal muscular atrophy: a deficiency in a ubiquitous protein; a motor neuron-specific disease." Neuron **48**(6): 885-896.
- Monani, U. R., M. Sendtner, D. D. Covert, D. W. Parsons, C. Andreassi, T. T. Le, S. Jablonka, B. Schrank, W. Rossoll, T. W. Prior, G. E. Morris and A. H. Burghes** (2000). "The human centromeric survival motor neuron gene (SMN2) rescues embryonic lethality in Smn (-/-) mice and results in a mouse with spinal muscular atrophy." Hum Mol Genet **9**(3): 333-339.
- Moreno-Mateos, M. A., C. E. Vejnar, J. D. Beaudoin, J. P. Fernandez, E. K. Mis, M. K. Khokha and A. J. Giraldez** (2015). "CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo." Nat Methods **12**(10): 982-988.

- Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer and B. Wold** (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq." Nat Methods **5**(7): 621-628.
- Mouaikel, J., C. Verheggen, E. Bertrand, J. Tazi and R. Bordonne** (2002). "Hypermethylation of the Cap Structure of Both Yeast snRNAs and snoRNAs Requires a Conserved Methyltransferase that Is Localized to the Nucleolus." Molecular Cell **9**: 891-901.
- Murray, L. M., L. H. Comley, D. Thomson, N. Parkinson, K. Talbot and T. H. Gillingwater** (2008). "Selective vulnerability of motor neurons and dissociation of pre- and post-synaptic pathology at the neuromuscular junction in mouse models of spinal muscular atrophy." Hum Mol Genet **17**(7): 949-962.
- Murray, L. M., S. Lee, D. Baumer, S. H. Parson, K. Talbot and T. H. Gillingwater** (2010). "Pre-symptomatic development of lower motor neuron connectivity in a mouse model of severe spinal muscular atrophy." Hum Mol Genet **19**(3): 420-433.
- Myers, P. Z., J. S. Eisen and M. Westerfield** (1986). "Development and Axonal Outgrowth of Identified Motoneurons in the zebrafish." The Journal of Neuroscience **6**(8): 2278-2289.
- Nakagawa, T., M. Setou, D. H. Seog, K. Ogasawara, N. Dohmae, K. Takio and N. Hirokawa** (2000). "A Novel Motor, KIF13A, Transports Mannose-6-Phosphate Receptor to Plasma Membrane through Direct Interaction with AP-1 Complex." Cell **103**: 569-581.
- Narayanan, U., T. Achsel, R. Luhrmann and A. G. Matera** (2004). "Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein." Mol Cell **16**(2): 223-234.
- Narayanan, U., J. K. Ospina, M. R. Frey, M. D. Hebert and A. Gregory** (2002). "SMN, the Spinal Muscular Atrophy Protein, Forms a Pre-Import Snrnp Complex with Snurportin1 and Importin  $\beta$ ." Hum Mol Genet **11**(15): 1785-1795.
- Narver, H. L., L. Kong, B. G. Burnett, D. W. Choe, M. Bosch-Marce, A. A. Taye, M. A. Eckhaus and C. J. Sumner** (2008). "Sustained improvement of spinal muscular atrophy mice treated with trichostatin A plus nutrition." Ann Neurol **64**(4): 465-470.
- Nave, K. A.** (2010). "Myelination and support of axonal integrity by glia." Nature **468**(7321): 244-252.
- Neuenkirchen, N., C. Englbrecht, J. Ohmer, T. Ziegenhals, A. Chari and U. Fischer** (2015). "Reconstitution of the human U snRNP assembly machinery reveals stepwise Sm protein organization." EMBO J **34**(14): 1925-1941.
- Ng, S. Y., B. S. Soh, N. Rodriguez-Muela, D. G. Hendrickson, F. Price, J. L. Rinn and L. L. Rubin** (2015). "Genome-wide RNA-Seq of Human Motor Neurons Implicates Selective ER Stress Activation in Spinal Muscular Atrophy." Cell Stem Cell **17**(5): 569-584.
- Ohno, M., A. Segref, A. Bachi, M. Wilm and I. W. Mattaj** (2000). "PHAX, a Mediator of U snRNA Nuclear Export Whose Activity Is Regulated by Phosphorylation." Cell **101**(2): 187-198.
- Ota, S., Y. Hisano, Y. Ikawa and A. Kawahara** (2014). "Multiple genome modifications by the CRISPR/Cas9 system in zebrafish." Genes Cells **19**(7): 555-564.

- Palacios, I., M. Hetzer, S. A. Adam and I. W. Mattaj** (1997). "Nuclear import of U snRNPs requires importin  $\beta$ ." EMBO J **16**(22): 6783-6792.
- Park, G. H., Y. Maeno-Hikichi, T. Awano, L. T. Landmesser and U. R. Monani** (2010). "Reduced survival of motor neuron (SMN) protein in motor neuronal progenitors functions cell autonomously to cause spinal muscular atrophy in model mice expressing the human centromeric (SMN2) gene." J Neurosci **30**(36): 12005-12019.
- Pearn, J. H., P. Hudgson and J. N. Walton** (1978). "A clinical and genetic study of Spinal Muscular Atrophy of adult onset- The autosomal recessive form as a discrete disease entity." Brain **101**: 591-606.
- Peippo, M. and J. Ignatius** (2011). "Pitt-Hopkins Syndrome." Molecular Syndromology.
- Pellizzoni, L., B. Charroux and G. Dreyfuss** (1999). "SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins." Proc Natl Acad Sci U S A **96**: 11167-11172.
- Pellizzoni, L., N. Kataoka, B. Charroux and G. Dreyfuss** (1998). "A Novel Function for SMN, the Spinal Muscular Atrophy Disease Gene Product, in Pre-mRNA Splicing." Cell **95**: 615-624.
- Pellizzoni, L., J. Yong and G. Dreyfuss** (2002). "Essential role for the SMN complex in the specificity of snRNP assembly." Science **298**(5599): 1775-1779.
- Pennisi, E.** (2013). "The CRISPR Craze." Science **341**: 833-836.
- Persson, E., S. Maday, M. M. Fu, A. J. Moughamian and E. L. Holzbaur** (2010). "Retrograde axonal transport: pathways to cell death?" Trends Neurosci **33**(7): 335-344.
- Pfeffer, S. R.** (1994). "Rab GTPases: master regulators of membrane trafficking." Curr Opin Cell Biol **6**(4): 522-526.
- Piazzon, N., F. Rage, F. Schlotter, H. Moine, C. Branlant and S. Massenet** (2008). "In vitro and in cellulo evidences for association of the survival of motor neuron complex with the fragile X mental retardation protein." J Biol Chem **283**(9): 5598-5610.
- Ponomareva, O. Y., K. W. Eliceiri and M. C. Halloran** (2016). "Charcot-Marie-Tooth 2b associated Rab7 mutations cause axon growth and guidance defects during vertebrate sensory neuron development." Neural Dev **11**: 2.
- Puls, I., C. Jonnakuty, B. H. LaMonte, E. L. Holzbaur, M. Tokito, E. Mann, M. K. Floeter, K. Bidus, D. Drayna, S. J. Oh, R. H. Brown, Jr., C. L. Ludlow and K. H. Fischbeck** (2003). "Mutant dynactin in motor neuron disease." Nat Genet **33**(4): 455-456.
- Qi, L. S., M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin and W. A. Lim** (2013). "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." Cell **152**(5): 1173-1183.
- Qian, X., Y. Ba, Q. Zhuang and G. Zhong** (2014). "RNA-Seq technology and its application in fish transcriptomics." OMICS **18**(2): 98-110.

- Qiu, S., S. Luo, O. Evgrafov, R. Li, G. P. Schroth, P. Levitt, J. A. Knowles and K. Wang** (2012). "Single-neuron RNA-Seq: technical feasibility and reproducibility." Front Genet **3**: 124.
- Reichardt, L. F.** (2006). "Neurotrophin-regulated signalling pathways." Philos Trans R Soc Lond B Biol Sci **361**(1473): 1545-1564.
- Richardson, G. M., J. Lannigan and I. G. Macara** (2015). "Does FACS perturb gene expression?" Cytometry A **87**(2): 166-175.
- Rosel, T. D., L. H. Hung, J. Medenbach, K. Donde, S. Starke, V. Benes, G. Ratsch and A. Bindereif** (2011). "RNA-Seq analysis in mutant zebrafish reveals role of U1C protein in alternative splicing regulation." EMBO J **30**(10): 1965-1976.
- Rossoll, W., S. Jablonka, C. Andreassi, A. K. Kroning, K. Karle, U. R. Monani and M. Sendtner** (2003). "Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of beta-actin mRNA in growth cones of motoneurons." J Cell Biol **163**(4): 801-812.
- Rossoll, W., A. K. Kroning, U. M. Ohndorf, C. Steegborn, S. Jablonka and M. Sendtner** (2002). "Specific interaction of Smn, the spinal muscular atrophy determining gene product, with hnRNP-R and gry-rbp-hnRNP-Q: a role for Smn in RNA processing in motor axons?" Hum Mol Genet **11**(1): 93-105.
- Rowen, L., J. Young, B. Birditt, A. Kaur, A. Madan, D. L. Philipps, S. Qin, P. Minx, R. K. Wilson, L. Hood and B. R. Graveley** (2002). "Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity." Genomics **79**(4): 587-597.
- Rozic-Kotliroff, G. and N. Zisapel** (2007). "Ca<sup>2+</sup>-dependent splicing of neurexin IIalpha." Biochem Biophys Res Commun **352**(1): 226-230.
- Ruiz, R., J. J. Casanas, L. Torres-Benito, R. Cano and L. Tabares** (2010). "Altered intracellular Ca<sup>2+</sup> homeostasis in nerve terminals of severe spinal muscular atrophy mice." J Neurosci **30**(3): 849-857.
- Saal, L., M. Briese, S. Kneitz, M. Glinka and M. Sendtner** (2014). "Subcellular transcriptome alterations in a cell culture model of spinal muscular atrophy point to widespread defects in axonal growth and presynaptic differentiation." RNA **20**(11): 1789-1802.
- Sanes, J. R. and J. W. Lichtman** (2001). "Induction, assembly, maturation and maintenance of a postsynaptic apparatus." Nature Reviews Neuroscience **2**: 791-805.
- Saxena, S., C. Bucci, J. Weis and A. Kruttgen** (2005). "The small GTPase Rab7 controls the endosomal trafficking and neuritogenic signaling of the nerve growth factor receptor TrkA." J Neurosci **25**(47): 10930-10940.
- See, K., P. Yadav, M. Giegerich, P. S. Cheong, M. Graf, H. Vyas, S. G. Lee, S. Mathavan, U. Fischer, M. Sendtner and C. Winkler** (2014). "SMN deficiency alters Nrnx2 expression and splicing in zebrafish and mouse models of spinal muscular atrophy." Hum Mol Genet **23**(7): 1754-1770.
- Seng, C. O., C. Magee, P. J. Young, C. L. Lorson and J. P. Allen** (2015). "The SMN structure reveals its crucial role in snRNP assembly." Hum Mol Genet **24**(8): 2138-2146.

- Setou, M., T. Nakagawa, D. H. Seog and N. Hirokawa** (2000). "Kinesin superfamily motor protein KIF17 and mLin-10 in NMDA receptor-containing vesicle transport." Science **288**(5472): 1796-1802.
- Shababi, M., C. L. Lorson and S. S. Rudnik-Schoneborn** (2013). "Spinal muscular atrophy: a motor neuron disorder or a multi-organ disease?" J Anat.
- Shah, A. N., C. F. Davey, A. C. Whitebirch, A. C. Miller and C. B. Moens** (2015). "Rapid reverse genetic screening using CRISPR in zebrafish." Nat Methods **12**(6): 535-540.
- Shanker, S., A. Paulson, H. J. Edenberg, A. Peak, A. Perera, Y. O. Alekseyev, N. Beckloff, N. J. Bivens, R. Donnelly, A. F. Gillaspay, D. Grove, W. Gu, N. Jafari, J. S. Kerley-Hamilton, R. H. Lyons, C. Tepper and C. M. Nicolet** (2015). "Evaluation of commercially available RNA amplification kits for RNA sequencing using very low input amounts of total RNA." J Biomol Tech **26**(1): 4-18.
- Shanmugarajan, S., K. J. Swoboda, S. T. Innaccone, W. L. Ries, B. L. Maria and S. V. Reddy** (2010). "Congenital bone fractures in spinal muscular atrophy- functional role for SMN protein in bone remodeling." Journal of Child Neurology **22**(8): 967-973.
- Shanmugarajan, S., E. Tsuruga, K. J. Swoboda, B. L. Maria, W. L. Ries and S. V. Reddy** (2009). "Bone loss in survival motor neuron (Smn<sup>(-/-)</sup> SMN2) genetic mouse model of spinal muscular atrophy." J Pathol **219**(1): 52-60.
- Shepard, P. J. and K. J. Hertel** (2009). "The SR protein family." Genome Biology **10**: 242.
- Shieh, S. Y., C. M. M. Stellrecht and M. J. Tsai** (1995). "Molecular characterization of the rat insulin enhancer-binding complex 3b2: Cloning of a binding factor with putative helicase motifs." The Journal of Biological Chemistry **270**(37): 21503-21508.
- Shihabuddin, L. S.** (2008). "Adult rodent spinal cord-derived neural stem cells: isolation and characterization." Methods in Molecular Biology **438**: 55-66.
- Shin, N., S. Lee, N. Ahn, S. A. Kim, S. G. Ahn, Z. YongPark and S. Chang** (2007). "Sorting nexin 9 interacts with dynamin 1 and N-WASP and coordinates synaptic vesicle endocytosis." J Biol Chem **282**(39): 28939-28950.
- Shoji, W. and M. Sato-Maeda** (2008). "Application of heat shock promoter in transgenic zebrafish." Dev Growth Differ **50**(6): 401-406.
- Sleeman, J. E. and A. I. Lamond** (1999). "Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway." Current Biology **9**: 1065-1074.
- Smith, I. W., M. Mikesch, Y. Lee and W. J. Thompson** (2013). "Terminal Schwann cells participate in the competition underlying neuromuscular synapse elimination." J Neurosci **33**(45): 17724-17736.
- Soehnge, H., X. Huang, M. Becker, P. Whitley, D. Conover and M. Stern** (1996). "A neurotransmitter transporter encoded by the *Drosophila* inebriated gene." Proc Natl Acad Sci U S A **93**: 13262-13267.
- Soulet, F., D. Yarar, M. Leonard and S. L. Schmid** (2005). "SNX9 regulates dynamin assembly and is required for efficient clathrin-mediated endocytosis." Mol Biol Cell **16**(4): 2058-2067.

**Sperling, R.** (2016). "The nuts and bolts of the endogenous spliceosome." Wiley Interdiscip Rev RNA.

**Strick, D. J. and L. A. Elferink** (2005). "Rab15 effector protein: a novel protein for receptor recycling from the endocytic recycling compartment." Mol Biol Cell **16**(12): 5699-5709.

**Strong, M. J., G. Xu, L. Morici, S. Splinter Bon-Durant, M. Baddoo, Z. Lin, C. Fewell, C. M. Taylor and E. K. Flemington** (2014). "Microbial contamination in next generation sequencing: implications for sequence-based analysis of clinical samples." PLoS Pathog **10**(11): e1004437.

**Sugnet, C. W., W. J. Kent, M. Ares, Jr. and D. Haussler** (2004). "Transcriptome and genome conservation of alternative splicing events in humans and mice." Pac Symp Biocomput: 66-77.

**Sullivan, K. A., B. Kim and E. L. Feldman** (2008). "Insulin-like growth factors in the peripheral nervous system." Endocrinology **149**(12): 5963-5971.

**Thomas, A. C., H. Williams, N. Seto-Salvia, C. Bacchelli, D. Jenkins, M. O'Sullivan, K. Mengrelis, M. Ishida, L. Ocaka, E. Chanudet, C. James, F. Lescai, G. Anderson, D. Morrogh, M. Ryten, A. J. Duncan, Y. J. Pai, J. M. Saraiva, F. Ramos, B. Farren, D. Saunders, B. Vernay, P. Gissen, A. Straatman-Iwanowska, F. Baas, N. W. Wood, J. Hershenson, H. Houlden, J. Hurst, R. Scott, M. Bitner-Glindzicz, G. E. Moore, S. B. Sousa and P. Stanier** (2014). "Mutations in SNX14 cause a distinctive autosomal-recessive cerebellar ataxia and intellectual disability syndrome." Am J Hum Genet **95**(5): 611-621.

**Todd, D. J., A. H. Lee and L. H. Glimcher** (2008). "The endoplasmic reticulum stress response in immunity and autoimmunity." Nat Rev Immunol **8**(9): 663-674.

**Trapp, B. D., Y. Itoyama, N. H. Sternberger, R. H. Quarles and H. Webster** (1981). "Immunocytochemical localization of P0 protein in Golgi complex membranes and myelin of developing rat Schwann cells." J Cell Biol **90**(1): 1-6.

**Trapp, B. D., G. J. Kidd, P. Hauer, E. Mulrenin, C. A. Haney and S. B. Andrews** (1995). "Polarization of myelinating Schwann cell surface membranes: role of microtubules and the trans-Golgi network." J Neurosci **15**(3 Pt 1): 1797-1807.

**Trapp, B. D. and R. H. Quarles** (1982). "Presence of the Myelin-associated Glycoprotein Correlates with Alterations in the Periodicity of Peripheral Myelin." J Cell Biol **92**(3): 877-882.

**Turunen, J. J., E. H. Niemela, B. Verma and M. J. Frilander** (2013). "The significant other: splicing by the minor spliceosome." Wiley Interdiscip Rev RNA **4**(1): 61-76.

**Uribe, R. A., T. Kwon, E. M. Marcotte and J. M. Gross** (2012). "Id2a functions to limit Notch pathway activity and thereby influence the transition from proliferation to differentiation of retinoblasts during zebrafish retinogenesis." Dev Biol **371**(2): 280-292.

**Vesterlund, L., H. Jiao, P. Unneberg, O. Hovatta and J. Kere** (2011). "The zebrafish transcriptome during early development." BMC Dev Biol **11**: 30.

**Vuyisich, M., A. Arefin, K. Davenport, S. Feng, C. Gleasner, K. McMurry, B. Parson-Quintana, J. Price, M. Scholz and P. Chain** (2014). "Facile, high quality sequencing of bacterial genomes from small amounts of DNA." Int J Genomics **2014**: 434575.

- Wang, H., H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang and R. Jaenisch** (2013). "One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering." Cell **153**(4): 910-918.
- Wang, X., Y. Zhao, X. Zhang, H. Badie, Y. Zhou, Y. Mu, L. S. Loo, L. Cai, R. C. Thompson, B. Yang, Y. Chen, P. F. Johnson, C. Wu, G. Bu, W. C. Mobley, D. Zhang, F. H. Gage, B. Ranscht, Y. W. Zhang, S. A. Lipton, W. Hong and H. Xu** (2013). "Loss of sorting nexin 27 contributes to excitatory synaptic dysfunction by modulating glutamate receptor recycling in Down's syndrome." Nat Med **19**(4): 473-480.
- Wang, Z., M. Gerstein and M. Snyder** (2009). "RNA-Seq- a revolutionary tool for transcriptomics." Nature Reviews Genetics **10**: 57-63.
- Wee, C. D., M. A. Havens, F. M. Jodelka and M. L. Hastings** (2014). "Targeting SR proteins improves SMN expression in spinal muscular atrophy cells." PLoS One **9**(12): e115205.
- Wiedenheft, B., S. H. Sternberg and J. A. Doudna** (2012). "RNA-guided genetic silencing systems in bacteria and archaea." Nature **482**(7385): 331-338.
- Will, C. L. and R. Luhrmann** (2001). "Spliceosomal UsnRNP biogenesis, structure and function." Curr Opin Cell Biol **13**(3): 290-301.
- Willems, B., S. Tao, T. Yu, A. Huyseune, P. E. Witten and C. Winkler** (2015). "The Wnt Co-Receptor Lrp5 Is Required for Cranial Neural Crest Cell Migration in Zebrafish." PLoS One **10**(6): e0131768.
- Winkler, C., C. Eggert, D. Gradl, G. Meister, M. Giegerich, D. Wedlich, B. Lagerbauer and U. Fischer** (2005). "Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy." Genes Dev **19**(19): 2320-2330.
- Wirth, B.** (2000). "An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA)." Human Mutation **15**: 228-237.
- Wittkopp, N., E. Huntzinger, C. Weiler, J. Sauliere, S. Schmidt, M. Sonawane and E. Izaurralde** (2009). "Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival." Mol Cell Biol **29**(13): 3517-3528.
- Xu, C. C., K. R. Denton, Z. B. Wang, X. Zhang and X. J. Li** (2016). "Abnormal mitochondrial transport and morphology as early pathological changes in human models of spinal muscular atrophy." Dis Model Mech **9**(1): 39-49.
- Young, P. J., C. J. DiDonato, D. Hu, R. Kothary, E. J. Androphy and C. L. Lorson** (2002). "SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 $\beta$ 1." Hum Mol Genet **11**(5): 577-587.
- Young, P. J., J. W. Francis, D. Lince, K. Coon, E. J. Androphy and C. L. Lorson** (2003). "The Ewing's sarcoma protein interacts with the Tudor domain of the survival motor neuron protein." Molecular Brain Research **119**(1): 37-49.
- Zahler, A. M., W. S. Lane, J. A. Stolk and M. B. Roth** (1992). "SR proteins: a conserved family of pre-mRNA splicing factors." Genes and Development **6**: 837-847.

- Zahler, A. M., K. M. Neugebauer, W. S. Lane and M. B. Roth** (1993). "Distinct functions of SR proteins in alternative pre-mRNA splicing." Science **260**: 219-222.
- Zetsche, B., J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, K. S. Makarova, P. Essletzbichler, S. E. Volz, J. Joung, J. van der Oost, A. Regev, E. V. Koonin and F. Zhang** (2015). "Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system." Cell **163**(3): 759-771.
- Zhang, Q., Y. C. J. Wang and E. A. Montalvo** (1999). "Subp-2 Represses the Epstein-Barr Virus Lytic Switch Promoter." Virology **255**: 160-170.
- Zhang, R., B. R. So, P. Li, J. Yong, T. Glisovic, L. Wan and G. Dreyfuss** (2011). "Structure of a Key Intermediate of the SMN Complex Reveals Gemin2's Crucial Function in snRNP Assembly." Cell **146**(3): 384-395.
- Zhang, Z., F. Lotti, K. Dittmar, I. Younis, L. Wan, M. Kasim and G. Dreyfuss** (2008). "SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing." Cell **133**(4): 585-600.
- Zhang, Z., A. M. Pinto, L. Wan, W. Wang, M. G. Berg, I. Oliva, L. N. Singh, C. Dengler, Z. Wei and G. Dreyfuss** (2013). "Dysregulation of synaptogenesis genes antecedes motor neuron pathology in spinal muscular atrophy." Proc Natl Acad Sci U S A **110**(48): 19348-19353.
- Zhiling, Y., E. Fujita, Y. Tanabe, T. Yamagata, T. Momoi and M. Y. Momoi** (2008). "Mutations in the gene encoding CADM1 are associated with autism spectrum disorder." Biochem Biophys Res Commun **377**(3): 926-929.
- Zuk, P. A. and L. A. Elferink** (2000). "Rab15 differentially regulates early endocytic trafficking." J Biol Chem **275**(35): 26754-26764.
- Zweier, C., E. K. de Jong, M. Zweier, A. Orrico, L. B. Ousager, A. L. Collins, E. K. Bijlsma, M. A. Oortveld, A. B. Ekici, A. Reis, A. Schenck and A. Rauch** (2009). "CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila." Am J Hum Genet **85**(5): 655-666.

