GAINING INSIGHTS FROM TRANSCRIPTOMIC SIGNATURES IN SPINAL MUSCULAR ATROPHY: IDENTIFICATION AND EXPLORATION OF NEW THERAPEUTIC TARGETS

Larissa GOLI

Department of Physiology, Anatomy, and Genetics

Keble College, University of Oxford

Thesis submitted for the

Degree of Doctor of Philosophy

Trinity 2022

Declaration

The work presented in this thesis was undertaken in Professor Matthew Wood's "Nucleic Acid Gene Therapy in Brain and Muscle" laboratory, at the Department of Physiology, Anatomy, and Genetics and at the Department of Paediatrics of the University of Oxford, between 2018 and 2022.

I, Larissa Goli, hereby declare that all work hereafter presented is my own, with the following exceptions:

Chapter 2: *The bioinformatics pipeline in Galaxy was suggested by Dr. Irina Chelysheva, who also helped with the pre-processing of the human RNAseq samples from fastq.gz to .count files. Implementation of clustering and correlation pipelines in Python were conducted with the assistance of Madison Tung.*

Chapter 3: *The Pip6a-PMO conjugates were produced by Dr. Frank Abendroth. Injections of Pip6a-PMO compounds into mice, as well as their dissections, were performed by Dr. Karin Meijboom, Dr. Suzan Hammond, and Dr. Melissa Bowerman. Laser-capture microdissection of motor neuron bodies was performed by Prof. Lyndsay Murray. TRIzol® RNA extraction from LCM-MN was performed by Dr. Karin Meijboom. The exon microarray was conducted at the Genomics Platform of the Karolinska Institute. Splicing analyses were implemented in Python with the assistance of Madison Tung.*

Chapter 5: *Dr. Yulia Lomonosova designed the five PMO sequences targeting mFoxo3. Pip8b2 synthesis and its conjugation to PMO moieties were conducted by Dr. Yahya Jad. Synthesis, purification, and LC-MS quality control on the FoxO3(DBE)-VHL PROTAC system compounds were performed with Dr. François Halloy.*

This work has not been submitted for any other degrees at any other University, educational institute, or institute of learning.

Acknowledgements

> *And they that be wise shall shine as the brightness of the firmament;*
> *and they that turn many to righteousness as the stars for ever and ever.*

I would like to address my heartful gratitude and thanks:

To my mother, Dr. Béatrice Goli, with whose example it all started; who gave me the two great gifts of a happy childhood and of love of learning; who, having repeated throughout my upbringing that my only dowry would be my degrees, instilled in me the importance of education; and whose reassurance and compassionate encouragements ensured that I completed this work.

To Lt. Madison Tung, whose energy, unextinguishable kindness, patience, and otherworldly brightness carried me as I was carrying this work around the world; who custom-built the desk upon which most of this thesis was written; who makes the journey a thousand times more exciting; who always makes me laugh at least once a day; and with whose love all continues.

To the teachers and researchers who first led me to pursue biomedical research, notably to Dr. Thien-Tri Lâm, to l'Ecole de l'INSERM-Liliane Bettencourt, and to Dr. Melissa Bowerman.

To all who have supervised this work: Dr. Yulia Lomonosova, Dr. Suzan Hammond, Prof. Matthew Wood, for the liberty, trust, and structure they have consistently provided. To Prof. Wood: thank you for believing in the potential of my proposal four years ago. To Yulia: I would not have been able to pursue this work without you. Your continued support and your ability to turn everything to its bright side is a gift and an inspiration.

To the other members of the Wood Group, most notably to Nina Ahlskog, Amarjit Bhomra, Katarzyna Chwalenia, Mathieu Fischer, Dr. François Halloy, Dr. Britt Hanson, Anna Kordala, Dr. Thomas Roberts, Jessica Stoodley, Nenad Svrzikapa, Dr. Audrey Winkelsas – you have set the bar very high for any future colleagues! To both Ania and Audrey: thank you for reminding me so often of what really matters in life and in scholarship. To Britt: what a treat to be your friend!

To past and to external collaborators whose conversations, advice, and occasional help supported this work, notably to Dr. Irina Chelysheva and Dr. Loïc Roux.

To the students whom I had the joy to supervise, most notably to Sai Ashraf and Bobby Gould, who were dedicated, hard-working, and curious – and inspired me to be the same.

To all who have helped with the revisions of this work.

To the Clarendon Fund, the Sloane-Robinson Scholarship of Keble College, the Department of Physiology, Anatomy and Genetics, the Keble Association, and the Goodger-Schorstein Scholarship, without whose funding this work would have neither begun at all, continued when the world had stopped, nor ended in good time.

To the administrators who have supported every step of the process, notably Sarah Noujaim.

To all who have made Keble College a second home and a place where love of neighbour and all sound learning forever flourish:
to the Rev. Dr. Nevsky Everett for his sincere and daily care for all, for the example of his faithfulness, and for his embodiment of gentleness, humour, scholarship, and kindness;
to friends in its Chapel Community – by name particularly grateful for its (to this day!) very chiefest apostle, Susannah Peppiatt; as well as to Jonathan Tanner and Amanda Westcott;
to past and present friends in its Chapel Choir, who, through two tours, three Early Music Festivals, three CD recordings, three and a half years of weekly services, and six BBC3 Choral Evensong broadcasts, endowed me with innumerable glorious memories, whose fellowship and musicality nourished me weekly, and for whose friendship I am indebted to this day – most notably its musical directors over the years, Matthew Martin and Paul Brough, to its administrator Pippa Thyne, and to Elizabeth Etches-Jones, Dr. Matthew Golesworthy, Bethany Lucas Ferguson, Jarek Jankowski, Benjamin Mills, and Dr. Lucy Oswald Mifsud;
to Paola Vargas Guttierez, for her honesty, support, and laughter in our Monday lunch breaks.

To those with whom I have trod and waited as the COVID-19 pandemic erupted, whose presence gave me certainty when the world had none – James Brahm, Diane Magnin, Matthew Rogers.

To those who have never stopped keeping me *rooted and grounded in love* – my partners in (large-scale) cooking crimes (‘*ora et labora et strawberries*’), favourite multiple offenders in (even larger scale) side projects, the best people with whom to scheme about what is good, true, and beautiful (amongst others: opera, movies, Dante, travels, countryside walks, candle-lit feasts) – the friends who have always helped me practice hope, and who have oh so consistently and oh so considerably reduced the weeping and gnashing of teeth experienced during this DPhil: Shanti Daffern, Dr. Carolyn Dreyer, and the Rev. James Mosher.

Thank you.

In te, Domine, speravi.
Dominus illuminatio mea.

Abstract

5q-spinal muscular atrophy (SMA) is a rare, autosomal recessive, progressive, and lethal neuromuscular disease, most frequently of paediatric onset, caused by the loss of function of the survival of motor neuron (*SMN1*) gene. *SMN1* codes for the SMN protein, whose best characterised role supports the biogenesis of the spliceosome and efficient splicing of pre-messenger RNAs. Twenty years of drug development have led to approval of three therapies in SMA, two of which are splice-switching therapies targeting the *SMN2* paralog and disease-modifying gene to *SMN1*. The ceiling effect of splice-switching therapies as well as issues associated with the specialised delivery of oligonucleotide therapeutics in neurological use calls for research towards the identification and functional exploration of new therapeutic targets beyond *SMN2* pre-mRNA.

In this work, I explore three avenues for the identification of new therapeutic targets in SMA, using transcriptomics data in motor neurons (MN) and in skeletal muscle. First, I study time- and treatment- dependent genome-wide transcriptomic changes in neuronal tissues in a target-agnostic way and develop a proof-of-concept meta-analysis of all published human RNAseq data in neuronal tissues. In addition, I report a microarray dataset in laser-captured microdissected motor neurons from a severe (Taiwanese) mouse model and provide the first demonstration of neuronal transcriptome-wide efficacy of an intravenously administered splice-switching therapy. Secondly, I investigate the changes in the transcription of a rare subpopulation of RNAs, circular RNAs (circRNA), derived from the *SMN1/2* loci. I show changes of expression of the most expressed *SMN*-derived circRNA, C2A-2B-3-4, in SMA fibroblasts, and provide computational evidence for their putative biological role. Finally, I design and develop two distinct pharmacological approaches to down-regulate FoxO3, a transcription factor transactivating the transcription of atrophy-inducing genes. I identify 8b2-(A) as a promising candidate for further *in vitro* development.

This work provides new insights in the neurodegenerative processes in MN by tackling the first meta-analysis of published neuronal transcriptomics datasets in SMA, revealing methodological gaps in the study of circular RNAs in a high-isoform context, and identifying novel groups of genes dysregulated in early post-natal MN development in SMA. The identified targets can be further explored for therapeutic development in SMA and in other neurodegenerative and muscle-wasting conditions.

Table of content

List of figures

List of tables

List of abbreviations

aa	aminoacid
AAV	adeno-associated virus
ABS	alternative back splicing
ACh	acetylcholine
AD	Alzheimer's disease
AIDS	acquired immunodeficiency syndrome
ALAT	alanine amino transferase
ALS	amyotrophic lateral sclerosis
AON	antisense oligonucleotide
ANOVA	analysis of variance
ASAT	aspartate amino transferase
ASO	antisense splice-switching oligonucleotide
ATP	adenosine 5'-triphosphate
bp	base pair
BCV	biological coefficient of variation
BMS	Biomedical Services
BSCB	blood-spinal cord barrier
BSJ	back-splice junction
CC	cellular compartment
cDNA	complementary DNA
circRNA	circular RNA
CNS	central nervous system
COPD	chronic obstructive pulmonary disease
CHOP INTEND	Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders
CPM	count per million
CSF	cerebrospinal fluid
Da	dalton
DAS	differentially alternatively spliced
DBD	DNA binding domain
DBE	*Daf-16* family member-binding element
DE	differentially expressed
DEPC	diethylpyrocarbonate
DEX	dexamethasone
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
EB	elution buffer
EBI	European Bioinformatics Institute
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
ELISA	enzyme-linked immunosorbent assay
EMBL	European Molecular Biology Laboratory
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
FBS	fetal bovine serum
FC	fold change
FDR	false discovery rate
FL	full length
FTD	frontotemporal dementia
GEO	Gene Expression Omnibus
GFF	general feature format
GO	gene ontology analysis
HCL	hierarchical clustering
HD	Huntington's disease
HFMSE	Hammersmith Functional Motor Scale Expanded

HINE-2	Hammersmith Infant Neurological Exam-Part 2
HPLC	high performance liquid chromatography
IP-3	inositol trisphosphate
iPSC	induced pluripotent stem cell
IRES	internal ribosomal entry site
ISS-N1	intronic splice site N1
IV	intravenous
IT	intrathecal
LCM	laser-capture microdissected
LC-MS	liquid chromatography mass spectrometry
LNA	locked nucleic acid
LPLC	low pressure liquid chromatography
MFE	minimum free energy
MHC	myosin heavy chain
miRNA/miR	microRNA
MOE	methoxyethyl
MN	motor neuron
mRNA	messenger RNA
MUP	major urinary protein
MW	molecular weight
NCBI	National Center for Biotechnology Information
NI	normalised intensity
NGS	next generation sequencing
NHS	N-hydroxysuccimide
NMJ	neuromuscular junction
nTPM	normalised transcript per million
nt	nucleotide
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCA	principal component analysis
PCR	polymerase chain reaction
PD	Parkinson's disease
PDB	Protein Data Bank
Pip	peptide nuclei acid/PMO internalisation peptide
PMO	phosphorodiamidate
PND	post-natal day
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROTAC	proteolysis targeting chimera
PS	penicillin/streptomycin
PVDF	polyvinylidene fluoride
RBP	RNA binding protein
rcf	relative centrifugal force
RIPA	radioimmunoprecipitation assay buffer
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RNP	ribonucleoprotein
RMA	robust multi-array average
rRNA	ribosomal RNA
RT	retrotranscription
SC	spinal cord
SCA	spinocerebellar ataxia
SDS	sodium dodecyl sulfate
SEM	standard error of the mean
SI	splicing index
SMA	spinal muscular atrophy
SMN/Smn	survival of motor neuron

snRNA	small nuclear ribonucleic acid
snRNP	small nuclear ribonuclear particle
snoRNP	small nucleolar ribonuclear particle
SRA	Sequence Read Archive
SRP	signal recognition particle
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TBE	Tris/Borate/EDTA
TF	transcription factor
TIMPSI	Test of Infant Motor Performance Screening Items
Tm	melting temperature
TMG	2,2,7-trimethylguanosine
t-SNE	t-distributed stochastic neighbor embedding
UPS	ubiquitin protease system
UTR	untranslated region
UV	ultraviolet
VHL	von Hippel Lindau
WT	wild type

1. Chapter 1: Spinal muscular atrophy: pathophysiology, RNA biology, and RNA-based therapeutics approaches

What is SMA?

Why focus on RNA in SMA?

What are the current therapies available in SMA?

Why are new therapeutic targets needed?

***

1.1. Spinal muscular atrophy (SMA)

1.1.1. SMA discovery, presentation, and natural history

Spinal muscular atrophy (SMA) is a rare, neuromuscular, degenerative disease of autosomal recessive inheritance, most frequently of paediatric onset. Today SMA is understood as a unified disease expressing itself on a spectrum of clinical presentations, but it has been discovered between 1891 and 1956 [1–9]as a series of cases and independent conditions (reviewed in [10,11]). Clinical features common to all SMA patients center around progressive proximal muscle weakness eventually leading to muscle atrophy, caused by the progressive degeneration of alpha motor neurons (MN) in the ventral horn of the spinal cord (SC) [12]. For paediatric patients, this presents as low muscle tone and/or low muscle power, decreased movement, and weakness in crying.

SMA clinical presentation varies in motor milestones achievement (rolling back to front, independent sitting, crawling on all fours, independent walking, running) as well as in disease onset, from antenatal to adult. This clinical spectrum has led to an initial unified classification in three [13–15] and later five major SMA subtypes (***Table*** 1-1) that is being revised towards a more functional classification (non-sitters, sitters, non-ambulant, ambulant) [16,17]. SMA type I is the most frequent type [18–20]. In addition to the impeded motor development, SMA patients present comorbidities including dysphagia [21–23], pulmonary morbidities [24–28], and musculoskeletal complications [29] (joint contractures [30–33], scoliosis [34–38]) which impact the ability to walk, eat, or breathe.

Table 1-1: SMA is classified in five types based on onset and achieved motor milestones

Type	Subtype	Onset	Symptoms & highest motor milestone	% SMA cases	Notes
0		Antenatal	Reduced/absent foetal movements, severe hypotonia, joint contractures No motor milestone (cannot sit nor roll)		
I	Ia Ib Ic	<1 month <3 months <6 months	Ia: joint contractures at birth Ic: may achieve head control [39] Sits with support only (severe hypotonia)	54.5% [18] – 60% [40]	Also known as: severe infantile acute SMA, Werdnig-Hoffmann disease (OMIM: 253300) Shortened lifespan
II		<18 months	Sits independently (proximal weakness, no independent standing or walking)	27% [40] - 31.9% [18]	Also known as infantile chronic SMA, Dubowitz disease (OMIM: 253550) Shortened lifespan
III	IIIa IIIb IIIc	<3 years > 3 years > 12 years	Independent walking (ability may decrease)	12% [40]-13.2% [18]	Also known as juvenile SMA, Wohlfart-Kugelberg-Welander disease (OMIM: 253400) Normal lifespan
IV		adulthood (20s/30s)	Walks in adulthood	0.4 % [18]	Also known as adult-onset SMA (OMIM: 271150) Normal lifespan

Natural history of SMA depends on the severity of the disease and management of comorbities, including gastrostomies and positive airway pressure devices [41]. Symptoms in SMA type I patients deteriorate most rapidly [42–45], with a slower decline in type II and type III patients [43,46].

1.1.2. SMA epidemiology

SMA is historically the leading genetic cause of infant death [47]. Reports of SMA incidence varies depending on type, molecular diagnostic method, size of a study's population, health care system characteristics, population with regards to ethnic origin [20] (carrier frequency is function of founders effects: 1 in 27 in North American Mennonites [41] – similar to other reports in worldwide founder populations [48–53], as opposed to 1 in 54 in a pan-ethnic context [54]), levels of consanguinity in the general population [55–57], and the presence (or lack of) national registries. Incidence calculated with pilot newborn screening programs (i.e. based on a genetic rather than clinical diagnosis) varies between 1 in 7,096 (Germany) [58,59] and 1 in 28,137 (State of New York, USA) [60]. In Europe, the prevalence of all SMA types is estimated to be approximately 1-2 people per 100,000 [19,20] (by contrast and to illustrate local reporting differences, France reports 120-150 new annual cases, or 15-19 per 100,000 [61]).

1.1.3. SMA molecular aetiology

Modern molecular biology research has revealed a shared aetiology between all SMA subtypes. Genetic linkage studies from the 1990's in type II and type III SMA patients showed that the gene responsible for SMA is located on chromosome 5q13 [62]. It was identified in 1995 as part of an 500 kb inverted segmented duplication with two highly homologous genes in the telomeric and centromeric regions of 5q13, respectively named *survival of motor neuron 1* and *2* (*SMN1* and *SMN2*) [63]. Loss of function of *SMN1* is responsible for 95% of SMA cases [64]. (The remaining 5% are phenocopies caused by the loss of function of genes other than *SMN1*, collectively classified as non-5q SMA [64–66]. Henceforth, 'SMA' will refer exclusively to 5q-SMA.)

The duplication of *SMN1* to *SMN2* occurred after the speciation of Neanderthals [67], and as such, *SMN2* is unique to *Homo sapiens* [68]. *SMN1* and *SMN2* differ by five nucleotides in the coding region [69]. A silent C-to-T transition in *SMN2* (position +6 of exon 7, i.e. in codon 280) is the only difference in the coding sequence between *SMN1/2* genes [70]. This transition creates an exon splicing silencer (ESS) in *SMN2* [71], which is a binding site for the known splicing repressor hnRNPA1 (heterogeneous nuclear ribonucleoprotein A1) [72–75]. In addition, it also disrupts the CACACGA heptameric binding motif of SF2/ASF, thereby abolishing the exon splicing enhancer

(ESE) [76,77]. This induces exon 7 skipping [69] in 90% of all *SMN2* pre-mRNA transcripts, and results in the synthesis of the SMN protein lacking exon 7 (SMNΔ7). SMNΔ7 protein isoforms lack some of the tyrosine-glycine domain (Y-G domain) [78] necessary for SMN oligomerization [79] and higher order SMN oligomers [80]. As the SMNΔ7 isoform is unstable and rapidly degraded [81], SMN self-association levels correlate with disease severity [82]. *SMN2* only produces ~10% of full-length SMN (FL-SMN) protein. As such, in SMA patients, copy numbers of *SMN2* act as the primary disease modifier, and correlate with clinical types.

Of note, *SMN2Δ7* is not the only isoform produced by alternative splicing of *SMN2*. *SMN1* is transcribed to 11 linear mRNA transcripts, and *SMN2* to 16 mRNA transcripts (as referenced in Ensembl) (***Table*** 1-2) – although other linear transcripts (including Δ5, Δ3, Δ3/4, and combinations thereof) have been reported [83].

Canonically, both *SMN1* and *SMN2* entail 9 exons (exon 1: 27aa, ex. 2a: 24aa, ex. 2b: 40 aa, ex. 3: 67 aa, ex. 4:51 aa, ex. 5: 32 aa, ex. 6: 37 aa, ex. 7: 16aa, ex. 8 in 3'UTR) [83]. Recent approaches have revealed novel exons, notably an exon 6B [84] and exons 9-12 in intergenic sequences downstream of exon 8 [85].

Table 1-2: Summary of linear transcripts from SMN1/2 genes

Transcript ID	Name	Length (bp)	Biotype	Translation ID	UNIPROT ID	Protein length (aa)
SMN1						
ENST00000380707.9 *(cf. NM_000344.4)*	SMN1-202	1482	Protein coding	ENSP00000370083.4	Q16637-1	294 = FL-SMN *(NP_000335.1)*
ENST00000506163.5	SMN1-204	1445		ENSP00000424926.1	Q16637-3	282
ENST00000514951.5	SMN1-209	1308		ENSP00000423298.1	B4DP61	227
ENST00000503079.6	SMN1-203	1219		ENSP00000428128.1	Q16637-2	262
ENST00000351205.8	SMN1-201	900		ENSP00000305857.5	Q16637-1	294
ENST00000625245.2	SMN1-211	885		ENSP00000486539.1	E7EQZ4	294
ENST00000506239.6	SMN1-205	1550	Nonsense mediated decay	ENSP00000422679.2	E7EQZ4	294
ENST00000507905.6	SMN1-206	431		ENSP00000430657.1	H0YBZ9	79
ENST00000510679.1	SMN1-207	1045	Processed transcript (no ORF)	No protein		
ENST00000513228.1	SMN1-208	499	Retained intron			
ENST00000518504.5	SMN1-210	453				
SMN2 (in grey, transcripts & protein products identical to SMN1)						
ENST00000380743.9 *(cf. NM_017411.4)*	SMN2-203	1482	Protein coding	ENSP00000370119.4	Q16637-1	294
ENST00000626847.2	SMN2-214	1440		ENSP00000486152.1	Q16637-3	282
ENST00000380742.8	SMN2-202	1388		ENSP00000370118.4	Q16637-2	262
ENST00000380741.8	SMN2-201	900		ENSP00000370117.5	Q16637-1	294
ENST00000628696.2	SMN2-215	885		ENSP00000486268.1	E7EQZ4	294
ENST00000511812.5	SMN2-210	867		ENSP00000424282.1	B4DP61	227
ENST00000638794.1	SMN2-216	846		ENSP00000492675.1	A0A1W2PRV5	282
ENST00000614240.4	SMN2-213	804		ENSP00000479279.1	Q16637-2	262
ENST00000506734.5	SMN2-206	1550	Nonsense mediated decay	ENSP00000424799.1	E7EQZ4	294
ENST00000507458.2	SMN2-207	1045		ENSP00000475331.1	U3KPX7	46
ENST00000511873.6	SMN2-211	431		ENSP00000475824.1	H0YBZ9	79
ENST00000503678.5	SMN2-204	761	Retained intron	No protein		
ENST00000505346.5	SMN2-205	707				
ENST00000514914.1	SMN2-212	662				
ENST00000509805.5	SMN2-209	499				
ENST00000508258.1	SMN2-208	211				

1.2. RNA biology in SMA

1.2.1. Roles of the SMN protein

The FL-SMN protein (294 aa) is highly conserved throughout evolution. It is expressed in *S. pombe* [86,87] and as well as in all metazoans [88,89]. Comparison of conservation thresholds between *SMN* homologs show that exons 2, 3, and 6 are essential to SMN function in mammals *in vivo* [89].

The first characterized function of SMN is as a chaperone of the biogenesis of small nuclear ribonuclear particles (snRNPs) [90,91], themselves responsible for the assembly of the spliceosome.

The spliceosome (reviewed in [92–96]) is a multimegadalton dynamic complex of snRNPs and more than 50 associated proteins. The major spliceosome consists of five snRNPs subcomplexes, called U1, U2, U4, U5, and U6 [97]. There also exists a so-called 'minor' spliceosome (reviewed in [98,99]) consisting of U12 [100], U11 [100], U4atac [101], U6atac [101], and U5 . Each snRNP consists of a snRNA and of associated proteins. snRNPs interact with each other and with numerous external proteins and compose themselves into a series of spliceosomal complexes (complex A: U1/U2 [102]; complex B: U1/U2-U4/U5/U6 [103,104]; catalytic complex C: U2/U5/U6 [105,106]) which accurately recognize and remove non-coding introns [107,108] from pre-messenger mRNAs. The major spliceosome recognizes consensus sites at the 5'-end (splice donor AG/gt), branch site (purine-A-pyrimidine), and 3'-end (splice acceptor: ag/purine) of introns. The splice donor and acceptor sites are different for the minor spliceosome [109]. The minor spliceosome is slower in splicing out introns. It excises 1 in 300 introns in the human genome [110].

SMN is primarily located both in the cytoplasm [111], where the SMN complex (consisting of SMN protein, Gemins 2-8, and Unrip [112]) loads Sm rings with pre-snRNA [113]and stabilizes the formation of the 2,2,7-trimethylguanosine (TMG) cap structure unto the pre-snRNA [114]. SMN facilitates nuclear re-import [115,116] to the nuclear bodies (including gems and Cajal bodies) for further snRNP maturation [117].

Besides its critical role in the spliceosomal assembly, SMN has a number of other functions in RNA biology, including: the intracellular traffic and axonal transport of mRNAs [111,118,119], the direct involvement in pre-mRNA splicing [120], the assembly of the U7 snRNP complex responsible for the 3' end processing of histone mRNAs [121,122], and the direct and stable interaction with a methyltransferase essential to snoRNPs biogenesis [123,124].

SMN also interacts directly with proteins involved in several other processes: DNA recombination and repair [125,126], telomerase biogenesis [127,128], transcription and transcription termination [129], co-translational targeting of secretory and membrane proteins to the endoplasmic reticulum via the ribonucleoprotein SRP (signal recognition particle) [130], in signal transduction, stress granule formation, and intracellular trafficking (reviewed in Singh et al. 2017 [131]). A growing body of evidence also supports the role of SMN in both general translation [132] and in the regulation of the translation of a subset of selenoprotein mRNAs [133]. Finally, a short isoform of SMN, called a-SMN (axonal SMN), which includes intron 3, is also involved in axonogenesis during motor neuron (MN) early development [134].

1.2.2. SMA as a neuronal spliceopathy

The crucial role of SMN in spliceosomal assembly and in pre-mRNA splicing led to multiple investigations of the transcriptome in SMA. Global transcriptomic changes have been described following the loss of function of the SMN protein. Moderate reduction in SMN levels reduces primarily U4, U5, U12, and U4atac snRNAs, whilst a 95% knockdown of SMN levels reduces most snRNAs except for U2 and U6 [135]. SMN preferentially reduces the accumulation of minor snRNPs [135,136], with several reports of inefficient U12-dependent splicing [98,137–139].

In messenger RNAs, SMN deficiency causes widespread intron retention [139,140] and exon skipping [135,141,142]. Pre-symptomatic transcriptomic changes (increased intron retention, alternative 3'- and 5'-splice sites) have been reported in the smn1(ok355) *C. elegans* model of SMA [143], as well as in laser-captured mouse MNs, with the dysregulation of synaptogenesis gene expression and/or splicing before disease onset [141,144,145], or at later stages [146].

Though SMN is ubiquitously expressed, and though multiple tissues are affected by transcriptome-wide expression and splicing changes in SMA [135,139], splicing changes particularly affect MN. Whilst non-MN cell modifiers of MN are likely involved [147], the heightened MN vulnerability to SMN deficiency may be linked to the greater proportion neuronal homeostasis genes among genes spliced by the minor spliceosome [148]. Alternatively, cell-type specific transcriptomics studies have identified a handful of MN-specific splicing events in genes with crucial roles in neurogenesis (*Chodl* [146]), neuromuscular junction maturation [149] *(Agrn* [141]*)* and axonal excitability (*Gria* [141]*, Neurexin2A* [150]). It must be noted here that there exist contradictions between reports, notably regarding the disease progression stage at which splicing changes occur. Furthermore, there is growing body of literature investigating the vulnerability factors specific to lower alpha motor neurons (as opposed to upper, cranial motor neurons) [144,151,152].

1.2.3. SMA as a muscle wasting disease

The clinical presentation in SMA relies heavily on the effect of the disease on muscle. There are two primary mechanisms which cause muscle atrophy in SMA: MN-dependent muscle survival, and intrinsic myogenesis defects in SMA.

First, muscle development is heavily dependent on its innervation by MN. The neuromuscular junction (NMJ) is the acetylcholine (ACh)-dependent synaptic connection between MNs and muscle. At the NMJ, the electric signal from the MN is converted to a chemical signal triggering muscle contraction [153]. NMJs present three principal defects in SMA: they do not mature properly [154,155], they accumulate neurofilaments [156,157], and they denervate out of muscle before the onset of neuronal death [149,156,158]. In these conditions, SMA muscle presents fewer NMJ [159] with defects in post-synaptic ACh-receptor clustering [160] (reviewed in Boido and Vercelli, 2016 [161]), and downstream postsynaptic dysfunctions such as slower muscle action potentials [162,163].

Second, there are intrinsic defects in muscle development. Myotubes are smaller in SMA type I foetuses at early stages of foetal development (15 weeks of gestational age) [164] and there is evidence for defective myogenic processes in several SMA mouse models and in human SMA muscle [165–171], suggesting both SMN-dependent defects in myoblast fusion [166,167], and a putative role of SMN in satellite cells (muscle stem cells) [172,173], possibly via a role for SMN in miRNA regulation [174].

1.3. SMA: existing animal models

The *SMN2* gene is unique to humans [67,68]. This has led to the development of numerous animal models for the study of SMA pathogenic processes. *Smn* full knockdown is embryonically lethal (at the morula stage) in mice [175]. Most approaches to SMA modelling in mouse balance visibility of phenotype with *Smn* knockdown lethality by incorporating mutated human *SMN1* and/or *SMN2* transgenes into the murine genome. Extensive reviews of SMA model development have been previously published [176–180]. The Taiwanese and the SMNΔ7 are the two main models in SMA pre-clinical studies and are very similar with regards to symptoms and natural history. Below, I briefly summarize the mouse animal models that will be mentioned in the rest of this thesis (Table 1-3).

Table 1-3: Summary of the SMA mouse models relevant to this thesis

Name & ref.	Genetics	symptom onset	salient features	survival
severe (Taiwanese) [181]	•*Smn*$^{-/-}$; *SMN2(2Hung)*$^{-/-}$ *(25%)* •*Smn*$^{-/-}$; *SMN2(2Hung)*$^{+/-}$ *(50%)* •*Smn*$^{-/-}$; *SMN2(2Hung)*$^{+/+}$ *(25%)*	depends on *SMN2* copy number •in most severe model: NMJ defects at day 2, MN at day 3 [178]	MN loss, decreased weight, decreased motor function, muscle atrophy	• <10 days or 13.5 days • 2-4 weeks • normal, with tail necrosis
SMNΔ7 [182]	*SMNΔ7(Smn*$^{-/-}$*); SMN2(89Ahmb)*$^{+/+}$*; SMNΔ7*$^{+/+}$*)*	weight loss at day 2 NMJ defects at day 5 muscle weakness day 6 [178]		13.5 days
Gogliotti [183,184]	•Tg(SMN2)2Hung$^{tg/0}$; Smn1$^{tm1Hung/WT}$ (50%) •Tg(SMN2)2Hung$^{tg/0}$; Smn1$^{tm1Hung/tm1Hung}$ (50%)	day 5		13.5 days
Smn$^{2B/-}$ [185,186]	Mutation of exonic enhancer in *Smn* exon 7	day 21	Reduced MN number, muscle weakness, atrophy	28 days
exon 7 floxed *NSE-Cre, Smn*$^{F7/\Delta7}$ [187]	*Cre*-mediated *Smn* exon 7 excision in neurons	day 14	severe motor defect, tremors, muscle atrophy, pronounced indentations of nuclear membrane in MN	25 days

1.4. SMA: available therapies

Until December 2016, treatment of SMA patients was exclusively symptomatic (ventilation support, gastrostomy, orthopaedic surgeries). In the past five years however, three disease-modifying therapies (two gene therapies and one small molecule) have been approved in SMA by major regulatory bodies [188–190] (***Table*** 1-4).

1.4.1. Nusinersen, an oligonucleotide-based splicing enhancer of *SMN2*

Nusinersen (Spinraza™, Biogen/Ionis®) is a naked 18-mer 2´-O-(2-methoxyethyl)-phosphothioate oligodeoxy antisense nucleotide which prevents splicing factors from excluding *SMN2* exon 7 by sterically occupying the intronic splice silencing site *ISS-N1* [191,192]. Nusinersen increases quantities of FL-SMN in the spinal fluid [193] and in the spinal cord [194–196].

Nusinersen is approved for treatment in all types of SMA. Nusinersen is administered by intrathecal administration only, with a regimen of repeated administrations of 12 mg (in 5 mL). Doses #1-#3 are administered at 14 days intervals, dose #4 is administered 30 days after the third dose, followed by maintenance doses every 4 months.

Recently, two other therapies have also been approved for use in SMA: onasemnogene abeparvovec (Zolgensma™, Novartis®, FDA approval in May 2019) and risdiplam (Evrysdi™, Genentech/Roche®, FDA approval in August 2020).

1.4.2. Onasemnogene abeparvovec-xioi, an *SMN1* replacement gene therapy

Onasemnogene abeparvovec-xioi (previously known as AVXS-101) is a self-complementary AAV9-based gene therapy approved to treat children with all types of SMA under 2 years of age, i.e. patients who have either been diagnosed with SMA type I (bi-allelic mutations of *SMN1*) or have up to 3 copies of *SMN2*. The therapy contains $2x10^{13}$ vector genomes/mL administered at a recommended dose of $1.1x10^{14}$ vector genome per kg of body weight by a single slow (60 minutes) intravenous infusion. The vector genome is the full-length *SMN1* cDNA under the control of a cytomegalovirus enhancer/chicken-β-actin hybrid promoter. An advantage of this type of delivery method is that the vector genome does not integrate into the host genome, unlike lentiviral-based approaches. Pre-treatment approval includes the baseline measurement of platelets counts, hepatic biochemistry (ASAT/ALAT, total bilirubin, prothrombin time), troponin I baseline, and AAV9 serology by ELISA. To reduce the inflammatory response, treatment with onasemnogene abeparvovec-xioi is preceded by 24h of systemic corticosteroids (1 mg/kg/day) which are continued for 30 days (in the occasional persistence of liver function abnormalities, corticosteroid treatment is continued till hepatic biochemistry becomes unremarkable, after which the corticosteroid dose is gradually tapered).

1.4.3. Risdiplam, a small-molecule exonic splicing enhancer of *SMN2*

Risdiplam is a small molecule (previously known as RG7916) approved for children and adult SMA patients of all ages. It is an optimised design from the SMN-C compound series described by PTC-Roche [197,198]. Mechanism studies on SMN-C3 and SMN-C5 compounds show it targets exonic sites in exon 7 [197], with a particular affinity for an AG-rich motif [199] (reviewed in [200]).

Risdiplam is administered daily, orally, at home. The daily dose varies between 0.20 mg/kg (for patients younger than 2 years), to 0.25 mg/kg (patients >2 years but <20 kg), to 0.5 mg (patients >20 kg). Most common adverse effects include pyrexia (55% in early-onset patients, 22% in later-onset patients), rash (resp. 29% and 16.7%), diarrhea (resp. 19.4% and 16.7%), and headache (20% in later-onset patients).

Table 1-4: Summary of effectiveness and patient outcomes for the three approved SMA treatments

Therapy names Regulatory approval date	Class/chemistry Mechanism of action	Administration, regimen	Clinical trial, population & key patient outcomes
Nusinersen (Spinraza™, ISIS-SMN$_{Rx}$) Dec. 2016 (FDA) May 2017 (EMA)	non-viral gene therapy (MOE-ASO) *SMN2* splice-switching	Intrathecal, repeated	well tolerated; ↑ SMN protein CSF & improved motor function (HFMSE score) 9–14 months post treatment [201] ENDEAR trial: SMA type I infants <6 months: broad CNS distribution, ↑ *FL-SMN2;* ↑ SMN protein 6 months post treatment. Improved motor function (HINE-2 score), survival, electrophysiological function [202,203]. CHERISH trial: SMA type II children (2-12 years). Improved motor function (HFMSE) in 57% (vs. 26% in sham-treated) children. Early termination due to effectiveness. [17] SHINE trial (ongoing): long-term study of outcomes from ENDEAR and CHERISH trials [204] NURTURE trial (ongoing): safety, tolerability & efficacy in pre-symptomatic SMA patients. After 2.9 years follow-up: 100% survival without permanent ventilation. 100% sitting without support. 92% walking with assistance. 88% walking independently. [205] EMBRACE trial: SMA infants and children ineligible for ENDEAR or CHERISH. 79% patients show functional improvement (HINE-2) [206] DEVOTE trial (ongoing): safety, tolerability, potential greater efficacy of higher doses (two loading doses of 50 mg 14 days apart, followed by a maintenance dose of 28 mg every four months) RESPOND trial (recruiting): benefit of nusinersen in SMA patients with unmet clinical needs after Zolgensma™ ASCEND trial (upcoming): nusinersen in SMA patients with unmet clinical needs after risdiplam
Onasemnogene abeparvovec-xioi (Zolgensma™, AVXS-101, OAV101) May 2019 (FDA)	viral gene therapy (scAAV9-SMN1) *SMN1* replacement	IV, unique	START trial (ongoing 15 year follow-up): safety & tolerability of IV Zolgensma™ [207]. SPR1NT trial: IV Zolgensma™ in pre-symptomatic SMA infants <6 weeks. At 18 months: 100% independent siting, 100% survival without permanent ventilation, improved motor function (CHOP INTEND & TIMPSI), 0% need for gastrostomy [208]

March 2020 (EMA)			STR1VE & STR1VE-EU trials: safety & efficacy in symptomatic type I infants <6 months. At 18 months of age: 97% survival without permanent ventilation, 44% independent sitting, 79% head control [209,210] SMART trial (recruiting): safety, tolerability, and efficacy of a single IV dose of Zolgensma™ in type II SMA in patients between 8.5 and 21 kg STRONG trial: safety and efficacy of a single IT dose of Zolgensma™ in type II SMA. Improved motor function (HFMSE) as of Feb. 2022. STEER trial (upcoming): safety, tolerability, efficacy of a single IT dose of Zolgensma™ in SMA sitters (non-ambulant).
Risdiplam (Evrysdi™, RG7916) August 2020 (FDA) (extension to patients <2 months old in May 2022) Feb. 2021 (EMA) for patients >2 months	small molecule *SMN2* splice-switching	Oral, daily	FIREFISH trial: infants with SMA type I. After 14.8 months: ↑ FL-SMN protein in the blood by 3-fold (high dose) or 2-fold (low dose) from baseline SMN levels. No clinical conclusions from post-hoc motor function analyses [211] SUNFISH trial: SMA patients type II and type III. Functional motor improvement (MFM32) after 12 months of treatment [212]. JEWELFISH trial: risdiplam in non-naïve SMA patients (6 months-60 years). No treatment-related safety issues. >2 fold-increase in median SMN protein levels after >30 months [213]. RAINBOWFISH trial: risdiplam in pre-symptomatic SMA patients. After 12 months of treatment: functional motor improvement (CHOP-INTEND >60), maintenance of swallowing and feeding abilities [214].

1.5. Rationale for searching for new therapeutic targets in SMA

Treatment for SMA has seen incredible progress in recent years – namely, the approval of the three aforementioned disease-modifying therapies between 2016 and 2022. It must first be noted that this progress is collapsed into the very recent past after more than 100 years of study – taking off in the last decade primarily due to progress in molecular biology tools (in particular, whole human genome sequencing [215–217], progress in SMA pre-clinical models, and developments in oligonucleotide chemistry modifications [218,219]) and to exceptionally collaborative attitudes between SMA patients organisations, academic research teams, and industrial partners. Furthermore, the recent bloom in available treatments for SMA stands in stark contrast to the wider landscape of neurodegenerative diseases, many of which to date have only limited disease-modifying treatments available [220–222]. It is, nonetheless, the core argument undergirding this thesis and the rationale for its research that *the current available treatments do not address all the unmet clinical needs of SMA, and that further research must be undertaken*. Below, I outline several reasons for the necessity of further research. The body of this thesis will present possible directions for such research and treatment.

First, for any *SMN2*-targeting therapy, there is an inherent ceiling effect in the form of the existing, and finite, pool of *SMN2* transcripts determined by the copy numbers of a given patient [223]. This warrants the investigation of SMN-independent therapeutic strategies.

Secondly, oligonucleotide-based therapies offer the greatest flexibility for translation to other monogenic diseases – as recently illustrated by the facilitated regulatory path and approval of milasen [224], based on nusinersen's chemistry. However, a major limiting factor is the delivery of oligonucleotides to the central nervous system. Repeated intrathecal administrations lead to side effects in up to a third of patients [203] and can be complicated by severe scoliosis in SMA patients [34,225–227]. This warrants the investigation of more convenient modalities of administration of splice-switching therapies for neurological applications.

Thirdly, in the absence of systematic newborn screening programs in many countries around the world [58–60,228–230], there will continue to be diagnostic delays and therefore to delays to treatment onset [231]. This leads to one of the major limitations in SMA therapeutic ventures, that is, the limited number of MN still available at the time of treatment. In fact, the earlier patients are treated (before severe neurodegeneration occurs), the better the outcome [228,232,233]. Till early neonatal treatment is universally achieved, there will continue to be a requirement for additional therapeutic options for patients with partial neurodegeneration. This warrants the investigation of new molecular targets that can be targeted *after* disease onset.

Finally, the onset of novel, SMN-centred therapies will lead to new phenotypes and new natural histories both at the molecular level (novel dynamics of changes of splicing defect), at the cellular level (new natural histories of MN cellular death), and at the clinical level. It remains unknown how each of these new therapies integrate with lower- and higher- level changes, how they affect long-term quality of life, whether they reveal new pathological conditions both in the neuromuscular system and beyond, and what new therapeutic needs will arise in middle- to long-term. This warrants the systematic 'omics'-level investigation of molecular phenotypes after SMN-centred therapies, and the development of theranostic markers of molecular treatment efficacy.

Most of the ongoing novel clinical development pipelines in SMA focus on increasing muscle growth – either by inhibiting negative regulators of muscle growth (incl. apitegromab, GYM329, and BIIB110, all targeting myostatin [234–236]), or by activating muscle contractility (this includes reldesemtiv, a fast skeletal muscle troponin activator; and NMD-670, a small molecule inhibitor of the muscle-specific chloride ion channel ClC-1 [237,238]). As muscle survival is tightly regulated to its innervation [239], myostatin such approaches are unlikely to benefit patients lest MN innervation is preserved – such as what has been observed in Duchenne Muscular Dystrophy [240,241]. As such, the research work undergone and presented here can be complementary of current therapeutic development endeavours.

1.6. Thesis aims and hypotheses

The scope of this thesis encompasses and focuses on the neuromuscular system in SMA and in health, in mouse and in human. This thesis has three primary aims: (1) synthesise available transcriptomic data in motor neurons in SMA; (2) evaluate the molecular efficacy of an intravenously administered splice-switching therapy at the whole-transcriptome level in motor neurons in SMA; and (3) identify and functionally explore new therapeutic target(s) in SMA.

These aims were pursued based on four hypotheses. First (1), transcriptomics data can be used to uncover new therapeutic targets. Second (2), transcriptomics data can be used to evaluate a new drug candidate. Third (3), bioinformatics analysis of novel, rare transcriptome products can yield insight into new therapeutic targets. Finally (4), new nucleic acid-based therapies can be developed for non-SMN targets with potential both in SMA and in other muscle wasting diseases.

The remainder of this thesis seeks to develop the aforementioned aims through in-depth exploration of the associated hypotheses. There are four results chapter in this thesis followed by a summary of results and considerations for further development. In Chapter 2, I tackle the first reported meta-analysis of RNAseq transcriptomic datasets in SMA. In Chapter 3, I evaluate the efficacy of Pip6a-PMO, an intravenously administered *SMN2* targeting splice-switching therapy, in a exome microarray dataset in laser-captured microdissected MN. In Chapter 4, I move away from classical, linear transcription products as our study of *SMN*-derived circRNAs reveals important methodological gaps in the quantification of *SMN*-derived circRNAs and of other circRNAs in a high-isoform context. In Chapter 5, I report two novel pharmacological approaches to target FOXO3 in muscle atrophy. The final chapter summarizes results, evaluates them against our initial aims, and places them within the broader drug development context in neurodegenerative and muscle-wasting diseases.

1.7. A note regarding the genesis and the circumstances of this work

I have conducted the experimental design, pursuit, and analysis for this thesis from October 2018 to October 2022 in Professor Matthew Wood's laboratory at the University of Oxford. This DPhil initially started with a particular focus on blood-brain barrier crossing oligonucleotide delivery platforms, in continuity with work on Pip6a-PMO transcriptomics signatures in SMA mouse muscle previously undertaken by Dr. Karin Meijboom and Dr. Suzan Hammond, who was my supervisor in 2018. This work is the seed of what is now this thesis' Chapter 3.

Two events have occurred during my second year of studies that have drastically affected both the continuity and the direction of the work.

Firstly, in March 2020, the UK entered a national lockdown in response to the COVID-19 worldwide pandemic. This resulted in the total closure of the experimental facilities (both molecular biology laboratory and animal facilities) for three months, to the limited accessibility of the laboratory for a further six months, and in a major shift in the world's supply chains for experimental reagents. Effectually, this has put a stop to most of the experimental work constituting Chapter 4, as well as some of the validation work in Chapter 3.

Secondly, in September 2020, the primary supervisor of this work, Dr. Suzan Hammond resigned from her position of ten years as a Senior Researcher and SMA Group Leader in the Wood Group to transition to the biotech industry. Primary supervision was overtaken by Dr. Yulia Lomonosova, whose primary area of expertise is muscle biology, with a particular focus on muscle atrophy and Duchenne muscular dystrophy. This has led to a gradual shift out of the neuronal aspects of SMA into the investigation of its muscular dysfunctions as a target for combinatorial therapies. Chapter 5 is the consequence of this change of supervision.

Finally, throughout the deep changes of 2020, I have undertaken a shift towards dry lab methodologies – which has resulted in the expansion of the literature review from Chapter 1 to a chapter of its own: Chapter 2, which the reader will explore next.

2. Chapter 2: Meta-analysis of transcriptome alterations in neuronal tissues in SMA

2.1. Introduction

Numerous studies at the cellular and transcriptomic level have identified dysregulated transcripts and protein products that contribute to the understanding of neurodegeneration and SMA pathological processes. Each study focuses on a specific aspect of SMA pathology and seeks to contribute to the lingering unknowns of SMA: why are MNs particularly vulnerable to SMN deficiency? Are pre-symptomatic abnormalities a feature of SMA? Are there developmental delays in SMA MNs? What SMA-dysregulated molecular processes cause MN death in SMA?

These questions have been pursued since the elucidation of the genetic origin of SMA in 1995 [63], the identification of the roles of the SMN protein for spliceosome biogenesis [90,91,242], and have continued as new murine SMA disease models were developed and implemented (including [181,182,185], reviewed in [177]). Mice models have been used extensively in this period given the (clinically justified) absence of live human spinal cord biopsies, the difficulty of collecting SMA post-mortem samples [196], and the lack of good *in vitro* human motor neuron models. (Of note, fly [138,243–246], worm [143,247], zebrafish [248,249], and yeast [86,87,246,250–252] models, reviewed in [180,253] have also been developed but exceed the scope of our study). The 2008 discovery of stem cell reprogramming [254] allowed for the modelling of human motor neurons (and other central nervous system cell types, including astrocytes and oligodendrocytes) *in vitro* from patient samples (e.g. in [255–257]). This technological leap is concomitant with others in the field of genomics and transcriptomic profiling, with different generations of microarrays since their first inception in 1997 [258], and the onset of RNA sequencing (reviewed in [259]).

As such, transcriptomics studies in neuronal tissues in SMA exist, but are heterogeneous. For example, for transcriptomics studies in murine neuronal tissues, the first microarray study of SMA spinal cord in a mild model of SMA (pups heterozygous for *Smn*) in 2007 [260] was followed by microarrays in the Δ7 [135,146,261] and the *Smn*$^{-/-}$;*SMN2*$^{+/0}$ [262] mouse models, then by RNA-seq studies in Δ7, *Smn*$^{2B/-}$, and Taiwanese mouse models [139,141,144,152]. Comparison between findings in these studies is made difficult by the diversity of technologies, technological biases (e.g. batch effects in RNA-seq analyses), reference genome annotation updates, choice of analyses packages and parameters, and choice of output cutoffs.

At the time of writing, there is no review of all the published transcriptomics studies undertaken on neuronal samples in human and mouse samples. Secundary research, and particularly meta-analysis of published transcriptomics data, can bring new insight into characteristics of disease progression, and can consolidate and integrate knowledge of transcriptome alterations in SMA in MN across different models of the same disease [263–266]. This is achieved by the comparable and systematised processing of raw data from individual samples – including standardised mapping, alignment, normalisation, and differential expression analysis parameters.

The goal of this chapter, therefore, is to provide a meta-analysis of published SMA transcriptomics studies in human and murine models in neuronal tissues, by:

1. Presenting a comprehensive review of a chosen subset of the available literature;
2. Collecting the raw data from publicly available repositories;
3. Selecting and pre-processing samples in a standardised manner;
4. Evaluating the consistency and robustness of the processed data; and
5. Integrating and comparing studies with each other to elucidate transcriptome alterations in SMA neuronal tissue.

2.2. Materials and methods

2.2.1. Literature review, collection, and classification of raw data

2.2.1.1. Literature review

Datasets were collected between June and October 2021 from the Gene Omnibus repository (GEO datasets) and Sequence Read Archive (SRA) [267] from NCBI, and from the ArrayExpress Archive of Functional Genomics Data from EMBL-EBI, searching for "*spinal muscular atrophy AND transcript* AND profil* AND neuron* NOT bulbar*".

Datasets not linked to a publication as well as datasets not including *.cel* or *.fastq* raw files were initially excluded.

2.2.1.2. Data classification

Individual samples were classified manually based on file description on GEO, SRA, and ArrayExpress, as well as from cross-linking with publications. Processes were reported according to the updated PRISMA guidelines [268].

2.2.2. Pre-processing of RNA-seq samples

Reads were downloaded and extracted in FASTQ format from NCBI SRA to Galaxy servers version 21.09.rc1 [269], using the 'fastq-dump' function from SRA. Their quality was evaluated using FastQC [270] and compared using MultiQC [271]. Alternatively, SRR1016945, SRR1016946, SRR1016948, SRR1016949, SRR1016950, SRR1016951, SRR1016952 were locally downloaded before upload to the Galaxy servers.

Adapters of Illumina NGS data were trimmed using Trimmomatic [272].

Latest alignments (mouse: GRCm39; human: GRCh38.p13) were recovered from Ensembl [273].

All genomes were aligned into .bam files using RNA Star [274].

Count-aligned reads in .bam files were counted into a .count file by overlapping features the GFF files from the latest annotations, using htseq-count [275].

2.2.3. Library normalisation, correlation calculations, and covariate adjustments

Count files were processed in Python 3 [276] using pandas [277], anaconda [278], jupyter notebooks [279], NumPy [280], matplotlib (v. 3.6.0) [281], SciPy [282], and scikit-learn [283]. RNAseq samples were normalised using counts per millions (CPM), with

$$CPM = \frac{number\ of\ reads\ mapped\ to\ gene\ \times 10^6}{total\ number\ of\ mapped\ reads}$$

CPM calculations were validated using bioinfokit [284].

Batch correction was conducted in R [285] using the ComBat-Seq package [286].

2.2.4. Differential expression analyses

Following batch-effect correction, differential expression analysis was conducted in R using an exact test with *post-hoc* Benjamini-Hochberg correction after the removal of low count genes (CPM<1 in at least *'number of SMA samples – 1'* samples), using the edgeR package [287]. Genes with a fold-change equal or greater than two ($|\text{logFC}| \geq 1$), and an adjusted p.value $\leq$ 5%, were considered statistically differentially expressed between WT and SMA samples.

GO annotation from Ensembl was conducted using the g:CONVERT tool (g;PROFILER) [288] and confirmed using BioMart [289] (www.ensembl.org). Venn diagrams were plotted with matplotlib [281]. Gene ontology analysis was performed on the 931 transcripts identified as differentially expressed in the meta-analysis, with the g:SCS algorithm to compute the multiple testing correction for p-values gained from GO and pathway enrichment analysis, with alpha = 5%; with the g:GOST tool (g:PROFILER), version e106_eg53_p16_65fcd97 (access to the query: https://biit.cs.ut.ee/gplink/l/5DWhPZDbTZ).

2.3. Results

2.3.1. Collection of datasets in bioinformatics repositories

The purpose of this meta-analysis of transcriptomes being the re-analysis of publicly available transcriptome datasets, I chose to gather datasets from the U.S. National Center for BioTechnology Information (NCBI) and the European Bioinformatics Institute (EBI). Microarrays were collected from GeneExpression Omnibus (GEO) platform in NCBI and the ArrayExpress platform in EBI. RNAseq datasets were collected using the Sequence Read Archive (SRA) in NCBI. The "SMA", "neuron*", and "transcript*" search terms yielded 27 microarray datasets in GEO (14 in SRA and 19 in ArrayExpress). I excluded 21 studies: those with irrelevant designs, duplicates deposited both to NCBI and to EBI repositories, and studies that revolved around either muscle or brain samples. Due to the relative paucity of datasets available, I decided to keep datasets for which I could not retrieve raw files (*.cel* or *.fastq*) (incl. E-MEXP-131: 14 microarray samples from WT and Δ7 SMA mice spinal cords) as well as datasets not traceable to an available publication within EBI (incl. GSE65470: 8 SMA-patient derived iPSC MN microarray samples). This left 14 studies (***Figure*** 2-1***, Table*** 2-1***, Table*** 2-2) for downstream meta-analysis.

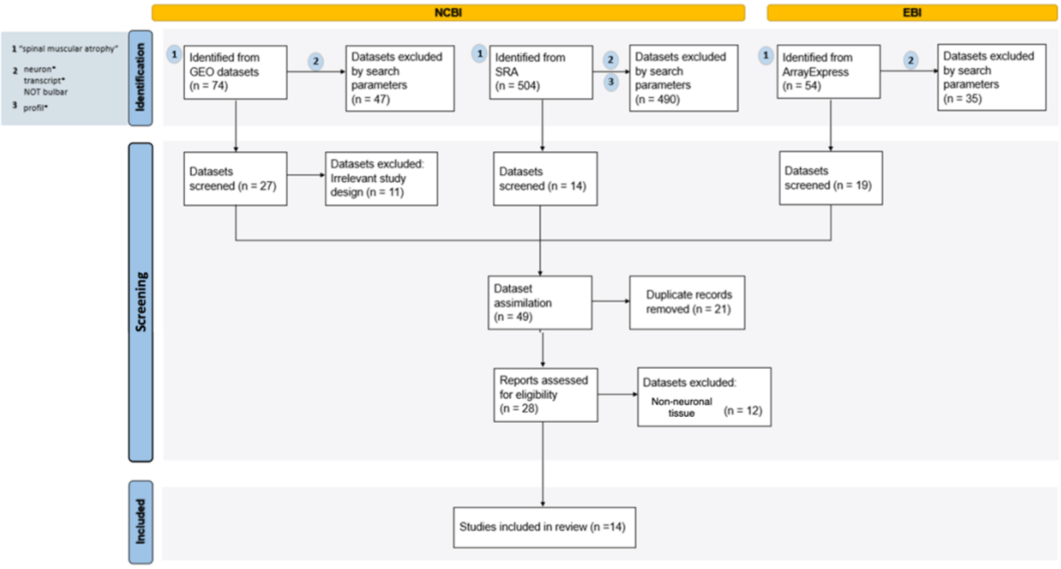


Figure 2-1: PRISMA flow diagram of the study selection process for human and mouse neuronal transcriptomic studies in SMA

Table 2-1: Review of NCBI-deposited transcriptomic studies in SMA in neuronal tissues in human and in mouse

	ORIGIN			MODEL							TECHNOLOGY		
	Reference	Deposit date	GSE	Species	Design	sample #	Genetics	Treatment	Control	Platform	Raw format	Processed	SRA Accession
1	Corti et al. 2012 [290]	Dec 2012	GSE27206	*Homo sapiens*	Microarray - SMA global expression profiles iPSC comparison to fibroblasts	9 (3/3/3)	iPSC patient: $SMN1\Delta7;SMN2^{+/+}$; iPSCs father: $SMN1\Delta7/+;SMN2^{+/+}$; control iPSC: $SMN1\Delta7/+;SMN2^{+/+}$		iPSC 19.9	GPL570 Affymetrix Human Genome U133 Plus 2.0 Array	.CEL	RMA, median polish	–
2	Yoshida et al. 2015 [160]	Apr 2018	GSE65470	*Homo sapiens*	Microarray analysis - patient derived iPSCs and fibroblasts control	8	SMA patient derived iPSCs		fibroblasts, biological triplicates	GPL17077 Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381 (Probe Name version)	.txt		–
3	Corti et al. 2012 [290]	Dec 2012	GSE27205	*Homo sapiens*	Microarray analysis - SMA global expression profile + splicing events + TGC-treatment	9 (3/3/3)	iPSC MN patient: $SMN1^{-/-};SMN2^{+/+}$ iPSC MN father: $SMN1^{+/-};SMN2^{+/+}$ iPSC MN treated: $SMN1^{-/-}$;SMN2+;SMN2(into1)	TGC treatment		GPL5175 Affymetrix Human Exon 1.0 ST Array	.CEL	RMA, moderated t-test	–
4	Boyd et al. 2017 [291]	Apr 2017	GSE86908	*Mus musculus*	Microarray analysis - total RNA from MN w/SMA background - specific motor neurons	9 (3/3/3)	normal FVB on SMA background		Biological triplicate	GPL20775 Affymetrix Mouse Transcriptome Array 1.0	.CEL	RMA normalised with Affymetrix	–
5	Hauser et al. 2020 [292]	Sep 2020	GSE157185	*Homo sapiens*	RNA-Seq across 5 cell types, healthy donors, validating MND models	10 (2/2/2/2/2)	GRCh37			GPL16791 Illumina HiSeq 2500	See SRA	FASTQ, STAR, featureCounts	SRP279443
6	Rizzo et al. 2019 [293]	Feb 2019	GSE108094	*Homo sapiens*	RNA-Seq of SMA IPSC MN to find common motif sequences (splicing/expressed genes)	8 (4/4)	SMA and control		Biological duplicates and technical duplicates	GPL11154 Illumina HiSeq 2000 (Homo sapiens)	See SRA	STAR2, cuffquant, cuffdiff, rMATS	SRP126773
7	Zhang et al. 2013 [141]	Nov 2013	GSE51735	*Mus musculus*	RNA-Seq transcriptome changes in SMA model before pathological changes, white matter glia and neurones	8 (4/4)	WT $SMN2^{+/+};SMN\Delta7^{+/+}$; $Smn^{-/-}$		WT	GPL13112 Illumina HiSeq 2000	See SRA	TopHat, BEDTools, DESeq, MISO	SRP032162
8	Maeda et al. 2014 [145]	Sep 2014	GSE56284	*Mus musculus*	RNA-Seq transcriptomes in SMA model and control using mESCs	6 (3/3)	Hb9 normal mESCs (WT and *$SMN2^{+/+};mSmn^{-/-}$)* A2 SMA mESCs (*$SMN2^{+/+};mSmn^{-/-}$*)		WT	GPL17021 Illumina HiSeq 2500	See SRA	FASTQC, CutAdapt, TOpHat, Cufflinks, Cuffdiff, CummeRbund	SRP040645
9	Yoshida et al. 2015 [160]	Apr 2015	GSE65508	*Homo sapiens*	RNA-Seq using patient-derived iPS MNs and VPA	4 (2/2)	iPSC derived motor neurone, h19 genome	1mM VPA for 6 days	Untreated and WT	GPL16791 Illumina HiSeq 2500	See SRA	BCL2FASTQ, Quality filtering, TopHat alignment, RPKM library	SRP053043

Table 2-2: Review of EBI-deposited transcriptomic studies in SMA in neuronal tissues in human and in mouse (SRA-duplicates in grey)

	ORIGIN			MODEL					TECHNOLOGY		
	Team	Dep. date	Accession	Species	Design	sample #	Genetics	Treatment	Control	Platform	Raw
2	Yoshida et al. 2015 [160]	Feb 2015	E-GEOD-65508	*Homo sapiens*	RNASeq - using patient-derived iPSC MNs and VPA	4	iPSC derived motor neurone, h19 genome	1mM VPA 6 days, change 1/2 medium every day	Untreated and WT	Illumina HiSeq 2500	SRA
2		Feb 2015	E-GEOD-65470	*Homo sapiens*	Microarray - patient derived iPSCs and fibroblasts control	8	SMA patient		fibroblasts, biological triplicates	A-GEOD-17077 - Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381 (Probe Name version)	.txt
8	Maeda et al. 2014 [145]	Dec 2014	E-GEOD-56284	*Mus musculus*	RNASeq - transcriptomes in SMA model and control using mESCs	6	Hb9 normal mESCs (WT and $SMN2^{+/+};mSmn^{-/-}$) A2 SMA mESCs ($SMN2^{+/+};mSmn^{-/-}$)		WT		SRA
6	Zhang et al. 2013 [141]	May 2014	E-GEOD-51735	*Mus musculus*	RNASeq - transcriptome changes in SMA model before pathological changes, white	8	WT SMN2$^{+/+}$;SMNΔ7$^{+/+}$;Smn$^{-/-}$		WT		
6	Superseries (5a) and (5b)		E-GEOD-27207	*Homo sapiens*							.CEL
6a	Rizzo et al. 2019 [293]	May 2014	E-GEOD-27206	*Homo sapiens*	Microarray - SMA global expression profiles iPSC comparison to fibroblasts	9	iPSC patient: $SMN1\Delta7;SMN2^{+/+}$; iPSCs father: $SMN1\Delta7/+;SMN2^{+/+}$; control iPSC: SMN1Δ7/+;SMN2$^{+/+}$		fibroblasts, biological triplicates		.CEL
6b		Jan 2013	E-GEOD-27205	*Homo sapiens*	Microarray - SMA global expression profile + splicing events + TGC-treatment	9	iPSC MN patient: $SMN1^{-/-};SMN2^{+/+}$ iPSC MN father: $SMN1^{+/-};SMN2^{+/+}$ iPSC MN treated: SMN1$^{-/-}$;SMN2+;SMN2(into1)	TGC treatment	Biological triplicate		.CEL
10	Corti et al. 2008 [294]	Sep 2008	E-GEOD-10224	*Mus musculus*	Microarray - potential application of stem cell therapy in SMA therapeutics	9 (3/3/3)	WT: SMN2$^{+/+}$;SmnΔ7$^{+/+}$;mSmn$^{+/-}$ SMA: SMN2$^{+/+}$;SmnΔ7$^{+/+}$; mSmn$^{-/-}$	Stem cell transplantation	Vehicle adminstered to wild type and SMA, n=3 in groups	GPL339 [MOE430A] Affymetrix Mouse Expression 430A Array	.CEL
11	Nizzardo et al. 2011 [295]	Dec 2013	E-GEOD-19674	*Mus musculus*	Microarray - WT, SMA and SMA with ceftriaxone treatment of spinal cord (lumbar region)	12 (4/4/4)	WT: SMN2$^{+/+}$;SmnΔ7$^{+/+}$;mSmn$^{+/-}$ SMA: SMN2$^{+/+}$;SmnΔ7$^{+/+}$; mSmn$^{-/-}$	Ceftriaxone admistered at P5; samples taken at P13	Vehicle adminstered to wild type and SMA, n=4 in all groups	GPL8321 [Mouse430A_2] Affymetrix Mouse Genome 430A 2.0 Array	.CEL
12	Murray et al. 2010 [261]	Nov 2009	E-MEXP-2428	*Mus musculus*	Microarray - SMA and WT spinal cord cells analysed at P1 and P5	15 (4/4/4/3)	SMA: mSmn$^{-/-}$;hSMN2		n = 4 for groups except SMA at P5 (n=3)	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	.CEL
13	Meyer et al. 2009 [296]	Mar 2015	E-MTAB-3170	*Mus musculus*	Microarray - comparison heterozygote carriers of SMA, SMA mice and SMA mice with U7 snRNA hSMN2 splicing correcting	12 (4/4/4)	mSmn$^{-/-}$; hSMN2+/+; U7-ESE-B1 & mSmn$^{-/-}$; hSMN2$^{+/+}$ & mSmn$^{+/-}$; hSMN2$^{+/+}$	U7 correction	Unaffected carrier, untreated, 4 repeats/condition	A-MEXP-2242 - Affymetrix Mouse GeneSplice Array	.CEL
14	Olaso et al. 2005 [297]	May 2005	E-MEXP-131	*Mus musculus*	Microarray – gene expression profile 10 days after Smn depletion in neurons	14 (6/8)	neuronal *(NSE-Cre, Smn*$^{F7/F7}$*)* mutant C57BL/6J mice		*Smn*$^{+/F7}$ littermates	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	.txt

2.3.2. Classification of datasets

The 14 studies included after the literature review consist of 9 microarray studies and 5 RNA-seq studies. 8 are murine and 6 are human. These studies include both SMA-centric studies as well as studies in non-SMA affected samples (or simply on SMA genetic background).

At this stage, I excluded treated samples from the 112 individual microarray samples and the 36 individual RNAseq samples (***Supplementary Table*** 7-1), leaving 82 mouse samples (68 mouse microarray samples, 14 mouse RNA seq samples) and 38 human samples (16 human microarray samples, 22 human RNAseq samples) (***Table*** 2-3).

Table 2-3: Classification of the 14 included studies according to species and technology.

SMA neuronal	Mouse	Human	Total
Microarrays	6 studies (n=68)	3 studies (n=16)	9 microarrays studies (n=84)
RNA-seq	2 studies (n=14)	3 studies (n=22)	5 RNA-seq studies (n=36)
Total	8 mouse studies (n=82)	6 human studies (n=38)	14 studies (n=120)

The 120 leftover samples are unevenly divided between early, pre-, and post-symptomatic stages. Late and pre-symptomatic samples constitute most of the data available (n=38 and n=37, respectively); early symptomatic samples represent only 7 samples (***Table*** 2-4). This summary confirms that to date, no single study systematically investigated the transcriptomic changes at all stages of the disease.

Table 2-4: Classification the 112 included samples according to species, technology, and symptomatic stage

	Type, GEO accession number, year, reference	Pre-symptomatic	Early symptomatic	Late symptomatic
Mouse	Microarrays			
	E-MEXP-131 (2005) [297]	n(WT)=3 n(SMA)=3		n(WT)=4 n(SMA)=4
	E-GEOD-10224 (2008) [294]			n(WT)=3 n(SMA)=3
	E-MEXP-2428 (2009) [261]	n(WT)=4 n(SMA)=4	n(WT)=4 n(SMA)=3	
	E-GEOD-19674 (2013) [295]			n(WT)=4 n(SMA)=4 n(SMAtreat)=4
	E-MTAB-3170 (2015) [296]			n(WT)=4 n(SMA)=4 n(SMAtreat)=4
	GSE86908 (2017) [291]	n(WT-MN)=9		
	RNA-seq			
	GSE51735 (2013) [141]	n(WT)=4 n(SMA)=4		
	GSE56284 (2014) [145]	n(WT)=3 n(SMA)=3		
Human	Microarrays			
	GSE27205 (2012) [290]			n(WT)=3 n(SMA)=3 n(SMAtreat)=3
	GSE27206 (2012) [290]			n(WT)=1 n(SMA)=2
	GSE65470 (2018) (not published)		n(WT)=2 n(SMA)=2	
	RNA-Seq			
	GSE65508 (2015) [160]	n(WT)=1 n(SMA)=1 n(WTtreated)=1 n(SMAtreat)=1		
	GSE108094 (2019) [293]			n(WT)=4 n(SMA)=4
	GSE157185 (2020) [292]			n(WT-neurons)=2 n(non-neurons)=8

With 120 transcriptomic samples ready to be included in the meta-analysis, I scoped the present study on a meta-analysis of the most recently published datapoints with the greatest technological coherence, that is, of the 5 SMA and 7 WT neuronal human RNA-seq datasets (excluding any treated samples). This limited analysis encompassed two types of sequencing technologies (Illumina HiSeq 2000 and Illumina HiSeq 2500) and served as a demonstration of feasibility for further cross-technological comparisons.

2.3.3. Pre-processing of individual human RNAseq samples

The 12 human RNAseq samples are taken from three studies dating from 2015 to 2020. They consist of three Illumina HiSeq studies. The SRP126773 dataset represents most of the datapoints (n=8) and includes biological replicates (***Table*** 2-5).

Table 2-5: Technical characteristics of the 12 samples included in the initial hRNAseq meta-analysis

ID (SRA, GEO, ref.)	Sample ID	Sample description	Sequencing technology	Library preparation
SRP126773 GSE108094 [293]	SRR6376963	SMA MN2 Replicate 2	Illumina HiSeq 2000	Illumina® TruSeq® Stranded mRNA Sample Preparation Kits enrichment of mRNAs by poly(A)-tail capture
	SRR6376962	SMA MN2 Replicate 1		
	SRR6376961	SMA MN1 Replicate 2		
	SRR6376960	SMA MN1 Replicate 1		
	SRR6376959	Control MN2 Replicate 2		
	SRR6376958	Control MN2 Replicate 1		
	SRR6376957	Control MN1 Replicate 2		
	SRR6376956	Control MN1 Replicate 1		
SRP279443 GSE157185 [292]	SRR12553375	WT iPSC2	Illumina HiSeq 2500	TruSeq mRNA v2Kit (polyA) – Stranded enrichment of mRNAs by poly(A)-tail capture
	SRR12553374	WT iPSC1		
SRP053043 GSE65508 [160]	SRR1784091	WT iPSC 201B7_MN VPA_minus	Illumina HiSeq 2500	TruSeq Stranded Total RNA Sample Prep Kit (Illumina) depletion of rRNAs by RiboZero
	SRR1784093	SMA iPSC 13_5_MN VPA_minus		

We used the web-based platform Galaxy for a systematic, parallelisable, and reproducible processing pipeline of all samples (***Figure*** 2-2). After upload, we ran quality control programs (fastqc and multiqc) to check the quality of the sequence reads and to compare the 12 files with each other. Samples of interest for this meta-analysis all passed quality control. We aligned reads on the latest version of the human genome available to date on Ensembl: GRCh38.p13 and counted using the latest GFF file from Ensembl. I normalised all counts as counts per million (CPM).

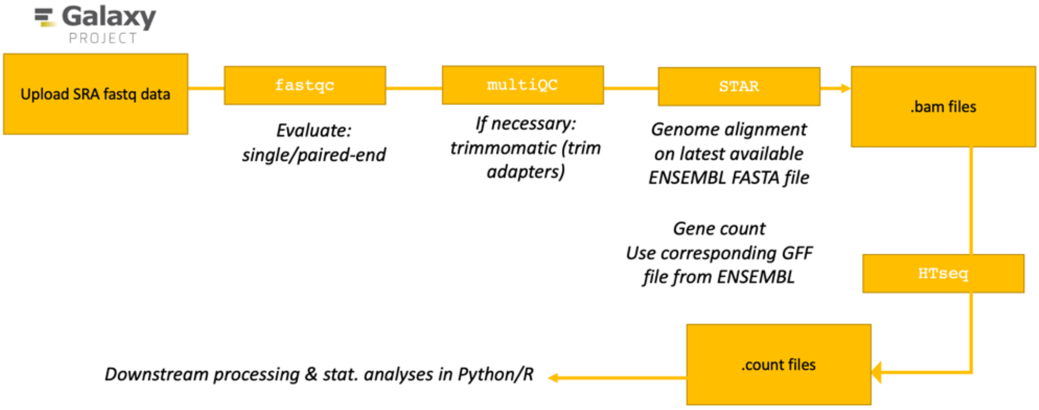


Figure 2-2: Schematic representation of the reproducible RNAseq data pre-processing pipeline implemented in the current study

The four main steps involved in the workflow process are data upload (`fastq`), quality assessment (`fastqc` and `multiqc`), alignment to the latest reference genome (`STAR`), and gene quantification (`HTseq`). The pipeline was implemented on Galaxy servers.

2.3.4. Human RNAseq samples require batch correction

Following the parallel processing of the 12 human RNAseq samples, individual samples cluster by study: SRP053043 clusters separately from the other two studies alongside the second PCA component, as does SRP126773. The first two principal components account for 82% of the total variance within the samples (***Figure*** 2-3).

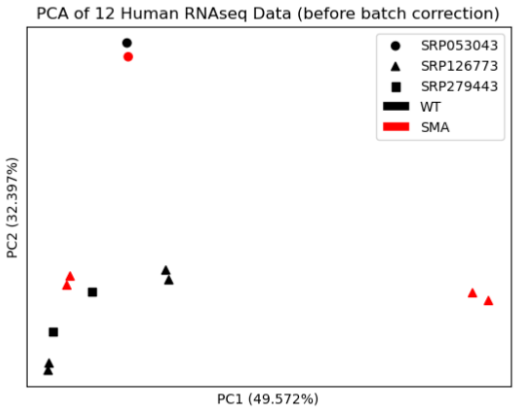


Figure 2-3: Principal component analysis (PCA) plot illustrating the relatedness of 12 human WT and SMA RNAseq neuronal datasets

PCA reduces higher-dimensional objects (here, RNA sequencing data) to principal components that explain most of the variance. The first two principal components are shown here, illustrating that prior to batch correction, samples segregate by study rather than by genotype (WT samples in black, SMA samples in red).

I then investigated the level of correlation between the seven WT samples taken from the three different studies by visually inspecting them against each other. I set a temporary arbitrary threshold of 90% to qualify WT samples as similar enough for direct downstream processing for differential expression analysis (***Figure*** 2-4).

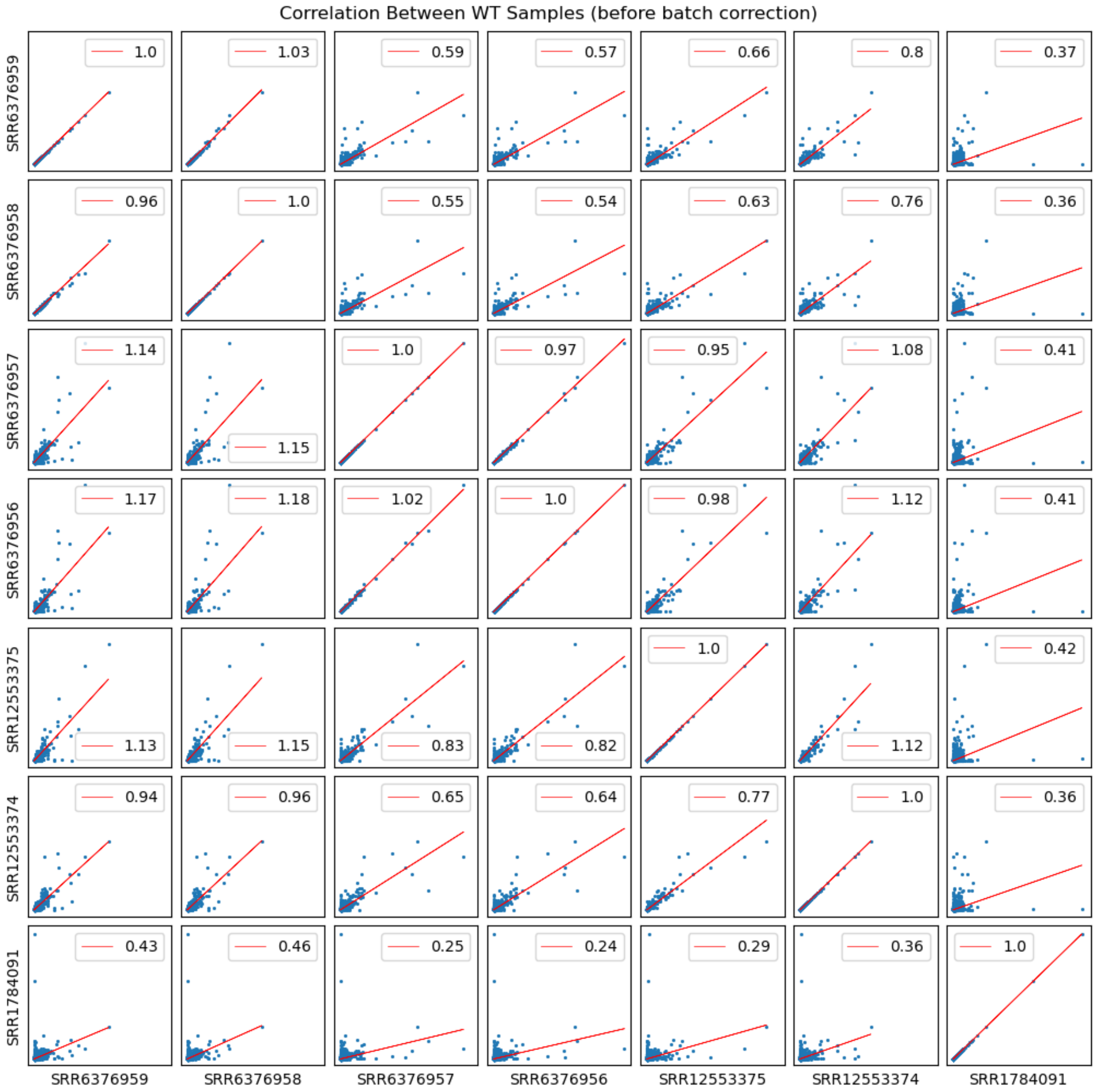


Figure 2-4: Genomic data correlation between the 7 WT samples before batch-effect correction

Scatter plots allow for visual inspection of the similarity (correlation) between variables. Here, the gene expression of a given WT dataset is plotted against the gene expression of all the other WT datasets, and a simple linear best fit model (y=ax+b) is applied to determine to what degree every pair of datasets is correlated: the closer to 1 the slope coefficient a (in red) is, the more similar the two datasets are. Technical replicates within the same study (e.g. SRP6376958 and SRP6376959) are very similar (a=1.03), whilst WT samples taken from different studies (e.g. SRP6376957 and SRP1784091) are much more dissimilar (a=0.25).

I confirm the PCA results showing that SRR1784091 (from study SRP053043) is very poorly correlated to any of the other WT samples, with linear best fit slopes between 0.24-0.46. Similarly, samples from the SRP279443 study do not correlate with the SRP126773 study's WT samples (linear best fit slopes: 0.64 and 0.65 with Control MN1 replicates 1 and 2). This suggests the existence of a batch-effect, that is, of variance in the data due to the unwanted *technical* variation between individual experiments (variation between handlers, experimental location, and/or reagents, etc.), rather than variance due to biological differences. As such, I decided to apply batch correction to all twelve samples.

2.3.5. Batch correction reveals non-linear relationship between samples

Methodologies to adjust for batch effect (hereafter referred as 'batch correction') are numerous and vary from reducing batch effect parameter estimates toward the overall mean of the batch effect estimates (e.g. [298]) to including batch variables as covariate in the linear regression models used in latter differential expression analysis (e.g. [287]).

I corrected for batch effect using the CombatSeq package, which uses negative binomial regression models (rather than Gaussian models) to model RNAseq count data [286]. I corrected batch effect by study, grouping by selected biological covariant (SMA or WT). PCA of the batch corrected data shows a reduction of the individual clustering of the SRP053043 samples (***Figure*** 2-5***,***

Supplementary Figure 7-1).

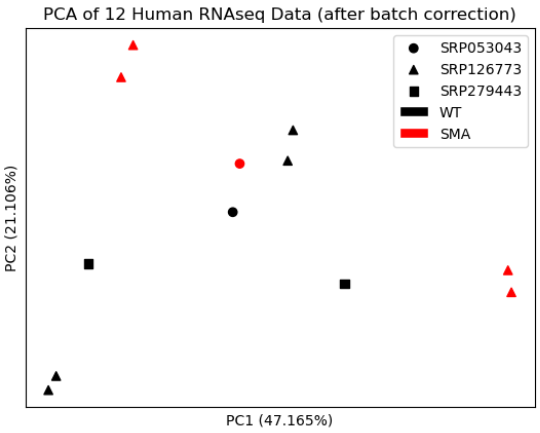


Figure 2-5: Principal component analysis (PCA) plot illustrating the relatedness of 12 human WT and SMA RNAseq neuronal datasets, following batch correction

The first two principal components of the dimensionally reduced, batch-corrected datasets still explain most of the variance of the twelve datasets, albeit to a mildly lesser degree than before batch-correction. Importantly, following *batch correction, samples do not cluster by study anymore, and some genotype clustering is emerging (WT in black, SMA in red).*

We then conducted a t-SNE (t-distributed stochastic neighbor embedding) analysis to assess whether the batch correction left enough differentiation between groups for the data to be worth pursuing for DE analyses, and to investigate any patterns, including clusters. PCA only investigates linear components of the relationships between samples, whilst t-SNE is a non-linear dimensionality reduction method which can capture non-linear relationships within the group of samples. T-SNE shows a consistent pattern separating WT and SMA samples in two separate clusters (particularly visible on perplexities 4 and 11, ***Figure*** 2-6). The grouped clustering of WT samples together within the 'horseshoe' pattern of SMA samples is confirmed by the correlation plots: the formerly very poorly correlated SRR1784091 sample (from study SRP053043) presents linear best fit slopes between 0.73-1.1 (vs. 0.24-0.46 before batch correction) (***Figure*** 2-7).

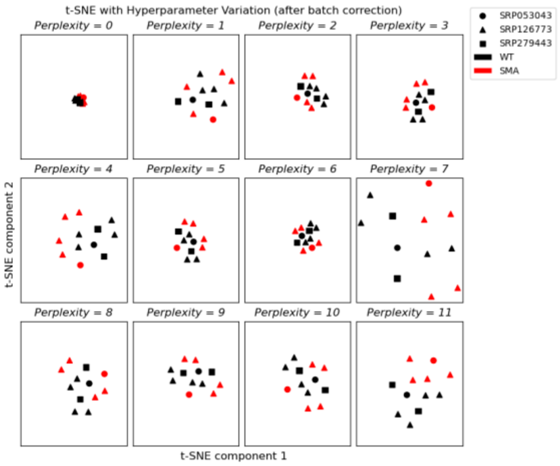


Figure 2-6: T-SNE plot of the clustering of batch-corrected samples

T-SNE visualisations from the twelve neuronal datasets, using different perplexity values. Perplexity is a tuneable parameter allowing to balance the focus on local vs. global aspects of the dimensionality reduction. SMA samples (in red) distinctly cluster from WT samples (in black) in a horseshoe pattern (esp. in perplexities 4, 7, and 11)

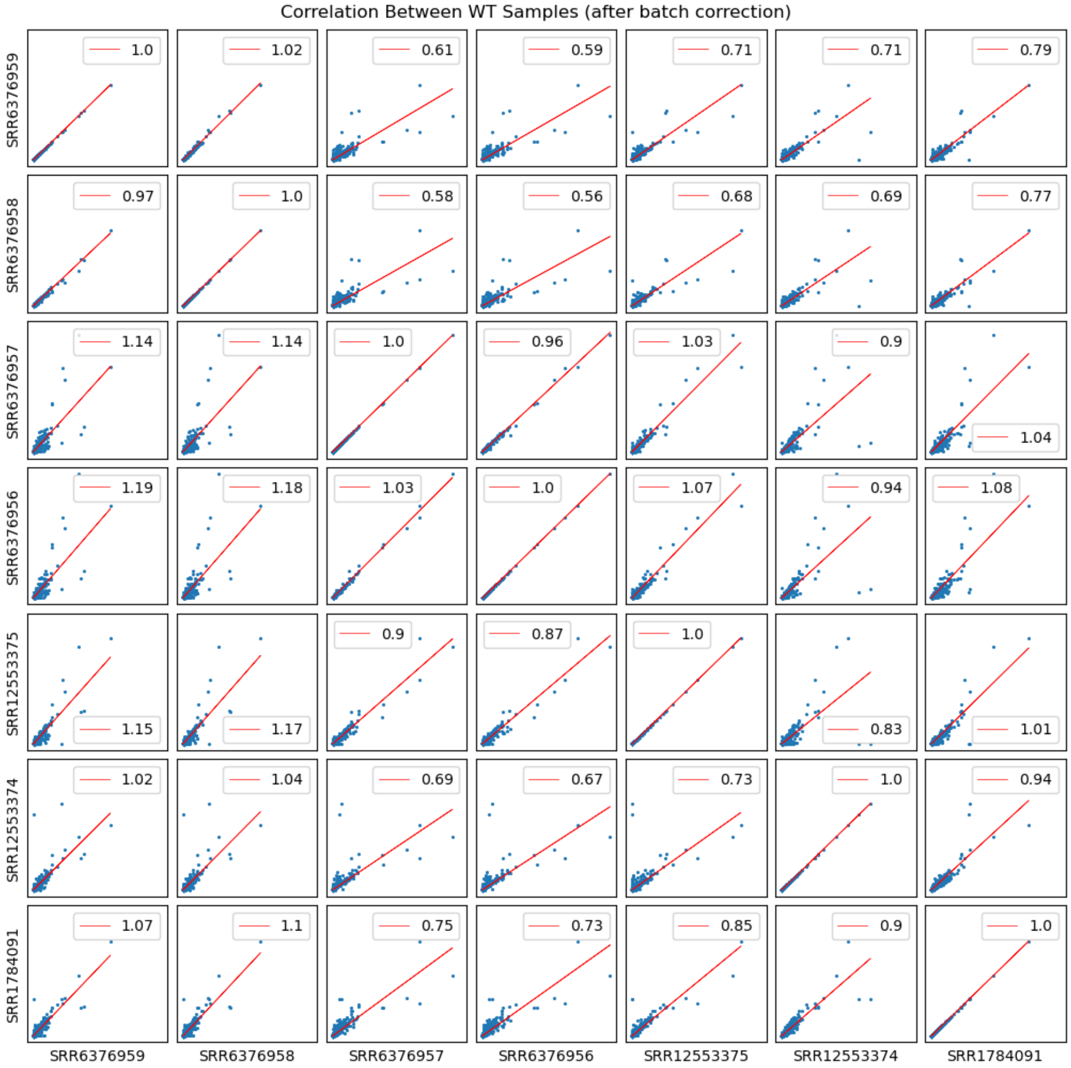


Figure 2-7: Genomic data correlation of the 7 WT samples after batch-effect correction

Visual inspection of the similarity (correlation) between studies in the correlation plots and the fitting of a simple linear best fit model (y=ax+b) shows that technical replicates within the same study (e.g. SRP6376958 and SRP6376959) are still very similar (a=1.02) to each other, and that similarity has increased after batch-effect correction between WT samples taken from different studies (e.g. SRP6376957 and SRP1784091) are much more similar (a=0.75) than before batch-effect correction.

2.3.6. Differential gene expression from meta-analysis

I then sought to reanalyse the aggregated SMA vs. WT groups for differential gene expression with an exact test. From the 16,034 genes left after the non-inclusion of genes of low expression, I find 437 genes significantly down-regulated and 494 genes significantly up-regulated between WT and SMA (|logFC|≥1 and fdr.pval.corr ≤5%), of expressions varying between CPM=0.57 and to CPM=5120. (***Figure*** 2-8).

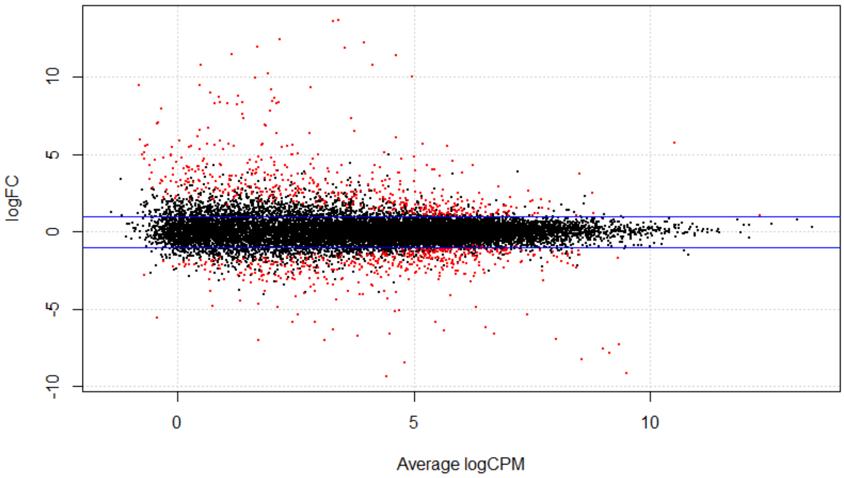


Figure 2-8: Significantly dysregulated in the meta-analysis between WT and SMA neuronal samples

MA plot showing the relationship between average gene expression (log CPM) and fold change (logFC) across all DE genes between WT and SMA neuronal samples. Each gene is represented by a dot. DE genes (fdr.pval.corr ≤5%), are colored in red. The blue line shows the |logFC|≥1 threshold (logFC>1 indicates upregulation: 494 genes; logFC<1, downregulation: 437 genes).

I validate the approach by identifying a number of well-known dysregulated genes in SMA amongst our hits, including *SMN1* (logFC=8.409, adj.p.val= 4.46E-22) and other transcripts previously described in SMA models, e.g. *PLA2G7* [145] (logFC=5.540, adj.p.val=2.46E-4).

Interestingly, there are several novel, not annotated transcripts amongst the top 50 upregulated genes between WT and SMA, including ENSG00000260834, ENSG00000254277, ENSG00000284294, and ENSG00000275216 (***Table*** 2-6). The meta-analysis also reveals a likely sex imbalance between the WT and SMA samples, with most of the top 50 down-regulated genes on the Y chromosome (***Table*** 2-7).

Table 2-6: Top 50 most differentially expressed upregulated genes

Gene.stable.ID	logFC	logCPM	PValue	FDR.pval.corr.	Protein.stable.ID	Gene.name	Transcript.name	Gene.description
ENSG00000204121	13.70293728	3.40161861	9.24E-09	2.28E-06		ECEL1P1	ECEL1P1-201	endothelin converting enzyme like 1 pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:14017]
ENSG00000249152	13.58584888	3.284758556	6.20E-12	5.85E-09		LNCPRESS2	LNCPRESS2-201	lncRNA p53 regulated and ESC associated 2 [Source:HGNC Symbol;Acc:HGNC:52747]
ENSG00000280511	12.46110591	2.161678089	3.49E-09	1.00E-06				novel transcript
ENSG00000280623	12.45794828	2.162181403	1.66E-20	1.33E-16		PCAT14	PCAT14-203	prostate cancer associated transcript 14 [Source:HGNC Symbol;Acc:HGNC:48977]
ENSG00000280623	12.45794828	2.162181403	1.66E-20	1.33E-16		PCAT14	PCAT14-202	prostate cancer associated transcript 14 [Source:HGNC Symbol;Acc:HGNC:48977]
ENSG00000280623	12.45794828	2.162181403	1.66E-20	1.33E-16		PCAT14	PCAT14-201	prostate cancer associated transcript 14 [Source:HGNC Symbol;Acc:HGNC:48977]
ENSG00000260834	12.23700682	3.939620667	4.84E-14	9.71E-11				novel transcript
ENSG00000260834	12.23700682	3.939620667	4.84E-14	9.71E-11				novel transcript
ENSG00000260834	12.23700682	3.939620667	4.84E-14	9.71E-11				novel transcript
ENSG00000234787	11.97618574	1.685268493	1.36E-13	2.18E-10		LINC00458	LINC00458-202	long intergenic non-protein coding RNA 458 [Source:HGNC Symbol;Acc:HGNC:42807]
ENSG00000234787	11.97618574	1.685268493	1.36E-13	2.18E-10		LINC00458	LINC00458-203	long intergenic non-protein coding RNA 458 [Source:HGNC Symbol;Acc:HGNC:42807]
ENSG00000234787	11.97618574	1.685268493	1.36E-13	2.18E-10		LINC00458	LINC00458-201	long intergenic non-protein coding RNA 458 [Source:HGNC Symbol;Acc:HGNC:42807]
ENSG00000234787	11.97618574	1.685268493	1.36E-13	2.18E-10		LINC00458	LINC00458-204	long intergenic non-protein coding RNA 458 [Source:HGNC Symbol;Acc:HGNC:42807]
ENSG00000234787	11.97618574	1.685268493	1.36E-13	2.18E-10		LINC00458	LINC00458-206	long intergenic non-protein coding RNA 458 [Source:HGNC Symbol;Acc:HGNC:42807]
ENSG00000234787	11.97618574	1.685268493	1.36E-13	2.18E-10		LINC00458	LINC00458-207	long intergenic non-protein coding RNA 458 [Source:HGNC Symbol;Acc:HGNC:42807]
ENSG00000234787	11.97618574	1.685268493	1.36E-13	2.18E-10		LINC00458	LINC00458-205	long intergenic non-protein coding RNA 458 [Source:HGNC Symbol;Acc:HGNC:42807]
ENSG00000133980	11.84846912	3.552114363	7.14E-13	8.81E-10	ENSP00000256362	VRTN	VRTN-201	vertebrae development associated [Source:HGNC Symbol;Acc:HGNC:20223]
ENSG00000133980	11.84846912	3.552114363	7.14E-13	8.81E-10	ENSP00000452158	VRTN	VRTN-202	vertebrae development associated [Source:HGNC Symbol;Acc:HGNC:20223]
ENSG00000244280	11.45419314	1.158990131	7.98E-07	9.76E-05		ECEL1P2	ECEL1P2-202	endothelin converting enzyme like 1 pseudogene 2 [Source:HGNC Symbol;Acc:HGNC:14019]
ENSG00000244280	11.45419314	1.158990131	7.98E-07	9.76E-05		ECEL1P2	ECEL1P2-201	endothelin converting enzyme like 1 pseudogene 2 [Source:HGNC Symbol;Acc:HGNC:14019]
ENSG00000241186	11.38686029	4.617087422	7.79E-13	8.93E-10	ENSP00000296145	TDGF1	TDGF1-201	teratocarcinoma-derived growth factor 1 [Source:HGNC Symbol;Acc:HGNC:11701]
ENSG00000241186	11.38686029	4.617087422	7.79E-13	8.93E-10	ENSP00000446375	TDGF1	TDGF1-205	teratocarcinoma-derived growth factor 1 [Source:HGNC Symbol;Acc:HGNC:11701]
ENSG00000241186	11.38686029	4.617087422	7.79E-13	8.93E-10		TDGF1	TDGF1-204	teratocarcinoma-derived growth factor 1 [Source:HGNC Symbol;Acc:HGNC:11701]
ENSG00000241186	11.38686029	4.617087422	7.79E-13	8.93E-10		TDGF1	TDGF1-202	teratocarcinoma-derived growth factor 1 [Source:HGNC Symbol;Acc:HGNC:11701]
ENSG00000241186	11.38686029	4.617087422	7.79E-13	8.93E-10		TDGF1	TDGF1-203	teratocarcinoma-derived growth factor 1 [Source:HGNC Symbol;Acc:HGNC:11701]
ENSG00000280142	10.79773216	0.502979661	7.29E-08	1.41E-05				TEC
ENSG00000254277	10.79508833	4.14026314	1.91E-10	9.18E-08				novel transcript
ENSG00000254277	10.79508833	4.14026314	1.91E-10	9.18E-08				novel transcript
ENSG00000254277	10.79508833	4.14026314	1.91E-10	9.18E-08				novel transcript
ENSG00000254277	10.79508833	4.14026314	1.91E-10	9.18E-08				novel transcript
ENSG00000162624	10.22729798	1.918745973	2.78E-06	0.000272225	ENSP00000512714	LHX8	LHX8-204	LIM homeobox 8 [Source:HGNC Symbol;Acc:HGNC:28838]
ENSG00000162624	10.22729798	1.918745973	2.78E-06	0.000272225	ENSP00000294638	LHX8	LHX8-201	LIM homeobox 8 [Source:HGNC Symbol;Acc:HGNC:28838]
ENSG00000162624	10.22729798	1.918745973	2.78E-06	0.000272225		LHX8	LHX8-203	LIM homeobox 8 [Source:HGNC Symbol;Acc:HGNC:28838]
ENSG00000162624	10.22729798	1.918745973	2.78E-06	0.000272225	ENSP00000348597	LHX8	LHX8-202	LIM homeobox 8 [Source:HGNC Symbol;Acc:HGNC:28838]
ENSG00000078596	10.02675259	4.962710459	3.09E-12	3.10E-09		ITM2A	ITM2A-206	integral membrane protein 2A [Source:HGNC Symbol;Acc:HGNC:6173]
ENSG00000078596	10.02675259	4.962710459	3.09E-12	3.10E-09		ITM2A	ITM2A-204	integral membrane protein 2A [Source:HGNC Symbol;Acc:HGNC:6173]
ENSG00000078596	10.02675259	4.962710459	3.09E-12	3.10E-09	ENSP00000415533	ITM2A	ITM2A-202	integral membrane protein 2A [Source:HGNC Symbol;Acc:HGNC:6173]
ENSG00000078596	10.02675259	4.962710459	3.09E-12	3.10E-09		ITM2A	ITM2A-205	integral membrane protein 2A [Source:HGNC Symbol;Acc:HGNC:6173]
ENSG00000078596	10.02675259	4.962710459	3.09E-12	3.10E-09	ENSP00000362395	ITM2A	ITM2A-201	integral membrane protein 2A [Source:HGNC Symbol;Acc:HGNC:6173]
ENSG00000078596	10.02675259	4.962710459	3.09E-12	3.10E-09		ITM2A	ITM2A-203	integral membrane protein 2A [Source:HGNC Symbol;Acc:HGNC:6173]
ENSG00000248131	9.964880582	1.656230208	3.59E-07	5.23E-05		LINC01194	LINC01194-203	long intergenic non-protein coding RNA 1194 [Source:HGNC Symbol;Acc:HGNC:37171]
ENSG00000248131	9.964880582	1.656230208	3.59E-07	5.23E-05		LINC01194	LINC01194-202	long intergenic non-protein coding RNA 1194 [Source:HGNC Symbol;Acc:HGNC:37171]
ENSG00000248131	9.964880582	1.656230208	3.59E-07	5.23E-05		LINC01194	LINC01194-201	long intergenic non-protein coding RNA 1194 [Source:HGNC Symbol;Acc:HGNC:37171]
ENSG00000248131	9.964880582	1.656230208	3.59E-07	5.23E-05		LINC01194	LINC01194-205	long intergenic non-protein coding RNA 1194 [Source:HGNC Symbol;Acc:HGNC:37171]
ENSG00000248131	9.964880582	1.656230208	3.59E-07	5.23E-05		LINC01194	LINC01194-204	long intergenic non-protein coding RNA 1194 [Source:HGNC Symbol;Acc:HGNC:37171]
ENSG00000284294	9.441655476	0.484684329	1.81E-09	5.91E-07				novel transcript
ENSG00000284294	9.441655476	0.484684329	1.81E-09	5.91E-07				novel transcript
ENSG00000284294	9.441655476	0.484684329	1.81E-09	5.91E-07				novel transcript
ENSG00000229150	9.436441184	-0.814772659	7.37E-07	9.16E-05		CRYGEP	CRYGEP-202	crystallin gamma E, pseudogene [Source:HGNC Symbol;Acc:HGNC:2412]
ENSG00000229150	9.436441184	-0.814772659	7.37E-07	9.16E-05		CRYGEP	CRYGEP-201	crystallin gamma E, pseudogene [Source:HGNC Symbol;Acc:HGNC:2412]

Table 2-7: Top 50 most differentially expressed downregulated genes

Gene.stable.ID	logFC	logCPM	PValue	FDR.pval.corr.	Protein.stable.ID	Gene.name	Transcript.name	Gene.description
ENSG00000219438	-9.374934652	4.419854077	1.50E-11	1.10E-08	ENSP00000385603	TAFA5	TAFA5-204	TAFA chemokine like family member 5 [Source:HGNC Symbol;Acc:HGNC:21592]
ENSG00000219438	-9.374934652	4.419854077	1.50E-11	1.10E-08		TAFA5	TAFA5-205	TAFA chemokine like family member 5 [Source:HGNC Symbol;Acc:HGNC:21592]
ENSG00000219438	-9.374934652	4.419854077	1.50E-11	1.10E-08	ENSP00000351043	TAFA5	TAFA5-202	TAFA chemokine like family member 5 [Source:HGNC Symbol;Acc:HGNC:21592]
ENSG00000219438	-9.374934652	4.419854077	1.50E-11	1.10E-08	ENSP00000383933	TAFA5	TAFA5-203	TAFA chemokine like family member 5 [Source:HGNC Symbol;Acc:HGNC:21592]
ENSG00000219438	-9.374934652	4.419854077	1.50E-11	1.10E-08	ENSP00000336812	TAFA5	TAFA5-201	TAFA chemokine like family member 5 [Source:HGNC Symbol;Acc:HGNC:21592]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-215	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-212	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-219	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-214	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-218	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-211	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-208	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-210	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-213	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-205	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-217	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-209	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-201	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-206	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-207	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-216	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-204	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-203	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000131002	-9.18798809	9.506946123	1.31E-05	0.000921974		TXLNGY	TXLNGY-202	taxilin gamma pseudogene, Y-linked [Source:HGNC Symbol;Acc:HGNC:18473]
ENSG00000234444	-8.498219126	4.8057859	8.32E-17	3.33E-13		ZNF736	ZNF736-205	zinc finger protein 736 [Source:HGNC Symbol;Acc:HGNC:32467]
ENSG00000234444	-8.498219126	4.8057859	8.32E-17	3.33E-13	ENSP00000347210	ZNF736	ZNF736-201	zinc finger protein 736 [Source:HGNC Symbol;Acc:HGNC:32467]
ENSG00000234444	-8.498219126	4.8057859	8.32E-17	3.33E-13		ZNF736	ZNF736-207	zinc finger protein 736 [Source:HGNC Symbol;Acc:HGNC:32467]
ENSG00000234444	-8.498219126	4.8057859	8.32E-17	3.33E-13	ENSP00000400852	ZNF736	ZNF736-202	zinc finger protein 736 [Source:HGNC Symbol;Acc:HGNC:32467]
ENSG00000234444	-8.498219126	4.8057859	8.32E-17	3.33E-13	ENSP00000391083	ZNF736	ZNF736-203	zinc finger protein 736 [Source:HGNC Symbol;Acc:HGNC:32467]
ENSG00000234444	-8.498219126	4.8057859	8.32E-17	3.33E-13		ZNF736	ZNF736-204	zinc finger protein 736 [Source:HGNC Symbol;Acc:HGNC:32467]
ENSG00000234444	-8.498219126	4.8057859	8.32E-17	3.33E-13		ZNF736	ZNF736-208	zinc finger protein 736 [Source:HGNC Symbol;Acc:HGNC:32467]
ENSG00000234444	-8.498219126	4.8057859	8.32E-17	3.33E-13		ZNF736	ZNF736-206	zinc finger protein 736 [Source:HGNC Symbol;Acc:HGNC:32467]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05		UTY	UTY-222	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000484698	UTY	UTY-215	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000485539	UTY	UTY-216	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000507919	UTY	UTY-221	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000372352	UTY	UTY-205	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000439922	UTY	UTY-209	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000506749	UTY	UTY-207	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000506900	UTY	UTY-219	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000355420	UTY	UTY-203	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000485013	UTY	UTY-213	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05		UTY	UTY-208	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000442047	UTY	UTY-212	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000328939	UTY	UTY-202	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000508171	UTY	UTY-217	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000483735	UTY	UTY-214	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000441943	UTY	UTY-211	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000445274	UTY	UTY-210	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]
ENSG00000183878	-8.239071992	8.557586111	1.27E-07	2.31E-05	ENSP00000507736	UTY	UTY-220	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked [Source:HGNC Symbol;Acc:HGNC:12638]

Before continuing and identifying the biological function of the identified DE genes, it was necessary to consider whether the meta-analysis approach yielded significant results when compared with the individual studies of which it is made.

2.3.7. Concurrence of the meta-analysis with studies already published

I proceeded to assess the value of the meta-analysis approach regarding biological variation amongst samples and regarding numbers of genes identified. I focused this methodological variation on Rizzo et al. (2019) (SRP126773), which is the only human study in our meta-analysis that contains enough (n ≥3) of both SMA (resp. WT) samples per group for a differential expression gene analysis following the exact same pipeline.

2.3.7.1. Biological coefficient of variation

The biological coefficient of variation (BCV) plots the variances across all genes, from lowest to highest expression (average logCPM across datasets). Normally, expressed genes of lower expression have larger variation compared to more highly expressed genes. For isogenic organisms in a laboratory, the BCV is expected around 10-15%. Interestingly, the aggregated human RNAseq datasets shows a lower biological coefficient of variation than Rizzo's study, with BCV_{meta} of 62.68% vs. BCV_{rizzo}=65.25% (***Figure*** 2-9).

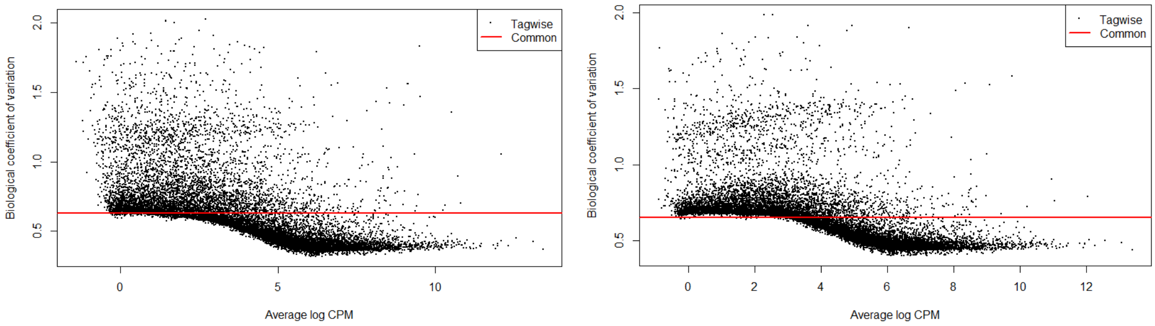


Figure 2-9: Batch-effect correction does not significantly increase the biological coefficient of variation

Biological coefficient of variation (BCV) of study (left: meta-analysis; right: Rizzo et al. 2019) plotted against the average abundance of each gene (log CPM), as estimated by edgeR. Coefficients are shown corresponding to the estimates for the common ('mean' dispersion across all genes – red line) and tagwise (empirical Bayes gene-specific dispersions – black dots) negative binomial dispersions.

2.3.7.2. Exclusion of genes with low counts

In our meta-analysis pipeline, I defined genes with sufficient count as genes with CPM>1 in at least 'number of SMA samples -1', i.e. in 4 samples for the meta-analysis (n_{SMA}=5). For a similar pipeline in the Rizzo study, I included only genes with CPM>1 in 3 samples.

This left 16,034 genes for down-stream statistical analysis in the meta-analysis, vs. 15,645 genes in the Rizzo et al. study – which represent the inclusion of 2.5% more genes in the meta-analysis.

2.3.7.3. Comparability of lists of differentially expressed genes

In this meta-analysis, I added four samples to the existing study by Rizzo et al. – one SMA sample and four WT samples. Using similar threshold viz. decision-making pipelines, the meta-analysis yielded 931 differentially expressed transcripts in 899 unique genes, vs. 868 differentially expressed transcripts in 837 unique genes in the original Rizzo et al. study.

There is overall a significant overlap between the identified genes, with 370 transcripts upregulated in both Rizzo et al. and in our meta-analysis, and 218 transcripts downregulated in both (***Figure*** 2-10). However, the meta-analysis also identifies 124 new upregulated transcripts and 219 new down-regulated transcripts.

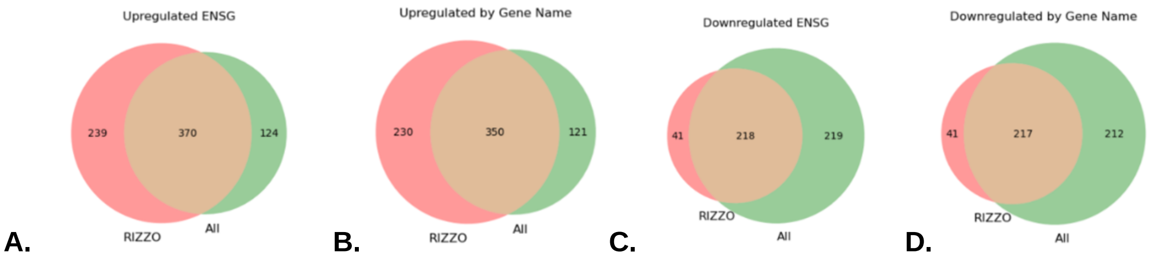


Figure 2-10: The meta-analysis of 12 human RNAseq samples identifies new DE transcripts and yields similar DE transcripts to Rizzo et. al.

Two-set Venn diagrams of the number of features at the intersection (in brown) of DE transcripts between Rizzo et al. (in red) and the present meta-analysis (in green). ***(A)*** *Venn diagram of upregulated DE transcripts.* ***(B)*** *Venn diagram of upregulated DE genes*. ***(C)*** *Venn diagram of downregulated DE transcripts.* ***(D)*** *Venn diagram of downregulated DE genes.*

I therefore proceeded with the list of DE transcripts from the whole meta-analysis for biological hypothesis generation.

2.3.8. Meta-analysis of SMA neuronal RNAseq studies yields new candidates for further biological exploration

Gene ontology (GO) analysis of the 931 DE transcripts identified in the meta-analysis study shows significant enrichment of calcium ion binding proteins (p.adj = 6.803x10E-5), sulfur binding proteins (3.984x10E-2), and laminin binding proteins (p.adj. = 4.129x10E-2). These molecular functions correlate with the cellular compartments particularly enriched within our dataset, in particular the plasma membrane (p.adj = 4.453x10-16), cell and cell-cell junctions (p.adj. respectively 8.508x10E-12 and 9.772x10-8), and the biological processes of cell adhesion (p.adj = 2.114x10E-18) and of regulation of cell communication (p.adj=8.614x10E-19). Interestingly, angiogenesis, vasculature- and blood-vessel development genes are particularly enriched (p.adj. between 7.109x10E-9 and 9.578x10E-13) (***Figure*** 2-11).

ID	Source	Term ID	Term Name	P_{adj} (query_1)
1	GO:BP	GO:0007155	cell adhesion	2.114×10^{-18}
2	GO:BP	GO:0010646	regulation of cell communication	8.614×10^{-19}
3	GO:BP	GO:0023051	regulation of signaling	3.549×10^{-18}
4	GO:CC	GO:0005886	plasma membrane	4.453×10^{-16}
5	GO:CC	GO:0030054	cell junction	8.508×10^{-12}
6	GO:MF	GO:0005509	calcium ion binding	6.803×10^{-5}
7	GO:MF	GO:0043236	laminin binding	4.129×10^{-2}
8	GO:MF	GO:0046872	metal ion binding	2.560×10^{-4}
9	GO:MF	GO:1901681	sulfur compound binding	3.984×10^{-2}
10	GO:BP	GO:0070887	cellular response to chemical stimulus	6.902×10^{-19}
11	GO:BP	GO:0072359	circulatory system development	1.685×10^{-14}
12	GO:BP	GO:0016477	cell migration	3.007×10^{-14}
13	GO:BP	GO:0098609	cell-cell adhesion	1.318×10^{-13}
14	GO:BP	GO:0001568	blood vessel development	9.578×10^{-13}
15	GO:BP	GO:0001944	vasculature development	3.016×10^{-12}
16	GO:BP	GO:0048514	blood vessel morphogenesis	1.699×10^{-11}
17	GO:BP	GO:0007166	cell surface receptor signaling pathway	4.132×10^{-10}
18	GO:BP	GO:0001525	angiogenesis	7.190×10^{-9}
19	GO:BP	GO:0030334	regulation of cell migration	1.092×10^{-8}
20	GO:MF	GO:0030020	extracellular matrix structural constituent conf...	1.585×10^{-2}
21	GO:CC	GO:0005911	cell-cell junction	9.772×10^{-8}
22	GO:CC	GO:0030424	axon	4.681×10^{-4}
23	GO:CC	GO:0045202	synapse	7.154×10^{-4}
24	GO:CC	GO:0015629	actin cytoskeleton	1.918×10^{-3}
25	GO:CC	GO:0098794	postsynapse	7.747×10^{-3}
26	GO:CC	GO:0030425	dendrite	1.704×10^{-2}
27	KEGG	KEGG:04020	Calcium signaling pathway	4.477×10^{-3}
28	KEGG	KEGG:04514	Cell adhesion molecules	2.035×10^{-2}
29	KEGG	KEGG:04974	Protein digestion and absorption	3.157×10^{-2}
30	TF	TF:M00649_1	Factor: MAZ; motif: GGGGAGGG; match class: 1	1.423×10^{-10}
31	TF	TF:M01104_1	Factor: MOVO-B; motif: GNGGGGG; match cla...	9.999×10^{-10}
32	TF	TF:M01104	Factor: MOVO-B; motif: GNGGGGG	9.817×10^{-9}

GO:MF		stats	
Term name	Term ID	P_{adj}	$-\log_{10}(P_{adj})$
calcium ion binding	GO:0005509	6.803×10^{-5}	
metal ion binding	GO:0046872	2.560×10^{-4}	
extracellular matrix structural constituent conferring te...	GO:0030020	1.585×10^{-2}	
sulfur compound binding	GO:1901681	3.984×10^{-2}	
laminin binding	GO:0043236	4.129×10^{-2}	

1 to 5

GO:BP		stats	
Term name	Term ID	P_{adj}	$-\log_{10}(P_{adj})$
cellular response to chemical stimulus	GO:0070887	6.902×10^{-19}	
regulation of cell communication	GO:0010646	8.614×10^{-19}	
cell adhesion	GO:0007155	2.114×10^{-18}	
regulation of signaling	GO:0023051	3.549×10^{-18}	
circulatory system development	GO:0072359	1.685×10^{-14}	
cell migration	GO:0016477	3.007×10^{-14}	
cell-cell adhesion	GO:0098609	1.318×10^{-13}	
blood vessel development	GO:0001568	9.578×10^{-13}	
vasculature development	GO:0001944	3.016×10^{-12}	
blood vessel morphogenesis	GO:0048514	1.699×10^{-11}	
cell surface receptor signaling pathway	GO:0007166	4.132×10^{-10}	
angiogenesis	GO:0001525	7.190×10^{-9}	
regulation of cell migration	GO:0030334	1.092×10^{-8}	

1 to 13

GO:CC		stats	
Term name	Term ID	P_{adj}	$-\log_{10}(P_{adj})$
plasma membrane	GO:0005886	4.453×10^{-16}	
cell junction	GO:0030054	8.508×10^{-12}	
cell-cell junction	GO:0005911	9.772×10^{-8}	
axon	GO:0030424	4.681×10^{-4}	
synapse	GO:0045202	7.154×10^{-4}	
actin cytoskeleton	GO:0015629	1.918×10^{-3}	
postsynapse	GO:0098794	7.747×10^{-3}	
dendrite	GO:0030425	1.704×10^{-2}	

1 to 8

KEGG		stats	
Term name	Term ID	P_{adj}	$-\log_{10}(P_{adj})$
Calcium signaling pathway	KEGG:04020	4.477×10^{-3}	
Cell adhesion molecules	KEGG:04514	2.035×10^{-2}	
Protein digestion and absorption	KEGG:04974	3.157×10^{-2}	

1 to 3

TF		stats	
Term name	Term ID	P_{adj}	$-\log_{10}(P_{adj})$
Factor: MAZ; motif: GGGGAGGG; match class: 1	TF:M00649_1	1.423×10^{-10}	
Factor: MOVO-B; motif: GNGGGGG; match class: 1	TF:M01104_1	9.999×10^{-10}	
Factor: MOVO-B; motif: GNGGGGG	TF:M01104	9.817×10^{-9}	

1 to 3

Figure 2-11: Gene ontology and pathway analysis in the DE transcripts from the meta-analysis

Manhattan plot of the enrichment of gene ontologies (GO) amongst the 931 DE transcripts. On the x-axis, GO are grouped and coloured according to: molecular function (MF), biological process (BP), cellular compartment (CC), KEGG pathway, and transcription factor (TF). On the y-axis, adjusted pval (in log10 scale). Selected top 32 GO terms are presented.

I further manually refined a list of candidates based on evidence level, and selected only genes for which the gene ontology is inferred either from experiments, direct assays, mutant phenotypes, and/or a genetic or a physical interaction. For further down-stream *in vitro* validation and future functional validation, I limited this list to a subset of GO terms of particular relevance in neurobiology: 'calcium ion binding', 'laminin binding', 'axon', 'dendrite' (***Table*** 2-8) and 'plasma membrane' genes (***Supplementary Table*** 7-2).

Table 2-8: Subset of high-level evidence DE genes for further exploration

GO term	**gene description**	**gene ID**
Calcium ion binding GO:0005509	calbindin 1 [Source:HGNC Symbol;Acc:HGNC:1434]	ENSG00000104327
	lipoprotein lipase [Source:HGNC Symbol;Acc:HGNC:6677]	ENSG00000175445
	aspartate beta-hydroxylase [Source:HGNC Symbol;Acc:HGNC:757]	ENSG00000198363
	SPARC (osteonectin), cwcv and kazal like domains proteoglycan 1 [Source:HGNC Symbol;Acc:HGNC:11251]	ENSG00000152377
	iduronate 2-sulfatase [Source:HGNC Symbol;Acc:HGNC:5389]	ENSG00000010404
	solute carrier family 25 member 13 [Source:HGNC Symbol;Acc:HGNC:10983]	ENSG00000004864
	reticulocalbin 3 [Source:HGNC Symbol;Acc:HGNC:21145]	ENSG00000142552
	annexin A4 [Source:HGNC Symbol;Acc:HGNC:542]	ENSG00000196975
	alkaline phosphatase, biomineralization associated [Source:HGNC Symbol;Acc:HGNC:438]	ENSG00000162551
	polypeptide N-acetylgalactosaminyltransferase 3 [Source:HGNC Symbol;Acc:HGNC:4125]	ENSG00000115339
	RUNX family transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:10471]	ENSG00000159216
	thrombospondin 4 [Source:HGNC Symbol;Acc:HGNC:11788]	ENSG00000113296
	dysferlin [Source:HGNC Symbol;Acc:HGNC:3097]	ENSG00000135636
	multiple C2 and transmembrane domain containing 2 [Source:HGNC Symbol;Acc:HGNC:25636]	ENSG00000140563
	histidine rich calcium binding protein [Source:HGNC Symbol;Acc:HGNC:5178]	ENSG00000130528
	troponin C1, slow skeletal and cardiac type [Source:HGNC Symbol;Acc:HGNC:11943]	ENSG00000114854
	sonic hedgehog signaling molecule [Source:HGNC Symbol;Acc:HGNC:10848]	ENSG00000164690
	glutamate metabotropic receptor 7 [Source:HGNC Symbol;Acc:HGNC:4599]	ENSG00000196277
laminin binding GO:0043236	galectin 3 [Source:HGNC Symbol;Acc:HGNC:6563]	ENSG00000131981
	extracellular matrix protein 1 [Source:HGNC Symbol;Acc:HGNC:3153]	ENSG00000143369
	basal cell adhesion molecule (Lutheran blood group) [Source:HGNC Symbol;Acc:HGNC:6722]	ENSG00000187244
	leucine rich repeat containing 15 [Source:HGNC Symbol;Acc:HGNC:20818]	ENSG00000172061
axon GO:0030424	calbindin 1 [Source:HGNC Symbol;Acc:HGNC:1434]	ENSG00000104327
	membrane metalloendopeptidase [Source:HGNC Symbol;Acc:HGNC:7154]	ENSG00000196549
	claudin 11 [Source:HGNC Symbol;Acc:HGNC:8514]	ENSG00000013297
dendrite GO:0030425	prion protein [Source:HGNC Symbol;Acc:HGNC:9449]	ENSG00000171867
	calbindin 1 [Source:HGNC Symbol;Acc:HGNC:1434]	ENSG00000104327
	EPH receptor A5 [Source:HGNC Symbol;Acc:HGNC:3389]	ENSG00000145242
	transmembrane protein 266 [Source:HGNC Symbol;Acc:HGNC:26763]	ENSG00000169758
	calcium/calmodulin dependent protein kinase II alpha [Source:HGNC Symbol;Acc:HGNC:1460]	ENSG00000070808
	glutamate metabotropic receptor 7 [Source:HGNC Symbol;Acc:HGNC:4599]	ENSG00000196277

I conclude this bioinformatic analysis by identifying 266 genes enriched for consensus sites of transcription factors (TF) targeting the MAZ motif (GGGGAGGG), and 335 genes enriched for consensus sites from TFs targeting the MOVO-B motif (GNGGGGG) (out of respectively 4,241 and 5,798 so annotated in the 19,958 genes with TF annotations) (***Supplementary Table*** 7-3***,*** ***Supplementary Table*** 7-4). Genes enriched for consensus binding sites for the MAZ TF span across numerous biological functions crucial to neuronal development, including ion channels (*KCNC4, SLC30A3, GRIA3, KCNQ3, SCN1A, SLC22A4, CACNG6*), adhesion molecules (*ICAM5, NCAM2, LAMB3, NECTIN2*), genes involved in the CNS development (*NGF, SLIT1, NGFR, ASTN1*) and proteins involved in the development and pruning of neuronal synapses (*SHANK1, C1R*). Similarly, genes enriched for consensus binding sites for the MOVO-B TF include extracellular matrix structural constituents conferring tensile strength, notably collagen (*COL23A1, COL24A1, COL11A2, COL18A1, COL9A3, COL7A1, COL6A3*), and neurogenesis and its regulation (*PRTG, BMP7, TTBK1, LEF1, SLIT1, EPHB2, MDK, EFNB3, HES1, RND2, SEMA4C, SMO, SMARCD3, ITPKA, VEGFA, VEGFC, ZNF365, CAMK2B, NGF*).

The exploration of the role of TF in the regulation of gene expression as another possible line for further biological exploration in SMA, though briefly discussed below, awaits results from the meta-analysis of murine neuronal RNAseq samples.

2.3.9. Future work on RNAseq SMA and WT neuronal datasets

At this stage, I have provided a proof-of-concept meta-analysis from human RNAseq SMA and WT samples. Future bioinformatic work will replicate the same analysis pipeline on the 14 murine RNA-seq samples I have previously identified (7 WT samples, 7 SMA samples). After DE analysis, I will continue this work by extracting orthologs from the murine RNAseq datasets and establish a list of mouse-human overlapping genes. Further bioinformatics studies include network analysis (to refine conserved candidates consistently dysregulated across disease model) and stratification by disease stage, as discussed below.

2.4. Discussion

In the discussion of this systematic review and meta-analysis, I will summarise the results from our pilot study of all human neuronal RNAseq datasets available in SMA; discuss the relevance of this methodology, its applicability to SMA; and its usefulness for further therapeutic candidate identification.

2.4.1. On the meta-analysis methodology

2.4.1.1. Advantages

There are four primary advantages to meta-analytic approaches of transcriptomic datasets.

First, it is a thorough approach. I scoped the meta-analysis following a literature review of datasets deposited in the world's two primary raw transcriptomic data repositories, NCBI's SRA and EBI's ArrayExpress. I included every publicly available SMA neuronal transcriptomic dataset as well as every WT neuronal transcriptomic dataset relevant to SMA studies. In particular, this allowed for the inclusion of WT samples that had never been compared to SMA samples, including the two WT samples from SRP279443 [292], whose original study focuses on the comparison of different WT cell types, as well as the two samples (one WT, one SMA) from SRP053043 [160] – whose study does not include more replicates and does not yield sufficient power for statistical analysis of differential gene expression.

Second, meta-analytic approach with parallelised re-analysis of raw data allows the inclusion of reference genome annotation updates. This has not been stressed in this pilot study, which rather occupies itself with human RNAseq samples published from 2015 to 2020. All the original studies will therefore have used the GRCh38.p13 human annotation, which was released in July 2014 (patched April 2022). However, the microarray samples that will be included in the follow-up analyses span a publication period from 2005-2017 (mouse microarrays) and 2012-2018 (human microarrays). During this time, murine genome annotation was built anew, in the form of GRCm39, whose gene build started August 2020 and was patched in April 2022. The first reported human microarray will also benefit from the latest human genome build. Using the latest genome annotations will allow the inclusion of newly annotated genes and the identification of new candidates from original publications. It will also ensure comparability between the studies for downstream analyses.

Third, the overall biological coefficient variation after the coalition of several studies does not increase the biological coefficient variation. Indeed, in our pilot study, the batch correction reduced the BCV of the meta-analytic study when compared to the BCV of the study from which most samples were taken.

Finally, the last step of the pipeline on RNA-seq analysis can be expanded using edgeR to include not only differential expression analyses, but alternative splicing analyses. Whilst this may not be applicable to include all the samples – notably depending on whether microarray platforms include exon-level analyses capabilities – such an analysis is easily expandable to both murine and human RNA-seq samples.

2.4.1.2. Limitations

Several limitations to our pilot study must be noted.

First, I must clarify a lexical ambiguity regarding the systematicity of the meta-analytic work. This study aimed at systematicity in the *data processing* from transcriptomic studies differing in their raw data processing, genome alignments, etc. – all done to be able to pool together samples and groups from studies separated in time, space, and protocols. Whilst the initial literature review was very much conducted with the goal to gather as many relevant studies as possible, the methodology used does not conform to all the standards of systematic *literature review* for secondary research – a field most systematised in clinical research [299]. Notably, (1) the screens of: title/abstract, whole data downloads, data extracted, and quality assessment, were conducted sequentially by two researchers, rather than independently by two-three of them; (2) the study was not pre-registered; and (3) the databases searched were limited to NCBI and EBI. Additionally, investigators were contacted to gather primary raw data were missing, but not parallel to the primary database search efforts.

Whilst this can be explained by the circumstances in which this work originated (see section 1.7), these limitations are worth acknowledging prior to considering the biological results presented. In particular, it is possible that further studies ought to be included, notably if unpublished, or if found on Google Scholar, Scopus, or Embase.

Second, whilst 'spinal muscular atrophy' was our first search term, it does not include WT samples from transcriptomic studies in other motor neuron conditions – including amyotrophic lateral sclerosis (ALS).

Third, the list of DE transcripts from our meta-analysis revealed that most of the top 50 down-regulated genes are linked to the Y-chromosome. Such disease-relevant sex-specific expression patterns have also been observed in other meta-analyses [266]. SMA and control fibroblasts from [292] are derived from females individuals. Yoshida et al. uses commercially available type 1 male patients from Coriell (Coriell ID GM03813) and the historical, Kyoto University 201B7 iPSC cell line derived from human dermal fibroblasts from a Caucasian female [254]. The sex of patients and of control individuals is unclear the largest study included [293] (itself referring to extraction and differentiation protocols from previous work from the same laboratory [290,300,301]). *Post-hoc* analysis of the expression (by CPM) of the X- and Y-borned genes *DDX3X* and *DDX3Y* reveals that all control MNs stem from female individuals, whilst there is one male and one female SMA MN sample. As such, all WT samples in this study are from female individuals (n_F=7), whilst the SMA MNs are both male (n'_M=3) and female (n'_F=2). As such, there is a sex imbalance between the WT and the SMA groups, which could account for the observed apparent down-regulation of Y-linked genes in SMA. Though there exists evidence for sex-specific disease modifiers in SMA [302], this suggests that further analyses might need to include a further stratification according to the sex of the sample donor/animal.

Finally, I focused this study on an overall WT vs. SMA comparison. As such, this pilot study is agnostic to the disease stage represented by the various samples it includes. There is only a limited amount of samples per species per technology per disease stage. As such, the systematic re-analysis of the raw microarray files, followed by the subsetting of the RNAseq files to include only genes present on the microarrays, will be necessary to establish a systematic transcriptomic natural history of SMA that makes use of all the available SMA transcriptomic information and that allows stratification by disease stage.

2.4.2. On the interest of meta-analysis for the dissection of SMA pathophysiological processes

2.4.2.1. On similar approaches used in other neurological or muscle wasting conditions

Whilst I report the first meta-analysis of SMA transcriptomics datasets, similar systems biology approaches have recently been applied to other neurological muscle wasting conditions.

Some of these meta-analyses have identified pathways shared between related conditions. For example, the meta-analysis of 2,600 samples from 26 Alzheimer's disease (AD), 21 Lewy body dementia, 13 ALS-FTD (ALS-fronto temporal dementia) human CNS transcriptomics datasets and age-matched controls revealed neuroinflammation, deficient energy metabolism, and proteostasis failure across neurodegeneration [303]. (Other AD and Parkinson's disease (PD) studies include [304,305]). Others focus on summarising the plethora of information available across samples and models: for example, the meta-analysis of 2,114 brain RNAseq datasets from 1,234 AD post-mortem human subjects and from 96 distinct mouse studies (including but not limited to AD) provides the most extensive cross-species AD-related transcriptomic resource available to date [266]. Finally, some meta-analyses provide new insight in underlying disease mechanisms. For example, a meta-analysis of 3 arrays containing samples from FTD patients' brains revealed found that familial and sporadic FTD genes were associated to different signaling pathways (respectively, Wnt vs. MAPK signaling pathway) [306]. Similar studies have been reported for human brain and blood samples in Huntington's disease [307].

Meta-analyses in muscle wasting conditions have also been reported [308–312], relevant to SMA disease characterisation (notably for the integration of recently published muscle transcriptomics datasets [313]), but out of the scope of this chapter's discussion.

2.4.2.2. On the integration of the several SMA mouse models contributions

One of the most attractive prospects in the pursuit of this project is the integration of biological information, not only across species but also across mouse models. The transcriptomic information available from published SMA transcriptomics mouse studies span across the SMNΔ7, the Taiwanese, and the $Smn^{2B/-}$ models. Such information integration will allow for better integration of new datasets with therapeutic interventions (see Chapter 3).

2.4.2.3. On the gene candidates identified in this pilot study

The initial aim of the meta-analysis was to integrate and compare SMA transcriptomic studies with each other to elucidate transcriptome alterations across disease progression in SMA motor neurons. I faced computational difficulties in the pre-processing of two of the SMA samples of the murine RNAseq samples that prevented the inclusion of a mouse-human overlap to our pilot project. These technical issues will be resolved in the subsequent expansion of this study, which include (1) integration of further human samples from micro-arrays and (2) the inclusion of murine RNAseq datasets. The latter will not only provide critical insight in the generation of a cross-model natural history of transcriptomic changes in SMA, but also yield mouse-human overlaps that will validate and narrow down the pool of candidates across model.

I discuss below the mechanistic and therapeutic potential of two of the primary findings from this meta-analysis: the potential role of MAZ and MOVO-B TF, and the role of calcium dysregulation.

2.4.2.3.1. *MAZ and MOVO-B transcription factors*

MAZ (Myc Associated Zing finger protein) is a nuclear six zinc finger protein with a G-rich binding motif [314]. MAZ has been reported to act as “insulator” element blocking the effects of a distal enhancer on a promoter, as well as blocking the advance of a transcribing RNA polymerase II, leading to alternative RNA splicing patterns [315]. In parallel, the mouse ortholog to MOVO-B (‘mouse homologue of *Drosophila* Ovo protein’), *Ovol2* (ortholog to Ovo like zinc finger 2) is normally expressed early during embryogenesis [316] and regulates cell fate decisions between neuroectoderm and mesendoderm [317]. It, too, binds to a G-rich sequence (5'-G(G/C/T)GGGGG-3’) [318].

It is uncertain at this stage whether the reported enrichment of genes with a MAZ (viz. MOVO-B) binding motif is due to any dysregulation of the MAZ (viz. MOVO-B) factor itself in SMN-deficient contexts (in which case, further investigation of possible indirect interactions between SMN and MAZ viz. MOVO-B are warranted), whether it has to do with the G-rich nature of the MAZ (viz. MOVO-B) motif sequence (which are known to form G-quadruplexes contributing to the formation of double-stranded break and of DNA damage, as previously reported in SMN-deficient context [140]), or whether the observed enrichment of MAZ motif-containing gene is an artifact of AS changes in SMA. Further mechanistic studies should be conducted to assess the therapeutic potential of MAZ- and MOVO-B targeted genes in SMA.

2.4.2.3.2. *Calcium dysregulation*

The present meta-analysis of human SMA RNAseq samples has provided new insights into novel gene candidates for therapeutic intervention in SMA. Amongst genes dysregulated in SMA, it shows a particular enrichment of both plasma membrane proteins as well as of calcium-binding proteins. Whilst Rizzo et al. ([293]) is the first report to confirm in human samples the differential expression of genes previously identified as mis-spliced in murine SMA MN (including calcium/calmodulin dependent protein kinase II delta (*CAMK2D*), it does not explore further its biological significance. Similarly, the manual curation of published reports shows that the majority of human studies identified in this chapter focus on membrane-related dysregulations but do not pursue calcium-binding proteins as candidates for deliberate downstream efforts on *in vitro*/*in vivo* validation, pathway dissection, nor for therapeutic development (***Table*** 2-9).

Altered calcic homeostasis, in particular with effects to (1) the calpain system and its regulation of neuronal apoptosis via *Cdk5* [319], (2) defects in axonal calcium-channel clustering, and (3) reduced entry of mitochondrial Ca^{2+} levels, with functional consequences on exo- and endo-cytosis in SMA motor terminals; have long been described in SMA mouse models [162,320,321]. In human SMA cellular models, calcic dysregulation has been reported not only in neurons but also in astrocytes [257] and cardiac cells [322]. Ratiometric live-cell calcium imaging of SMA iPSC astrocytes reveals an increase in basal intracellular calcium levels and a reduced calcium response to ATP, with functional consequences on downstream pathways (incl. the upregulation of pERK1/2) [257]. In SMA iPSC cardiomyocytes, the sarco-endoplasmic reticulum ATPase (SERCA) pump, which extrudes Ca^{2+} into the endoplasmic reticulum (ER) lumen is downregulated early in the disease course [322]. These processes appear to be cell autonomous [322] (i.e. not secondary to MN degeneration). Though the pleiotropy of Ca^{2+} dysregulation suggests at least some degree of SMN dependency, calcic dysregulation is not fully accounted by changes in alternative splicing [322]. Whilst a recent report highlights the role of calcium on the regulation of SMN splicing (possibly due to the modulation of the RNA-binding capacity of splicing factors) [323], the mechanistic understanding of how SMN deficiency precisely affects the different steps of calcium signalling is still elusive and remains to be further explored.

This is the first study centred on human SMA neurons that identifies calcium-binding proteins as candidates for therapeutic development in SMA. Interestingly, calcium signalling dysregulation is known as one of the early–stage and key processes in the pathogenesis of multiple neurodegenerative diseases including AD, PD, ALS, HD, and spinocerebellar ataxias (SCA) [324]. In other neurodegenerative disorders, including AD, not only neuronal but also astrocytic and microglial calcium dysregulation has been pursued as a therapeutic target [325]. The scope of experimental therapeutic interventions includes targeting plasma membrane receptors and channels, cytosolic calcium, endoplasmic reticulum calcium, mitochondrial calcium, and lysosomal calcium [325], and has yielded treatments for patients, including memantine (a glutamate receptor and mild calcic channel antagonist) in AD. In *Smn*-deficient murine MNs, a cAMP analog does restore the morphological and functional alterations of MN axons by increasing the frequency of Ca^{2+} transients currents [320]. cAMP is broken down by PDE-4, for which there are three drugs approved for human use: apremilast, crisaborole, and roflumilast [326]. Their repositioning in SMA to improve the calcic dysregulation is one potential avenue for therapeutic development and remains to be assessed, in particular with regards to their ability to cross the neurovascular unit.

In conclusion, I have presented the first meta-analysis of transcriptomics meta-analysis conducted in SMA. Design, processing, and downstream analyses have focused on a proof-of-concept using human RNAseq datasets from three individual studies. I have built a robust analysis pipeline for differential gene expression analysis and demonstrated feasibility by identifying enriched cellular components (including cellular membranes), biological functions (including calcium-binding), and DNA binding motifs (including MAZ and MOVO-B motifs).

Future analytic steps will include alternative splicing analyses, mouse-human ortholog overlaps, and co-expression network analysis. This work provides the first step to integrate transcriptomic information across models and species and build a systematic molecular natural history of neurodegeneration in spinal muscular atrophy. It also provides background for the integration of transcriptomic datasets such as the one I present in Chapter 3.

Table 2-9: Manual curation of published SMA transcriptomics paper for plasma membrane or calcium-binding protein validation

	Reference	Technology	Plasma membrane candidates	Calcium-binding candidates
H. sapiens	Corti et al. 2012 [290] (GSE27205)	microarray	membrane bound or organelle: 19% of the 586 DE genes PLP1 validated at the protein level	
	Corti et al. 2012 [290] (GSE272056)	microarray		
	GSE65470 (2018)	microarray	not applicable	
	Yoshida et al. 2015 [160]	RNA-Seq	study of AchR clusters on iPSC-MN/C2C12 myotubes cocultures: synapse loss in SMA MNs surviving neurodegeneration	
	Rizzo et al. 2019 [293]	RNA-Seq	identification of axon-related proteins, synapse proteins, and K^+ channels Cellular component (CC) 'membrane' significantly dysregulated	calcium/calmodulin dependent protein kinase II delta
	Hauser et al. 2020 [292]	RNA-Seq	not applicable – no SMA samples	
M. musculus	Nichterwitz et al. 2020 [152]	RNA-seq	CC 'synapse' and 'neuron projection' significantly dysregulated	
	Doktor et al. 2017 [139]	RNA-seq	'axon guidance' significantly dysregulated	U12- dependent intron retention in subunits to voltage-gated Ca^{2+} channels: *Cacna1a, Cacna1b, Cacna1c, Cacna1e* and *Cacna1h*
	Murray et al. 2015 [144]	RNA-seq	in SMA vulnerable MNs (vs. SMA resistant MNs): 1.73-fold downregulation of 'membrane' 1.38-fold upregulation of mitochondrial inner membrane in SMA vulnerable MNs (vs. WT vulnerable MNs): 2.7 downregulation of 'mitochondrial membrane' 1.52 downregulation of 'focal adhesion'	in SMA vulnerable MNs (vs. SMA resistant MNs): 3.07-fold downregulation of 'metal ion binding' in SMA vulnerable MNs (vs. WT vulnerable MNs): 3.2 upregulation of 'metal ion binding'
	Staropoli et al. 2016 [262]	3' & exon microarray		upregulation of cyclin dependent kinase inhibitor 1A (p21) note: specifically cleavage of Cdk1A by CASP3-like caspases (calcium dependent)
	Zhang et al. 2013 [141]	RNA-seq	defects in maturation and neurotransmission of NMJ and MN deafferentation at PND5 study at PND1: no significant enrichment of any pathways or biological processes for genes up-regulated, down-regulated, or affected by splicing in MNs or white matter function-modifying splicing changes in *Gria2* and *Gria4* (core subunits of AMPA-type glutamate)	at PND2: upregulation of all subunits of the C1q complex in SMA MNs (*C1qa*, *C1qb*, and *C1qc*) note: calcium-dependent first step of the complement system (C1q)
	Murray et al. 2010 [261]	Exon microarray	not applicable – comparison of MNs innervation different muscle groups	
	Bäumer et al. 2009 [146]	Exon microarray	1.5-fold downregulation of *Tmprss11d* at PND7 1.6-fold upregulation of Laminin alpha 2 (*Lama2*) at PND1 alternative splicing of the chondrolectin *Chodl*	upregulation of *C1qc* at PND7 upregulation of *Gp49* at PND1
	Zhang et al. 2008 [135]	Exon microarray	several extracellular matrix & cell adhesion protein-coding genes dysregulated: *Cadherin1, Postn, thrombospondin 3, sarcoglycan epsilon,*	*Cacna1d* 2-fold upregulation validated with RT-qPCR

3. Chapter 3: Treatment- and time-dependent transcriptomic changes in laser-captured microdissected motor neurons, in a severe mouse model of SMA

How efficient is Pip6a-PMO, a new cell-penetrating peptide conjugated, SMN-targeting splice-switching therapy, at correcting the molecular phenotype in SMA-affected motor neurons?

What is the course of splicing changes through disease progression in motor neurons in a severe model of SMA?

Is combinatorial therapy needed in MN, and if so, for what targets?

***

3.1. Introduction

In Chapter 2, I tackled a proof-of-concept meta-analysis of published and deposited transcriptomics human RNAseq datasets in SMA, with the goal of developing a systematic molecular natural history of quantitative and qualitative transcriptomic alterations during SMA neurodegeneration. In this chapter, I report a novel dataset exploring time- and splice-switching treatment- transcriptomic alterations in laser-captured microdissected motor neurons (LCM-MN) in a severe mouse model of SMA.

3.1.1. Splice-switching therapy on *ISS-N1* locus in *SMN2* gene

As presented in Chapter 1, recent therapeutic advances in SMA include the commercialisation in 2016/2017 of nusinersen. Nusinersen is a naked, 2'-O-methoxyethyl group-bearing, phosphothioate-backboned, splice-switching 18-mer oligonucleotide targeting the intronic splice site N1 (*ISS-N1)* locus in the *SMN2* gene, and increases the inclusion of exon 7 [192]. Nusinersen extends survival and quality of life in a cohort of patients with SMA [195]. The major factor affecting treatment efficacy is its precocity: the earlier patients are treated (before severe neurodegeneration occurs), the better the outcome [228,232,233].

The primary cell type affected by SMA are spinal cord alpha motor neurons in the spinal cord (SC) parenchyma. Some SMA type I patients display specific thalamic defects [14,15] justifying the additional cerebral parenchymal distribution for any disease-modifying treatment. The spinal cord is physiologically separated by the blood-spinal cord barrier (BSCB) from the blood flow of vertebral arteries [327]. Nusinersen's size (molecular weight 7,127 Daltons) and overall negative charge exclude the possibility of crossing the neurovascular unit [328]. As such, nusinersen requires a repeated intrathecal injection regimen (three initial loading doses, then every 4 months, for life). This presents several challenges.

3.1.2. Challenges of intrathecal injections in SMA patients

Intrathecal injections require anaesthesia, specialised clinical teams (neuroradiologist or paediatric anaesthesiologist), and specialised administrations centres [203]. The mode of administration is in itself responsible for side effects experienced by patients, of whom 12% report headaches, 11% back pain, and 10% post-lumbar puncture syndrome [203]. In addition, patients with SMA often present orthopaedic defects, including neuromuscular scoliosis [35], which can complicate the fluoroscopic guidance most commonly used for lumbar punctures [329]. The high effort of the procedure and the procedure-related side effects are named as causes for treatment discontinuation [330].

3.1.3. New splice-switching platform candidate: Pip6a-PMO

At the onset of this work in 2015, given the variety of challenges associated with intrathecal injections, the development of an effective and easily administrable treatment for this devastating disease was still of clinical importance in SMA. It is unlikely that all spliceopathies (many of which are rare diseases) will benefit from the type of extensive oral drug development programs that have led to the approval of an oral splice-switching small molecule (risdiplam) in SMA. Therefore, at the time of the writing of this thesis (2022), splice-switching antisense oligonucleotide (AON) treatments still hold considerable promise as a therapeutic platform for neurological conditions [219]. Intravenous (IV) administration does not hold the same limits as intrathecal administration. One of the strategies for IV administration is the formulation of AON with a delivery moiety – which supposes AON chemistry compatible with such modifications.

Based on previous studies on cationic-rich peptides [331–334], followed by polyarginine peptides [331], the Wood lab has been investigating derivatives of penetratin [335], called peptide nucleic acid/PMO-internalization peptides (Pip) [336]. Further optimisation of the different Pip series (Pip5, Pip6) has led to the identification of two peptides, Pip6a and Pip6d, that are able to reach the central nervous system tissues [336].

Pip6a can be conjugated to AON with no charge on their backbone (therefore excluding phosphorothioate backbones such as nusinersen) – such as phosphorodiamidate morpholino oligomers (PMO). Following these developments, the Wood laboratory developed a Pip6a-PMO conjugate targeting the *ISS-N1* locus of *SMN2*.

3.1.4. Hypothesis and goals

Previous work in the Wood lab assessed the efficacy of Pip6a-PMO delivered pre-symptomatically to a severe mouse model of SMA for the correction of splicing defects in muscle [313]. This data set was used to conduct genomic-led drug discovery using connectivity mapping [337] . No study has been reported on the efficacy of Pip6a-PMO to correct the global transcriptomics defects (described in Chapter 1) resulting from SMN deficiency following presymptomatic treatment. We therefore chose to conduct an exome-level transcriptomic analysis from laser-captured microdissected motor neurons (LCM-MN) in a severe mouse model of SMA.

There are three primary goals to this study. The efficacy of Pip6a-PMO on transcriptomics defects in neurons is unknown. The first goal of this study is to evaluate the efficacy of Pip6a-

PMO, a novel intravenously delivered splice-switching therapy, on transcriptomic defects in a severe mouse model of SMA.

As outlined in Chapter 2, the discrepancy in onset of alternative splicing changes is one of the existing areas without a consensus in the SMA literature. The second goal of this study is to use this dataset to study the course of splicing changes during SMA and identify novel gene candidates involved in molecular processes at play in neurodegeneration.

By comparing the transcriptomic differences of oligonucleotide-treated samples to those different stages of SMA disease progression, one can identify defects left uncorrected by splice-switching therapies that could be targeted by a combinatorial therapy. The third goal of this study is to evaluate the need for combinatorial therapy in SMA nervous system following splice-swiching therapy and provide gene candidates, if appropriate.

3.2. Material and Methods

3.2.1. Pip6a-PMO synthesis

Morpholino oligonucleotide (PMO) were purchased from Gene Tools. Peptide synthesis reagents and solvents purchased from Merck, SigmaAldrich and FluoroChem. All solvents are HPLC grade and used without further purification. Synthesis, conjugation, and purification of Pip6a (Ac-(RXRRBRRXRYQFLIRXRBRXRB)-COOH) [336] to the ISS-N1(-10-27) targeting PMO sequence (5'-ATTCACTTTCATAATGCTGG-3') and or the scrambled sequence (5'-ATTGTCTATCAAAATCCTGC-3') [338] were conducted following the protocol outlined in the recent patents WO2020030927A1 [339]. Pip6a-PMO and Pip6a-PMO_scr were diluted in 0.9% saline.

3.2.2. SMA animal model

All animal procedures were carried out at the BMS facility in Oxford, in accordance with procedures authorised by the UK Home Office (PIL no: IF5D73644). For all experiments, litters were randomly assigned at birth and whole litters composed of both sexes were used.

The severe 'Taiwanese' model (*Smn*$^{-/-}$;*hSMN2*$^{+/-}$) is an SMA-like mouse strain (FVB.Cg-Smn1tm1HungTg(SMN2)2Hung/J) in which in the mouse *Smn* exon 7 is deleted and animals carry two copies of human *SMN2* [181]. These mice typically start showing symptoms around PND5 [239,340] and reach their humane end point between PND7 and PND9.

3.2.3. Genotyping

Each ear- or tail- clip was incubated in 200 µL of DirectPCR Lysis Reagent (Mouse Tail) (Viagen Biotech, 102-T) and 5 µl of proteinase K for 5-8 hour (or overnight at 65 °C), followed by 10 minutes at 100°C. Samples were spun for 1' at 8,000 x*g*.

Genotyping for *Smn* was performed on 2.5 µL lysed reagents, using DreamTaq™ Green Buffer (10X) (Thermoscientific, B71). Products were separated on 2 % agarose gel with 0.05 µl/ml ethidium bromide at 120 V for ~75 mins, with 50 bp ladder for reference if needed. Genotyping for sex was performed using the primers and cycling conditions described in [341] (***Table*** 3-1).

Table 3-1: PCR primers for Smn and sex genotyping.

121_mSmnTg	AGC CTG AAG AAC GAG ATC AGC
122_SmnTg_fw	GCG TTG ATG GCA TTT CAC TA
123_SmnTg_rv	TGC TCG TGG AAG TCA GTG AG
Sex_mouse_Tunster_fwd	5'-CACCTTAAGAACAAGCCAATACA
Sex_mouse_Tunster_rev	5'-GGCTTGTCCTGAAAACATTTGG

3.2.4. Drug administration

The treated groups were injected intravenously (in the facial vein) with 2x10 µg/mL Pip6a-PMO or Pip6a-PMO_scrambled at PND0 and PND2, at a volume of 5 µL/g of body weight (n=3 per group).

3.2.5. Tissue collection & laser-capture microdissection (LCM)

Mice pups (all below PND10) were culled at pre-symptomatic (post-natal day 2, PND2) or symptomatic stages (post-natal day 7, PND 7) by decapitation. Death was verified by exsanguination.

Embedded spinal cords were subsequently sectioned at 10 µm on a cryostat (Leica). Initial sections were collected on standard microscope slides and stained using a 1% toluidine blue dye to confirm the presence of MN cell bodies. Following sections were collected on membrane-coated slides required for the LCM microscope (MemberaneSlide 1.0 Pen (D), Zeiss). Slides were immediately processed for LCM by Prof. Lyndsay M Murray (University of Edinburgh).

Sections were stained with 1% toluidine blue and dehydrated using ethanol. The protocol used was as follows:

Step 1: Wash in DEPC water for approximately 5 seconds

Step 2: Immerse in 1% toluidine blue for approximately 1 minute

Step 3: Wash in DEPC water for approximately 5 seconds

Step 4: Immerse in 70% ethanol approximately 20 seconds

Step 5: Immerse in 95% ethanol approximately 20 seconds

Step 6: Immerse in 100% ethanol approximately 20 seconds

As noted, the purpose of this exercise was to obtain RNA and therefore steps were taken to minimise RNAse exposure during this process. DEPC water was made per manufacturer instructions (Sigma) and was used to prepare all ethanol solutions and the 1% toluidine blue dye. In addition, the containers used to hold these solutions were cleaned using RNAzap and rinsed with DEPC water.

LCM was carried out on the PALM MicroBeam microscope (Zeiss). Individual MNs were outlined and marked for collection by the LCM microscope. The microscope cuts and propels tissue into the Zeiss AdhesiveCap tube. Collected tissue was snap frozen on dry ice to be stored at -80°C.

3.2.6. RNA extraction

Total RNA was isolated using guanidinium thiocyanate acid phenol/chloroform extraction (Trizol®, Life Technologies), resuspended in nuclease-free water. RNA from LCM-MN was then prepared using the Affymetrix GeneChip™ WT Pico Kit (Thermofischer Scientific, cat. no. 902622).

3.2.7. Microarray analysis

The exome was analysed on GeneChip® Mouse Transcriptome Assay 1.0 (Affymetrix) by the platform of the Karolinska Instituet, with 100 ng RNA per sample. Annotations for the Mouse Transcript Array 1.0 at the transcript, gene, and probe levels were obtained from the Affymetrix website (http://www.affymetrix.com/products_services/arrays/specific/mo_trans_assay.affx#1_4).

3.2.8. Bioinformatics analyses

3.2.8.1. Pre-processing, quality control & normalisation

Bioinformatics processing and analyses of the `.cel` files were performed using R [285] in the R Studio environment[342], with Bioconductor 3.15 packages; and in Python 3.10.6.

Quality control, background correction and robust multi-array average (RMA) normalisation [343] were performed at the probe level using `oligo` R package [344]. For principal component analysis

I used the `prcomp` R function on the RMA normalised gene expression data at the gene level. For differential expression analysis, I used `limma` R package [345]. P values were adjusted using the Benjamini & Hochberg correction [346].

3.2.8.2. Differential expression analyses

The level of a given exon is a result of the transcription rate of the gene and the inclusion rate of the exon (exon level = transcription * splicing). The sorting of differential expression of genes (DEG) was performed following RMA at the exon level. To consider both the biological relevance and the statistical significance in the ranking of transcripts, I calculated *a posteriori* the π-value [347]:

$$\pi\text{-value}_i = |\log FC_i| * -\log_{10}(\text{adj.P.Val}_i)$$

I considered genes with π-value > 1.3 as differentially expressed.

3.2.8.3. Differential splicing analyses

Using a custom splicing index pipeline in Python, I define the normalised intensity (NI) at the gene-level, as:

$$NI = \frac{\text{probe set intensity}}{\text{expression level of the "gene"}}$$

Using NI, I then define SI the splicing index of an exon, as:

$$SI = \log 2 \frac{\text{NI (sample condition 1)}}{\text{NI (sample condition 2)}}$$

An exon was considered alternatively spliced if |SI| > 0.5.

3.2.8.4. Functional enrichment and network analyses

Functional enrichment analysis, including gene ontology, was performed using the g:Profiler web server [288]. Networks were visualised using Cytoscape (version 3.8.2) [348], using the stringApp plugin [349] for the study of the functional associations and/or physical protein/protein interactions between protein-protein; the BinGO plugin to determine to determine which Gene Ontology (GO) categories are statistically overrepresented in a set of genes [350], and the CentiScaPe plugin for the analysis of networks' parameters [351].

3.3. Results

3.3.1. Tissue collection, sample preparation, and microarray quality control

We performed a study of the transcriptomic changes in the spinal cord MNs between pre-symptomatic (post-natal day 2, PND2) and symptomatic (PND7) SMA pups. Tissues were collected from WT or SMA mice, at PND2 and PND7, either not treated, or treated at PND0 and PND2 with Pip6a-PMO or Pip6a-scrambled_PMO as controls (***Figure*** 3-1). Pip6a-PMO targets the same *ISS-N1* sequence as nusinersen, with a different oligonucleotide backbone. For delivery to the central nervous system, the oligonucleotide moiety is conjugated to Pip6a, a cell-penetrating peptide.

Litters consisted of both male and female pups. The sex of each pup was confirmed by a *post-hoc* analysis of three X-linked markers (***Supplementary Figure*** 7-2, ***Supplementary Table*** 7-5). Of note, the 'SMA PND2' and the 'Pip6a-PMO-treated' groups both consist of males only.

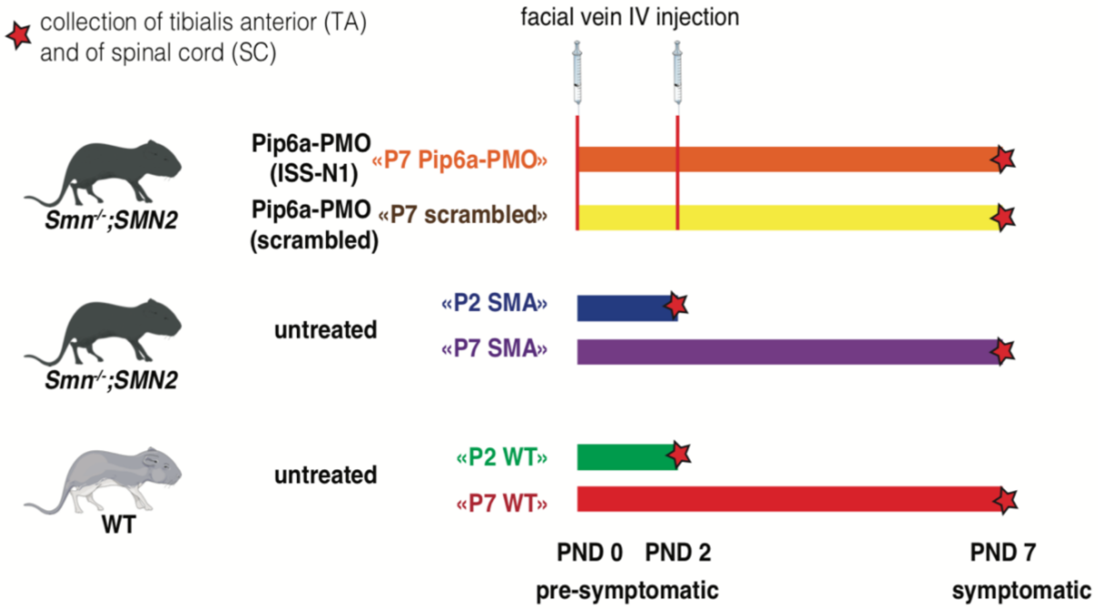


Figure 3-1: Experimental design for the study of the time- and oligonucleotide treatment-dependent transcriptomic changes in wild-type (WT) and Taiwanese SMA mice

Schematic representation of experimental design and sample collection. Six mice groups (n=3 per group) were analysed. Taiwanese SMA mice ($mSmn1^{-/-};hSMN2^{+/+}$) of the treated groups (Pip6a-PMO$_{ISS-N1}$ and Pip6a-PMO$_{scrambled}$) were injected with 10 µg/mL in the facial vein at post-natal day (PND) 0 and PND2. Pups were sacrificed at PND2 or PND7, as shown.

MN bodies were isolated from the adjacent white matter by laser-capture microscopy (***Figure*** 3-2). Number of captured MN varies between 200-300 per sample (***Table*** 3-2). RNA was extracted and transcriptomics levels were measured with Affymetrix® microarrays, which bear 4 probes per exon and an average of 40 probes per gene. Quality control was assessed by generating chip pseudoimages and boxplot distributions before and after RMA normalisation (***Supplementary Figure*** 7-3).

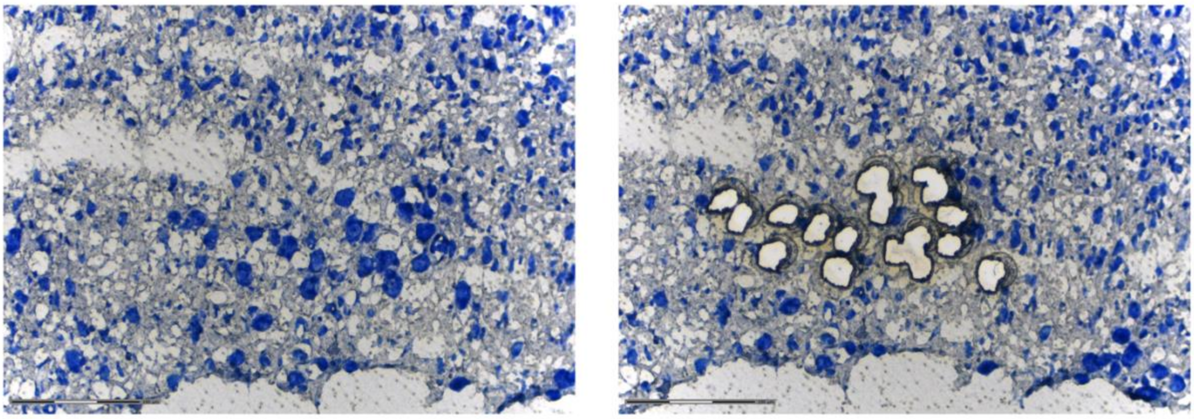


Figure 3-2: Precise capture of motor neurons by laser-capture microdissection
Sequential representative images of the embedded spinal cord 10 µm section before (left) and after (right) LCM of MN. The 1% toluidine-blue dye confirms the presence of MN and guides the outlining and collection process. The cut MN bodies are collected and immediately snap-frozen prior to further processing. Scale bar: 150 µm.

Table 3-2: Summary of the laser-capture microdissections on motor neurons

Genotype	Age	Treatment	Number of MN
SMA	P2	-	256
			248
			241
WT	P2	-	219
			240
			255
SMA	P7	-	219
			257
			218
WT	P7	-	193
			247
			250
SMA	P7	Pip6a-PMO	359
			286
			260
SMA	P7	Scrambled-PMO	194
			239
			228
			195

3.3.2. The LCM-captured cell population is enriched in neuronal cells and motor neurons

I confirmed at the RNA level that the cell population is enriched in *Nefh, Nefl, Gabra1, Syt1, Snap25* (neuronal markers) and in *Mnx1*, *Chat*, *Isl1*, *Slc18a3* (motor neurons markers) (***Figure*** 3-3).

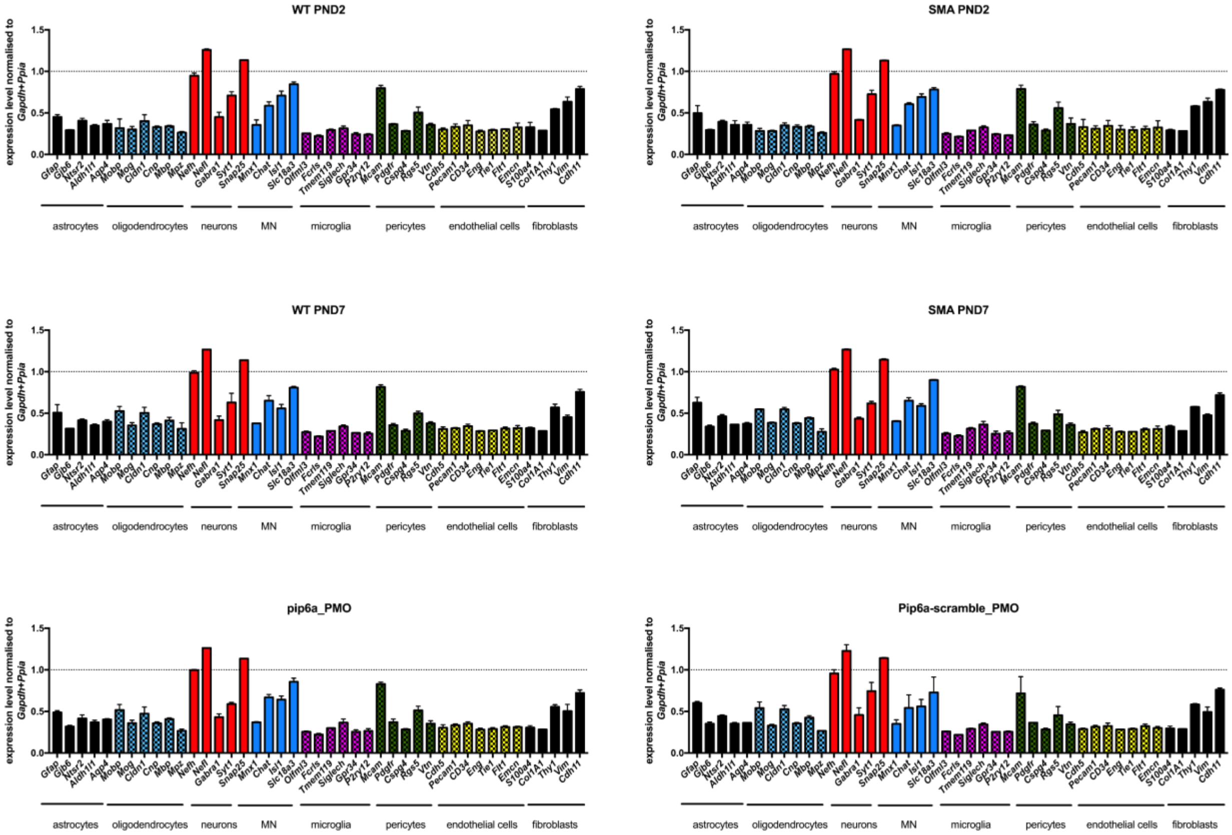


Figure 3-3: Quantification of the relative expression levels of cell-lineage specific markers in each LCM-MN group confirms the neuronal and MN enrichment of the captured samples.

Post-hoc *transcriptomic quantification of the expression of known lineage markers, normalised on a hybrid normalisor (*Gapdh+Ppia*). Each panel corresponds to an experimental treatment/timepoint group.*
*X-axis: array of known markers for astrocytic lineage (*Gfap, Gib6, Ntsr2, Aldh1l1, Aqp4*), oligodendrocytic lineage (*Mobp, Mog, Cldn1, Cnp, Mbp, Mpz*), neuronal lineage (*Nefh, Gabra1, Syt1, Snap25*), MN lineage (*Mnx1, Chat, Isl1, Slc18a3*), microglial linage (*Olfml3, Fcrls, Tmem119, Siglech, Gpr34, P2ry12*), pericytic lineage (*Mcam, Pdgfr, Cspg4, Rgs5, Vtn*), endothelial lineage (*Cdh5, Pecam1, CD34, Eng, Tie1, Fit1, Emcn*), and fibroblastic lineage (*S100a4, Col1A1, Thy1, Vim, Cdh11*).*
Y-axis: normalised expression level, n=3 MN samples per group.

3.3.3. Biological replicates show larger intra-replicate variance than inter-group variance

Hierarchical clustering and principal component analysis (PCA) reveal heterogeneity at the global transcriptomic level. The first three components explain only 41% of the samples' variations (***Figure*** 3-4). The first dimension does not appear to discriminate samples by timepoint or disease status. The second dimension separates PND2 samples, and WT_PND7, Pip6a-PMO-treated, and Pip6a-scramble-treated samples. Several of the biological groups present a vast scattering of samples, including both pre-symptomatic groups WT_PND2 and SMA_PND2. This yields low power for differential expression analysis. Treatment with Pip6a-PMO shifts the centre of gravity of that group towards the WT_PND7 group.

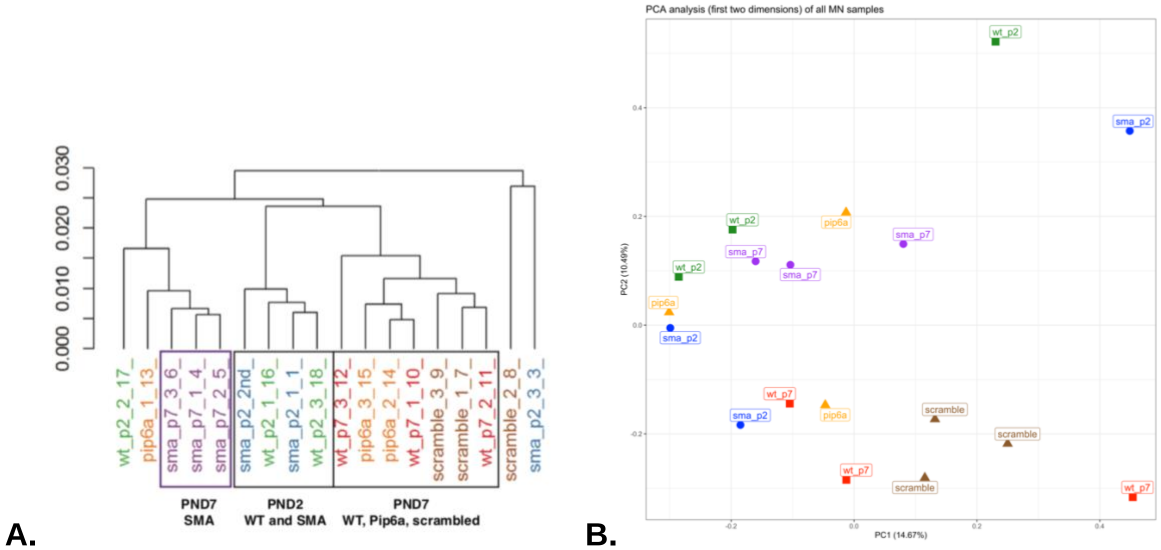
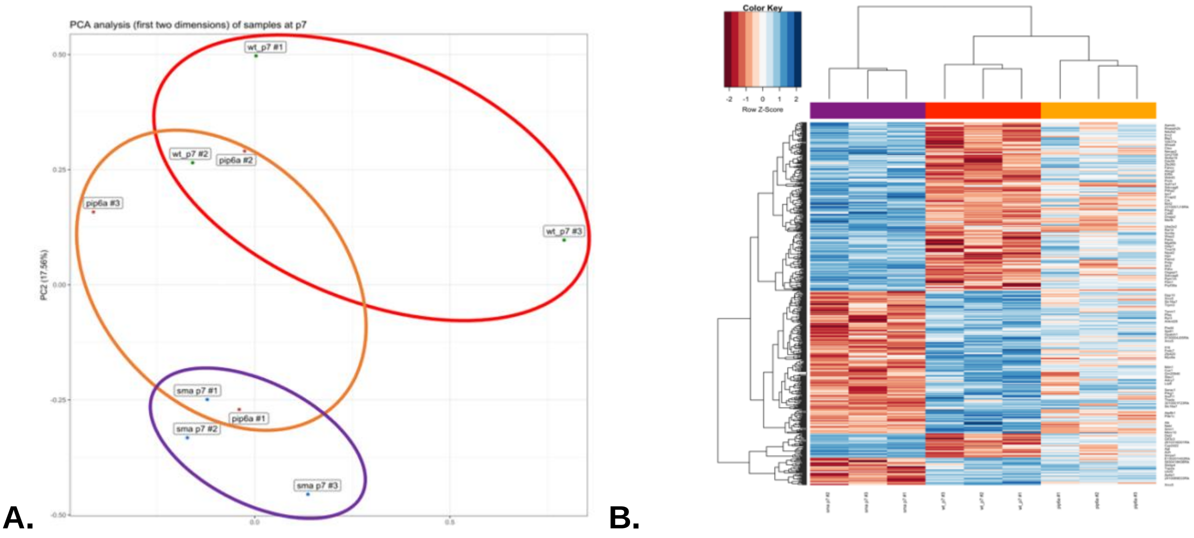


Figure 3-4: Clustering and dimensionality reduction of the normalised LCM-MN transcriptomic data.

Samples are identified by their colour: WT_PND2 (green). WT_PND7 (red); SMA_PND2 (blue); SMA_PND7 (purple); Pip6a-PMO (orange); Pip6a-scramble_PMO (maroon). Each dot represents an individual sample. ***(A)*** *The hierarchical cluster-derived dendrogram shows three distinct clusters: SMA_PND7 animals, presymptomatic animals (both WT and SMA), and Pip6a-PMO animals, who cluster alongside WT animals. Interestingly, Pip6a-PMO*$_{ISS\text{-}N1}$ *#1 does not cluster alongside its groupmates.* ***(B)*** *PCA plot of the individual MN samples using whole-transcript gene expression profiling. The first two principal PCA components (x-axis: PC1=14.67%, y-axis: PC2=10.49%) explain a minority of the variance between all the samples.*

3.3.4. Pip6-PMO efficiently changes the transcriptomic landscape of motor neurons after neonatal administration in SMA pups

After two neonatal 10 µg/g IV doses of Pip6a-PMO, the transcriptomic landscape of SMA pups is closer to that of WT pups than that of untreated SMA pups (***Figure*** 3-5). Noticeably, one of the Pip6a-PMO-treated pups (Pip6a-#1) clusters with untreated SMA pups, suggesting either a fault in the injection of Pip6a-PMO or differences in the gene expression changes following treatment in that animal. Hierarchical clustering similarly highlights that Pip6a-PMO-treated pups cluster more with WT pups than with their genetically similar SMA counterparts (***Figure*** 3-5).



C.	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Proportion of Variance	0.2207	0.1756	0.1211	0.1079	0.106	0.0936	0.08942	0.08577	0.00E+00
Cumulative Proportion	0.2207	0.3963	0.5174	0.6253	0.7312	0.825	0.91423	1	1.00E+00

Figure 3-5: Pip6a-PMO$_{ISS-N1}$ neonatal treatment of Taiwanese SMA mouse pups shifts the transcriptomic landscape from SMA-like to WT-like

Samples are identified by their colour: WT_PND7 (red); SMA_PND7 (purple); Pip6a-PMO (orange). Each dot represents an individual sample. ***(A)*** *PCA plot of the individual MN samples using whole-transcript gene expression profiling. X-axis: PC1 (22.07%). Y-axis: PC2 (17.56%). Dimensionality reduction shows that the Pip6a-PMO$_{ISS-N1}$ treated group is further away from the SMA group following natural disease progression, but still does not cluster completely with the WT group.* ***(B)*** *Microarray heatmap and clustering based on the hierarchical clustering of the top DE genes. Pip6a-PMO$_{ISS-N1}$ treated animals cluster alongside WT animals and are further away from SMA animals. However, the gene expression levels are not completely restored in Pip6a-PMO$_{ISS-N1}$ treated animals. Red: down-regulated genes. Blue: up-regulated genes.* ***(C)*** *Proportion of variance and cumulative proportion of the variance explained by the principal components derived from the PCA of samples at PND7. The first two dimensions PC1 and PC2 are the axes of* ***(A)****.*

Pip6a-PMO effectively restores >99% of the early transcriptomic expression changes occurring in SMA pups. Treatment with 2x 10 µg/g neonatal Pip6a-PMO treatment abolishes the early changes in transcript expression in SMA-treated pups and restores WT-like levels of expression for some of the most dysregulated probes in SMA, including *Sulta1a1* and *Nav3* (***Figure*** 3-6).

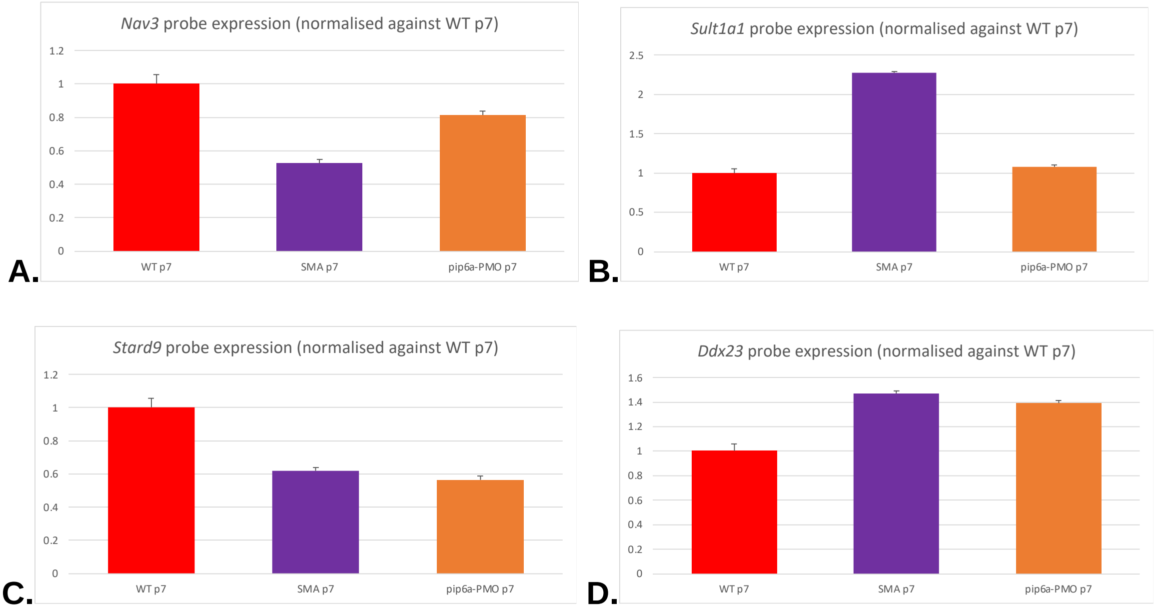
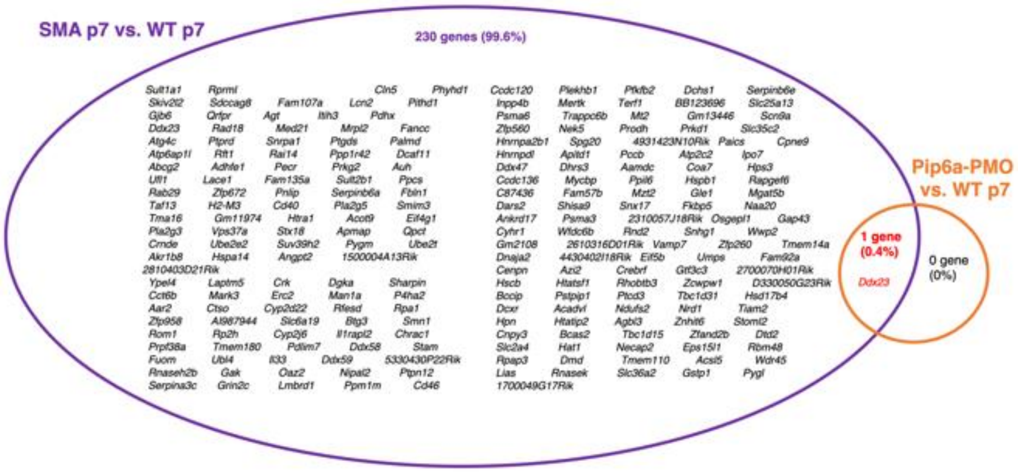


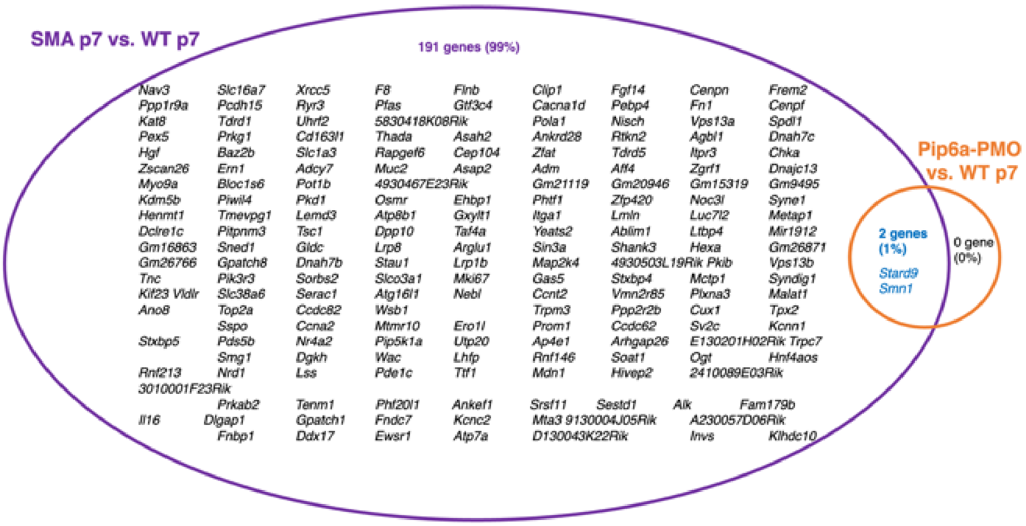
Figure 3-6: Pip6a-PMO restores the expression of some of the most dysregulated probes at early SMA symptomatic stages (Nav3, Sult1a1) but fails to correct Stard9 and Ddx23.

Post-hoc *transcriptomic quantification of the RMA-normalised probe expression from two of the most SMA dysregulated genes:* ***(A)*** Nav3*,* ***(B)*** Sult1a1*; and of two genes not corrected by Pip6a-PMO$_{ISS\text{-}N1}$ treatment:* ***(C)*** Stard9*, and* ***(D)*** Ddx23. *Probe expression level has been normalised to the WT-PND7 group. Groups are identified by their colour: WT_PND7 (red); SMA_PND7 (purple); Pip6a-PMO (orange).*

Differential expression (DE) analysis at early symptomatic stages (PND7) is here conducted using normalisation at the probeset level to allow for downstream splicing index analysis. DE analysis between SMA and WT pups shows 552 differentially expressed transcripts within 419 unique genes with changes in expression, 231 of which are up-regulated, and 193 of which are down-regulated. (A gene is considered differentially expressed (DE) if at least one exon in it has a π-value >1.3) (***Figure*** 3-7).



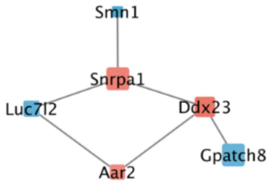
A.



B.

RefSeq	gene_name	pivalue	AveExpr	seqname	strand
NM_001252629	**Smn1**	3.109094324	4.039007632	chr13	+
ENSMUST00000140843	**Stard9**	2.983076342	4.436676284	chr2	+
ENSMUST00000003450	**Ddx23**	2.618097824	7.604374284	chr15	-

C.



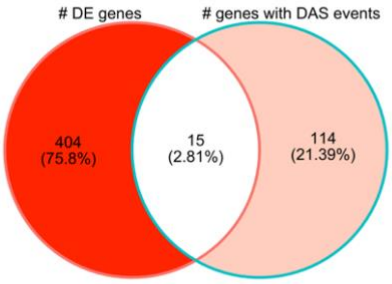
D.

Figure 3-7: Pip6a-PMO-treatment corrects >99% of the early symptomatic SMA gene expression dysregulation in SMA Taiwanese mouse pups.

(A) *Venn diagram of the intersection (in red) between DE genes upregulated in SMA (vs. WT animals) at PND7 (purple) and DE genes upregulated between Pip6a-PMO*$_{ISS-N1}$ *(vs. WT animals) at PND7 (orange).* ***(B)*** *Venn diagram of the intersection (in blue) between DE genes downregulated in SMA (vs. WT animals) at PND7 (purple) and DE genes downregulated between Pip6a-PMO*$_{ISS-N1}$ *(vs. WT animals) at PND7 (orange).* ***(C)*** *Table summarising the name, π-value, chromosomal location, and average expression of the three genes identified in the intersections of the Venn diagrams.* ***(D)*** *Protein interaction network of DE genes based on the STRING online database. Node colour indicates upregulation (in red) or downregulation (in blue) when compared to WT PND7 animals.*

Following treatment with Pip6a-PMO, three genes are left differentially expressed in Pip6a-PMO treated SMA pups (***Table*** 3-3).

Table 3-3: Summary of expression and splicing changes in LCM-MNs of SMA pups with and without Pip6a-PMO treatment at PND7

		SMA vs. WT	
		PND7 SMA vs. WT	PND7 Pip6a-PMO vs. WT
# unique genes with ≥ 1 DE transcript	π-val > 1.3	419	3
# DE transcripts	π-val > 1.3	552	3
# genes with transcripts with ≥ 1 AS event	SI>|0.5| & p.val <0.05	129	937
# AS events	SI>|0.5| & p.val <0.05	144	1077

DE: differentially expressed. AS: alternative splicing. SI: splicing index. The Venn diagram summarises the intersections between DE genes and genes with DAS events, considering DE genes between the SMA vs. WT groups at PND7, in LCM-MN.

I therefore proceeded to compare the 'SMA vs. WT' list of DE genes to the 'Pip6a-PMO-treated vs. WT' list of DE genes. I identified only 3 transcripts left with alternative splicing changes: one upregulated transcript, in *Ddx23*, and one downregulated transcript, in *Stard9* (***Figure*** 3-6***,*** ***Figure*** 3-7). *Smn* is still described as downregulated in SMA pups because the microarray only bears probes for the murine *Smn* gene and does not have probes for the human *SMN2* transgene borne by the Taiwanese mouse model.

I then considered the effect of Pip6a-PMO on splicing defects observed in SMA samples. An exon is considered differentially alternatively spliced (DAS) if |SI| > 0.5 and p.value < 5%. Most transcriptome related changes at PND7 are expression-related, not splicing related: of the 419 DE genes at early symptomatic changes in SMA, only 15 (2.81%) also present at least one alternative splicing event (***Table*** 3-3).

Interestingly, Pip6a-PMO treated SMA pups still exhibit 1077 splicing events found in 937 unique transcripts (***Table*** 3-3***,*** ***Figure*** 3-8). Those are splicing events both naturally present in SMA and splicing events that may be induced by treatment.

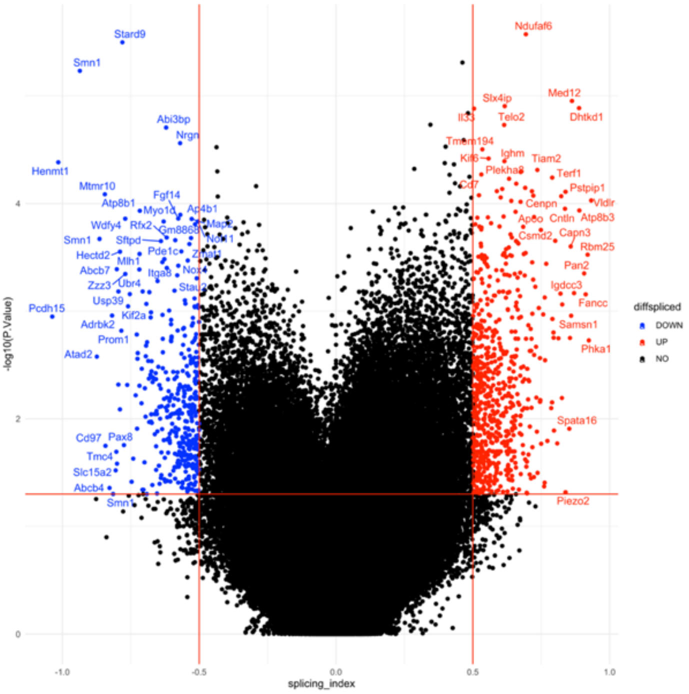


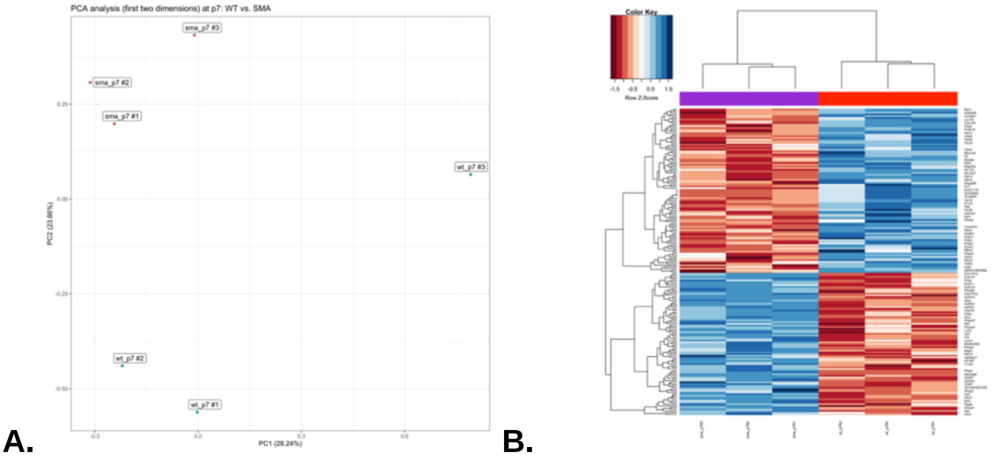
Figure 3-8: Alternative splicing changes in SMA pups following treatment with Pip6a-PMO.

Volcano plot for differential alternative splicing. The changes of alternative splicing in the Pip6a-PMO$_{ISS\text{-}N1}$ *group are quantified against levels in the WT PND7 group. X-axis: splicing index (SI). Y-axis: statistical significance (pval) on the logarithmic scale. In red: genes with a statistically significant upregulation of splicing in treated group. In blue: genes with a statistically significant downregulation of splicing in treated group. In black: genes with no significant difference in splicing between treated and control group. Red lines: defined thresholds for significance, here:* $|SI| > 0.5$ *and p.value* $< 5\%$*.*

3.3.5. Expression and splicing effects at the early symptomatic stage are varied and of small amplitude

The study of the correction of expression and splicing changes following Pip6a-PMO treatment led us to seek to understand the scope of transcriptomic changes early in SMA neurodegeneration – a question that remains widely discussed – by investigating the untreated PND7 samples.

SMA and WT pups segregate very clearly in two clusters both using PCA and HCL (***Figure*** 3-9). In this analysis, I conducted the normalisation at the gene level to focus on expression changes that could be further followed-up by protein level analyses.

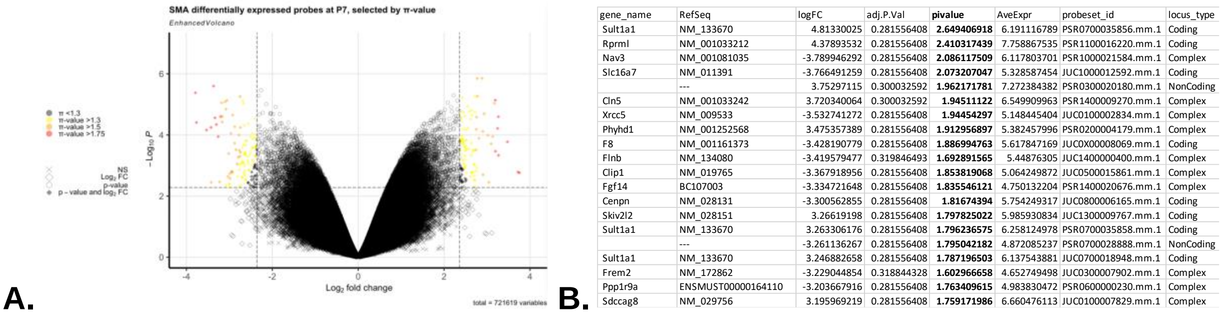


C.	PC1	PC2	PC3	PC4	PC5	PC6
Proportion of Variance	0.2824	0.2386	0.1747	0.1617	0.1427	0.00E+00
Cumulative Proportion	0.2824	0.5209	0.6956	0.8573	1	1.00E+00

Figure 3-9: Clustering of differentially expressed probes (π-val >1.3) at early symptomatic stages of SMA (PND7)

Samples are identified by their colour: WT_PND7 (red) and SMA_PND7 (purple). Each dot represents an individual sample. ***(A)*** *PCA plot of the individual MN samples using whole-transcript gene expression profiling. X-axis: PC1 (28.24%). Y-axis: PC2 (23.86%). Dimensionality reduction shows a clear segregation of samples by genotype (SMA vs. WT) even at early stages of disease progression.* ***(B)*** *Microarray heatmap and clustering based on the hierarchical clustering of the top DE genes. At PND7, there already exist clear differences of gene expression between SMA and WT LCM-MN. Red: down-regulated genes. Blue: up-regulated genes.* ***(C)*** *Proportion of variance and cumulative proportion of the variance explained by the principal components derived from the PCA of samples at PND7. The first two dimensions PC1 and PC2 are the x and y axes of* ***(A).***

At PND7, following microarray gene level normalisation, 183 probes are DE expressed between SMA and WT pups, in 140 unique genes (***Figure*** 3-10). 95.6% of those probes are in coding and alternatively spliced exons. Most strongly dysregulated genes include *Sult1a1*, *Rpml1*, *Nav3*, and *Slc16a7* (***Figure*** 3-10).

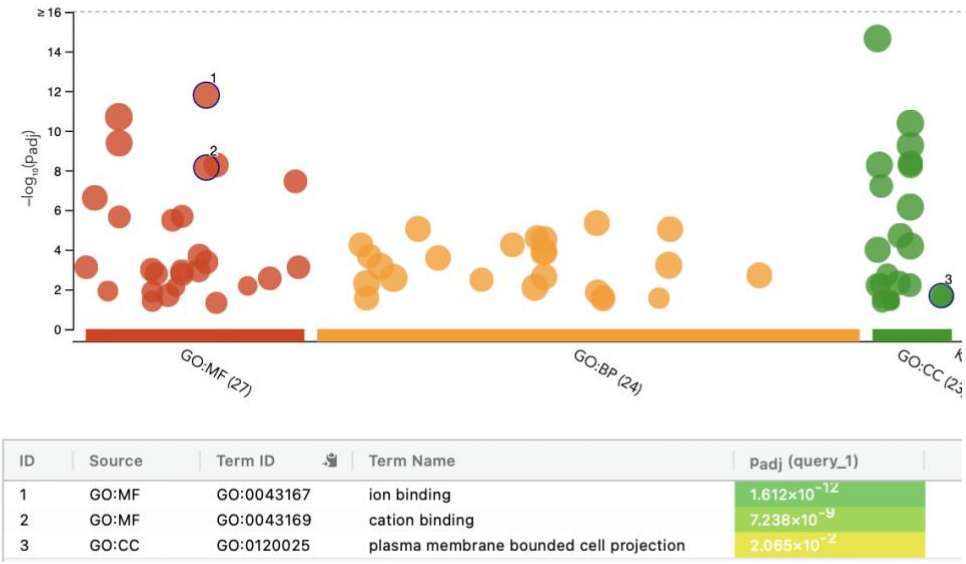


gene_name	RefSeq	logFC	adj.P.Val	**pivalue**	AveExpr	probeset_id	locus_type
Sult1a1	NM_133670	4.81330025	0.281556408	**2.649406918**	6.191116789	PSR0700035856.mm.1	Coding
Rprml	NM_001033212	4.37893532	0.281556408	**2.410317439**	7.758867535	PSR1100016220.mm.1	Coding
Nav3	NM_001081035	-3.789946292	0.281556408	**2.086117509**	6.117803701	PSR1000021584.mm.1	Complex
Slc16a7	NM_011391	-3.766491259	0.281556408	**2.073207047**	5.328587454	JUC1000012592.mm.1	Coding
	---	3.75297115	0.300032592	**1.962171781**	7.272384382	PSR0300020180.mm.1	NonCoding
Cln5	NM_001033242	3.720340064	0.300032592	**1.94511122**	6.549909963	PSR1400009270.mm.1	Complex
Xrcc5	NM_009533	-3.532741272	0.281556408	**1.94454297**	5.148445404	JUC0100002834.mm.1	Complex
Phyhd1	NM_001252568	3.475357389	0.281556408	**1.912956897**	5.382457996	PSR0200004179.mm.1	Complex
F8	NM_001161373	-3.428190779	0.281556408	**1.886994763**	5.617847169	JUC0X00008069.mm.1	Coding
Flnb	NM_134080	-3.419579477	0.319846493	**1.692891565**	5.44876305	JUC1400000400.mm.1	Complex
Clip1	NM_019765	-3.367918956	0.281556408	**1.853819068**	5.064249872	JUC0500015861.mm.1	Complex
Fgf14	BC107003	-3.334721648	0.281556408	**1.835546121**	4.750132204	PSR1400020676.mm.1	Complex
Cenpn	NM_028131	-3.300562855	0.281556408	**1.81674394**	5.754249317	JUC0800006165.mm.1	Coding
Skiv2l2	NM_028151	3.26619198	0.281556408	**1.797825022**	5.985930834	JUC1300009767.mm.1	Coding
Sult1a1	NM_133670	3.263306176	0.281556408	**1.796236575**	6.258124978	PSR0700035858.mm.1	Coding
	---	-3.261136267	0.281556408	**1.795042182**	4.872085237	PSR0700028888.mm.1	NonCoding
Sult1a1	NM_133670	3.246882658	0.281556408	**1.787196503**	6.137543881	JUC0700018948.mm.1	Coding
Frem2	NM_172862	-3.229044854	0.318844328	**1.602966658**	4.652749498	JUC0300007902.mm.1	Complex
Ppp1r9a	ENSMUST00000164110	-3.203667916	0.281556408	**1.763409615**	4.983830472	PSR0600000230.mm.1	Complex
Sdccag8	NM_029756	3.195969219	0.281556408	**1.759171986**	6.660476113	JUC0100007829.mm.1	Complex

Figure 3-10: Top 20 differentially expressed probes between SMA and WT pups at early symptomatic stage

(A) *Volcano plot for DE probes between SMA and WT PND7 pups. The changes of expression in the SMA group are quantified against levels in the WT group. X-axis: logFC of probe expression. Y-axis: statistical significance (π-value) on the logarithmic scale. Genes are ranked by their π-value: red (π-val>1.75), orange (π-val>1.5), yellow (π-val>1.3). Grey dots symbolise genes with no significant difference in expression between SMA and WT groups. Dotted lines: defined thresholds for significance, with -log10(π-value) > 2.72 in order to exclude genes with π-val<1.3.* ***(B)*** *Table of the names, average expression, logFC, and locus type of the 20 top DE probes between SMA and WT LCM-MN at PND7, ranking genes by decreasing π-value.*

Gene ontology analysis shows an enrichment of ion channels (ion and cation channels) (***Figure*** 3-11), confirmed using the BinGO plugin in the network analysis (***Supplementary Figure*** 7-4).



ID	Source	Term ID	Term Name	Padj (query_1)
1	GO:MF	GO:0043167	ion binding	1.612×10^{-12}
2	GO:MF	GO:0043169	cation binding	7.238×10^{-9}
3	GO:CC	GO:0120025	plasma membrane bounded cell projection	2.065×10^{-2}

Figure 3-11: Gene Ontology analysis of the DE probes in SMA at early symptomatic stage

Manhattan plot of the enrichment of gene ontologies (GO) amongst the 140 unique genes. On the x-axis, GO are grouped and coloured according to: molecular function (MF), biological process (BP), and cellular compartment (CC). On the y-axis, adjusted pval (in log10 scale). Selected top 3 GO terms are presented.

I then selected a 'high interest genes' subset amongst the 140 unique DE genes, by cross-referencing it with orthologs from a DE gene list obtained from an unpublished RNA-seq study in SMA iPSC cells (Dr. Jakub Scaber, Prof. Talbot laboratory, University of Oxford). Four of the 140 unique genes DE in murine LCM-MN at p7 are also DE in human SMA iPSC: *ATP8B1*, *ITPR3*, *HTRA1*, and *SMN1* (***Table*** 3-4).

Table 3-4: Four of the genes DE in murine LCM-MN at p7 are also DE in human SMA iPCs

	symbol	**description**	**Base Mean**	**log2Fold Change**	**padj**
ENSG00000081923	**ATP8B1**	ATPase phospholipid transporting 8B1	98.89	-0.64	0.01
ENSG00000096433	**ITPR3**	inositol 1,4,5-trisphosphate receptor type 3	307.44	-1.01	0.00
ENSG00000172062	**SMN1**	survival of motor neuron 1, telomeric	1090.61	-2.95	0.00
ENSG00000166033	**HTRA1**	HtrA serine peptidase 1	1544.50	0.76	0.00

ITPR3 is the gene with the highest fold change in human SMA iPSC (down-regulated 1.01-times when compared to WT SMA iPSCs) and is down-regulated 2.72 times in murine LCM-MN (π-value -1.46). The down-regulated probe in the microarray is located on the junction of exon 49-50 of *Itpr3* (nt 7131-7162 of NM_080553.3). This portion codes for a helical (transmembrane) domain (aa 2368-2388) in the C-terminal, channel domain of *Itpr3*. *Itpr3* is an endoplasmic reticulum membrane protein involved in the release of intracellular calcium into the cytosol and functionally interacts with other calcium-related proteins, including Ryr3 and Cacna1d (***Figure*** 3-12).

I therefore conjecture that *Itpr3* splicing defect in its transmembrane domain affects the function of *Itpr3*. Validation of the impact of this splicing defect at the protein level as well further study of the down-stream effects of the downregulation of *Itpr3* on IP-3, ATP, and calcium-binding remain for further functional investigation.

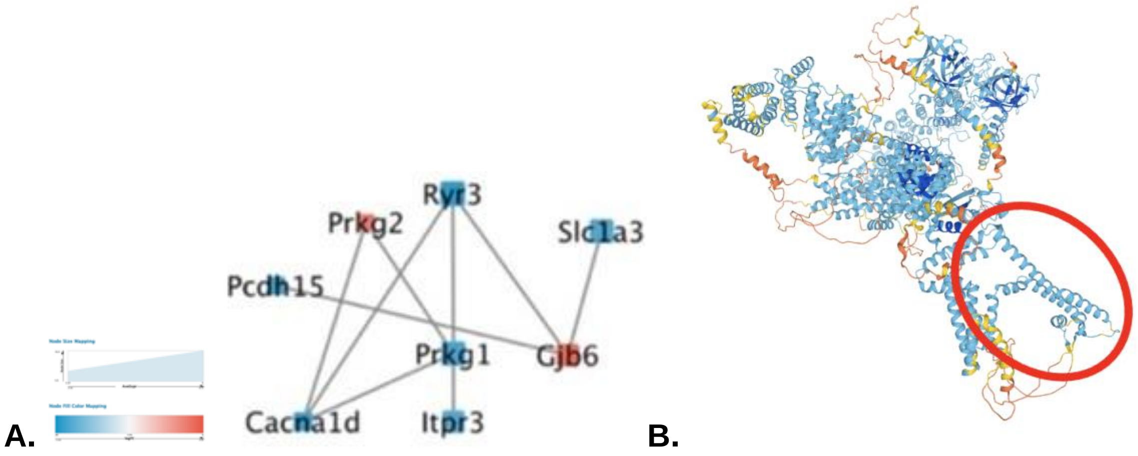
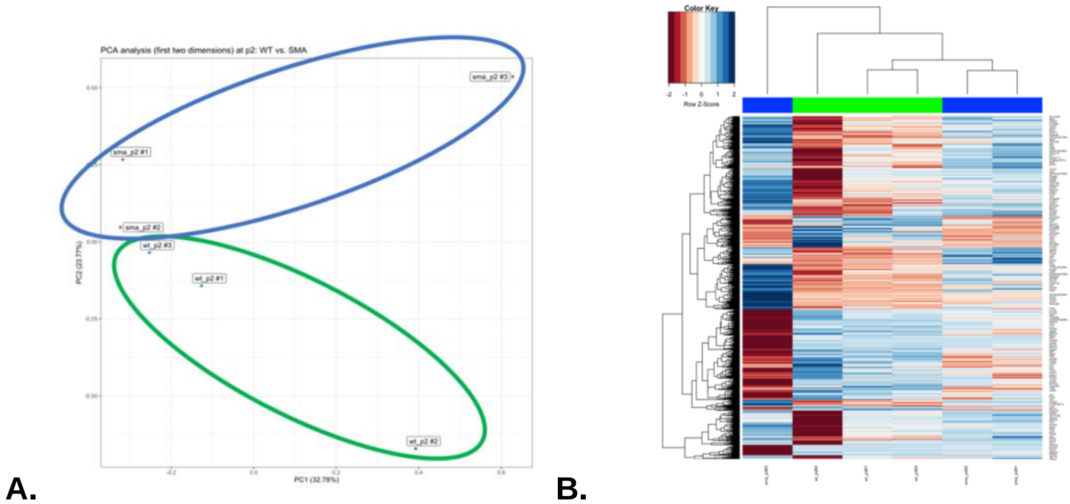


Figure 3-12: Itpr3 is a transmembrane protein involved in calcium transport with splicing defects in its transmembrane domain

(A) *Subset of the protein interaction network of top DE genes (π-val >1.3) between SMA and WT pups at PND7. Node size is proportional to the average expression in the sample. Node colour indicates upregulation (in red) or downregulation (in blue) when compared to WT PND7 animals.* ***(B)*** *Alpha-fold prediction of the structure of Itpr3 (PDB AF_AFP70227F1). The helical (transmembrane) domain (aa 2368-2388) in the C-terminal, channel domain of Itpr3 is circled in red.*

3.3.6. Expression and splicing alterations at the presymptomatic stage

I then proceeded to analyze whether any expression or splicing changes were reported at the early presymptomatic stage (SMA PND2). I found that SMA and WT pups cluster together on the second principal component (***Figure*** 3-13) and are not distinguishable when clustering samples hierarchically (***Figure*** 3-13). When conducting DE analyses, I find that the maximum π-value = 0.03, is well below the 1.3 threshold I have used for the comparisons at PND7.



C.	PC1	PC2	PC3	PC4	PC5	PC6
Proportion of Variance	0.3278	0.2377	0.1743	0.1348	0.1254	0.00E+00
Cumulative Proportion	0.3278	0.5655	0.7398	0.8746	1	1.00E+00

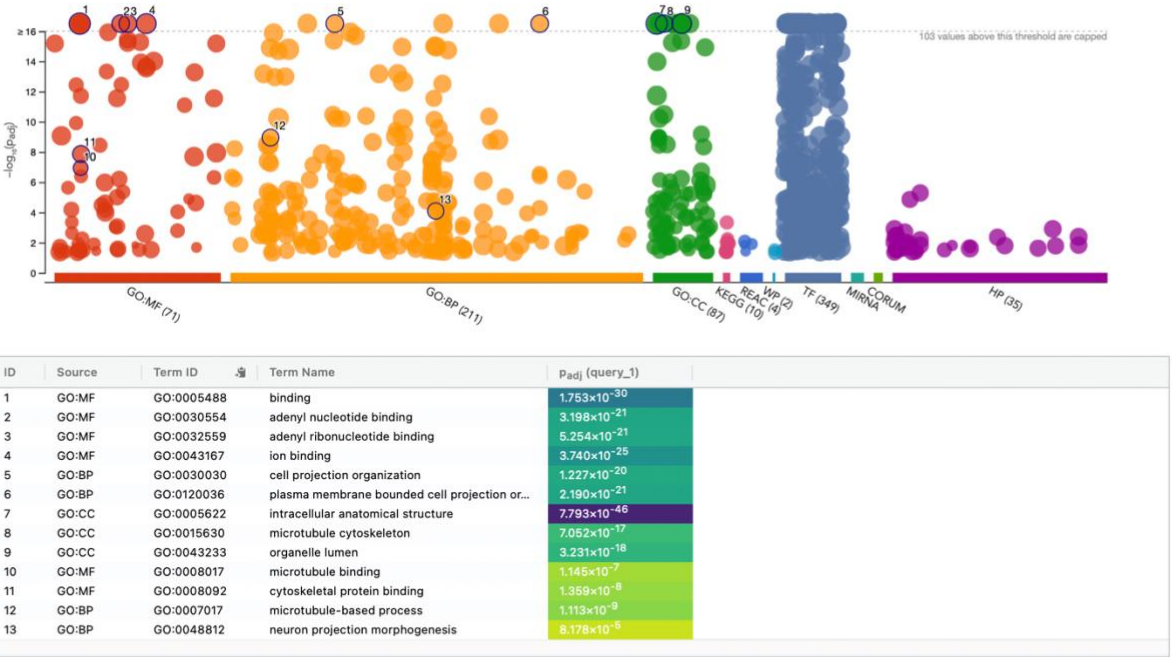
Figure 3-13: Clustering of presymptomatic (PND2) LCM-MN SMA and WT samples

Samples are identified by their colour: WT_PND2 (green) and SMA_PND2 (blue). Each dot represents an individual sample. ***(A)*** *PCA plot of the individual MN samples using whole-transcript gene expression profiling. X-axis: PC1 (32.78%). Y-axis: PC2 (23.77%). Dimensionality reduction fails to show segregation of samples by genotype (SMA vs. WT) at the presymptomatic stage and reveals some intra-group heterogeneity (WT_PND2#2, SMA_PND2#3).* ***(B)*** *Microarray heatmap and clustering based on the hierarchical clustering of the top DE genes. At PND2, there are no clear differences of gene expression between SMA and WT LCM-MN. Red: down-regulated genes. Blue: up-regulated genes.* ***(C)*** *Proportion of variance and cumulative proportion of the variance explained by the principal components derived from the PCA of samples at PND2. The first two dimensions PC1 and PC2 are the x and y axes of* ***(A).***

When calculating the splicing indeces for the probes, I identify the presence of alternative splicing changes at the presymptomatic stage SMA p2 vs. WT p2: 1,454 probes show changes in splicing greater than 50% (SI > 0.5) in 968 unique genes; and 826 probes with SI < -0.5 in 621 unique genes (***Table*** 3-5). The splicing changes are enriched in genes related to neuron projection organisation, including ion binding, cell projection organisation, and microtubule organisation (***Figure*** 3-14).

Table 3-5: Number of motor neuron splicing defects at pre-symptomatic stages

Motor neurons	PND2 SMA vs. WT
# unique genes with >= 1 DE transcript (π-val > 1.3)	0
# DE transcripts (π-val > 1.3)	0
# genes with transcripts with >=1 AS event (SI>|0.5| & p.val <0.05)	1085
# AS events (SI>|0.5| & p.val <0.05)	1331



ID	Source	Term ID	Term Name	P_{adj} (query_1)
1	GO:MF	GO:0005488	binding	1.753×10^{-30}
2	GO:MF	GO:0030554	adenyl nucleotide binding	3.198×10^{-21}
3	GO:MF	GO:0032559	adenyl ribonucleotide binding	5.254×10^{-21}
4	GO:MF	GO:0043167	ion binding	3.740×10^{-25}
5	GO:BP	GO:0030030	cell projection organization	1.227×10^{-20}
6	GO:BP	GO:0120036	plasma membrane bounded cell projection or...	2.190×10^{-21}
7	GO:CC	GO:0005622	intracellular anatomical structure	7.793×10^{-46}
8	GO:CC	GO:0015630	microtubule cytoskeleton	7.052×10^{-17}
9	GO:CC	GO:0043233	organelle lumen	3.231×10^{-18}
10	GO:MF	GO:0008017	microtubule binding	1.145×10^{-7}
11	GO:MF	GO:0008092	cytoskeletal protein binding	1.359×10^{-8}
12	GO:BP	GO:0007017	microtubule-based process	1.113×10^{-9}
13	GO:BP	GO:0048812	neuron projection morphogenesis	8.178×10^{-5}

Figure 3-14: Splicing defects at early pre-symptomatic changes reveal that most genes affected are related to neuronal processes

Manhattan plot of the enrichment of gene ontologies (GO) amongst the 1,454 probes DE in SMA at PND2. On the x-axis, GO are grouped and coloured according to: molecular function (MF), biological process (BP), cellular compartment (CC), KEGG pathway, Reactome (REAC), WikiPathways (WP), transcription factor (RF), miRNA (MIRNA), CORUM protein complexes, and Human Phenotype Ontology (HP). On the y-axis, adjusted pval (in log10 scale). Selected top 13 GO terms are presented.

3.3.7. Transcriptomic signature of MN during development

Considering the number of DE genes with splicing defects at an early, pre-symptomatic stage, I proceeded to compare how normal developmental processes at the critical period before and after symptom onset (between PND2 and PND7) change in SMA. After confirming that SMA PND2 and SMA PND7 samples (and resp. WT PND2 and WT PND7) do not cluster together (***Supplementary Figure*** 7-5), I compared the toptable lists of genes DE between SMA (resp., WT) PND2 and PND7 and selected genes with a π-val>1.3. During normal (WT) development on MN, 43 genes are DE. In SMA conditions, 269 genes are DE between PND2 and PND7 (***Figure*** 3-15***, Supplementary Table*** 7-6).

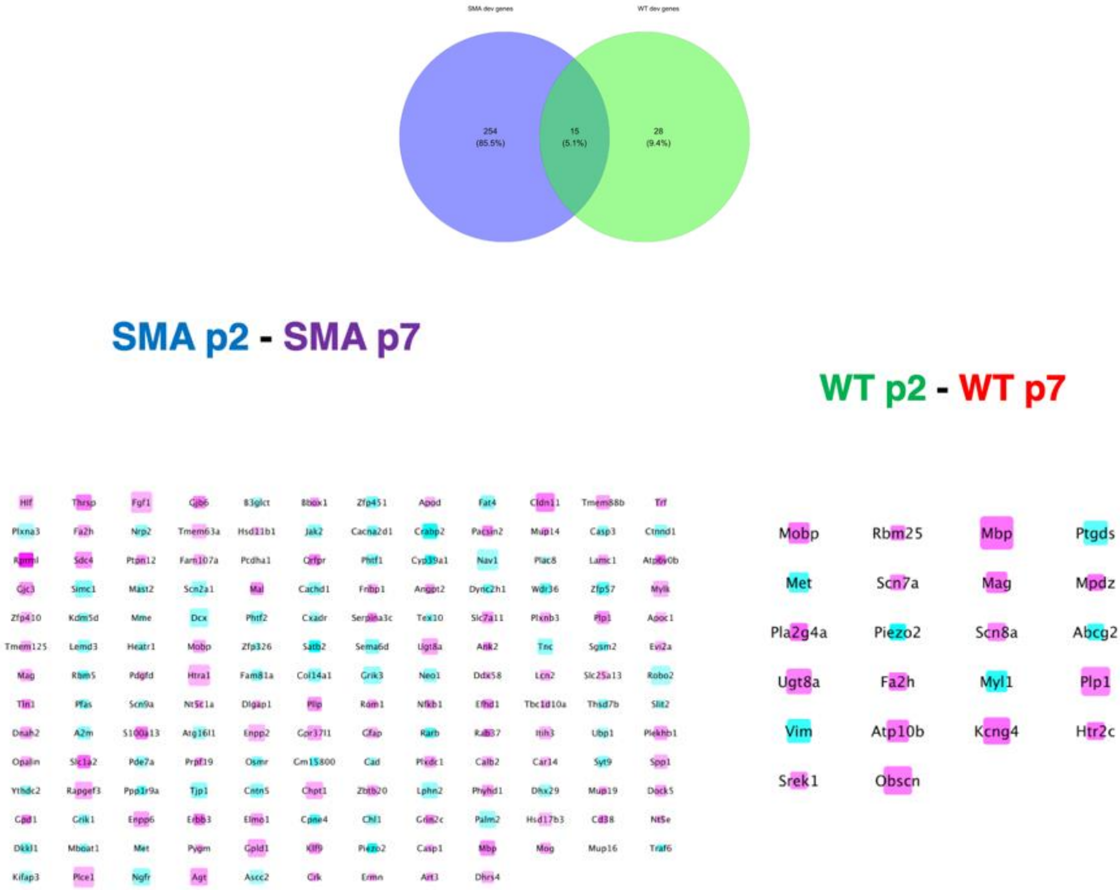


Figure 3-15: Venn Diagram of DE genes (π-value>1.3) during early post-natal development in SMA vs. WT animals.

The Venn diagram (top) presents the 15 genes at intersection between the 254 genes DE during SMA MN development (blue circle), and the 43 genes DE during WT MN development (green circle). The names of those genes is listed under, with node size proportional to their expression level, and colour representing the direction of their differential expression at PND7 vs. PN2 (cyan: downregulated, magenta: upregulated).

Amongst the 269 genes DE between SMA PND2 and PND7 (***Figure*** 3-15), the vast majority (254, i.e. 85.5%) are unique to SMA pathogenic development. Gene ontology enrichment analyses reveal the SMA-specific upregulation of odorant-binding proteins, including *Mup1, 2, 7, 10, 13, 14, 15* & *Olfr152*. Membrane depolarisation processes and the regulation of long-term synaptic potentiation are also dysregulated (***Figure*** 3-16).

There are 28 genes unique to WT PND2 to WT PND7 and not found to be DE in SMA. Those genes include the voltage-gated sodium channel complex (GO:0001518) *Scn7a* and *Scn8a*, as well genes related to 'retinoid binding' (GO:0005501). 'Phosphatidylinositol phosphate binding' (GO:0035091) is also an enriched process amongst dysregulated genes.

A minority of gene expression changes between PND2 and PND7 is unaffected in SMA. Those 15 genes mostly include myelination-related genes and galactosylceramide biosynthetic processes (***Figure*** 3-17).

ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0005550	pheromone binding	1.023×10^{-6}
2	KEGG	KEGG:04151	PI3K-Akt signaling pathway	2.730×10^{-2}
3	GO:MF	GO:0046872	metal ion binding	4.687×10^{-3}
4	GO:MF	GO:0043169	cation binding	2.244×10^{-3}
5	GO:BP	GO:0048856	anatomical structure development	4.670×10^{-13}

Figure 3-16: Gene Ontology enrichment in DE genes during the PND2-PND7 SMA symptom progression period.

Manhattan plot of the enrichment of gene ontologies (GO) amongst the 269 genes DE during early post-natal development in SMA processes. On the x-axis, GO are grouped and coloured according to: molecular function (MF), biological process (BP), cellular compartment (CC), KEGG pathway, and Reactome (REAC). On the y-axis, adjusted pval (in log10 scale). Selected top 5 GO terms are presented.

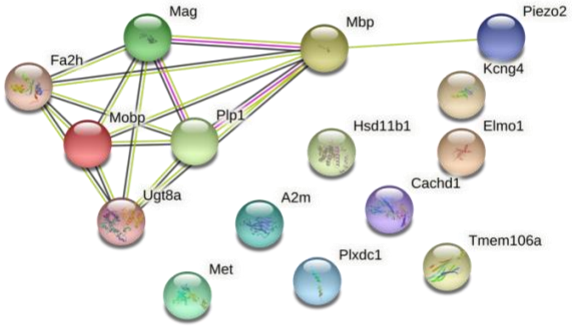


Figure 3-17: STRING network of DE genes from PND2 till PND7 in both WT and SMA animals

Protein interaction network of the 15 genes DE similarly in SMA and WT pups during early post-natal MN development (PND2 to PND7), based on the STRING online database.

3.4. Discussion

In this study, I generated a novel transcriptomic dataset from laser-captured microdissected MNs from control and SMA mice. The choice of timepoints in this study (presymptomatic, PND2, and early symptomatic, PND7) was driven by a drug-discovery goal aiming to identify transcripts, genes, and/or pathways affected at early stages of symptoms and not corrected by early splice-switching therapy.

3.4.1. Value of the initial data set

There are multiple points of value to this dataset despite its power-related shortcomings (see below). First, whilst being the second report of *ISS-N1*-targeting ASO-treated murine population [139], this is the first study focusing on MNs from treated animals (rather than whole spinal cord). It is also the first report of transcriptomics data in neurons following intravenous administration of a splice-switching therapy. As such, it can be further used as a comparison dataset, e.g. against the only other report of LCM-MN in SMA mice, which was published during the course of this research project [152].

In addition, our study is extracted from the same individual animals as the previous report in muscle [313]. As such, it provides a unique examination point for both comparing disease progression in two SMA relevant tissues, and to compare the efficacy of the same treatment. Notably, this clarifies that the Pip6a-PMO treated animal showing no correction ('pip6a#1') is not due to a defect in the I.V. administration (there is extreme homogeneity in the transcriptomic landscapes of Pip6a-PMO treated muscle samples), but rather due to a potential issue either at the neurovascular interface or within the brain.

Finally, by providing datasets from animals that were treated neonatally (thus mimicking an ideal clinical scenario of neonatal diagnosis and treatment onset), this dataset provides motor neuron-centric presentation of the transcriptomic landscape on an SMA genetic background following SMN-centric treatment. This study therefore inserts itself in the broader exploration of new (molecular) phenotypes deriving from early treatment modalities.

3.4.1. Limitations of the initial data set

The primary limitations of this dataset relate to (1) species-specific effects, (2) power limitations, and (3) considerations of neurovascular development timelines.

Taken on its own, the differential expression and splicing analyses conducted reflect changes in mice and are not necessarily informative of changes in human patients. For example, *Chodl* splicing (a gene involved in axonogenesis in zebrafish [352]) is significantly altered in mouse but not in human samples [146].

Perhaps most critical to this study are the power limitations. Experiment design was limited at the onset of this study by financial limitations and the high number of groups necessary to answer the experimental questions, and as such, compromises were made, notably regarding (1) choice of control groups, (2) sex stratification, and (3) power calculations.

Firstly, previous studies by the same laboratory [338] had already compared several regimens of Pip6a-PMO against PMO-only treatment. The lack of a control group of mice treated with nusinersen and focus on functional molecular consequences of Pip6a-PMO only is one of the compromises led by the financial limitations at the onset of the study. Should the same experiment be conducted today, a proper nusinersen (or at least an *ISS-N1* targeting PMO) control group would clarify the contribution of the Pip6a moiety to CNS penetration and neuronal effect.

Secondly, whilst litters have been randomised between groups, further effort should have been conducted to avoid the excessive representation of either males or females within an already small number of animals per group (as seen in this study for the all-male 'SMA PND2' and 'Pip6a-PMO-treated' groups).

Finally, the absolute changes investigated in this study have been quite small, with a gene considered DE if π-value > 1.3 (which translates to a 2-fold change at $\alpha=5\%$ [347]) and DAS if $|SI| > 0.5$ and p.value $< 5\%$. *A priori* power analysis calculations [353] with G*Power3 [354] for a linear multiple regression model (such as deployed by the `limma` package), two-tailed test, with $\alpha=5\%$, a desired power of 80%, an effect size of 2, and an attrition rate in SMA pups of 10%, suggest a sample size no lesser than n=9 per group to identify changes in gene expression. (Of note, should this study be conducted today, RNA-seq methodologies (rather than microarrays) would be conducted, which affects the statistical assumptions and warrants using dedicated power analysis calculation methodologies [355,356]).

One important issue relates to species-dependent neurovascular unit maturation timelines. It is very clear that the BBB is completely formed at birth (reviewed in [357]): whilst developing cerebral vessels (particularly, capillaries [358,359]) are not fully adult-like, and whilst their increased susceptibility to pro-inflammatory signals (including lipopolysaccharides) during early post-natal days (evidenced in rat [360,361], mice [362], and suggested in human as reviewed in [357]) can affect the BBB permeability to proteins, the tight junctions and active transporters are functional and effective, and allow for a well-controlled internal cerebral homeostasis since early stages in development [363]. In parallel, early post-natal development in marsupials and rodents (both rat and mouse) is accompanied by changes in the morphology, size, and molecular expression of choroid plexuses [364–367], and as such, the choroid plexus is thought to be the primary locus of blood-to-brain (via CSF) macromolecular transfer during foetal and early postnatal development [364,365]. Consequently, the blood-to-brain transfers of macromolecules at the time of our study (PND0 and PND2) occurs within a system with peculiar, time-limited characteristics. This suggests that the delivery of the Pip6a-PMO may occur independently of any Pip6a-moiety delivery ability. However, this discussion does not affect the interpretation of the data, since the PMO itself is responsible for the transcriptional changes. Finally, and importantly, the cross-comparison of developmental benchmarks between humans and mice suggests that rodent brain development between PND1 and PND5 corresponds roughly to second trimester gestation in humans ([368], cited in [369]). As such, to gauge the relevance of the molecular correction observed in murine SMA MN to a neonatal intervention in human SMA patients, it would be interesting to repeat this set of experiments at later stages, e.g. with injections at PND7-PND10 [368].

3.4.2. Concordance of identified DE transcripts with previously reported transcriptomic studies

In this study of time- and treatment-dependent transcriptomic changes in an SMA mouse model at pre- and early- symptomatic stages in LCM-MN, I found numerous splice-switching changes at PND2 (1,085 genes with transcripts containing at least one significant splicing change) but no expression changes at the early symptomatic stage. Whilst the lack of detection of pre-symptomatic expression changes could be attributed to the lack of power in our study (n=3 per group, with mixed sexes within the groups), our results at the splicing level agree with previously reported study at similar disease progression timepoints, notably studies reporting early changes in synaptogenesis genes [141] (***Table*** 3-6).

Similarly, at the early-symptomatic change, I found 419 differentially expressed transcripts between SMA and WT pups, with 184 unique genes differentially expressed. Interestingly, only 4 genes have overlap with human orthologs from SMA iPSC-derived MNs: *Smn1*, *Htra1*, *Atp8b1*, and *Itpr3*. *ITPR3* is also amongst the 'plasma membrane' protein-coding genes identified in the meta-analysis of human SMA RNAseq datasets in Chapter 2. Most interestingly, *ITPR3* is directly downregulated by miR-1248 [370], a microRNA with putative bindings sites both within the 5' UTR and within exon 7 of *SMN* mRNA [371]. Whether (and how) the downregulation of *Itpr3* expression is directly linked to a dysregulation in miR-1248 binding on SMN transcripts remains to be further explored.

I have presented how a splicing change of *Itpr3* can be an avenue for further neuropathological exploration. Similarly, the other genes identified in this study can provide new hypotheses for the understanding of the molecular biology of motor neurons. For example, *Htra1* is a serine protease with a variety of targets, acting intracellularly on TGF-beta signaling. By degrading TSC2, it activates its downstream targets [372] that in neurons are involved in the control of neuronal polarity [373]. The *in vitro* downregulation of *Htra1* also leads to the inhibition of neurite extension [374]. Further overlap studies with lists of DE genes from the meta-analysis approach (Chapter 2) may consolidate and/or extend the subset of candidate genes hereby reported.

Table 3-6: Manual review of all published transcriptomics mouse studies in SMA

Study	Analysis type	Tissue Disease model	Pre-symptomatic (vs. time-point WT control)	Early symptomatic (vs. time-point WT control)	Notes
152	RNA-seq	LCM-MN Δ7 mice	PND2 Expression 17 changes	PND5 Expression 49 changes	Study also at PND10 Study also from MN in cranial and facial nuclei
139	RNA-seq	Whole SC Taiwanese mice	PND1 Expression 23 up 10 down	PND5 Expression 1429 up 1611 down	–
144	RNA-seq	LCM-MN Smn$^{2B/-}$ mice	PND10 Expression 623 up 654 down	–	Study distinguishes vulnerable (abdominal/thoracic) and resistant (cranial) MN
262	3' & exon microarray	Whole SC Smn$^{-/-}$;SMN2$^{+/0}$ mice	PND1 Expression 5 up 20 down	PND5 Expression 14 up 39 down	Also study in *Smn*$^{-/-}$; *SMN2*$^{+/+}$ mice with induced Smn knockdown
141	RNA-seq	LCM-MN Δ7 mice	PND1 Expression 138 up 110 down	–	Study also on white matter
261	Exon microarray	Whole SC Smn$^{-/-}$;SMN2 mice Δ7 mice	PND1 Expression 1 up 2 down 5 up 6 down	PND5 Expression 0 up 39 down	–
146	Exon microarray	Whole SC Δ7 mice	PND1 Splicing 16 changes	PND7 Splicing 5 changes	Study also at PND13
135	Exon microarray	Brain Δ7 mice	–	PND6 Diminished snRNP assembly	Also study at PND11 in brain and SC

In summary, whilst this present study has a suboptimal statistical power, it supports previously published data reporting numerous splicing changes before the onset of symptoms in SMA to be responsible for changes in neuron projection morphogenesis that impact the longer development of motor neurons. This is an additional argument to support early, pre-symptomatic treatment of SMA patients.

3.4.3. *Ddx23* and *Stard9* as new potential targets for combinatorial therapy

This novel microarray was used to identify any genes not corrected after neonatal treatment by Pip6a-PMO. Despite its low number of animals per group is small (n=3), it provides insight into potential targets for combinatorial therapy with SMN-centered treatment.

Contrary to the muscle microarray data of [313], the LCM-MN samples present significant heterogeneity between samples of a same group, which reduces the statistical power of the differential expression analyses. I chose the π-value metric to allow for the extraction and ranking of information based both on fold change and statistical significance [347]. I chose the *ad-hoc* threshold of π-value=1.3 because it corresponds to a p.value of 0.05 of a protein with a two-fold expression change [347]. With those parameters of analysis, I found that Pip6a-PMO treatment restores all expression changes in PND7 SMA pups except for two genes: the RNA helicase *Ddx23* (DEAD-Box Helicase 23), which is still up-regulated after Pip6a-PMO treatment, and the kinesin-like protein *Stard9,* still down-regulated after Pip6a-PMO treatment. Of these, *Ddx23* is a particularly interesting candidate as it is involved in R-loop resolution. R-loops have been involved in SMA pathogenesis [375] via an intron-retention mechanism in situations of low SMN function [140]. *Ddx23* is involved in RNA splicing and genome stability and its mutations have recently been linked to neurodevelopmental disorders with a varied clinical presentation including hypotonia and autistic-like features [376]. The STARD9 (also known as Kif16a) kinesin is expressed in the central nervous system [377] where it associates with mitotic microtubules and regulates spindle pole assembly [378]. Its knockdown induces a strong depolymerisation of the interphase microtubules [379].

The depletion of STARD9 induces spindle assembly defects in human culture cells and has been reported in a human patient with intellectual disability [380].

The functional relevance of the non-correction of *Ddx23* and *Stard9* expression in Pip6a-PMO treated SMA pups is unknown, as is the reason for which those two genes are the only ones not corrected following splice-switching therapy. Future work on this should include the validation of these results in Pip6a-PMO treated SMA iPSCs as well as in samples from nusinersen-treated patients.

3.4.4. Pip6a-PMO efficacy, delivery, and side-effects

The comparison of these results with the previously reported time- and treatment-dependent transcriptomic changes in SMA mouse muscle [313] raises the question of the untangling of the efficacy of the *ISS-N1* targeting splice-switching PMO moiety from the efficacy of Pip6a as its delivery system [336,381].

Previous studies have demonstrated the pre-clinical efficacy of the Pipa-PMO neonatal treatment [338] without investigating the molecular effects, in particular at the whole-transcriptomic level of the treatment. Whilst SMN levels in spinal cord/MN has not been quantified in this study (partially following the interruption of experimental work by the COVID-19 pandemic), it has been reported that neonatal treatment of this mouse model with 2x10 µg/g Pip6a-PMO increases the expression of SMN in the spinal cord at the mRNA and protein levels [338]. Our study confirms that Pip6a-PMO is overall correcting >99% of the SMA-induced expression changes at early symptomatic changes. It also suggests that the Pip6a-PMO treatment itself induces splicing changes in genes otherwise not affected in SMA. Amongst these 1,077 genes, there is a particular enrichment of genes associated with the GO term ‘microtubule cytoskeleton’ (78 genes), ‘cytoskeleton organisation’ (90 genes) and ‘microtubule cytoskeleton organisation’ (44 genes), which are processes quite ubiquitous to cellular function (both in terms of nuclear organisation, neuronal morphology, and neuronal exocytic functions). Reassuringly, as for risdiplam [382], Pip6a-

PMO does not significantly affect the splicing of critical isoforms involved in cell division and chromosome missegregation (*Foxmb1*) or toxic glycolipid degradation (*Galc*). Known off-target genes of risdiplam do not exhibit significant (|SI| >0.5) splicing changes: neither *Poln, Papd4*, nor *Pdxdc1* have a probe with an |SI| greater than 0.3 (±30% change). It is unclear at this stage what underlines these changes, whether they are consequences of the binding of the Pip6a cationic residues within the cell to splicing factors, and/or whether they are off-site effects of the PMO splicing effects. The mechanism of these splicing changes, as well as their long-term consequences, are one of the avenues of studies opened by this study.

Whilst onasemnogene abeparvovec-xioi (which is administered IV) has shown great results in the clinic, some recent communications suggest the importance of the pipeline of delivery of SMN-targeting treatment. Notably, AAV9-SMN gene replacement therapy targets different disease mechanisms whether administered intravenously or intracerebroventricularly. In particular, whilst the survival, weight, and motor outcomes are similar, there are strong differences in spinal SMN correction pattern, and neuroprotection of MN being achieved to a greater degree for centrally-delivered AAV9-SMN (motor neuron count twice lower in IV-treated than in ICV-treated $Smn^{2B/-}$ mice, neurofilament light chain count a third lower in IV-treated than in ICV-treated animals) [383]. The previously published effect of Pip6a-PMO on SMN spinal cord level expression increase [338] therefore suggests that the correction of SMA expression and splicing changes at PND7 is SMN-dependent and does not depend on peripheral systems from e.g. muscle.

3.4.5. Possible role of Major Urinary Proteins and Olfactory Receptors during neurodegeneration

As described in Chapter 1, during the time of this study, two novel therapies have been approved in SMA – Zolgensma™ and risdiplam, which both present significant advantages of delivery when compared to an *ISS-N1*-targeting ASO. As such, at present, the value of this dataset for SMA is derived from the theoretical contributions it provides to motor neuron biology and to the pathways affected in early SMA neuronal development.

I focused part of our analyses on identifying the processes in early postnatal MN development most affected by SMA. Interestingly, I identified a number of olfactory-related processes (members of the *Mup* family as well as the olfactory receptor *Olfr152)* that have also been reported (and not further investigated) in [261]. Particularly notable is the downregulation of *Olfr1288* at PND1 in *Smn*$^{-/-}$*;SMN2* as well as the downregulation of *Olfr1393/1392* and of *Oflr978* in Δ7 SMA pups [261]. In addition, *Olfr420/636/323*, and *Olfr978* are down-regulated in SMA spinal cords (at PND7 and PND1 respectively) [146]. Olfactory receptor (*Olfr*) genes are the largest gene family in mice [384]. Little, however, is known about their role in MN development.

Major Urinary Proteins (MUPs) is another family of gene over-represented in SMA developing MNs and absent from normal MN development. The major urinary protein (*Mup*) cluster (ortholog to the lipocalin family in humans) is primarily known for its role in chemical communication and pheromone carrier [385]. Interestingly, MUP proteins have been suggested to play a role in glucose metabolism and have been linked to lipid disorders [386]. The mechanism behind these two families' enrichment, whether any of these results correlate to human neurodegenerative pathology, or whether they merely reflect the specialised expression of olfactory systems in mouse for kin recognition, remains to be investigated.

In conclusion, I have demonstrated the efficacy at the whole transcriptome level of an intravenously administered splice-switching potential treatment for spinal muscular atrophy. Whilst the study bears less statistical power than similar studies in the field, it corroborates the pathogenic role of early splicing changes in neuronal process defects, identifies the intriguing dysregulation of *Mup* and *Olfr* transcripts in the critical early postnatal MN development period, and suggests two genes left uncorrected after Pip6a-PMO treatment that warrant further molecular exploration.

In the next chapter, I will continue our transcriptome-centred approach to the identification and exploration of new therapeutic targets in SMA by investigating circular RNAs – a type of RNAs left aside by both RNA-seq studies and exon microarray.

4. Chapter 4: *SMN*-derived circRNAs in health and in SMA

Are* SMN*-derived circular RNAs differentially expressed in SMA?

What methodological gaps persist in the study of circular RNAs in a high isoform context?

What computationally defined properties of* SMN*-derived circRNAs can be exploited for their potential exploitation as a target in SMA?

4.1. Introduction

The two preceding results chapters have focused on the linear transcription products of the *SMN* genes – whether including the poly-(A) enriched and/or ribosomal RNA depleted subsets of the whole transcriptome (Chapter 2), or exclusively the whole exome (Chapter 3). The emphasis on linear transcription products stemmed from Chapter 1, in which I presented multiple linear transcripts generated from the *SMN1* and *SMN2* loci. In this present chapter, I will focus on another family of transcription products from *SMN* genes with radically different biogenesis, expression, and functional properties – namely, *SMN1*- and *SMN2*-derived circRNAs. In order to evaluate their relevance in SMA neurobiology and their potential in drug development, I will explore their expression in SMA cell models and characterise their potential biological function.

4.1.1. What are circRNAs?

Circular RNAs (circRNAs), first observed by electron microscopy in the cytoplasm of eukaryotic cells (then thought to be viral infection by-products [387]), are covalently closed, single-stranded RNAs produced by the backsplicing of the downstream 3'-splicing donor site to an upstream 5' accepting site of a pre-mRNA [388]. Enzymatic enrichment using RNAseR and the development of novel bioinformatics pipeline (notably, of pipelines assembling back-splice junctions (BSJ)[389,390] and more recently of alternative back-splicing (ABS)[391]) have identified that ca. 14% of genes expressed in human fibroblasts produce circRNAs [389].

CircRNAs have been described in nematode, insects [392], and mammals [393], including non-human primates and humans [394]. They are particularly favoured in *Alu*-rich loci [389], which provide for the pairing of inverted repeat elements (so called ‘pairing-driven circularisation’ [389]), and/or binding of certain RNA binding proteins such as QK1 [395], FUS [396], and Sam68 [397]. As such, the presence of circRNAs across several lineages is likely an evolutionary convergence resulting from the independent insertion of transposable elements [398]. Gene ontology studies have revealed that certain biological processes (signalling, neurogenesis, neurodifferentiation), molecular functions (transcription factor activity, protein kinase activity), and cellular components (plasma membrane, synapse) are more frequent amongst circRNA-producing genes [389,392].

A vast majority (83-85%) of circRNAs find their origin in mRNA-coding genes [388,399]. They can be classified according to their exonic, exonic-intronic, or intronic origin. Circularisation patterns can be complex and involve alternative circularisation patterns [392]. Most circRNAs result from *cis*-events, though *trans*- circRNAs have been reported [400,401]. 90% of circRNAs are estimated to generate in competition with linear splicing events [402]. The remaining 10% of circRNAs result from lariat-driven circularisation [403]. Long and repeat rich intronic sequences favour circRNA formation under conditions of reduced spliceosomal activity [404].

The circular nature of circRNAs as well as certain base pair modifications (e.g. N6-methyladenosine) make them resistant to exonucleases and dramatically increases their stability [388,405,406]. RNA edition in circRNAs is cell type-dependent [407,408] and is more frequent in circRNAs containing a greater number on exons [408].

CircRNAs are highly expressed in the CNS throughout the animal kingdom [392]. More than 35,000 circRNAs have been identified in the human frontal cortex (vs. more than 20,000 in the liver, and 15,000 in the heart); 600 of which are highly expressed [409]. At the subcellular level, brain-expressed circRNAs are enriched in synaptic neuropils [401] and accumulate in the synaptosomes [410].

CircRNA expression is upregulated during neuronal differentiation [410]. Several circRNAs are upregulated in the murine hippocampus of the SAMP8 Alzheimer's mouse model [399]. The accumulation of circRNAs in aging brains has been described in *C. elegans* [411], *D. melanogaster* [392], and *M. musculus* [393].

In the CNS, circRNAs are produced by astrocytes and by neuronal stem cells – both in *in vitro* studies [412,413] and in *in vivo* models [414]. The functional roles of certain circRNAs have been elucidated in neurological processes – notably, circHECTD1 sponging miR-142 off the *TIPARP* mRNA in the astrocytic activation during the cerebral infarction in a mouse stroke model [415]; and the network between the *Cyrano* long non-coding RNA, the circular RNA *CDR1as*, and miR-671 and miR-7 [416]. Other regulatory loops have been suggested in human neurological pathology, including a network with four circRNAs, 11 miRNAs and 49 genes in human brain astrocytes collected from biopsies of patients with Alzheimer disease [417].

Several functions have been explored for circRNAs in such networks. circRNAs can modulate gene expression by sponging on miRNAs – with up to 74 miR-7 seed matches found in the *CDR1as* circRNA [388,418]. CircRNA have also been suggested as scaffolds or sponges for RNA-binding proteins. In doing so, they enter into competition with other RNA targets [419]. They are also thought to be able to regulate *Pol II*-dependent transcription (reviewed in [420]) and/or compete with the linear splicing of pre-mRNA [403]. Finally, certain circRNAs also include internal ribosomal entry sites (IRES) and are thought to be cap-translated into short peptides [421–423]– with the possible involvement of base modifications as drivers and/or regulators [394,423]. When identified, the molecular function of such a translation product (e.g. circ-ZNF609) remains unknown [423].

4.1.2. What are *SMN*-derived circRNAs?

The highly homologous *SMN1* and *SMN2* loci (to which I hereafter refer indiscriminately as either *SMN1/2* loci, or simply *SMN* loci) on chromosome 5q13 are rich in *Alu* elements, which represent ca. 40% of the transcribed *SMN1/2* region [424]. However, until recently, description of any circRNA derived from the *SMN1/2* loci was limited to circRNAs identified from BSJ-centred algorithms on RNAseq data. Those *SMN*-derived circRNAs include hsa_circ_0002251, a 407 bp *SMN2*-derived circRNA identified in HS68 fibroblast cells [389], and hsa_circ_0072853, and 207 bp *SMN2*-derived circRNA identified in the H1 human embryonic stem cell line [425].

A recent publication from a single group reports that backsplicing on the *SMN* loci generates a repertoire of 53 *SMN*-derived circRNAs, facilitated by canonical U2-type splice sites of both known and of newly identified cryptic exons [85]. At the onset of this work, and to our knowledge, the proportion of *SMN*-derived circRNAs within the pool of total *SMN1* and *SMN2* transcripts is unknown. In addition, cross-regulation of *SMN*-derived circRNAs on each other is suggested [85]. However, the effects of the numerous backsplicing events in *SMN* loci on the number of pre-mRNA transcripts available for splicing into linear *SMN1* and *SMN2* isoforms is unknown – as are the downstream functions, if any, of *SMN*-derived circRNAs. Finally, at the onset of this work, no specific biochemical property – be it miR sponging, RBP binding, interaction with PolII polymerase, or *SMN*-derived circRNA-derived peptide – had been identified for any of the 53 *SMN*-derived circRNAs.

4.1.3. Why investigate *SMN*-derived circRNAs in SMA?

There are three primary reasons for the investigation of *SMN*-derived circular RNAs in SMA. First, as described in Chapter 1, the SMN protein controls RNP assembly (reviewed in [426]) and SMN deficiency causes widespread splicing defects [135,145,146]. At the onset of this work, there was no existing report on the effect of the reduced spliceosomal efficiency of SMA on the expression of circRNAs. Following the recent report of a numerous repertoire of *SMN*-derived circRNAs, I chose to scope this study of the circular RNA repertoire in the low spliceosome efficiency context that is SMA; that is, on *SMN*-derived circRNA in health and in SMA.

Second, SMA is a neurological condition. Investigating *SMN*-derived circRNAs in SMA is relevant considering the neuronal tropism of both circRNA expression and of degeneration in SMA. It is unknown whether *SMN*-derived circRNAs are involved in regulatory and/or survival molecular loops in neurons – as other circRNAs have been reported. It is also unknown whether *SMN*-derived circRNAs have any function at all.

Finally, at the time when I undertook this study, a single publication reported the existence of a *SMN*-derived circRNA repertoire [85]. In it, authors report the relative abundance of *SMN*-derived circRNAs in different human, healthy, human tissues – but do not investigate their expression in SMA patients. Similarly, their report of *Smn*-derived murine circRNAs is limited to murine cell cultures. As such, there is to date no data concerning the expression of *SMN*-derived circRNAs in SMA-affected cells or patient, and nor data concerning *SMN*- or *Smn*- derived circRNA expression in any existing SMA mouse model.

4.1.4. Hypothesis and goals

The existence of a vast repertoire of new, uncharacterised transcripts from *SMN1/2* loci led us to investigate them further to establish whether they would be a target with therapeutic, diagnostic, and/or theranostic potential in SMA.

Two main limits arose to gain insights from SMA-related *SMN*-circRNA-omics and apply it drug discovery in our laboratory. The first is the use of a labour-intensive, tailored approach to describe the *SMN*-derived circRNA repertoire from *in vitro* cell cultures [85], starting with RNAse R enrichment, exon-specific reverse transcription, exon-specific divergent primers for PCR, electrophoretic separation of the RT-PCR products on native polyacrylamide gels, and an in-house radioactivity-based exon-skipping assay [83]. The second is the fall from favour of radioactivity-based applications in the UK. As such, there is a lack of a streamlineable, sensitive, specific, quantitative method to study individual *SMN*-derived circRNAs.

This led us to refine our research goals concerning *SMN*-derived circRNAs in health and in SMA. More specifically, in this chapter, I describe the work undertaken to:

(1) characterise the expression of a subset of the *SMN*-derived circRNAs described in [85], in control and SMA patient-derived fibroblasts with varying numbers of *SMN1/2* copies;

(2) develop a methodology for quantitative, streamlineable, sensitive, methodology to study circRNAs in a high-isoform context, specifically, a TaqMan™-based qPCR assay for *SMN*-derived circRNAs.

(3) use existing computational pipelines to evaluate whether four, reportedly more highly expressed, *SMN*-derived circRNAs would be suitable candidates for further *in vitro* functional exploration – including investigation of potential miR binding, translation, and RBP binding.

4.2. Material and methods

4.2.1. Cellular models

All cells were subcultured following manufacturers' instructions unless otherwise specified.

SH-SY5Y CRL-2266™ human neuroblastoma cells were purchased from ATCC. Human primary astrocytes from the spinal cord (HA-sc) were purchased from Innoprot® (cat. no. P10255). Human astrocytes were grown in DMEM/F12 (+ (L)-Gln, + bicarbonate) with 1% P/S ad 3% of Astrocyte Growth Supplement (Cell Applications, Inc.) following [427]. Human control fibroblasts (male patient, cat. no. AG02261), SMA type I fibroblasts (male patient, homozygous deletion of exons 7 and 8 in *SMN1*, 2 copies of *SMN2*, cat. no. GM00232) and SMA type II fibroblasts (male patient, homozygous deletion of exons 7 and 8, 3 copies of *SMN2*, cat. no. GM03813) were purchased from Coriell Institute.

4.2.2. RNA isolation

Human spinal cord total RNA was purchased from TakaraBio (catalogue no. 636554).

RNA was extracted by guanidinium thiocyanate-phenol-chloroform extraction using TriZOL® (Invitrogen). During the optimisation and comparison steps of this study, RNA was also extracted using Maxwell® RSC SimplyRNA cells kit (Promega) following manufacturer's instructions, including DNAse I treatment (10' at room temperature).

Total RNA was treated for 30' at room temperature using the *E. coli* 3' to 5' exonuclease RNAse R (Biovision® M1228-500), on 5 µg in 20 µL final volume.

Exon-specific cDNA synthesis was performed from 500 ng RNA per 5 µL (Vf = 10 µL) using the High-Capacity cDNA Kit (Life Technologies) following manufacturer's protocol, and primers ***(Table*** 4-1***)*** ordered from IDT.

Table 4-1: Primers used for exon-specific cDNA synthesis

Name	Primer sequence (5'-3')
230_C2A-2B-3-4_E2a_RT	TGCTTTTATCAGTGCTGTATCAT
233_C2A-2B-3-4_E2b_RT	GGCTTTTATTCTTCTTAGCAGG
236_C2A-2B-3-4_E3_RT	TCCATATCCAGTGTAAACCACAAC
239_C2A-2B-3-4_E4_RT	GTTATCTGATTTATTTCCAGGAGACC
242_C6-7-8A_E6_RT	ATTAACATACTTCCCAAAGCATCAG
245_C6-7-8A_E7_RT	GCACCTTCCTTCTTTTTGATTTTG
248_C6-7-8A_E8_RT	CTTCACATTCCAGATCTGTCTG
252_C6-6B-7-8A_E6B_RT	GATCACGGGGTCAGGAGATTG

4.2.3. Polymerase chain amplification

PCR on cDNA complementary to circRNAs was performed using DreamTaq Polymerase (ThemoFischer) and divergent primers (***Table*** 4-2***,***

Table 4-3), using between 25 and 40 cycles: 2' at 94°C, [20 seconds at 94°C, 20 seconds at 60°C, 60 seconds at 72°C] - 5 minutes at 72°C, and stored at 4°C. PCR on linear transcripts was performed following the same protocol and relevant primers (***Table*** 4-4). Quantitative PCR was performed using FastSYBR Green Master Mix (Applied Biosystems ref. 385612).

4.2.1. Semi-quantitative gels

PCR products were resolved in 3% agarose gels in TBE at 130 V for 100-120 minutes alongside the GeneRuler 50 bp DNA Ladder (ThermoFischer Scientific, cat. no. SM0371), stained for 20' with ethidium bromide (final concentration in ddH_2O: 0.5 μg/mL) and rinsed with ddH_2O for 20') after electrophoresis. Gels were imaged using UV fluorescence.

During optimisation of the protocols, native 6% polyacrylamide gels were handcast for electrophoresis following BioRad® protocols and stained with ethidium bromide.

Table 4-2: Divergent primers used for polymerase chain reaction of SMN-derived circRNAs

Name	Primer sequence (5’-3’) [85]
228_C2A-2B-3-4_E2a_end	GCATATGATAAAGCTGTGGGCTTC
229_C2A-2B-3-4_E2a_start	CCCAAATGTCAGAATCATCGC
231_C2A-2B-3-4_E2b_end	GAATACTGCAGCTTCCTTACAAC
232_C2A-2B-3-4_E2b_start	CAAATGTCACCATTCTTTAGAGC
234_C2A-2B-3-4_E3_end	CTTTCCCCAATCTGTGAAGTAGCTA
235_C2A-2B-3-4_E3_start	CAACCGTCTTCTGACCAAATGG
237_C2A-2B-3-4_E4_end	GGCCAAGACTGGGACCAGG
238_C2A-2B-3-4_E4_start	TCACTTTCATCTGTTGAAACTTGG
240_C6-7-8A_E6_end	TCATGGTACATGAGTGGCTATCA
241_C6-7-8A_E6_start	CATATGGGAGGTGGTGGG
243_C6-7-8A_E7_end	GTGCTCACATTCCTTAAATTAAGG
244_C6-7-8A_E7_start	CTTCTTTTTGATTTTGTCTGAAACC
246_C6-7-8A_E8A_end	GTTGGGAAAGAAAAAAGGAAGTGG
247_C6-7-8A_E8_start	GGTGTCATTTAGTGCTGCTCTAT
249_C6-7-8A_E8A_start	CGATCAGACAGATCTGGAATGTG
250_C6-6B-7-8A_E6B_end	ATGAAATTTTTTTAATGCACAAAGATCTG
251_C6-6B-7-8A_E6B_start	GGCTAACACAGTGAAACCCTG

Table 4-3: Divergent primers amplifying other circRNAs

Name	Primer sequence 5’ to 3’
295_Barrett_CDR1as_fwd	CGTCTCCAGTGTGCTGATCT
296_Barrett_CDR1as_rev	AAGACCCGGAGTTGTTGGAA
297_Hanniford_CDR1as_fwd	CCCAGTCTTCCATCAACTGGCT
298_Hanniford_CDR1as_rev	AAGACCTTGACACAGGTGCCAT
299_Hansen_ciRS-7_fwd	ACGTCTCCAGTGTGCTGA
300_Hansen_ciRS-7_rev	CTTGACACAGGTGCCATC
hsa_circ471_fwd	CAAAGACCTCCTCCTCCACA-
hsa_circ471_rev	TCAGTGCTGAACACAATGCC
hsa_circ284_fwd	GTCGGCCAGTCATGTATCAA
hsa_circ284_rev	ACCAAGACTTGTGAGGCCAT

Table 4-4: Primers used for the amplification of linear RNAs

Name	Primer sequence 5’ to 3’
224_hGAPDH_Fw	AACAGCGACACCCACTCCTC
225_hGAPDH_Rv	CATACCAGGAAATGAGCTTGACAA
226_h18s_Fw	GGCCCTGTAATTGGAATGAGTC
227_h18s_Rv	CCAAGATCCAACTACGAGCTT

4.2.2. Cloning, sequencing, and sequence analyses

Bands of PCR products were cut from the agarose gel, purified using the QiaQuick Gel Purification kit (QIAquick Gel Extraction Kit (50) Cat No./ID: 28704) and eluted in 30 µL EB buffer, quantified with Nanodrop, and inserted into the TOPO™ TA Cloning™ Kit for Sequencing, without competent cells (Cat# 450030, Thermofisher) following manufacturer's instructions. Plasmids from bacteria (Subcloning Efficiency™ DH5α Competent Cells Cat#18265017) were isolated using Qiagen Miniprep kit (QIAprep Spin Miniprep Kit, Cat #27104) and sent to Source Bioscience (Oxford, UK) for sequencing.

All sequence analyses and alignments were conducted using SnapGene software (www.snapgene.com).

4.2.3. Bioinformatic analyses of *SMN*-derived circRNAs

Identification of RNA secondary structures was performed using the RNAfold software (version 2.4.13) [428] which uses energy parameters taken from [429,430], as well as using the mFold program [431]. The presence of ORFs was evaluated using SnapGene software (www.snapgene.com). Identification of RBP was performed using circInteractome [432] and the RBPsuite [433]. Analysis of miR seed binding sites was performed using circInteractome [432] and miRDB [434].

4.2.4. qPCR primer design

Primers were designed using PrimerQuest Tool (Integrated DNA Technologies), by entering sequences manually, selecting the '2 primers + probe' option, and inputing the location of all exon-exon junctions in the 'custom design parameters' section.

4.3. Results

4.3.1. Semi-quantitative *SMN*-derived circRNAs expression in healthy and SMA cells

The first part of the work aimed at characterising the expression of a subset of *SMN*-derived circRNAs in control and SMA patient-derived fibroblasts with varying numbers of *SMN1/2* copies. The SMA type I fibroblasts bear a homozygous deletion of exons 7 and 8 in *SMN1* and 2 copies of *SMN2*. The SMA type II fibroblasts bear a homozygous deletion of exons 7 and 8 and 3 copies of *SMN2*.

I limited this analysis to two *SMN*-derived circRNAs, named C2A-2B-3-4 and C6-7-8A after their exon content, and chosen both for their reported expression levels in the initial publication [85], and for their exon composition. C2A-2B-3-4 is backspliced early in the *SMN1/2* transcripts and is highly homologous to other 'early' *SMN*-derived circRNAs. C6-7-8A includes the loci responsible for the alternative splicing between *SMN1* and *SMN2* transcripts and is of functional importance in SMA pathophysiology. To eschew any retrotranscription-specific efficiency bias, I assessed the expression of each circRNA from parallel retrotranscription in 'number of exons -1' loci – that is, three parallel RT reactions for C2A-2B-3-4, and two parallel RT reactions for C6-7-8A.

C2A-2B-3-4 and C6-7-8A are expressed in both type I and type II SMA fibroblasts (***Figure* 4-1**). From visual semi-quantitative estimation, their expression is greater in SMA type II fibroblasts than in SMA type I fibroblasts. This data also suggests that *SMN2* copy number appears to affect the intensity of *SMN*-derived circRNA expression in SMA fibroblasts.

Interestingly, there is no consistent pattern of over- or under-expression of C2A-2B-3-4 in SMA fibroblasts when compared to control fibroblasts. Semi-quantitation is not consistent across exons of retro-transcription: only PCR products from C6-7-8A derived from exon 7 does not yield signal in type I SMA fibroblasts.

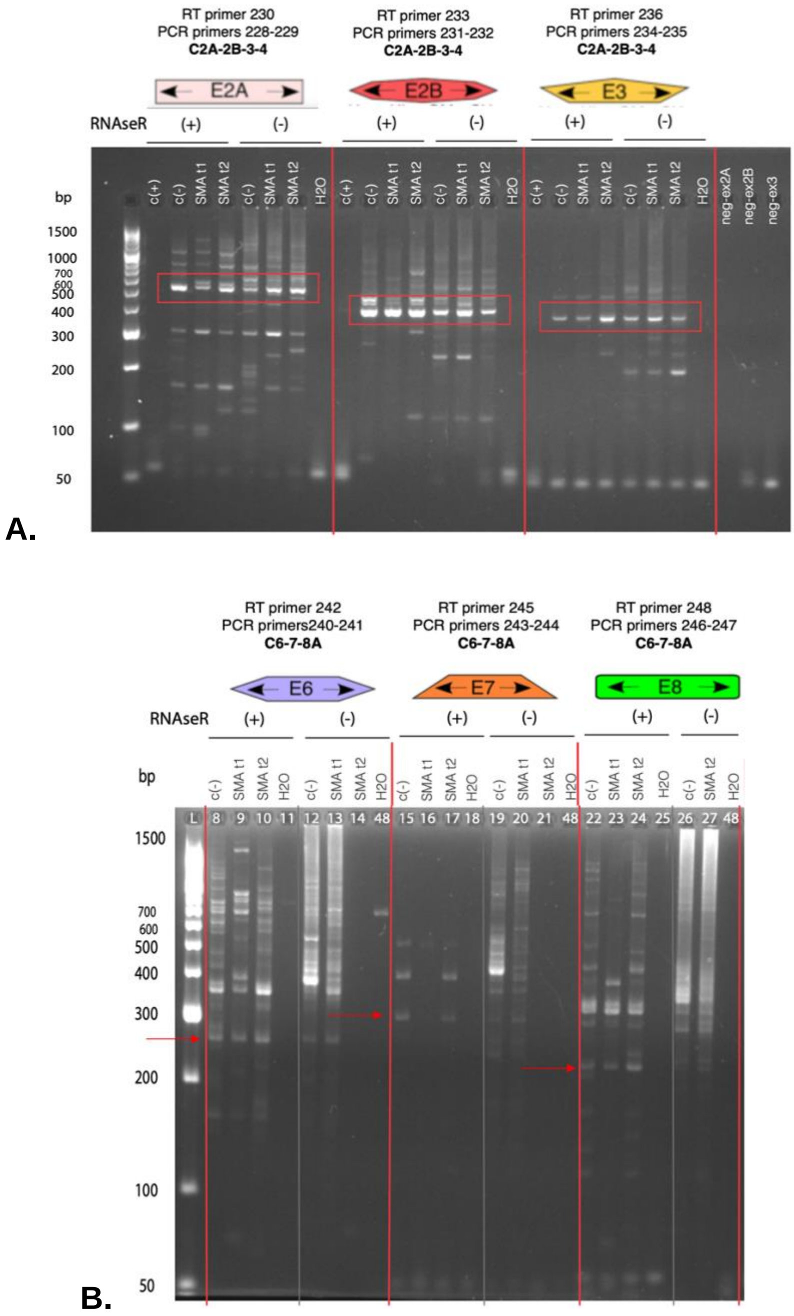


Figure 4-1: SMN-C2A-2B-3-4 and SMN-circ6-7-8A expression in human control and SMA fibroblasts

Ethidium-bromide visualisation of PCR products from the divergent amplification of SMN exons (3% agarose gel). Trizol®-extracted total RNA from human control fibroblasts ('c(-)'), SMA patient type I fibroblasts ('SMA t1'), and SMA patient type II fibroblasts ('SMA t2') was treated (+) or not treated (-) with RNAseR to deplete the samples of linear transcripts prior to exon-specific cDNA synthesis. H2O: PCR control performed with water in lieu of cDNA. ***(A)*** *PCR amplicons from divergent amplification of exons within* SMN-C2A-2B-3-4 *(theoretical size: 546 bp). Expected amplicon size (marked with red boxes) for exon 2A ('E2A', primers 228-229): 486 bp; exon 2B ('E2B', primers 231-232): 477 bp; exon 3 ('E3', primers 234-235): 443 bp.* ***(B)*** *PCR amplicons from divergent amplification of exons within* SMN-C6-7-8A *(theoretical size: 307 bp). Expected amplicon size (marked with red arrows) for exon 6 ('E6', primers 240-241): 258 bp; exon 7 ('E7', primers 243-244): 303 bp; exon 8 ('E8', primers 246-247): 225 bp. Numbers on the wells identified the samples during their processing.*

4.3.2. Semi-quantitative *SMN*-derived circRNAs expression in human spinal cord

SMN-derived circRNA including exon 2B (which includes C2A-2B-3-4 as well as other circRNAs, such as C2B-3-4) are expressed in human spinal cord samples (***Figure*** 4-2). The expression level is very low and could not be detected using other exon-specific retro-transcription reactions (data not shown). Albeit there is a PCR product of ca. 450 bp, no amplicon of a size corresponding to an amplicon from C6-7-8A can be observed for products derived from the divergent amplification of exon 6 (***Figure*** 4-2).

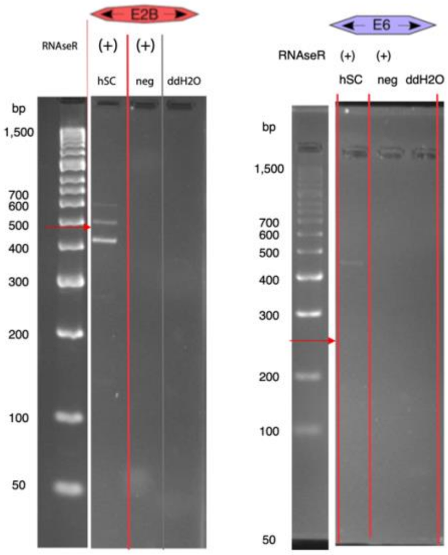


Figure 4-2: SMN-derived circRNAs containing exon 2B and exon 6 in human spinal cord.

Ethidium-bromide visualisation of PCR products from the divergent amplification of SMN exons 2B (left) and E6 (right). Expected amplicon size (marked with red arrows) for C2A-2B-3-4: exon 2B ('E2B', primers 231-232): 477 bp; and for C6-7-8A: exon 6 ('E6', primers 240-241): 258 bp. The 3% agarose gels were cropped for clarity.

4.3.3. Methodological limits to the study of *SMN*-derived circRNAs

Whilst initial data suggests that *SMN*-derived circRNA expression is affected in severe types of SMA, this data remains semi-quantitative, of low-signal intensity, and not consistent across different retrotranscription reactions. I therefore focused our efforts on methodological optimisation and the identification of limiting factors.

4.3.3.1. Limits in the RNA extraction and circRNA enrichment methods

I first compared commercially available automated kit for RNA extraction (Promega) with TriZol-based extractions (data not shown). RT-PCR performed using TriZol always yielded stronger bands than the Promega automated kit's counterpart.

I confirmed that RNAse R treated samples show fewer smears than their RNAseR (-) counterparts, and that RNAseR treatment should be systematic on an initial mass of no less than 1,5 µg RNA (***Figure*** 4-3 and ***Figure*** 4-4).

I then investigated whether exon-specific primers for cDNA synthesis really yielded better RT-PCR results than cDNA synthesis performed using random primers. As in Ottesen et al., the exon-specific amplifications produce multiple non-specific bands. Using C2A-2B-3-4 as a reference, RT-PCR performed using random primers is feasible but yields weaker signal (***Figure*** 4-3 ***A***) than for RT-PCR performed on cDNA synthesised using exon-specific primers (***Figure*** 4-3 ***B***). As for E6 divergent amplification in hSC samples, the amplicon size (and/or migration) differs from its expected theoretical size (that is, the assumption that PCR products derive from C2A-2B-3-4). Of note, amplicon size differs for the three different PCRs performed from three different exon-specific cDNAs.

4.3.3.2. Limits in PCR product resolution

The nucleotide-level of resolution imposed by the high number of circRNAs generated from the *SMN1/2* loci imposes the use of high-density gels. I therefore compared the resolution capacity of 3% agarose gels in TBE and native 6% PAGE. Using the same PCR products, 6% PAGE yields clearer band resolution but weaker signal than 3% agarose gels in TBE. At this stage of our analysis, the limiting factor to our semi-quantitation is signal strength. I therefore pursued with 3% agarose gels in TBE, which yield stronger signal than their polyacrylamide counterpart – in addition to being easier to cast and handle (***Figure*** 4-4).

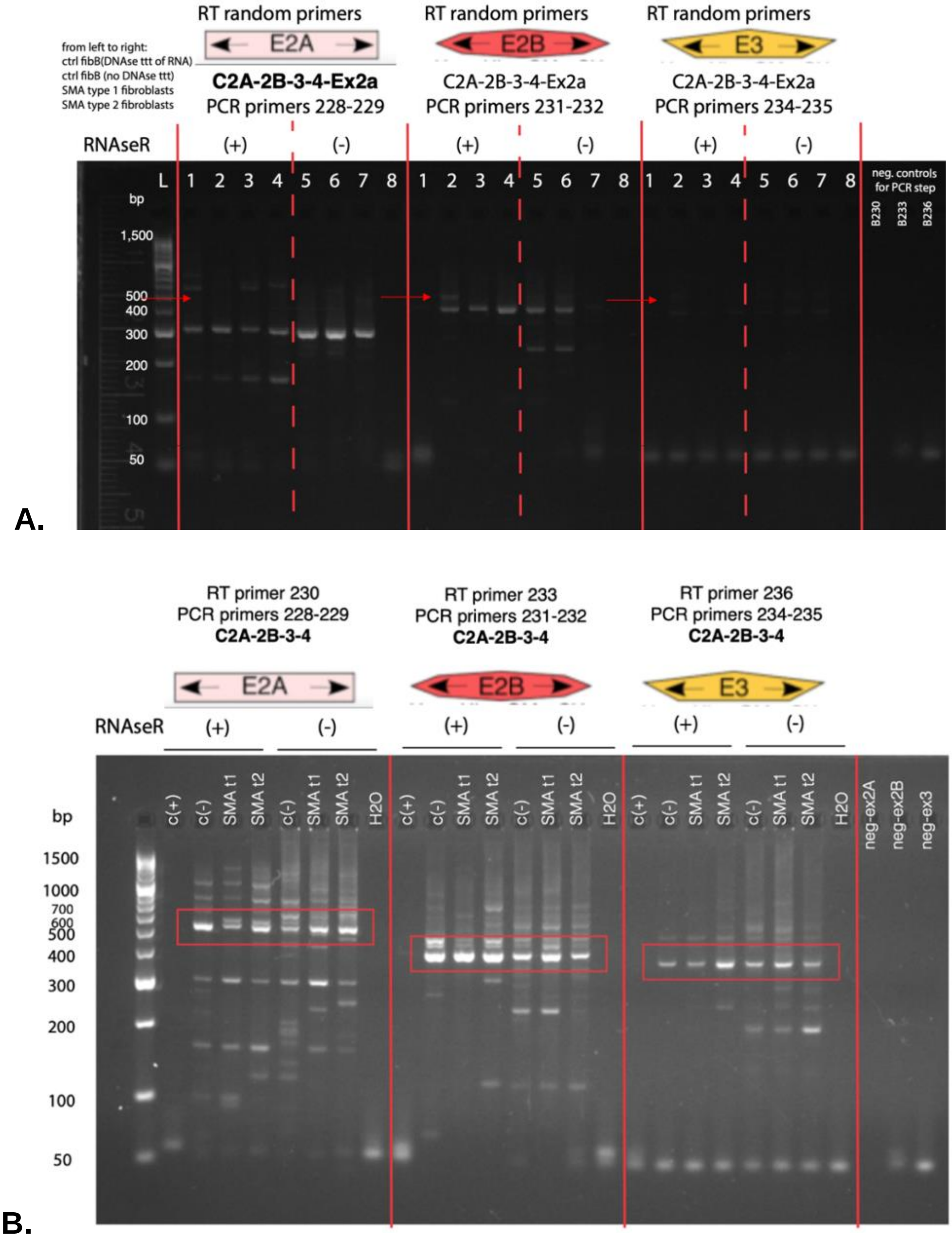


Figure 4-3: Divergent amplification of SMNcirc2A-2B-3-4 following random vs. exon-specific retrotranscription

Ethidium-bromide visualisation of PCR products from the divergent amplification of SMN exons (3% agarose gel). Trizol®-extracted total RNA from human control fibroblasts treated with DNAse ('c(+)') or not ('c(-)'), SMA patient type I fibroblasts ('SMA t1'), and SMA patient type II fibroblasts ('SMA t2') was treated (+) or not treated (-) with RNAseR to deplete the samples of linear transcripts prior to exon-specific cDNA synthesis. H2O: PCR control performed with water in lieu of cDNA. Neg-ex2A: exon 2A retrotranscription control with water in lieu of total RNA. Neg-ex2B: exon 2B retrotranscription control with water in lieu of total RNA. Neg-ex3: exon 3 retrotranscription control with water in lieu of total RNA. Expected amplicon size (marked with red arrows or boxes) for exon 2A ('E2A', primers 228-229): 486 bp; exon 2B ('E2B', primers 231-232): 477 bp; exon 3 ('E3', primers 234-235): 443 bp. ***(A)*** *PCR amplicons from divergent amplification of exons after cDNA synthesis using random primers.* ***(B)*** *PCR amplicons from divergent amplification of exons after cDNA synthesis using exon-specific primers.*

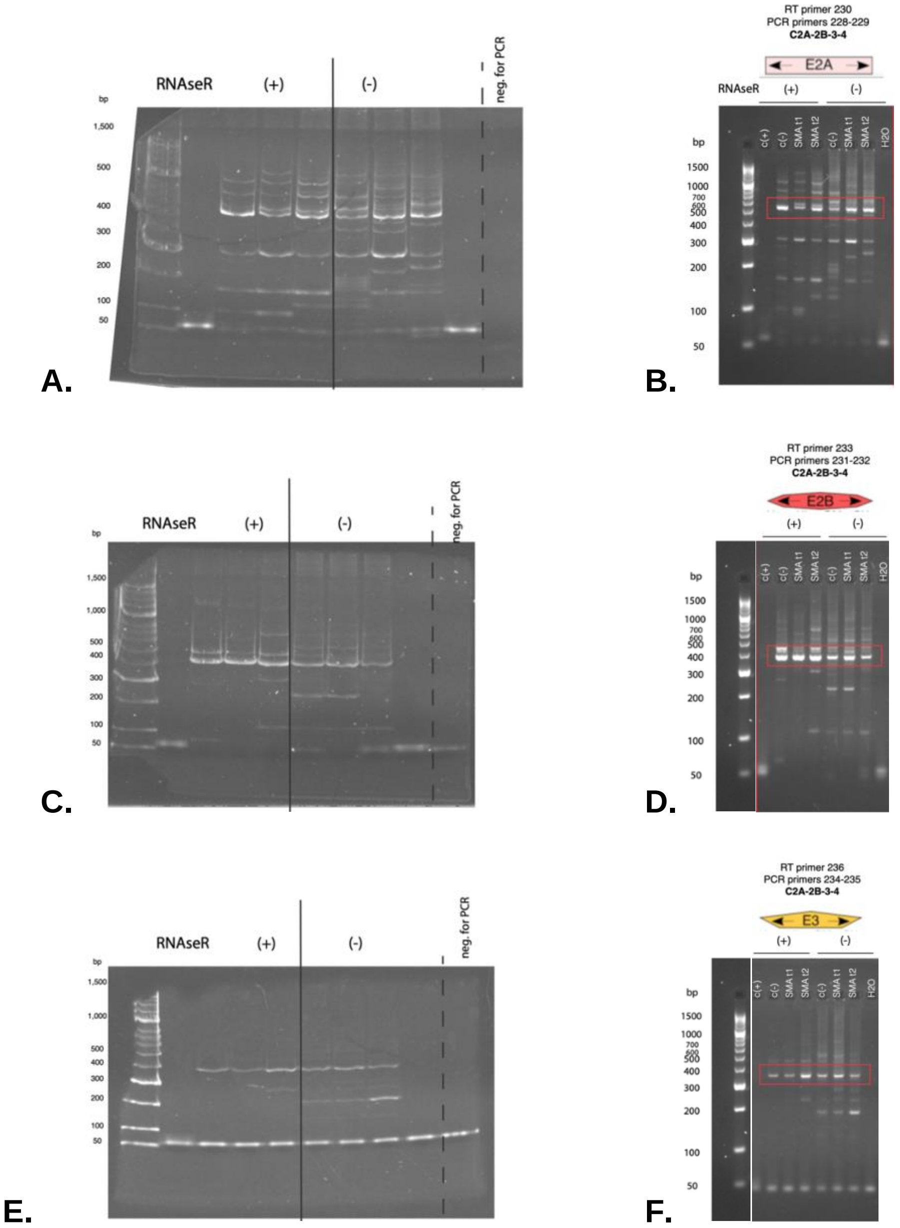


Figure 4-4: Differential resolution of products from divergent amplification in 6% native PAGE and 3% agarose gels.

Ethidium-bromide visualisation of PCR products from the divergent amplification of exons present in C2A-2B-3-4. Trizol®-extracted total RNA from human control fibroblasts treated with DNAse ('c(+)') or not ('c(-)'), SMA patient type I fibroblasts ('SMA t1'), and SMA patient type II fibroblasts ('SMA t2') was treated (+) or not treated (-) with RNAseR to deplete the samples of linear transcripts prior to exon-specific cDNA synthesis. H2O: PCR control performed with water in lieu of cDNA. PCR amplicons from divergent amplification of exons present within SMN-C2A-2B-3-4 *(theoretical size: 546 bp).* ***(A), (C), (E)*** *PCR product resolution in 6% native polyacrylamide gels (PAGE).* ***(B), (D), (F)*** *PCR product resolution in 3% agarose gels.*
(A), (B) *Expected amplicon size (marked with red boxes) for exon 2A ('E2A', primers 228-229): 486 bp.*
(C), (D) *Expected amplicon size (marked with red boxes) for exon 2B ('E2B', primers 231-232): 477 bp.*
(E), (F) *Expected amplicon size (marked with red boxes) for exon 3 ('E3', primers 234-235): 443 bp.*
(B), (D), (F) *were run on the same gel and share the same ladder lane, added here* post-hoc *for lisibility.*

4.3.3.3. Defining a loading control

The first report of *SMN*-derived circRNAs by Ottesen et al. is semi-quantitative and evaluates the presence of linear RNAs using *18s* and *GAPDH* as controls – both controls that are quantifiable *before* treatment with RNAse R (respectively, rRNA and mRNA). I therefore investigated whether hsa_circ471 and hsa_circ284, two recently reported [435] human circRNA internal control candidates, could be used as controls for relative quantitation between samples. Such a loading should be specific (i.e. yield a single amplicon) and expressed across cell lines.

hsa_circ284 is not expressed at high enough levels in control fibroblasts to be used as an internal circRNA control (***Figure*** 4-5). The other candidate, hsa_circ471, is expressed significantly in control fibroblasts. However, the PCR reaction is not specific to a unique circRNA locus and yields three products (***Figure*** 4-5). In addition, hsa_circ471 is not expressed at sufficient levels in two central nervous system cells of interest, the neuroblastoma cell line SH-SY5Y and human spinal cord astrocytes HA-sp, to be used across cell lines (***Figure*** 4-5).

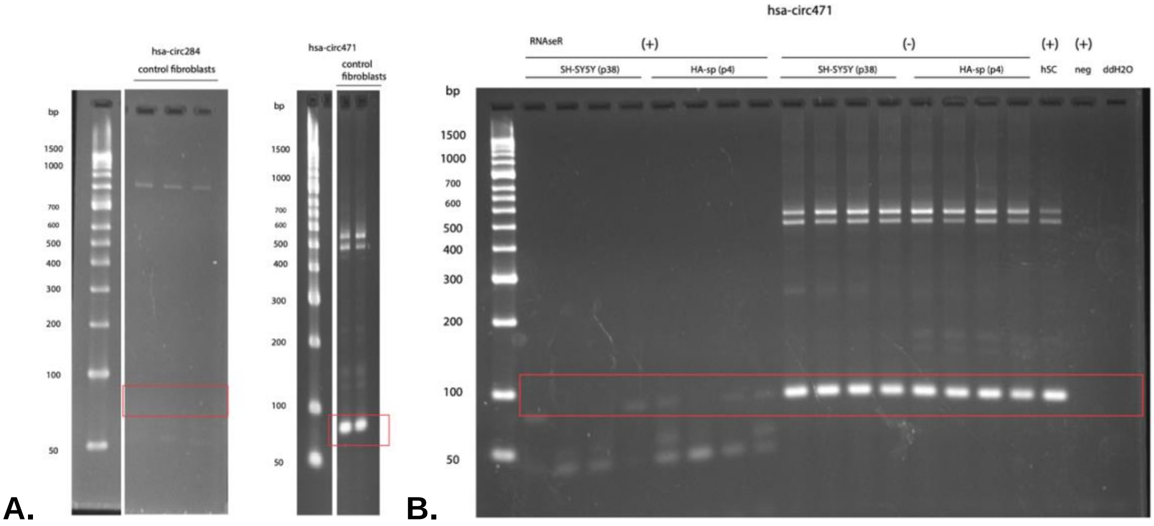


Figure 4-5: Expression of hsa-circ284 and hsa-circ471 in control fibroblasts and in central nervous system cells.

*Ethidium-bromide visualisation of PCR products from the divergent amplification of SMN exons (3% agarose gel). Trizol®-extracted total RNA from **(A)** RNAse-R (-) human control fibroblasts , and **(B)** either RNAseR(+) or RNAse(-)SH-SY5Y cells, HA-sp cells, and human SC. H2O: PCR control performed with water in lieu of cDNA. Expected amplicon size (marked with red boxes) for hsa-circ284: 82 bp. Expected amplicon size (marked with red boxes) for hsa-circ471: 77 bp.*

4.3.4. Perspectives on the optimisation of circRNA quantitation methodology in a highly homologous context

Thus far, I have shown that quantitation of *SMN*-derived circRNAs is labour-intensive, semi-quantitative, is not easy to streamline with usual controls for RNA quantitation, and does not allow for a fast resolution of different, highly similar isoforms. The low levels of PCR products generated from divergent PCRs containing circRNAs are also a limiting factor to the validation of the backsplice junction by sequencing (data not shown).

All these issues are compounded by two particular facts. The first is the rare nature of backsplicing events (an estimated 2% of splicing events are backsplicing events [436]): to this date, there is no good estimation of what proportion of *SMN* pre-mRNA yield circular vs. linear mature transcripts. The second compounding factor is the high number of isoforms within the *SMN* circRNAs repertoire. Based on our experience testing other circRNA primers using divergent primer amplification, the *SMN* locus is not the only one; hsa_circ471 also yielded several PCR products.

To circumvent these generation- and locus-related issues, I aimed at developing a specific, sensitive, quantitative method to distinguish highly homologous circRNAs derived from the *SMN* loci. I focus development on four circRNAs chosen as a subset of highly homologous circRNAs with high expression levels [85]. The closure of laboratories during the COVID-19 pandemic significantly impacted the experimental application of these designs, and I therefore present them as perspectives for further development.

4.3.4.1. Characteristics of the *SMN*-derived circRNAs chosen for proof of concept

I chose four circular RNAs derived from the *SMN* loci. They present a subset summarising several of the issues from the whole repertoire, including the fact that several circRNAs share the same backsplice junction (BSJ). For example, C2A-2B-3-4 shares the exon 4-2A BSJ with C2A-2B-4; and C6-7-8A shares the exon 8A-6 BSJ with C6-6B-7-8A and with C6-8A (***Table*** 4-5).

Table 4-5: exonic and splice junction characteristics of four SMN-derived circRNAs

SMN-circRNA	Length (bp)	BSJ	SMN circRNAs sharing the BSJ
C2A-2B-3-4	546	Exon 4 to 2A	C2A-2B-3-4 C2A-2B-4
C2B-3-4	474	Exon 4 to 2B	n/a
C3-4	354	Exon 4 to 3	n/a
C6-7-8A	307	Exon 8A to 6	C6-6B-7-8A C6-8A

4.3.4.2. On technologies and primer design for quantitative PCR of *SMN*-derived circRNAs

4.3.4.2.1. SYBR Green PCR quantitation of SMN-derived circRNAs

SYBR Green qPCRs rely on the intercalation of a dye to double-stranded DNA strands synthesised by *Taq* polymerase. SYBR Green qPCR has been reported for *SMN*-derived circRNAs [85,397]. In our laboratory however, SYBR Green qPCR reactions that were performed with divergent primers on RT-products from RNAse R-treated samples have yielded products with multiple Tm. This prevented their use for quantitation (data not shown). I therefore chose to pursue the design of a TaqMan™ based system.

4.3.4.2.2. TaqMan™-based PCR quantitation of SMN-derived circRNAs

The TaqMan™ assay relies on the presence of a hydrolysis probe bearing a fluorophore located on the top strand, between the forward and reverse primers. The signal of the fluorophore is quenched until the probe is digested by a 5'-3' exonuclease upon synthesis of the nascent strand by the *Taq* polymerase. The hydrolysis probe adds specificity to the assay. The signal read is based on the fluorophore borne on the hydrolysis probe. As such, it is possible to multiplex assays and measure the amplification of two different sequences within the same reaction.

I designed sets of TaqMan™ primers and probes for the quantitation of highly homologous *SMN*-derived circRNAs. This approach still relies on the combination of RNAseR enrichment and probe/primer design over the BSJ to select circRNAs. For primer/probe design using IDTdna's PrimerQuest Tool, I followed the following guiding principles: (1) The TaqMan™ hydrolysis probe is to be located upon the BSJ to yield increased specificity. (2) Sequences are entered manually based on theoretical BSJ, and the location of all exon-exon junctions is put in the 'custom design parameters' section. (3) Where the PrimerQuest program yields different candidate sets, I select those with a maximum coverage of different exons within the amplicon. (4) Should the sequence allow it, positioning one or both primers to cover a forward splice junction allows the selective amplification of RNA spliced products only, and the exclusion of any left-over genomic DNA.

4.3.4.3. Perspectives for relative quantitation of highly homologous circRNAs

I considered two (not mutually excluding) possibilities for circRNA normalisation: (i) the spike-in of an artificial circRNA; or (ii) the use of a well-characterised circRNA expressed in neuronal tissues generated from a locus with lesser genomic complexity than *SMN*. I set on *CDR1as* as a widely studied circRNA [418,437–440]. Relative quantitation of *SMN*-derived circRNA transcripts to linear *SMN* transcripts is possible using different probe fluorophores in the TaqMan™ based quantitation method.

The experimental validation of these primer sets and of normalisation methods did not occur during the time of our experimental work due to laboratory closures during the COVID-19 pandemic.

4.3.4.4. Other possible strategies for the direct quantitation of circRNAs

In parallel to design of a nucleic amplification-based method for the quantitation of *SMN*-derived circRNAs, I also prepared the development of another method, direct-hybridisation based method from an initial cellular lysate. First, the cell lysate is incubated with sets of *n* probes with perfect homology to *n* number of splice-junctions (both linear and BSJ). The chemistry of the probe remains to be determined, though probes with locked nucleic acid (LNA) have very high Tm that guaranties tight and stable association. Each probe is coupled to a given fluorophore. Second, negatively charged nucleic acids are separated from all positive residues (proteins, membranes) using a cation-exchange chromatography, and are separated by size and shape using a low-pressure size-exclusion chromatography (LPLC). Third, fractions from the size-exclusion chromatography are quantified using a spectrophotometer, exciting the probes at their respective wavelength. A calibration curve is run in parallel to allow for absolute circRNA quantification. Finally, chromatography fractions with highest fluorescence signals are run on a gel to validate circRNA size. After linearisation and PCR amplification, it can be sequenced to BSJ identity.

Whilst this LPLC-based, hybridisation-dependent method relies on specialised equipment, it is not destructive of the total pool of RNA and does not rely on enzymatic activity. Again, the experimental validation of these primer sets and of normalisation methods did not occur during the time of our experimental work due to laboratory closures during the COVID-19 pandemic.

4.3.5. *In silico* predictions of *SMN*-derived circRNAs functions

At this stage of our study, all report and most ongoing research on the numerous *SMN*-derived circRNAs focused on circRNA categorisation and manipulation. At the onset of our research, there was no report of any research into their possible function(s). I focused our investigation of *SMN*-derived circRNAs on the elucidation of their function.

I undertook to use sequences (derived from NM0000344.3) of the most strongly expressed *SMN1*-derived circular RNAs derived from early exons to assess whether any would (1) be a template for translation, either canonically (via ORFs) or non-canonically (via internal ribosomal entry sites, IRES) and could be a template for translation; (2) what their secondary structures entailed; (3) whether any *SMN*-derived circRNA contained seeds for miR binding; and (4) whether any of their sequences predicted for the binding of RNA-binding proteins (RBP).

4.3.5.1. Prediction of canonical translation of *SMN*-derived circRNAs

Using RefSeq sequences (NG008691.1 for *hSMN1* and NG008723.2 for *hSMN2*), I used the SnapGene software to identify open reading frames (ORF). ORF were defined using three assumptions: ATG start codon, standard genetic code, peptides of minimum 75 aa length.

SMN-derived circRNAs from later exons, including all *SMN*-derived circRNAs including exon 6, yield between 2 and 10 predicted ORFs (***Table*** 4-6). Some circular RNAs include no ORFs (C2A-2B-4, C2B-3, C3-4).

Table 4-6: Number of ORF as predicted in vitro using SnapGene

SMN-circRNA	Number of ORF
C2A-2B-3-4	2
C2A-2B-4	0
C2B-3	0
C2B-3-4	2
C3-4	0
C2A-2B-3-4-5-6	4
C2A-2B-3-4-6	4
C2A-2B-4-6	3
C2A-2B-4-6(61)	2
C2A-2B-6	2
C4-5-6	10
C5-6-7	10
C6-7-8A-9	10
C6-7-8A-9tr1	10
C6-8A-9	10
C6-6B-7-8A-9tr1	10
C5-6-6B-7-8A-9tr1	10

4.3.5.2. Secundary structures of *SMN*-derived circRNAs

Structure governs function, be it in terms of availability of given sequences (e.g. for base pairing) or of structural motives recognised by other molecular actors (e.g. RNA binding proteins).

I used RNAfold for computational prediction of the RNA secondary structures of four most highly expressed *SMN*-derived circRNAs (C2A-2B-3-4, C2B-3-4, C3-4, and C6-7-8A), based on thermodynamics, and defined as sets of base-pairs with hydrogen bonds between the nucleotides. Specifically, it predicts duplex stability with reasonable accuracy by adding the free energies from simpler modelling that accounts for the identity of each base pair and its nearest neighbours [441].

C2A-2B-3-4, C2B-3-4, C3-4, and C6-7-8A all exhibit areas of high probability of double-stranded RNA stems. C2A-2B-3-4 and C2B-3-4 both present long sequences of hairpin loops without internal base pairing. This suggests high flexibility of C2A-2B-3-4 and possible changes in secondary and tertiary structures, as well as the availability of those sequences for pairing with other short nucleotide sequences (***Figure*** 4-6). Similar data emerges from using mFold (now amalgamated with DINAMelt into UNAFold), with C2A-2B-3-4 adopting several secondary structures, with free energy changes between -120.80 and -124.41 (***Figure*** 4-7***, Supplementary Figure*** 7-7).

hSMN1, C2A-2B-3-4 hSMN1, C2B-3-4 hSMN1, C3-4 hSMN1, C6-7-8A

Figure 4-6: RNA-fold structural predictions of SMN-derived circular species

The structural predictions for C2A-2B-3-4, C2B-3-4, C3-4, and C6-7-8A (from left to right), were made using the RNAfold server. The centroid structure drawing encodes base-pair probabilities (low probabilities, towards 0, in blue; high probabilities, towards 1, in red); using minimum free energy (MFE) and partition function, and avoiding isolated base-pairs.

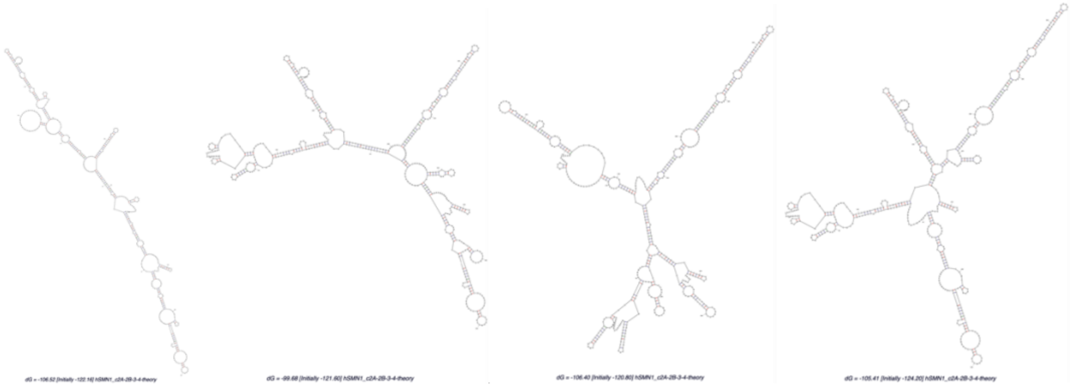


Figure 4-7: mFold predicts four possible configurations for C2A-2B-3-4

Structural predictions for C2A-2B-3-4 using the mfold server shows for possible configurations with varying free energies: dG=-106.52, dG=-99.68; dG=-106.40, and dG=-105.41 (from left to right).

4.3.5.3. Prediction of RNA-binding proteins associations with *SMN*-derived circRNAs

I used the RBP-suite RNA-binding site 'general model' prediction *in silico* tool [433] to search for consensus RNA-binding protein (RBP) motives. The RBPsuite is a deep-learning algorithm that has been trained to identify the consensus motives of 37 RBP from circInteractome – including AGO proteins, FUS, LIN28 proteins, QKI, TIA and TIAL1, and TDP43.

C2A-2B-3-4 bears verified (i.e. binding sites that have been previously validated experimentally with CLIP-seq data) sites for 10 different RNA-binding proteins, with highest scores for FOX2 and IGF2BP2. All four circRNAs of interest bind to FOX2, IGF2BP2, IGF2BP3, LIN28A, LIN28B, and PTB with at least one binding site (***Figure*** 4-8***, Supplementary Table*** 7-7). No identified verified RBPs binding site is located on the BSJ of either C2A-2B-3-4 (***Supplementary Figure*** 7-8 ***to Supplementary*** Figure 7-16) or C6-7-8A (***Supplementary Figure*** 7-17 ***to Supplementary*** Figure 7-28).

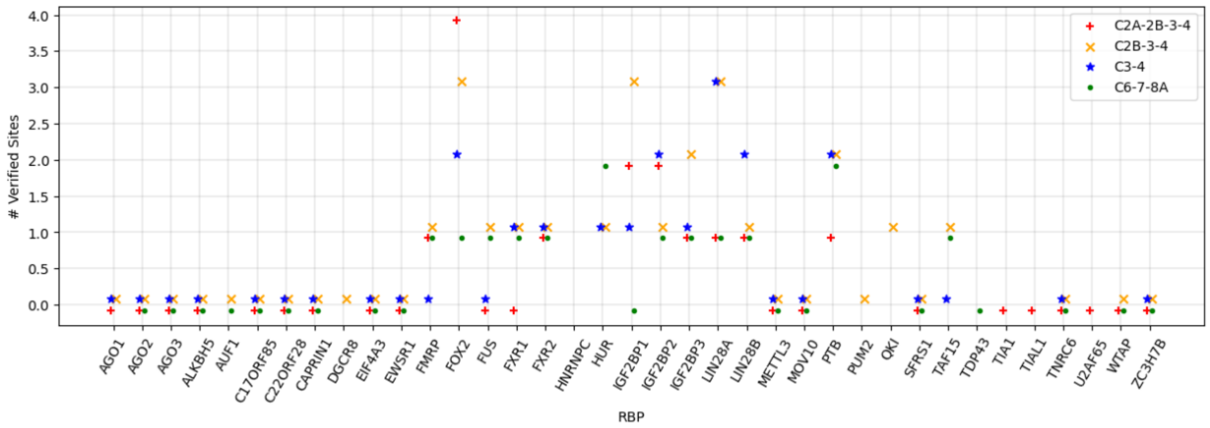


Figure 4-8: Number of verified RBP binding sites per SMN-derived circRNA

The x-axis plots the 37 RBPs for which the RBPsuite algorithm has been trained to identify binding sites on given RNA sequences. The y-axis plots the number of verified sites identified by the algorithm for C2A-2B-3-4 (red), C2B-3-4 (orange), C3-4 (blue), and C6-7-8A (green).

Using the Human Protein Atlas, I selected for proteins expressed in the brain (with evidence available at the protein level) (***Table*** 4-7). 30 of the 37 RBPs investigated with RBPsuite are expressed in the brain. Of those, PTB is the most highly expressed (34.4 nTPM). The two RBPs predicted to bind to all four investigated *SMN*-derived circRNAs are IGF2BP2 (1.5 nTPM) and FOX2 (1.1 nTPM) – both expressed in the brain, but at low levels. IGF2BP3, LIN28A, LIN28B do not appear expressed in the spinal cord using the Human Protein Atlas.

Table 4-7 : RBPs predicted to bind to SMN-derived circRNAs are expressed in the central nervous system but at low levels.

Gene	Ensembl	Gene description	RNA transcript counts in spinal cord [nTPM]
QKI	ENSG00000112531	QKI, KH domain containing RNA binding	304.5
HNRNPC	ENSG00000092199	Heterogeneous nuclear ribonucleoprotein C	295.2
TARDBP (TDB43)	ENSG00000120948	TAR DNA binding protein	74.8
U2AF2 (U2AF65)	ENSG00000063244	U2 small nuclear RNA auxiliary factor 2	72
FXR1	ENSG00000114416	FMR1 autosomal homolog 1	60.5
EWSR1	ENSG00000182944	EWS RNA binding protein 1	59.7
TAF15	ENSG00000270647	TATA-box binding protein associated factor 15	54.3
HNRNPD (AUF1)	ENSG00000138668	Heterogeneous nuclear ribonucleoprotein D	54.1
SRSF1	ENSG00000136450	Serine and arginine rich splicing factor 1	47.4
FUS	ENSG00000089280	FUS RNA binding protein	45.7
WTAP	ENSG00000146457	WT1 associated protein	36.9
PTBP1	ENSG00000011304	Polypyrimidine tract binding protein 1	34.4
FXR2	ENSG00000129245	FMR1 autosomal homolog 2	33.4
CAPRIN1	ENSG00000135387	Cell cycle associated protein 1	32.3
ZC3H7B	ENSG00000100403	Zinc finger CCCH-type containing 7B	30.2
PUM2	ENSG00000055917	Pumilio RNA binding family member 2	29.1
EIF4A3	ENSG00000141543	Eukaryotic translation initiation factor 4A3	22.7
FMR1	ENSG00000102081	FMRP translational regulator 1	22.7
TIA1	ENSG00000116001	TIA1 cytotoxic granule associated RNA binding protein	21.4
ALKBH5	ENSG00000091542	AlkB homolog 5, RNA demethylase	19.3
ELAVL1 (HUR)	ENSG00000066044	ELAV like RNA binding protein 1	16.2
TIAL1	ENSG00000151923	TIA1 cytotoxic granule associated RNA binding protein like 1	14.7
METTL3	ENSG00000165819	Methyltransferase like 3	13.1
AGO1	ENSG00000092847	Argonaute RISC component 1	11.3
DGCR8	ENSG00000128191	DGCR8 microprocessor complex subunit	4.7
MOV10	ENSG00000155363	Mov10 RISC complex RNA helicase	4.1
TNRC6C	ENSG00000078687	Trinucleotide repeat containing adaptor 6C	2.5
IGF2BP2	ENSG00000073792	Insulin like growth factor 2 mRNA binding protein 2	1.5
RBFOX2 (**FOX2**)	ENSG00000100320	RNA binding fox-1 homolog 2	1.1

nTPM: transcript per million, normalised.

4.3.5.4. miR binding of *SMN*-derived circRNAs

The best characterised function of circRNAs in literature is their role in binding to microRNAs. In the following section, I report evidence supporting any interaction between the four *SMN*-derived circRNAs of interest (C2A-2B-3-4, C2B-3-4, C3-4, and C6-7-8A) and known miRs.

I investigate three lines of evidence: (1) It has been reported for other circRNAs that there exist interactions of miR with circRNAs via AGO ([442,443] reviewed in [444]): I therefore summarise evidence regarding binding of AGO proteins with circRNAs. (2) I summarise any known binding seeding sites for miRs that do not differ between the linear and circRNAs loci. (3) Finally, I explore any miRs seeds created by the backsplice junctions of C2A-2B-3-4, C2B-3-4, C3-4, and C6-7-8A.

4.3.5.4.1. AGO proteins and SMN-derived circRNAs

I previously reported (using RBPsuite) that neither C2A-2B-3-4, C2B-3-4, C3-4, nor C6-7-8A bear a binding site to AGO1, AGO2, or AGO3 (***Figure*** 4-8).

Using circInteractome, another online web server, I found that *SMN*-C2B-3-4 reportedly binds to AGO1 with 25 bp located in exon 3 (***Figure*** 4-9). SMN-C5-6 (hsa-circ78253) also binds to AGO1. Of note, circInteractome only takes input from a given format of previously reported in UCSC with the *hsa_circ_XXXX* standardised nomenclature. I was therefore unable to test RBP for other *SMN*-derived circRNAs.

A.

CircRNA ID	hsa_circ_0002251	Location	chr5:70237215-70238697
Genomic Length	1482 bp	Spliced Seq Length	474 bp
Best Transcript	NM_017411 Primers	Gene Symbol	SMN2
Samples	Hs68_RNase, Hs68_control	Study	Jeck2013
GenomicSeq	hsa_circ_0002251	Mature Seq	hsa_circ_0002251

RNA-binding protein sites matching to circRNAs

RBP	BindingSites
AGO1	1
FUS	1

CircRNA ID	hsa_circ_0072853	Location	chr5:70240484-70242003
Genomic Length	1519 bp	Spliced Seq Length	207 bp
Best Transcript	NM_017411 Primers	Gene Symbol	SMN2
Samples	H1hesc	Study	Salzman2013
GenomicSeq	hsa_circ_0072853	Mature Seq	hsa_circ_0072853

RNA-binding protein sites matching to circRNAs

RBP	BindingSites
AGO1	1
EIF4A3	1
LIN28B	1

B.

Circular RNA	Tag Name	% Identity	Alignment Length	Mismatches	Gap Openings	Tag Start	Tag End	Circular RNA Start	Circular RNA ENd
hsa_circ_0002251	HPRO6_60062_G71907.1_69362859_25	100.00	25	0	0	1	25	208	232

C.

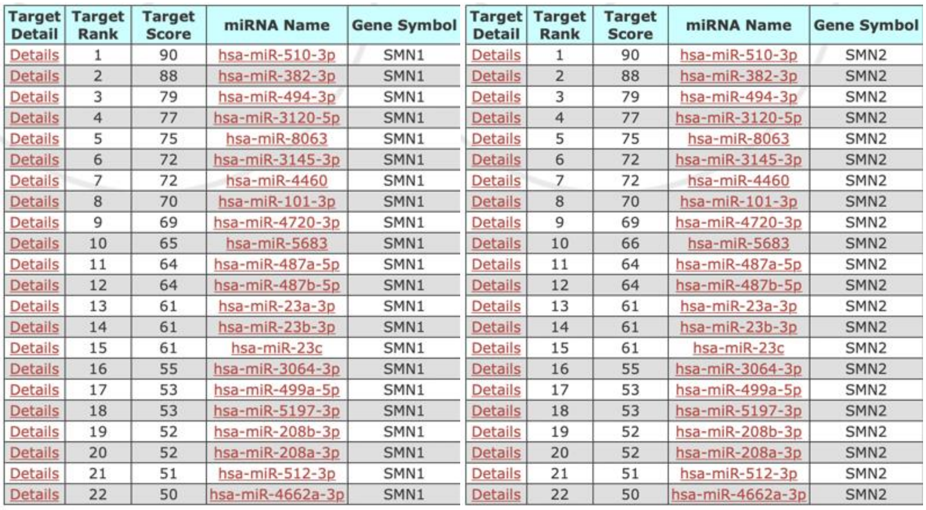
Figure 4-9: In silico evidence for the binding of SMN-C2B-3-4 (hsa-circ-2251) to AGO1.

(A) *circInteractome ("circular RNA" tab) synthesis of the available RBP binding sites for hsa-circ-2251 (C2B-3-4, left) and hsa-circ72853 (C5-6, right), based on their original reporting studies.* ***(B)*** *Based on crosslinking-immunoprecipitation (CLIP) datasets, circInteractome computationally identifies potential AGO1 binding sites only on hsa-circ-2251 (C2B-3-4).* ***(C)*** *The potential AGO1 25-bp binding site is located in exon 3 of C2B-3-4.*

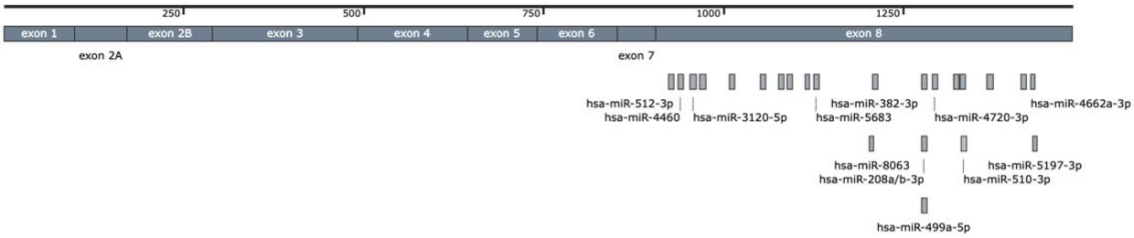
4.3.5.4.2. *miR binding on SMN linear locus*

Next, I sought to establish any binding of miRs to *SMN*-derived circRNAs from sequences already present in the linear SMN loci. miRDB compiles miRs predicted by the MirTarget bioinformatics tool analysing of miRNA-target interactions from high-throughput sequencing experiments [434]. It lists 22 miRs binding to both the *SMN1* and *SMN2* loci with a target score >50 (***Figure*** 4-10). All 22 miR locate to the 3'UTR sequence of the *SMN1/2* loci. Using miRwalk[445], miR-512-3p binds to bp 53-70 of exon 2A with a free energy of -17.1 (***Table*** 4-8). Of note, miR-512-3p also binds to exon 1 (bases 92-134).

A.



Target Detail	Target Rank	Target Score	miRNA Name	Gene Symbol	Target Detail	Target Rank	Target Score	miRNA Name	Gene Symbol
Details	1	90	hsa-miR-510-3p	SMN1	Details	1	90	hsa-miR-510-3p	SMN2
Details	2	88	hsa-miR-382-3p	SMN1	Details	2	88	hsa-miR-382-3p	SMN2
Details	3	79	hsa-miR-494-3p	SMN1	Details	3	79	hsa-miR-494-3p	SMN2
Details	4	77	hsa-miR-3120-5p	SMN1	Details	4	77	hsa-miR-3120-5p	SMN2
Details	5	75	hsa-miR-8063	SMN1	Details	5	75	hsa-miR-8063	SMN2
Details	6	72	hsa-miR-3145-3p	SMN1	Details	6	72	hsa-miR-3145-3p	SMN2
Details	7	72	hsa-miR-4460	SMN1	Details	7	72	hsa-miR-4460	SMN2
Details	8	70	hsa-miR-101-3p	SMN1	Details	8	70	hsa-miR-101-3p	SMN2
Details	9	69	hsa-miR-4720-3p	SMN1	Details	9	69	hsa-miR-4720-3p	SMN2
Details	10	65	hsa-miR-5683	SMN1	Details	10	66	hsa-miR-5683	SMN2
Details	11	64	hsa-miR-487a-5p	SMN1	Details	11	64	hsa-miR-487a-5p	SMN2
Details	12	64	hsa-miR-487b-5p	SMN1	Details	12	64	hsa-miR-487b-5p	SMN2
Details	13	61	hsa-miR-23a-3p	SMN1	Details	13	61	hsa-miR-23a-3p	SMN2
Details	14	61	hsa-miR-23b-3p	SMN1	Details	14	61	hsa-miR-23b-3p	SMN2
Details	15	61	hsa-miR-23c	SMN1	Details	15	61	hsa-miR-23c	SMN2
Details	16	55	hsa-miR-3064-3p	SMN1	Details	16	55	hsa-miR-3064-3p	SMN2
Details	17	53	hsa-miR-499a-5p	SMN1	Details	17	53	hsa-miR-499a-5p	SMN2
Details	18	53	hsa-miR-5197-3p	SMN1	Details	18	53	hsa-miR-5197-3p	SMN2
Details	19	52	hsa-miR-208b-3p	SMN1	Details	19	52	hsa-miR-208b-3p	SMN2
Details	20	52	hsa-miR-208a-3p	SMN1	Details	20	52	hsa-miR-208a-3p	SMN2
Details	21	51	hsa-miR-512-3p	SMN1	Details	21	51	hsa-miR-512-3p	SMN2
Details	22	50	hsa-miR-4662a-3p	SMN1	Details	22	50	hsa-miR-4662a-3p	SMN2



B.

Figure 4-10: MicroRNAs primarily bind to the 3'UTR segments of SMN1/2 loci

(A) *miRNAs candidates (target score, that is, probability score reflecting the statistical confidence of the prediction, >50) for binding on the* SMN1 *(left panel) and* SMN2 *(right panel) loci as ranked by the MirTarget algorithm, which combines information from CLIP binding and target downregulation data.*
(B) *Binding sites of the 22 miRNAs identified by MirTarget, all of which bind to SMN1/2 3'UTR (exon 8).*

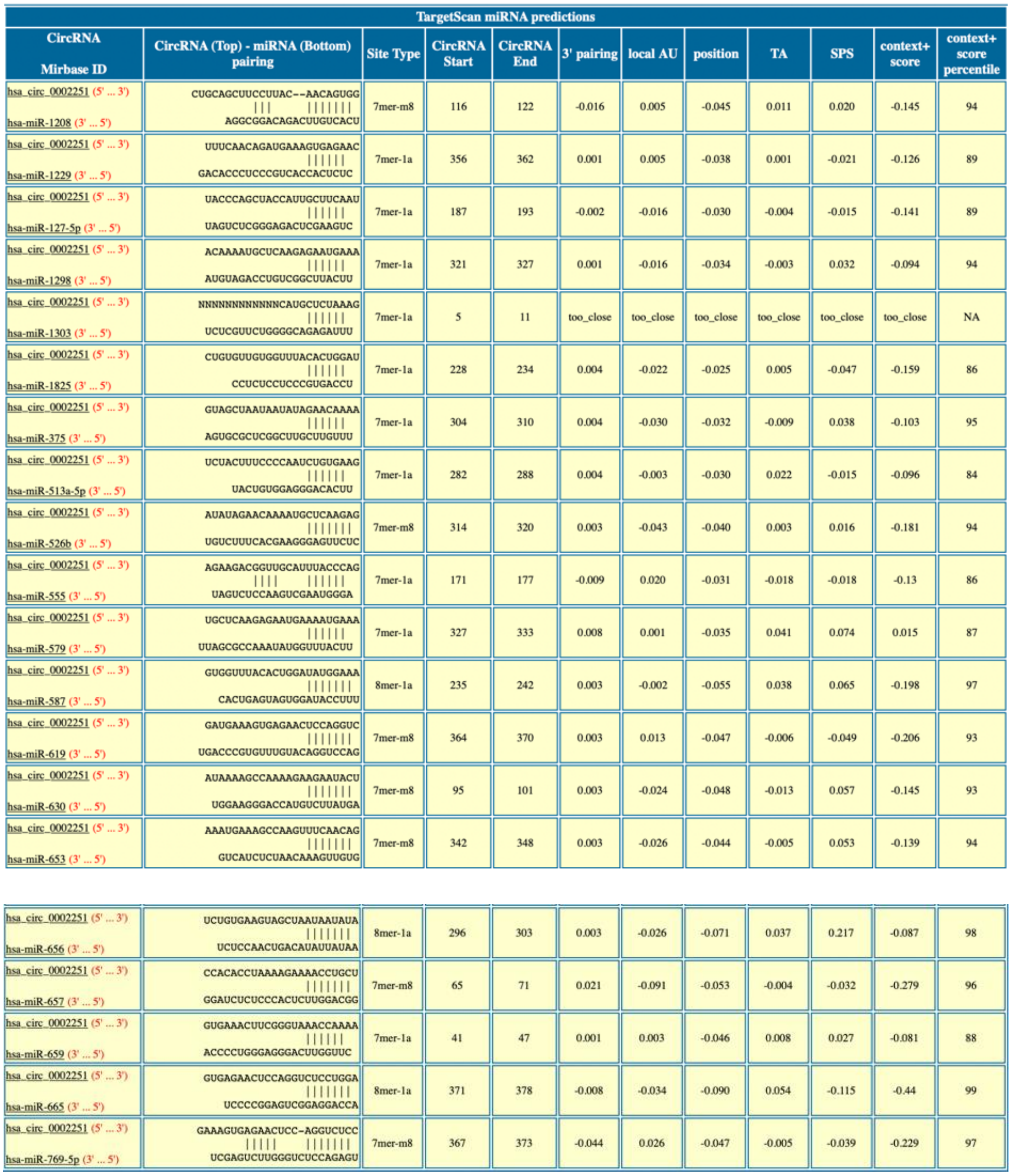
Table 4-8: MiR binding to SMN1 linear locus, assessed by miRwalkmiR binding to SMN1 linear locus, assessed by miRwalk

mirnaid	refseqid	genesymbol	duplex	start	end	bindingp	energy	seed	accessibility
hsa-miR-512-3p	NM_022874	SMN1	AAGTGCTGTCATAGCTGAGGTC#AGCTGTGGCTTCATTTA#(((((..((((((((......#))))))))..))))).	297	314	0.84615385	-17.1	0	0.0002836
hsa-miR-3120-5p	NM_022874	SMN1	CCTGTCTGTGCCTGCTGTACA#TGTTCCGGCGCGGCACAGGCCAGA#.(((((((((((((((.(((#)))..))))).))))))))).)).	221	245	0.84615385	-25.2	1	3.41E-05
hsa-miR-512-3p	NM_000344	SMN1	AAGTGCTGTCATAGCTGAGGTC#GGCCAGAGCGATGATTCTGACATTTGGGATGATACAGCACTG#.((((((((((.(((..((((#))))..))).)))).....................))))))).	92	134	0.84615385	-21.1	1	0.0121589
hsa-miR-512-3p	NM_000344	SMN1	AAGTGCTGTCATAGCTGAGGTC#GATAAAGCTGTGGCTTCATTTA#(((((..(((((((((...(((#)))..)))))))))..))))).	146	168	0.84615385	-18.3	0	0.0002836

4.3.5.4.3. miR binding on SMN exons in circRNAs

Using circInteractome, 20 miRs are reported to bear one binding site on C2B-3-4 (hsa-circ0002251): miR-1208, miR-1229, miR-127-5p, miR-1298, miR-1303, miR-1825, miR-375, miR-513a-5p, miR-526b, miR-555, miR-579, miR-587, miR-619 miR-630, miR-653, miR-656, miR-657, miR-659, miR-665 and miR-719-5p (***Table*** 4-9). 9 miRNAs are predicted to bind to C5-6 (hsa_circ_0072853): miR-1252, miR-1287, miR-198, miR-330-3p, miR-335, miR-338-3p, miR-625, miR-876-3p, and miR-944. Notably, C5-6 bears three binding sites for miR-876-3p (***Table*** 4-10).

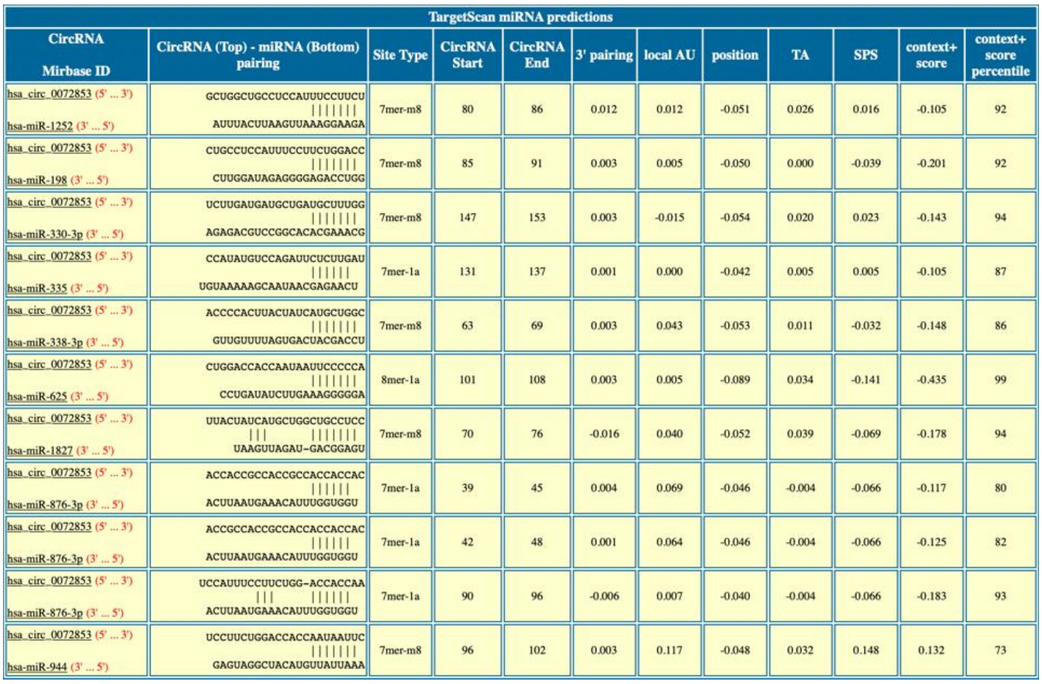
Table 4-9: MiR predicted to bind to C2B-3-4 (circInteractome)



TargetScan miRNA predictions

CircRNA Mirbase ID	CircRNA (Top) - miRNA (Bottom) pairing	Site Type	CircRNA Start	CircRNA End	3' pairing	local AU	position	TA	SPS	context+ score	context+ score percentile
hsa_circ_0002251 (5' ... 3') hsa-miR-1208 (3' ... 5')	CUGCAGCUUCCUUAC--AACAGUGG ||| ||||||| AGGCGGACAGACUUGUCACU	7mer-m8	116	122	-0.016	0.005	-0.045	0.011	0.020	-0.145	94
hsa_circ_0002251 (5' ... 3') hsa-miR-1229 (3' ... 5')	UUUCAACAGAUGAAAGUGAGAAC |||||| GACACCCUCCCGUCACCACUCUC	7mer-1a	356	362	0.001	0.005	-0.038	0.001	-0.021	-0.126	89
hsa_circ_0002251 (5' ... 3') hsa-miR-127-5p (3' ... 5')	UACCCAGCUACCAUUGCUUCAAU |||||| UAGUCUCGGGAGACUCGAAGUC	7mer-1a	187	193	-0.002	-0.016	-0.030	-0.004	-0.015	-0.141	89
hsa_circ_0002251 (5' ... 3') hsa-miR-1298 (3' ... 5')	ACAAAAUGCUCAAGAGAAUGAAA |||||| AUGUAGACCUGUCGGCUUACUU	7mer-1a	321	327	0.001	-0.016	-0.034	-0.003	0.032	-0.094	94
hsa_circ_0002251 (5' ... 3') hsa-miR-1303 (3' ... 5')	NNNNNNNNNNNNCAUGCUCUAAAG |||||| UCUCGUUCUGGGGCAGAGAUUU	7mer-1a	5	11	too_close	too_close	too_close	too_close	too_close	too_close	NA
hsa_circ_0002251 (5' ... 3') hsa-miR-1825 (3' ... 5')	CUGUGUUGUGGUUUACACUGGAU |||||| CCUCUCCUCCCGUGACCU	7mer-1a	228	234	0.004	-0.022	-0.025	0.005	-0.047	-0.159	86
hsa_circ_0002251 (5' ... 3') hsa-miR-375 (3' ... 5')	GUAGCUAAUAAUAUAGAACAAAA |||||| AGUGCGCUCGGCUUGCUUGUUU	7mer-1a	304	310	0.004	-0.030	-0.032	-0.009	0.038	-0.103	95
hsa_circ_0002251 (5' ... 3') hsa-miR-513a-5p (3' ... 5')	UCUACUUUCCCCAAUCUGUGAAG |||||| UACUGUGGAGGGACACUU	7mer-1a	282	288	0.004	-0.003	-0.030	0.022	-0.015	-0.096	84
hsa_circ_0002251 (5' ... 3') hsa-miR-526b (3' ... 5')	AUAUAGAACAAAAUGCUCAAGAG ||||||| UGUCUUUCACGAAGGGAGUUCUC	7mer-m8	314	320	0.003	-0.043	-0.040	0.003	0.016	-0.181	94
hsa_circ_0002251 (5' ... 3') hsa-miR-555 (3' ... 5')	AGAAGACGGUUGCAUUUACCCAG |||| |||||| UAGUCUCCAAGUCGAAUGGGA	7mer-1a	171	177	-0.009	0.020	-0.031	-0.018	-0.018	-0.13	86
hsa_circ_0002251 (5' ... 3') hsa-miR-579 (3' ... 5')	UGCUCAAGAGAAUGAAAAUGAAA |||||| UUAGCGCCAAAUAUGGUUUACUU	7mer-1a	327	333	0.008	0.001	-0.035	0.041	0.074	0.015	87
hsa_circ_0002251 (5' ... 3') hsa-miR-587 (3' ... 5')	GUGGUUUACACUGGAUAUGGAAA ||||||| CACUGAGUAGUGGAUACCUUU	8mer-1a	235	242	0.003	-0.002	-0.055	0.038	0.065	-0.198	97
hsa_circ_0002251 (5' ... 3') hsa-miR-619 (3' ... 5')	GAUGAAAGUGAGAACUCCAGGUC ||||||| UGACCCGUGUUUGUACAGGUCCAG	7mer-m8	364	370	0.003	0.013	-0.047	-0.006	-0.049	-0.206	93
hsa_circ_0002251 (5' ... 3') hsa-miR-630 (3' ... 5')	AUAAAAGCCAAAAGAAGAAUACU ||||||| UGGAAGGGACCAUGUCUUAUGA	7mer-m8	95	101	0.003	-0.024	-0.048	-0.013	0.057	-0.145	93
hsa_circ_0002251 (5' ... 3') hsa-miR-653 (3' ... 5')	AAAUGAAAGCCAAGUUUCAACAG ||||||| GUCAUCUCUAACAAAGUUGUG	7mer-m8	342	348	0.003	-0.026	-0.044	-0.005	0.053	-0.139	94
hsa_circ_0002251 (5' ... 3') hsa-miR-656 (3' ... 5')	UCUGUGAAGUAGCUAAUAAUAUA ||||||| UCUCCAACUGACAUAUUAUAA	8mer-1a	296	303	0.003	-0.026	-0.071	0.037	0.217	-0.087	98
hsa_circ_0002251 (5' ... 3') hsa-miR-657 (3' ... 5')	CCACACCUAAAAGAAAACCUGCU ||||||| GGAUCUCUCCCACUCUUGGACGG	7mer-m8	65	71	0.021	-0.091	-0.053	-0.004	-0.032	-0.279	96
hsa_circ_0002251 (5' ... 3') hsa-miR-659 (3' ... 5')	GUGAAACUUCGGGUAAACCAAAA |||||| ACCCCUGGGAGGGACUUGGUUC	7mer-1a	41	47	0.001	0.003	-0.046	0.008	0.027	-0.081	88
hsa_circ_0002251 (5' ... 3') hsa-miR-665 (3' ... 5')	GUGAGAACUCCAGGUCUCCUGGA ||||||| UCCCCGGAGUCGGAGGACCA	8mer-1a	371	378	-0.008	-0.034	-0.090	0.054	-0.115	-0.44	99
hsa_circ_0002251 (5' ... 3') hsa-miR-769-5p (3' ... 5')	GAAAGUGAGAACUCC-AGGUCUCC ||||| ||||||| UCGAGUCUUGGGUCUCCAGAGU	7mer-m8	367	373	-0.044	0.026	-0.047	-0.005	-0.039	-0.229	97

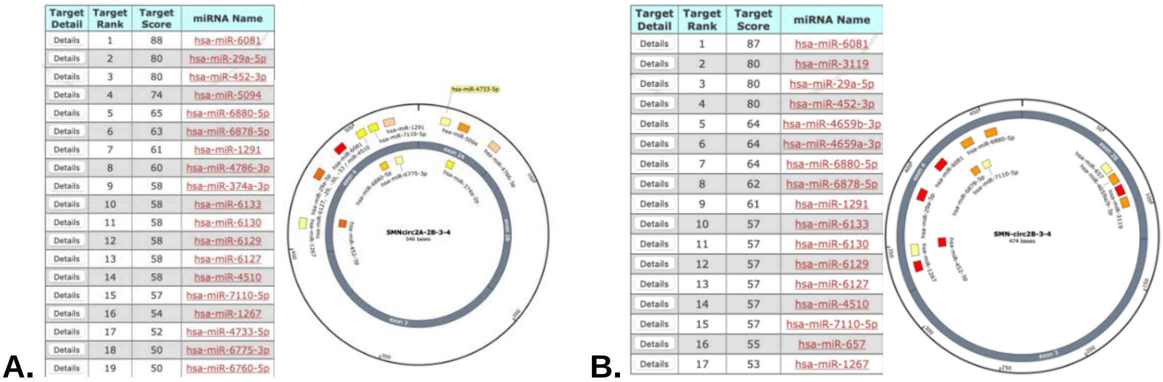
Table 4-10: MiR predicted to bind to C5-6 (circInteractome)



TargetScan miRNA predictions											
CircRNA Mirbase ID	CircRNA (Top) - miRNA (Bottom) pairing	Site Type	CircRNA Start	CircRNA End	3' pairing	local AU	position	TA	SPS	context+ score	context+ score percentile
hsa_circ_0072853 (5' ... 3') hsa-miR-1252 (3' ... 5')	GCUGGCUGCCUCCAUUUCCUUCU AUUUACUUAAGUUAAAGGAAGA	7mer-m8	80	86	0.012	0.012	-0.051	0.026	0.016	-0.105	92
hsa_circ_0072853 (5' ... 3') hsa-miR-198 (3' ... 5')	CUGCCUCCAUUUCCUUCUGGACC CUUGGAUAGAGGGGAGACCUGG	7mer-m8	85	91	0.003	0.005	-0.050	0.000	-0.039	-0.201	92
hsa_circ_0072853 (5' ... 3') hsa-miR-330-3p (3' ... 5')	UCUUGAUGAUGCUGAUGCUUUGG AGAGACGUCCGGCACACGAAACG	7mer-m8	147	153	0.003	-0.015	-0.054	0.020	0.023	-0.143	94
hsa_circ_0072853 (5' ... 3') hsa-miR-335 (3' ... 5')	CCAUAUGUCCAGAUUCUCUUGAU UGUAAAAAGCAAUAACGAGAACU	7mer-1a	131	137	0.001	0.000	-0.042	0.005	0.005	-0.105	87
hsa_circ_0072853 (5' ... 3') hsa-miR-338-3p (3' ... 5')	ACCCCACUUACUAUCAUGCUGGC GUUGUUUUAGUGACUACGACCU	7mer-m8	63	69	0.003	0.043	-0.053	0.011	-0.032	-0.148	86
hsa_circ_0072853 (5' ... 3') hsa-miR-625 (3' ... 5')	CUGGACCACCAAUAAUUCCCCCA CCUGAUAUCUUGAAAGGGGGA	8mer-1a	101	108	0.003	0.005	-0.089	0.034	-0.141	-0.435	99
hsa_circ_0072853 (5' ... 3') hsa-miR-1827 (3' ... 5')	UUACUAUCAUGCUGGCUGCCUCC UAAGUUAGAU-GACGGAGU	7mer-m8	70	76	-0.016	0.040	-0.052	0.039	-0.069	-0.178	94
hsa_circ_0072853 (5' ... 3') hsa-miR-876-3p (3' ... 5')	ACCACCGCCACCGCCACCACCAC ACUUAAUGAAACAUUUGGUGGU	7mer-1a	39	45	0.004	0.069	-0.046	-0.004	-0.066	-0.117	80
hsa_circ_0072853 (5' ... 3') hsa-miR-876-3p (3' ... 5')	ACCGCCACCGCCACCACCACCAC ACUUAAUGAAACAUUUGGUGGU	7mer-1a	42	48	0.001	0.064	-0.046	-0.004	-0.066	-0.125	82
hsa_circ_0072853 (5' ... 3') hsa-miR-876-3p (3' ... 5')	UCCAUUUCCUUCUGG-ACCACCAA ACUUAAUGAAACAUUUGGUGGU	7mer-1a	90	96	-0.006	0.007	-0.040	-0.004	-0.066	-0.183	93
hsa_circ_0072853 (5' ... 3') hsa-miR-944 (3' ... 5')	UCCUUCUGGACCACCAAUAAUUC GAGUAGGCUACAUGUUAUUAAA	7mer-m8	96	102	0.003	0.117	-0.048	0.032	0.148	0.132	73

4.3.5.4.4. *miR binding on the BSJ of SMN-derived circRNAs*

I used the miRDB algorithm on custom entries for the back splice junction exon 4-exon 2A (from C2A-2B-3-4) and for the BSJ exon 4-exon2B (from C2B-3-4) and mapped hits with a score ≥50% to the circular RNA's sequence. I found that none of the 19 miRNAs predicted binding sites to C2A-2B-3-4, and none of the 17 miRNAs predicted binding sites to C2B-3-4, associate with either circRNA's backsplice junction (***Figure*** 4-11).



A.

Target Detail	Target Rank	Target Score	miRNA Name
Details	1	88	hsa-miR-6081
Details	2	80	hsa-miR-29a-5p
Details	3	80	hsa-miR-452-3p
Details	4	74	hsa-miR-5094
Details	5	65	hsa-miR-6880-5p
Details	6	63	hsa-miR-6878-5p
Details	7	61	hsa-miR-1291
Details	8	60	hsa-miR-4786-3p
Details	9	58	hsa-miR-374a-3p
Details	10	58	hsa-miR-6133
Details	11	58	hsa-miR-6130
Details	12	58	hsa-miR-6129
Details	13	58	hsa-miR-6127
Details	14	58	hsa-miR-4510
Details	15	57	hsa-miR-7110-5p
Details	16	54	hsa-miR-1267
Details	17	52	hsa-miR-4733-5p
Details	18	50	hsa-miR-6775-3p
Details	19	50	hsa-miR-6760-5p

B.

Target Detail	Target Rank	Target Score	miRNA Name
Details	1	87	hsa-miR-6081
Details	2	80	hsa-miR-3119
Details	3	80	hsa-miR-29a-5p
Details	4	80	hsa-miR-452-3p
Details	5	64	hsa-miR-4659b-3p
Details	6	64	hsa-miR-4659a-3p
Details	7	64	hsa-miR-6880-5p
Details	8	62	hsa-miR-6878-5p
Details	9	61	hsa-miR-1291
Details	10	57	hsa-miR-6133
Details	11	57	hsa-miR-6130
Details	12	57	hsa-miR-6129
Details	13	57	hsa-miR-6127
Details	14	57	hsa-miR-4510
Details	15	57	hsa-miR-7110-5p
Details	16	55	hsa-miR-657
Details	17	53	hsa-miR-1267

Figure 4-11: MiR seeding sites identified by custom search for the BSJ on miRDB.org

miRNAs candidates (with target score >50) for binding on the custom BSJ sites of ***(A)*** *the* C2A-2B-3-4 *(19 identified potential miR interactors) and* ***(B)*** the *C2B-3-4 sequences (17 identified potential miR interactors), as ranked by the MirTarget algorithm.*

4.4. Discussion

This chapter presented three primary efforts to investigate circular RNAs in SMA. First, I investigated the expression of a subset of the *SMN*-derived circRNAs described in [85] in a SMA cellular model. Second, I aimed to develop a methodology for quantitative, streamlineable, sensitive, methodology to study circRNAs in a high-isoform context, specifically, a TaqMan™-based qPCR assay for *SMN*-derived circRNAs. Finally, I used existing computational pipelines to evaluate whether four *SMN*-derived circRNAs (C2A-2B-3-4, C2B-3-4, C3-4 and C6-7-8A) – chosen based on their reported expression levels – would be good candidates for exploration of their function (with regards to miR binding, RBP binding, and translation).

4.4.1. On the expression of *SMN*-derived circRNAs in SMA

I have successfully replicated most of the methodology applied by Ottesen and colleagues to isolate and amplify sequences from *SMN*-derived circular RNAs. The technical limitations in our laboratory have led us to deviate from the methodology of Ottesen et al. only by substituting detection of the PCR products by radioisotope-labelled nucleobases [83] with identification with intercalating dyes (ethidium bromide) of high detection threshold (0.5 to 5.0 ng/band in 3% agarose [446]).

This is the first report comparing the expression of *SMN*-derived circRNAs in cellular models of both type I and type II SMA ([397] focuses only on type I fibroblasts, without non-SMA control). I have fallen short of identifying the back-splice junctions by sequencing of the PCR amplicons, which is a function of (1) the low amounts purified from the isolated bands from PCR products and (2) the interruption of this project by the COVID19 pandemic. In the future, I will try repeating the PCRs and optimising the extraction of the PCR product either by stabbing the PCR product and amplifying it again using the same primers, following which it will be extracted by direct PCR purification; or by designing nested PCR primers.

The existence of *SMN*-derived circRNAs has been confirmed since the onset of this work by other approaches, including computational ones. While the first RNA-seq based identifications of circRNAs in the whole transcriptome uncovered only two *SMN*-derived circRNAs (C2B-3-4 and C5-6, respectively known as hsa-circ2251 and hsa-circ72853), one additional study by another group has confirmed the expression of part of the *SMN*-derived circRNA repertoire [397].

Importantly, for circRNAs others than *SMN*-derived circRNAs, tissue-specific expression has been reported as particularly abundant in the central nervous system viz. in neuronal cells. This high expression has been linked to (1) preferential back-splicing of genes involved in neural processes [392,447], (2) increased expression of circRNAs in mammalian brains compared to other tissues [394,409,417], and (3) lack of clearance mechanisms (viz. cell-division related dilution processes) in non-proliferating, fully differentiated neurons [392,393,411]. Future computational work to confirm the existence of the whole of the diverse repertoire of *SMN*-derived circRNAs not only in SMA fibroblasts but also in SMA neurons and MN is required and could be based on the recently published isoform quantification of circRNAs based on long read Nanopore sequencing [448] – e.g. if conducted on SMA iPSC-MNs.

4.4.2. On the functions of *SMN*-derived circRNAs

I have found that *SMN*-derived circRNAs bear some seeding sites for microRNAs, notably three binding sites for miR-876-3p on C5-6. miR-876-3p has been reported to regulate cell proliferation and glucose metabolism, and is known to be involved in other circRNA/miR/mRNA regulatory loops [449–452].

Crucially, all of the miRs I have identified as binding to C2A-2B-3-4, C2B-3-4, C3-4 and C6-7-8A have also been confirmed and reported in a follow-up paper published by Ottesen and colleagues [453]. I have not identified new miR binding sites on the BSJ in C2A-2B-3-4 nor in C2B-3-4. (Future work will replicate this approach for both C3-4 and C6-7-8A). None of the *SMN*-derived circRNAs I have studied bear the same number of miR seeds as the circRNAs that have miR sponging abilities. For example, the circRNA CDR1as bears 130 sites for miR-7 [416]. As such and given the number of circRNAs in the *SMN* circular transcripts repertoire, it appears unlikely that *SMN*-derived circRNAs have a role in complex gene regulation networks involving microRNA pool regulation.

I used two tools to study whether C2A-2B-3-4, C2B-3-4, C3-4 and C6-7-8A bear binding sites for RBP. The first tool, circInteractome, is limited to circRNAs have been reported in RNA-seq studies – and thus, for SMN, to C2B-3-4 and C5-6. I therefore used the recently reported RBPsuite which allows for input of given circular sequences. Using RBPsuite, I found that *SMN*-derived circRNAs bear verified binding sites to six RNA-binding proteins: FOXP2, IGF2BP1-3, LIN28A, and PTB.

FOX2, also known as FOXP2, is a transcription factor expressed in the brain. FOXP2 regulates gene networks implicated in neurite outgrowth during brain development [454]. It is also known to downregulate the expression of *SRPX2*, which reduces dendritic spine length and shape [455]. Interestingly, the expression of FOXP2 dramatically increases in early post-natal motor neuron development, peaking at PND21 [456].

The IGF2BP proteins 1-3 (also known as IF2B1) are involved in RNA transport and increase RNA stability [457]. Loss of function of LIN28A has been associated with Parkinson's disease [458]. It has also been reported as essential in the differentiation program of skeletal muscle, by increasing the translation of IGF2 and MYOD1 mRNAs to which it binds [459].

At this stage, we do not have the tools to quantify the differences of expression of C2A-2B-3-4, C2B-3-4, C3-4 and C6-7-8A between healthy and SMA patients. Should there be a significant difference and should the association of C2A-2B-3-4, C2B-3-4, C3-4 and C6-7-8A with those proteins be verified by cross-linking and immunoprecipitation (CLIP), FOXP2 represents an interesting avenue of investigation.

Finally, I have identified that several *SMN*-derived circRNAs bear open reading frames. It is at this stage unknown whether they are indeed translated to small peptides. Internal ribosomal entry sites (IRES) can only be assessed experimentally and do not bear a consensus sequence [460]. It remains to be assessed *in vitro* whether *SMN*-derived circRNAs produce small peptides, and what function, if any, those would bear.

4.4.3. Methodological contributions to the field of circRNAs

Only a handful of circRNAs from the wide repertoire of *SMN*-derived circRNAs are commonly reported RNA-seq data deposited since the onset of this study: C2B-3-4 and C6-7-8A in [448], C2B-3-4 and C5-6 in [389]. A custom-based approach like Ottesen and colleagues therefore still has its place. While [85] and [397] use SYBRGreen approach for the quantification of *SMN*-derived circRNAs, I have not been able to replicate this method in our laboratory due to the several isoforms amplified within each reaction. I speculate that the successful SYBR-based amplification of certain circRNAs may be explained by ultra-low expression levels of the circular RNAs sharing sequences with them (low enough, indeed, to be below their threshold-of-detection after 30 PCR cycles).

Other studies – particularly in the early days of circRNA research – have used TaqMan™ assay, and have successfully amplified circRNAs derived from *HPK3*, *ASXL1*, *KIAA0182*, and *LPAR1* [389]. I take advantage of existing tools for primer design (IDT-DNA PrimerQuest tool) and of the possibility to customise primer and probe design over given junctions. Remaining work includes to quantification of efficiency, specificity, and range of detection of the primers I have designed.

Circular RNAs are generated in *Alu*-rich sequences via processes of backsplicing. The unspecific nature of backsplicing yields to the likelihood that *SMN* is not the only locus with a vast repertoire of circRNAs. Indeed, a recent novel computational pipeline estimates that 24.6% of circRNAs bear multiple isoforms [461]. As such, a qPCR methodology able to resolve between highly homologous circRNAs isoforms can be useful to study of other circRNA repertoires.

Finally, another perspective must be underlined. *SMN1/2* are one of the human genome loci richest in *Alu* elements [397]. As pointed out before, RNA-seq based studies repeatedly fail to reveal the richness of the repertoire of circRNAs generating from them. It is therefore crucial to be able to develop sensitive quantitative methods to estimate to what degree the number of published RNA-seq studies on circRNAs under-estimate the amount of existing circular RNAs, and further elucidate what effect *SMN*-derived circRNAs have on the transcription of *SMN1* and *SMN2* derived linear transcripts.

4.4.4. Future directions

Once the TaqMan assay™ has been validated for sensitivity, efficiency, and range of detection, we would like to pursue quantitative analysis of the four circRNAs most highly expressed in the initial report of their repertoire in SMA iPSC MNs and in WT iPSC MN.

Furthermore, the Taiwanese SMA mouse model bears a human *SMN2* gene. Whether *SMN*-derived circRNAs can be detected in Taiwanese mice tissue is of importance to assess the relevance of circRNAs as a biomarker of theranostic interest for splice modifying therapies.

Finally, the future improvement of analytical methods for circRNAs will enable further studies on effect of the reduced spliceosomal function in SMA beyond the *SMN*-derived circRNA landscape, in the genome-wide circRNA landscape.

4.5. Conclusion

In conclusion, the investigation of circular RNAs derived from the *SMN1/2* loci has revealed another layer of complexity of the *SMN* gene transcriptional regulation. The transcriptome of *SMN* is more complex than what has been studied in Chapter 2 (coding and linear non-coding transcriptome) and Chapter 3 (linear coding transcriptome). This chapter has clearly underlined the methodological issues behind the lack of identification of new therapeutic targets within the *SMN*-derived circRNA repertoire. Whether *SMN*-derived circRNAs bear therapeutic or theranostic interest remains to be explored in further studies.

5. Chapter 5: Pharmacological targeting of Foxo3-dependent muscle atrophy

5.1. Introduction

In the past three result chapters, I have focused the identification and exploration of new therapeutic targets on the neuronal manifestations of SMA. As outlined in Chapter 1 however, SMA is a neuromuscular disorder in which the muscular atrophy is not only MN degeneration-dependent, but also linked to intrinsic myogenesis defects. In this chapter, I explore muscle atrophy independently of motor neuron defects – that is, as a targetable phenotype of its own. In particular, I suggest two oligonucleotide-based approaches to target a transcription factor critical in atrophy and present splice-switching therapeutic candidates with potential to modulate muscular atrophy in SMA.

5.1.1. Skeletal muscle atrophy: definition and classification

Muscle atrophy is defined as a reduction in muscle mass and in muscle function [462]. A classification system for muscular atrophy was established during the 2009 Sarcopenia Consensus Conference held in Rome, Italy, resulting in three widely defined types of muscle atrophy: sarcopenia, cachexia, and disuse atrophy [463] (***Table*** 5-1). However, there is an increasing movement towards pathway-based classification of muscle atrophy [464].

Table 5-1: Classification of skeletal muscle loss processes

	Sarcopenia-related atrophy	Cachexia-related atrophy	Disuse atrophy
Associated condition	Aging	Underlying illness and inflammation: cancer, AIDS, COPD, etc.	Mechanical unloading resulting from joint inactivity (bedrest) and/or neural activity (disease, spinal cord injury)
Clinical presentation	(1) Low muscle strength (2) Low muscle quantity or quality (3) Low physical performance	Variable definition: weight loss ≥5% (resp. 10% [465]) in six (resp. 12 [463]) months with other symptom associated: fatigue, early satiation, anorexia [466]	
Metabolic processes	Increase in protein degradation and decrease in protein synthesis	Increase in protein degradation	Decreased rate of muscle protein synthesis
Reviewed in:	Consensus on definition and diagnosis in [467].	[468]	[469,470]

5.1.2. Molecular processes involved in skeletal muscle atrophy

Muscle metabolism relies on the balance between anabolic and catabolic systems [471,472]. The anabolic systems in muscle [473] depend on: endocrinal control (both endocrine factors, e.g. anabolic steroids like testosterone, and general anabolic endocrine factors, e.g. insulin [474]) [475,476], growth factor-dependent translational [477–479] and transcriptional control (in particular, the IGF1-dependent IGF-1R/PI3K/Akt pathway [480,481]) [482,483], mechanotransduction [484,485], and satellite cell-related stem cell anabolism [486–488]. Catabolic systems in muscle [489] depend on the ubiquitin protease system (UPS) [490,491], the autophagy-lysosome system [492,493], and other proteolytic processes either upstream (including calpains [494,495] and caspases [496,497]) or downstream (tripeptidyl-peptidase II [498] and aminopeptidases [499]) of the UPS.

Some of the catabolic molecular actors are specific to skeletal muscle atrophy and are therefore called atrogenes. Atrogenes are defined as "gene[s] which mRNA is up-or down-regulated during the atrophying phase and which return to normal levels during recovery" [500]. Atrogenes from the proteolytic systems include striated-muscle specific E3 ligases: *Atrogin-1* (*MAFbx, Fbxo32*), *MuRF-1* (*Trim63*), and *Fbxo40* [501]. Atrogenes are translated to atroproteins. A recent proteomic meta-analysis shows that the atroproteins Fbxo32 and Trim63 are upregulated in both disuse atrophy and cancer-related cachexia [311].

5.1.3. The role of FoxO factors in regulating atrogenes and skeletal muscle atrophy

Foxo proteins are part of the large 'Forkhead box' family of transcription factors (which in humans encompasses more than a hundred proteins from FOXA to FOXR [502]). The O class of Fox proteins is characterised by IGF-1/PI3K/Akt pathway-dependent inhibition and downregulation after Akt-dependent direct phosphorylation [502,503]. In mammals (both mice and human), there are four members in the Foxo/FOXO sub-family: *Foxo1/FOXO1* (formerly known as *FHKR*) [504,505], *Foxo3/FOXO3a* (formerly *FHKRL1*) [506], *Foxo4/FOXO4* (formerly *AFX*) [507,508], and *Foxo6/FOXO6* [509].

Each Foxo/FOXO proteins can be phosphorylated by Akt/AKT (and more incidently, by other AGC family kinases) at three (two for Foxo6/FOXO6) conserved Thr/Ser residues [503]. Despite debates regarding their effect on DNA-binding abilities [510], phosphorylation of FoxO proteins at these sites is thought to mediate binding with 14-3-3, leading to the export of phospho-Foxo/FOXO-14-3-3 from the nucleus to cytoplasm [511]. Regardless of the mechanism, phosphorylation of Foxo/FOXO by Akt/AKT depletion prevents Foxo/FOXO activity as a transcription factor.

Foxo/FOXO proteins have numerous transcriptional targets and regulate a vast array of cellular processes, including apoptosis initiation (via transcriptional activation of the pro-apoptotic factors FasL and Bim [511–513]), cell-cycle arrest (via the upregulation of p27kip1 [513] and GADD45 [514]), stress resistance (upregulation of catalase genes[515]), DNA damage repair (upregulation of GADD45 [514] and DDB1 [516]), glucose metabolism and energy homeostasis [517], cell differentiation of multiple cell types [518–520], and muscle atrophy. Foxo1/FOXO1 and Foxo3/FOXO3a regulate muscle atrophy by inducing the autophagy–lysosome and ubiquitin–proteasome systems [521,522] including transcription of atrogenes including *Trim63* and *MAFbx* [522,523].

Thus, FoxO protein expression and function are of interest when studying muscle atrophy in SMA.

5.1.4. Muscle atrophy in SMA

SMA clinical presentation varies depending on subtype severity. It presents with typically proximal [524] and typically symmetrical muscle weakness, fatigue (though not detectable physiologically [525]), and fatigability [526]. Muscle weakness is most severe in hip adductors, hip flexors, and quadriceps [526]. In addition to the role of SMN in myoblast physiology [166,167], SMA symptom onset and functional decline is caused by the progressive denervation of muscle at the neuromuscular junctions (NMJ) [156,158]. Molecular atrophy pathways in SMA are FoxO1- and FoxO3- dependent, with increased expression of the E3 ligases *MAF-bx* (four-fold upregulation) and *Trim63* in both *Smn*$^{2B/-}$ and *Smn*$^{-/-}$*;SMN2* mice models [527]. In a Taiwanese SMA mouse model, the *Trim63* mRNA is upregulated (logFC=2.4) at early symptomatic stage in tibialis anterior muscle (reanalysis of exon microarray from [313]).

5.1.5. Rationale for targeting muscle atrophy by pharmacological inhibition of Foxo1/3 in SMA

Muscle has been a very early target in SMA therapeutic development [528]. Current therapeutic efforts to pharmacologically target skeletal muscle atrophy include the search for molecules enhancing muscle contractability (including reldesemtiv (CK-2127107, Cytokinetics®)[529] and amifampridine (Catalyst Biopharmaceuticals®)[530]), as well as muscle mass regulators (apitegromab (SRK-015, ScholarRock)[235], a selective inhibitor of the activation of latent myostatin).

Because muscle atrophy is a condition shared across numerous diseases, multiple attempts have been conducted to target it with pharmacological tools [531], including anabolic drugs, enzyme inhibitors, anti-inflammatory drugs [532], anti-oxidants [533], and more targeted approaches, e.g. myostatin inhibition [234,240,534,535].

Down-regulation of the ubiquitin proteasome system is protective against muscle atrophy. Knockout of *MAFbx* and of *Trim63* [536] in mice does not yield a visible phenotype but protects from denervation induced atrophy[472]. The double knockout of *Trim63* and of *Trim55* (*Murf2*, coding for a protein contribution to sarcomere assembly) in mouse causes muscle hypertrophy [537], as does the knockout of *Fbxo40* [538]. Overexpression of *Foxo1* in mice shows reduced skeletal muscle mass and increases skeletal protein degradation in a *Mafbx-, Trim63-* (ubiquitine ligases) and *cathepsin L-* (lysosomal protease) dependent manner [539]. When conditionally overexpressed in the heart, Foxo3 induces cardiac atrophy via autophagic mechanisms [540]. Temporary downregulation of *Foxo3* reduces autophagy in muscle [541]. In all these studies, the deletion of Foxo factors is permissive for normal muscle function [521].

FOXO proteins have been investigated for therapeutic development beyond muscle atrophy applications, though the mediation of their activity by binding to either other proteins or to DNA makes them not easily druggable by small molecules [542]. I therefore proceeded to investigate oligonucleotides to target Foxo proteins. In the mouse skeletal myoblast cell line C2C12, antisense RNA molecules targeting the mouse *Foxo-1* mRNA sequence significantly downregulate *Foxo-1* mRNA expression by 81% compared with the control [543].

5.1.6. Hypotheses and goals

In this chapter, I present work aiming at developing new pharmacological tools to down-regulated FOXO factors and reduce skeletal muscle atrophy. I hypothesise that interfering with the splicing of *Foxo3* pre-mRNA down-regulates Foxo3 levels and can reduce skeletal muscle atrophy.

I aim at identifying the nucleotide sequence best suited for a splice-switching approach to Foxo3 downregulation, and at presenting molecular evidence of the effect of such an approach on the ubiquitin-proteasome pathway. I also aim at presenting fundamentals for another strategy for FOXO3 downregulation.

5.2. Materials and Methods

5.2.1. Pip8b2-PMO compounds

Foxo3-targeting splice-switching sequences were designed to aim for: an optimal length of 20 nt, Tm of at least 48°C, GC% between 40% and 60%, avoiding stretches of three or more G/C, using SnapGene software v.6.1.1. (www.snapgene.com), RNA Folding Form (www.unafold.org), and ESE-Finder and RESCUE-ESE tools (following [544]). Melting temperatures were estimated using OligoAnalyzer (www.idtdna.com) and the Oligonucleotide Proprieties Calculator (www.cail.cn).

Morpholino oligonucleotides (PMO) were purchased from Gene Tools. Peptide synthesis reagents and solvents purchased from Merck, SigmaAldrich and FluoroChem. All solvents are HPLC grade and used without further purification. Synthesis, conjugation, and purification of Pip8b2del01 (RBRR YQFLI RBRXRB-COOH, with X= 6-aminohexanoic acid and B= beta-alanine) [545] to the *Foxo3* targeting PMO sequences were conducted following the protocol outlined in the recent patent WO2020030927A1 [339]. Pip8b2-PMOs were diluted in nuclease-free water.

5.2.2. Cell culture

5.2.2.1. Cells

C2C12 immortalised mouse myoblasts (ATCC cat. no. CRL-1772™) are maintained in DMEM (1X) + GlutaMax-1 supplemented with 20% FBS and 1% P/S ("growth medium"). The "differentiation medium" consists of DMEM (1X) + GlutaMax-1 supplemented with 1% horse serum and 1% penicillin/streptomycin (PS), and is changed every 48h. Cells were maintained between p6 and p13 passages, and were allowed to grow for a full passage before being split for experimental work.

Cells are frozen down in 50% FBS, 10% DMSO, 40% growth medium, using a Mr. Frosty™.

5.2.2.2. Atrophy induction

Glucocorticoids such as dexamethasone (DEX) can increase the transcription of atrogenes and model muscle atrophy *in vitro* [522]. Dexamethasone is dissolved in ethanol and further diluted in differentiation medium to a final concentration of 100 μM. Medium is changed every 24h. Control cells are incubated with 0.03% (v/v) ethanol in differentiation medium.

5.2.2.3. Treatment regimen

Cells were seeded in 24-well plates at 2x10E5 cells/well (0.5 mL/well) or in 6-well plates at 3x10E5 cells/well (2.0 mL/well) and grown for 24h in growth medium before differentiation. Cells differentiated for 72 hours were treated in biological triplicates for 48h with daily medium change (Pip8b2-PMO ± dexamethasone 100 μM) and optical microscope viability monitoring.

5.2.3. RNA processing and analyses

RNA was collected 24h or 48h after treatment onset and extracted using Maxwell® RSC simplyRNA Tissue Kit (Promega, cat#AS1340). cDNA was prepared from 500 ng RNA with the High Capacity cDNA Kit (Life Technologies) according to the manufacturer's instructions. After 1:5 dilution, the cDNA template was amplified on a StepOnePlus Real-Time PCR Thermocycler (Life Technologies) with SYBR Green Master Mix (Applied Biosystems, cat no#4385612) or TaqMan™ Gene Expression Master Mix (ThermoFischer Scientific, cat. no. 4369016) using primers listed in ***Table*** 5-2 (ordered from Integrated DNA Technologies (IDT)).

qPCR data was analyzed using the StepOne Software v2.3 (Applied Biosystems). Relative gene expression was quantified using the Pfaffl method [546]. Primer efficiencies were calculated with the LinRegPCR software [547]. Gene expression was normalised against *PolJ*.

Table 5-2: Primers used to study Foxo1, Foxo3, and E3 ubiquitin ligases

Target (gene)	sequence
mFoxo1_ex1-2 (unskipped)	fwd: 5' CGT GCC CTA CTT CAA GGA TAA G 3' rev: 5' TCC TTC ATT CTG CAC TCG AAT AA 3' probe: /56-FAM/TT CGC CAC A/Zen/A TCT GTC CCT TCA CA/3IABkFQ/
mFoxo1_ex2-3 (total)	fwd: 5' CAT GGA TGG AGA TAC CTT GGA TT 3' rev: 5' AAT GTA GCC TGC TCA CTA ACT C 3' probe: /56-FAM/AT GTG TTG C/Zen/C CAA CCA AAG CTT CC/3IABkFQ/
mFoxo3_ex3-4 (total)	fwd: 5' AAG CAG ACC CTC AAA CTG AC 3' rev: 5' CAA AGG TGT CAA GCT GTA AAC G 3' probe: /56-FAM/TG CTC TCA G/Zen/C AAG TGG ACA GTG AT/3IABkFQ/
mFoxo3_ex1-3 (skipped)	fwd: 5' AGG AGC CGA GAA CTC CAT 3' rev: 5' TTG CCC GTG CCT TCA TT 3' probe: /56-FAM/TC CCT GCA C/Zen/A GCC GCT TCA T/3IABkFQ/
MuRF-1 (*Trim63*)	fwd: 5′ - CTGCAGAGTGACCAAGGAGA - 3′ rev: 5′ - TTTTCCAGCTGCTCCCTG - 3′
Mafbx/atrogin-1 (*Fbxo32*)	fwd : 5′ - ACTGTCTGTTCTGTAGGCCC - 3′ rev : 5′ - CCCCACCCCAGGAATTAACT - 3′
PolJ	5' ACC ACA CTC TGG GGA ACA TC 3' 5' CTC GCT GAT GAG GTC TGT GA 3'

5.2.4. Protein processing and analyses

Total protein was collected from 6-well plates by adding 100 µl lysis buffer (***Table*** 5-3) to cell pellets, pipetting up and down, puting on ice for 20 min on a shaker, centrifuging at 15,000 x*g*, 10 min. Protein lysates were aliquoted at 4°C and stored at -20°C.

Table 5-3: Lysis buffer composition for protein extraction

Composition	reference/catalogue number	for 250 µL
RIPA buffer	ThermoFisher Scientific cat. 89901	180.5 µl
Triton X-100 10% in aqueous solution	SigmaAldrich cat. no. 93443	2.5 µL
Phosphatase Inhibitor Mini Tablets (10X)	ThermoFisher cat. no. A32957	25 µL
cOmplete mini, EDTA-free protease inhibitor cocktail (7X)	Roche cat. no. 11836153001	38 µL
phenylmethylsulfonyl fluoride (PMSF protease inhibitor) (0.1 M)	ThermoFischer cat. 36978	4 µL

Proteins were quantified with the bicinchoninic acid assay (Lowry protein assay) [548], using the DC™ Protein Assay Reagent, (BioRad cat. no. 500-0113) and a BSA standard curve. Samples were resuspended to 20 µg per sample in Laemmli Sample Buffer 4X (BioRad) and ran on 10% Mini-PROTEAN® TGX Stain-Free™ Protein Gels (Bio-Rad Cat no. 4568033) alongside the Full Range Rainbow MW Marker (Cytiva Life Sciences, cat. no. RPN800E) in Novex™ Tris-Glycine/SDS Running Buffer (ThermoFischer cat. no. LC2675) at 200 V for 40 minutes. Gels were transferred to a PVDF 0.45 µm membrane using the Mixed setting (2.5A, 25V, 7 min) of the Trans-Blot® Turbo™ Transfer System (BioRad, cat. no. 1704150) and the Trans-Blot Turbo Midi PVDF Transfer Packs (1704157). Equal transfer of proteins was assessed using FastGreen for 20 minutes (then washed 3x 10' using FastGreen washing solution). Membranes were blocked for 1 hour using Odyssey Blocking Buffer (Li-COR, USA). Primary antibodies (***Table*** 5-4) were incubated overnight at 4°C in Odyssey Blocking Buffer. Membranes were washed 3x 10' with T-BST before incubation of the secondary antibodies at 1:1,000 at room temperature for 1 hour (***Table*** 5-4) in TBS-T. Blots were imaged and analysed using the C-DiGit Blot Scanner and Image Studio Digits software (LI-COR).

Table 5-4: Antibodies used for Western blots

Protein	Antibody	Dilution
FOXO3A (ca. 82 kDa)	Anti-FOXO3A (T32) antibody (ab26649)	1: 500
mouse GAPDH (37kDa)	primary GAPDH (D4C6R) Mouse mAb (Cell Signaling #97166)	1: 200

5.2.5. Statistical analyses

Results were displayed (mean ± SEM) and analysed using Prism v.9.3.1 (GraphPad software, San Diego, CA). Differences between groups were compared using either multiple t-tests or one-way ANOVA and corrected *post-hoc* for multiple comparisons. Differences were considered significant at alpha = 5%.

5.3. Results

5.3.1. Validation of the cellular models

I began by validating that the differentiated C2C12 'myotubes' are a relevant cellular model for the study of Foxo-induced atrogene-related mechanisms. C2C12 mouse myoblasts can be differentiated to myotube-like structures upon reduction of serum concentration from 20% FBS to 2% horse serum. Fused myotubes are clearly visible upon optic microscopy after 72h in differentiation medium (***Figure*** 5-1 ***A***). Reduction in number and in size of the myotubes, which mimics muscle atrophy, sets on 72h of treatment with 100 µM dexamethasone (***Figure*** 5-1 ***B***).

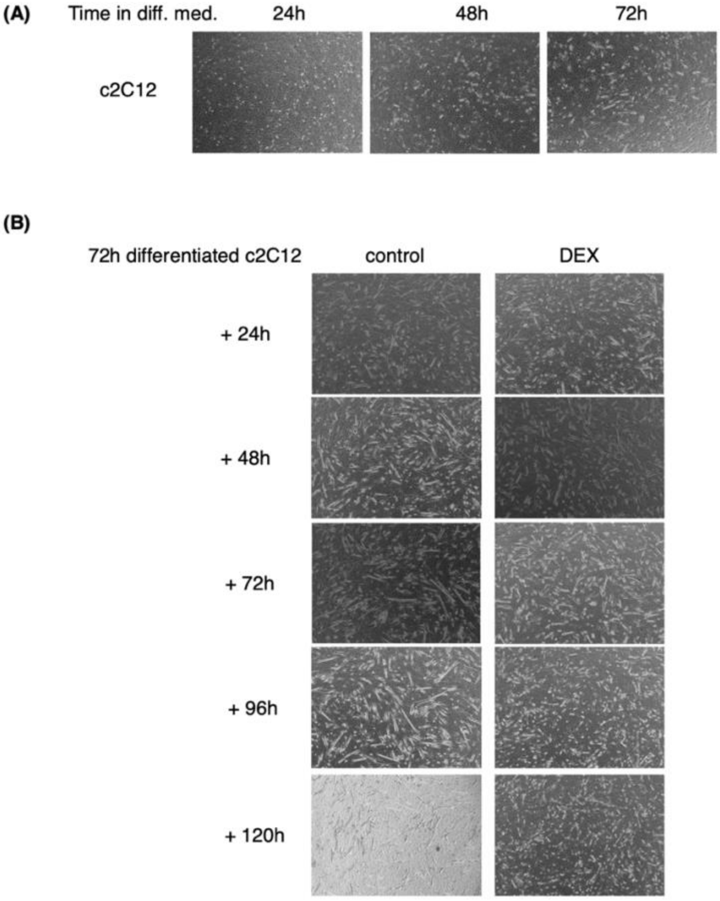
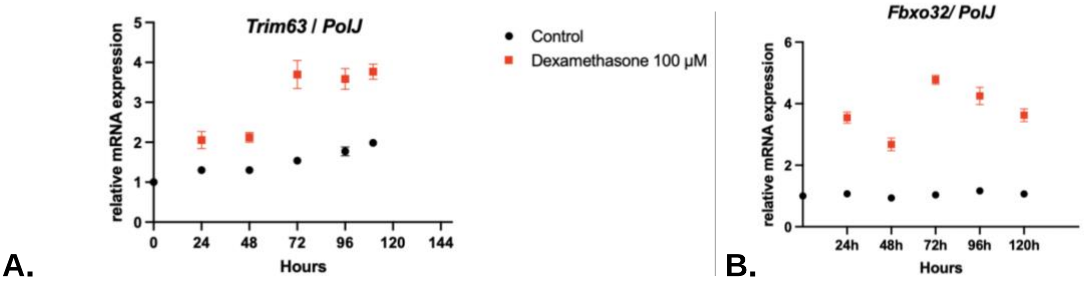


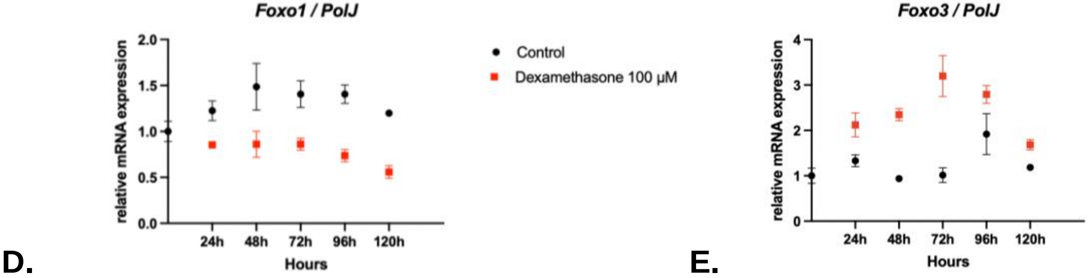
Figure 5-1: Dexamethasone treatment changes the morphology of differentiated C2C12 myotubes

(A) *Optic microscopy imaging of C2C12 cells differentiated for 24h, 48h, and 72h by reducing the serum concentration from 20% FBS to 2% horse serum. Upon differentiation, cells adopt myotube-like structures (polynucleated after fusion, elongated).* ***(B)*** *Optic microscopy imaging of C2C12 cells differentiated for 24h, 48h, 72h, 96h, up to 120h, in either 2% horse serum or 2% horse serum + 100 µM dexamethasone (DEX). DEX treatment reduces the number and the size of myotube-like cells, thus mimicking muscle atrophy.*

At the molecular level, transcription of the atrogenes *Trim63* and *Fbxo32* in differentiated c2c12 myotubes is significantly upregulated after 48h of dexamethasone treatment (adj.P.value = 0.043 and 0.046 respectively), with fold increases of 2.5 and 4.56, respectively. This correlates with a significant three-fold increase of *Foxo3* transcription after 48h of dexamethasone treatment (adj.P.value = 0.011). Interestingly, *Foxo1* transcription is not significantly affected by dexamethasone treatment at 72h (adj.P.value = 0.24) (***Figure*** 5-2).



C.	*Trim63/PolJ*				*Fbxo32/Polj*		
	mean(control)	mean(DEX)	adj.P.value		mean(control)	mean(DEX)	adj.P.value
0h	1.000				1.000		
24h	1.303	2.057	0.296532		1.073	3.547	0.012239
48h	1.303	2.123	0.042985		0.9367	2.680	0.046994
72h	1.540	3.697	0.117731		1.037	4.777	0.005686
96h	1.773	3.587	0.054708		1.167	4.250	0.031950
120h	1.987	3.767	0.053652		1.067	3.627	0.029386



F.	*Foxo1/PolJ*				*Foxo3/Polj*		
	mean(control)	mean(DEX)	adj.P.value		mean(control)	mean(DEX)	adj.P.value
0h	1.000				1.000		
24h	1.227	0.8533	0.31		1.330	2.123	0.39
48h	1.487	0.8600	0.59		0.9400	2.347	0.01
72h	1.407	0.8600	0.24		1.013	3.197	0.14
96h	1.407	0.7367	0.04		1.920	2.797	0.91
120h	1.200	0.5600	0.02		1.187	1.683	0.22

Figure 5-2: Dexamethasone treatment of differentiated C2C12s recapitulates skeletal muscle atrophy features.

Relative expression (normalised to PolJ*), as measured by qPCR, of a time-course experiment of DEX (100 µM) treatment of C2C12 differentiation myotubes. Cells were treated with DEX for 0h, 24h, 48h, 72h, 96h, or 120h. Transcripts of interest include atrogenes Trim63* ***(A)*** *and Fbxo32* ***(B)*** *and the Foxo1* ***(D)*** *and Foxo3* ***(E)*** *TFs. Red: DEX 100µM. Black: untreated myotube-like C2C12s.* ***(C), (F)****: Tables summarising the mean expression level in each timepoint group, and the adjusted p-value (unpaired t test with Welch correction, Bonferroni-Dunn method, alpha = 5%).*

In light of the three-fold increase of total *Foxo3* expression in differentiated myotubes, I therefore continued all experiments with a 72h initial differentiation, followed by a 72h period of dexamethasone atrophy induction.

5.3.2. Design and synthesis of Foxo3-targeting compounds

I proceeded with the design, synthesis, and screening of PMO oligonucleotide sequences to induce splice-switching changes in the 6,787 bp-long *Foxo3* transcript variant 1 (NM_001376967) or 6,783 bp-long *Foxo3* transcript variant 2 (NM_019740) that would either introduce early stop codons or generate mature mRNA transcripts lacking essential Foxo3 domains. We designed five sequences in exon 2 of *Foxo3,* named Foxo3-(A) to (E). Of note, the 5' end of Foxo3-(E) targets exon 2 whilst its 3' end targets a splicing site in the flanking intron. The sequences of the five sequences designed have been withheld from this thesis due to upcoming patent application (***Figure*** 5-3). Splicing out of exon 2 yields a truncated Foxo3(ex1-3-4) transcript that cannot encode for the DNA-binding domain (within Forkhead) and cannot be phosphorylated at the threonine T32.

A.	length (nt)	sequence targeted	GC content (%)	Tm (IDTdna.com)	Tm (cail.cn)
Foxo3-(A)	25	exon 2	56%	62.8	61
Foxo3-(B)	25	exon 2	56%	63.0	61
Foxo3-(C)	25	exon 2	64%	66.3	64
Foxo3-(D)	25	exon 2	52%	61.5	59
Foxo3-(E)	22	exon 3 (in 5')	60%	60.0	59

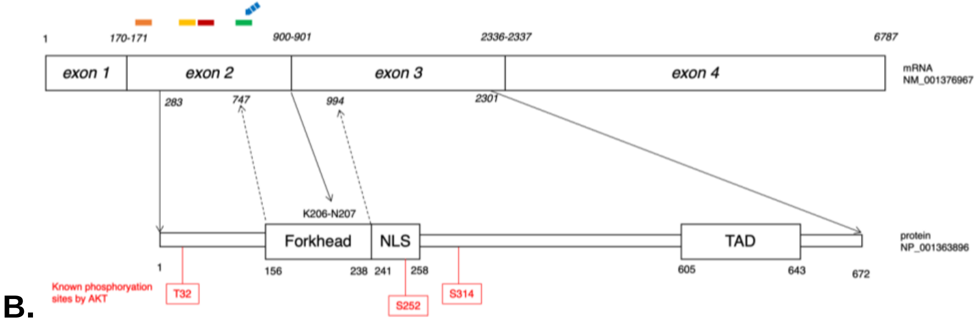


Figure 5-3: FoxO3 transcript and protein characteristics defined splice-switching sequence design

(A) *Nucleotide composition characteristics of the five initial Foxo3 splice-switching candidates, with the melting temperature Tm assessed by two different algorithms (IDTdna.com and cail.cn).* ***(B)*** *Schematic (not to scale) of the transcript and protein organisation, including know threonine (T) and serine (S) phosphorylation sites and functional domains. The coloured rectangles above exon 2 represent the location of the Foxo3-(A) to –(E) sequences.*

During this initial sequence-screening stage, I focused our efforts on finding a sequence region targetable for the induction of pharmacological skipping of exon 2. All sequences bear different GC% content and are of different lengths, making a single scrambled sequence highly time- and resource- ineffective at this stage of the process. Therefore, we did not include a unique scramble sequence in the initial study design.

PMO(Foxo3-A to Foxo3-E) moities were then conjugated to the Pip8b3del01 cell penetrating peptide, purified, and resuspended in water for *in vitro* assays (***Supplementary Table*** 7-8).

5.3.3. Pip8b2-PMO(Foxo3) candidates affect Foxo3 splicing, atrogene levels, and cell morphology

I proceeded to screen all the Pip8b2-PMO(Foxo3) conjugates for their efficacy on exon 2 skipping in *Foxo3*, for any downstream effects on *Trim63* and *Fbxo32*, and for their effect (if any) on the homologous sequences of *Foxo1*. C2C12 myoblasts were differentiated for 72h before being treated for either 24h or 48h by 8b2-Foxo3 candidates at 5 μM.

8b2-(A) has the best exon 2 skipping efficiency: after 48h of treatment, up to 1.5% of all mFoxo3 isoforms have skipped exon 2 ($p<0.0001$). 8b2-(A) does not significantly change the total amount of *mFoxo3* mRNA after 48h of treatment. Interestingly, whilst all 8b2-PMO(Foxo3) candidates upregulate total Foxo1 levels after 24h, 8b2-(A) is the only candidate for which this effect is transient. All candidates, include 8b2-(A), have a *Foxo3*-specific effect on splicing and do not affect *mFoxo1* splicing at either 24h nor 48h. 8b2-(A) down-regulates of *Trim63* mRNA levels in the dexamethasone-treated c2c12 myotubes to levels similar to control myotubes at 48h ($p=0.0006$). It also downregulates *Fbxo32* levels from DEX at 48h, though not significantly ($p=0.48$) (***Figure*** 5-4).

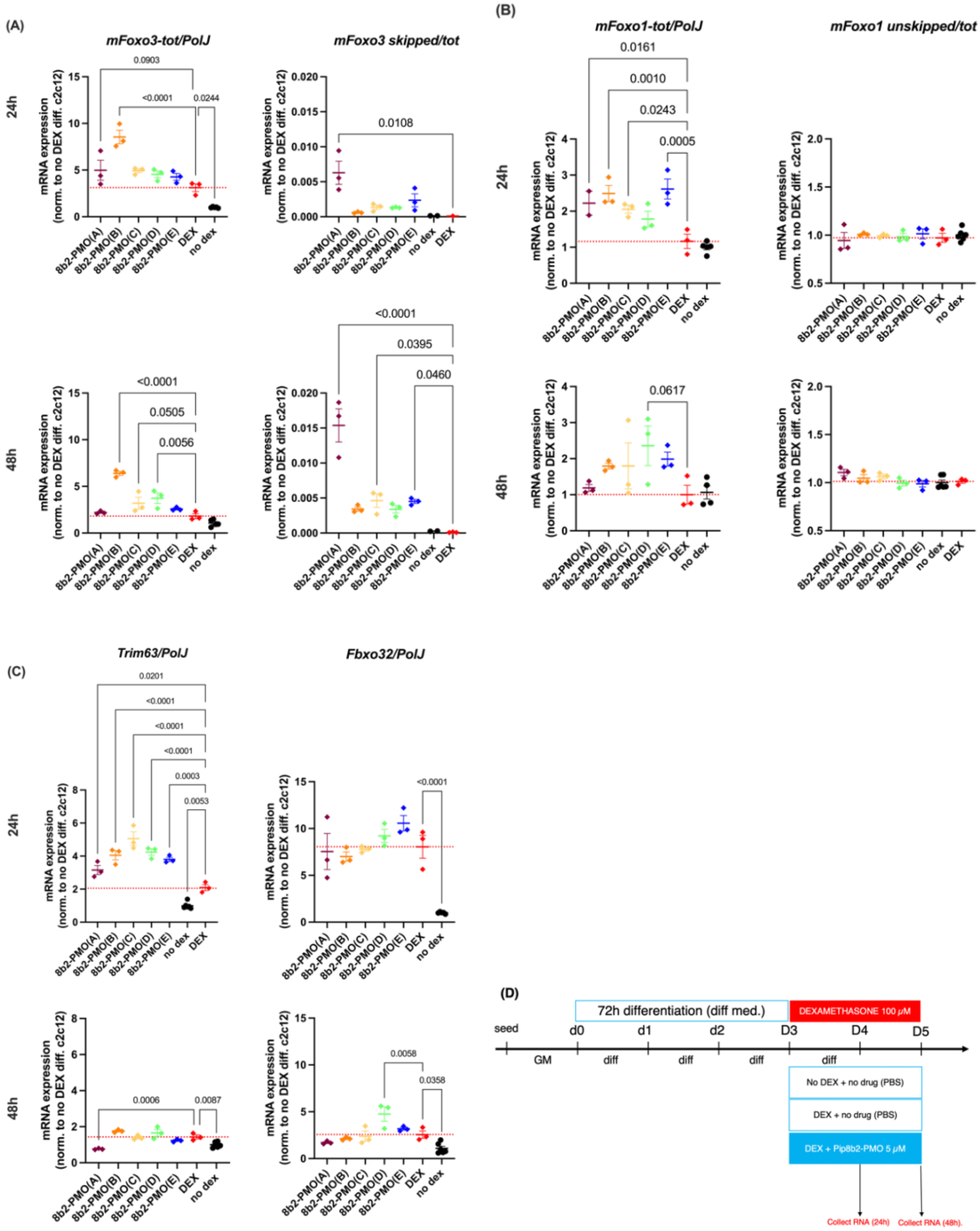


Figure 5-4: Initial screening of PPMO sequences in differentiated C212 myotubes.

Quantitative RT-PCR *results of the expression changes of total (left half of the panel) and spliced (right half of the panel) forms of **(A)*** mFoxo3, ***(B)*** mFoxo1, *and of **(C)** the total amount of atrogenes* Trim63 *and* Fbxo32, *after 24h (top half of each panel) or 48h (bottom half of each panel) treatment with 5μM 8b2-PMO candidates, in differentiated C2C12 myotubes. Statistical analysis: one way ANOVA, Dunnett's multiple comparisons test comparing to DEX, alpha = 5%. **(D)** Schematic of the experimental design.*

8b2-(C) is a secondary positive hit for further exploration, with the second-best exon 2 skipping efficiency 48h after treatment (up to 0.5% of all transcripts exclude exon 2, adj.pval = 0.0395).

Comparison of dexamethasone + PBS treated c2c12 myotubes with dexamethasone + 8b2-PMO myotubes suggests a certain toxicity of the 8b2-PMO compound at the 5 µM concentration. Interestingly, 8b2-(A) and 8b2-(B) both seem to be the least affected by this toxicity and best maintain the remaining c2c12 myotubes (***Figure*** 5-5). As such, these results need to be validated using histochemistry.

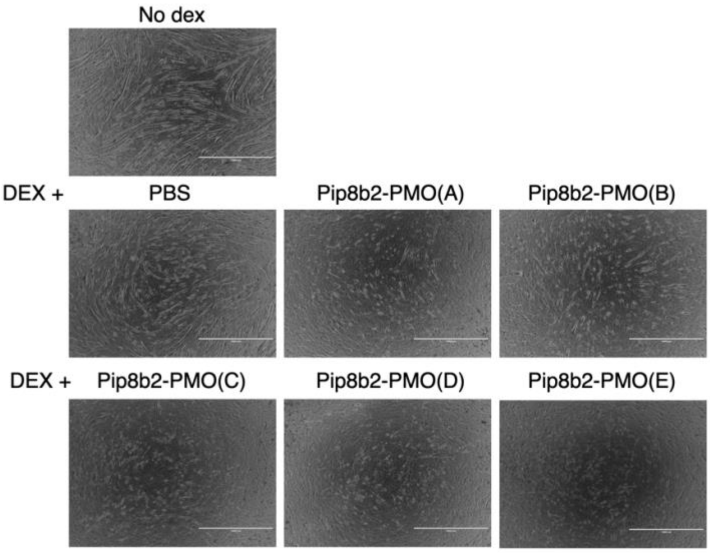


Figure 5-5: Treatment with Pip8b2-PMO(Foxo3) candidates affect c2c12 morphology.

Representative optic microscopy imaging of differentiated C2C12 cells, treated with DEX (lower two lines), treated with each Pip8b2-PMO candidate at 5 µM for 48h. Scale bar: 1,000 µm.

5.3.4. Dose-response study on the leading Pip8b2-PMO candidate

The first screening identified two candidates for further study: 8b2-(A) and 8b2-(C) based on amounts of *mFoxo3* skipped after 48h treatment. I further investigated the dose-response of *Foxo3*, *Foxo1*, *Trim63* and *Fbxo32* splicing and transcription at lower doses. Due to limiting factors in amounts of 8b2-(C) at the time of this study, I started with 8b2-(A).

8b2-(A) decreases the inclusion of *Foxo3* exon 2 splicing from 250 nM, with a hundred-fold increase when compared to non-treated DEX c2c12 myotubes ($p<0.0001$), in a dose-dependent fashion. The compared effect on *mFoxo1* exon skipping is minimal (within the same order of magnitude, vs. the *Foxo3* hundred-fold exon 2 skipping increase).

At these lower doses, the correlation between increased splice changes in *mFoxo3* and atrogene transcription initially observed at 5 µM disappears: treatment with 125 nM or 250 nM of 8b2-(A) moderately increases the transcription of *Trim63* (up to x1.5 fold) and *Fbxo32* (up to x2.2 fold) (***Figure*** 5-6).

Finally, at these lower doses (and contrary to what was observed at 5 µM), 48h of treatment with 8b2-(A) increases the total transcription of *Foxo3* (up to x1.5 fold) (***Figure*** 5-6). However, this does not correlate to an increase in FoxO3 protein levels, where FoxO3 seems to decrease in a non-significant manner after 8b2-PMO(A) treatment (***Figure*** 5-7).

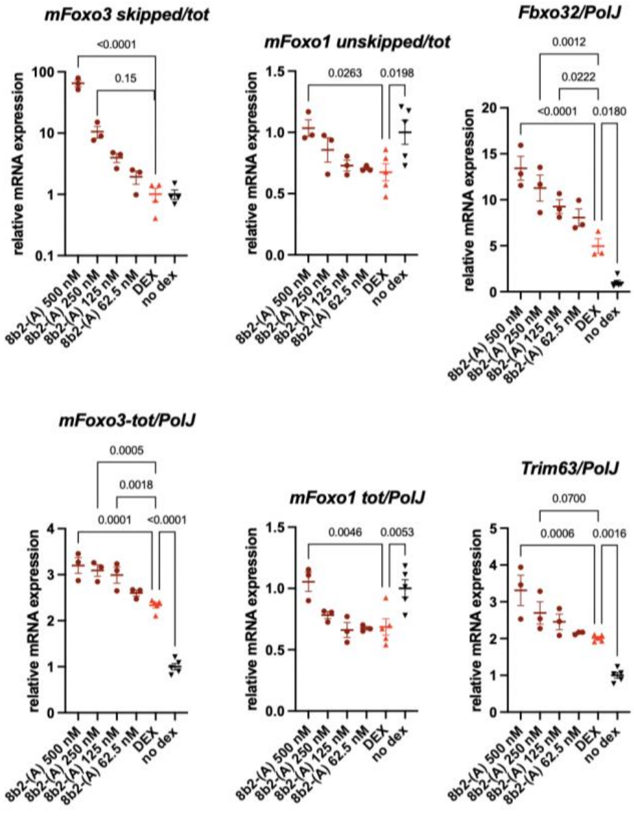


Figure 5-6: Gene expression changes in dose-response to 48h treatment with 8b2-PMO(A)

Quantitative RT-PCR results of the expression changes of total and spliced (right half of the panel) forms of mFoxo3, mFoxo1*, and of the total amount of atrogenes* Trim63 *and* Fbxo32*, after 48h treatment with increasing doses (62.5 nM to 500 nM) of the 8b2-PMO(A) candidate, in differentiated C2C12 myotubes. Statistical analysis: one way ANOVA, Dunnett's multiple comparisons test comparing to DEX, alpha = 5%.*

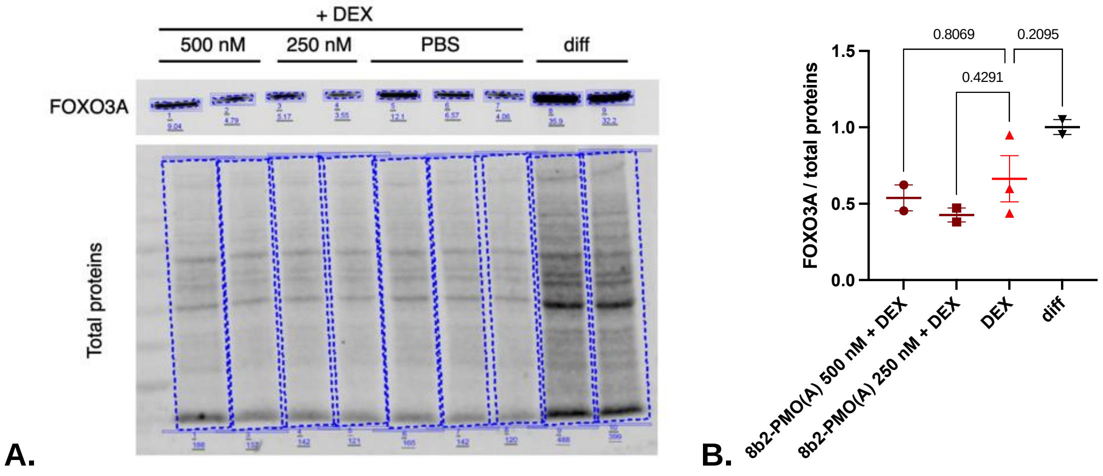


Figure 5-7: 8b2-PMO(A) does not increase FoxO3 protein levels

(A) *FOXO3A-centred (top) and full length, total protein (bottom) Western Blot images following the SDS-PAGE of total cell lysates extracted from differentiated C2C12 treated or not with 100 µM DEX, treated or not with 250 nM or 500 nM of the 8b2-(A) candidate.* ***(B)*** *Quantification of FOXO3A (normalised to total proteins). No comparison is statistically significant. Western Blot: n=1, on 2-3 technical replicates.*

Further experimental work in this first pharmacological approach to FOXO3 downregulation for muscular atrophy will be discussed later on in this chapter and will include: (i) validation of these results with additional replicates; (ii) immunohistochemistry experiments to more precisely quantify effects on myotube morphology; (iii) dose-response studies on 8b2-(C); (iv) the design of second-generation PMO sequences in the region targeted by Foxo3-(A) and Foxo3-(C); and (v) the investigation and characterisation of the splicing regulators in *mFoxo3* and *mFoxo1* transcripts.

5.3.5. Perspective: PROTAC compounds to pharmacologically downregulate FOXO3

5.3.5.1. Rationale for the development of a PROTAC system against FOXO3

In parallel to the development of a exon-skipping approach to downregulate levels of *Foxo3* mRNA and of Foxo3 downstream atrogenes, and to the identification of Pip8b2-(A) and –(C) as promising (albeit first generation) candidates, I pursued the design of another pharmacological strategy aiming at Foxo3 protein levels themselves. In the following section, I briefly summarise our rationale, our approach, and future experimental plans.

PROTAC (proteolysis targeting chimera) is an event-driven pharmacological class of bifunctional molecules constraining a target of choice into a ternary complex with E3 ligases and leads to the target's 26S proteasomal degradation [549,550]. PROTAC systems consist of three moieties covalently linked: (1) a moiety targeting the protein of interest; (2) a linker; and (3) the E3-ligase addressing moiety, also called 'degron'.

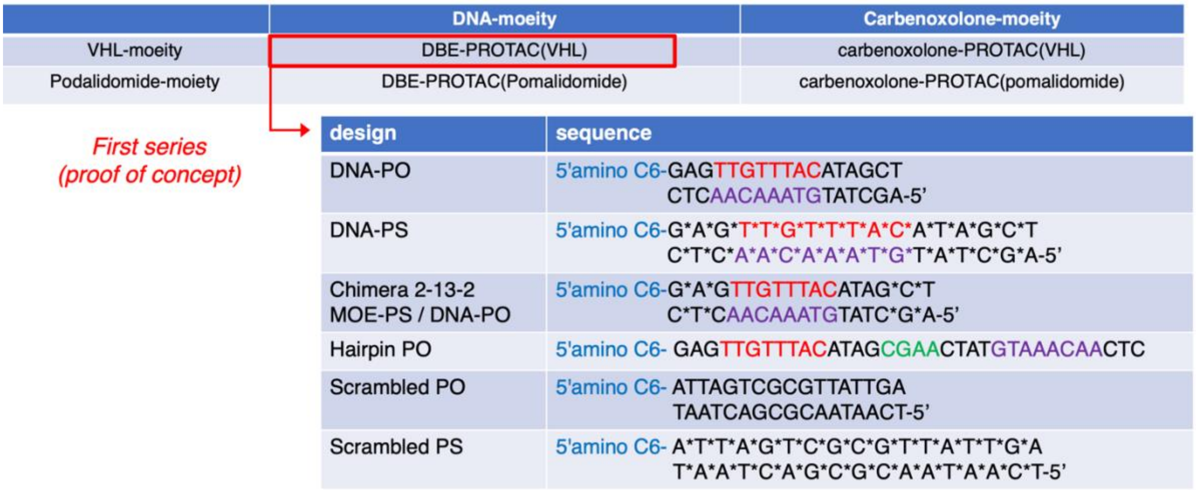
I designed two types of FoxO3-targeting PROTAC systems, which target FoxO3 using either carbenoxolone [551], a small-molecule ligand to FoxO3; or an oligonucleotide-based moiety targeting the consensus DBE DNA-binding domain [510,552]. Each of those systems is to be tested with two different known degrons: pomalidomide – a ligand to the E3 ubiquitin ligase Cereblon, CRBN [553]; and von-Hippel Lindau – ligand to the Cullin2 RING E3 ligase Cullin2, CRL2$^{\mathrm{VHL}}$ [554] (***Table*** 5-5).

5.3.5.2. Design considerations

I designed the oligonucleotide-based DBE-VHL series first, as a proof of concept.

To balance oligonucleotide stability with DNA/protein affinity, I designed both phosphodiester and phosphothioate candidates, integrated base modifications (MOE) in one of the candidates, and experimented with different modalities of double-stranded DNA binding (hairpin vs. duplex) (***Table*** 5-5). Sequence length was defined based both on the highly evolutionarily conserved [555] DBE consensus motif (8 nucleotides: 5'-TTGTTTAC-3' [556]) and on the necessity of a Tm greater than 37°C to allow for double-stranded DNA binding in a cellular or animal system. I therefore extended the 13-mer DNA strand used in the crystallisation of human FOXO3a DNA-binding domain (PDB structure: 2UZK), as the DNA major groove interacts with aminoacids from Forkhead's Helix 3 (H3), Wing 1, and C-terminal coil (notably, Ser253, the target of Akt phosphorylation [557]. (Of note, human and murine FoxO3 DNA binding element DBE is 100% homologous [557] (***Supplementary Figure*** 7-29)).

Table 5-5: Design of FoxO3-PROTAC targeting series



	DNA-moeity	Carbenoxolone-moeity
VHL-moeity	DBE-PROTAC(VHL)	carbenoxolone-PROTAC(VHL)
Podalidomide-moiety	DBE-PROTAC(Pomalidomide)	carbenoxolone-PROTAC(pomalidomide)

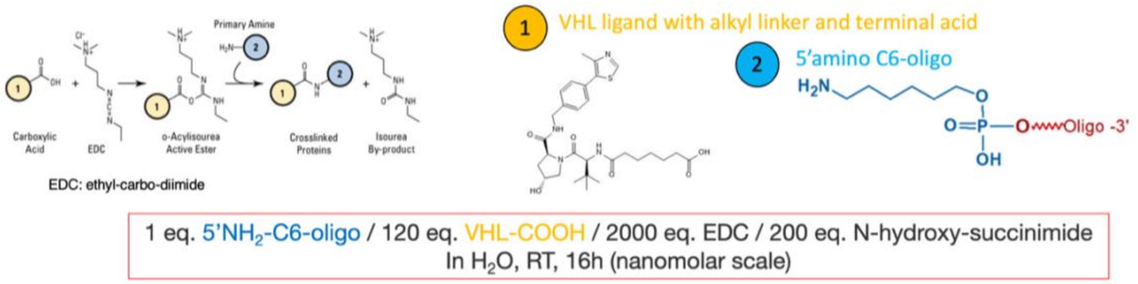
First series (proof of concept)

design	sequence
DNA-PO	5'amino C6-GAGTTGTTTACATAGCT CTCAACAAATGTATCGA-5'
DNA-PS	5'amino C6-G*A*G*T*T*G*T*T*T*A*C*A*T*A*G*C*T C*T*C*A*A*C*A*A*A*T*G*T*A*T*C*G*A-5'
Chimera 2-13-2 MOE-PS / DNA-PO	5'amino C6-G*A*GTTGTTTACATAG*C*T C*T*CAACAAATGTATC*G*A-5'
Hairpin PO	5'amino C6- GAGTTGTTTACATAGCGAACTATGTAAACAACTC
Scrambled PO	5'amino C6- ATTAGTCGCGTTATTGA TAATCAGCGCAATAACT-5'
Scrambled PS	5'amino C6- A*T*T*A*G*T*C*G*C*G*T*T*A*T*T*G*A T*A*A*T*C*A*G*C*G*C*A*A*T*A*A*C*T-5'

5.3.5.3. Synthesis of the first series of Foxo3(DBE)-VHL PROTAC system

The VHL ligand with alkyl linker and terminal acid VH032 amide-alkylC5-acid (Cat. No. 7217) was purchased from Tocris Bioscience (Bristol, UK). Synthetic modified oligonucleotides were ordered from Integrated DNA Technologies (Coralville, IA), at the 25 nmol scale (or 100 nmol scale for oligos with modified nucleobases), with standard desalting. Oligonucleotides from IDT were spun for 3 minutes at 2,000 x*g* at 4°C and resuspended in pure water to a concentration of 100 µM. Following vortexing and spinning on a table-top centrifuge, the concentration of oligonucleotides was determined by ultraviolet absorption at 260 nm wavelength.

DNA-PROTAC compounds were conjugated at the nanomolar scale, in liquid phase, at room temperature, for 16 hours, following the protocol outlined by [558], with the following stoichiometry: 1 eq. 5'NH_2-C6-oligo / 120 eq. VHL-COOH / 2000 eq. EDC / 200 eq. N-hydroxy-succinimide (***Figure*** 5-8). DNA-PROTAC compounds were desalted using the Glen Gel-Pak™ 1.0 desalting column (cat. no. 61-5010-XX), following manufacturer's instructions, with IDTE (10 mM Tris, 0.1 mM EDTA) pH=8 (cat. no. 11-01-02-05, IDT, Iowa) as the buffer.



1 eq. 5'NH$_2$-C6-oligo / 120 eq. VHL-COOH / 2000 eq. EDC / 200 eq. N-hydroxy-succinimide
In H$_2$O, RT, 16h (nanomolar scale)

Figure 5-8: Schematic of chemical conjugation of the Foxo3(DBE)-VHL PROTAC system

The aqueous phase conjugation of the PROTAC system exploits classic carbodiimide crosslinker chemistry. The carboxylic acid moiety at the end of the alkyl linker conjugated to the VHL ligand (item (1), in yellow) is activated by EDC to an ester intermediate (O-acylisourea active ester). A nucleophilic attack from the primary amine on the 5' end of the C6-oligo (item (2), in blue) yields an amide bond between (1) and (2), and a soluble urea EDC by-product.

5.3.5.4. Preliminary synthesis and quality control results

Using 5'-OH-oligonucleotide as a negative control, we show that the amide coupling is specific to the 5' amine (and does not yield VHL-nucleobase products) as well as quantitative (***Figure*** 5-9). Delay between the end of the reaction and the LC-MS run may explain EDC adducts; and further quality controls is required before column desalting, HPLC purification, and *in vitro* assays.

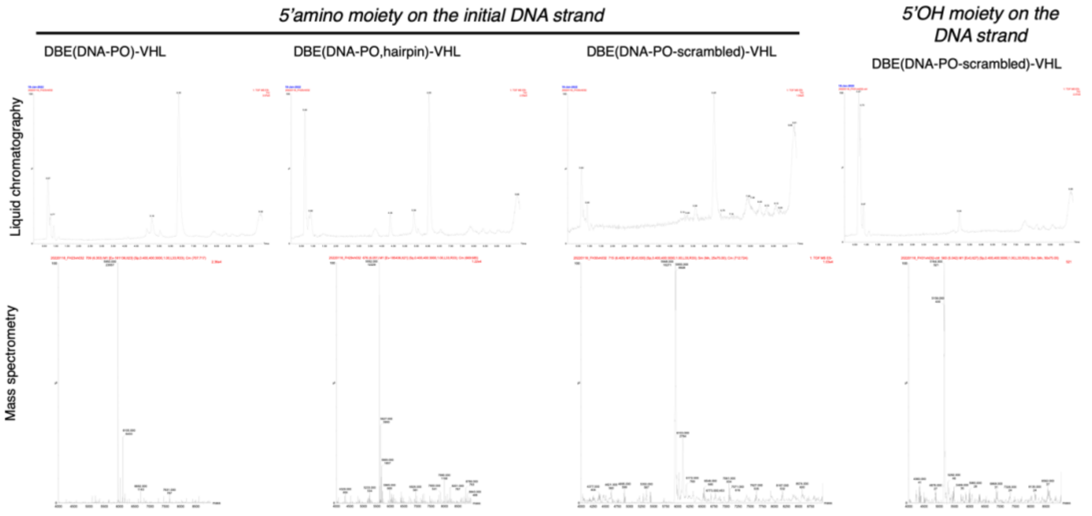


Figure 5-9: LC-MS spectra confirm specific conjugation of the VHL ligand to the 5'end of the DNA moiety

LC/MS analysis of the VHL amide-coupled DNA phosphodiester (PO) strands. Top: liquid chromatograms. Bottom: MS spectra. From left to right, the VHL moiety was coupled with: DBE(DNA, PO, single strand, 5' amino moiety); DBE(DNA, PO, hairpin, 5' amino moiety); DBE(DNA, PO, single strand, scrambled sequence, 5' amino moiety); and DBE(DNA, PO, single strand, 5' OH moiety). The presence of a 5' amino moiety on the PO strand is necessary to obtain the secund, larger chromatography peak characteristic of a longer retention, and therefore of coupling.

5.3.5.5. Future experimental work

PROTAC systems have seen considerable extension of their targets since the first report of PROTAC in 2001 [559]. Oligonucleotide-based PROTAC has been developed for RNA-binding proteins [560]. A DNA-approach to target the DBD of FoxO3 is promising: two cancer-related transcription factors have successfully been targeted by DNA-based PROTAC systems [561].

Further experimental work on Foxo3(DBE-VHL) PROTAC system will include measuring FOXO3 protein levels (as well as intermediary phenotypic atrophy readouts at the mRNA level: *Trim63* and *Fbxo32*) in dexamethasone-treated atrophied C2C12 myotubes, 48h after lipofectamine transfection of 10 nM, 100 nM, and 1 μM PROTAC compounds. If successful in decreasing FOXO3 protein levels and improving visible myotube phenotype, translation to a human *in vitro* model (human myoblasts) will be facilitated by the aforementioned perfect murine-human DBE-sequence homology.

5.4. Discussion

5.4.1. Further steps in experimental development of the 8b2-Foxo3 candidates

I began this work with aim to investigate whether an exon-skipping approach could be used to pharmacologically target Foxo3 in muscular atrophy. I have identified two candidates for further development, targeting contiguous sequences in *Foxo3* exon 2. The first dose response study on 8b2-(A) shows dose-dependent exon skipping with doses as low as 250 nM. This system includes a cellular delivery system (the Pip8b2del01 moeity) and does not require any transfection to cells. Further work is required to (1) improve the measurement of phenotypic changes upon treatment, (2) refine sequence targeting withing *mFoxo3* transcript for improved downstream effect specificity, and (3) clarify the unexpected functional preliminary findings of the initial dose-response experiments.

First, further work will ensure better molecular characterisation of phenotypic changes upon treatment. The current evaluation of effects of the 8b2-PMO system on C2C12 atrophied myotubes relies on visual estimations of cell morphology parameters, including number of total myotubes and the area they cover. Besides the addition of replicate experiments and the study of dose-response studies on 8b2-(C), further experimental work will include immunohistochemistry staining of 8b2-(A) and -(C) treated cells for the contractile protein myosin heavy chain MHC.

Next, further work will refine and optimise the sequence targeted by the PMO. Foxo3-(A) and Foxo3-(C) are contiguous in the *Foxo3* exon 2 sequence. The design of second-generation PMO sequences will therefore focus on this locus. These compounds will also yield additional insight in the characterisation of the underlying splicing regulating sequences in *mFoxo3* and *mFoxo1* transcripts. Particularly, such studies may contribute to the mechanistic understanding behind *Foxo1* and *Foxo3* transcription upregulation at lower doses (250 nM-500 nM).

Therefore, further work will also resolve the current discrepancy between the current preliminary findings and the original hypothesis: namely, that at lower doses (500 nM and under), the current dose-response experiment with 8b2-(A) show the upregulation of atrogene (*Fbxo32* and *Trim63*) relative mRNA amounts upon increased levels on skipped *Foxo3* transcripts. In that experiment, the increase of skipped *mFoxo3* is accompanied by an unexplained increase in total *mFoxo3* levels. This could be a first explanation for the increased transcription of downstream Foxo3 effectors. (Of note, this increase is specific to the lower dosing specimen and had not been observed at 5 µM dosing). Interestingly, within the dose-response experiment, *mFoxo1* expression patterns also differ in DEX-only treated control myotubes from their counterparts in the 5 µM experiment. This suggests the possibility of a batch-specific artifact and calls for the repeating of this experiment.

Should an increase of atrogene mRNA expression be confirmed upon repeated experiments, I raise two additional hypotheses: (1) crosstalk between splicing of *mFoxo3* transcripts and their transcription process; and (2) self-regulation of *Foxo3* transcription by *Foxo3* transcript pool and/or FoxO3 protein. Transcription by RNA polymerase II and pre-mRNA splicing are intricately related processes [562–564] with splicing being transcription-rate dependent [564], and relying on tertiary parameters including the local epigenetic markers [565]. Whether the occupation by 8b2-PMO(Foxo) of a putative exon skipping inhibitor sequence on nascent *Foxo3* pre-mRNAs affects RNA polymerase II rates remains to be explored. Alternatively, there is a conserved FOXO-binding site in the promoter of the FOXO1 gene and it has been demonstrated that the transcription of FoxO genes is stimulated by FOXO3 [566]. Whether a feedback loop is affected by an exon-skipping compound targeting FoxO3 likewise remains to be further investigated.

5.4.2. Oligonucleotides to target FOXO3 and other transcription factors

Foxo3 is a transcription factor (TFs) at the heart of multiple cellular processes [542], with a complex regulation system itself. Drugability has traditionally been defined based on the processes used in small molecule ligand identification, including *in silico* pocket drugability statistical predictions [567], virtual fragment screening [568], and binding site description, alignment, and comparison tools [569]. Though they are well-characterised structurally, FOXO factors in particular are canonically classified within the aforementioned parameters as poorly druggable compounds by both strict ('rule of five') and more relaxed (chemical tractability) prediction methods [570].

Both pharmacological strategies presented in this chapter overcome this limit. PROTAC systems in general target often undruggable compounds [571,572]. Our Foxo(DBE)-PROTAC system is designed to target muscle atrophy and inserts itself into a growing ecosystem of oligonucleotide-based PROTAC platforms targeting TFs, including TF-PROTAC (which successfully targeted NF-κB and E2F) [573], TRAFTAC (which successfully targeted NF-κB and brachyury) [574], and O'PROTAC (which successfully targeted ERG and LEF1) [561] – all of which rely on the well-characterised DNA binding consensus sites, therefore avoiding *in silico* prediction limits.

The Pip8b2-PMO dianophore [218] does not rely on protein conformational interactions but acts directly at the pre-mRNA level. The data generated in this study provides background for more insightful compound design, as the investigation of the effect of second-generation compounds will also provide novel insights into the downstream effects divide between Foxo3 and Foxo1.

5.4.3. Applicability to SMA, to muscle-wasting diseases, and to other conditions

The approach of this chapter does not tackle SMA underlying genetics, but further application of optimised compounds would eventually yield compounds to be used in combination with SMN-centered approaches. Whilst our initial results identify two candidates for further development, 8b2-(A) shows an increase of *Foxo1* and *Foxo3* transcription at therapeutically relevant (nanomolar) doses, do not show any reduced expression of FOXO3 at the protein level, and upregulates atrogenes in a dose-response fashion. It remains to be seen how atrophied c2c12 myotubes respond to 8b2-(C), and how second-generation designs function with regards to splicing vs. transcription effects.

Surprisingly, the dose-response study on 8b2-(A) shows the upregulation of the atrogenes *Trim63* and *Mafbx* at lower doses – in contradiction to observations at the micromolar dose. Both those E3-ligases are targets not only of FOXO3 but of FOXO1 [575]. As such, it is possible that the upregulation of *Trim63* and *Mafbx* at lower doses is linked not to a lack of effect of 8b2-(A) on *Foxo3* exon 2 skipping, but rather to the observed upregulation of *Foxo3* and *Foxo1* total transcripts. This makes 8b2-(A) an interesting pharmacological tool that can be further investigated in conditions where FoxO1 and FoxO3 are downregulated. FoxO TFs are tumour suppressors [542,576,577] that have been reported dysregulated in breast cancer, prostate cancer, and glioblastoma [578], and their upregulation in gastric cancer improves prognosis [579,580]. What effect 8b2-(A), if any, has on tumor growth, remains to be investigated in further studies.

In conclusion, this chapter has presented two oligonucleotide-based approaches to down-regulate Foxo3 and reduce atrophy markers in a cellular model. I identified two exon-skipping candidates for further sequence development, which can also be used as tools to further dissect the transcriptomic regulation of the highly homologous *Foxo1* and *Foxo3* loci. Further work needs to be undertaken to increase *Foxo3* over *Foxo1* specificity and demonstrate efficacy on atrophy reduction at pharmacologically relevant doses for SMA and other wasting diseases.

6. Discussion on the use of transcriptomics datasets to identify and explore new therapeutic targets in SMA

6.1. Summary of results

This work started in the year following the regulatory approval of nusinersen, the *SMN2*-targeting splice-switching oligonucleotide therapy for SMA. Chapter 1 described how the availability of disease-modifying therapeutics does not preclude, and indeed rather encourages research into new therapeutic targets besides *SMN2*. The following chapters have all used developed from insights gained from transcriptomics datasets in motor neurons and in muscle.

Chapter 2 described first meta-analysis of transcriptomics human neuronal datasets in SMA. I showed that it is a powerful approach to integrate information from different studies on motor neuron biology. Withing the scope of Chapter 1 (human iPSCs MNs), I identified processes linked to plasma membrane, calcium regulation, and neuronal processes to be particularly enriched within genes differentially expressed across all available SMA human iPSCs MNs, regardless of stage of disease progression *in vitro*. I also identified that genes containing MAZ motifs and MOVO-B motifs are particularly enriched amongst the differentially expressed genes. This is a new avenue for downstream biological research in SMA. The meta-analytic approach is a powerful tool to integrate heterogeneity of SMA patients' samples within a rigorous and systematic study. When conducted to its fullest (after the integration of murine samples on one hand, and of systematically re-analyzed microarray on the other), the full meta-analysis of all reviewed SMA neuronal transcriptomic datasets promises to yield unique insight into the natural history of neurodegeneration in SMA at the transcriptomic level (both in terms of expression and of alternative splicing), across models and across species.

In Chapter 3, I demonstrated the exome-wide correction ability of a novel, intravenously administered splice-switching therapy to restore gene expression levels in motor neurons of severe SMA mice to levels of WT MN expression. I found that *Itpr3*/*ITPR3*, a candidate gene differentially expressed in both Chapter 3's microarray and in an unpublished collaborator's SMA iPSC-MNs RNAseq data, also features in the DE gene list from the meta-analysis. I have suggested that a splicing defect in the portion of *Itpr3* pre-mRNA coding for its transmembrane domain should be investigated for its functional effects on Itpr3's ability to release calcium from the endoplasmic reticulum to the cytosol. It will be interesting to integrate the newly generated dataset into the expansion of Chapter 2's meta-analysis to generate more hypotheses into the underlying individual molecular mechanism dysregulations behind neurodegeneration in SMA.

In Chapter 4, I expanded our study of the transcriptomic landscape in SMA to include circular RNAs. I show important methodological gaps in the wet-lab, targeted study of individual circRNAs within high isoform circRNA repertoires. I focused my work on the application of existing bioinformatics pipelines to generate hypotheses regarding the function of four *SMN*-derived circRNAs of reported high expression. I showed that *SMN*-derived circRNAs bear verified binding sites to six RNA-binding proteins (FOXP2, IGF2BP1-3 family, LIN28A, and PTB), and that some of them contain up to ten open-reading frames. The computational integration of the study of circRNAs into Chapter 2's meta-analysis is more challenging: RNAseq analyses of circRNAs rely on different initial technical pipelines designed around the biochemical characteristics of circRNAs. Overall, these results provide some additional insight into the complexity of the transcriptional regulation of *SMN1/2* loci, and, with a specific and sensitive quantitative analysis method, has the potential to open new avenues of research into theranostic markers of *SMN2* splice-correction efficacy.

Finally, in Chapter 5, I exploited transcriptomics information available from past studies in atrophied muscle to investigate FOXO3 as a *SMN*-independent therapeutic target to reduce muscle atrophy in muscle-wasting conditions like SMA. I presented two nucleic acid-based strategies, targeting FoxO3 either as a pre-mRNA (Pip8b2-PMO approach) or as a protein (FOXO3(DBE)-VHL PROTAC system). I showed that a segment of *Foxo3* exon 2 can be targeted to induce the exon skipping of the sequence coding for important protein regulatory domains. I identified two candidates as bases for second-generation splice-switching compounds against FOXO3. Whilst at lower (nanomolar) doses the shortlisted candidate 8b2-(A) induces the upregulation of both *Foxo3* and *Foxo1* total levels and of their downstream atrogenes *Trim63* and *Mafbx*, I proposed its repositioning as a therapeutic approach in conditions where the upregulation of FoxO1/3 can be beneficial.

6.2. Contributions to neuropathological dissection of SMA processes

As such, this work provides hypotheses-generating results applicable to two aspects SMA molecular neuropathology: calcium-signaling dysregulation at early symptomatic stages, and SMA-specific alterations of early post-natal MN development.

6.2.1. Insights from transcriptomics approaches

The pilot meta-analysis on human RNAseq neuronal SMA datasets identified the enrichment of membrane- and calcium-related genes. This result is corroborated in the novel mouse LCM-MN SMA dataset I presented in Chapter 3. I suggest the further exploration of the effect of splicing changes in the pre-mRNA of the endoplasmic reticulum transmembrane protein *Iptr3* and of its role in calcium dysregulation during neurodegeneration [324].

The comparison of SMA vs. WT early MN developmental processes (between PND2 and PND5) in a severe mouse model of SMA, and the manual curation of eight other MN/SC-centered mouse transcriptomic datasets at either or both of those post-natal developmental processes, has identified two intriguing gene families: the olfactory receptors (*Olfr*) genes and the major urinary protein (*Mup*) genes. Future work to explore this avenue of research ought to include the review of the recently published comprehensive study of the transcriptional and chromatin accessibility dynamic of motor neurons from E10.5 to PND21 in healthy mice [456] to investigate the normal dynamics of *Olfr* and *Mup* gene expression, as well as the integration of their datasets into the continuation of our meta-analysis study.

At present, the principal biomarker of therapeutic efficiency (so called 'theranostic' marker) used in SMA clinical studies is the measure of FL-SMN protein concentration in blood [211,213,581,582]. Whilst this biomarker correlates with *SMN2* mRNA levels in type I and type II patients (respective R_2 of 0.93 and 0.63), it fails to do so in type III patients ($R_2 = 0.23$) [581]. RNA-based biomarkers rely primarily on chain amplification methods, which makes them more sensitive than protein biomarkers and more cost-effective [583]. Linear RNA biomarkers (both mRNA and non-coding linear RNAs, e.g. miRNA) are sensitive to exonuclease degradation and alkaline conditions. CircRNAs in contrast present remarkable stability (half-line >48h, vs. <10h for linear RNA) [390]. My investigation of *SMN*-derived circRNAs suggests that their expression in SMA tissues is *SMN2* copy number dependent. Should the technical developments outlined in this thesis be continued, result in a sensitive quantitative method resolving between individual isoforms of the vast *SMN*-derived circRNA repertoire, and confirm our observations, the further study of *SMN*-derived circRNA could result in the development of additional theranostic biomarkers for *SMN2* splicing correction.

6.2.2. SMN deficiency beyond its effects on the transcriptome

Given the thoroughly characterised role of SMN in splicing regulation (both in the assembly of the spliceosome, and in pre-mRNA splicing itself), much of this thesis has focused on the SMA as a spliceopathy, and therefore has considered readouts on the changes and corrections in mRNA levels. It is however worth mentioning that the description of SMN protein function was recently expanded to translation regulation [132], notably in synergy with eIF3e [584]. SMN can bind to actively translating ribosomes [132]. The so-called 'SMN-primed ribosomes' preferentially translate a subset of mRNAs enriched for translational enhancer sequences in the 5'UTR and for rare codons at the beginning of their sequences [132], with functional consequences including lower levels of translation of acetylcholinesterase [132], a critical enzyme for neurotransmitter turnover at the NMJ. I mention this to stress the complex and multilevel regulation of gene and protein expression in SMA, and suggest that this would warrant further investigation of sequence-specific features to refine lists of candidate genes.

6.3. Perspectives on drug development in SMA beyond the motor neurons

6.3.1. Ongoing trends in drug development pipelines for SMA

In Chapter 5, I investigate FOXO3 as a new, *SMN*-independent therapeutic target to reduce muscle atrophy in muscle-wasting conditions like SMA. This line of work is concomitant to a broader shift towards muscle-centric therapies in SMA drug development pipelines.

The approval of two gene therapies in SMA (nusinersen and onasemnogene abeparvovec) has been followed by significant efforts to develop small, orally available molecules, including risdiplam. As such, all three therapies that are currently clinically available to patients aim at increasing levels of the SMN protein. Nusinersen (from Biogen™) is explicitly delivered to the central nervous system while both Zolgensma® and risdiplam also have a broader, peripheral distribution. (It must be noted that the STRONG and STEER clinicals trials investigate AAV9-SMN1 replacement gene therapy delivered intrathecally rather than intravenously.) As of September 2021, Biogen™ reports 11,740 patients worldwide are receiving nusinersen (including patients in Expanded Access Programs and in clinical trials) [585].

Much of the ongoing clinical therapeutic development focus either on combinatorial approaches (see below for apitegromab, GYM329), and/or on targeting the muscle (apitegromab, reldesemtiv, GYM329, NMD-670, BBIB110).

Apitegromab (SRK-015), from ScholarRock™ is a selective inhibitor of the activation of latent myostatin [234,235]. This muscle-targeting candidate is currently in phase III clinical trials (SAPPHIRE study, NCT05156320) for type II and type III SMA patients receiving either risdiplam or nusinersen. Reldesemtiv (CK-2127107), from Cytokinetics™ is a fast skeletal muscle troponin activator (increase of muscle contractility) ongoing phase II clinical trial (NCT02644668) for patients with type II-IV SMA. GYM329 (RO7204239), from Roche/Genentech, is an anti-myostatin antibody in a phase II/III trial (MANATEE study, NCT NCT05115110) in combination with risdiplam for ambulant children (2-10 years) with SMA. NMD-670, from NMD-Pharma®, is a small molecule inhibitor of the muscle-specific chloride ion channel ClC-1 [237,238]. It is expected that the first patient will be treated in a phase I/IIa trial by the end of 2022. Finally, BBIB110, from Biogen, is myostatin inhibitor (activin receptor type IIA/B ligand trap) and is announced to start a phase I trial.

In future studies, it will be interesting to consider how both these clinical-stage candidates fare on their own and as combinatorial therapies to their *SMN*-targeting counterparts. This also ought to inform further drug design for the application of FoxO3-targeting therapies in SMA.

6.3.2. Newborn screening programs and the new clinical SMA phenotypes

The regulatory approval of nusinersen provided the possibility of early intervention and of a treatment decreasing morbidity and mortality. This has lifted ethical limitations to the establishment of systematic newborn screening programs (NBS) for SMA. In turn, NBS allow for rapid molecular diagnosis of SMA. In addition to providing new insights into SMA population genetics [58,60,228], NBS permit early, pre-symptomatic treatment with disease-modifying therapies [233].

New phenotypes also appear in patients with post-symptomatic treatment onset. This has led to radically new phenotypes (e.g., when to consider surgery for scoliosis for orthopaedic management; how peripheral organ development is affected in SMA patients treated with intrathecal therapies; whether sensory-motor circuit dysfunctions change upon SMN splice-switching or replacement therapies) and new SMA natural histories, for which we still lack the hindsight necessary for their qualification and the understanding of novel unmet therapeutic needs.

Our transcriptomic study of PPMO-treated Taiwanese mice shows that the correction of expression levels is accompanied by an increase in splicing changes at the whole-exome level. This is a novel insight into the new natural histories (at the molecular level, in a pre-clinical animal model) that are appearing with the expansion of treatment amongst SMA patient populations. A refined analysis is warranted to further understand the longer-term consequences of splice-switching therapies on neuronal homeostasis and survival in an *SMN1*-null context.

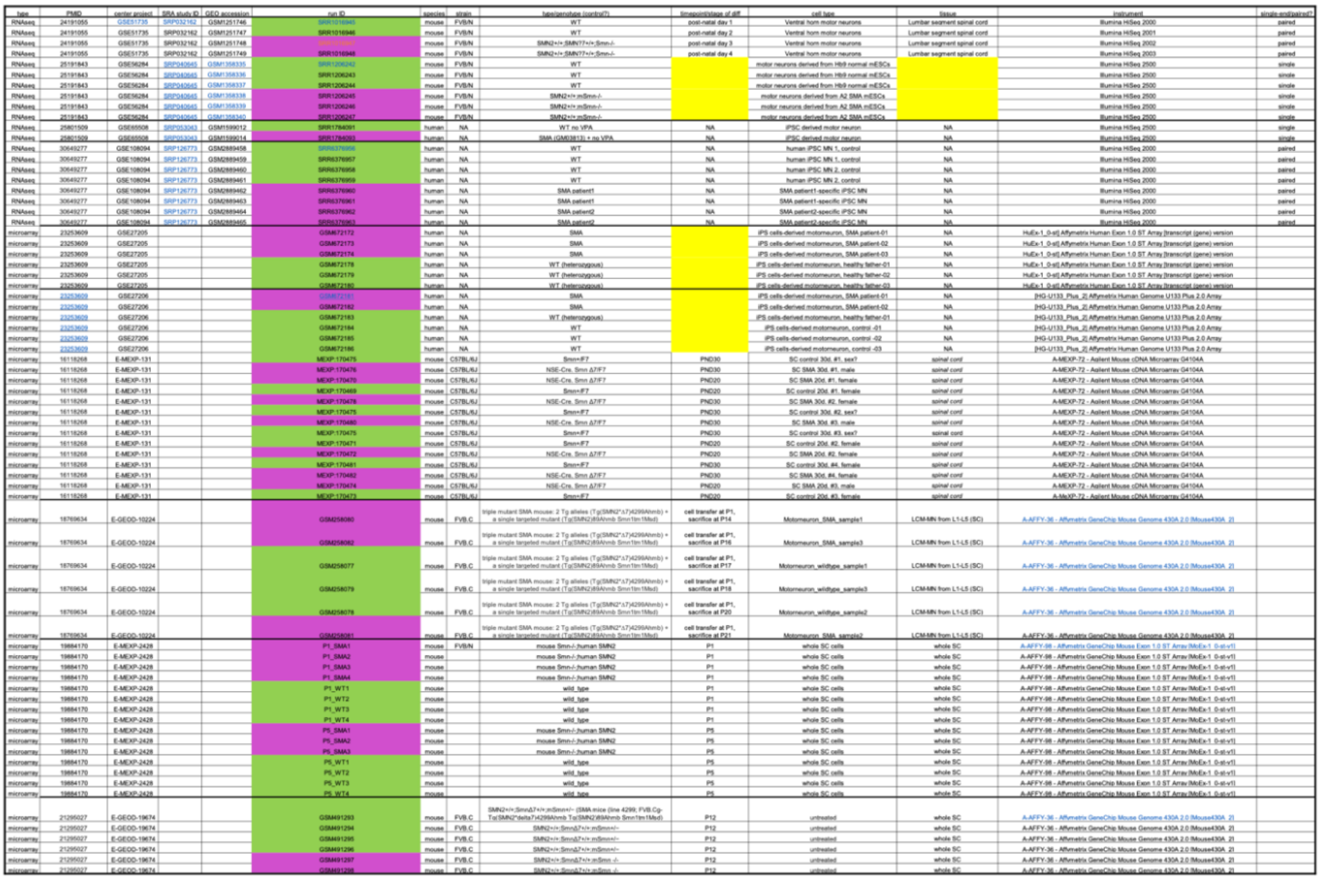
6.4. Concluding remarks

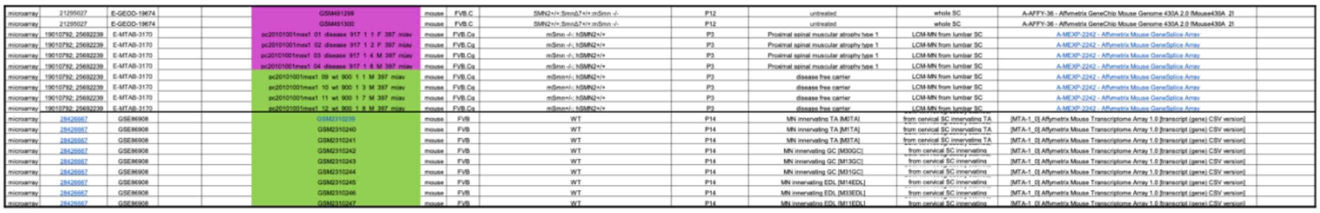
In conclusion, this thesis contributes to several aspects of critical basic research avenues to SMA therapeutic development. Our meta-analysis of human RNAseq neuronal datasets in SMA provides insight into critical functions of SMA in neuronal processes. I have demonstrated the efficacy in MN at the whole-exome level of an intravenously delivered splice-switching SMN-targeting strategy. I have used a transcriptomic dataset to identify genes that are left uncorrected by such a treatment. I have designed and initiated two pharmacological strategies to down-regulate two atrophy-controlling transcription factors in SMA, FoxO1 and FoxO3, with the aim of increasing muscle strength without necessarily targeting MN.

Taken as a whole, the use of transcriptomic-level approaches in this thesis provides new avenues of research both to further the understanding of fundamental neurodevelopment processes in SMA, and to identify candidates for the development of new therapies addressing unmet therapeutic needs.

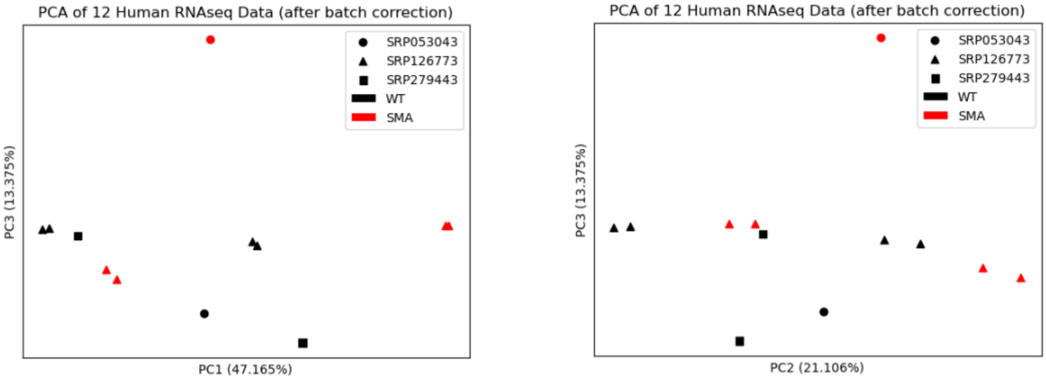
7. Supplementary Materials

Supplementary Table 7-1: detailed characteristics of all samples included in the transcriptomics meta-analysis





type	PMID	center project	SRA study ID	GEO accession	run ID	species	strain	type/genotype (control?)	timepoint/stage of diff	cell type	tissue	instrument	single-end/paired?
RNAseq	24191055	GSE51735	SRP032162	GSM1251746	SRR1016945	mouse	FVB/N	WT	post-natal day 1	Ventral horn motor neurons	Lumbar segment spinal cord	Illumina HiSeq 2000	paired
RNAseq	24191055	GSE51735	SRP032162	GSM1251747	SRR1016946	mouse	FVB/N	WT	post-natal day 2	Ventral horn motor neurons	Lumbar segment spinal cord	Illumina HiSeq 2001	paired
RNAseq	24191055	GSE51735	SRP032162	GSM1251748	SRR1016947	mouse	FVB/N	SMN2+/+;SMN?7+/+;Smn-/-	post-natal day 3	Ventral horn motor neurons	Lumbar segment spinal cord	Illumina HiSeq 2002	paired
RNAseq	24191055	GSE51735	SRP032162	GSM1251749	SRR1016948	mouse	FVB/N	SMN2+/+;SMN?7+/+;Smn-/-	post-natal day 4	Ventral horn motor neurons	Lumbar segment spinal cord	Illumina HiSeq 2003	paired
RNAseq	25191843	GSE56284	SRP040645	GSM1358335	SRR1206242	mouse	FVB/N	WT		motor neurons derived from Hb9 normal mESCs		Illumina HiSeq 2500	single
RNAseq	25191843	GSE56284	SRP040645	GSM1358336	SRR1206243	mouse	FVB/N	WT		motor neurons derived from Hb9 normal mESCs		Illumina HiSeq 2500	single
RNAseq	25191843	GSE56284	SRP040645	GSM1358337	SRR1206244	mouse	FVB/N	WT		motor neurons derived from Hb9 normal mESCs		Illumina HiSeq 2500	single
RNAseq	25191843	GSE56284	SRP040645	GSM1358338	SRR1206245	mouse	FVB/N	SMN2+/+;mSmn-/-		motor neurons derived from A2 SMA mESCs		Illumina HiSeq 2500	single
RNAseq	25191843	GSE56284	SRP040645	GSM1358339	SRR1206246	mouse	FVB/N	SMN2+/+;mSmn-/-		motor neurons derived from A2 SMA mESCs		Illumina HiSeq 2500	single
RNAseq	25191843	GSE56284	SRP040645	GSM1358340	SRR1206247	mouse	FVB/N	SMN2+/+;mSmn-/-		motor neurons derived from A2 SMA mESCs		Illumina HiSeq 2500	single
RNAseq	25801509	GSE65508	SRP053043	GSM1599012	SRR1784091	human	NA	WT no VPA	NA	iPSC derived motor neuron	NA	Illumina HiSeq 2500	single
RNAseq	25801509	GSE65508	SRP053043	GSM1599014	SRR1784093	human	NA	SMA (GM03813) + no VPA	NA	iPSC derived motor neuron	NA	Illumina HiSeq 2500	single
RNAseq	30649277	GSE108094	SRP126773	GSM2889458	SRR6376956	human	NA	WT	NA	human iPSC MN 1, control	NA	Illumina HiSeq 2000	paired
RNAseq	30649277	GSE108094	SRP126773	GSM2889459	SRR6376957	human	NA	WT	NA	human iPSC MN 1, control	NA	Illumina HiSeq 2000	paired
RNAseq	30649277	GSE108094	SRP126773	GSM2889460	SRR6376958	human	NA	WT	NA	human iPSC MN 2, control	NA	Illumina HiSeq 2000	paired
RNAseq	30649277	GSE108094	SRP126773	GSM2889461	SRR6376959	human	NA	WT	NA	human iPSC MN 2, control	NA	Illumina HiSeq 2000	paired
RNAseq	30649277	GSE108094	SRP126773	GSM2889462	SRR6376960	human	NA	SMA patient1	NA	SMA patient1-specific iPSC MN	NA	Illumina HiSeq 2000	paired
RNAseq	30649277	GSE108094	SRP126773	GSM2889463	SRR6376961	human	NA	SMA patient1	NA	SMA patient1-specific iPSC MN	NA	Illumina HiSeq 2000	paired
RNAseq	30649277	GSE108094	SRP126773	GSM2889464	SRR6376962	human	NA	SMA patient2	NA	SMA patient2-specific iPSC MN	NA	Illumina HiSeq 2000	paired
RNAseq	30649277	GSE108094	SRP126773	GSM2889465	SRR6376963	human	NA	SMA patient2	NA	SMA patient2-specific iPSC MN	NA	Illumina HiSeq 2000	paired
microarray	23253609	GSE27205			GSM672172	human	NA	SMA		iPS cells-derived motorneuron, SMA patient-01	NA	HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version	
microarray	23253609	GSE27205			GSM672173	human	NA	SMA		iPS cells-derived motorneuron, SMA patient-02	NA	HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version	
microarray	23253609	GSE27205			GSM672174	human	NA	SMA		iPS cells-derived motorneuron, SMA patient-03	NA	HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version	
microarray	23253609	GSE27205			GSM672178	human	NA	WT (heterozygous)		iPS cells-derived motorneuron, healthy father-01	NA	HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version	
microarray	23253609	GSE27205			GSM672179	human	NA	WT (heterozygous)		iPS cells-derived motorneuron, healthy father-02	NA	HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version	
microarray	23253609	GSE27205			GSM672180	human	NA	WT (heterozygous)		iPS cells-derived motorneuron, healthy father-03	NA	HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version	
microarray	23253609	GSE27206			GSM672181	human	NA	SMA		iPS cells-derived motorneuron, SMA patient-01	NA	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	
microarray	23253609	GSE27206			GSM672182	human	NA	SMA		iPS cells-derived motorneuron, SMA patient-02	NA	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	
microarray	23253609	GSE27206			GSM672183	human	NA	WT (heterozygous)		iPS cells-derived motorneuron, healthy father-01	NA	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	
microarray	23253609	GSE27206			GSM672184	human	NA	WT		iPS cells-derived motorneuron, control -01	NA	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	
microarray	23253609	GSE27206			GSM672185	human	NA	WT		iPS cells-derived motorneuron, control -02	NA	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	
microarray	23253609	GSE27206			GSM672186	human	NA	WT		iPS cells-derived motorneuron, control -03	NA	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	
microarray	16118268	E-MEXP-131			MEXP:170475	mouse	C57BL/6J	Smn+/F7	PND30	SC control 30d, #1, sex?	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170476	mouse	C57BL/6J	NSE-Cre, Smn Δ7/F7	PND30	SC SMA 30d, #1, male	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170470	mouse	C57BL/6J	NSE-Cre, Smn Δ7/F7	PND20	SC SMA 20d, #1, female	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170469	mouse	C57BL/6J	Smn+/F7	PND20	SC control 20d, #1, female	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170478	mouse	C57BL/6J	NSE-Cre, Smn Δ7/F7	PND30	SC SMA 30d, #2, female	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170475	mouse	C57BL/6J	Smn+/F7	PND30	SC control 30d, #2, sex?	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170480	mouse	C57BL/6J	NSE-Cre, Smn Δ7/F7	PND30	SC SMA 30d, #3, male	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170475	mouse	C57BL/6J	Smn+/F7	PND30	SC control 30d, #3, sex?	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170471	mouse	C57BL/6J	Smn+/F7	PND20	SC control 20d, #2, female	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170472	mouse	C57BL/6J	NSE-Cre, Smn Δ7/F7	PND20	SC SMA 20d, #2, female	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170481	mouse	C57BL/6J	Smn+/F7	PND30	SC control 30d, #4, female	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170482	mouse	C57BL/6J	NSE-Cre, Smn Δ7/F7	PND30	SC SMA 30d, #4, female	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170474	mouse	C57BL/6J	NSE-Cre, Smn Δ7/F7	PND20	SC SMA 20d, #3, male	spinal cord	A-MEXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	16118268	E-MEXP-131			MEXP:170473	mouse	C57BL/6J	Smn+/F7	PND20	SC control 20d, #3, female	spinal cord	A-MeXP-72 - Agilent Mouse cDNA Microarray G4104A	
microarray	18769634	E-GEOD-10224			GSM258080	mouse	FVB.C	triple mutant SMA mouse: 2 Tg alleles (Tg(SMN2*Δ7)4299Ahmb) + a single targeted mutant (Tg(SMN2)89Ahmb Smn1tm1Msd)	cell transfer at P1, sacrifice at P14	Motorneuron_SMA_sample1	LCM-MN from L1-L5 (SC)	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	18769634	E-GEOD-10224			GSM258082	mouse	FVB.C	triple mutant SMA mouse: 2 Tg alleles (Tg(SMN2*Δ7)4299Ahmb) + a single targeted mutant (Tg(SMN2)89Ahmb Smn1tm1Msd)	cell transfer at P1, sacrifice at P16	Motorneuron_SMA_sample3	LCM-MN from L1-L5 (SC)	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	18769634	E-GEOD-10224			GSM258077	mouse	FVB.C	triple mutant SMA mouse: 2 Tg alleles (Tg(SMN2*Δ7)4299Ahmb) + a single targeted mutant (Tg(SMN2)89Ahmb Smn1tm1Msd)	cell transfer at P1, sacrifice at P17	Motorneuron_wildtype_sample1	LCM-MN from L1-L5 (SC)	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	18769634	E-GEOD-10224			GSM258079	mouse	FVB.C	triple mutant SMA mouse: 2 Tg alleles (Tg(SMN2*Δ7)4299Ahmb) + a single targeted mutant (Tg(SMN2)89Ahmb Smn1tm1Msd)	cell transfer at P1, sacrifice at P18	Motorneuron_wildtype_sample3	LCM-MN from L1-L5 (SC)	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	18769634	E-GEOD-10224			GSM258078	mouse	FVB.C	triple mutant SMA mouse: 2 Tg alleles (Tg(SMN2*Δ7)4299Ahmb) + a single targeted mutant (Tg(SMN2)89Ahmb Smn1tm1Msd)	cell transfer at P1, sacrifice at P20	Motorneuron_wildtype_sample2	LCM-MN from L1-L5 (SC)	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	18769634	E-GEOD-10224			GSM258081	mouse	FVB.C	triple mutant SMA mouse: 2 Tg alleles (Tg(SMN2*Δ7)4299Ahmb) + a single targeted mutant (Tg(SMN2)89Ahmb Smn1tm1Msd)	cell transfer at P1, sacrifice at P21	Motorneuron_SMA_sample2	LCM-MN from L1-L5 (SC)	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	19884170	E-MEXP-2428			P1_SMA1	mouse	FVB/N	mouse Smn-/-;human SMN2	P1	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P1_SMA2	mouse		mouse Smn-/-;human SMN2	P1	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P1_SMA3	mouse		mouse Smn-/-;human SMN2	P1	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P1_SMA4	mouse		mouse Smn-/-;human SMN2	P1	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P1_WT1	mouse		wild_type	P1	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P1_WT2	mouse		wild_type	P1	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P1_WT3	mouse		wild_type	P1	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P1_WT4	mouse		wild_type	P1	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P5_SMA1	mouse		mouse Smn-/-;human SMN2	P5	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P5_SMA2	mouse		mouse Smn-/-;human SMN2	P5	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P5_SMA3	mouse		mouse Smn-/-;human SMN2	P5	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P5_WT1	mouse		wild_type	P5	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P5_WT2	mouse		wild_type	P5	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P5_WT3	mouse		wild_type	P5	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	19884170	E-MEXP-2428			P5_WT4	mouse		wild_type	P5	whole SC cells	whole SC	A-AFFY-98 - Affymetrix GeneChip Mouse Exon 1.0 ST Array [MoEx-1_0-st-v1]	
microarray	21295027	E-GEOD-19674			GSM491293	mouse	FVB.C	SMN2+/+;SmnΔ7+/+;mSmn+/− (SMA mice (line 4299; FVB.Cg-Tg(SMN2*delta7)4299Ahmb Tg(SMN2)89Ahmb Smn1tm1Msd)	P12	untreated	whole SC	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	21295027	E-GEOD-19674			GSM491294	mouse	FVB.C	SMN2+/+;SmnΔ7+/+;mSmn+/−	P12	untreated	whole SC	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	21295027	E-GEOD-19674			GSM491295	mouse	FVB.C	SMN2+/+;SmnΔ7+/+;mSmn+/−	P12	untreated	whole SC	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	21295027	E-GEOD-19674			GSM491296	mouse	FVB.C	SMN2+/+;SmnΔ7+/+;mSmn+/−	P12	untreated	whole SC	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	21295027	E-GEOD-19674			GSM491297	mouse	FVB.C	SMN2+/+;SmnΔ7+/+;mSmn -/-	P12	untreated	whole SC	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	21295027	E-GEOD-19674			GSM491298	mouse	FVB.C	SMN2+/+;SmnΔ7+/+;mSmn -/-	P12	untreated	whole SC	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	21295027	E-GEOD-19674			GSM491299	mouse	FVB.C	SMN2+/+;SmnΔ7+/+;mSmn -/-	P12	untreated	whole SC	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	21295027	E-GEOD-19674			GSM491300	mouse	FVB.C	SMN2+/+;SmnΔ7+/+;mSmn -/-	P12	untreated	whole SC	A-AFFY-36 - Affymetrix GeneChip Mouse Genome 430A 2.0 [Mouse430A_2]	
microarray	19010792; 25692239	E-MTAB-3170			pc20101001mex1 01 disease 917 1 1 F 397 miav	mouse	FVB.Cg	mSmn -/-; hSMN2+/+	P3	Proximal spinal muscular atrophy type 1	LCM-MN from lumbar SC	A-MEXP-2242 - Affymetrix Mouse GeneSplice Array	
microarray	19010792; 25692239	E-MTAB-3170			pc20101001mex1 02 disease 917 1 2 F 397 miav	mouse	FVB.Cg	mSmn -/-; hSMN2+/+	P3	Proximal spinal muscular atrophy type 1	LCM-MN from lumbar SC	A-MEXP-2242 - Affymetrix Mouse GeneSplice Array	
microarray	19010792; 25692239	E-MTAB-3170			pc20101001mex1 03 disease 917 1 4 M 397 miav	mouse	FVB.Cg	mSmn -/-; hSMN2+/+	P3	Proximal spinal muscular atrophy type 1	LCM-MN from lumbar SC	A-MEXP-2242 - Affymetrix Mouse GeneSplice Array	
microarray	19010792; 25692239	E-MTAB-3170			pc20101001mex1 04 disease 917 1 8 M 397 miav	mouse	FVB.Cg	mSmn -/-; hSMN2+/+	P3	Proximal spinal muscular atrophy type 1	LCM-MN from lumbar SC	A-MEXP-2242 - Affymetrix Mouse GeneSplice Array	
microarray	19010792; 25692239	E-MTAB-3170			pc20101001mex1 09 wt 900 1 1 M 397 miav	mouse	FVB.Cg	mSmn+/-; hSMN2+/+	P3	disease free carrier	LCM-MN from lumbar SC	A-MEXP-2242 - Affymetrix Mouse GeneSplice Array	
microarray	19010792; 25692239	E-MTAB-3170			pc20101001mex1 10 wt 900 1 3 M 397 miav	mouse	FVB.Cg	mSmn+/-; hSMN2+/+	P3	disease free carrier	LCM-MN from lumbar SC	A-MEXP-2242 - Affymetrix Mouse GeneSplice Array	
microarray	19010792; 25692239	E-MTAB-3170			pc20101001mex1 11 wt 900 1 7 M 397 miav	mouse	FVB.Cg	mSmn+/-; hSMN2+/+	P3	disease free carrier	LCM-MN from lumbar SC	A-MEXP-2242 - Affymetrix Mouse GeneSplice Array	
microarray	19010792; 25692239	E-MTAB-3170			pc20101001mex1 12 wt 900 1 8 M 397 miav	mouse	FVB.Cg	mSmn+/-; hSMN2+/+	P3	disease free carrier	LCM-MN from lumbar SC	A-MEXP-2242 - Affymetrix Mouse GeneSplice Array	
microarray	28426667	GSE86908			GSM2310239	mouse	FVB	WT	P14	MN innervating TA [M0TA]	from cervical SC innervating TA	[MTA-1_0] Affymetrix Mouse Transcriptome Array 1.0 [transcript (gene) CSV version]	
microarray	28426667	GSE86908			GSM2310240	mouse	FVB	WT	P14	MN innervating TA [M1TA]	from cervical SC innervating TA	[MTA-1_0] Affymetrix Mouse Transcriptome Array 1.0 [transcript (gene) CSV version]	
microarray	28426667	GSE86908			GSM2310241	mouse	FVB	WT	P14	MN innervating TA [M3TA]	from cervical SC innervating TA	[MTA-1_0] Affymetrix Mouse Transcriptome Array 1.0 [transcript (gene) CSV version]	
microarray	28426667	GSE86908			GSM2310242	mouse	FVB	WT	P14	MN innervating GC [M30GC]	from cervical SC innervating	[MTA-1_0] Affymetrix Mouse Transcriptome Array 1.0 [transcript (gene) CSV version]	
microarray	28426667	GSE86908			GSM2310243	mouse	FVB	WT	P14	MN innervating GC [M13GC]	from cervical SC innervating	[MTA-1_0] Affymetrix Mouse Transcriptome Array 1.0 [transcript (gene) CSV version]	
microarray	28426667	GSE86908			GSM2310244	mouse	FVB	WT	P14	MN innervating GC [M31GC]	from cervical SC innervating	[MTA-1_0] Affymetrix Mouse Transcriptome Array 1.0 [transcript (gene) CSV version]	
microarray	28426667	GSE86908			GSM2310245	mouse	FVB	WT	P14	MN innervating EDL [M14EDL]	from cervical SC innervating	[MTA-1_0] Affymetrix Mouse Transcriptome Array 1.0 [transcript (gene) CSV version]	
microarray	28426667	GSE86908			GSM2310246	mouse	FVB	WT	P14	MN innervating EDL [M33EDL]	from cervical SC innervating	[MTA-1_0] Affymetrix Mouse Transcriptome Array 1.0 [transcript (gene) CSV version]	
microarray	28426667	GSE86908			GSM2310247	mouse	FVB	WT	P14	MN innervating EDL [M11EDL]	from cervical SC innervating	[MTA-1_0] Affymetrix Mouse Transcriptome Array 1.0 [transcript (gene) CSV version]	


PCA of 12 Human RNAseq Data (after batch correction)

PCA of 12 Human RNAseq Data (after batch correction)

Supplementary Figure 7-1: further PCA components comparisons after batch correction

Supplementary Table 7-2: 'plasma membrane' (GO:0005886) subset of high-level evidence DE genes for further exploration

plasma membrane	EPH receptor A5 [Source:HGNC Symbol;Acc:HGNC:3389]	ENSG00000145242
	glutamate metabotropic receptor 7 [Source:HGNC Symbol;Acc:HGNC:4599]	ENSG00000196277
	LDL receptor related protein 1 [Source:HGNC Symbol;Acc:HGNC:6692]	ENSG00000123384
	karyopherin subunit alpha 2 [Source:HGNC Symbol;Acc:HGNC:6395]	ENSG00000182481
	podocalyxin like [Source:HGNC Symbol;Acc:HGNC:9171]	ENSG00000128567
	calpain 2 [Source:HGNC Symbol;Acc:HGNC:1479]	ENSG00000162909
	niban apoptosis regulator 2 [Source:HGNC Symbol;Acc:HGNC:25282]	ENSG00000136830
	serpin family E member 2 [Source:HGNC Symbol;Acc:HGNC:8951]	ENSG00000135919
	aspartate beta-hydroxylase [Source:HGNC Symbol;Acc:HGNC:757]	ENSG00000198363
	acyl-CoA synthetase long chain family member 4 [Source:HGNC Symbol;Acc:HGNC:3571]	ENSG00000068366
	Jun proto-oncogene, AP-1 transcription factor subunit [Source:HGNC Symbol;Acc:HGNC:6204]	ENSG00000177606
	selenophosphate synthetase 1 [Source:HGNC Symbol;Acc:HGNC:19685]	ENSG00000086475
	cathepsin L [Source:HGNC Symbol;Acc:HGNC:2537]	ENSG00000135047
	adhesion G protein-coupled receptor L2 [Source:HGNC Symbol;Acc:HGNC:18582]	ENSG00000117114
	solute carrier family 2 member 1 [Source:HGNC Symbol;Acc:HGNC:11005]	ENSG00000117394
	glypican 4 [Source:HGNC Symbol;Acc:HGNC:4452]	ENSG00000076716
	LIM domain and actin binding 1 [Source:HGNC Symbol;Acc:HGNC:24636]	ENSG00000050405
	UDP-N-acetylglucosamine pyrophosphorylase 1 [Source:HGNC Symbol;Acc:HGNC:12457]	ENSG00000117143
	inositol 1,4,5-trisphosphate receptor type 3 [Source:HGNC Symbol;Acc:HGNC:6182]	ENSG00000096433
	jagged canonical Notch ligand 1 [Source:HGNC Symbol;Acc:HGNC:6188]	ENSG00000101384
	transforming growth factor beta receptor 1 [Source:HGNC Symbol;Acc:HGNC:11772]	ENSG00000106799
	ATP binding cassette subfamily A member 2 [Source:HGNC Symbol;Acc:HGNC:32]	ENSG00000107331
	reversion inducing cysteine rich protein with kazal motifs [Source:HGNC Symbol;Acc:HGNC:11345]	ENSG00000122707
	junction plakoglobin [Source:HGNC Symbol;Acc:HGNC:6207]	ENSG00000173801
	G protein signaling modulator 1 [Source:HGNC Symbol;Acc:HGNC:17858]	ENSG00000160360
	TNF receptor superfamily member 21 [Source:HGNC Symbol;Acc:HGNC:13469]	ENSG00000146072
	adhesion G protein-coupled receptor B2 [Source:HGNC Symbol;Acc:HGNC:944]	ENSG00000121753
	syntaxin binding protein 1 [Source:HGNC Symbol;Acc:HGNC:11444]	ENSG00000136854
	smoothened, frizzled class receptor [Source:HGNC Symbol;Acc:HGNC:11119]	ENSG00000128602
	alanyl aminopeptidase, membrane [Source:HGNC Symbol;Acc:HGNC:500]	ENSG00000166825
	solute carrier family 9 member A7 [Source:HGNC Symbol;Acc:HGNC:17123]	ENSG00000065923
	cell migration inducing hyaluronidase 1 [Source:HGNC Symbol;Acc:HGNC:29213]	ENSG00000103888
	protocadherin 7 [Source:HGNC Symbol;Acc:HGNC:8659]	ENSG00000169851
	desmoglein 2 [Source:HGNC Symbol;Acc:HGNC:3049]	ENSG00000046604

	erythrocyte membrane protein band 4.1 like 3 [Source:HGNC Symbol;Acc:HGNC:3380]	ENSG00000082397
	serine incorporator 5 [Source:HGNC Symbol;Acc:HGNC:18825]	ENSG00000164300
	Ras association domain family member 2 [Source:HGNC Symbol;Acc:HGNC:9883]	ENSG00000101265
	SLC9A3 regulator 1 [Source:HGNC Symbol;Acc:HGNC:11075]	ENSG00000109062
	AHNAK nucleoprotein 2 [Source:HGNC Symbol;Acc:HGNC:20125]	ENSG00000185567
	thymine DNA glycosylase [Source:HGNC Symbol;Acc:HGNC:11700]	ENSG00000139372
	neurofascin [Source:HGNC Symbol;Acc:HGNC:29866]	ENSG00000163531
	clathrin light chain B [Source:HGNC Symbol;Acc:HGNC:2091]	ENSG00000175416
	adhesion G protein-coupled receptor G6 [Source:HGNC Symbol;Acc:HGNC:13841]	ENSG00000112414
	platelet derived growth factor receptor alpha [Source:HGNC Symbol;Acc:HGNC:8803]	ENSG00000134853
	galectin 3 [Source:HGNC Symbol;Acc:HGNC:6563]	ENSG00000131981
	solute carrier family 7 member 2 [Source:HGNC Symbol;Acc:HGNC:11060]	ENSG00000003989
	ATPase phospholipid transporting 11A [Source:HGNC Symbol;Acc:HGNC:13552]	ENSG00000068650
	membrane metalloendopeptidase [Source:HGNC Symbol;Acc:HGNC:7154]	ENSG00000196549
	claudin 11 [Source:HGNC Symbol;Acc:HGNC:8514]	ENSG00000013297
	N-myc downstream regulated 1 [Source:HGNC Symbol;Acc:HGNC:7679]	ENSG00000104419
	dedicator of cytokinesis 4 [Source:HGNC Symbol;Acc:HGNC:19192]	ENSG00000128512
	raftlin, lipid raft linker 1 [Source:HGNC Symbol;Acc:HGNC:30278]	ENSG00000131378
	protein kinase C alpha [Source:HGNC Symbol;Acc:HGNC:9393]	ENSG00000154229
	junctional cadherin 5 associated [Source:HGNC Symbol;Acc:HGNC:29283]	ENSG00000165757
	adenylate cyclase 9 [Source:HGNC Symbol;Acc:HGNC:240]	ENSG00000162104
	cadherin 1 [Source:HGNC Symbol;Acc:HGNC:1748]	ENSG00000039068
	annexin A4 [Source:HGNC Symbol;Acc:HGNC:542]	ENSG00000196975
	solute carrier family 16 member 3 [Source:HGNC Symbol;Acc:HGNC:10924]	ENSG00000141526
	glycine decarboxylase [Source:HGNC Symbol;Acc:HGNC:4313]	ENSG00000178445
	BMP and activin membrane bound inhibitor [Source:HGNC Symbol;Acc:HGNC:30251]	ENSG00000095739
	F11 receptor [Source:HGNC Symbol;Acc:HGNC:14685]	ENSG00000158769
	N-myristoyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:7858]	ENSG00000152465
	peripheral myelin protein 22 [Source:HGNC Symbol;Acc:HGNC:9118]	ENSG00000109099
	caveolin 2 [Source:HGNC Symbol;Acc:HGNC:1528]	ENSG00000105971
	growth arrest specific 1 [Source:HGNC Symbol;Acc:HGNC:4165]	ENSG00000180447
	alkaline phosphatase, biomineralization associated [Source:HGNC Symbol;Acc:HGNC:438]	ENSG00000162551
	plasminogen activator, urokinase receptor [Source:HGNC Symbol;Acc:HGNC:9053]	ENSG00000011422
	motile sperm domain containing 2 [Source:HGNC Symbol;Acc:HGNC:28381]	ENSG00000130150
	dipeptidyl peptidase 4 [Source:HGNC Symbol;Acc:HGNC:3009]	ENSG00000197635
	ATPase Na+/K+ transporting subunit alpha 2 [Source:HGNC Symbol;Acc:HGNC:800]	ENSG00000018625
	butyrophilin subfamily 3 member A1 [Source:HGNC Symbol;Acc:HGNC:1138]	ENSG00000026950
	calcium voltage-gated channel subunit alpha1 H [Source:HGNC Symbol;Acc:HGNC:1395]	ENSG00000196557
	integrin subunit alpha 1 [Source:HGNC Symbol;Acc:HGNC:6134]	ENSG00000213949
	cannabinoid receptor interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:24546]	ENSG00000119865
	cadherin EGF LAG seven-pass G-type receptor 1 [Source:HGNC Symbol;Acc:HGNC:1850]	ENSG00000075275

	proteolipid protein 1 [Source:HGNC Symbol;Acc:HGNC:9086]	ENSG00000123560
	gasdermin E [Source:HGNC Symbol;Acc:HGNC:2810]	ENSG00000105928
	teratocarcinoma-derived growth factor 1 [Source:HGNC Symbol;Acc:HGNC:11701]	ENSG00000241186
	cingulin [Source:HGNC Symbol;Acc:HGNC:17429]	ENSG00000143375
	neurotensin receptor 1 [Source:HGNC Symbol;Acc:HGNC:8039]	ENSG00000101188
	APC membrane recruitment protein 1 [Source:HGNC Symbol;Acc:HGNC:26837]	ENSG00000184675
	EPH receptor B2 [Source:HGNC Symbol;Acc:HGNC:3393]	ENSG00000133216
	integrin subunit alpha 8 [Source:HGNC Symbol;Acc:HGNC:6144]	ENSG00000077943
	leupaxin [Source:HGNC Symbol;Acc:HGNC:14061]	ENSG00000110031
	solute carrier family 1 member 1 [Source:HGNC Symbol;Acc:HGNC:10939]	ENSG00000106688
	ATPase copper transporting beta [Source:HGNC Symbol;Acc:HGNC:870]	ENSG00000123191
	desmocollin 2 [Source:HGNC Symbol;Acc:HGNC:3036]	ENSG00000134755
	beta-secretase 2 [Source:HGNC Symbol;Acc:HGNC:934]	ENSG00000182240
	syntaxin 1A [Source:HGNC Symbol;Acc:HGNC:11433]	ENSG00000106089
	calcium voltage-gated channel auxiliary subunit beta 4 [Source:HGNC Symbol;Acc:HGNC:1404]	ENSG00000182389
	SH3 and multiple ankyrin repeat domains 1 [Source:HGNC Symbol;Acc:HGNC:15474]	ENSG00000161681
	integrin subunit beta 8 [Source:HGNC Symbol;Acc:HGNC:6163]	ENSG00000105855
	MAM domain containing glycosylphosphatidylinositol anchor 1 [Source:HGNC Symbol;Acc:HGNC:19267]	ENSG00000112139
	thrombospondin type 1 domain containing 7A [Source:HGNC Symbol;Acc:HGNC:22207]	ENSG00000005108
	adhesion G protein-coupled receptor D1 [Source:HGNC Symbol;Acc:HGNC:19893]	ENSG00000111452
	nerve growth factor receptor [Source:HGNC Symbol;Acc:HGNC:7809]	ENSG00000064300
	potassium voltage-gated channel subfamily Q member 3 [Source:HGNC Symbol;Acc:HGNC:6297]	ENSG00000184156
	APC down-regulated 1 [Source:HGNC Symbol;Acc:HGNC:15718]	ENSG00000154856
	neural cell adhesion molecule 2 [Source:HGNC Symbol;Acc:HGNC:7657]	ENSG00000154654
	TNF receptor superfamily member 8 [Source:HGNC Symbol;Acc:HGNC:11923]	ENSG00000120949
	PDZ domain containing 2 [Source:HGNC Symbol;Acc:HGNC:18486]	ENSG00000133401
	fibroblast activation protein alpha [Source:HGNC Symbol;Acc:HGNC:3590]	ENSG00000078098
	claudin 4 [Source:HGNC Symbol;Acc:HGNC:2046]	ENSG00000189143
	pleckstrin and Sec7 domain containing 4 [Source:HGNC Symbol;Acc:HGNC:19096]	ENSG00000125637
	SH3GL interacting endocytic adaptor 1 [Source:HGNC Symbol;Acc:HGNC:25412]	ENSG00000118473
	RELT like 1 [Source:HGNC Symbol;Acc:HGNC:27379]	ENSG00000181826
	Xg glycoprotein (Xg blood group) [Source:HGNC Symbol;Acc:HGNC:12806]	ENSG00000124343
	potassium calcium-activated channel subfamily N member 4 [Source:HGNC Symbol;Acc:HGNC:6293]	ENSG00000104783
	dysferlin [Source:HGNC Symbol;Acc:HGNC:3097]	ENSG00000135636
	CD74 molecule [Source:HGNC Symbol;Acc:HGNC:1697]	ENSG00000019582
	V-set immunoregulatory receptor [Source:HGNC Symbol;Acc:HGNC:30085]	ENSG00000107738
	cyclic GMP-AMP synthase [Source:HGNC Symbol;Acc:HGNC:21367]	ENSG00000164430
	desmin [Source:HGNC Symbol;Acc:HGNC:2770]	ENSG00000175084
	solute carrier family 7 member 14 [Source:HGNC Symbol;Acc:HGNC:29326]	ENSG00000013293
	C1q and TNF related 1 [Source:HGNC Symbol;Acc:HGNC:14324]	ENSG00000173918
	gamma-glutamyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:4250]	ENSG00000100031

	prostaglandin F receptor [Source:HGNC Symbol;Acc:HGNC:9600]	ENSG00000122420
	vesicle associated membrane protein 8 [Source:HGNC Symbol;Acc:HGNC:12647]	ENSG00000118640
	bradykinin receptor B2 [Source:HGNC Symbol;Acc:HGNC:1030]	ENSG00000168398
	lysophosphatidic acid receptor 4 [Source:HGNC Symbol;Acc:HGNC:4478]	ENSG00000147145
	caspase recruitment domain family member 11 [Source:HGNC Symbol;Acc:HGNC:16393]	ENSG00000198286
	claudin 10 [Source:HGNC Symbol;Acc:HGNC:2033]	ENSG00000134873
	histamine receptor H1 [Source:HGNC Symbol;Acc:HGNC:5182]	ENSG00000196639
	solute carrier organic anion transporter family member 1A2 [Source:HGNC Symbol;Acc:HGNC:10956]	ENSG00000084453
	EPH receptor A5 [Source:HGNC Symbol;Acc:HGNC:3389]	ENSG00000145242
	serine/threonine/tyrosine kinase 1 [Source:HGNC Symbol;Acc:HGNC:18889]	ENSG00000060140
	TraB domain containing 2B [Source:HGNC Symbol;Acc:HGNC:44200]	ENSG00000269113
	filaggrin [Source:HGNC Symbol;Acc:HGNC:3748]	ENSG00000143631
	occludin [Source:HGNC Symbol;Acc:HGNC:8104]	ENSG00000197822
	sodium voltage-gated channel alpha subunit 1 [Source:HGNC Symbol;Acc:HGNC:10585]	ENSG00000144285
	cilia and flagella associated protein 95 [Source:HGNC Symbol;Acc:HGNC:31422]	ENSG00000204711
	G protein-coupled receptor 143 [Source:HGNC Symbol;Acc:HGNC:20145]	ENSG00000101850
	anoctamin 1 [Source:HGNC Symbol;Acc:HGNC:21625]	ENSG00000131620
	claudin 19 [Source:HGNC Symbol;Acc:HGNC:2040]	ENSG00000164007
	ATP binding cassette subfamily G member 2 (Junior blood group) [Source:HGNC Symbol;Acc:HGNC:74]	ENSG00000118777
	desmocollin 3 [Source:HGNC Symbol;Acc:HGNC:3037]	ENSG00000134762
	LCK proto-oncogene, Src family tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:6524]	ENSG00000182866
	selectin P ligand [Source:HGNC Symbol;Acc:HGNC:10722]	ENSG00000110876
	cadherin 5 [Source:HGNC Symbol;Acc:HGNC:1764]	ENSG00000179776
	androgen dependent TFPI regulating protein [Source:HGNC Symbol;Acc:HGNC:21214]	ENSG00000111863
	modulator of VRAC current 1 [Source:HGNC Symbol;Acc:HGNC:17082]	ENSG00000100427
	chloride intracellular channel 2 [Source:HGNC Symbol;Acc:HGNC:2063]	ENSG00000155962
	interleukin 2 receptor subunit beta [Source:HGNC Symbol;Acc:HGNC:6009]	ENSG00000100385
	solute carrier family 22 member 4 [Source:HGNC Symbol;Acc:HGNC:10968]	ENSG00000197208
	transmembrane protein 266 [Source:HGNC Symbol;Acc:HGNC:26763]	ENSG00000169758
	claudin 5 [Source:HGNC Symbol;Acc:HGNC:2047]	ENSG00000184113
	CD70 molecule [Source:HGNC Symbol;Acc:HGNC:11937]	ENSG00000125726
	glutamyl aminopeptidase [Source:HGNC Symbol;Acc:HGNC:3355]	ENSG00000138792
	corin, serine peptidase [Source:HGNC Symbol;Acc:HGNC:19012]	ENSG00000145244
	5-hydroxytryptamine receptor 3A [Source:HGNC Symbol;Acc:HGNC:5297]	ENSG00000166736
	CEA cell adhesion molecule 1 [Source:HGNC Symbol;Acc:HGNC:1814]	ENSG00000079385
	solute carrier family 16 member 12 [Source:HGNC Symbol;Acc:HGNC:23094]	ENSG00000152779
	adherens junctions associated protein 1 [Source:HGNC Symbol;Acc:HGNC:30801]	ENSG00000196581
	purinergic receptor P2X 7 [Source:HGNC Symbol;Acc:HGNC:8537]	ENSG00000089041
	potassium voltage-gated channel subfamily E regulatory subunit 5 [Source:HGNC Symbol;Acc:HGNC:6241]	ENSG00000176076
	calcium voltage-gated channel auxiliary subunit gamma 6 [Source:HGNC Symbol;Acc:HGNC:13625]	ENSG00000130433

	solute carrier family 52 member 3 [Source:HGNC Symbol;Acc:HGNC:16187]	ENSG00000101276
	serpin family A member 5 [Source:HGNC Symbol;Acc:HGNC:8723]	ENSG00000188488
	transmembrane serine protease 5 [Source:HGNC Symbol;Acc:HGNC:14908]	ENSG00000166682
	cytochrome P450 family 2 subfamily C member 8 [Source:HGNC Symbol;Acc:HGNC:2622]	ENSG00000138115
	folate receptor alpha [Source:HGNC Symbol;Acc:HGNC:3791]	ENSG00000110195
	G protein-coupled bile acid receptor 1 [Source:HGNC Symbol;Acc:HGNC:19680]	ENSG00000179921
	major histocompatibility complex, class II, DO alpha [Source:HGNC Symbol;Acc:HGNC:4936]	ENSG00000204252
	glutamate metabotropic receptor 7 [Source:HGNC Symbol;Acc:HGNC:4599]	ENSG00000196277
	- unc-93 homolog A [Source:HGNC Symbol;Acc:HGNC:12570]	- ENSG00000112494
	dipeptidase 1 [Source:HGNC Symbol;Acc:HGNC:3002]	ENSG00000015413

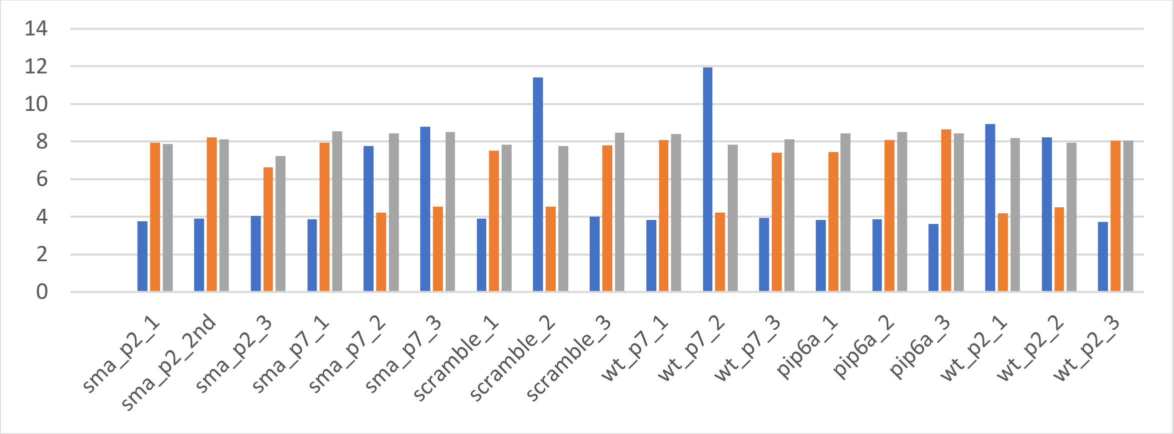
Supplementary Table 7-3: list of DE genes enriched for MAZ motif

MAZ motif			
initial_alias	converted_alias	name	description
ENSG00000170561	ENSG00000170561	IRX2	iroquois homeobox 2 [Source:HGNC Symbol;Acc:HGNC:14359]
ENSG00000178409	ENSG00000178409	BEND3	BEN domain containing 3 [Source:HGNC Symbol;Acc:HGNC:23040]
ENSG00000115594	ENSG00000115594	IL1R1	interleukin 1 receptor type 1 [Source:HGNC Symbol;Acc:HGNC:5993]
ENSG00000105928	ENSG00000105928	GSDME	gasdermin E [Source:HGNC Symbol;Acc:HGNC:2810]
ENSG00000241186	ENSG00000241186	TDGF1	teratocarcinoma-derived growth factor 1 [Source:HGNC Symbol;Acc:HGNC:11701]
ENSG00000143375	ENSG00000143375	CGN	cingulin [Source:HGNC Symbol;Acc:HGNC:17429]
ENSG00000101188	ENSG00000101188	NTSR1	neurotensin receptor 1 [Source:HGNC Symbol;Acc:HGNC:8039]
ENSG00000168646	ENSG00000168646	AXIN2	axin 2 [Source:HGNC Symbol;Acc:HGNC:904]
ENSG00000112773	ENSG00000112773	TENT5A	terminal nucleotidyltransferase 5A [Source:HGNC Symbol;Acc:HGNC:18345]
ENSG00000184675	ENSG00000184675	AMER1	APC membrane recruitment protein 1 [Source:HGNC Symbol;Acc:HGNC:26837]
ENSG00000116396	ENSG00000116396	KCNC4	potassium voltage-gated channel subfamily C member 4 [Source:HGNC Symbol;Acc:HGNC:6236]
ENSG00000105376	ENSG00000105376	ICAM5	intercellular adhesion molecule 5 [Source:HGNC Symbol;Acc:HGNC:5348]
ENSG00000133216	ENSG00000133216	EPHB2	EPH receptor B2 [Source:HGNC Symbol;Acc:HGNC:3393]
ENSG00000118276	ENSG00000118276	B4GALT6	beta-1,4-galactosyltransferase 6 [Source:HGNC Symbol;Acc:HGNC:929]
ENSG00000146757	ENSG00000146757	ZNF92	zinc finger protein 92 [Source:HGNC Symbol;Acc:HGNC:13168]
ENSG00000186577	ENSG00000186577	SMIM29	small integral membrane protein 29 [Source:HGNC Symbol;Acc:HGNC:1340]
ENSG00000110031	ENSG00000110031	LPXN	leupaxin [Source:HGNC Symbol;Acc:HGNC:14061]
ENSG00000219438	ENSG00000219438	TAFA5	TAFA chemokine like family member 5 [Source:HGNC Symbol;Acc:HGNC:21592]
ENSG00000166831	ENSG00000166831	RBPMS2	RNA binding protein, mRNA processing factor 2 [Source:HGNC Symbol;Acc:HGNC:19098]
ENSG00000101144	ENSG00000101144	BMP7	bone morphogenetic protein 7 [Source:HGNC Symbol;Acc:HGNC:1074]
ENSG00000128872	ENSG00000128872	TMOD2	tropomodulin 2 [Source:HGNC Symbol;Acc:HGNC:11872]
ENSG00000118298	ENSG00000118298	CA14	carbonic anhydrase 14 [Source:HGNC Symbol;Acc:HGNC:1372]
ENSG00000114853	ENSG00000114853	ZBTB47	zinc finger and BTB domain containing 47 [Source:HGNC Symbol;Acc:HGNC:26955]
ENSG00000091844	ENSG00000091844	RGS17	regulator of G protein signaling 17 [Source:HGNC Symbol;Acc:HGNC:14088]
ENSG00000129946	ENSG00000129946	SHC2	SHC adaptor protein 2 [Source:HGNC Symbol;Acc:HGNC:29869]
ENSG00000137203	ENSG00000137203	TFAP2A	transcription factor AP-2 alpha [Source:HGNC Symbol;Acc:HGNC:11742]
ENSG00000138641	ENSG00000138641	HERC3	HECT and RLD domain containing E3 ubiquitin protein ligase 3 [Source:HGNC Symbol;Acc:HGNC:48
ENSG00000182240	ENSG00000182240	BACE2	beta-secretase 2 [Source:HGNC Symbol;Acc:HGNC:934]
ENSG00000092969	ENSG00000092969	TGFB2	transforming growth factor beta 2 [Source:HGNC Symbol;Acc:HGNC:11768]
ENSG00000170412	ENSG00000170412	GPRC5C	G protein-coupled receptor class C group 5 member C [Source:HGNC Symbol;Acc:HGNC:13309]
ENSG00000115194	ENSG00000115194	SLC30A3	solute carrier family 30 member 3 [Source:HGNC Symbol;Acc:HGNC:11014]
ENSG00000143786	ENSG00000143786	CNIH3	cornichon family AMPA receptor auxiliary protein 3 [Source:HGNC Symbol;Acc:HGNC:26802]
ENSG00000148204	ENSG00000148204	CRB2	crumbs cell polarity complex component 2 [Source:HGNC Symbol;Acc:HGNC:18688]
ENSG00000188766	ENSG00000188766	SPRED3	sprouty related EVH1 domain containing 3 [Source:HGNC Symbol;Acc:HGNC:31041]
ENSG00000158050	ENSG00000158050	DUSP2	dual specificity phosphatase 2 [Source:HGNC Symbol;Acc:HGNC:3068]
ENSG00000161681	ENSG00000161681	SHANK1	SH3 and multiple ankyrin repeat domains 1 [Source:HGNC Symbol;Acc:HGNC:15474]
ENSG00000078114	ENSG00000078114	NEBL	nebulette [Source:HGNC Symbol;Acc:HGNC:16932]
ENSG00000125675	ENSG00000125675	GRIA3	glutamate ionotropic receptor AMPA type subunit 3 [Source:HGNC Symbol;Acc:HGNC:4573]
ENSG00000187122	ENSG00000187122	SLIT1	slit guidance ligand 1 [Source:HGNC Symbol;Acc:HGNC:11085]
ENSG00000064300	ENSG00000064300	NGFR	nerve growth factor receptor [Source:HGNC Symbol;Acc:HGNC:7809]
ENSG00000184156	ENSG00000184156	KCNQ3	potassium voltage-gated channel subfamily Q member 3 [Source:HGNC Symbol;Acc:HGNC:6297]
ENSG00000152092	ENSG00000152092	ASTN1	astrotactin 1 [Source:HGNC Symbol;Acc:HGNC:773]
ENSG00000140807	ENSG00000140807	NKD1	NKD inhibitor of WNT signaling pathway 1 [Source:HGNC Symbol;Acc:HGNC:17045]
ENSG00000150630	ENSG00000150630	VEGFC	vascular endothelial growth factor C [Source:HGNC Symbol;Acc:HGNC:12682]
ENSG00000054690	ENSG00000054690	PLEKHH1	pleckstrin homology, MyTH4 and FERM domain containing H1 [Source:HGNC Symbol;Acc:HGNC:17
ENSG00000133980	ENSG00000133980	VRTN	vertebrae development associated [Source:HGNC Symbol;Acc:HGNC:20223]
ENSG00000154654	ENSG00000154654	NCAM2	neural cell adhesion molecule 2 [Source:HGNC Symbol;Acc:HGNC:7657]
ENSG00000176697	ENSG00000176697	BDNF	brain derived neurotrophic factor [Source:HGNC Symbol;Acc:HGNC:1033]
ENSG00000126368	ENSG00000126368	NR1D1	nuclear receptor subfamily 1 group D member 1 [Source:HGNC Symbol;Acc:HGNC:7962]
ENSG00000177374	ENSG00000177374	HIC1	HIC ZBTB transcriptional repressor 1 [Source:HGNC Symbol;Acc:HGNC:4909]
ENSG00000171291	ENSG00000171291	ZNF439	zinc finger protein 439 [Source:HGNC Symbol;Acc:HGNC:20873]
ENSG00000133401	ENSG00000133401	PDZD2	PDZ domain containing 2 [Source:HGNC Symbol;Acc:HGNC:18486]
ENSG00000198353	ENSG00000198353	HOXC4	homeobox C4 [Source:HGNC Symbol;Acc:HGNC:5126]
ENSG00000139835	ENSG00000139835	GRTP1	growth hormone regulated TBC protein 1 [Source:HGNC Symbol;Acc:HGNC:20310]
ENSG00000169946	ENSG00000169946	ZFPM2	zinc finger protein, FOG family member 2 [Source:HGNC Symbol;Acc:HGNC:16700]
ENSG00000181826	ENSG00000181826	RELL1	RELT like 1 [Source:HGNC Symbol;Acc:HGNC:27379]
ENSG00000138795	ENSG00000138795	LEF1	lymphoid enhancer binding factor 1 [Source:HGNC Symbol;Acc:HGNC:6551]
ENSG00000108950	ENSG00000108950	FAM20A	FAM20A golgi associated secretory pathway pseudokinase [Source:HGNC Symbol;Acc:HGNC:23015
ENSG00000135480	ENSG00000135480	KRT7	keratin 7 [Source:HGNC Symbol;Acc:HGNC:6445]
ENSG00000064692	ENSG00000064692	SNCAIP	synuclein alpha interacting protein [Source:HGNC Symbol;Acc:HGNC:11139]
ENSG00000204248	ENSG00000204248	COL11A2	collagen type XI alpha 2 chain [Source:HGNC Symbol;Acc:HGNC:2187]
ENSG00000025708	ENSG00000025708	TYMP	thymidine phosphorylase [Source:HGNC Symbol;Acc:HGNC:3148]
ENSG00000130055	ENSG00000130055	GDPD2	glycerophosphodiester phosphodiesterase domain containing 2 [Source:HGNC Symbol;Acc:HGNC:2
ENSG00000135636	ENSG00000135636	DYSF	dysferlin [Source:HGNC Symbol;Acc:HGNC:3097]
ENSG00000135454	ENSG00000135454	B4GALNT1	beta-1,4-N-acetyl-galactosaminyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:4117]
ENSG00000160973	ENSG00000160973	FOXH1	forkhead box H1 [Source:HGNC Symbol;Acc:HGNC:3814]
ENSG00000204219	ENSG00000204219	TCEA3	transcription elongation factor A3 [Source:HGNC Symbol;Acc:HGNC:11615]
ENSG00000171502	ENSG00000171502	COL24A1	collagen type XXIV alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:20821]
ENSG00000147596	ENSG00000147596	PRDM14	PR/SET domain 14 [Source:HGNC Symbol;Acc:HGNC:14001]
ENSG00000171246	ENSG00000171246	NPTX1	neuronal pentraxin 1 [Source:HGNC Symbol;Acc:HGNC:7952]
ENSG00000135472	ENSG00000135472	FAIM2	Fas apoptotic inhibitory molecule 2 [Source:HGNC Symbol;Acc:HGNC:17067]
ENSG00000013619	ENSG00000013619	MAMLD1	mastermind like domain containing 1 [Source:HGNC Symbol;Acc:HGNC:2568]
ENSG00000178222	ENSG00000178222	RNF212	ring finger protein 212 [Source:HGNC Symbol;Acc:HGNC:27729]
ENSG00000174749	ENSG00000174749	FAM241A	family with sequence similarity 241 member A [Source:HGNC Symbol;Acc:HGNC:26813]
ENSG00000100031	ENSG00000100031	GGT1	gamma-glutamyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:4250]
ENSG00000118137	ENSG00000118137	APOA1	apolipoprotein A1 [Source:HGNC Symbol;Acc:HGNC:600]
ENSG00000168918	ENSG00000168918	INPP5D	inositol polyphosphate-5-phosphatase D [Source:HGNC Symbol;Acc:HGNC:6079]
ENSG00000164161	ENSG00000164161	HHIP	hedgehog interacting protein [Source:HGNC Symbol;Acc:HGNC:14866]
ENSG00000103710	ENSG00000103710	RASL12	RAS like family 12 [Source:HGNC Symbol;Acc:HGNC:30289]
ENSG00000134070	ENSG00000134070	IRAK2	interleukin 1 receptor associated kinase 2 [Source:HGNC Symbol;Acc:HGNC:6113]
ENSG00000165730	ENSG00000165730	STOX1	storkhead box 1 [Source:HGNC Symbol;Acc:HGNC:23508]
ENSG00000127578	ENSG00000127578	WFIKKN1	WAP, follistatin/kazal, immunoglobulin, kunitz and netrin domain containing 1 [Source:HGNC Sym
ENSG00000167244	ENSG00000167244	IGF2	insulin like growth factor 2 [Source:HGNC Symbol;Acc:HGNC:5466]
ENSG00000196878	ENSG00000196878	LAMB3	laminin subunit beta 3 [Source:HGNC Symbol;Acc:HGNC:6490]
ENSG00000134873	ENSG00000134873	CLDN10	claudin 10 [Source:HGNC Symbol;Acc:HGNC:2033]
ENSG00000080854	ENSG00000080854	IGSF9B	immunoglobulin superfamily member 9B [Source:HGNC Symbol;Acc:HGNC:32326]
ENSG00000260428	ENSG00000260428	SCX	scleraxis bHLH transcription factor [Source:HGNC Symbol;Acc:HGNC:32322]
ENSG00000261308	ENSG00000261308	FIGNL2	fidgetin like 2 [Source:HGNC Symbol;Acc:HGNC:13287]
ENSG00000060140	ENSG00000060140	STYK1	serine/threonine/tyrosine kinase 1 [Source:HGNC Symbol;Acc:HGNC:18889]
ENSG00000269113	ENSG00000269113	TRABD2B	TraB domain containing 2B [Source:HGNC Symbol;Acc:HGNC:44200]
ENSG00000159733	ENSG00000159733	ZFYVE28	zinc finger FYVE-type containing 28 [Source:HGNC Symbol;Acc:HGNC:29334]
ENSG00000188730	ENSG00000188730	VWC2	von Willebrand factor C domain containing 2 [Source:HGNC Symbol;Acc:HGNC:30200]
ENSG00000152784	ENSG00000152784	PRDM8	PR/SET domain 8 [Source:HGNC Symbol;Acc:HGNC:13993]
ENSG00000142273	ENSG00000142273	CBLC	Cbl proto-oncogene C [Source:HGNC Symbol;Acc:HGNC:15961]
ENSG00000144285	ENSG00000144285	SCN1A	sodium voltage-gated channel alpha subunit 1 [Source:HGNC Symbol;Acc:HGNC:10585]
ENSG00000087495	ENSG00000087495	PHACTR3	phosphatase and actin regulator 3 [Source:HGNC Symbol;Acc:HGNC:15833]
ENSG00000167178	ENSG00000167178	ISLR2	immunoglobulin superfamily containing leucine rich repeat 2 [Source:HGNC Symbol;Acc:HGNC:292
ENSG00000138311	ENSG00000138311	ZNF365	zinc finger protein 365 [Source:HGNC Symbol;Acc:HGNC:18194]
ENSG00000162624	ENSG00000162624	LHX8	LIM homeobox 8 [Source:HGNC Symbol;Acc:HGNC:28838]
ENSG00000203805	ENSG00000203805	PLPP4	phospholipid phosphatase 4 [Source:HGNC Symbol;Acc:HGNC:23531]
ENSG00000181444	ENSG00000181444	ZNF467	zinc finger protein 467 [Source:HGNC Symbol;Acc:HGNC:23154]
ENSG00000130528	ENSG00000130528	HRC	histidine rich calcium binding protein [Source:HGNC Symbol;Acc:HGNC:5178]
ENSG00000076344	ENSG00000076344	RGS11	regulator of G protein signaling 11 [Source:HGNC Symbol;Acc:HGNC:9993]
ENSG00000171227	ENSG00000171227	TMEM37	transmembrane protein 37 [Source:HGNC Symbol;Acc:HGNC:18216]
ENSG00000185519	ENSG00000185519	FAM131C	family with sequence similarity 131 member C [Source:HGNC Symbol;Acc:HGNC:26717]
ENSG00000131668	ENSG00000131668	BARX1	BARX homeobox 1 [Source:HGNC Symbol;Acc:HGNC:955]
ENSG00000253313	ENSG00000253313	C1orf210	chromosome 1 open reading frame 210 [Source:HGNC Symbol;Acc:HGNC:28755]
ENSG00000159166	ENSG00000159166	LAD1	ladinin 1 [Source:HGNC Symbol;Acc:HGNC:6472]
ENSG00000197208	ENSG00000197208	SLC22A4	solute carrier family 22 member 4 [Source:HGNC Symbol;Acc:HGNC:10968]
ENSG00000134259	ENSG00000134259	NGF	nerve growth factor [Source:HGNC Symbol;Acc:HGNC:7808]
ENSG00000125726	ENSG00000125726	CD70	CD70 molecule [Source:HGNC Symbol;Acc:HGNC:11937]
ENSG00000102575	ENSG00000102575	ACP5	acid phosphatase 5, tartrate resistant [Source:HGNC Symbol;Acc:HGNC:124]
ENSG00000159314	ENSG00000159314	ARHGAP27	Rho GTPase activating protein 27 [Source:HGNC Symbol;Acc:HGNC:31813]
ENSG00000204335	ENSG00000204335	SP5	Sp5 transcription factor [Source:HGNC Symbol;Acc:HGNC:14529]
ENSG00000145692	ENSG00000145692	BHMT	betaine--homocysteine S-methyltransferase [Source:HGNC Symbol;Acc:HGNC:1047]
ENSG00000170689	ENSG00000170689	HOXB9	homeobox B9 [Source:HGNC Symbol;Acc:HGNC:5120]
ENSG00000128713	ENSG00000128713	HOXD11	homeobox D11 [Source:HGNC Symbol;Acc:HGNC:5134]
ENSG00000079385	ENSG00000079385	CEACAM1	CEA cell adhesion molecule 1 [Source:HGNC Symbol;Acc:HGNC:1814]
ENSG00000152779	ENSG00000152779	SLC16A12	solute carrier family 16 member 12 [Source:HGNC Symbol;Acc:HGNC:23094]
ENSG00000151812	ENSG00000151812	SLC35F4	solute carrier family 35 member F4 [Source:HGNC Symbol;Acc:HGNC:19845]
ENSG00000168453	ENSG00000168453	HR	HR lysine demethylase and nuclear receptor corepressor [Source:HGNC Symbol;Acc:HGNC:5172]
ENSG00000159231	ENSG00000159231	CBR3	carbonyl reductase 3 [Source:HGNC Symbol;Acc:HGNC:1549]
ENSG00000130433	ENSG00000130433	CACNG6	calcium voltage-gated channel auxiliary subunit gamma 6 [Source:HGNC Symbol;Acc:HGNC:13625]
ENSG00000179855	ENSG00000179855	GIPC3	GIPC PDZ domain containing family member 3 [Source:HGNC Symbol;Acc:HGNC:18183]
ENSG00000118194	ENSG00000118194	TNNT2	troponin T2, cardiac type [Source:HGNC Symbol;Acc:HGNC:11949]
ENSG00000185915	ENSG00000185915	KLHL34	kelch like family member 34 [Source:HGNC Symbol;Acc:HGNC:26634]
ENSG00000168993	ENSG00000168993	CPLX1	complexin 1 [Source:HGNC Symbol;Acc:HGNC:2309]
ENSG00000189409	ENSG00000189409	MMP23B	matrix metallopeptidase 23B [Source:HGNC Symbol;Acc:HGNC:7171]
ENSG00000162543	ENSG00000162543	UBXN10	UBX domain protein 10 [Source:HGNC Symbol;Acc:HGNC:26354]
ENSG00000283930	ENSG00000283930	PLD5P1	PLD5 pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:55072]
ENSG00000168421	ENSG00000168421	RHOH	ras homolog family member H [Source:HGNC Symbol;Acc:HGNC:686]
ENSG00000161896	ENSG00000161896	IP6K3	inositol hexakisphosphate kinase 3 [Source:HGNC Symbol;Acc:HGNC:17269]
ENSG00000177272	ENSG00000177272	KCNA3	potassium voltage-gated channel subfamily A member 3 [Source:HGNC Symbol;Acc:HGNC:6221]
ENSG00000141744	ENSG00000141744	PNMT	phenylethanolamine N-methyltransferase [Source:HGNC Symbol;Acc:HGNC:9160]
ENSG00000179921	ENSG00000179921	GPBAR1	G protein-coupled bile acid receptor 1 [Source:HGNC Symbol;Acc:HGNC:19680]
ENSG00000015413	ENSG00000015413	DPEP1	dipeptidase 1 [Source:HGNC Symbol;Acc:HGNC:3002]
ENSG00000196277	ENSG00000196277	GRM7	glutamate metabotropic receptor 7 [Source:HGNC Symbol;Acc:HGNC:4599]
ENSG00000143627	ENSG00000143627	PKLR	pyruvate kinase L/R [Source:HGNC Symbol;Acc:HGNC:9020]
ENSG00000067048	ENSG00000067048	DDX3Y	DEAD-box helicase 3 Y-linked [Source:HGNC Symbol;Acc:HGNC:2699]
ENSG00000123384	ENSG00000123384	LRP1	LDL receptor related protein 1 [Source:HGNC Symbol;Acc:HGNC:6692]
ENSG00000114374	ENSG00000114374	USP9Y	ubiquitin specific peptidase 9 Y-linked [Source:HGNC Symbol;Acc:HGNC:12633]
ENSG00000182481	ENSG00000182481	KPNA2	karyopherin subunit alpha 2 [Source:HGNC Symbol;Acc:HGNC:6395]
ENSG00000128567	ENSG00000128567	PODXL	podocalyxin like [Source:HGNC Symbol;Acc:HGNC:9171]
ENSG00000126457	ENSG00000126457	PRMT1	protein arginine methyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:5187]
ENSG00000136830	ENSG00000136830	NIBAN2	niban apoptosis regulator 2 [Source:HGNC Symbol;Acc:HGNC:25282]
ENSG00000198467	ENSG00000198467	TPM2	tropomyosin 2 [Source:HGNC Symbol;Acc:HGNC:12011]
ENSG00000171867	ENSG00000171867	PRNP	prion protein [Source:HGNC Symbol;Acc:HGNC:9449]
ENSG00000106624	ENSG00000106624	AEBP1	AE binding protein 1 [Source:HGNC Symbol;Acc:HGNC:303]
ENSG00000110492	ENSG00000110492	MDK	midkine [Source:HGNC Symbol;Acc:HGNC:6972]
ENSG00000086475	ENSG00000086475	SEPHS1	selenophosphate synthetase 1 [Source:HGNC Symbol;Acc:HGNC:19685]
ENSG00000152377	ENSG00000152377	SPOCK1	SPARC (osteonectin), cwcv and kazal like domains proteoglycan 1 [Source:HGNC Symbol;Acc:HGNC
ENSG00000117114	ENSG00000117114	ADGRL2	adhesion G protein-coupled receptor L2 [Source:HGNC Symbol;Acc:HGNC:18582]
ENSG00000117394	ENSG00000117394	SLC2A1	solute carrier family 2 member 1 [Source:HGNC Symbol;Acc:HGNC:11005]
ENSG00000121741	ENSG00000121741	ZMYM2	zinc finger MYM-type containing 2 [Source:HGNC Symbol;Acc:HGNC:12989]
ENSG00000182871	ENSG00000182871	COL18A1	collagen type XVIII alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:2195]
ENSG00000112715	ENSG00000112715	VEGFA	vascular endothelial growth factor A [Source:HGNC Symbol;Acc:HGNC:12680]
ENSG00000177606	ENSG00000177606	JUN	Jun proto-oncogene, AP-1 transcription factor subunit [Source:HGNC Symbol;Acc:HGNC:6204]
ENSG00000013810	ENSG00000013810	TACC3	transforming acidic coiled-coil containing protein 3 [Source:HGNC Symbol;Acc:HGNC:11524]
ENSG00000090372	ENSG00000090372	STRN4	striatin 4 [Source:HGNC Symbol;Acc:HGNC:15721]
ENSG00000188486	ENSG00000188486	H2AX	H2A.X variant histone [Source:HGNC Symbol;Acc:HGNC:4739]
ENSG00000008083	ENSG00000008083	JARID2	jumonji and AT-rich interaction domain containing 2 [Source:HGNC Symbol;Acc:HGNC:6196]
ENSG00000142669	ENSG00000142669	SH3BGRL3	SH3 domain binding glutamate rich protein like 3 [Source:HGNC Symbol;Acc:HGNC:15568]
ENSG00000120738	ENSG00000120738	EGR1	early growth response 1 [Source:HGNC Symbol;Acc:HGNC:3238]
ENSG00000165821	ENSG00000165821	SALL2	spalt like transcription factor 2 [Source:HGNC Symbol;Acc:HGNC:10526]
ENSG00000130202	ENSG00000130202	NECTIN2	nectin cell adhesion molecule 2 [Source:HGNC Symbol;Acc:HGNC:9707]
ENSG00000175550	ENSG00000175550	DRAP1	DR1 associated protein 1 [Source:HGNC Symbol;Acc:HGNC:3019]
ENSG00000197535	ENSG00000197535	MYO5A	myosin VA [Source:HGNC Symbol;Acc:HGNC:7602]
ENSG00000114270	ENSG00000114270	COL7A1	collagen type VII alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:2214]
ENSG00000130203	ENSG00000130203	APOE	apolipoprotein E [Source:HGNC Symbol;Acc:HGNC:613]
ENSG00000106976	ENSG00000106976	DNM1	dynamin 1 [Source:HGNC Symbol;Acc:HGNC:2972]
ENSG00000096433	ENSG00000096433	ITPR3	inositol 1,4,5-trisphosphate receptor type 3 [Source:HGNC Symbol;Acc:HGNC:6182]
ENSG00000101384	ENSG00000101384	JAG1	jagged canonical Notch ligand 1 [Source:HGNC Symbol;Acc:HGNC:6188]
ENSG00000161011	ENSG00000161011	SQSTM1	sequestosome 1 [Source:HGNC Symbol;Acc:HGNC:11280]
ENSG00000168758	ENSG00000168758	SEMA4C	semaphorin 4C [Source:HGNC Symbol;Acc:HGNC:10731]
ENSG00000144730	ENSG00000144730	IL17RD	interleukin 17 receptor D [Source:HGNC Symbol;Acc:HGNC:17616]
ENSG00000107331	ENSG00000107331	ABCA2	ATP binding cassette subfamily A member 2 [Source:HGNC Symbol;Acc:HGNC:32]
ENSG00000106367	ENSG00000106367	AP1S1	adaptor related protein complex 1 subunit sigma 1 [Source:HGNC Symbol;Acc:HGNC:559]
ENSG00000102401	ENSG00000102401	ARMCX3	armadillo repeat containing X-linked 3 [Source:HGNC Symbol;Acc:HGNC:24065]
ENSG00000132718	ENSG00000132718	SYT11	synaptotagmin 11 [Source:HGNC Symbol;Acc:HGNC:19239]
ENSG00000118971	ENSG00000118971	CCND2	cyclin D2 [Source:HGNC Symbol;Acc:HGNC:1583]
ENSG00000160360	ENSG00000160360	GPSM1	G protein signaling modulator 1 [Source:HGNC Symbol;Acc:HGNC:17858]
ENSG00000167123	ENSG00000167123	CERCAM	cerebral endothelial cell adhesion molecule [Source:HGNC Symbol;Acc:HGNC:23723]
ENSG00000111252	ENSG00000111252	SH2B3	SH2B adaptor protein 3 [Source:HGNC Symbol;Acc:HGNC:29605]
ENSG00000074219	ENSG00000074219	TEAD2	TEA domain transcription factor 2 [Source:HGNC Symbol;Acc:HGNC:11715]
ENSG00000101115	ENSG00000101115	SALL4	spalt like transcription factor 4 [Source:HGNC Symbol;Acc:HGNC:15924]
ENSG00000091436	ENSG00000091436	MAP3K20	mitogen-activated protein kinase kinase kinase 20 [Source:HGNC Symbol;Acc:HGNC:17797]
ENSG00000107281	ENSG00000107281	NPDC1	neural proliferation, differentiation and control 1 [Source:HGNC Symbol;Acc:HGNC:7899]
ENSG00000136854	ENSG00000136854	STXBP1	syntaxin binding protein 1 [Source:HGNC Symbol;Acc:HGNC:11444]
ENSG00000128602	ENSG00000128602	SMO	smoothened, frizzled class receptor [Source:HGNC Symbol;Acc:HGNC:11119]
ENSG00000166825	ENSG00000166825	ANPEP	alanyl aminopeptidase, membrane [Source:HGNC Symbol;Acc:HGNC:500]
ENSG00000066322	ENSG00000066322	ELOVL1	ELOVL fatty acid elongase 1 [Source:HGNC Symbol;Acc:HGNC:14418]
ENSG00000105486	ENSG00000105486	LIG1	DNA ligase 1 [Source:HGNC Symbol;Acc:HGNC:6598]
ENSG00000065923	ENSG00000065923	SLC9A7	solute carrier family 9 member A7 [Source:HGNC Symbol;Acc:HGNC:17123]
ENSG00000168056	ENSG00000168056	LTBP3	latent transforming growth factor beta binding protein 3 [Source:HGNC Symbol;Acc:HGNC:6716]
ENSG00000006327	ENSG00000006327	TNFRSF12A	TNF receptor superfamily member 12A [Source:HGNC Symbol;Acc:HGNC:18152]
ENSG00000169851	ENSG00000169851	PCDH7	protocadherin 7 [Source:HGNC Symbol;Acc:HGNC:8659]
ENSG00000046604	ENSG00000046604	DSG2	desmoglein 2 [Source:HGNC Symbol;Acc:HGNC:3049]
ENSG00000082397	ENSG00000082397	EPB41L3	erythrocyte membrane protein band 4.1 like 3 [Source:HGNC Symbol;Acc:HGNC:3380]
ENSG00000102996	ENSG00000102996	MMP15	matrix metallopeptidase 15 [Source:HGNC Symbol;Acc:HGNC:7161]
ENSG00000170345	ENSG00000170345	FOS	Fos proto-oncogene, AP-1 transcription factor subunit [Source:HGNC Symbol;Acc:HGNC:3796]
ENSG00000179361	ENSG00000179361	ARID3B	AT-rich interaction domain 3B [Source:HGNC Symbol;Acc:HGNC:14350]
ENSG00000004864	ENSG00000004864	SLC25A13	solute carrier family 25 member 13 [Source:HGNC Symbol;Acc:HGNC:10983]
ENSG00000116017	ENSG00000116017	ARID3A	AT-rich interaction domain 3A [Source:HGNC Symbol;Acc:HGNC:3031]
ENSG00000163531	ENSG00000163531	NFASC	neurofascin [Source:HGNC Symbol;Acc:HGNC:29866]
ENSG00000078177	ENSG00000078177	N4BP2	NEDD4 binding protein 2 [Source:HGNC Symbol;Acc:HGNC:29851]
ENSG00000154122	ENSG00000154122	ANKH	ANKH inorganic pyrophosphate transport regulator [Source:HGNC Symbol;Acc:HGNC:15492]
ENSG00000159403	ENSG00000159403	C1R	complement C1r [Source:HGNC Symbol;Acc:HGNC:1246]
ENSG00000009413	ENSG00000009413	REV3L	REV3 like, DNA directed polymerase zeta catalytic subunit [Source:HGNC Symbol;Acc:HGNC:9968]
ENSG00000147852	ENSG00000147852	VLDLR	very low density lipoprotein receptor [Source:HGNC Symbol;Acc:HGNC:12698]
ENSG00000145703	ENSG00000145703	IQGAP2	IQ motif containing GTPase activating protein 2 [Source:HGNC Symbol;Acc:HGNC:6111]
ENSG00000197457	ENSG00000197457	STMN3	stathmin 3 [Source:HGNC Symbol;Acc:HGNC:15926]
ENSG00000068650	ENSG00000068650	ATP11A	ATPase phospholipid transporting 11A [Source:HGNC Symbol;Acc:HGNC:13552]
ENSG00000173442	ENSG00000173442	EHBP1L1	EH domain binding protein 1 like 1 [Source:HGNC Symbol;Acc:HGNC:30682]
ENSG00000196549	ENSG00000196549	MME	membrane metalloendopeptidase [Source:HGNC Symbol;Acc:HGNC:7154]
ENSG00000135525	ENSG00000135525	MAP7	microtubule associated protein 7 [Source:HGNC Symbol;Acc:HGNC:6869]
ENSG00000198853	ENSG00000198853	RUSC2	RUN and SH3 domain containing 2 [Source:HGNC Symbol;Acc:HGNC:23625]
ENSG00000011347	ENSG00000011347	SYT7	synaptotagmin 7 [Source:HGNC Symbol;Acc:HGNC:11514]
ENSG00000107819	ENSG00000107819	SFXN3	sideroflexin 3 [Source:HGNC Symbol;Acc:HGNC:16087]
ENSG00000101236	ENSG00000101236	RNF24	ring finger protein 24 [Source:HGNC Symbol;Acc:HGNC:13779]
ENSG00000165757	ENSG00000165757	JCAD	junctional cadherin 5 associated [Source:HGNC Symbol;Acc:HGNC:29283]
ENSG00000152284	ENSG00000152284	TCF7L1	transcription factor 7 like 1 [Source:HGNC Symbol;Acc:HGNC:11640]
ENSG00000110074	ENSG00000110074	FOXRED1	FAD dependent oxidoreductase domain containing 1 [Source:HGNC Symbol;Acc:HGNC:26927]
ENSG00000039068	ENSG00000039068	CDH1	cadherin 1 [Source:HGNC Symbol;Acc:HGNC:1748]
ENSG00000167779	ENSG00000167779	IGFBP6	insulin like growth factor binding protein 6 [Source:HGNC Symbol;Acc:HGNC:5475]
ENSG00000171223	ENSG00000171223	JUNB	JunB proto-oncogene, AP-1 transcription factor subunit [Source:HGNC Symbol;Acc:HGNC:6205]
ENSG00000010818	ENSG00000010818	HIVEP2	HIVEP zinc finger 2 [Source:HGNC Symbol;Acc:HGNC:4921]
ENSG00000166173	ENSG00000166173	LARP6	La ribonucleoprotein 6, translational regulator [Source:HGNC Symbol;Acc:HGNC:24012]
ENSG00000120129	ENSG00000120129	DUSP1	dual specificity phosphatase 1 [Source:HGNC Symbol;Acc:HGNC:3064]
ENSG00000169635	ENSG00000169635	HIC2	HIC ZBTB transcriptional repressor 2 [Source:HGNC Symbol;Acc:HGNC:18595]
ENSG00000148737	ENSG00000148737	TCF7L2	transcription factor 7 like 2 [Source:HGNC Symbol;Acc:HGNC:11641]
ENSG00000105699	ENSG00000105699	LSR	lipolysis stimulated lipoprotein receptor [Source:HGNC Symbol;Acc:HGNC:29572]
ENSG00000008513	ENSG00000008513	ST3GAL1	ST3 beta-galactoside alpha-2,3-sialyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:10862]
ENSG00000249992	ENSG00000249992	TMEM158	transmembrane protein 158 [Source:HGNC Symbol;Acc:HGNC:30293]
ENSG00000151883	ENSG00000151883	PARP8	poly(ADP-ribose) polymerase family member 8 [Source:HGNC Symbol;Acc:HGNC:26124]
ENSG00000152465	ENSG00000152465	NMT2	N-myristoyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:7858]
ENSG00000112559	ENSG00000112559	MDFI	MyoD family inhibitor [Source:HGNC Symbol;Acc:HGNC:6967]
ENSG00000056998	ENSG00000056998	GYG2	glycogenin 2 [Source:HGNC Symbol;Acc:HGNC:4700]
ENSG00000092758	ENSG00000092758	COL9A3	collagen type IX alpha 3 chain [Source:HGNC Symbol;Acc:HGNC:2219]
ENSG00000108947	ENSG00000108947	EFNB3	ephrin B3 [Source:HGNC Symbol;Acc:HGNC:3228]
ENSG00000157570	ENSG00000157570	TSPAN18	tetraspanin 18 [Source:HGNC Symbol;Acc:HGNC:20660]
ENSG00000085433	ENSG00000085433	WDR47	WD repeat domain 47 [Source:HGNC Symbol;Acc:HGNC:29141]
ENSG00000173559	ENSG00000173559	NABP1	nucleic acid binding protein 1 [Source:HGNC Symbol;Acc:HGNC:26232]
ENSG00000180447	ENSG00000180447	GAS1	growth arrest specific 1 [Source:HGNC Symbol;Acc:HGNC:4165]
ENSG00000148200	ENSG00000148200	NR6A1	nuclear receptor subfamily 6 group A member 1 [Source:HGNC Symbol;Acc:HGNC:7985]
ENSG00000162551	ENSG00000162551	ALPL	alkaline phosphatase, biomineralization associated [Source:HGNC Symbol;Acc:HGNC:438]
ENSG00000162545	ENSG00000162545	CAMK2N1	calcium/calmodulin dependent protein kinase II inhibitor 1 [Source:HGNC Symbol;Acc:HGNC:24190
ENSG00000134138	ENSG00000134138	MEIS2	Meis homeobox 2 [Source:HGNC Symbol;Acc:HGNC:7001]
ENSG00000198598	ENSG00000198598	MMP17	matrix metallopeptidase 17 [Source:HGNC Symbol;Acc:HGNC:7163]
ENSG00000082014	ENSG00000082014	SMARCD3	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, membe
ENSG00000197635	ENSG00000197635	DPP4	dipeptidyl peptidase 4 [Source:HGNC Symbol;Acc:HGNC:3009]
ENSG00000176834	ENSG00000176834	VSIG10	V-set and immunoglobulin domain containing 10 [Source:HGNC Symbol;Acc:HGNC:26078]
ENSG00000136383	ENSG00000136383	ALPK3	alpha kinase 3 [Source:HGNC Symbol;Acc:HGNC:17574]
ENSG00000105472	ENSG00000105472	CLEC11A	C-type lectin domain containing 11A [Source:HGNC Symbol;Acc:HGNC:10576]
ENSG00000133131	ENSG00000133131	MORC4	MORC family CW-type zinc finger 4 [Source:HGNC Symbol;Acc:HGNC:23485]
ENSG00000150907	ENSG00000150907	FOXO1	forkhead box O1 [Source:HGNC Symbol;Acc:HGNC:3819]
ENSG00000131435	ENSG00000131435	PDLIM4	PDZ and LIM domain 4 [Source:HGNC Symbol;Acc:HGNC:16501]
ENSG00000093010	ENSG00000093010	COMT	catechol-O-methyltransferase [Source:HGNC Symbol;Acc:HGNC:2228]
ENSG00000008441	ENSG00000008441	NFIX	nuclear factor I X [Source:HGNC Symbol;Acc:HGNC:7788]
ENSG00000196557	ENSG00000196557	CACNA1H	calcium voltage-gated channel subunit alpha1 H [Source:HGNC Symbol;Acc:HGNC:1395]
ENSG00000152495	ENSG00000152495	CAMK4	calcium/calmodulin dependent protein kinase IV [Source:HGNC Symbol;Acc:HGNC:1464]
ENSG00000198105	ENSG00000198105	ZNF248	zinc finger protein 248 [Source:HGNC Symbol;Acc:HGNC:13041]
ENSG00000069188	ENSG00000069188	SDK2	sidekick cell adhesion molecule 2 [Source:HGNC Symbol;Acc:HGNC:19308]
ENSG00000104081	ENSG00000104081	BMF	Bcl2 modifying factor [Source:HGNC Symbol;Acc:HGNC:24132]
ENSG00000122008	ENSG00000122008	POLK	DNA polymerase kappa [Source:HGNC Symbol;Acc:HGNC:9183]
ENSG00000063180	ENSG00000063180	CA11	carbonic anhydrase 11 [Source:HGNC Symbol;Acc:HGNC:1370]
ENSG00000119865	ENSG00000119865	CNRIP1	cannabinoid receptor interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:24546]

Supplementary Table 7-4: list of DE genes enriched for MOVO-B motif

MOVO-B motif			
initial_alias	converted_alias	name	description
ENSG00000150907	ENSG00000150907	FOXO1	forkhead box O1 [Source:HGNC Symbol;Acc:HGNC:3819]
ENSG00000074416	ENSG00000074416	MGLL	monoglyceride lipase [Source:HGNC Symbol;Acc:HGNC:17038]
ENSG00000196155	ENSG00000196155	PLEKHG4	pleckstrin homology and RhoGEF domain containing G4 [Source:HGNC Symbol;Acc:HGNC:24501]
ENSG00000131435	ENSG00000131435	PDLIM4	PDZ and LIM domain 4 [Source:HGNC Symbol;Acc:HGNC:16501]
ENSG00000093010	ENSG00000093010	COMT	catechol-O-methyltransferase [Source:HGNC Symbol;Acc:HGNC:2228]
ENSG00000167778	ENSG00000167778	SPRYD3	SPRY domain containing 3 [Source:HGNC Symbol;Acc:HGNC:25920]
ENSG00000008441	ENSG00000008441	NFIX	nuclear factor I X [Source:HGNC Symbol;Acc:HGNC:7788]
ENSG00000196557	ENSG00000196557	CACNA1H	calcium voltage-gated channel subunit alpha1 H [Source:HGNC Symbol;Acc:HGNC:1395]
ENSG00000132205	ENSG00000132205	EMILIN2	elastin microfibril interfacer 2 [Source:HGNC Symbol;Acc:HGNC:19881]
ENSG00000103742	ENSG00000103742	IGDCC4	immunoglobulin superfamily DCC subclass member 4 [Source:HGNC Symbol;Acc:HGNC:13770]
ENSG00000153823	ENSG00000153823	PID1	phosphotyrosine interaction domain containing 1 [Source:HGNC Symbol;Acc:HGNC:26084]
ENSG00000111275	ENSG00000111275	ALDH2	aldehyde dehydrogenase 2 family member [Source:HGNC Symbol;Acc:HGNC:404]
ENSG00000069188	ENSG00000069188	SDK2	sidekick cell adhesion molecule 2 [Source:HGNC Symbol;Acc:HGNC:19308]
ENSG00000156925	ENSG00000156925	ZIC3	Zic family member 3 [Source:HGNC Symbol;Acc:HGNC:12874]
ENSG00000170561	ENSG00000170561	IRX2	iroquois homeobox 2 [Source:HGNC Symbol;Acc:HGNC:14359]
ENSG00000178409	ENSG00000178409	BEND3	BEN domain containing 3 [Source:HGNC Symbol;Acc:HGNC:23040]
ENSG00000075275	ENSG00000075275	CELSR1	cadherin EGF LAG seven-pass G-type receptor 1 [Source:HGNC Symbol;Acc:HGNC:1850]
ENSG00000128849	ENSG00000128849	CGNL1	cingulin like 1 [Source:HGNC Symbol;Acc:HGNC:25931]
ENSG00000115594	ENSG00000115594	IL1R1	interleukin 1 receptor type 1 [Source:HGNC Symbol;Acc:HGNC:5993]
ENSG00000196247	ENSG00000196247	ZNF107	zinc finger protein 107 [Source:HGNC Symbol;Acc:HGNC:12887]
ENSG00000172216	ENSG00000172216	CEBPB	CCAAT enhancer binding protein beta [Source:HGNC Symbol;Acc:HGNC:1834]
ENSG00000185022	ENSG00000185022	MAFF	MAF bZIP transcription factor F [Source:HGNC Symbol;Acc:HGNC:6780]
ENSG00000168646	ENSG00000168646	AXIN2	axin 2 [Source:HGNC Symbol;Acc:HGNC:904]
ENSG00000128973	ENSG00000128973	CLN6	CLN6 transmembrane ER protein [Source:HGNC Symbol;Acc:HGNC:2077]
ENSG00000112773	ENSG00000112773	TENT5A	terminal nucleotidyltransferase 5A [Source:HGNC Symbol;Acc:HGNC:18345]
ENSG00000116396	ENSG00000116396	KCNC4	potassium voltage-gated channel subfamily C member 4 [Source:HGNC Symbol;Acc:HGNC:6236]
ENSG00000105376	ENSG00000105376	ICAM5	intercellular adhesion molecule 5 [Source:HGNC Symbol;Acc:HGNC:5348]
ENSG00000133216	ENSG00000133216	EPHB2	EPH receptor B2 [Source:HGNC Symbol;Acc:HGNC:3393]
ENSG00000118276	ENSG00000118276	B4GALT6	beta-1,4-galactosyltransferase 6 [Source:HGNC Symbol;Acc:HGNC:929]
ENSG00000077943	ENSG00000077943	ITGA8	integrin subunit alpha 8 [Source:HGNC Symbol;Acc:HGNC:6144]
ENSG00000181234	ENSG00000181234	TMEM132C	transmembrane protein 132C [Source:HGNC Symbol;Acc:HGNC:25436]
ENSG00000219438	ENSG00000219438	TAFA5	TAFA chemokine like family member 5 [Source:HGNC Symbol;Acc:HGNC:21592]
ENSG00000166831	ENSG00000166831	RBPMS2	RNA binding protein, mRNA processing factor 2 [Source:HGNC Symbol;Acc:HGNC:19098]
ENSG00000101144	ENSG00000101144	BMP7	bone morphogenetic protein 7 [Source:HGNC Symbol;Acc:HGNC:1074]
ENSG00000114853	ENSG00000114853	ZBTB47	zinc finger and BTB domain containing 47 [Source:HGNC Symbol;Acc:HGNC:26955]
ENSG00000175445	ENSG00000175445	LPL	lipoprotein lipase [Source:HGNC Symbol;Acc:HGNC:6677]
ENSG00000150510	ENSG00000150510	FAM124A	family with sequence similarity 124 member A [Source:HGNC Symbol;Acc:HGNC:26413]
ENSG00000091844	ENSG00000091844	RGS17	regulator of G protein signaling 17 [Source:HGNC Symbol;Acc:HGNC:14088]
ENSG00000129946	ENSG00000129946	SHC2	SHC adaptor protein 2 [Source:HGNC Symbol;Acc:HGNC:29869]
ENSG00000123191	ENSG00000123191	ATP7B	ATPase copper transporting beta [Source:HGNC Symbol;Acc:HGNC:870]
ENSG00000114315	ENSG00000114315	HES1	hes family bHLH transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:5192]
ENSG00000175352	ENSG00000175352	NRIP3	nuclear receptor interacting protein 3 [Source:HGNC Symbol;Acc:HGNC:1167]
ENSG00000118503	ENSG00000118503	TNFAIP3	TNF alpha induced protein 3 [Source:HGNC Symbol;Acc:HGNC:11896]
ENSG00000182240	ENSG00000182240	BACE2	beta-secretase 2 [Source:HGNC Symbol;Acc:HGNC:934]
ENSG00000159216	ENSG00000159216	RUNX1	RUNX family transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:10471]
ENSG00000168477	ENSG00000168477	TNXB	tenascin XB [Source:HGNC Symbol;Acc:HGNC:11976]
ENSG00000170412	ENSG00000170412	GPRC5C	G protein-coupled receptor class C group 5 member C [Source:HGNC Symbol;Acc:HGNC:13309]
ENSG00000135083	ENSG00000135083	CCNJL	cyclin J like [Source:HGNC Symbol;Acc:HGNC:25876]
ENSG00000155816	ENSG00000155816	FMN2	formin 2 [Source:HGNC Symbol;Acc:HGNC:14074]
ENSG00000166780	ENSG00000166780	BMERB1	bMERB domain containing 1 [Source:HGNC Symbol;Acc:HGNC:19213]
ENSG00000198780	ENSG00000198780	FAM169A	family with sequence similarity 169 member A [Source:HGNC Symbol;Acc:HGNC:29138]
ENSG00000115194	ENSG00000115194	SLC30A3	solute carrier family 30 member 3 [Source:HGNC Symbol;Acc:HGNC:11014]
ENSG00000143669	ENSG00000143669	LYST	lysosomal trafficking regulator [Source:HGNC Symbol;Acc:HGNC:1968]
ENSG00000187123	ENSG00000187123	LYPD6	LY6/PLAUR domain containing 6 [Source:HGNC Symbol;Acc:HGNC:28751]
ENSG00000162804	ENSG00000162804	SNED1	sushi, nidogen and EGF like domains 1 [Source:HGNC Symbol;Acc:HGNC:24696]
ENSG00000046889	ENSG00000046889	PREX2	phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 2 [Source:HGNC Symbol;A
ENSG00000105755	ENSG00000105755	ETHE1	ETHE1 persulfide dioxygenase [Source:HGNC Symbol;Acc:HGNC:23287]
ENSG00000188766	ENSG00000188766	SPRED3	sprouty related EVH1 domain containing 3 [Source:HGNC Symbol;Acc:HGNC:31041]
ENSG00000158050	ENSG00000158050	DUSP2	dual specificity phosphatase 2 [Source:HGNC Symbol;Acc:HGNC:3068]
ENSG00000161681	ENSG00000161681	SHANK1	SH3 and multiple ankyrin repeat domains 1 [Source:HGNC Symbol;Acc:HGNC:15474]
ENSG00000105855	ENSG00000105855	ITGB8	integrin subunit beta 8 [Source:HGNC Symbol;Acc:HGNC:6163]
ENSG00000152154	ENSG00000152154	TMEM178A	transmembrane protein 178A [Source:HGNC Symbol;Acc:HGNC:28517]
ENSG00000078114	ENSG00000078114	NEBL	nebulette [Source:HGNC Symbol;Acc:HGNC:16932]
ENSG00000125675	ENSG00000125675	GRIA3	glutamate ionotropic receptor AMPA type subunit 3 [Source:HGNC Symbol;Acc:HGNC:4573]
ENSG00000130363	ENSG00000130363	RSPH3	radial spoke head 3 [Source:HGNC Symbol;Acc:HGNC:21054]
ENSG00000112139	ENSG00000112139	MDGA1	MAM domain containing glycosylphosphatidylinositol anchor 1 [Source:HGNC Symbol;Acc:HGNC:19
ENSG00000187122	ENSG00000187122	SLIT1	slit guidance ligand 1 [Source:HGNC Symbol;Acc:HGNC:11085]
ENSG00000158122	ENSG00000158122	PRXL2C	peroxiredoxin like 2C [Source:HGNC Symbol;Acc:HGNC:16881]
ENSG00000147408	ENSG00000147408	CSGALNACT1	chondroitin sulfate N-acetylgalactosaminyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:24290]
ENSG00000184156	ENSG00000184156	KCNQ3	potassium voltage-gated channel subfamily Q member 3 [Source:HGNC Symbol;Acc:HGNC:6297]
ENSG00000152092	ENSG00000152092	ASTN1	astrotactin 1 [Source:HGNC Symbol;Acc:HGNC:773]
ENSG00000140807	ENSG00000140807	NKD1	NKD inhibitor of WNT signaling pathway 1 [Source:HGNC Symbol;Acc:HGNC:17045]
ENSG00000150630	ENSG00000150630	VEGFC	vascular endothelial growth factor C [Source:HGNC Symbol;Acc:HGNC:12682]
ENSG00000167992	ENSG00000167992	VWCE	von Willebrand factor C and EGF domains [Source:HGNC Symbol;Acc:HGNC:26487]
ENSG00000054690	ENSG00000054690	PLEKHH1	pleckstrin homology, MyTH4 and FERM domain containing H1 [Source:HGNC Symbol;Acc:HGNC:17
ENSG00000197442	ENSG00000197442	MAP3K5	mitogen-activated protein kinase kinase kinase 5 [Source:HGNC Symbol;Acc:HGNC:6857]
ENSG00000100918	ENSG00000100918	REC8	REC8 meiotic recombination protein [Source:HGNC Symbol;Acc:HGNC:16879]
ENSG00000075643	ENSG00000075643	MOCOS	molybdenum cofactor sulfurase [Source:HGNC Symbol;Acc:HGNC:18234]
ENSG00000170801	ENSG00000170801	HTRA3	HtrA serine peptidase 3 [Source:HGNC Symbol;Acc:HGNC:30406]
ENSG00000182379	ENSG00000182379	NXPH4	neurexophilin 4 [Source:HGNC Symbol;Acc:HGNC:8078]
ENSG00000154654	ENSG00000154654	NCAM2	neural cell adhesion molecule 2 [Source:HGNC Symbol;Acc:HGNC:7657]
ENSG00000113296	ENSG00000113296	THBS4	thrombospondin 4 [Source:HGNC Symbol;Acc:HGNC:11788]
ENSG00000173926	ENSG00000173926	MARCHF3	membrane associated ring-CH-type finger 3 [Source:HGNC Symbol;Acc:HGNC:28728]
ENSG00000041880	ENSG00000041880	PARP3	poly(ADP-ribose) polymerase family member 3 [Source:HGNC Symbol;Acc:HGNC:273]
ENSG00000129757	ENSG00000129757	CDKN1C	cyclin dependent kinase inhibitor 1C [Source:HGNC Symbol;Acc:HGNC:1786]
ENSG00000154553	ENSG00000154553	PDLIM3	PDZ and LIM domain 3 [Source:HGNC Symbol;Acc:HGNC:20767]
ENSG00000134668	ENSG00000134668	SPOCD1	SPOC domain containing 1 [Source:HGNC Symbol;Acc:HGNC:26338]
ENSG00000189143	ENSG00000189143	CLDN4	claudin 4 [Source:HGNC Symbol;Acc:HGNC:2046]
ENSG00000185666	ENSG00000185666	SYN3	synapsin III [Source:HGNC Symbol;Acc:HGNC:11496]
ENSG00000198353	ENSG00000198353	HOXC4	homeobox C4 [Source:HGNC Symbol;Acc:HGNC:5126]
ENSG00000050767	ENSG00000050767	COL23A1	collagen type XXIII alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:22990]
ENSG00000139835	ENSG00000139835	GRTP1	growth hormone regulated TBC protein 1 [Source:HGNC Symbol;Acc:HGNC:20310]
ENSG00000169946	ENSG00000169946	ZFPM2	zinc finger protein, FOG family member 2 [Source:HGNC Symbol;Acc:HGNC:16700]
ENSG00000138795	ENSG00000138795	LEF1	lymphoid enhancer binding factor 1 [Source:HGNC Symbol;Acc:HGNC:6551]
ENSG00000108950	ENSG00000108950	FAM20A	FAM20A golgi associated secretory pathway pseudokinase [Source:HGNC Symbol;Acc:HGNC:23015
ENSG00000135480	ENSG00000135480	KRT7	keratin 7 [Source:HGNC Symbol;Acc:HGNC:6445]
ENSG00000064692	ENSG00000064692	SNCAIP	synuclein alpha interacting protein [Source:HGNC Symbol;Acc:HGNC:11139]
ENSG00000204248	ENSG00000204248	COL11A2	collagen type XI alpha 2 chain [Source:HGNC Symbol;Acc:HGNC:2187]
ENSG00000136052	ENSG00000136052	SLC41A2	solute carrier family 41 member 2 [Source:HGNC Symbol;Acc:HGNC:31045]
ENSG00000132622	ENSG00000132622	HSPA12B	heat shock protein family A (Hsp70) member 12B [Source:HGNC Symbol;Acc:HGNC:16193]
ENSG00000135636	ENSG00000135636	DYSF	dysferlin [Source:HGNC Symbol;Acc:HGNC:3097]
ENSG00000135454	ENSG00000135454	B4GALNT1	beta-1,4-N-acetyl-galactosaminyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:4117]
ENSG00000248746	ENSG00000248746	ACTN3	actinin alpha 3 [Source:HGNC Symbol;Acc:HGNC:165]
ENSG00000162722	ENSG00000162722	TRIM58	tripartite motif containing 58 [Source:HGNC Symbol;Acc:HGNC:24150]
ENSG00000204219	ENSG00000204219	TCEA3	transcription elongation factor A3 [Source:HGNC Symbol;Acc:HGNC:11615]
ENSG00000171502	ENSG00000171502	COL24A1	collagen type XXIV alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:20821]
ENSG00000239264	ENSG00000239264	TXNDC5	thioredoxin domain containing 5 [Source:HGNC Symbol;Acc:HGNC:21073]
ENSG00000171246	ENSG00000171246	NPTX1	neuronal pentraxin 1 [Source:HGNC Symbol;Acc:HGNC:7952]
ENSG00000170837	ENSG00000170837	GPR27	G protein-coupled receptor 27 [Source:HGNC Symbol;Acc:HGNC:4482]
ENSG00000135472	ENSG00000135472	FAIM2	Fas apoptotic inhibitory molecule 2 [Source:HGNC Symbol;Acc:HGNC:17067]
ENSG00000013619	ENSG00000013619	MAMLD1	mastermind like domain containing 1 [Source:HGNC Symbol;Acc:HGNC:2568]
ENSG00000175084	ENSG00000175084	DES	desmin [Source:HGNC Symbol;Acc:HGNC:2770]
ENSG00000178222	ENSG00000178222	RNF212	ring finger protein 212 [Source:HGNC Symbol;Acc:HGNC:27729]
ENSG00000013293	ENSG00000013293	SLC7A14	solute carrier family 7 member 14 [Source:HGNC Symbol;Acc:HGNC:29326]
ENSG00000100031	ENSG00000100031	GGT1	gamma-glutamyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:4250]
ENSG00000119686	ENSG00000119686	FLVCR2	FLVCR heme transporter 2 [Source:HGNC Symbol;Acc:HGNC:20105]
ENSG00000118137	ENSG00000118137	APOA1	apolipoprotein A1 [Source:HGNC Symbol;Acc:HGNC:600]
ENSG00000137841	ENSG00000137841	PLCB2	phospholipase C beta 2 [Source:HGNC Symbol;Acc:HGNC:9055]
ENSG00000138675	ENSG00000138675	FGF5	fibroblast growth factor 5 [Source:HGNC Symbol;Acc:HGNC:3683]
ENSG00000164161	ENSG00000164161	HHIP	hedgehog interacting protein [Source:HGNC Symbol;Acc:HGNC:14866]
ENSG00000122420	ENSG00000122420	PTGFR	prostaglandin F receptor [Source:HGNC Symbol;Acc:HGNC:9600]
ENSG00000176244	ENSG00000176244	ACBD7	acyl-CoA binding domain containing 7 [Source:HGNC Symbol;Acc:HGNC:17715]
ENSG00000165730	ENSG00000165730	STOX1	storkhead box 1 [Source:HGNC Symbol;Acc:HGNC:23508]
ENSG00000167244	ENSG00000167244	IGF2	insulin like growth factor 2 [Source:HGNC Symbol;Acc:HGNC:5466]
ENSG00000140563	ENSG00000140563	MCTP2	multiple C2 and transmembrane domain containing 2 [Source:HGNC Symbol;Acc:HGNC:25636]
ENSG00000198286	ENSG00000198286	CARD11	caspase recruitment domain family member 11 [Source:HGNC Symbol;Acc:HGNC:16393]
ENSG00000134873	ENSG00000134873	CLDN10	claudin 10 [Source:HGNC Symbol;Acc:HGNC:2033]
ENSG00000205683	ENSG00000205683	DPF3	double PHD fingers 3 [Source:HGNC Symbol;Acc:HGNC:17427]
ENSG00000080854	ENSG00000080854	IGSF9B	immunoglobulin superfamily member 9B [Source:HGNC Symbol;Acc:HGNC:32326]
ENSG00000260428	ENSG00000260428	SCX	scleraxis bHLH transcription factor [Source:HGNC Symbol;Acc:HGNC:32322]
ENSG00000215845	ENSG00000215845	TSTD1	thiosulfate sulfurtransferase like domain containing 1 [Source:HGNC Symbol;Acc:HGNC:35410]
ENSG00000261308	ENSG00000261308	FIGNL2	fidgetin like 2 [Source:HGNC Symbol;Acc:HGNC:13287]
ENSG00000111679	ENSG00000111679	PTPN6	protein tyrosine phosphatase non-receptor type 6 [Source:HGNC Symbol;Acc:HGNC:9658]
ENSG00000269113	ENSG00000269113	TRABD2B	TraB domain containing 2B [Source:HGNC Symbol;Acc:HGNC:44200]
ENSG00000159733	ENSG00000159733	ZFYVE28	zinc finger FYVE-type containing 28 [Source:HGNC Symbol;Acc:HGNC:29334]
ENSG00000188730	ENSG00000188730	VWC2	von Willebrand factor C domain containing 2 [Source:HGNC Symbol;Acc:HGNC:30200]
ENSG00000139832	ENSG00000139832	RAB20	RAB20, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:18260]
ENSG00000152784	ENSG00000152784	PRDM8	PR/SET domain 8 [Source:HGNC Symbol;Acc:HGNC:13993]
ENSG00000123892	ENSG00000123892	RAB38	RAB38, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9776]
ENSG00000131620	ENSG00000131620	ANO1	anoctamin 1 [Source:HGNC Symbol;Acc:HGNC:21625]
ENSG00000087495	ENSG00000087495	PHACTR3	phosphatase and actin regulator 3 [Source:HGNC Symbol;Acc:HGNC:15833]
ENSG00000141574	ENSG00000141574	SECTM1	secreted and transmembrane 1 [Source:HGNC Symbol;Acc:HGNC:10707]
ENSG00000138311	ENSG00000138311	ZNF365	zinc finger protein 365 [Source:HGNC Symbol;Acc:HGNC:18194]
ENSG00000107282	ENSG00000107282	APBA1	amyloid beta precursor protein binding family A member 1 [Source:HGNC Symbol;Acc:HGNC:578]
ENSG00000128918	ENSG00000128918	ALDH1A2	aldehyde dehydrogenase 1 family member A2 [Source:HGNC Symbol;Acc:HGNC:15472]
ENSG00000203805	ENSG00000203805	PLPP4	phospholipid phosphatase 4 [Source:HGNC Symbol;Acc:HGNC:23531]
ENSG00000183018	ENSG00000183018	SPNS2	sphingolipid transporter 2 [Source:HGNC Symbol;Acc:HGNC:26992]
ENSG00000137825	ENSG00000137825	ITPKA	inositol-trisphosphate 3-kinase A [Source:HGNC Symbol;Acc:HGNC:6178]
ENSG00000058404	ENSG00000058404	CAMK2B	calcium/calmodulin dependent protein kinase II beta [Source:HGNC Symbol;Acc:HGNC:1461]
ENSG00000161940	ENSG00000161940	BCL6B	BCL6B transcription repressor [Source:HGNC Symbol;Acc:HGNC:1002]
ENSG00000081277	ENSG00000081277	PKP1	plakophilin 1 [Source:HGNC Symbol;Acc:HGNC:9023]
ENSG00000198133	ENSG00000198133	TMEM229B	transmembrane protein 229B [Source:HGNC Symbol;Acc:HGNC:20130]
ENSG00000185519	ENSG00000185519	FAM131C	family with sequence similarity 131 member C [Source:HGNC Symbol;Acc:HGNC:26717]
ENSG00000131668	ENSG00000131668	BARX1	BARX homeobox 1 [Source:HGNC Symbol;Acc:HGNC:955]
ENSG00000102109	ENSG00000102109	PCSK1N	proprotein convertase subtilisin/kexin type 1 inhibitor [Source:HGNC Symbol;Acc:HGNC:17301]
ENSG00000127528	ENSG00000127528	KLF2	Kruppel like factor 2 [Source:HGNC Symbol;Acc:HGNC:6347]
ENSG00000159166	ENSG00000159166	LAD1	ladinin 1 [Source:HGNC Symbol;Acc:HGNC:6472]
ENSG00000197208	ENSG00000197208	SLC22A4	solute carrier family 22 member 4 [Source:HGNC Symbol;Acc:HGNC:10968]
ENSG00000134259	ENSG00000134259	NGF	nerve growth factor [Source:HGNC Symbol;Acc:HGNC:7808]
ENSG00000172794	ENSG00000172794	RAB37	RAB37, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:30268]
ENSG00000175874	ENSG00000175874	CREG2	cellular repressor of E1A stimulated genes 2 [Source:HGNC Symbol;Acc:HGNC:14272]
ENSG00000134470	ENSG00000134470	IL15RA	interleukin 15 receptor subunit alpha [Source:HGNC Symbol;Acc:HGNC:5978]
ENSG00000151650	ENSG00000151650	VENTX	VENT homeobox [Source:HGNC Symbol;Acc:HGNC:13639]
ENSG00000169758	ENSG00000169758	TMEM266	transmembrane protein 266 [Source:HGNC Symbol;Acc:HGNC:26763]
ENSG00000170608	ENSG00000170608	FOXA3	forkhead box A3 [Source:HGNC Symbol;Acc:HGNC:5023]
ENSG00000164690	ENSG00000164690	SHH	sonic hedgehog signaling molecule [Source:HGNC Symbol;Acc:HGNC:10848]
ENSG00000204335	ENSG00000204335	SP5	Sp5 transcription factor [Source:HGNC Symbol;Acc:HGNC:14529]
ENSG00000170689	ENSG00000170689	HOXB9	homeobox B9 [Source:HGNC Symbol;Acc:HGNC:5120]
ENSG00000128713	ENSG00000128713	HOXD11	homeobox D11 [Source:HGNC Symbol;Acc:HGNC:5134]
ENSG00000203867	ENSG00000203867	RBM20	RNA binding motif protein 20 [Source:HGNC Symbol;Acc:HGNC:27424]
ENSG00000196581	ENSG00000196581	AJAP1	adherens junctions associated protein 1 [Source:HGNC Symbol;Acc:HGNC:30801]
ENSG00000151812	ENSG00000151812	SLC35F4	solute carrier family 35 member F4 [Source:HGNC Symbol;Acc:HGNC:19845]
ENSG00000168453	ENSG00000168453	HR	HR lysine demethylase and nuclear receptor corepressor [Source:HGNC Symbol;Acc:HGNC:5172]
ENSG00000171794	ENSG00000171794	UTF1	undifferentiated embryonic cell transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:12634]
ENSG00000176076	ENSG00000176076	KCNE5	potassium voltage-gated channel subfamily E regulatory subunit 5 [Source:HGNC Symbol;Acc:HGN
ENSG00000092345	ENSG00000092345	DAZL	deleted in azoospermia like [Source:HGNC Symbol;Acc:HGNC:2685]
ENSG00000171817	ENSG00000171817	ZNF540	zinc finger protein 540 [Source:HGNC Symbol;Acc:HGNC:25331]
ENSG00000168993	ENSG00000168993	CPLX1	complexin 1 [Source:HGNC Symbol;Acc:HGNC:2309]
ENSG00000189409	ENSG00000189409	MMP23B	matrix metallopeptidase 23B [Source:HGNC Symbol;Acc:HGNC:7171]
ENSG00000119147	ENSG00000119147	ECRG4	ECRG4 augurin precursor [Source:HGNC Symbol;Acc:HGNC:24642]
ENSG00000180767	ENSG00000180767	CHST13	carbohydrate sulfotransferase 13 [Source:HGNC Symbol;Acc:HGNC:21755]
ENSG00000147647	ENSG00000147647	DPYS	dihydropyrimidinase [Source:HGNC Symbol;Acc:HGNC:3013]
ENSG00000278558	ENSG00000278558	TMEM191B	transmembrane protein 191B [Source:HGNC Symbol;Acc:HGNC:33600]
ENSG00000141744	ENSG00000141744	PNMT	phenylethanolamine N-methyltransferase [Source:HGNC Symbol;Acc:HGNC:9160]
ENSG00000074803	ENSG00000074803	SLC12A1	solute carrier family 12 member 1 [Source:HGNC Symbol;Acc:HGNC:10910]
ENSG00000183654	ENSG00000183654	MARCHF11	membrane associated ring-CH-type finger 11 [Source:HGNC Symbol;Acc:HGNC:33609]

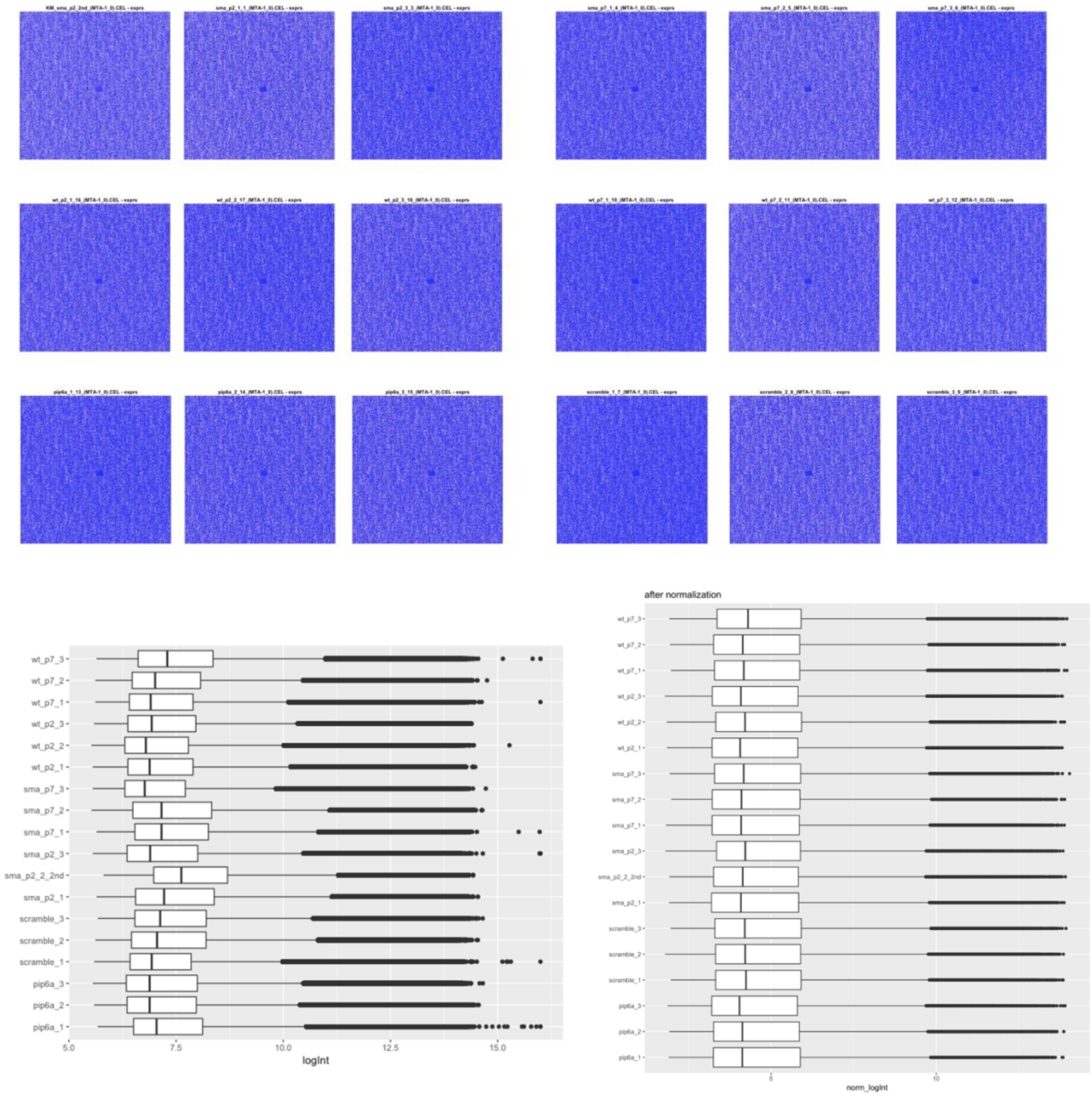
MOVO-B motif			
initial_alias	converted_alias	name	description
ENSG00000106153	ENSG00000106153	CHCHD2	coiled-coil-helix-coiled-coil-helix domain containing 2 [Source:HGNC Symbol;Acc:HGNC:21645]
ENSG00000182481	ENSG00000182481	KPNA2	karyopherin subunit alpha 2 [Source:HGNC Symbol;Acc:HGNC:6395]
ENSG00000128567	ENSG00000128567	PODXL	podocalyxin like [Source:HGNC Symbol;Acc:HGNC:9171]
ENSG00000170421	ENSG00000170421	KRT8	keratin 8 [Source:HGNC Symbol;Acc:HGNC:6446]
ENSG00000052802	ENSG00000052802	MSMO1	methylsterol monooxygenase 1 [Source:HGNC Symbol;Acc:HGNC:10545]
ENSG00000163359	ENSG00000163359	COL6A3	collagen type VI alpha 3 chain [Source:HGNC Symbol;Acc:HGNC:2213]
ENSG00000126457	ENSG00000126457	PRMT1	protein arginine methyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:5187]
ENSG00000136830	ENSG00000136830	NIBAN2	niban apoptosis regulator 2 [Source:HGNC Symbol;Acc:HGNC:25282]
ENSG00000198363	ENSG00000198363	ASPH	aspartate beta-hydroxylase [Source:HGNC Symbol;Acc:HGNC:757]
ENSG00000198467	ENSG00000198467	TPM2	tropomyosin 2 [Source:HGNC Symbol;Acc:HGNC:12011]
ENSG00000186480	ENSG00000186480	INSIG1	insulin induced gene 1 [Source:HGNC Symbol;Acc:HGNC:6083]
ENSG00000171867	ENSG00000171867	PRNP	prion protein [Source:HGNC Symbol;Acc:HGNC:9449]
ENSG00000100097	ENSG00000100097	LGALS1	galectin 1 [Source:HGNC Symbol;Acc:HGNC:6561]
ENSG00000106624	ENSG00000106624	AEBP1	AE binding protein 1 [Source:HGNC Symbol;Acc:HGNC:303]
ENSG00000116260	ENSG00000116260	QSOX1	quiescin sulfhydryl oxidase 1 [Source:HGNC Symbol;Acc:HGNC:9756]
ENSG00000110492	ENSG00000110492	MDK	midkine [Source:HGNC Symbol;Acc:HGNC:6972]
ENSG00000086475	ENSG00000086475	SEPHS1	selenophosphate synthetase 1 [Source:HGNC Symbol;Acc:HGNC:19685]
ENSG00000117394	ENSG00000117394	SLC2A1	solute carrier family 2 member 1 [Source:HGNC Symbol;Acc:HGNC:11005]
ENSG00000149428	ENSG00000149428	HYOU1	hypoxia up-regulated 1 [Source:HGNC Symbol;Acc:HGNC:16931]
ENSG00000121741	ENSG00000121741	ZMYM2	zinc finger MYM-type containing 2 [Source:HGNC Symbol;Acc:HGNC:12989]
ENSG00000182871	ENSG00000182871	COL18A1	collagen type XVIII alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:2195]
ENSG00000159217	ENSG00000159217	IGF2BP1	insulin like growth factor 2 mRNA binding protein 1 [Source:HGNC Symbol;Acc:HGNC:28866]
ENSG00000068366	ENSG00000068366	ACSL4	acyl-CoA synthetase long chain family member 4 [Source:HGNC Symbol;Acc:HGNC:3571]
ENSG00000111057	ENSG00000111057	KRT18	keratin 18 [Source:HGNC Symbol;Acc:HGNC:6430]
ENSG00000112715	ENSG00000112715	VEGFA	vascular endothelial growth factor A [Source:HGNC Symbol;Acc:HGNC:12680]
ENSG00000177606	ENSG00000177606	JUN	Jun proto-oncogene, AP-1 transcription factor subunit [Source:HGNC Symbol;Acc:HGNC:6204]
ENSG00000013810	ENSG00000013810	TACC3	transforming acidic coiled-coil containing protein 3 [Source:HGNC Symbol;Acc:HGNC:11524]
ENSG00000090372	ENSG00000090372	STRN4	striatin 4 [Source:HGNC Symbol;Acc:HGNC:15721]
ENSG00000188486	ENSG00000188486	H2AX	H2A.X variant histone [Source:HGNC Symbol;Acc:HGNC:4739]
ENSG00000162434	ENSG00000162434	JAK1	Janus kinase 1 [Source:HGNC Symbol;Acc:HGNC:6190]
ENSG00000008083	ENSG00000008083	JARID2	jumonji and AT-rich interaction domain containing 2 [Source:HGNC Symbol;Acc:HGNC:6196]
ENSG00000153904	ENSG00000153904	DDAH1	dimethylarginine dimethylaminohydrolase 1 [Source:HGNC Symbol;Acc:HGNC:2715]
ENSG00000120738	ENSG00000120738	EGR1	early growth response 1 [Source:HGNC Symbol;Acc:HGNC:3238]
ENSG00000165821	ENSG00000165821	SALL2	spalt like transcription factor 2 [Source:HGNC Symbol;Acc:HGNC:10526]
ENSG00000050405	ENSG00000050405	LIMA1	LIM domain and actin binding 1 [Source:HGNC Symbol;Acc:HGNC:24636]
ENSG00000175550	ENSG00000175550	DRAP1	DR1 associated protein 1 [Source:HGNC Symbol;Acc:HGNC:3019]
ENSG00000114270	ENSG00000114270	COL7A1	collagen type VII alpha 1 chain [Source:HGNC Symbol;Acc:HGNC:2214]
ENSG00000119772	ENSG00000119772	DNMT3A	DNA methyltransferase 3 alpha [Source:HGNC Symbol;Acc:HGNC:2978]
ENSG00000130203	ENSG00000130203	APOE	apolipoprotein E [Source:HGNC Symbol;Acc:HGNC:613]
ENSG00000106976	ENSG00000106976	DNM1	dynamin 1 [Source:HGNC Symbol;Acc:HGNC:2972]
ENSG00000134324	ENSG00000134324	LPIN1	lipin 1 [Source:HGNC Symbol;Acc:HGNC:13345]
ENSG00000108854	ENSG00000108854	SMURF2	SMAD specific E3 ubiquitin protein ligase 2 [Source:HGNC Symbol;Acc:HGNC:16809]
ENSG00000161011	ENSG00000161011	SQSTM1	sequestosome 1 [Source:HGNC Symbol;Acc:HGNC:11280]
ENSG00000168758	ENSG00000168758	SEMA4C	semaphorin 4C [Source:HGNC Symbol;Acc:HGNC:10731]
ENSG00000107331	ENSG00000107331	ABCA2	ATP binding cassette subfamily A member 2 [Source:HGNC Symbol;Acc:HGNC:32]
ENSG00000122707	ENSG00000122707	RECK	reversion inducing cysteine rich protein with kazal motifs [Source:HGNC Symbol;Acc:HGNC:11345]
ENSG00000106367	ENSG00000106367	AP1S1	adaptor related protein complex 1 subunit sigma 1 [Source:HGNC Symbol;Acc:HGNC:559]
ENSG00000235162	ENSG00000235162	C12orf75	chromosome 12 open reading frame 75 [Source:HGNC Symbol;Acc:HGNC:35164]
ENSG00000118971	ENSG00000118971	CCND2	cyclin D2 [Source:HGNC Symbol;Acc:HGNC:1583]
ENSG00000160360	ENSG00000160360	GPSM1	G protein signaling modulator 1 [Source:HGNC Symbol;Acc:HGNC:17858]
ENSG00000111252	ENSG00000111252	SH2B3	SH2B adaptor protein 3 [Source:HGNC Symbol;Acc:HGNC:29605]
ENSG00000109805	ENSG00000109805	NCAPG	non-SMC condensin I complex subunit G [Source:HGNC Symbol;Acc:HGNC:24304]
ENSG00000107281	ENSG00000107281	NPDC1	neural proliferation, differentiation and control 1 [Source:HGNC Symbol;Acc:HGNC:7899]
ENSG00000121753	ENSG00000121753	ADGRB2	adhesion G protein-coupled receptor B2 [Source:HGNC Symbol;Acc:HGNC:944]
ENSG00000136854	ENSG00000136854	STXBP1	syntaxin binding protein 1 [Source:HGNC Symbol;Acc:HGNC:11444]
ENSG00000128602	ENSG00000128602	SMO	smoothened, frizzled class receptor [Source:HGNC Symbol;Acc:HGNC:11119]
ENSG00000166825	ENSG00000166825	ANPEP	alanyl aminopeptidase, membrane [Source:HGNC Symbol;Acc:HGNC:500]
ENSG00000066322	ENSG00000066322	ELOVL1	ELOVL fatty acid elongase 1 [Source:HGNC Symbol;Acc:HGNC:14418]
ENSG00000105486	ENSG00000105486	LIG1	DNA ligase 1 [Source:HGNC Symbol;Acc:HGNC:6598]
ENSG00000065923	ENSG00000065923	SLC9A7	solute carrier family 9 member A7 [Source:HGNC Symbol;Acc:HGNC:17123]
ENSG00000168056	ENSG00000168056	LTBP3	latent transforming growth factor beta binding protein 3 [Source:HGNC Symbol;Acc:HGNC:6716]
ENSG00000006327	ENSG00000006327	TNFRSF12A	TNF receptor superfamily member 12A [Source:HGNC Symbol;Acc:HGNC:18152]
ENSG00000103888	ENSG00000103888	CEMIP	cell migration inducing hyaluronidase 1 [Source:HGNC Symbol;Acc:HGNC:29213]
ENSG00000079308	ENSG00000079308	TNS1	tensin 1 [Source:HGNC Symbol;Acc:HGNC:11973]
ENSG00000048028	ENSG00000048028	USP28	ubiquitin specific peptidase 28 [Source:HGNC Symbol;Acc:HGNC:12625]
ENSG00000169851	ENSG00000169851	PCDH7	protocadherin 7 [Source:HGNC Symbol;Acc:HGNC:8659]
ENSG00000082397	ENSG00000082397	EPB41L3	erythrocyte membrane protein band 4.1 like 3 [Source:HGNC Symbol;Acc:HGNC:3380]
ENSG00000102996	ENSG00000102996	MMP15	matrix metallopeptidase 15 [Source:HGNC Symbol;Acc:HGNC:7161]
ENSG00000170345	ENSG00000170345	FOS	Fos proto-oncogene, AP-1 transcription factor subunit [Source:HGNC Symbol;Acc:HGNC:3796]
ENSG00000088882	ENSG00000088882	CPXM1	carboxypeptidase X, M14 family member 1 [Source:HGNC Symbol;Acc:HGNC:15771]
ENSG00000179361	ENSG00000179361	ARID3B	AT-rich interaction domain 3B [Source:HGNC Symbol;Acc:HGNC:14350]
ENSG00000185567	ENSG00000185567	AHNAK2	AHNAK nucleoprotein 2 [Source:HGNC Symbol;Acc:HGNC:20125]
ENSG00000116017	ENSG00000116017	ARID3A	AT-rich interaction domain 3A [Source:HGNC Symbol;Acc:HGNC:3031]
ENSG00000174136	ENSG00000174136	RGMB	repulsive guidance molecule BMP co-receptor b [Source:HGNC Symbol;Acc:HGNC:26896]
ENSG00000136111	ENSG00000136111	TBC1D4	TBC1 domain family member 4 [Source:HGNC Symbol;Acc:HGNC:19165]
ENSG00000163531	ENSG00000163531	NFASC	neurofascin [Source:HGNC Symbol;Acc:HGNC:29866]
ENSG00000125247	ENSG00000125247	TMTC4	transmembrane O-mannosyltransferase targeting cadherins 4 [Source:HGNC Symbol;Acc:HGNC:25
ENSG00000078177	ENSG00000078177	N4BP2	NEDD4 binding protein 2 [Source:HGNC Symbol;Acc:HGNC:29851]
ENSG00000175416	ENSG00000175416	CLTB	clathrin light chain B [Source:HGNC Symbol;Acc:HGNC:2091]
ENSG00000108830	ENSG00000108830	RND2	Rho family GTPase 2 [Source:HGNC Symbol;Acc:HGNC:18315]
ENSG00000112414	ENSG00000112414	ADGRG6	adhesion G protein-coupled receptor G6 [Source:HGNC Symbol;Acc:HGNC:13841]
ENSG00000166450	ENSG00000166450	PRTG	protogenin [Source:HGNC Symbol;Acc:HGNC:26373]
ENSG00000142552	ENSG00000142552	RCN3	reticulocalbin 3 [Source:HGNC Symbol;Acc:HGNC:21145]
ENSG00000009413	ENSG00000009413	REV3L	REV3 like, DNA directed polymerase zeta catalytic subunit [Source:HGNC Symbol;Acc:HGNC:9968]
ENSG00000131981	ENSG00000131981	LGALS3	galectin 3 [Source:HGNC Symbol;Acc:HGNC:6563]
ENSG00000135631	ENSG00000135631	RAB11FIP5	RAB11 family interacting protein 5 [Source:HGNC Symbol;Acc:HGNC:24845]
ENSG00000145703	ENSG00000145703	IQGAP2	IQ motif containing GTPase activating protein 2 [Source:HGNC Symbol;Acc:HGNC:6111]
ENSG00000197457	ENSG00000197457	STMN3	stathmin 3 [Source:HGNC Symbol;Acc:HGNC:15926]
ENSG00000068650	ENSG00000068650	ATP11A	ATPase phospholipid transporting 11A [Source:HGNC Symbol;Acc:HGNC:13552]
ENSG00000173442	ENSG00000173442	EHBP1L1	EH domain binding protein 1 like 1 [Source:HGNC Symbol;Acc:HGNC:30682]
ENSG00000128512	ENSG00000128512	DOCK4	dedicator of cytokinesis 4 [Source:HGNC Symbol;Acc:HGNC:19192]
ENSG00000135525	ENSG00000135525	MAP7	microtubule associated protein 7 [Source:HGNC Symbol;Acc:HGNC:6869]
ENSG00000131378	ENSG00000131378	RFTN1	raftlin, lipid raft linker 1 [Source:HGNC Symbol;Acc:HGNC:30278]
ENSG00000221968	ENSG00000221968	FADS3	fatty acid desaturase 3 [Source:HGNC Symbol;Acc:HGNC:3576]
ENSG00000110583	ENSG00000110583	NAA40	N-alpha-acetyltransferase 40, NatD catalytic subunit [Source:HGNC Symbol;Acc:HGNC:25845]
ENSG00000138061	ENSG00000138061	CYP1B1	cytochrome P450 family 1 subfamily B member 1 [Source:HGNC Symbol;Acc:HGNC:2597]
ENSG00000198853	ENSG00000198853	RUSC2	RUN and SH3 domain containing 2 [Source:HGNC Symbol;Acc:HGNC:23625]
ENSG00000011347	ENSG00000011347	SYT7	synaptotagmin 7 [Source:HGNC Symbol;Acc:HGNC:11514]
ENSG00000154229	ENSG00000154229	PRKCA	protein kinase C alpha [Source:HGNC Symbol;Acc:HGNC:9393]
ENSG00000156017	ENSG00000156017	CARNMT1	carnosine N-methyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:23435]
ENSG00000164506	ENSG00000164506	STXBP5	syntaxin binding protein 5 [Source:HGNC Symbol;Acc:HGNC:19665]
ENSG00000136490	ENSG00000136490	LIMD2	LIM domain containing 2 [Source:HGNC Symbol;Acc:HGNC:28142]
ENSG00000164932	ENSG00000164932	CTHRC1	collagen triple helix repeat containing 1 [Source:HGNC Symbol;Acc:HGNC:18831]
ENSG00000101236	ENSG00000101236	RNF24	ring finger protein 24 [Source:HGNC Symbol;Acc:HGNC:13779]
ENSG00000165757	ENSG00000165757	JCAD	junctional cadherin 5 associated [Source:HGNC Symbol;Acc:HGNC:29283]
ENSG00000162104	ENSG00000162104	ADCY9	adenylate cyclase 9 [Source:HGNC Symbol;Acc:HGNC:240]
ENSG00000139546	ENSG00000139546	TARBP2	TARBP2 subunit of RISC loading complex [Source:HGNC Symbol;Acc:HGNC:11569]
ENSG00000152284	ENSG00000152284	TCF7L1	transcription factor 7 like 1 [Source:HGNC Symbol;Acc:HGNC:11640]
ENSG00000083223	ENSG00000083223	TUT7	terminal uridylyl transferase 7 [Source:HGNC Symbol;Acc:HGNC:25817]
ENSG00000171223	ENSG00000171223	JUNB	JunB proto-oncogene, AP-1 transcription factor subunit [Source:HGNC Symbol;Acc:HGNC:6205]
ENSG00000129009	ENSG00000129009	ISLR	immunoglobulin superfamily containing leucine rich repeat [Source:HGNC Symbol;Acc:HGNC:6133]
ENSG00000149809	ENSG00000149809	TM7SF2	transmembrane 7 superfamily member 2 [Source:HGNC Symbol;Acc:HGNC:11863]
ENSG00000158246	ENSG00000158246	TENT5B	terminal nucleotidyltransferase 5B [Source:HGNC Symbol;Acc:HGNC:28273]
ENSG00000010818	ENSG00000010818	HIVEP2	HIVEP zinc finger 2 [Source:HGNC Symbol;Acc:HGNC:4921]
ENSG00000166173	ENSG00000166173	LARP6	La ribonucleoprotein 6, translational regulator [Source:HGNC Symbol;Acc:HGNC:24012]
ENSG00000079931	ENSG00000079931	MOXD1	monooxygenase DBH like 1 [Source:HGNC Symbol;Acc:HGNC:21063]
ENSG00000169635	ENSG00000169635	HIC2	HIC ZBTB transcriptional repressor 2 [Source:HGNC Symbol;Acc:HGNC:18595]
ENSG00000148737	ENSG00000148737	TCF7L2	transcription factor 7 like 2 [Source:HGNC Symbol;Acc:HGNC:11641]
ENSG00000140451	ENSG00000140451	PIF1	PIF1 5'-to-3' DNA helicase [Source:HGNC Symbol;Acc:HGNC:26220]
ENSG00000103449	ENSG00000103449	SALL1	spalt like transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:10524]
ENSG00000105699	ENSG00000105699	LSR	lipolysis stimulated lipoprotein receptor [Source:HGNC Symbol;Acc:HGNC:29572]
ENSG00000151498	ENSG00000151498	ACAD8	acyl-CoA dehydrogenase family member 8 [Source:HGNC Symbol;Acc:HGNC:87]
ENSG00000095739	ENSG00000095739	BAMBI	BMP and activin membrane bound inhibitor [Source:HGNC Symbol;Acc:HGNC:30251]
ENSG00000008513	ENSG00000008513	ST3GAL1	ST3 beta-galactoside alpha-2,3-sialyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:10862]
ENSG00000249992	ENSG00000249992	TMEM158	transmembrane protein 158 [Source:HGNC Symbol;Acc:HGNC:30293]
ENSG00000104177	ENSG00000104177	MYEF2	myelin expression factor 2 [Source:HGNC Symbol;Acc:HGNC:17940]
ENSG00000177283	ENSG00000177283	FZD8	frizzled class receptor 8 [Source:HGNC Symbol;Acc:HGNC:4046]
ENSG00000106772	ENSG00000106772	PRUNE2	prune homolog 2 with BCH domain [Source:HGNC Symbol;Acc:HGNC:25209]
ENSG00000056998	ENSG00000056998	GYG2	glycogenin 2 [Source:HGNC Symbol;Acc:HGNC:4700]
ENSG00000174306	ENSG00000174306	ZHX3	zinc fingers and homeoboxes 3 [Source:HGNC Symbol;Acc:HGNC:15935]
ENSG00000092758	ENSG00000092758	COL9A3	collagen type IX alpha 3 chain [Source:HGNC Symbol;Acc:HGNC:2219]
ENSG00000108947	ENSG00000108947	EFNB3	ephrin B3 [Source:HGNC Symbol;Acc:HGNC:3228]
ENSG00000157570	ENSG00000157570	TSPAN18	tetraspanin 18 [Source:HGNC Symbol;Acc:HGNC:20660]
ENSG00000105971	ENSG00000105971	CAV2	caveolin 2 [Source:HGNC Symbol;Acc:HGNC:1528]
ENSG00000085433	ENSG00000085433	WDR47	WD repeat domain 47 [Source:HGNC Symbol;Acc:HGNC:29141]
ENSG00000169826	ENSG00000169826	CSGALNACT2	chondroitin sulfate N-acetylgalactosaminyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:24292]
ENSG00000187244	ENSG00000187244	BCAM	basal cell adhesion molecule (Lutheran blood group) [Source:HGNC Symbol;Acc:HGNC:6722]
ENSG00000180447	ENSG00000180447	GAS1	growth arrest specific 1 [Source:HGNC Symbol;Acc:HGNC:4165]
ENSG00000162551	ENSG00000162551	ALPL	alkaline phosphatase, biomineralization associated [Source:HGNC Symbol;Acc:HGNC:438]
ENSG00000115525	ENSG00000115525	ST3GAL5	ST3 beta-galactoside alpha-2,3-sialyltransferase 5 [Source:HGNC Symbol;Acc:HGNC:10872]
ENSG00000162545	ENSG00000162545	CAMK2N1	calcium/calmodulin dependent protein kinase II inhibitor 1 [Source:HGNC Symbol;Acc:HGNC:24190
ENSG00000134138	ENSG00000134138	MEIS2	Meis homeobox 2 [Source:HGNC Symbol;Acc:HGNC:7001]
ENSG00000198598	ENSG00000198598	MMP17	matrix metallopeptidase 17 [Source:HGNC Symbol;Acc:HGNC:7163]
ENSG00000130150	ENSG00000130150	MOSPD2	motile sperm domain containing 2 [Source:HGNC Symbol;Acc:HGNC:28381]
ENSG00000082014	ENSG00000082014	SMARCD3	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, membe
ENSG00000165895	ENSG00000165895	ARHGAP42	Rho GTPase activating protein 42 [Source:HGNC Symbol;Acc:HGNC:26545]
ENSG00000168209	ENSG00000168209	DDIT4	DNA damage inducible transcript 4 [Source:HGNC Symbol;Acc:HGNC:24944]
ENSG00000176834	ENSG00000176834	VSIG10	V-set and immunoglobulin domain containing 10 [Source:HGNC Symbol;Acc:HGNC:26078]
ENSG00000133131	ENSG00000133131	MORC4	MORC family CW-type zinc finger 4 [Source:HGNC Symbol;Acc:HGNC:23485]



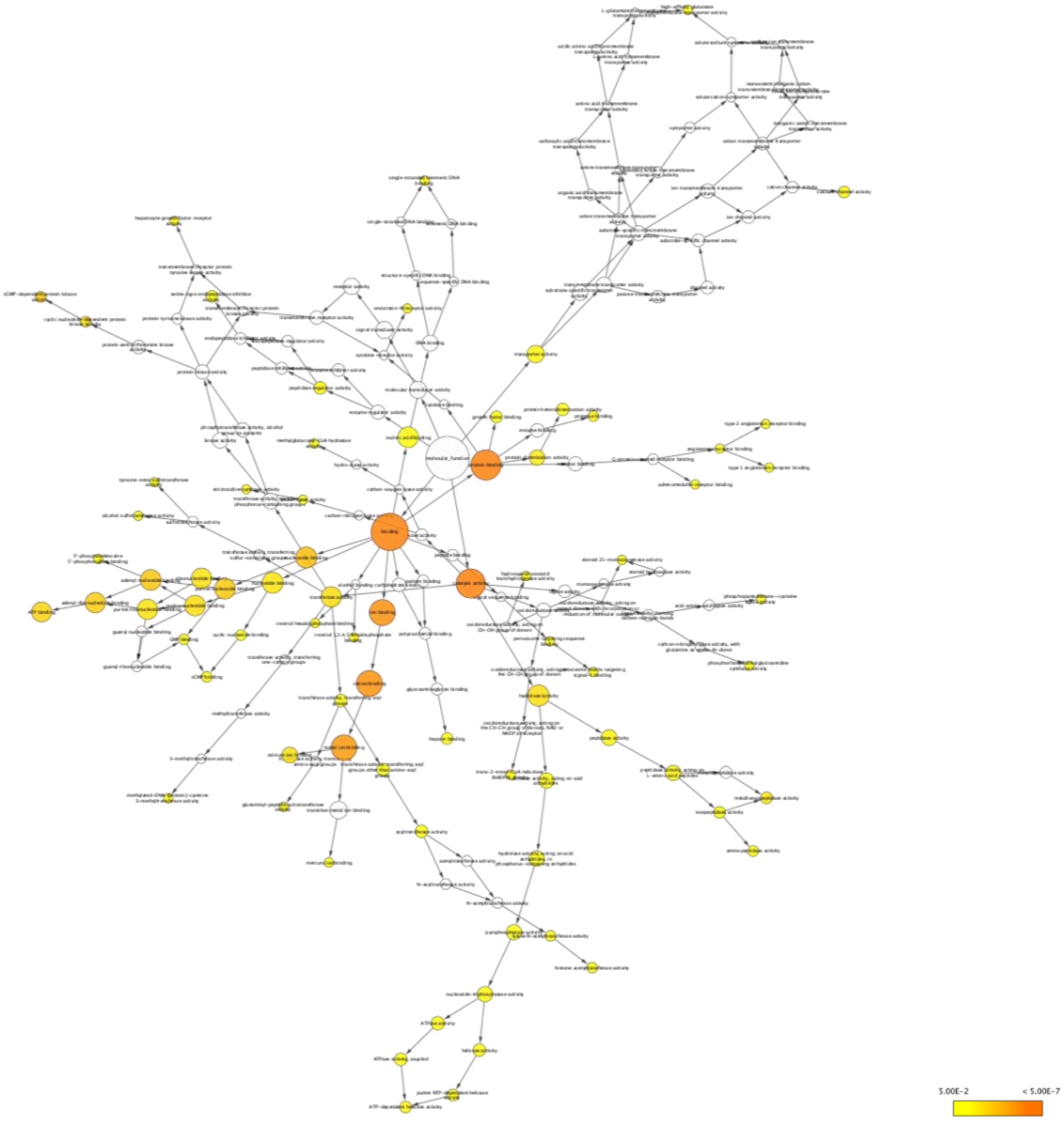
Supplementary Figure 7-2: expression levels of three X-linked markers: Xist (blue), Ddx3y (orange) and Ddx3x (grey)

Supplementary Table 7-5: Sex distribution within the groups

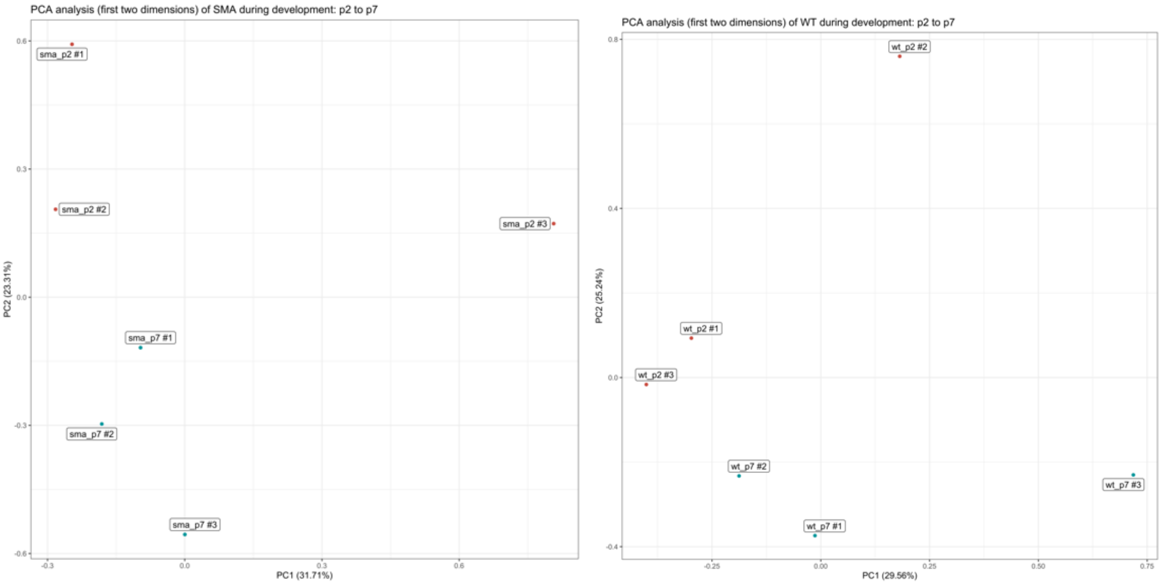
- Males (XY)	- (Females XX)
- Sma_p2_1, sma_p2_2, sma_p2_3	-
- Sma_p7_1	- Sma_p7_2, Sma_p7_3
- Scramble_1, scramble_3	- Scramble_2
- WT_p7_1, WT_p7_3	- Wt_p7_2
- Pip6a_1, Pip6a_2, Pip6a_3	-
- wt_p2_3	- Wt_p2_1, wt_p2_2



Supplementary Figure 7-3: quality control of the reads, as assessed by chip pseudoarrays of the exon microarrays from LCM-MN samples (top); and by the boxplot of the log(probe Intensity) before and after RMA normalisation.



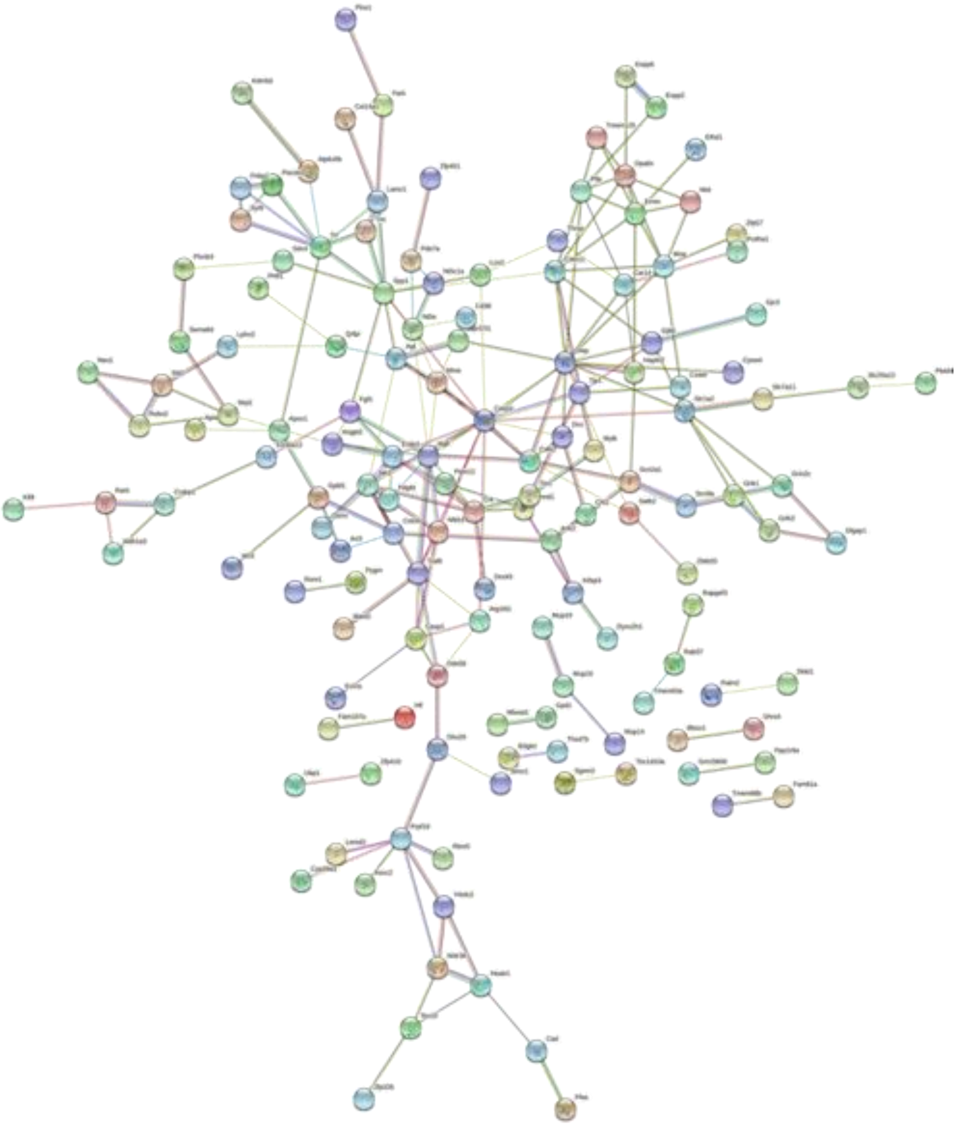
Supplementary Figure 7-4 : GO-term enrichment analysis of the DE genes at SMA p7 (BiNGO plug-in in Cytoscape)



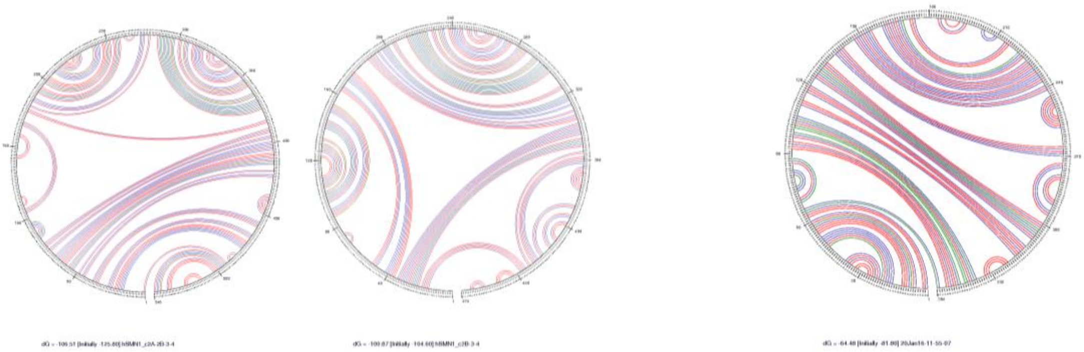
Supplementary Figure 7-5: Principal Component Analysis of SMA (left) and WT (right) animals from p2 to p7

Supplementary Table 7-6: Genes differentially (π-value >1.3) expressed throughout the P2 to P7 development of SMA and WT mice

SMA p2-p7 specific genes							**Intercept genes**	**WT p2-p7 specific genes**
Mog	Gpd1	Dync2h1	Mup13	Gm5062	Crybb1	4932438A13Rik	Plp1	Pla2g4a
Rprml	1700040L02Rik	Cxadr	Mup14	2310001H17Rik	Nrp2	Hsd17b3	Kcng4	Atp8b3
Crabp2	Cpne4	Stk36	Mup19	Sema6d	Rnf152	Ascc2	Met	Mpdz
Gjc3	Trf	Wdr36	Mup15	Phtf1	Tbc1d10a	Ubp1	Mobp	Myl1
Cyp39a1	Thsd7b	Dkkl1	Tmem63a	Nos1ap	Lphn2	Dcbld2	Mag	Syt5
Sult1a1	Pdgfd	Evi2a	Rom1	Zbtb20	Wfdc2	Plx3	Mbp	Htr2c
Satb2	Lemd3	Efhd1	Crk	2410089E03Rik	Scn2a1	Kifap3	Piezo2	Stt3a
Car14	Ank2	Eif2ak4	Angpt2	Lamc1	Ctage5		A2m	Abcg2
Enpp6	Ralgapa1	Gm2260	2810468N07Rik	Pde7a	Exoc6b		Plxdc1	Ogn
Plce1	Fam107a	Gm2274	Nt5c1a	Dh2	Cad		Cachd1	Tesk2
9630013A20Rik	Rarb	E130309F12Rik	Mast2	Cntn5	Folh1		Ugt8a	Atp10b
Lcn2	Serpi3c	Scn9a	Smim3	Ythdc2	F630042J09Rik		Hsd11b1	Vim
Apod	Slc6a6	Klf9	Slit2	Paf1	Chpt1		Tmem106a	Rbm25
Cyp2d22	Ctnnd1	Zfp326	Magi1	Swap70	Cep170		Fa2h	Usp48
Itih3	Ctns	Tjp1	Wiz	St8sia6	Adam15		Elmo1	Ptgds
C130080G10Rik	Pgap1	1110054M08Rik	Fam81a	Pcp4l1	Plac8			Pnlip
Cldn11	Plxnb3	Ice2	Npas1	Fnbp1	Palm2			Obscn
Tnc	B230206L02Rik	Idi1	Mboat1	2310057J18Rik	Heatr1			Sla
Trpc7	4930506M07Rik	Bbox1	Tex10	Zfp248	Wsb1			Smyd4
Bnc2	Mme	Arhgap9	Adck3	Acadsb	Mylk			Scn7a
Slc1a2	Ppp1r9a	Plekhb1	Ngfr	Chl1	Zfp410			Stag1
Kdm5d	Htra1	Ttc37	Ptpn12	Pygm	Hlf			Ppm1j
Kcnj14	Cd38	Gm13629	Tsc22d4	Sept6	Dhx29			Herc6
Rab37	Pacsin2	Pllp	Dlgap1	Amhr2	Dcx			Rlbp1
Vps13a	Abr	Prpf19	Pfas	Gpld1	Tes			Fam185a
Erbb3	Tln1	Cac2d1	Simc1	Grik3	Atg16l1			Scn8a
Atp6v0b	Lrp1b	Stc1	Dock5	Syt9	Gm17081			Srek1
Cyp2d13	Tdrd5	Mup-ps21	Fgf1	Hn1l	Nfrkb			Skiv2l2
S100a13	Mal	v2	Fam160b1	Gpr37l1	Dhrs4			
Thrsp	Traf6	Sdc4	Creg2	Ermn	Sike1			
Stag2	Neo1	Sgsm2	Mt2	Ccdc28a	Slc25a13			
Phyhd1	Osmr	Slc7a11	Nfkb1	v1	Apoc1			
Agt	Rbm5	Zfp451	Gm13301	Casp1	Calb2			
Kcnh8	Med21	Mup7	Nt5e	Jak2	Pcdhac2			
Ebf3	Coch	Mup2	Art3	Ddx58	B3glct			
Atp13a4	Zfp57	Mup9	Tmem88b	Kazald1	Gfap			
Qrfpr	Spp1	Mup8	5330417C22Rik	Cyp2d40	Robo2			
Gjb6	Phtf2	Mup10	Casp3	Rapgef3	Enpp2			
Gm13446	Cnpy3	Mup12	Col14a1	Tmem125	Gm17094			
Opalin	Grik1	Gm2083	Fat4	B930095G15Rik	Pdp1			
Grin2c	Gm15800	Mup1	Asxl3	Arrdc3	Ogdhl			



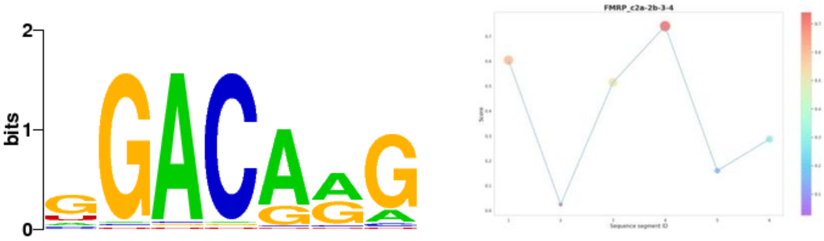
Supplementary Figure 7-6: STRING analyses of the networks of genes DE during the p2-p7 symptom progression in SMA



Supplementary Figure 7-7: mFold predictions of secondary structures of C2A-2B-3-4 and of C3-4

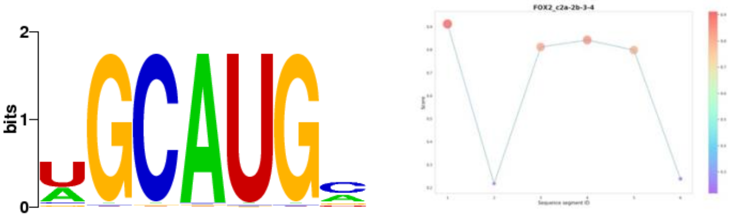
Supplementary Table 7-7: list of RBP binding sites to four SMN-derived circRNAs, as identified with RBPsuite

Name	C2A-2B-3-4	C2A-2B-3-4	C2A-2B-3-4	C2B-3-4	C2B-3-4	C2B-3-4	C3-4	C3-4	C3-4	C6-7-8A	C6-7-8A	C6-7-8A
description	# sites > 0.5	# sites verified	# verified > 0.8	# sites > 0.5	# sites verified	# verified > 0.8	# sites > 0.5	# sites verified	# verified > 0.8	# sites > 0.5	# sites verified	# verified > 0.8
AGO1	1	0	-	1	0	-	2	0	-	0	-	-
AGO2	5	0	-	3	0	-	3	0	-	2	0	-
AGO3	3	0	-	1	0	-	3	0	-	1	0	-
ALKBH5	3	0	-	1	0	-	2	0	-	2	0	-
AUF1	0	-	-	1	0	-	0	-	-	1	0	-
C17ORF85	2	0	-	3	0	-	3	0	-	2	0	-
C22ORF28	2	0	-	2	0	-	3	0	-	1	0	-
CAPRIN1	3	0	-	4	0	-	3	0	-	2	0	-
DGCR8	0	-	-	2	0	-	0	-	-	0	-	-
EIF4A3	4	0	-	4	0	-	4	0	-	2	0	-
EWSR1	2	0	-	1	0	-	1	0	-	2	0	-
FMRP	3	1	0	1	1	1	2	0	-	1	1	1
FOX2	4	4	3	4	3	2	3	2	2	3	1	1
FUS	1	0	-	1	1	1	2	0	-	1	1	0
FXR1	1	0	-	1	1	1	1	1	1	1	1	1
FXR2	1	1	1	2	1	0	3	1	0	1	1	0
HNRNPC	0	-	-	0	-	-	0	-	-	0	-	-
HUR	0	-	-	2	1	1	2	1	0	2	2	0
IGF2BP1	2	2	0	3	3	0	3	1	0	1	0	-
IGF2BP2	3	2	1	3	1	0	3	2	1	1	1	1
IGF2BP3	3	1	0	4	2	1	3	1	1	1	1	1
LIN28A	2	1	0	4	3	2	3	3	1	3	1	0
LIN28B	2	1	0	2	1	0	2	2	0	2	1	0
METTL3	1	0	-	2	0	-	2	0	-	1	0	-
MOV10	1	0	-	1	0	-	1	0	-	1	0	-
PTB	2	1	0	2	2	1	2	2	0	2	2	0
PUM2	0	-	-	1	0	-	0	-	-	0	-	-
QKI	0	-	-	1	1	0	0	-	-	0	-	-
SFRS1	1	0	-	3	0	-	1	0	-	2	0	-
TAF15	0	-	-	1	1	1	1	0	-	1	1	0
TDP43	0	-	-	0	-	-	0	-	-	1	0	-
TIA1	1	0	-	0	-	-	0	-	-	0	-	-
TIAL1	1	0	-	0	-	-	0	-	-	0	-	-
TNRC6	5	0	-	1	0	-	3	0	-	2	0	-
U2AF65	1	0	-	0	-	-	0	-	-	0	-	-
WTAP	1	0	-	1	0	-	0	-	-	2	0	-
ZC3H7B	3	0	-	4	0	-	3	0	-	2	0	-



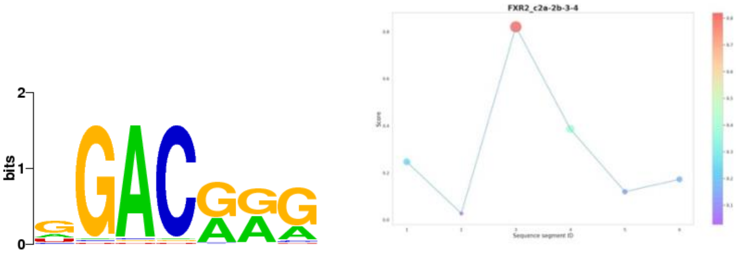
Sequence name	Sequence	Score
c2a-2b-3-4_4	GGAUAUGGAAAUAGAGAGGAGCAAAAUCUGUCCGAUCUACUUUCCCCAAUCUGUGAAGUAGCUAAUAAUAUAGAACAAAAUGCUCAAGAGAAUGAAAAUGA	0.74115115
c2a-2b-3-4_1	AGCGAUGAUUCUGACAUUUGGGAUGAUACAGCACUGAUAAAAGCAUAUGAUAAAGCUGUGGCUUCAUUUAAGCAUGCUCUAAAGAAUGGUGACAUUUGUGA	0.6041987
c2a-2b-3-4_3	GGGACAAAUGUUCUGCCAUUUGGUCAGAAGACGGUUGCAUUUACCCAGCUACCAUUGCUUCAAUUGAUUUUAAGAGAGAAACCUGUGUUGUGGUUUACACU	0.514854

Supplementary Figure 7-8: FMRP binding sites to C2A-2B-3-4



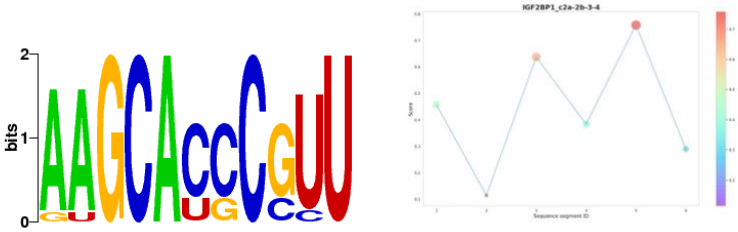
Sequence name	Sequence	Score
c2a-2b-3-4_1	AGCGAUGAUUCUGACAUUUGGGAUGAUACAGCACUGAUAAAAGCAUAUGAUAAAGCUGUGGCUUCAUUUAAGCAUGCUCUAAAGAAUGGUGACAUUUGUGA	0.912342
c2a-2b-3-4_4	GGAUAUGGAAAUAGAGAGGAGCAAAAUCUGUCCGAUCUACUUUCCCCAAUCUGUGAAGUAGCUAAUAAUAUAGAACAAAAUGCUCAAGAGAAUGAAAAUGA	0.84265256
c2a-2b-3-4_3	GGGACAAAUGUUCUGCCAUUUGGUCAGAAGACGGUUGCAUUUACCCAGCUACCAUUGCUUCAAUUGAUUUUAAGAGAGAAACCUGUGUUGUGGUUUACACU	0.81225586
c2a-2b-3-4_5	AAGCCAAGUUUCAACAGAUGAAAGUGAGAACUCCAGGUCUCCUGGAAAUAAAUCAGAUAACAUCAAGCCCAAAUCUGCUCCAUGGAACUCUUUUCUCCCUC	0.79897183

Supplementary Figure 7-9: FOX2 binding sites to C2A-2B-3-4



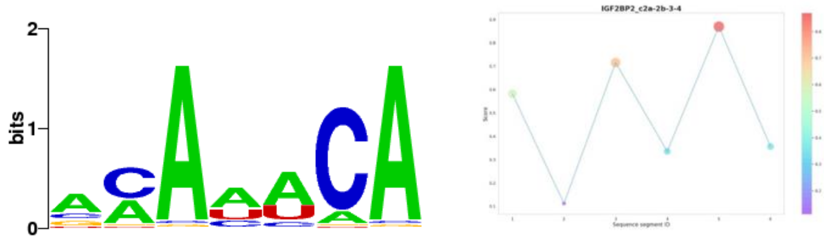
Sequence name	Sequence	Score
c2a-2b-3-4_3	GGGACAAAUGUUCUGCCAUUUGGUCAGAAGACGGUUGCAUUUACCCAGCUACCAUUGCUUCAAUUGAUUUUAAGAGAGAAACCUGUGUUGUGGUUUACACU	0.8207747

Supplementary Figure 7-10 : FXR2 binding sites to C2A-2B-3-4



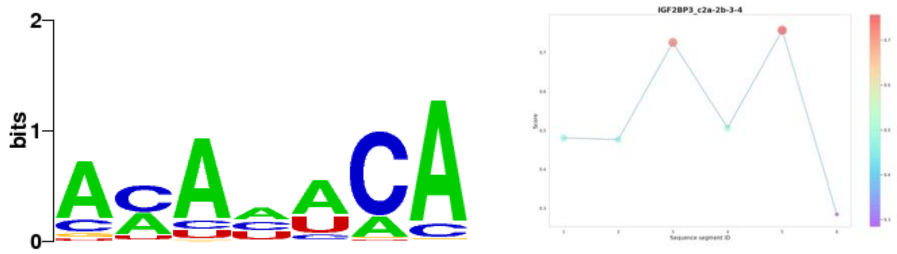
Sequence name	Sequence	Score
c2a-2b-3-4_5	AAGCCAAGUUUCAACAGAUGAAAGUGAGAACUCCAGGUCUCCUGGAAAUAAAUCAGAUAACAUCAAGCCCAAAUCUGCUCCAUGGAACUCUUUUCUCCCUC	0.75783306
c2a-2b-3-4_3	GGGACAAAUGUUCUGCCAUUUGGUCAGAAGACGGUUGCAUUUACCCAGCUACCAUUGCUUCAAUUGAUUUUAAGAGAGAAACCUGUGUUGUGGUUUACACU	0.63756895

Supplementary Figure 7-11: IGF2BP1 binding sites to C2A-2B-3-4



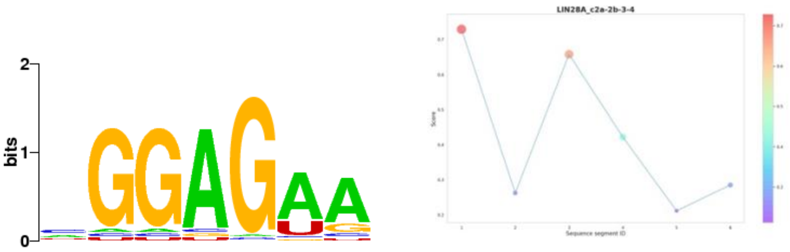
Sequence name	Sequence	Score
c2a-2b-3-4_5	AAGCCAAGUUUCAACAGAUGAAAGUGAGAACUCCAGGUCUCCUGGAAAUAAAUCAGAUAACAUCAAGCCCAAAUCUGCUCCAUGGAACUCUUUUCUCCCUC	0.87006724
c2a-2b-3-4_3	GGGACAAAUGUUCUGCCAUUUGGUCAGAAGACGGUUGCAUUUACCCAGCUACCAUUGCUUCAAUUGAUUUUAAGAGAGAAACCUGUGUUGUGGUUUACACU	0.7158649
c2a-2b-3-4_1	AGCGAUGAUUCUGACAUUUGGGAUGAUACAGCACUGAUAAAAGCAUAUGAUAAAGCUGUGGCUUCAUUUAAGCAUGCUCUAAAGAAUGGUGACAUUUGUGA	0.58252096

Supplementary Figure 7-12: IGF2BP2 binding sites to C2A-2B-3-4



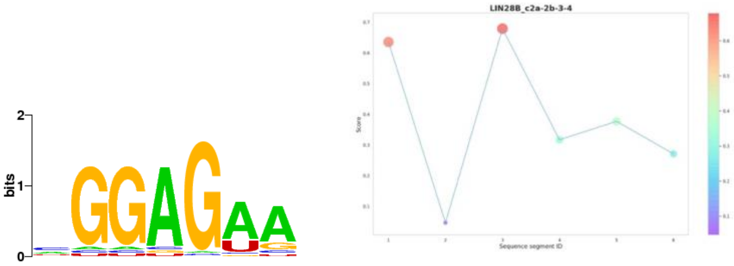
Sequence name	Sequence	Score
c2a-2b-3-4_5	AAGCCAAGUUUCAACAGAUGAAAGUGAGAACUCCAGGUCUCCUGGAAAUAAAUCAGAUAACAUCAAGCCCAAAUCUGCUCCAUGGAACUCUUUUCUCCCUC	0.7567844
c2a-2b-3-4_3	GGGACAAAUGUUCUGCCAUUUGGUCAGAAGACGGUUGCAUUUACCCAGCUACCAUUGCUUCAAUUGAUUUUAAGAGAGAAACCUGUGUUGUGGUUUACACU	0.7256034
c2a-2b-3-4_4	GGAUAUGGAAAUAGAGAGGAGCAAAAUCUGUCCGAUCUACUUUCCCCAAUCUGUGAAGUAGCUAAUAAUAUAGAACAAAAUGCUCAAGAGAAUGAAAAUGA	0.5065128

Supplementary Figure 7-13: IGF2BP3 binding sites to C2A-2B-3-4



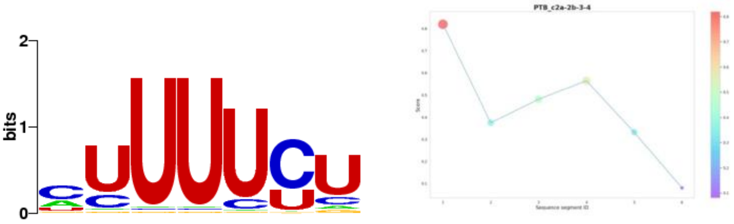
Sequence name	Sequence	Score
c2a-2b-3-4_1	AGCGAUGAUUCUGACAUUUGGGAUGAUACAGCACUGAUAAAAGCAUAUGAUAAAGCUGUGGCUUCAUUUAAGCAUGCUCUAAAGAAUGGUGACAUUUGUGA	0.72908294
c2a-2b-3-4_3	GGGACAAAUGUUCUGCCAUUUGGUCAGAAGACGGUUGCAUUUACCCAGCUACCAUUGCUUCAAUUGAUUUUAAGAGAGAAACCUGUGUUGUGGUUUACACU	0.657796

Supplementary Figure 7-14: LIN28A binding sites to C2A-2B-3-4



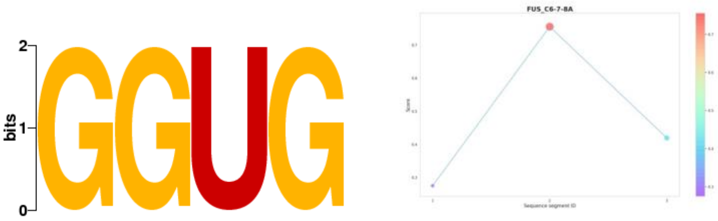
Sequence name	Sequence	Score
c2a-2b-3-4_3	GGGACAAAUGUUCUGCCAUUUGGUCAGAAGACGGUUGCAUUUACCCAGCUACCAUUGCUUCAAUUGAUUUUAAGAGAGAAACCUGUGUUGUGGUUUACACU	0.67888254
c2a-2b-3-4_1	AGCGAUGAUUCUGACAUUUGGGAUGAUACAGCACUGAUAAAAGCAUAUGAUAAAGCUGUGGCUUCAUUUAAGCAUGCUCUAAAGAAUGGUGACAUUUGUGA	0.63536114

Supplementary Figure 7-15: LIN28B binding sites to C2A-2B-3-4



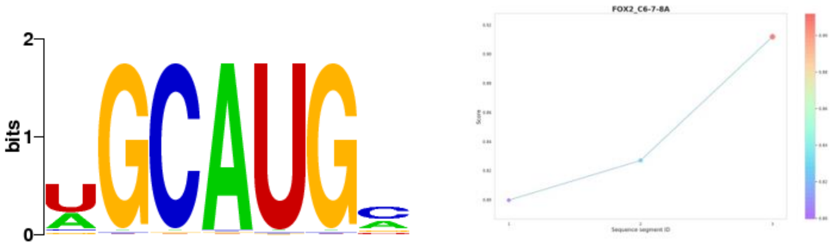
Sequence name	Sequence	Score
c2a-2b-3-4_1	AGCGAUGAUUCUGACAUUUGGGAUGAUACAGCACUGAUAAAAGCAUAUGAUAAAGCUGUGGCUUCAUUUAAGCAUGCUCUAAAGAAUGGUGACAUUUGUGA	0.81997895
c2a-2b-3-4_4	GGAUAUGGAAAUAGAGAGGAGCAAAAUCUGUCCGAUCUACUUUCCCCAAUCUGUGAAGUAGCUAAUAAUAUAGAACAAAAUGCUCAAGAGAAUGAAAAUGA	0.56508404

Supplementary Figure 7-16: PTB binding sites to C2A-2B-3-4



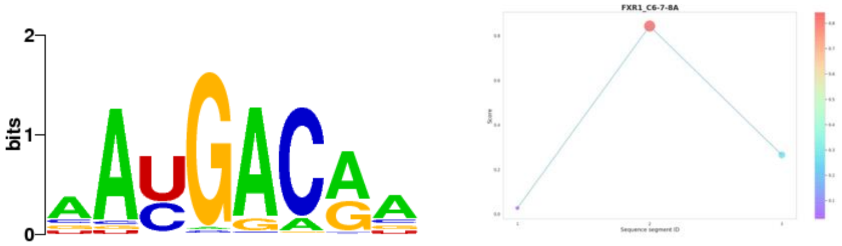
Sequence name	Sequence	Score
C6-7-8A_2	CUAUUAUAUGGGUUUUAGACAAAAUCAAAAAGAAGGAAGGUGCUCACAUUCCUUAAAUUAAGGAGAAAUGCUGGCAUAGAGCAGCACUAAAUGACACCACU	0.7555943

Supplementary Figure 7-17: FUS binding sites to C6-7-8A



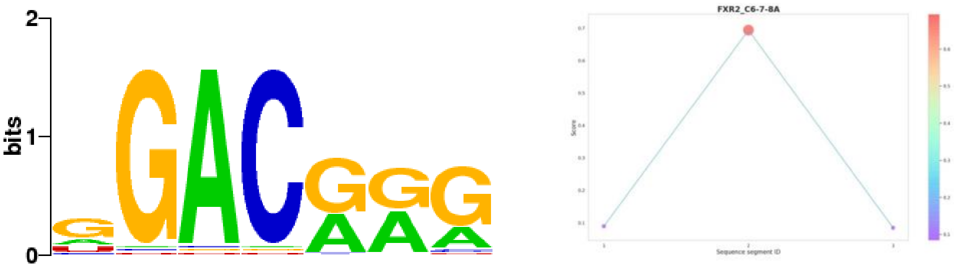
Sequence name	Sequence	Score
C6-7-8A_3	AAAGAAACGAUCAGACAGAUCUGGAAUGUGAAGCGUUAUAGAAGAUAACUGGCCUCAUUUCUUCAAAAUAUCAAGUGUUGGGAAAGAAAAAAGGAAGUGGA	0.9119439
C6-7-8A_2	CUAUUAUAUGGGUUUUAGACAAAAUCAAAAAGAAGGAAGGUGCUCACAUUCCUUAAAUUAAGGAGAAAUGCUGGCAUAGAGCAGCACUAAAUGACACCACU	0.8270332
C6-7-8A_1	AUAAUUCCCCCACCACCUCCCAUAUGUCCAGAUUCUCUUGAUGAUGCUGAUGCUUUGGGAAGUAUGUUAAUUUCAUGGUACAUGAGUGGCUAUCAUACUGG	0.79971385

Supplementary Figure 7-18: FOX2 binding sites to C6-7-8A



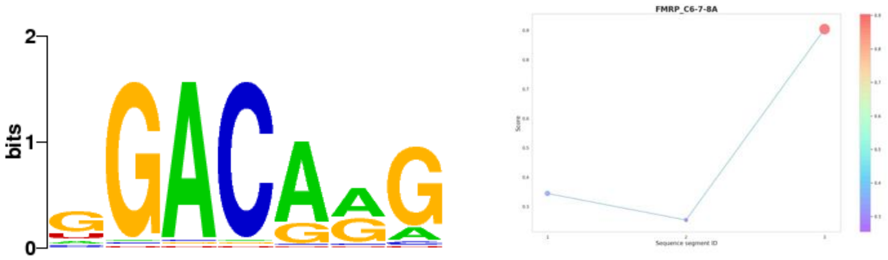
Sequence name	Sequence	Score
C6-7-8A_2	CUAUUAUAUGGGUUUUAGACAAAAUCAAAAAGAAGGAAGGUGCUCACAUUCCUUAAAUUAAGGAGAAAUGCUGGCAUAGAGCAGCACUAAAUGACACCACU	0.84378

Supplementary Figure 7-19: FXR1 binding sites to C6-7-8A



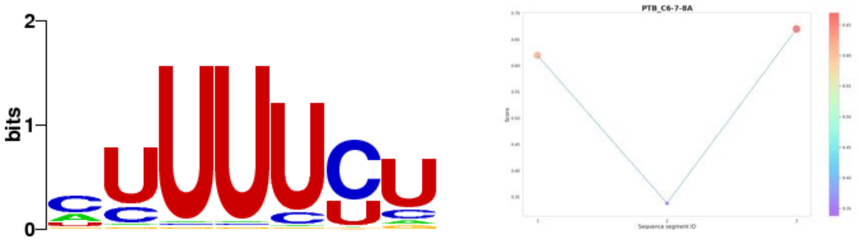
Sequence name	Sequence	Score
C6-7-8A_2	CUAUUAUAUGGGUUUUAGACAAAAUCAAAAAGAAGGAAGGUGCUCACAUUCCUUAAAUUAAGGAGAAAUGCUGGCAUAGAGCAGCACUAAAUGACACCACU	0.69398874

Supplementary Figure 7-20: FXR2 binding sites to C6-7-8A



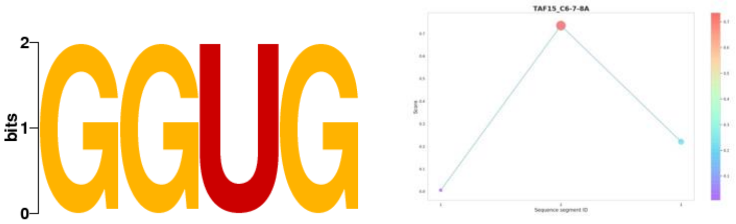
Sequence name	Sequence	Score
C6-7-8A_3	AAAGAAACGAUCAGACAGAUCUGGAAUGUGAAGCGUUAUAGAAGAUAACUGGCCUCAUUUCUUCAAAAUAUCAAGUGUUGGGAAAGAAAAAAGGAAGUGGA	0.90439624

Supplementary Figure 7-21: FMRP binding sites to C6-7-8A



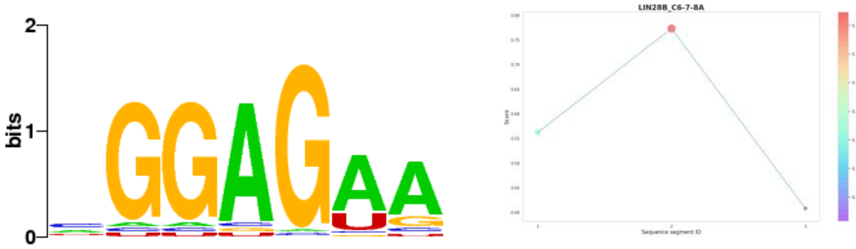
Sequence name	Sequence	Score
C6-7-8A_3	AAAGAAACGAUCAGACAGAUCUGGAAUGUGAAGCGUUAUAGAAGAUAACUGGCCUCAUUUCUUCAAAAUAUCAAGUGUUGGGAAAGAAAAAAGGAAGUGGA	0.6700308
C6-7-8A_1	AUAAUUCCCCCACCACCUCCCAUAUGUCCAGAUUCUCUUGAUGAUGCUGAUGCUUUGGGAAGUAUGUUAAUUUCAUGGUACAUGAGUGGCUAUCAUACUGG	0.61968285

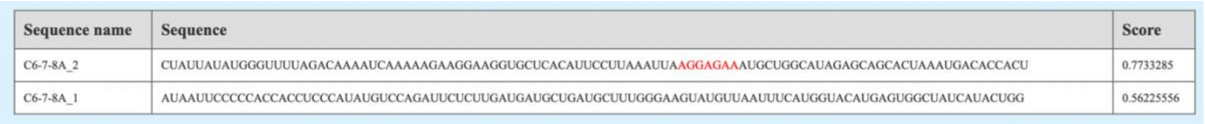
Supplementary Figure 7-22: PTB binding sites to C6-7-8A



Sequence name	Sequence	Score
C6-7-8A_2	CUAUUAUAUGGGUUUUAGACAAAAUCAAAAAGAAGGAAGGUGCUCACAUUCCUUAAAUUAAGGAGAAAUGCUGGCAUAGAGCAGCACUAAAUGACACCACU	0.7355249

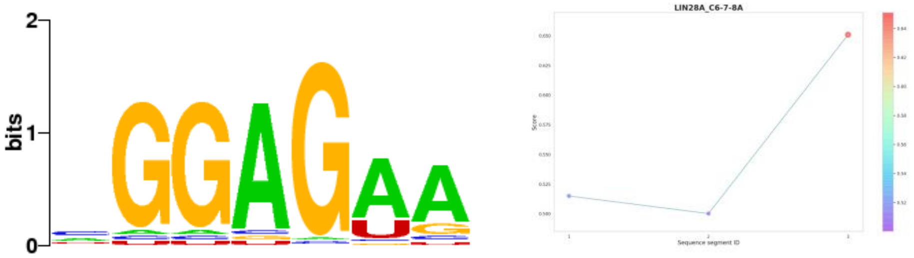
Supplementary Figure 7-23: TAF15 binding sites to C6-7-8A





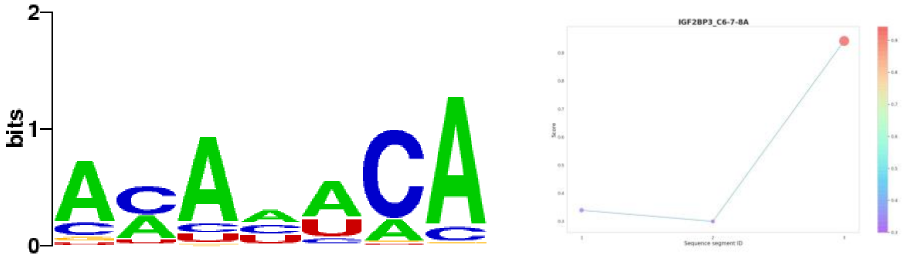
Sequence name	Sequence	Score
C6-7-8A_2	CUAUUAUAUGGGUUUUAGACAAAAUCAAAAAGAAGGAAGGUGCUCACAUUCCUUAAAUUAAGGAGAAAUGCUGGCAUAGAGCAGCACUAAAUGACACCACU	0.7733285
C6-7-8A_1	AUAAUUCCCCCACCACCUCCCAUAUGUCCAGAUUCUCUUGAUGAUGCUGAUGCUUUGGGAAGUAUGUUAAUUUCAUGGUACAUGAGUGGCUAUCAUACUGG	0.56225556

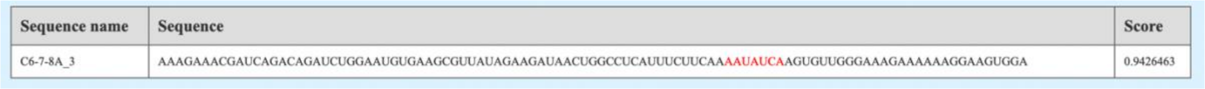
Supplementary Figure 7-24: LIN28B binding sites to C6-7-8A



Sequence name	Sequence	Score
C6-7-8A_3	AAAGAAACGAUCAGACAGAUCUGGAAUGUGAAGCGUUAUAGAAGAUAACUGGCCUCAUUUCUUCAAAAUAUCAAGUGUUGGGAAAGAAAAAAGGAAGUGGA	0.65092516
C6-7-8A_1	AUAAUUCCCCCACCACCUCCCAUAUGUCCAGAUUCUCUUGAUGAUGCUGAUGCUUUGGGAAGUAUGUUAAUUUCAUGGUACAUGAGUGGCUAUCAUACUGG	0.51509947
C6-7-8A_2	CUAUUAUAUGGGUUUUAGACAAAAUCAAAAAGAAGGAAGGUGCUCACAUUCCUUAAAUUAAGGAGAAAUGCUGGCAUAGAGCAGCACUAAAUGACACCACU	0.500279

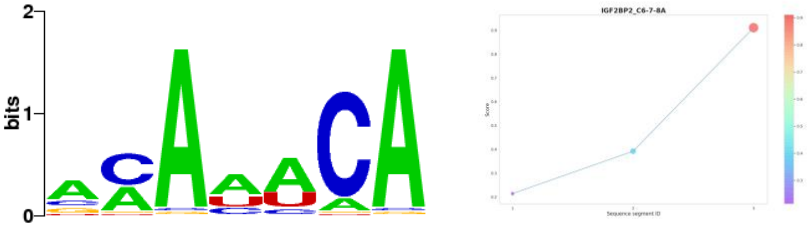
Supplementary Figure 7-25: LIN28A binding sites to C6-7-8A

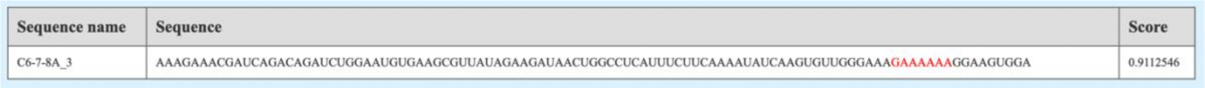




Sequence name	Sequence	Score
C6-7-8A_3	AAAGAAACGAUCAGACAGAUCUGGAAUGUGAAGCGUUAUAGAAGAUAACUGGCCUCAUUUCUUCAAAAUAUCAAGUGUUGGGAAAGAAAAAAGGAAGUGGA	0.9426463

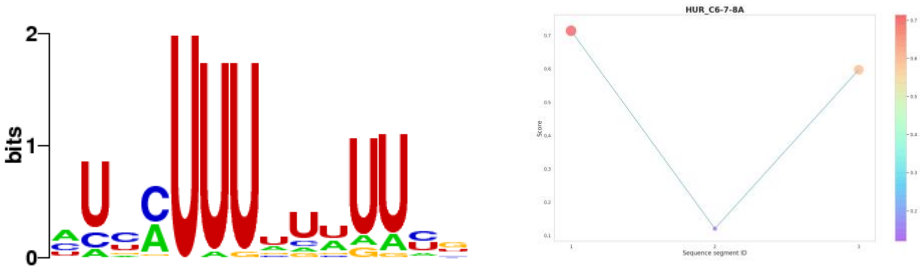
Supplementary Figure 7-26: IGF2BP3 binding sites to C6-7-8A

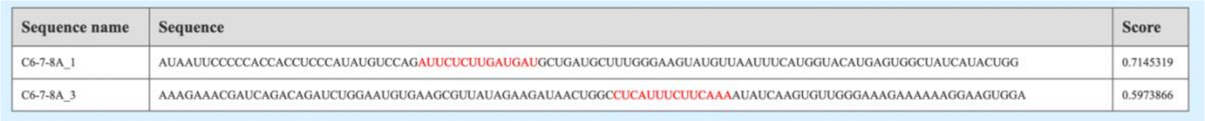




Sequence name	Sequence	Score
C6-7-8A_3	AAAGAAACGAUCAGACAGAUCUGGAAUGUGAAGCGUUAUAGAAGAUAACUGGCCUCAUUUCUUCAAAAUAUCAAGUGUUGGGAAAGAAAAAAGGAAGUGGA	0.9112546

Supplementary Figure 7-27: IGF2BP2 binding sites to C6-7-8A



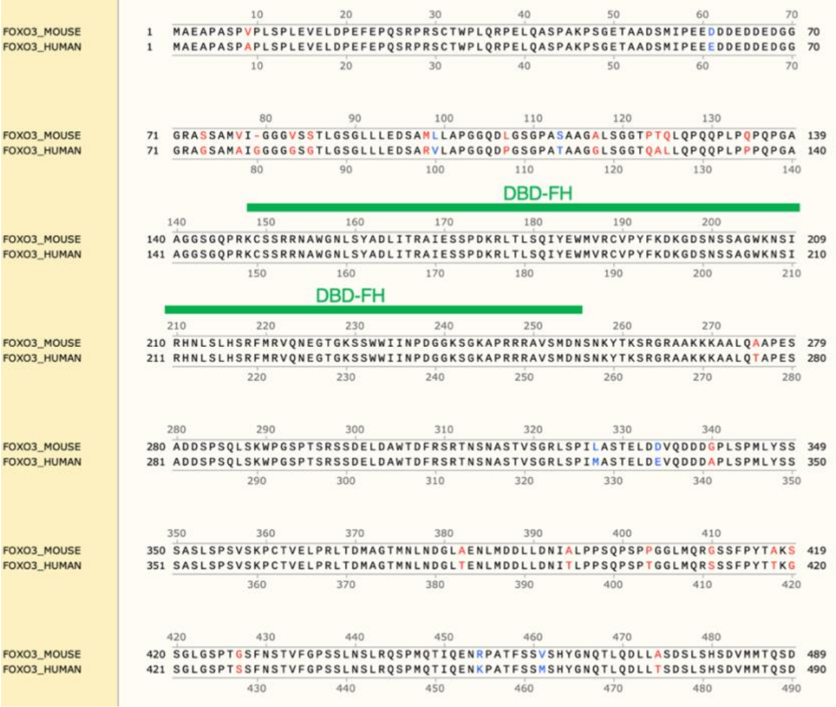


Sequence name	Sequence	Score
C6-7-8A_1	AUAAUUCCCCCACCACCUCCCAUAUGUCCAGAUUCUCUUGAUGAUGCUGAUGCUUUGGGAAGUAUGUUAAUUUCAUGGUACAUGAGUGGCUAUCAUACUGG	0.7145319
C6-7-8A_3	AAAGAAACGAUCAGACAGAUCUGGAAUGUGAAGCGUUAUAGAAGAUAACUGGCCUCAUUUCUUCAAAAUAUCAAGUGUUGGGAAAGAAAAAAGGAAGUGGA	0.5973866

Supplementary Figure 7-28: HUR binding sites to C6-7-8A

Supplementary Table 7-8: yields of the conjugation from Pip8b2 to PMO(A-E) moieties

PPMO	MW	Mol Abs	yield (nmol)	yield (mg)
Pip8b2-Foxo3_A	10697.93	240080	83.9	0.9
Pip8b2-Foxo3_B	10737.96	241280	101.7	1.1
Pip8b2-Foxo3_C	10806.03	247440	49.4	0.5
Pip8b2-Foxo3_D	10958.16	261180	62.4	0.7
Pip8b2-Foxo3_E	9798.2	224060	95.0	0.9



Supplementary Figure 7-29: protein sequences of human and murine FoxO3 are highly homologous (Black: identical aa / Blue: similar aa / Red: not similar aa)

References

1. Werdnig, G. Zwei frühinfantile hereditäre Fälle von progressiver Muskelatrophie unter dem Bilde der Dystrophie, aber auf neurotischer Grundlage. *Arch. fur Psychiatr. und Nervenkrankheiten* **22**, 437–481 (1891).
2. Hoffman, J. Über chronische spinale Muskelatrophie im Kindesalter auf familiärer Basis. *Dtsch. Z. Nervenheilkd.* **3**, 427–470 (1893).
3. Hoffman, J. Weiterer Beiträge zur Lehre von der hereditären progressiven spinalen Muskelatrophie im Kindesalter. *Dtsch. Z. Nervenheilkd.* **10**, 292–320 (1900).
4. Thomson, J. & Bruce, A. Progressive muscular atrophy in a child with a spinal lesion. *Edinburgh Hosp. Reports* **1**, 372 (1893).
5. Beevor, C. E. A case of congenital spinal muscular atrophy (family type) and a case of hemorrhage into the spinal cord at birth, giving similar symptoms. *Brain* **25**, 85–108 (1902).
6. Hutchinson, R. *Lectures on diseases of children.* (1910).
7. Brandt, S. Werdnig–Hoffmann's infantile progressive muscular atrophy. in *Opera ex domo biologiae hereditariae humanae universitatis hafniensis* (ed. Munksgaard, E.) vol. 22 (1950).
8. Wohlfart, G., Fex, J. & Eliasson, S. Hereditary proximal spinal muscular atrophy – a clinical entity simulating progressive muscular dystrophy. *Acta Psychiatr Neurol* **30**, 385–406 (1955).
9. Kugelberg, E. & Welander, L. Heredofamilial juvenile muscular atrophy simulating muscular dystrophy. *Arch. Neurol. Psychiatry* **75**, (1956).
10. Dubowitz, V. Ramblings in the history of spinal muscular atrophy. *Neuromuscul Disord* **19**, 69–73 (2009).
11. Kolb, S. J. & Kissel, J. T. Spinal muscular atrophy: a timely review. *Arch Neurol* **68**, 979–984 (2011).
12. Lunn, M. R. & Wang, C. H. Spinal muscular atrophy. *Lancet* **371**, 2120–2133 (2008).
13. Dubowitz, V. Infantile muscular atrophy--a broad spectrum. *Clin. Proc. Child. Hosp. Dist. Columbia* **23**, 223–239 (1967).

14. Munsat, T. L. International SMA Collaboration. *Neuromuscul Disord* **1**, 81 (1991).

15. Munsat, T. L. & Davies, K. E. International SMA consortium meeting. (26-28 June 1992, Bonn, Germany). *Neuromuscul Disord* **2**, 423–428 (1992).

16. Finkel, R. S. *et al.* 218th ENMC International Workshop:: Revisiting the consensus on standards of care in SMA Naarden, The Netherlands, 19–21 February 2016. *Neuromuscul. Disord.* **27**, 596–605 (2017).

17. Mercuri, E. *et al.* Nusinersen versus Sham Control in Later-Onset Spinal Muscular Atrophy. *N. Engl. J. Med.* **378**, 625–635 (2018).

18. Jones, C. C. *et al.* Spinal Muscular Atrophy (SMA) Subtype Concordance in Siblings: Findings from the Cure SMA Cohort. *J. Neuromuscul. Dis.* **7**, 33–40 (2020).

19. Verhaart, I. E. C. *et al.* A multi-source approach to determine SMA incidence and research ready population. *J. Neurol.* **264**, 1465–1473 (2017).

20. Verhaart, I. E. C. *et al.* Prevalence, incidence and carrier frequency of 5q-linked spinal muscular atrophy - A literature review. *Orphanet J. Rare Dis.* **12**, 1–15 (2017).

21. Van Den Engel-Hoek, L. *et al.* Dysphagia in spinal muscular atrophy type II: More than a bulbar problem? *Neurology* **73**, 1787–1791 (2009).

22. van der Heul, A. M. B. *et al.* Mastication in Patients with Spinal Muscular Atrophy Types 2 and 3 is Characterized by Abnormal Efficiency, Reduced Endurance, and Fatigue. *Dysphagia* (2021) doi:10.1007/s00455-021-10351-y.

23. McGrattan, K., Graham, R., Didonato, C. J. & Darras, B. T. Dysphagia Phenotypes in Spinal Muscular Atrophy: The Past, Present, and Promise for the Future. *Am. J. Speech-Language Pathol.* **30**, 1–15 (2021).

24. Samaha, F. J. *et al.* Pulmonary Function in Spinal Muscular Atrophy. *J. Child Neurol.* **9**, 326–329 (1994).

25. CureSMA. Breathing Risks and Care. https://www.curesma.org/treatment/ (2022).

26. Finkel, R. S. *et al.* Diagnosis and management of spinal muscular atrophy: Part 2: Pulmonary

and acute care; medications, supplements and immunizations; other organ systems; and ethics. *Neuromuscul. Disord.* **28**, 197–207 (2018).

27. Lin, L.-C. & Jong, Y.-J. Pulmonary function assessment in patients with spinal muscular atrophy type II and type III. *Acta Paediatr. Taiwan* **45**, 15–18 (2004).
28. Wijngaarde, C. A. *et al.* Natural history of lung function in spinal muscular atrophy. *Orphanet J. Rare Dis.* **15**, 1–11 (2020).
29. Mercuri, E. *et al.* Diagnosis and management of spinal muscular atrophy: Part 1: Recommendations for diagnosis, rehabilitation, orthopedic and nutritional care. *Neuromuscul. Disord.* **28**, 103–115 (2018).
30. Wang, H. Y., Ju, Y. H., Chen, S. M., Lo, S. K. & Jong, Y. J. Joint range of motion limitations in children and young adults with spinal muscular atrophy. *Arch. Phys. Med. Rehabil.* **85**, 1689–1693 (2004).
31. Fujak, A., Kopschina, C., Gras, F., Forst, R. & Forst, J. Contractures of the upper extremities in spinal muscular atrophy type II. Descriptive clinical study with retrospective data collection. *Ortop. Traumatol. Rehabil.* **12**, 410–419 (2010).
32. Fujak, A., Kopschina, C., Gras, F., Forst, R. & Forst, J. Contractures of the lower extremities in spinal muscular atrophy type II. Descriptive clinical study with retrospective data collection. *Ortop. Traumatol. Rehabil.* **13**, 27–36 (2011).
33. Dhawan, S. R., Saini, L., Ramachandran, R. P. & Sankhyan, N. Joint Hyperlaxity, Proximal Contractures, and Facial Weakness in Child With Spinal Muscular Atrophy. *J. Clin. Neuromuscul. Dis.* **20**, 138–140 (2019).
34. Fujak, A. *et al.* Natural course of scoliosis in proximal spinal muscular atrophy type II and IIIa: Descriptive clinical study with retrospective data collection of 126 patients. *BMC Musculoskelet. Disord.* **14**, 0–4 (2013).
35. Wijngaarde, C. A. *et al.* Natural course of scoliosis and lifetime risk of scoliosis surgery in spinal muscular atrophy. *Neurology* **93**, E149–E158 (2019).

36. Dunaway Young, S. *et al.* Should motor function determine the timing of scoliosis surgery in spinal muscular atrophy? *Neuromuscul. Disord.* **26**, S104 (2016).

37. Loughenbury, P. R. & Tsirikos, A. I. Current concepts in the treatment of neuromuscular scoliosis: clinical assessment, treatment options, and surgical outcomes. *Bone Jt. Open* **3**, 85–92 (2022).

38. Al Amrani, F. *et al.* Scoliosis in Spinal Muscular Atrophy Type 1 in the Nusinersen Era. *Neurol. Clin. Pract.* **12**, 279–287 (2022).

39. Farrar, M. A. *et al.* Emerging therapies and challenges in spinal muscular atrophy. *Ann. Neurol.* **81**, 355–368 (2017).

40. Ogino, S., Wilson, R. B. & Gold, B. New insights on the evolution of the SMN1 and SMN2 region: Simulation and meta-analysis for allele and haplotype frequency calculations. *Eur. J. Hum. Genet.* **12**, 1015–1023 (2004).

41. Carson, V. J. *et al.* Spinal muscular atrophy within Amish and Mennonite populations: Ancestral haplotypes and natural history. *PLoS One* **13**, 1–18 (2018).

42. Tizzano, E. F. & Finkel, R. S. Spinal muscular atrophy: A changing phenotype beyond the clinical trials. *Neuromuscul. Disord.* **27**, 883–889 (2017).

43. Zerres, K. & Rudnik-Schöneborn, S. Natural History in Proximal Spinal Muscular Atrophy. *Arch. Neurol.* **52**, 518–523 (1995).

44. Mercuri, E. *et al.* Longitudinal natural history of type i spinal muscular atrophy: A critical review. *Orphanet J. Rare Dis.* **15**, 1–7 (2020).

45. Cances, C. *et al.* Natural history of Type 1 spinal muscular atrophy: a retrospective, global, multicenter study. *Orphanet J. Rare Dis.* **17**, 300 (2022).

46. Annoussamy, M. *et al.* Natural history of Type 2 and 3 spinal muscular atrophy: 2-year NatHis-SMA study. *Ann. Clin. Transl. Neurol.* **8**, 359–373 (2021).

47. Murphy, S. L., Xu, J., Kochanek, K. D., Curtin, S. C. & Arias, E. Deaths: Final Data for 2015. *Natl Vital Stat Rep* **66**, (2017).

48. Cuscó, I. *et al.* A genetic and phenotypic analysis in Spanish spinal muscular atrophy patients with c.399_402del AGAG, the most frequently found subtle mutation in the SMN1 gene. *Hum. Mutat.* **22**, 136–143 (2003).

49. Jordanova, A. *et al.* Spinal muscular atrophy among the Roma (Gypsies) in Bulgaria and Hungary. *Neuromuscul. Disord.* **12**, 378–385 (2002).

50. Chong, J. X. *et al.* A common spinal muscular atrophy deletion mutation is present on a single founder haplotype in the US Hutterites. *Eur. J. Hum. Genet.* **19**, 1045–1051 (2011).

51. Fried, K. & Mundel, G. High incidence of spinal muscular atrophy type I (Werdnig - Hoffmann disease) in the Karaite community in Israel. *Clin. Genet.* **12**, 250–251 (1977).

52. Basel-Vanagaite, L. *et al.* Genetic carrier screening for spinal muscular atrophy and spinal muscular atrophy with respiratory distress 1 in an isolated population in Israel. *Genet. Test.* **12**, 53–56 (2008).

53. P<sc>ascalet-Guidon, M. -J *et al.* Cluster of acute infantile spinal muscular atrophy (Werdnig-Hoffmann disease) in a limited area of Reunion Island. *Clin. Genet.* **26**, 39–42 (1984).

54. Sugarman, E. A. *et al.* Pan-ethnic carrier screening and prenatal diagnosis for spinal muscular atrophy: clinical laboratory analysis of >72,400 specimens. *Eur J Hum Genet* **20**, 27–32 (2012).

55. Radhakrishnan, K., Thacker, A. K. & Maloo, J. C. A clinical, epidemiological and genetic study of hereditary motor neuropathies in Benghazi, Libya. *J. Neurol.* **235**, 422–424 (1988).

56. Ibrahim, S., Moatter, T. & Saleem, A. F. Spinal muscular atrophy: clinical spectrum and genetic mutations in Pakistani children. *Neurol. India* **60**, 294–298 (2012).

57. Badoe, E. V. Consanguinity and rare neurological disease. A five year experience from the Korle Bu Teaching Hospital, Accra, Ghana. *eNeurologicalSci* **3**, 33–34 (2016).

58. Czibere, L. *et al.* High-throughput genetic newborn screening for spinal muscular atrophy by rapid nucleic acid extraction from dried blood spots and 384-well qPCR. *Eur. J. Hum.*

Genet. **28**, 23–30 (2020).

59. Vill, K. *et al.* One year of newborn screening for SMA – Results of a German pilot project. *J. Neuromuscul. Dis.* **6**, 503–515 (2019).

60. Kay, D. M. *et al.* Implementation of population-based newborn screening reveals low incidence of spinal muscular atrophy. *Genet. Med.* **22**, 1296–1302 (2020).

61. Andoni- Urtizberea, J., Ferroudja, D. & Filnemus. Combien de patients atteints de SMA en France ? *Médecine/Sciences* **34**, 32–34 (2018).

62. Melki, J. *et al.* Gene for chronic proximal spinal muscular atrophies maps to chromosome 5q. *Nature* **344**, 767–768 (1990).

63. Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80**, 155–165 (1995).

64. Zerres, K. & Rudnik-Schöneborn, S. 93rd ENMC international workshop: Non-5q-spinal muscular atrophies (SMA) - Clinical picture (6-8 April 2001, Naarden, The Netherlands). *Neuromuscul. Disord.* **13**, 179–183 (2003).

65. Parkinson, N. J., Baumer, D., Rose-Morris, A. & Talbot, K. Candidate screening of the bovine and feline spinal muscular atrophy genes reveals no evidence for involvement in human motor neuron disorders. *Neuromuscul. Disord.* **18**, 394–397 (2008).

66. Axente, M. *et al.* Clinical features and genetics in non-5q spinal muscular atrophy caused by acid ceramidase deficiency. *J. Med. Life* **14**, 424–428 (2021).

67. Dennis, M. Y. *et al.* The evolution and population diversity of human-specific segmental duplications. *Nat. Ecol. Evol.* **1**, 1–10 (2017).

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

69. Lorson, C. L., Hahnen, E., Androphy, E. J. & Wirth, B. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci.*

U. S. A. **96**, 6307–6311 (1999).

70. Monani, U. R. *et al.* A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum Mol Genet* **1999**, 7 (1999).

71. Kashima, T. & Manley, J. L. A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat. Genet.* **34**, 460–463 (2003).

72. Caputi, M., Mayeda, A., Krainer, A. R. & Zahler, A. M. hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *EMBO J.* **18**, 4060–4067 (1999).

73. Hou, V. C. *et al.* Decrease in hnRNP A/B expression during erythropoiesis mediates a pre-mRNA splicing switch. *EMBO J.* **21**, 6195–6204 (2002).

74. Burd, C. G. & Dreyfuss, G. RNA binding specificity of hnRNP A1: Significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* **13**, 1197–1204 (1994).

75. Bilodeau, P. S., Domsic, J. K., Mayeda, A., Krainer, A. R. & Stoltzfus, C. M. RNA Splicing at Human Immunodeficiency Virus Type 1 3' Splice Site A2 Is Regulated by Binding of hnRNP A/B Proteins to an Exonic Splicing Silencer Element. *J. Virol.* **75**, 8487–8497 (2001).

76. Cartegni, L., Hastings, M. L., Calarco, J. A., De Stanchina, E. & Krainer, A. R. Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. *Am. J. Hum. Genet.* **78**, 63–77 (2006).

77. Cartegni, L. & Krainer, A. R. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN. *Nat. Genet.* **30**, 377–384 (2002).

78. Talbot, K. *et al.* Missense mutation clustering in the survival motor neuron gene: a role for a conserved tyrosine and glycine rich region of the protein in RNA metabolism? *Hum Mol Genet* **6**, 497–500 (1997).

79. Lorson, C. L. *et al.* SMN oligomerization defect correlates with spinal muscular atrophy severity. *Nat Genet* **19**, 63–66 (1998).

80. Gupta, K. *et al.* Assembly of higher-order SMN oligomers is essential for metazoan viability and requires an exposed structural motif present in the YG zipper dimer. *Nucleic Acids Res.* **49**, 7644–7664 (2021).

81. Burnett, B. G. *et al.* Regulation of SMN protein stability. *Mol. Cell. Biol.* **29**, 1107–1115 (2009).

82. Lorson, C. L. & Androphy, E. J. An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. *Hum Mol Genet* **9**, 259–265 (2000).

83. Singh, N. N., Seo, J., Rahn, S. J. & Singh, R. N. A Multi-Exon-Skipping Detection Assay Reveals Surprising Diversity of Splice Isoforms of Spinal Muscular Atrophy Genes. *PLoS One* **7**, (2012).

84. Seo, J., Singh, N. N., Ottesen, E. W., Lee, B. M. & Singh, R. N. A novel human-specific splice isoform alters the critical C-terminus of Survival Motor Neuron protein. *Sci. Rep.* **6**, 1–14 (2016).

85. Ottesen, E. W., Luo, D., Seo, J., Singh, R. N. & Singh, N. N. Human Survival Motor Neuron genes generate a vast repertoire of circular RNAs. *Nucleic Acids Res.* **1**, 1–22 (2019).

86. Paushkin, S. *et al.* The survival motor neuron protein of Schizosacharomyces pombe: Conservation of survival motor neuron interaction domains in divergent organisms. *J. Biol. Chem.* **275**, 23841–23846 (2000).

87. Veepaschit, J., Viswanathan, A., Bordonné, R., Grimm, C. & Fischer, U. Identification and structural analysis of the Schizosaccharomyces pombe SMN complex . *Nucleic Acids Res.* 1–17 (2021) doi:10.1093/nar/gkab158.

88. Kolb, S. J., Battle, D. J. & Dreyfuss, G. Molecular functions of the SMN complex. *J Child Neurol* **22**, 990–994 (2007).

89. Osman, E. Y. *et al.* Functional characterization of SMN evolution in mouse models of SMA. *Sci. Rep.* **9**, 1–12 (2019).

90. Pellizzoni, L., Kataoka, N., Charroux, B. & Dreyfuss, G. A novel function for SMN, the

spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell* **95**, 615–624 (1998).

91. Pellizzoni, L., Yong, J. & Dreyfuss, G. Essential role for the SMN complex in the specificity of snRNP assembly. *Science (80-.).* **298**, 1775–1779 (2002).

92. Wahl, M. C., Will, C. L. & Lührmann, R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **136**, 701–718 (2009).

93. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, 1–2 (2011).

94. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* **15**, 108–121 (2014).

95. Shi, Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.* **18**, 655–670 (2017).

96. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).

97. Berget, S. M. & Robberson, B. L. U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for in vitro splicing but not polyadenylation. *Cell* **46**, 691–696 (1986).

98. Turunen, J. J., Niemelä, E. H., Verma, B. & Frilander, M. J. The significant other: Splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA* **4**, 61–76 (2013).

99. Steitz, J. A. *et al.* Where in the cell is the minor spliceosome? *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8485–8486 (2008).

100. Montzka, K. A. & Steitz, J. A. Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8885–8889 (1988).

101. Tarn, W. Y. & Steitz, J. A. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science (80-.).* **273**, 1824–1832 (1996).

102. Shao, W., Kim, H.-S., Cao, Y., Xu, Y.-Z. & Query, C. C. A U1-U2 snRNP Interaction Network during Intron Definition. *Mol. Cell. Biol.* **32**, 470–478 (2012).

103. Bertram, K. *et al.* Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell* **170**, 701-713.e11 (2017).

104. Plaschka, C., Lin, P.-C. & Nagain, K. Structure of a pre-catalytic spliceosome. *Nature* **546**, 617–621 (2017).

105. Zhang, X. *et al.* Structure of the human activated spliceosome in three conformational states. *Cell Res.* **28**, 307–322 (2018).

106. Jurica, M. S., Sousa, D., Moore, M. J. & Grigorieff, N. Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat. Struct. Mol. Biol.* **11**, 265–269 (2004).

107. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 3171–3175 (1977).

108. Chow, L. T., Roberts, J. M., Lewis, J. B. & Broker, T. R. A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. *Cell* **11**, 819–836 (1977).

109. Jackson, L. J. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* **19**, 3795–3798 (1991).

110. Sheth, N. *et al.* Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* **34**, 3955–3967 (2006).

111. Pagliardini, S. *et al.* Subcellular localization and axonal transport of the survival motor neuron (SMN) protein in the developing rat spinal cord. *Hum Mol Genet* **9**, 47–56 (2000).

112. Gubitz, A. K., Feng, W. & Dreyfuss, G. The SMN complex. *Exp Cell Res* **296**, 51–56 (2004).

113. Battle, D. J. *et al.* The Gemin5 Protein of the SMN Complex Identifies snRNAs. *Mol. Cell* **23**, 273–279 (2006).

114. Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J. & Bordonné, R. Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol. Cell* **9**, 891–901 (2002).

115. Mahmoudi, S. *et al.* Wrap53 Is Essential for Cajal Body Formation and for Targeting the

Survival of Motor Neuron Complex To Cajal Bodies. *PLoS Biol.* **8**, 1–16 (2010).

116. Narayanan, U., Achsel, T., Lührmann, R. & Matera, A. G. Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein. *Mol. Cell* **16**, 223–234 (2004).

117. Raker, V. A., Hartmuth, K., Kastner, B. & Lührmann, R. Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol Cell Biol.* **19**, 6554–6565 (1999).

118. Fallini, C., Bassell, G. J. & Rossoll, W. Spinal muscular atrophy: the role of SMN in axonal mRNA regulation. *Brain Res* **1462**, 81–92 (2012).

119. Fallini, C. *et al.* Dynamics of survival of motor neuron (SMN) protein interaction with the mRNA-binding protein IMP1 facilitates its trafficking into motor neuron axons. *Dev Neurobiol* **74**, 319–332 (2014).

120. Makarov, E. M., Owen, N., Bottrill, A. & Makarova, O. V. Functional mammalian spliceosomal complex e contains SMN complex proteins in addition to U1 and U2 snRNPs. *Nucleic Acids Res.* **40**, 2639–2652 (2012).

121. Tisdale, S. *et al.* SMN is essential for the biogenesis of U7 Small nuclear ribonucleoprotein and 3'-end formation of Histone mRNAs. *Cell Rep.* **5**, 1187–1195 (2013).

122. Pillai, R. S. *et al.* Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. *Genes Dev.* **17**, 2321–2333 (2003).

123. Pellizzoni, L., Baccon, J., Charroux, B. & Dreyfuss, G. The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. *Curr. Biol.* **11**, 1079–1088 (2001).

124. Jones, K. W. *et al.* Direct Interaction of the Spinal Muscular Atrophy Disease Protein SMN with the Small Nucleolar RNA-associated Protein Fibrillarin. *J. Biol. Chem.* **276**, 38645–38651 (2001).

125. Kannan, A., Bhatia, K., Branzei, D. & Gangwani, L. Combined deficiency of Senataxin and

DNA-PKcs causes DNA damage accumulation and neurodegeneration in spinal muscular atrophy. *Nucleic Acids Res.* **46**, 8326–8346 (2018).

126. Karyka, E. *et al.* SMN-deficient cells exhibit increased ribosomal DNA damage. *Life Sci. Alliance* **5**, 1–19 (2022).

127. Bachand, F., Boisvert, F.-M., Côté, J., Richard, S. & Autexier, C. The product of the Survival of Motor Neuron (SMN) gene is a human telomerase-associated protein. *Mol. Biol. Cell* **13**, 3192–3202 (2002).

128. Poole, A. R. & Hebert, M. D. SMN and coilin negatively regulate dyskerin association with telomerase RNA. *Biol. Open* **5**, 726–735 (2016).

129. Yanling Zhao, D. *et al.* SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature* **529**, 48–53 (2016).

130. Piazzon, N. *et al.* Implication of the SMN complex in the biogenesis and steady state level of the Signal Recognition Particle. *Nucleic Acids Res.* **41**, 1255–1272 (2013).

131. Singh, R. N., Howell, M. D., Ottesen, E. W. & Singh, N. N. Diverse role of survival motor neuron protein. *Biochim Biophys Acta* **1860**, 299–315 (2017).

132. Lauria, F. *et al.* SMN-primed ribosomes modulate the translation of transcripts related to spinal muscular atrophy. *Nat. Cell Biol.* **22**, 1239–1251 (2020).

133. Wurth, L. *et al.* Hypermethylated-capped selenoprotein mRNAs in mammals. *Nucleic Acids Res.* **42**, 8663–8677 (2014).

134. Setola, V. *et al.* Axonal-SMN (a-SMN), a protein isoform of the survival motor neuron gene, is specifically involved in axonogenesis. *Proc Natl Acad Sci U S A* **104**, 1959–1964 (2007).

135. Zhang, Z. *et al.* SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* **133**, 585–600 (2008).

136. Gabanella, F. *et al.* Ribonucleoprotein assembly defects correlate with spinal muscular atrophy severity and preferentially affect a subset of spliceosomal snRNPs. *PLoS One* **2**, e921 (2007).

137. Boulisfane, N. *et al.* 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**, 641–648 (2011).

138. Lotti, F. *et al.* A SMN-Dependent U12 Splicing Event Essential for Motor Circuit Function. *Cell* **151**, 440–454 (2012).

139. Doktor, T. K. *et al.* RNA-sequencing of a mouse-model of spinal muscular atrophy reveals tissue-wide changes in splicing of U12-dependent introns. *Nucleic Acids Res.* **45**, 395–416 (2017).

140. Jangi, M. *et al.* SMN deficiency in severe models of spinal muscular atrophy causes widespread intron retention and DNA damage. *Proc Natl Acad Sci U S A* E2347–E2356 (2017).

141. Zhang, Z. *et al.* Dysregulation of synaptogenesis genes antecedes motor neuron pathology in spinal muscular atrophy. *Proc Natl Acad Sci U S A* **110**, 19348–19353 (2013).

142. Van Alstyne, M., Simon, C. M., Sardi, S. P. & Shihabuddin, L. S. Mdm2 and Mdm4 splicing dysregulation underlies motor neuron death in SMA. *Genes Dev.* 1–15 (2018) doi:10.1101/gad.316059.118.5.

143. Gao, X. *et al.* Defective Expression of Mitochondrial , Vacuolar H + -ATPase and Histone Genes in a C . elegans Model of SMA. *Front Genet* **10**, 1–14 (2019).

144. Murray, L. M., Beauvais, A., Gibeault, S., Courtney, N. L. & Kothary, R. Transcriptional profiling of differentially vulnerable motor neurons at pre-symptomatic stage in the Smn (2b/-) mouse model of spinal muscular atrophy. *Acta Neuropathol. Commun.* **3**, 55 (2015).

145. Maeda, M. *et al.* Transcriptome profiling of spinal muscular atrophy motor neurons derived from mouse embryonic stem cells. *PLoS One* **9**, (2014).

146. Bäumer, D. *et al.* Alternative splicing events are a late feature of pathology in a mouse model of spinal muscular atrophy. *PLoS Genet* **5**, e1000773 (2009).

147. Tu, W. Y., Simpson, J. E., Highley, J. R. & Heath, P. R. Spinal muscular atrophy: Factors

that modulate motor neuron vulnerability. *Neurobiol. Dis.* **102**, 11–20 (2017).

148. Onodera, O. *et al.* Minor splicing pathway is not minor any more: Implications for the pathogenesis of motor neuron diseases. *Neuropathology* **34**, 99–107 (2014).

149. Murray, L. M. *et al.* 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**, 949–962 (2008).

150. See, K. *et al.* SMN deficiency alters Nrxn2 expression and splicing in zebrafish and mouse models of spinal muscular atrophy. *Hum. Mol. Genet.* **23**, 1754–1770 (2014).

151. Comley, L. H. *et al.* Motor neurons with differential vulnerability to degeneration show distinct protein signatures in health and ALS. *Neuroscience* **291**, 216–229 (2015).

152. Nichterwitz, S. *et al.* LCM-seq reveals unique transcriptional adaptation mechanisms of resistant neurons and identifies protective pathways in spinal muscular atrophy. *Genome Res.* 1083–1096 (2020) doi:10.1101/gr.265017.120.

153. Rodríguez Cruz, P. M., Cossins, J., Beeson, D. & Vincent, A. The Neuromuscular Junction in Health and Disease: Molecular Mechanisms Governing Synaptic Formation and Homeostasis. *Front. Mol. Neurosci.* **13**, 1–22 (2020).

154. Kong, L. *et al.* Impaired synaptic vesicle release and immaturity of neuromuscular junctions in spinal muscular atrophy mice. *J. Neurosci.* **29**, 842–851 (2009).

155. Kariya, S. *et al.* Reduced SMN protein impairs maturation of the neuromuscular junctions in mouse models of spinal muscular atrophy. *Hum. Mol. Genet.* **17**, 2552–2569 (2008).

156. Ling, K. K. Y., Gibbs, R. M., Feng, Z. & Ko, C. P. Severe neuromuscular denervation of clinically relevant muscles in a mouse model of spinal muscular atrophy. *Hum. Mol. Genet.* **21**, 185–195 (2012).

157. Boido, M. *et al.* Increasing agrin function antagonizes muscle atrophy and motor impairment in spinal muscular atrophy. *Front. Cell. Neurosci.* **12**, 1–14 (2018).

158. Swoboda, K. J. *et al.* Natural history of denervation in SMA: Relation to age, SMN2 copy

number, and function. *Ann. Neurol.* **57**, 704–712 (2005).

159. Harding, B. *et al.* Spectrum of Neuropathophysiology in Spinal Muscular Atrophy. *J Neuropath Exper Neurol* **74**, 15–24 (2015).

160. Yoshida, M. *et al.* Modeling the early phenotype at the neuromuscular junction of spinal muscular atrophy using patient-derived iPSCs. *Stem Cell Reports* **4**, 561–568 (2015).

161. Boido, M. & Vercelli, A. Neuromuscular junctions as key contributors and therapeutic targets in spinal muscular atrophy. *Front. Neuroanat.* **10**, 1–10 (2016).

162. Ruiz, R., Casañas, J. J., Torres-Benito, L., Cano, R. & Tabares, L. Altered intracellular Ca2+ homeostasis in nerve terminals of severe spinal muscular atrophy mice. *J. Neurosci.* **30**, 849–857 (2010).

163. Wadman, R. I., Vrancken, A. F. J. E., Van Den Berg, L. H. & Van Der Pol, W. L. Dysfunction of the neuromuscular junction in spinal muscular atrophy types 2 and 3. *Neurology* **79**, 2050–2055 (2012).

164. Martínez-Hernández, R. *et al.* The developmental pattern of myotubes in spinal muscular atrophy indicates prenatal delay of muscle maturation. *J. Neuropathol. Exp. Neurol.* **68**, 474–481 (2009).

165. Ripolone, M. *et al.* Impaired muscle mitochondrial biogenesis and myogenesis in spinal muscular atrophy. *JAMA Neurol.* **72**, 666–675 (2015).

166. McCormack, N. M. *et al.* Survival motor neuron deficiency slows myoblast fusion through reduced myomaker and myomixer expression. *J. Cachexia. Sarcopenia Muscle* 1–19 (2021) doi:10.1002/jcsm.12740.

167. Shafey, D., Côté, P. D. & Kothary, R. Hypomorphic Smn knockdown C2C12 myoblasts reveal intrinsic defects in myoblast fusion and myotube morphology. *Exp. Cell Res.* **311**, 49–61 (2005).

168. Sansa, A. *et al.* Spinal Muscular Atrophy autophagy profile is tissue-dependent: differential regulation between muscle and motoneurons. *Acta Neuropathol. Commun.* **9**, 1–15 (2021).

169. Hellbach, N. *et al.* Impaired myogenic development, differentiation and function in hESC-derived SMA myoblasts and myotubes. *PLoS One* **13**, 1–21 (2018).

170. Boyer, J. G. *et al.* Myogenic program dysregulation is contributory to disease pathogenesis in spinal muscular atrophy. *Hum. Mol. Genet.* **23**, 4249–4259 (2014).

171. Ando, S. *et al.* SMN Protein Contributes to Skeletal Muscle Cell Maturation Via Caspase-3 and Akt Activation. *In Vivo (Brooklyn).* **34**, 3247–3254 (2020).

172. Hayhurst, M., Wagner, A. K., Cerletti, M., Wagers, A. J. & Rubin, L. L. A cell-autonomous defect in skeletal muscle satellite cells expressing low levels of survival of motor neuron protein. *Dev. Biol.* **368**, 323–334 (2012).

173. Nicole, S. *et al.* Intact satellite cells lead to remarkable protection against Smn gene defect in differentiated skeletal muscle. *J. Cell Biol.* **161**, 571–582 (2003).

174. Mecca, J. *et al.* Role of muscle satellite cells in Spinal muscular atrophy physiopathology. *Neuromuscul. Disord.* **27**, S136 (2017).

175. Schrank, B. *et al.* Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 9920–9925 (1997).

176. Monani, U. R., Coovert, D. D. & Burghes, A. H. M. Animal models of spinal muscular atrophy. *Hum. Mol. Genet.* **9**, 2451–2457 (2000).

177. Bebee, T. W., Dominguez, C. E. & Chandler, D. S. Mouse models of SMA: tools for disease characterization and therapeutic development. *Hum Genet* **131**, 1277–1293 (2012).

178. Sleigh, J. N., Gillingwater, T. H. & Talbot, K. The contribution of mouse models to understanding the pathogenesis of spinal muscular atrophy. *Dis Model Mech* **4**, 457–467 (2011).

179. Anderson, K. N., Baban, D., Oliver, P. L., Potter, A. & Davies, K. E. Expression profiling in spinal muscular atrophy reveals an RNA binding protein deficit. *Neuromuscul. Disord.* **14**, 711–722 (2004).

180. Edens, B. M., Ajroud-Driss, S., Ma, L. & Ma, Y. C. Molecular mechanisms and animal models of spinal muscular atrophy. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1852**, 685–692 (2015).

181. Hsieh-Li, H. M. *et al.* A mouse model for spinal muscular atrophy. *Nat Genet* **24**, 66–70 (2000).

182. Le, T. T. *et al.* SMNΔ7, 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**, 845–857 (2005).

183. Gogliotti, R. G., Hammond, S. M., Lutz, C. & Didonato, C. J. Molecular and phenotypic reassessment of an infrequently used mouse model for spinal muscular atrophy. *Biochem Biophys Res Commun* **391**, 517–522 (2010).

184. Hammond, S. M. *et al.* Mouse survival motor neuron alleles that mimic SMN2 splicing and are inducible rescue embryonic lethality early in development but not late. *PLoS One* **5**, (2010).

185. Bowerman, M. *et al.* SMN, profilin IIa and plastin 3: A link between the deregulation of actin dynamics and SMA pathogenesis. *Mol. Cell. Neurosci.* **42**, 66–74 (2009).

186. Bowerman, M., Beauvais, A., Anderson, C. L. & Kothary, R. Rho-kinase inactivation prolongs survival of an intermediate SMA mouse model. *Hum Mol Genet* **19**, 1468–1478 (2010).

187. Frugier, T. *et al.* Nuclear targeting defect of SMN lacking the C-terminus in a mouse model of spinal muscular atrophy. *Hum. Mol. Genet.* **9**, 849–858 (2000).

188. Mullard, A. FDA approves SMA gene therapy. *Nat. Rev. Drug Discov.* **18**, 488 (2019).

189. FDA Approves Genentech's Evrysdi (risdiplam) for Treatment of Spinal Muscular Atrophy (SMA) in Adults and Children 2 Months and Older. https://www.gene.com/media/press-releases/14866/2020-08-07/fda-approves-genentechs-evrysdi-risdipla (2020).

190. EMA. First medicine for spinal muscular atrophy.

https://www.ema.europa.eu/en/news/first-medicine-spinal-muscular-atrophy (2017).

191. Singh, N. K., Singh, N. N., Androphy, E. J. & Singh, R. N. Splicing of a critical exon of human Survival Motor Neuron is regulated by a unique silencer element located in the last intron. *Mol Cell Biol* **26**, 1333–1346 (2006).

192. Singh, N. N., Howell, M. D., Androphy, E. J. & Singh, R. N. How the discovery of ISS-N1 led to the first medical therapy for spinal muscular atrophy. *Gene Ther* **24**, 520–526 (2017).

193. Chiriboga, C. A. *et al.* Results from a phase 1 study of nusinersen (ISIS-SMN(Rx)) in children with spinal muscular atrophy. *Neurology* **86**, 890–897 (2016).

194. Finkel, R. S. *et al.* Treatment of infantile-onset spinal muscular atrophy with nusinersen: final report of a phase 2, open-label, multicentre, dose-escalation study. *Lancet Child Adolesc. Heal.* **5**, 491–500 (2021).

195. Finkel, R. S. *et al.* Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. *Lancet* **388**, 3017-3026. (2016).

196. Ramos, D. M. *et al.* Age-dependent SMN expression in disease-relevant tissue and implications for SMA treatment. *J. Clin. Invest.* (2019) doi:10.1172/jci124120.

197. Naryshkin, N. A. *et al.* Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science (80-.).* **345**, 688–693 (2014).

198. Ratni, H. *et al.* Discovery of Risdiplam, a Selective Survival of Motor Neuron-2 (SMN2) Gene Splicing Modifier for the Treatment of Spinal Muscular Atrophy (SMA). *J. Med. Chem.* **61**, 6501–6517 (2018).

199. Wang, J., Schultz, P. G. & Johnson, K. A. Mechanistic studies of a small-molecule modulator of SMN2 splicing. *Proc Natl Acad Sci U S A* **115**, E4604–E4612 (2018).

200. Singh, R. N., Ottesen, E. W. & Singh, N. N. The First Orally Deliverable Small Molecule for the Treatment of Spinal Muscular Atrophy. *Neurosci. Insights* **15**, (2020).

201. Chiriboga, C. A. *et al.* Results from a phase 1 study of nusinersen (ISIS-SMN Rx) in children

with spinal muscular atrophy. *Neurology* **86**, 890–897 (2016).

202. Finkel, R. S. *et al.* Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. *N. Engl. J. Med.* **377**, 1723–1732 (2017).

203. Hache, M. *et al.* Intrathecal Injections in Children With Spinal Muscular Atrophy: Nusinersen Clinical Trial Experience. *J Child Neurol* **31**, 899–906 (2016).

204. Chiriboga, C. *et al.* Longer-term treatment with nusinersen: Results in later-onset spinal muscular atrophy from the SHINE study. *Neurology* **94**, S67 (2020).

205. De Vivo, D. C. *et al.* Nusinersen initiated in infants during the presymptomatic stage of spinal muscular atrophy: Interim efficacy and safety results from the Phase 2 NURTURE study. *Neuromuscul. Disord.* **29**, 842–856 (2019).

206. Acsadi, G. *et al.* Safety and efficacy of nusinersen in spinal muscular atrophy: The EMBRACE study. *Muscle and Nerve* **63**, 668–677 (2021).

207. Mendell, J. R. *et al.* Five-Year Extension Results of the Phase 1 START Trial of Onasemnogene Abeparvovec in Spinal Muscular Atrophy. *JAMA Neurol.* **43205**, 1–8 (2021).

208. Strauss, K. A. *et al.* Onasemnogene abeparvovec for presymptomatic infants with two copies of SMN2 at risk for spinal muscular atrophy type 1: the Phase III SPR1NT trial. *Nat. Med.* **28**, 1381–1389 (2022).

209. Day, J. W. *et al.* Onasemnogene abeparvovec gene therapy for symptomatic infantile-onset spinal muscular atrophy in patients with two copies of SMN2 (STR1VE): an open-label, single-arm, multicentre, phase 3 trial. *Lancet Neurol.* **20**, 284–293 (2021).

210. Mercuri, E. *et al.* Onasemnogene abeparvovec gene therapy for symptomatic infantile-onset spinal muscular atrophy type 1 (STR1VE-EU): an open-label, single-arm, multicentre, phase 3 trial. *Lancet Neurol.* **20**, 832–841 (2021).

211. Baranello, G. *et al.* Risdiplam in Type 1 Spinal Muscular Atrophy. *N. Engl. J. Med.* **384**, 915–923 (2021).

212. Mercuri, E. *et al.* Safety and efficacy of once-daily risdiplam in type 2 and non-ambulant type 3 spinal muscular atrophy (SUNFISH part 2): a phase 3, double-blind, randomised, placebo-controlled trial. *Lancet Neurol.* **21**, 42–52 (2022).

213. Chiriboga, C. A. *et al.* JEWELFISH: Safety and pharmacodynamic data in non-naïve patients with spinal muscular atrophy receiving treatment with risdiplam (RG7916). *Neurology* **96**, 30 (2021).

214. Finkel, R. S. *et al.* RAINBOWFISH: A study of risdiplam in infants with presymptomatic spinal muscular atrophy (SMA). *Neurology* **96**, (2021).

215. Craig Venter, J. *et al.* The sequence of the human genome. *Science (80-.).* **291**, 1304–1351 (2001).

216. International Human Genome Sequencing Consortium *et al.* Initial sequencing and analysis of the human genome. *Nature* **412**, 565–566 (2001).

217. Harbeck, E. L. *et al.* The complete sequence of a human genome. *Science (80-.).* **58**, 1–14 (2022).

218. Khvorova, A. & Watts, J. K. The chemical evolution of oligonucleotide therapies of clinical utility. *Nat. Biotechnol.* **35**, 238–248 (2017).

219. Goli, L. Macromolécules thérapeutiques et barrière hémato-encéphaliques: exemples tirés de la recherche et du développement d'oligonucléotides antisens. (Université Paris Descartes, 2019).

220. Abramov, A. Y. & Bachurin, S. O. Neurodegenerative disorders—Searching for targets and new ways of diseases treatment. *Med. Res. Rev.* **41**, 2603–2605 (2021).

221. Lamptey, R. N. L. *et al.* A Review of the Common Neurodegenerative Disorders: Current Therapeutic Approaches and the Potential Role of Nanotherapeutics. *Int. J. Mol. Sci.* **23**, (2022).

222. Sun, J. & Roy, S. Gene-based therapies for neurodegenerative diseases. *Nat. Neurosci.* **24**, 297–311 (2021).

223. Winkelsas, A. M. *et al.* Targeting the 5' untranslated region of SMN2 as a therapeutic strategy for spinal muscular atrophy. *Mol. Ther. - Nucleic Acids* **23**, 731–742 (2021).

224. Yu, T. W. *et al.* PgmNr 3570: Patient-customized oligonucleotide therapy for an ultra-rare genetic disease. in *ASHG 2018 Annual Meeting* (2018).

225. Stolte, B. *et al.* Feasability and safety of intrathecal treatment with nusinersen in adult patients with spinal muscular atrophy. *Ther. Adv. Neurol. Disord.* **11**, 1–9 (2018).

226. Veerapandiyan, A. *et al.* Cervical puncture to deliver nusinersen in patients with spinal muscular atrophy. *Neurology* **91**, e620 LP-e624 (2018).

227. Zhang, J. *et al.* Ultrasound-guided nusinersen administration for spinal muscular atrophy patients with severe scoliosis: an observational study. *Orphanet J. Rare Dis.* **16**, 1–8 (2021).

228. Dangouloff, T. *et al.* Newborn screening programs for spinal muscular atrophy worldwide: Where we stand and where to go. *Neuromuscul. Disord.* **31**, 574–582 (2021).

229. Eggermann, K., Gläser, D., Abicht, A. & Wirth, B. Spinal muscular atrophy (5qSMA): Best practice of diagnostics, newborn screening and therapy. *Medizinische Genet.* **32**, 263–272 (2020).

230. Lopes, J. M. SMA Added to List of Recommended Screenings for Disease Given to Newborns in US. *SMA News Today* https://smanewstoday.com/2018/07/16/sma-added-to-us-list-of-diseases-recommended-for-newborn-screening/ (2016).

231. Lin, C. W., Kalb, S. J. & Yeh, W. S. Delay in Diagnosis of Spinal Muscular Atrophy: A Systematic Literature Review. *Pediatr Neurol* **53**, 293–300 (2015).

232. Kariyawasam, D. S. T., Russell, J. S., Wiley, V., Alexander, I. E. & Farrar, M. A. The implementation of newborn screening for spinal muscular atrophy: the Australian experience. *Genet. Med.* **22**, 557–565 (2020).

233. Dangouloff, T. & Servais, L. Clinical evidence supporting early treatment of patients with spinal muscular atrophy: Current perspectives. *Ther. Clin. Risk Manag.* **15**, 1153–1161 (2019).

234. Abati, E., Manini, A., Comi, G. Pietro & Corti, S. Inhibition of myostatin and related

signaling pathways for the treatment of muscle atrophy in motor neuron diseases. *Cell. Mol. Life Sci.* **79**, 1–21 (2022).

235. Crawford, T. *et al.* Apitegromab in Spinal Muscular Atrophy (SMA): An Analysis of Multiple Efficacy Endpoints in the TOPAZ Trial (P15-5.005). *Neurology* **98**, 1859 (2022).

236. Day, J. W. *et al.* Advances and limitations for the treatment of spinal muscular atrophy. *BMC Pediatr.* **22**, 1–15 (2022).

237. Pedersen, T. H., Riisager, A., de Paoli, F. V., Chen, T. Y. & Nielsen, O. B. Role of physiological ClC-1 Cl- ion channel regulation for the excitability and function of working skeletal muscle. *J. Gen. Physiol.* **147**, 291–308 (2016).

238. Pedersen, T. H., Macdonald, W. A., Broch-Lips, M., Halldorsdottir, O. & Bækgaard Nielsen, O. Chloride channel inhibition improves neuromuscular function under conditions mimicking neuromuscular disorders. *Acta Physiol.* **233**, e13690 (2021).

239. Lin, T. L. *et al.* Selective neuromuscular denervation in Taiwanese severe SMA mouse can be reversed by morpholino antisense oligonucleotides. *PLoS One* **11**, 1–15 (2016).

240. Rybalka, E. *et al.* The Failed Clinical Story of Myostatin Inhibitors against Duchenne Muscular Dystrophy: Exploring the Biology behind the Battle. *Cells* **9**, (2020).

241. Mariot, V. *et al.* Downregulation of myostatin pathway in neuromuscular diseases may explain challenges of anti-myostatin therapeutic approaches. *Nat. Commun.* **8**, 6–13 (2017).

242. Zhang, R. *et al.* Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. *Cell* **146**, 384–395 (2011).

243. Garcia, E. L., Lu, Z., Meers, M. P., Praveen, K. & Matera, A. G. Developmental arrest of Drosophila survival motor neuron (Smn) mutants accounts for differences in expression of minor intron-containing genes. *Rna* **19**, 1510–1516 (2013).

244. Praveen, K., Wen, Y. & Materaa, A. G. A Drosophila model of Spinal Muscular Atrophy uncouples the snRNP biogenesis functions of survival motor neuron from locomotion and viability defects. *Cell Rep.* **1**, 624–631 (2012).

245. Garcia, E. L., Wen, Y., Praveen, K. & Matera, A. G. Transcriptomic comparison of Drosophila snRNP biogenesis mutants reveals mutant-specific changes in pre-mRNA processing: Implications for spinal muscular atrophy. *Rna* **22**, 1215–1227 (2016).

246. Borg, R. M., Bordonne, R., Vassallo, N. & Cauchi, R. J. Genetic interactions between the members of the SMN-Gemins complex in Drosophila. *PLoS One* **10**, 1–23 (2015).

247. Briese, M. *et al.* Deletion of smn-1, the Caenorhabditis elegans ortholog of the spinal muscular atrophy gene, results in locomotor dysfunction and reduced lifespan. *Hum. Mol. Genet.* **18**, 97–104 (2009).

248. Oprea, G. E. *et al.* Plastin 3 is a protective modifier of autosomal recessive spinal muscular atrophy. *Science (80-.).* **320**, 524–527 (2008).

249. Rossoll, W. & Bassell, G. J. Spinal muscular atrophy and a model for survival of motor neuron protein function in axonal ribonucleoprotein complexes. *Results Probl Cell Differ* **48**, 289–326 (2009).

250. Mier, P. & Pérez-Pulido, A. J. Fungal Smn and Spf30 homologues are mainly present in filamentous fungi and genomes with many introns: Implications for spinal muscular atrophy. *Gene* **491**, 135–141 (2012).

251. Campion, Y., Neel, H., Gostan, T., Soret, J. & Bordonné, R. Specific splicing defects in S. pombe carrying a degron allele of the survival of motor neuron gene. *EMBO J.* **29**, 1817–1829 (2010).

252. Antoine, M. *et al.* Splicing Defects of the Profilin Gene Alter Actin Dynamics in an S. pombe SMN Mutant. *iScience* **23**, (2020).

253. Schmid, A. & DiDonato, C. J. Animal models of spinal muscular atrophy. *J. Child Neurol.* **22**, 1004–1012 (2007).

254. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131**, 861–872 (2007).

255. Patitucci, T. N. & Ebert, A. D. SMN deficiency does not induce oxidative stress in SMA

iPSC-derived astrocytes or motor neurons. *Hum. Mol. Genet.* **25**, 514–523 (2016).

256. Martin, J. E. *et al.* Decreased motor neuron support by SMA astrocytes due to diminished MCP1 secretion. *J. Neurosci.* **37**, 5309–5318 (2017).

257. Mcgivern, J. V. *et al.* Spinal muscular atrophy astrocytes exhibit abnormal calcium regulation and reduced growth factor production. *Glia* **61**, 1418–1428 (2013).

258. Lashkari, D. A. *et al.* Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 13057–13062 (1997).

259. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509–1517 (2008).

260. Balabanian, S., Gendron, N. H. & MacKenzie, A. E. Histologic and transcriptional assessment of a mild SMA model. *Neurol. Res.* **29**, 413–424 (2007).

261. Murray, L. M. *et al.* Pre-symptomatic development of lower motor neuron connectivity in a mouse model of severe spinal muscular atrophy. *Hum. Mol. Genet.* **19**, 420–433 (2010).

262. Staropoli, J. F. *et al.* Rescue of gene-expression changes in an induced mouse model of spinal muscular atrophy by an antisense oligonucleotide that promotes inclusion of SMN2 exon 7. *Genomics* **105**, 220–228 (2015).

263. Sudmant, P. H., Alexis, M. S. & Burge, C. B. Meta-analysis of RNA-seq expression data across species, tissues and studies. *Genome Biol.* **16**, 1–11 (2015).

264. Buti, M. *et al.* A meta-analysis of comparative transcriptomic data reveals a set of key genes involved in the tolerance to abiotic stresses in rice. *Int. J. Mol. Sci.* **20**, 1–30 (2019).

265. Poli-Neto, O. B., Meola, J., Rosa-e-Silva, J. C. & Tiezzi, D. Transcriptome meta-analysis reveals differences of immune profile between eutopic endometrium from stage I-II and III-IV endometriosis independently of hormonal milieu. *Sci. Rep.* **10**, 1–17 (2020).

266. Wan, Y. W. *et al.* Meta-Analysis of the Alzheimer's Disease Human Brain Transcriptome and Functional Dissection in Mouse Models. *Cell Rep.* **32**, (2020).

267. Leinonen, R., Sugawara, H. & Shumway, M. The sequence read archive. *Nucleic Acids Res.* **39**, 2010–2012 (2011).

268. Page, M. J. *et al.* The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **372**, (2021).

269. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, W537–W544 (2018).

270. Andrews, S. FastQC: A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

271. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

272. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

273. Howe, K. L. *et al.* Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).

274. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

275. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

276. Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual.* (Scotts Valley, CA: CreateSpace, 2009).

277. {The pandas development team}. *pandas-dev/pandas: Pandas.* (Zenodo, 2020). doi:10.5281/zenodo.3509134.

278. {Anaconda Software Distribution}. anaconda. *Anaconda Doc.* (2020).

279. Kluyver, T. *et al.* Jupyter Notebooks—a publishing format for reproducible computational workflows. *Position. Power Acad. Publ. Play. Agents Agendas - Proc. 20th Int. Conf. Electron. Publ. ELPUB 2016* 87–90 (2016) doi:10.3233/978-1-61499-649-1-87.

280. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).

281. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

282. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

283. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

284. Bedre, R. *reneshbedre/bioinfokit: Bioinformatics data analysis and visualization toolkit*. (Zenodo, 2020).

285. {R Core Team}. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2022).

286. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genomics Bioinforma.* **2**, 1–10 (2020).

287. Chen, Y., McCarthy, D., Ritchie, M., Robinson, M. & Smyth, G. EdgeR: differential analysis of sequence read count data. *R Packag.* 1–119 (2008).

288. Raudvere, U. *et al.* G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).

289. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).

290. Corti, S. *et al.* Genetic correction of human induced pluripotent stem cells from patients with spinal muscular atrophy. *Sci. Transl. Med.* **4**, (2012).

291. Boyd, P. J. *et al.* Bioenergetic status modulates motor neuron vulnerability and pathogenesis in a zebrafish model of spinal muscular atrophy. *PLOS Genet.* **13**, e1006744 (2017).

292. Hauser, S. *et al.* Comparative Transcriptional Profiling of Motor Neuron Disorder-Associated Genes in Various Human Cell Culture Models. *Front. Cell Dev. Biol.* **8**, 1–13 (2020).

293. Rizzo, F. *et al.* Key role of SMN/SYNCRIP and RNA-Motif 7 in spinal muscular atrophy: RNA-Seq and motif analysis of human motor neurons. *Brain* **142**, 1–19 (2019).

294. Corti, S. *et al.* Neural stem cell transplantation can ameliorate the phenotype of a mouse model of spinal muscular atrophy. *J. Clin. Invest.* **118**, 3316–3330 (2008).

295. Nizzardo, M. *et al.* Beta-lactam antibiotic offers neuroprotection in a spinal muscular atrophy model by multiple mechanisms. *Exp. Neurol.* **229**, 214–225 (2011).

296. Meyer, K. *et al.* Rescue of a severe mouse model for spinal muscular atrophy by U7 snRNA-mediated splicing modulation. *Hum. Mol. Genet.* **18**, 546–555 (2009).

297. Olaso, R. *et al.* Activation of RNA metabolism-related genes in mouse but not human tissues deficient in SMN. *Physiol. Genomics* **24**, 97–104 (2006).

298. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).

299. Tawfik, G. M. *et al.* A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Trop. Med. Health* **47**, 1–9 (2019).

300. Rizzo, F. *et al.* Genome-wide RNA-seq of iPSC-derived motor neurons indicates selective cytoskeletal perturbation in Brown-Vialetto disease that is partially rescued by riboflavin. *Sci. Rep.* **7**, 1–13 (2017).

301. Rizzo, F. *et al.* Selective mitochondrial depletion, apoptosis resistance, and increased mitophagy in human Charcot-Marie-Tooth 2A motor neurons. *Hum. Mol. Genet.* **25**, 4266–4281 (2016).

302. Singh, N. N., Hoffman, S., Reddi, P. P. & Singh, R. N. Spinal muscular atrophy: Broad disease spectrum and sex-specific phenotypes. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1867**, 166063 (2021).

303. Noori, A., Mezlini, A. M., Hyman, B. T., Serrano-Pozo, A. & Das, S. Systematic review and meta-analysis of human transcriptomics reveals neuroinflammation, deficient energy metabolism, and proteostasis failure across neurodegeneration. *Neurobiol. Dis.* **149**, 105225 (2021).

304. Patel, H., Dobson, R. J. B. & Newhouse, S. J. A meta-Analysis of Alzheimer's disease brain

transcriptomic data. *J. Alzheimer's Dis.* **68**, 1635–1656 (2019).

305. Kelly, J., Moyeed, R., Carroll, C., Albani, D. & Li, X. Gene expression meta-analysis of Parkinson's disease and its relationship with Alzheimer's disease. *Mol. Brain* **12**, 1–10 (2019).

306. Bottero, V., Alrafati, F., Santiago, J. A. & Potashkin, J. A. Transcriptomic and Network Meta-Analysis of Frontotemporal Dementias. *Front. Mol. Neurosci.* **14**, 1–13 (2021).

307. Seefelder, M. & Kochanek, S. A meta-analysis of transcriptomic profiles of Huntington's disease patients. *PLoS One* **16**, 1–28 (2021).

308. Deane, C. S. *et al.* Transcriptomic meta-analysis of disuse muscle atrophy vs. resistance exercise-induced hypertrophy in young and older humans. *J. Cachexia. Sarcopenia Muscle* **12**, 629–645 (2021).

309. Calura, E. *et al.* Meta-analysis of expression signatures of muscle atrophy: Gene interaction networks in early and late stages. *BMC Genomics* **9**, 1–20 (2008).

310. Walsh, C. J. *et al.* Comprehensive multi-cohort transcriptional meta-analysis of muscle diseases identifies a signature of disease severity. *Sci. Rep.* **12**, 1–14 (2022).

311. Hunt, L. C. *et al.* Integrated genomic and proteomic analyses identify stimulus-dependent molecular changes associated with distinct modes of skeletal muscle atrophy. *Cell Rep.* **37**, 109971 (2021).

312. Bazile, X. J. *et al.* Molecular signatures of muscle growth and composition deciphered by the meta-analysis of age-related public transcriptomics data. *Physiol. Genomics* **52**, 322–332 (2020).

313. Meijboom, K. E. *et al.* Combining multiomics and drug perturbation profiles to identify muscle-specific treatments for spinal muscular atrophy. **6**, (2021).

314. Pyrc, J. J., Moberg, K. H. & Hall, D. J. Isolation of a novel cDNA encoding a zinc-finger protein that binds to two sites within the c-myc promoter. *Biochemistry* **31**, 4102–4110 (1992).

315. Xiao, T., Li, X. & Felsenfeld, G. The Myc-associated zinc finger protein (MAZ) works together with CTCF to control cohesin positioning and genome organization. *Proc. Natl.*

Acad. Sci. U. S. A. **118**, 1–10 (2021).

316. MacKay, D. R., Hu, M., Li, B., Rhéaume, C. & Dai, X. The mouse Ovol2 gene is required for cranial neural tube development. *Dev. Biol.* **291**, 38–52 (2006).

317. Zhang, T. *et al.* The zinc finger transcription factor Ovol2 acts downstream of the bone morphogenetic protein pathway to regulate the cell fate decision between neuroectoderm and mesendoderm. *J. Biol. Chem.* **288**, 6166–6177 (2013).

318. Unezaki, S. *et al.* Characterization of the isoforms of MOVO zinc finger protein, a mouse homologue of Drosophila Ovo, as transcription factors. *Gene* **336**, 47–58 (2004).

319. Lee, M. S. *et al.* Neurotoxicity induces cleavage of p35 to p25 by calpain. *Nature* **405**, 360–364 (2000).

320. Jablonka, S., Beck, M., Lechner, B. D., Mayer, C. & Sendtner, M. Defective Ca2+ channel clustering in axon terminals disturbs excitability in motoneurons in spinal muscular atrophy. *J. Cell Biol.* **179**, 139–149 (2007).

321. Lopez-Manzaneda, M., Franco-Espin, J., Tejero, R., Cano, R. & Tabares, L. Calcium is reduced in presynaptic mitochondria of motor nerve terminals during neurotransmission in SMA mice. *Hum. Mol. Genet.* **30**, 629–643 (2021).

322. Khayrullina, G. *et al.* SMN-deficiency disrupts SERCA2 expression and intracellular Ca2+ signaling in cardiomyocytes from SMA mice and patient-derived iPSCs. *Skelet. Muscle* **10**, 1–13 (2020).

323. Choi, K. *et al.* Regulation of survival motor neuron gene expression by calcium signaling. *Int. J. Mol. Sci.* **22**, (2021).

324. Bezprozvanny, I. B. Calcium Signaling and Neurodegeneration. *Acta Naturae* **2**, 72–80 (2010).

325. Calvo-Rodriguez, M., Kharitonova, E. K. & Bacskai, B. J. Therapeutic Strategies to Target Calcium Dysregulation in Alzheimer's Disease. *Cells* **9**, (2020).

326. Phillips, J. E. Inhaled Phosphodiesterase 4 (PDE4) Inhibitors for Inflammatory Respiratory

Diseases. *Front. Pharmacol.* **11**, 1–7 (2020).

327. Somers, E. *et al.* Vascular Defects and Spinal Cord Hypoxia in Spinal Muscular Atrophy. *Ann Neurol* **79**, 217–230 (2016).

328. Wong, A. D. *et al.* The blood-brain barrier: an engineering perspective. *Front Neuroeng* **6**, 7 (2013).

329. Weaver, J. J. *et al.* Transforaminal intrathecal delivery of nusinersen using cone-beam computed tomography for children with spinal muscular atrophy and extensive surgical instrumentation: early results of technical success and safety. *Pediatr. Radiol.* **48**, 392–397 (2018).

330. Osmanovic, A. *et al.* Treatment satisfaction in 5q-spinal muscular atrophy under nusinersen therapy. *Ther. Adv. Neurol. Disord.* **14**, 1–19 (2021).

331. Mitchell, D. J., Kim, D. T., Steinman, L., Fathman, C. G. & Rothbard, J. B. Polyarginine enters cells more efficiently than other polycationic homopolymers. *J. Pept. Res. Off. J. Am. Pept. Soc.* **56**, 318–325 (2000).

332. Zaro, J. L. & Shen, W.-C. Cationic and amphipathic cell-penetrating peptides (CPPs): Their structures and in vivo studies in drug delivery. *Front. Chem. Sci. Eng.* **9**, 407–427 (2015).

333. Ezzat, K. *et al.* Scavenger receptor-mediated uptake of cell-penetrating peptide nanocomplexes with oligonucleotides. *FASEB J* **26**, 1172–1180 (2012).

334. Ciobanasu, C., Siebrasse, J. P. & Kubitscheck, U. Cell-penetrating HIV1 TAT peptides can generate pores in model membranes. *Biophys J* **99**, 153–162 (2010).

335. Perez, F. *et al.* Antennapedia homeobox as a signal for the cellular internalization and nuclear addressing of a small exogenous peptide. *J. Cell Sci.* **102 (Pt 4**, 717–722 (1992).

336. Betts, C. A. *et al.* Pip6-PMO, A New Generation of Peptide-oligonucleotide Conjugates With Improved Cardiac Exon Skipping Activity for DMD Treatment. *Mol Ther Nucleic Acids* **1**, e38 (2012).

337. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small

molecules, genes, and disease. *Science (80-.).* **313**, 1929–1935 (2006).

338. Hammond, S. M. *et al.* Systemic peptide-mediated oligonucleotide therapy improves long-term survival in spinal muscular atrophy. *Proc Natl Acad Sci U S A* **113**, 10962–10967 (2016).
339. Wood, M. *et al.* Cell-penetrating peptides. (2018).
340. Schreml, J. *et al.* Severe SMA mice show organ impairment that cannot be rescued by therapy with the HDACi JNJ-26481585. *Eur. J. Hum. Genet.* **21**, 643–652 (2013).
341. Tunster, S. J. Genetic sex determination of mice by simplex PCR. *Biol. Sex Differ.* **8**, 6–9 (2017).
342. {RStudioTeam}. RStudio: Integrated Development for R. *RStudio, PBC* (2020).
343. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
344. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363–2367 (2010).
345. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
346. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
347. Xiao, Y. *et al.* A novel significance score for gene selection and ranking. *Bioinformatics* **30**, 801–807 (2014).
348. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models. *Genome Res.* **13**, 2498–2504 (2003).
349. Doncheva, N. T., Morris, J. H., Gorodkin, J. & Jensen, L. J. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J. Proteome Res.* **18**, 623–632 (2019).
350. Maere, S., Heymans, K. & Kuiper, M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* **21**, 3448–3449 (2005).

351. Scardoni, G., Petterlini, M. & Laudanna, C. Analyzing biological network parameters with CentiScaPe. *Bioinformatics* **25**, 2857–2859 (2009).

352. Oprişoreanu, A. M. *et al.* Interaction of Axonal Chondrolectin with Collagen XIXa1 Is Necessary for Precise Neuromuscular Junction Formation. *Cell Rep.* **29**, 1082-1098.e10 (2019).

353. Charan, J. & Kantharia, N. How to calculate sample size in animal studies? *J. Pharmacol. Pharmacother.* **4**, 303–306 (2013).

354. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).

355. Ching, T., Huang, S. & Garmire, L. X. Power analysis and sample size estimation for RNA-Seq differential expression. *Rna* **20**, 1684–1696 (2014).

356. Su, K., Wu, Z. & Wu, H. Simulation, power evaluation and sample size recommendation for single-cell RNA-seq. *Bioinformatics* **36**, 4860–4868 (2020).

357. Saunders, N. R., Liddelow, S. A. & Dziegielewska, K. M. Barrier mechanisms in the developing brain. *Front. Pharmacol.* **3 MAR**, 1–18 (2012).

358. Cornford, E. M. & Cornford, M. E. Nutrient transport and the blood-brain barrier in developing animals. *Fed. Proc.* **45**, 2065–2072 (1986).

359. Xu, J. & Ling, E. A. Studies of the ultrastructure and permeability of the blood-brain barrier in the developing corpus callosum in postnatal rat brain using electron dense tracers. *J. Anat.* **184 (Pt 2**, 227–37 (1994).

360. Stolp, H. B. *et al.* Breakdown of the blood-brain barrier to proteins in white matter of the developing brain following systemic inflammation. *Cell Tissue Res.* **320**, 369–378 (2005).

361. Stolp, H. B., Dziegielewska, K. M., Ek, C. J., Potter, A. M. & Saunders, N. R. Long-term changes in blood–brain barrier permeability and white matter following prolonged systemic inflammation in early development in the rat. *Eur. J. Neurosci.* **22**, 2805–2816 (2005).

362. Zhao, Q. *et al.* Prenatal disruption of blood-brain barrier formation via cyclooxygenase activation leads to lifelong brain inflammation. *Proc. Natl. Acad. Sci. U. S. A.* **119**, 1–12 (2022).

363. Gröntoft, O. Intracranial haemorrhage and blood-brain barrier problems in the new-born; a pathologico-anatomical and experimental investigation. *Acta Pathol. Microbiol. Scand. Suppl.* **100**, 8–109 (1954).

364. Johansson, P. A., Dziegielewska, K. M., Liddelow, S. A. & Saunders, N. R. The blood–CSF barrier explained: when development is not immaturity. *BioEssays* **30**, 237–248 (2008).

365. Liddelow, S. A. *et al.* Cellular transfer of macromolecules across the developing choroid plexus of Monodelphis domestica. *Eur. J. Neurosci.* **29**, 253–266 (2009).

366. Liddelow, S. A. *et al.* Molecular characterisation of transport mechanisms at the developing mouse blood-CSF interface: A transcriptome approach. *PLoS One* **7**, (2012).

367. Liddelow, S. A. Development of the choroid plexus and blood-CSF barrier. *Front. Neurosci.* **9**, 1–13 (2015).

368. Semple, B. D., Blomgren, K., Gimlin, K., Ferriero, D. M. & Noble-Haeusslein, L. J. Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. *Prog. Neurobiol.* **106–107**, 1–16 (2013).

369. Coelho-Santos, V. & Shih, A. Y. Postnatal development of cerebrovascular structure and the neurogliovascular unit. *Wiley Interdiscip. Rev. Dev. Biol.* **9**, 1–20 (2020).

370. Jang, S. I. *et al.* Dual function of miR-1248 links interferon induction and calcium signaling defects in Sjögren's syndrome. *EBioMedicine* **48**, 526–538 (2019).

371. Singh, N. N., Ottesen, E. W. & Singh, R. N. A survey of transcripts generated by spinal muscular atrophy genes. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1863**, 194562 (2020).

372. Campioni, M. *et al.* The serine protease HtrA1 specifically interacts and degrades the tuberous sclerosis complex 2 protein. *Mol. Cancer Res.* **8**, 1248–1260 (2010).

373. Wildonger, J., Jan, L. Y. & Jan, Y. N. The Tscl-Tsc2 complex influences neuronal polarity

by modulating TORC1 activity and SAD levels. *Genes Dev.* **22**, 2447–2453 (2008).

374. Chen, J. *et al.* BMP-responsive protease HtrA1 is differentially expressed in astrocytes and regulates astrocytic development and injury response. *J. Neurosci.* **38**, 3840–3857 (2018).

375. Kannan, A., Jiang, X., He, L., Ahmad, S. & Gangwani, L. ZPR1 prevents R-loop accumulation, upregulates SMN2 expression and rescues spinal muscular atrophy. *Brain* 69–93 (2019) doi:10.1093/brain/awz373.

376. Burns, W. *et al.* Syndromic neurodevelopmental disorder associated with de novo variants in DDX23. *Am. J. Med. Genet. Part A* **185**, 2863–2872 (2021).

377. Halama, N., Grauling-Halama, S. A. & Jäger, D. Identification and characterization of the human StARD9 gene in the LGMD2A-region on chromosome 15q15 by in silico methods. *Int. J. Mol. Med.* **18**, 653–656 (2006).

378. Torres, J. Z. *et al.* The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly. *Cell* **147**, 1309–1323 (2011).

379. Srivastava, S. & Panda, D. A centrosomal protein STARD9 promotes microtubule stability and regulates spindle microtubule dynamics. *Cell Cycle* **17**, 2052–2068 (2018).

380. Okamoto, N. *et al.* A novel genetic syndrome with STARD9 mutation and abnormal spindle morphology. *Am. J. Med. Genet. Part A* **173**, 2690–2696 (2017).

381. Shabanpoor, F. & Gait, M. J. Development of a general methodology for labelling peptide-morpholino oligonucleotide conjugates using alkyne-azide click chemistry. *Chem Commun* **49**, 10260–10262 (2013).

382. Ando, S. *et al.* Discovery of a CNS penetrant small molecule SMN2 splicing modulator with improved tolerability for spinal muscular atrophy. *Sci. Rep.* **10**, 17472 (2020).

383. Reilly, A. *et al.* Central and peripheral delivered AAV9-SMN are both efficient but target different pathomechanisms in a mouse model of spinal muscular atrophy. *Gene Ther.* (2022) doi:10.1038/s41434-022-00338-1.

384. Shum, E. Y., Espinoza, J. L., Ramaiah, M. & Wilkinson, M. F. Identification of novel post-

transcriptional features in olfactory receptor family mRNAs. *Nucleic Acids Res.* **43**, 9314–9326 (2015).

385. Chamero, P. *et al.* Identification of protein pheromones that promote aggressive behaviour. *Nature* **450**, 899–902 (2007).

386. Charkoftaki, G. *et al.* Update on the human and mouse lipocalin (LCN) gene family, including evidence the mouse Mup cluster is result of an 'evolutionary bloom'. *Hum. Genomics* **13**, 11 (2019).

387. Coca-Prados, M. & Hsu, M.-T. Electron microscopic evidence for circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* **280**, 0–1 (1979).

388. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).

389. Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141–157 (2013).

390. Jeck, W. R. & Sharpless, N. E. Detecting and characterizing circular RNAs. *Nat Biotechnol* **32**, 453–461 (2014).

391. Zhang, P. *et al.* Comprehensive identification of alternative back-splicing in human tissue transcriptomes. *Nucleic Acids Res.* **48**, 1779–1789 (2020).

392. Westholm, J. O. *et al.* Genome-wide Analysis of Drosophila Circular RNAs Reveals Their Structural and Sequence Properties and Age-Dependent Neural Accumulation. *Cell Rep.* **9**, 1966–1980 (2014).

393. Gruner, H., Cortés-López, M., Cooper, D. A., Bauer, M. & Miura, P. CircRNA accumulation in the aging mouse brain. *Sci. Rep.* **6**, 1–14 (2016).

394. Ji, P. *et al.* Expanded Expression Landscape and Prioritization of Circular RNAs in Mammals. *Cell Rep.* **26**, 3444-3460.e5 (2019).

395. Conn, S. J. *et al.* The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**, 1125–1134 (2015).

396. Errichelli, L. *et al.* FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat. Commun.* **8**, 1–11 (2017).

397. Pagliarini, V. *et al.* Sam68 binds Alu-rich introns in SMN and promotes pre-mRNA circularization. *Nucleic Acid Res.* **48**, 633–645 (2020).

398. Gruhl, F., Janich, P., Kaessmann, H. & Gatfield, D. Circular RNA repertoires are associated with evolutionarily young transposable elements. *Elife* **10**, 1–33 (2021).

399. Huang, J. L. *et al.* Comprehensive analysis of differentially expressed profiles of Alzheimer's disease associated circular rnas in an Alzheimer's disease mouse model. *Aging (Albany. NY).* **10**, 253–265 (2018).

400. Chwalenia, K., Facemire, L. & Li, H. Chimeric RNAs in cancer and normal physiology. *Wiley Interdiscip. Rev. RNA* **8**, (2017).

401. Chen, W. & Schuman, E. Circular RNAs in Brain and Other Tissues: A Functional Enigma. *Trends Neurosci.* **39**, 597–604 (2016).

402. Aufiero, S. *et al.* Cardiac circRNAs arise mainly from constitutive exons rather than alternatively spliced exons. *Rna* **24**, 815–827 (2018).

403. Aufiero, S., Reckman, Y. J., Pinto, Y. M. & Creemers, E. E. Circular RNAs open a new chapter in cardiovascular biology. *Nat. Rev. Cardiol.* (2019) doi:10.1038/s41569-019-0185-2.

404. Wang, M., Hou, J., Müller-McNicoll, M., Chen, W. & Schuman, E. M. Long and Repeat-Rich Intronic Sequences Favor Circular RNA Formation under Conditions of Reduced Spliceosome Activity. *iScience* **20**, 237–247 (2019).

405. Zhang, Y. *et al.* The Biogenesis of Nascent Circular RNAs. *Cell Rep.* **15**, 611–624 (2016).

406. Li, Z. *et al.* Tumor-released exosomal circular RNA PDE8A promotes invasive growth via the miR-338/MACC1/MET pathway in pancreatic cancer. *Cancer Lett.* **432**, 237–250 (2018).

407. Reddy, A. S. *et al.* A Comprehensive Analysis of Cell Type–Specific Nuclear RNA From Neurons and Glia of the Brain. *Biol. Psychiatry* **81**, 252–264 (2017).

408. Zhou, C. *et al.* Genome-Wide Maps of m6A circRNAs Identify Widespread and Cell-Type-Specific Methylation Patterns that Are Distinct from mRNAs. *Cell Rep.* **20**, 2262–2276 (2017).

409. Sekar, S. & Liang, W. S. Circular RNA expression and function in the brain. *Non-coding RNA Res.* **4**, 23–29 (2019).

410. Rybak-Wolf, A. *et al.* Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol. Cell* **58**, 870–885 (2014).

411. Cortés-López, M. *et al.* Global accumulation of circRNAs during aging in Caenorhabditis elegans. *BMC Genomics* **19**, 1–12 (2018).

412. Zhang, S. B. *et al.* CircAnks1a in the spinal cord regulates hypersensitivity in a rodent model of neuropathic pain. *Nat. Commun.* **10**, (2019).

413. Bian, A. *et al.* Circular RNA Complement Factor H (CFH) Promotes Glioma Progression by Sponging miR-149 and Regulating AKT1. *Med. Sci. Monit.* **24**, 5704–5712 (2018).

414. Meganck, R. M. *et al.* Tissue-Dependent Expression and Translation of Circular RNAs with Recombinant AAV Vectors In Vivo. *Mol. Ther. - Nucleic Acids* **13**, 89–98 (2018).

415. Han, B. *et al.* Novel insight into circular RNA HECTD1 in astrocyte activation via autophagy by targeting MIR142-TIPARP: implications for cerebral ischemic stroke. *Autophagy* **14**, 1164–1184 (2018).

416. Kleaveland, B., Shi, C. Y., Stefano, J. & Bartel, D. P. A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. *Cell* **174**, 350-362.e17 (2018).

417. Sekar, S., Cuyugan, L., Adkins, J., Geiger, P. & Liang, W. S. Circular RNA expression and regulatory network prediction in posterior cingulate astrocytes in elderly subjects. *BMC Genomics* **19**, 1–11 (2018).

418. Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388 (2013).

419. Chen, Y. G. *et al.* Sensing Self and Foreign Circular RNAs by Intron Identity. *Mol. Cell* **67**,

228-238.e5 (2017).

420. Bose, R. & Ain, R. Regulation of Transcription by Circular RNAs. *Adv. Exp. Med. Biol.* **1087**, 81–94 (2018).

421. Liang, D. *et al.* The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. *Mol. Cell* **68**, 940-954.e3 (2017).

422. Tatomer, D. C. & Wilusz, J. E. An Unchartered Journey for Ribosomes: Circumnavigating Circular RNAs to Produce Proteins. *Mol. Cell* **66**, 1–2 (2017).

423. Legnini, I. *et al.* Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cell* **66**, 22-37.e9 (2017).

424. Ottesen, E. W., Seo, J., Singh, N. N. & Singh, R. N. A multilayered control of the human Survival Motor Neuron gene expression by Alu elements. *Front. Microbiol.* **8**, 1–12 (2017).

425. Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L. & Brown, P. O. Cell-Type Specific Features of Circular RNA Expression. *PLoS Genet.* **9**, (2013).

426. Li, D. K., Tisdale, S., Lotti, F. & Pellizzoni, L. SMN control of RNP assembly: from post-transcriptional gene regulation to motor neuron disease. *Semin Cell Dev Biol* **32**, 22–29 (2014).

427. Guest, P. C. *et al.* MK-801 treatment affects glycolysis in oligodendrocytes more than in astrocytes and neuronal cells: Insights for schizophrenia. *Front. Cell. Neurosci.* **9**, 1–10 (2015).

428. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, (2011).

429. Mathews, D. H., Moss, W. N. & Turner, D. H. Folding and Finding RNA Secondary Structure. *Cold Spring Harb Perspect Biol* **2**, 1–15 (2010).

430. Turner, D. H. & Mathews, D. H. NNDB: The nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.* **38**, 2009–2011 (2009).

431. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

432. Dudekula, D. B. *et al.* Circinteractome: A web tool for exploring circular RNAs and their

interacting proteins and microRNAs. *RNA Biol.* **13**, 34–42 (2016).

433. Pan, X., Fang, Y., Li, X., Yang, Y. & Shen, H. Bin. RBPsuite: RNA-protein binding sites prediction suite based on deep learning. *BMC Genomics* **21**, 1–8 (2020).

434. Chen, Y. & Wang, X. MiRDB: An online database for prediction of functional microRNA targets. *Nucleic Acids Res.* **48**, D127–D131 (2020).

435. Zhong, S. *et al.* Identification of internal control genes for circular RNAs. *Biotechnol. Lett.* **0**, (2019).

436. Lasda, E. & Parker, R. Circular RNAs: diversity of form and function. *RNA* **20**, 1829–1842 (2014).

437. Piwecka, M. *et al.* Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science (80-.).* **357**, (2017).

438. Sang, M. *et al.* Circular RNA ciRS-7 accelerates ESCC progression through acting as a miR-876-5p sponge to enhance MAGE-A family expression. *Cancer Lett.* **426**, 37–46 (2018).

439. Barrett, S., Parker, K., Horn, C., Mata, M. & Salzman, J. ciRS-7 exonic sequence is embedded in a long non-coding RNA locus. *PLoS Genet.* 169508 (2017) doi:10.1101/169508.

440. Hanniford, D. *et al.* *Epigenetic Silencing of CDR1as Drives IGF2BP3-Mediated Melanoma Invasion and Metastasis*. *Cancer Cell* vol. 37 (2020).

441. Zuber, J. & Mathews, D. H. Estimating Uncertainty in Predicted Folding Free Energy Changes of RNA Secondary Structures. *RNA* **25**, 747–754 (2019).

442. Li, Z. *et al.* Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* **22**, 256–264 (2015).

443. Ashwal-Fluss, R. *et al.* CircRNA Biogenesis competes with Pre-mRNA splicing. *Mol. Cell* **56**, 55–66 (2014).

444. Du, W. W. *et al.* Identifying and characterizing circRNA-protein interaction. *Theranostics* **7**, 4183–4191 (2017).

445. Sticht, C., De La Torre, C., Parveen, A. & Gretz, N. Mirwalk: An online resource for prediction of microrna binding sites. *PLoS One* **13**, 1–6 (2018).

446. ThermoFisher Scientific. Ethidium Bromide (EtBr) Dye for DNA and RNA Detection. https://www.thermofisher.com/dk/en/home/life-science/dna-rna-purification-analysis/nucleic-acid-gel-electrophoresis/dna-stains/etbr.html (2015).

447. You, X. *et al.* Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat. Neurosci.* **18**, 603–610 (2015).

448. Xin, R. *et al.* isoCirc catalogs full-length circular RNA isoforms in human transcriptomes. *Nat. Commun.* **12**, 1–11 (2021).

449. Ma, H., Li, M., Jia, Z., Chen, X. & Bu, N. miR-876-3p suppresses the progression of colon cancer and correlates the prognosis of patients. *Exp. Mol. Pathol.* **122**, 104682 (2021).

450. Guo, J. *et al.* The HOXC-AS2/miR-876-5p/HKDC1 axis regulates endometrial cancer progression in a high glucose-related tumor microenvironment. *Cancer Sci.* **113**, 2297–2310 (2022).

451. Rajan, S. *et al.* miR-876-3p regulates glucose homeostasis and insulin sensitivity by targeting adiponectin. *J. Endocrinol.* **239**, 1–17 (2018).

452. Yan, S. *et al.* CircTP53 promotes colorectal cancer by acting as a miR-876-3p sponge to increase cyclin-dependent kinase-like 3 expression. *Cell. Signal.* **78**, 109845 (2021).

453. Ottesen, E. W. & Singh, R. N. Characteristics of circular RNAs generated by human Survival Motor Neuron genes. *Cell. Signal.* **73**, 109696 (2020).

454. Vernes, S. C. *et al.* FOXP2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet.* **7**, (2011).

455. Sia, G., Clem, R. & Huganir, R. The human language and epilepsy associated gene SRPX2 regulates synapse formation and vocalization in mice. *Science (80-.).* **342**, 987–991 (2013).

456. Patel, T. *et al.* Transcriptional dynamics of murine motor neuron maturation in vivo and in vitro. *Nat. Commun.* **13**, 5427 (2022).

457. Huang, H. *et al.* Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat Cell Biol* **20**, 285–295 (2018).

458. Chang, M. *et al.* LIN 28A loss of function is associated with Parkinson's disease pathogenesis . *EMBO J.* **38**, 1–17 (2019).

459. Polesskaya, A. *et al.* Lin-28 binds IGF-2 mRNA and participates in skeletal myogenesis by increasing translation efficiency. *Genes Dev.* **21**, 1125–1138 (2007).

460. Sunnie Thompson. So you want to know if your message has an IRES? *Wiley Interdiscip. Rev. RNA* **3**, 697–705 (2012).

461. Zheng, Y., Ji, P., Chen, S., Hou, L. & Zhao, F. Reconstruction of full-length circular RNAs enables isoform-level quantification. *Genome Med.* **11**, 1–20 (2019).

462. Fanzani, A., Conraads, V. M., Penna, F. & Martinet, W. Molecular and cellular mechanisms of skeletal muscle atrophy: An update. *J. Cachexia. Sarcopenia Muscle* **3**, 163–179 (2012).

463. Evans, W. J. Skeletal muscle loss: Cachexia, sarcopenia, and inactivity. *Am. J. Clin. Nutr.* **91**, 1123–1127 (2010).

464. Zizzo, J. Muscle Atrophy Classification: The Need for a Pathway-Driven Approach. *Curr. Pharm. Des.* **27**, 3012–3019 (2021).

465. Bozzetti, F. & Mariani, L. Defining and classifying cancer cachexia: A proposal by the SCRINIO Working Group. *J. Parenter. Enter. Nutr.* **33**, 361–367 (2009).

466. Fearon, K. *et al.* Definition and classification of cancer cachexia: An international consensus. *Lancet Oncol.* **12**, 489–495 (2011).

467. Cruz-Jentoft, A. J. *et al.* Sarcopenia: Revised European consensus on definition and diagnosis. *Age Ageing* **48**, 16–31 (2019).

468. Ni, J. & Zhang, L. Cancer cachexia: Definition, staging, and emerging treatments. *Cancer Manag. Res.* **12**, 5597–5605 (2020).

469. Gao, Y., Arfat, Y., Wang, H. & Goswami, N. Muscle atrophy induced by mechanical unloading: Mechanisms and potential countermeasures. *Front. Physiol.* **9**, (2018).

470. Bodine, S. C. Disuse-induced muscle wasting. *Int J Biochem Cell Biol* **45**, (2013).

471. McKinnell, I. W. & Rudnicki, M. A. Molecular mechanisms of muscle atrophy. *Cell* **119**, 907–910 (2004).

472. Bonaldo, P. & Sandri, M. Cellular and molecular mechanisms of muscle atrophy. *DMM Dis. Model. Mech.* **6**, 25–39 (2013).

473. Schiaffino, S., Reggiani, C., Akimoto, T. & Blaauw, B. Molecular Mechanisms of Skeletal Muscle Hypertrophy. *J. Neuromuscul. Dis.* **8**, 169–183 (2021).

474. Dimitriadis, G., Mitron, P., Lambadiari, V., Maratou, E. & Raptis, S. A. Insulin effects in muscle and adipose tissue. *Diabetes Res. Clin. Pract.* **93**, 52–59 (2011).

475. Solomon, A. M. & Bouloux, P. M. G. Modifying muscle mass - The endocrine perspective. *J. Endocrinol.* **191**, 349–360 (2006).

476. Gharahdaghi, N. *et al.* Links Between Testosterone, Oestrogen, and the Growth Hormone/Insulin-Like Growth Factor Axis and Resistance Exercise Muscle Adaptations. *Front. Physiol.* **11**, 1–12 (2021).

477. Miyake, M. *et al.* Skeletal muscle-specific eukaryotic translation initiation factor 2α phosphorylation controls amino acid metabolism and fibroblast growth factor 21-mediated non-cell-autonomous energy metabolism. *FASEB J.* **30**, 798–812 (2016).

478. Roux, P. P. & Topisirovic, I. Signaling Pathways Involved in the Regulation of mRNA Translation. *Mol. Cell. Biol.* **38**, 1–26 (2018).

479. Egerman, M. A. & Glass, D. J. Signaling pathways controlling skeletal muscle mass. *Crit. Rev. Biochem. Mol. Biol.* **49**, 59–68 (2014).

480. Ahmad, S. S., Ahmad, K., Lee, E. J., Lee, Y. H. & Choi, I. Implications of Insulin-Like Growth Factor-1 in Skeletal Muscle and Various Diseases. *Cells* **9**, 1–15 (2020).

481. Kunkel, S. D. *et al.* mRNA expression signatures of human skeletal muscle atrophy identify a natural compound that increases muscle mass. *Cell Metab* **13**, 627–638 (2011).

482. Black, B. L. & Olson, E. N. Transcriptional control of muscle development by myocyte

enhancer factor-2 (MEF2) proteins. *Annu. Rev. Cell Dev. Biol.* **14**, 167–196 (1998).

483. Yu, S. J., Zheng, L., Trinh, D. K., Asa, S. L. & Ezzat, S. Distinct transcriptional control and action of fibroblast growth factor receptor 4 in differentiating skeletal muscle cells. *Lab. Investig.* **84**, 1571–1580 (2004).

484. Hornberger, T. A. & Esser, K. A. Mechanotransduction and the regulation of protein synthesis in skeletal muscle. *Proc. Nutr. Soc.* **63**, 331–335 (2004).

485. Tidball, J. G. Invited review: Mechanical signal transduction in skeletal muscle growth and adaptation. *J. Appl. Physiol.* **98**, 1900–1908 (2005).

486. Snijders, T. *et al.* Satellite cells in human skeletal muscle plasticity. *Front. Physiol.* **6**, 1–21 (2015).

487. Relaix, F. *et al.* Perspectives on skeletal muscle stem cells. *Nat. Commun.* **12**, 1–11 (2021).

488. Yin, H., Price, F. & Rudnicki, M. A. Satellite cells and the muscle stem cell niche. *Physiol. Rev.* **93**, 23–67 (2013).

489. Attaix, D. *et al.* The ubiquitin-proteasome system and skeletal muscle wasting. *Essays Biochem.* 173–186 (2005).

490. Kitajima, Y., Yoshioka, K. & Suzuki, N. The ubiquitin-proteasome system in regulation of the skeletal muscle homeostasis and atrophy: From basic science to disorders. *J. Physiol. Sci.* **70**, (2020).

491. Murton, A. J., Constantin, D. & Greenhaff, P. L. The involvement of the ubiquitin proteasome system in human skeletal muscle remodelling and atrophy. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1782**, 730–743 (2008).

492. Xia, Q. *et al.* The Role of Autophagy in Skeletal Muscle Diseases. *Front. Physiol.* **12**, (2021).

493. Sandri, M. Protein breakdown in muscle wasting: Role of autophagy-lysosome and ubiquitin-proteasome. *Int. J. Biochem. Cell Biol.* **45**, 2121–2129 (2013).

494. Ono, Y. & Sorimachi, H. Calpains - An elaborate proteolytic system. *Biochim. Biophys. Acta - Proteins Proteomics* **1824**, 224–236 (2012).

495. Huang, J. & Zhu, X. The molecular mechanisms of calpains action on skeletal muscle atrophy. *Physiol. Res.* **65**, 547–560 (2016).

496. Skinner, S. K. *et al.* Mitochondrial permeability transition causes mitochondrial reactive oxygen species-and caspase 3-dependent atrophy of single adult mouse skeletal muscle fibers. *Cells* **10**, (2021).

497. Hyatt, H., Deminice, R., Yoshihara, T. & Powers, S. K. Mitochondrial dysfunction induces muscle atrophy during prolonged inactivity: A review of the causes and effects. *Arch. Biochem. Biophys.* **662**, 49–60 (2019).

498. Wray, C. J., Tomkinson, B., Robb, B. W. & Hasselgren, P. O. Tripeptidyl-peptidase II expression and activity are increased in skeletal muscle during sepsis. *Biochem. Biophys. Res. Commun.* **296**, 41–47 (2002).

499. White, M. G., Stoward, P. J., Christie, K. N. & Anderson, J. M. Proteases in normal and diseased human skeletal muscle: a preliminary histochemical survey. *Histochem. J.* **17**, 819–832 (1985).

500. Taillandier, D. & Polge, C. Skeletal muscle atrogenes: From rodent models to human pathologies. *Biochimie* **166**, 251–269 (2019).

501. Rom, O. & Reznick, A. Z. The role of E3 ubiquitin-ligases MuRF-1 and MAFbx in loss of skeletal muscle mass. *Free Radic. Biol. Med.* **98**, 218–230 (2016).

502. Carter, M. E. & Brunet, A. Quick guide FOXO transcription factors. *Curr. Biol.* **17**, 113–114 (2007).

503. Tzivion, G., Dobson, M. & Ramakrishnan, G. FoxO transcription factors; Regulation by AKT and 14-3-3 proteins. *Biochim. Biophys. Acta - Mol. Cell Res.* **1813**, 1938–1945 (2011).

504. Shapiro, D. N., Sublett, J. E., Li, B., Downing, J. R. & Naeve, C. W. Fusion of PAX3 to a member of the forkhead family of transcription factors in human alveolar rhabdomyosarcoma. *Cancer Res.* **53**, 5108–5112 (1993).

505. Galili, N. *et al.* Fusion of a fork head domain gene to PAX3 in the solid tumour. *Nat. Gene*

5, 1–6 (1993).

506. Hillion, J., Le Coniat, M., Jonveaux, P., Berger, R. & Bernard, O. A. AF6q21, a novel partner of the MLL gene in t(6;11)(q21;q23), defines a forkhead transcriptional factor subfamily. *Blood* **90**, 3714–3719 (1997).

507. Parry, P., Wei, Y. & Evans, G. Cloning and characterization of the t(X;II) breakpoint from a leukemic cell line identify a new member of the forkhead gene family. *Genes, Chromosom. Cancer* **11**, 79–84 (1994).

508. Borkhardt, A. *et al.* Cloning and characterization of AFX, the gene that fuses to MLL in acute leukemias with a t(X;11)(q13;q23). *Oncogene* **14**, 195–202 (1997).

509. Jacobs, F. M. J. *et al.* FoxO6, a novel member of the FoxO class of transcription factors with distinct shuttling dynamics. *J. Biol. Chem.* **278**, 35959–35967 (2003).

510. Obsil, T. & Obsilova, V. Structural basis for DNA recognition by FOXO proteins. *Biochim. Biophys. Acta - Mol. Cell Res.* **1813**, 1946–1953 (2011).

511. Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a forkhead transcription factor. *Cell* **96**, 857–868 (1999).

512. Gilley, J., Coffer, P. J. & Ham, J. FOXO transcription factors directly activate bim gene expression and promote apoptosis in sympathetic neurons. *J. Cell Biol.* **162**, 613–622 (2003).

513. Stahl, M. *et al.* The Forkhead Transcription Factor FoxO Regulates Transcription of p27 Kip1 and Bim in Response to IL-2 . *J. Immunol.* **168**, 5024–5031 (2002).

514. Furukawa-Hibi, Y., Yoshida-Araki, K., Ohta, T., Ikeda, K. & Motoyama, N. FOXO forkhead transcription factors induce G2-M checkpoint in response to oxidative stress. *J. Biol. Chem.* **277**, 26729–26732 (2002).

515. Awad, H., Nolette, N., Hinton, M. & Dakshinamurti, S. AMPK and FoxO1 regulate catalase expression in hypoxic pulmonary arterial smooth muscle. *Pediatr. Pulmonol.* **49**, 885–897 (2014).

516. Chen, Y.-R. *et al.* Epstein-Barr Virus Latent Membrane Protein 1 Represses DNA Repair

through the PI3K/Akt/FOXO3a Pathway in Human Epithelial Cells. *J. Virol.* **82**, 8124–8137 (2008).

517. Kousteni, S. FoxO1, the transcriptional chief of staff of energy metabolism. *Bone* **50**, 437–443 (2012).

518. Ouyang, W. *et al.* Foxo proteins cooperatively control the differentiation of Foxp3 + regulatory T cells. *Nat. Immunol.* **11**, 618–627 (2010).

519. Kim, D. Y., Hwang, I., Muller, F. L. & Paik, J. H. Functional regulation of FoxO1 in neural stem cell differentiation. *Cell Death Differ.* **22**, 2034–2045 (2015).

520. Chen, J., Lu, Y., Tian, M. & Huang, Q. Molecular mechanisms of foxo1 in adipocyte differentiation. *J. Mol. Endocrinol.* **62**, R239–R253 (2019).

521. Milan, G. *et al.* Regulation of autophagy and the ubiquitin-proteasome system by the FoxO transcriptional network during muscle atrophy. *Nat. Commun.* **6**, (2015).

522. Sandri, M. *et al.* Foxo transcription factors induce the atrophy-related ubiquitin ligase atrogin-1 and cause skeletal muscle atrophy. *Cell* **117**, 399–412 (2004).

523. Kang, S. H. *et al.* Forkhead box O3 plays a role in skeletal muscle atrophy through expression of E3 ubiquitin ligases MuRF-1 and atrogin-1 in Cushing's syndrome. *Am. J. Physiol. - Endocrinol. Metab.* **312**, E495–E507 (2017).

524. Crawford, T. O. & Pardo, C. A. The neurobiology of childhood spinal muscular atrophy. *Neurobiol. Dis.* **3**, 97–110 (1996).

525. Iannaccone, S. T. *et al.* Muscle fatigue in spinal muscular atrophy. *J. Child Neurol.* **12**, 321–326 (1997).

526. Montes, J. *et al.* Leg muscle function and fatigue during walking in spinal muscular atrophy type 3. *Muscle and Nerve* **50**, 34–39 (2014).

527. Deguise, M. O. *et al.* Differential induction of muscle atrophy pathways in two mouse models of spinal muscular atrophy. *Sci. Rep.* **6**, 1–13 (2016).

528. Guettier-Sigrist, S., Coupin, G., Braun, S., Warter, J. M. & Poindron, P. Muscle could be

the therapeutic target in SMA treatment. *J. Neurosci. Res.* **53**, 663–669 (1998).

529. Rudnicki, S. A. *et al.* Reldesemtiv in Patients with Spinal Muscular Atrophy: a Phase 2 Hypothesis-Generating Study. *Neurotherapeutics* **18**, 2130 (2021).

530. Bonanno, S. *et al.* Amifampridine safety and efficacy in spinal muscular atrophy ambulatory patients: a randomized, placebo-controlled, crossover phase 2 trial. *J. Neurol.* (2022) doi:10.1007/s00415-022-11231-7.

531. Chen, L., Zhang, H., Chi, M., Yang, Q. & Guo, C. Drugs for the Treatment of Muscle Atrophy. in *Background and Management of Muscular Atrophy* (ed. Cseri, J.) (IntechOpen, 2021). doi:10.5772/intechopen.93503.

532. Wu, C. *et al.* Salidroside attenuates denervation-induced skeletal muscle atrophy through negative regulation of pro-inflammatory cytokine. *Front. Physiol.* **10**, 1–11 (2019).

533. Asami, Y., Aizawa, M., Kinoshita, M., Ishikawa, J. & Sakuma, K. Resveratrol attenuates denervation-induced muscle atrophy due to the blockade of atrogin-1 and p62 accumulation. *Int. J. Med. Sci.* **15**, 628–637 (2018).

534. Smith, R. C. & Lin, B. K. Myostatin inhibitors as therapies for muscle wasting associated with cancer and other disorders. *Curr. Opin. Support. Palliat. Care* **7**, 352–360 (2013).

535. Long, K. K. *et al.* Specific inhibition of myostatin activation is beneficial in mouse models of SMA therapy. *Hum Mol Genet* **00**, 1–14 (2018).

536. Bodine, S. C. *et al.* Identification of ubiquitin ligases required for skeletal Muscle Atrophy. *Science (80-.).* **294**, 1704–1708 (2001).

537. Witt, C. C. *et al.* Cooperative control of striated muscle mass and metabolism by MuRF1 and MuRF2. *EMBO J.* **27**, 350–360 (2008).

538. Shi, J., Luo, L., Eash, J., Ibebunjo, C. & Glass, D. J. The SCF-Fbxo40 complex induces IRS1 ubiquitination in skeletal muscle, limiting IGF1 signaling. *Dev. Cell* **21**, 835–847 (2011).

539. Kamei, Y. *et al.* Skeletal muscle FOXO1 (FKHR) transgenic mice have less skeletal muscle mass, down-regulated type I (slow twitch/red muscle) fiber genes, and impaired glycemic

control. *J. Biol. Chem.* **279**, 41114–41123 (2004).

540. Schips, T. G. *et al.* FoxO3 induces reversible cardiac atrophy and autophagy in a transgenic mouse model. *Cardiovasc. Res.* **91**, 587–597 (2011).

541. Mammucari, C. *et al.* FoxO3 Controls Autophagy in Skeletal Muscle In Vivo. *Cell Metab.* **6**, 458–471 (2007).

542. Calissi, G., Lam, E. W. F. & Link, W. Therapeutic strategies targeting FOXO transcription factors. *Nat. Rev. Drug Discov.* **20**, 21–38 (2021).

543. Liu, C. M. *et al.* Effect of RNA oligonucleotide targeting Foxo-1 on muscle growth in normal and cancer cachexia mice. *Cancer Gene Ther.* **14**, 945–952 (2007).

544. Aartsma-Rus, A. Overview on AON design. *Methods Mol. Biol.* **867**, 117–129 (2012).

545. Gait, M. J. *et al.* Cell-Penetrating Peptide Conjugates of Steric Blocking Oligonucleotides as Therapeutics for Neuromuscular Diseases from a Historical Perspective to Current Prospects of Treatment. *Nucleic Acid Ther.* **29**, 1–12 (2019).

546. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acid Res.* **29**, e45 (2001).

547. Ramakers, C., Ruijter, J. M., Lekanne Deprez, R. H. & Moorman, A. F. M. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* **339**, 62–66 (2003).

548. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**, 265–275 (1951).

549. Pettersson, M. & Crews, C. M. PROteolysis TArgeting Chimeras (PROTACs) — Past, present and future. *Drug Discov. Today Technol.* **31**, 15–27 (2019).

550. Chamberlain, P. P. & Hamann, L. G. Development of targeted protein degradation therapeutics. *Nat. Chem. Biol.* **15**, 937–944 (2019).

551. Salcher, S. *et al.* A drug library screen identifies Carbenoxolone as novel FOXO inhibitor that overcomes FOXO3-mediated chemoprotection in high-stage neuroblastoma. *Oncogene*

39, 1080–1097 (2020).

552. Brent, M. M., Anand, R. & Marmorstein, R. Structural basis for DNA recognition by FoxO1 and its regulation by post-translational modification. *Structure* **16**, 1407–1416 (2008).

553. Fuchs, O., Bokorová, R., Vostrý, M., Koste\vcka, A. & Polák, J. Cereblon and Its Role in the Treatment of Multiple Myeloma by Lenalidomide or Pomalidomide. in (2014).

554. Myung, J., Kim, K. B. & Crews, C. M. The ubiquitin-proteasome pathway and proteasome inhibitors. *Med. Res. Rev.* **21**, 245–273 (2001).

555. Webb, A. E., Kundaje, A. & Brunet, A. Characterization of the direct targets of FOXO transcription factors throughout evolution. *Aging Cell* **15**, 673–685 (2016).

556. Furuyama, T., Nakazawa, T., Nakano, I. & Mori, N. Identification of the differential distribution patterns of mRNAs and consensus binding sequences for mouse DAF-16 homologues. *Biochem. J.* **349**, 629–634 (2000).

557. Tsai, K. L. *et al.* Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification. *Nucleic Acids Res.* **35**, 6984–6994 (2007).

558. Carter, J. D. & LaBean, T. H. Coupling strategies for the synthesis of peptide-oligonucleotide conjugates for patterned synthetic biomineralization. *J. Nucleic Acids* **2011**, (2011).

559. Sakamoto, K. M. *et al.* Protacs: Chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8554–8559 (2001).

560. Ghidini, A., Cléry, A., Halloy, F., Allain, F. H. T. & Hall, J. RNA-PROTACs: Degraders of RNA-Binding Proteins. *Angew. Chemie - Int. Ed.* **60**, 3163–3169 (2021).

561. Shao, J. *et al.* Destruction of DNA-binding proteins by programmable O'PROTAC: Oligonucleotide-based PROTAC. *bioRxiv* 2021.03.08.434493 (2021).

562. Kornblihtt, A. R., De La Mata, M., Fededa, J. P., Muñoz, M. J. & Nogués, G. Multiple links between transcription and splicing. *Rna* **10**, 1489–1498 (2004).

563. Kornblihtt, A. R. *et al.* Alternative splicing: A pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14**, 153–165 (2013).

564. Dujardin, G. *et al.* How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping. *Mol. Cell* **54**, 683–690 (2014).

565. Luco, R. F. *et al.* Regulation of alternative splicing by histone modifications. *Science (80-.).* **327**, 996–1000 (2010).

566. Essaghir, A., Dif, N., Marbehant, C. Y., Coffer, P. J. & Demoulin, J. B. The transcription of FOXO genes is stimulated by FOXO3 and repressed by growth factors. *J. Biol. Chem.* **284**, 10334–10342 (2009).

567. Abi Hussein, H. *et al.* Global vision of druggability issues: applications and perspectives. *Drug Discov. Today* **22**, 404–415 (2017).

568. Edfeldt, F. N. B., Folmer, R. H. A. & Breeze, A. L. Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discov. Today* **16**, 284–287 (2011).

569. Desaphy, J., Azdimousa, K., Kellenberger, E. & Rognan, D. Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. *J. Chem. Inf. Model.* **52**, 2287–2299 (2012).

570. Patel, M. N., Halling-Brown, M. D., Tym, J. E., Workman, P. & Al-Lazikani, B. Objective assessment of cancer genes for drug discovery. *Nat. Rev. Drug Discov.* **12**, 35–50 (2013).

571. Samarasinghe, K. T. G. & Crews, C. M. Targeted protein degradation: A promise for undruggable proteins. *Cell Chem. Biol.* **28**, 934–951 (2021).

572. Poso, A. The Future of Medicinal Chemistry, PROTAC, and Undruggable Drug Targets. *J. Med. Chem.* **64**, 10680–10681 (2021).

573. Liu, J. *et al.* TF-PROTACs Enable Targeted Degradation of Transcription Factors. *J. Am. Chem. Soc.* **143**, 8902–8910 (2021).

574. Samarasinghe, K. T. G. *et al.* Targeted degradation of transcription factors by TRAFTACs: TRAnscription Factor TArgeting Chimeras. *Cell Chem. Biol.* **28**, 648-661.e5 (2021).

575. Vilchinskaya, N., Altaeva, E. & Lomonosova, Y. Gaining insight into the role of FoxO1 in the progression of disuse-induced skeletal muscle atrophy. *Adv. Biol. Regul.* **85**, 100903 (2022).

576. Dansen, T. B. & Burgering, B. M. T. Unravelling the tumor-suppressive functions of FOXO proteins. *Trends Cell Biol.* **18**, 421–429 (2008).

577. Link, W. & Fernandez-Marcos, P. J. FOXO transcription factors at the interface of metabolism and cancer. *Int. J. Cancer* **141**, 2379–2391 (2017).

578. Myatt, S. S. & Lam, E. W. F. The emerging roles of forkhead box (Fox) proteins in cancer. *Nat. Rev. Cancer* **7**, 847–859 (2007).

579. Yu, S. *et al.* Activation of FOXO3a suggests good prognosis of patients with radically resected gastric cancer. *Int. J. Clin. Exp. Pathol.* **8**, 2963–2970 (2015).

580. Hu, M. C. T. *et al.* IκB Kinase Promotes Tumorigenesis through Inhibition of Forkhead FOXO3a. *Cell* **129**, 1427–1428 (2007).

581. Czech, C. *et al.* Biomarker for spinal muscular atrophy: Expression of SMN in peripheral blood of SMA patients and healthy controls. *PLoS One* **10**, 1–11 (2015).

582. Alves, C. R. R. *et al.* Whole blood survival motor neuron protein levels correlate with severity of denervation in spinal muscular atrophy. *Muscle and Nerve* **62**, 351–357 (2020).

583. Xi, X. *et al.* RNA biomarkers: Frontier of precision medicine for cancer. *Non-coding RNA* **3**, (2017).

584. Hayek, H., Eriani, G. & Allmang, C. eIF3 Interacts with Selenoprotein mRNAs. *Biomolecules* **12**, 1–16 (2022).

585. Biogen. *Third quarter 2021 Financial Results and Business Update*. *Biogen* (2021) doi:10.1111/oet.12888.