

Aus dem Center for Regenerative Therapies Dresden

Direktor: Prof. Dr. Michael Brand

Role of alternative splicing in neurogenic commitment

DISSERTATIONSSCHRIFT

zur Erlangung des akademischen Grades

Doctor of Philosophy (Ph.D.)

vorgelegt

der Medizinischen Fakultät Carl Gustav Carus

der Technischen Universität Dresden

von

M.Sc Neuroscience Leila Haj Abdullah Alieh

aus Brescia, Italien

Dresden 2020

1. Gutachter: Prof. Dr. Gerd Kempermann
DZNE German Center for Neurodegenerative Diseases

2. Gutachter: Prof. Dr. Marius Ader
Center for Regenerative Therapies Dresden

Tag der mündlichen Prüfung:

gez.: _____
Vorsitzender der Promotionskommission

Anmerkung:

Die Eintragung der Gutachter und Tag der mündlichen Prüfung (Verteidigung) erfolgt nach Festlegung von Seiten der Medizinischen Fakultät Carl Gustav Carus der TU Dresden. Sie wird durch die Promovenden nach der Verteidigung zwecks Übergabe der fünf Pflichtexemplare an die Zweigbibliothek Medizin in gedruckter Form oder handschriftlich vorgenommen.

Abstract

Role of alternative splicing in neurogenic commitment

To form complex organisms characterized by different tissues with specialized functions, cells must acquire distinct identities during development. Yet, all the cells of an organism are equipped with the same genomic information. Elucidating the mechanisms that regulate the determination of a cell identity, i.e. the cell-fate commitment, is a main purpose in developmental biology. Numerous studies focused on genes that are activated or repressed at each stage of differentiation, identifying several key regulators of development. However, this approach ignores the transcript variability derived from alternative splicing, the transcriptional process by which different gene coding segments, i.e. exons, are combined giving rise to multiple transcripts and proteins from the same gene. With the advent of novel sequencing technologies, it is becoming clear that alternative splicing is widespread in higher organisms, regulates several processes and presents tissue- and cell-specificity. In mammals, the brain shows the highest degree of alternative splicing, with neurons expressing a high variety of splice variants. In this project I investigated whether and how alternative splicing could regulate cell-fate determination in the context of the embryonic development of the mouse neocortex, a highly complex structure presenting several different neuronal subtypes generated at specific time points. For this purpose, I analyzed transcriptome data of cells of the neurogenic lineage isolated from the developing mouse neocortex at subsequent stages of differentiation. I showed that the expression pattern of the proteins regulating splicing, i.e. the splicing factors, changes during neocortical development. By employing several bioinformatic tools, I described the splicing profile that characterizes each differentiation stage and, for the first time, I identified the splicing events that mark cell-fate commitment to a neurogenic identity. Alternative splicing mostly involved genes with a role in nervous system development, cell growth and signaling, mainly leading to the production of alternative protein isoforms. Splicing choices taken during the neurogenic commitment were kept throughout neurogenesis. Thus, exons that start to be included during cell-fate determination are always included in post-mitotic neurons. Exons gained during neurogenic commitment were characterized by strong features in their upstream intron, presented a general short length with an overrepresentation of microexons in the 3-27 nucleotides length range and showed an enrichment for binding motifs of the neural splicing factor nSR100. *In vivo* manipulation in the embryonic mouse neocortex highlighted isoform-specific effects on neocortical development, strongly suggesting a

causal relationship between alternative splicing choices and cell-fate commitment. Moreover, the higher cell-specificity offered by the present dataset, compared to similar studies, allowed a better understanding of previously identified splicing events that characterize the nervous system and the relationships between neural-specific splicing factors.

Zusammenfassung

Die Rolle von alternativem Spleißen bei der neurogenen Festlegung

Um einen komplexen Organismus zu bilden, der charakterisiert werden kann durch verschiedene Gewebe mit spezialisierten Funktionen, müssen Zellen während der Entwicklung bestimmte Identitäten annehmen. Dennoch sind alle Zellen eines Organismus mit denselben genomischen Informationen ausgestattet. Einer der wichtigsten Ziele der Entwicklungsbiologie ist die Aufklärung von Mechanismen, die die Bestimmung der Zellidentität, d.h. Zellschicksals-festlegung, regulieren. Zahlreiche Studien haben sich auf Gene fokussiert, die bei einem Entwicklungsstadium aktiviert oder reprimiert werden, was zur Identifizierung von einigen Schlüsselregulatoren der Entwicklung geführt hat. Dieser Ansatz ignoriert jedoch die Transkriptvariabilität bedingt durch alternatives Spleißen – ein Transkriptionsprozess, bei dem verschiedene Gen-kodierende Segmente, sogenannte Exons, unterschiedlich kombiniert werden, was zu multiplen Transkripten und Proteinen vom selben Gen führt. Mit dem Aufkommen von neuen Sequenzierungstechnologien wurde klar, dass alternatives Spleißen in höheren Organismen weit verbreitet ist, mehrere Prozesse reguliert und Gewebe- und Zellspezifität aufweist. Bei Säugern zeigt das Gehirn den höchsten Grad an alternativem Spleißen und Neurone exprimieren eine hohe Vielfalt an Spleißvarianten. In diesem Projekt habe ich untersucht, ob und wie alternatives Spleißen Zellschicksalsbestimmungen regulieren kann im Kontext der Embryonalentwicklung des Maus-Neokortex, einer hoch komplexen Struktur, die mehrere unterschiedliche neuronale Subtypen zu spezifischen Zeitpunkten generiert. Zu diesem Zweck habe ich Transkriptionsdaten von Zellen der neurogenen Linie in aufeinanderfolgenden Differenzierungsstadien analysiert, die vom sich entwickelnden Maus-Neokortex isoliert worden waren. Ich habe gezeigt, dass sich das Expressionsmuster von Proteinen, die das Spleißen regulieren, d.h. Spleißfaktoren, während der neokortikalen Entwicklung ändert. Durch die Verwendung von einigen bioinformatischen Programmen habe ich das Spleißprofil beschrieben, was jeden Differenzierungsschritt charakterisiert und ich habe zum ersten Mal Spleißereignisse identifiziert, die die Zellschicksalsfestlegung zu einer neurogenen Identität kennzeichnen. Alternatives Spleißen war hauptsächlich involviert in Gene mit einer Rolle in Nervensystementwicklung, Zellwachstum und –signalkaskaden, die meistens zu einer Produktion von alternativen Proteinisoformen geführt haben. Spleißentscheidungen, die während der neurogenen Festlegung getroffen wurden, wurden

während der Neurogenese durchgängig beibehalten. Demzufolge waren Exons, die zu Beginn der Zellentwicklungsbestimmung eingeschlossen wurden, in post-mitotischen Neuronen immer eingeschlossen. Dazugewonnene Exons während der neurogenen Festlegung wurden charakterisiert durch starke Features im upstream Intron, wiesen allgemein eine kurze Länge mit einer Überrepräsentierung von Mikroexons mit einer Längenspanne von 3-27 Nukleotiden auf und zeigten eine Anreicherung von Bindungsmotiven des neuralen Spleißfaktors nSR100. Die *in vivo* Manipulation im embryonalen Maus-Neokortex hat die Isoform-spezifischen Effekte auf die neokortikale Entwicklung hervorgehoben und legt einen kausalen Zusammenhang zwischen alternativen Spleißentscheidungen und Zellschicksalsfestlegung sehr nahe. Darüber hinaus hat die höhere Zellspezifität durch den aktuellen Datensatz verglichen mit ähnlichen Studien ein besseres Verständnis von zuvor identifizierten Spleißereignissen ermöglicht, die das Nervensystem und die Zusammenhänge zwischen neural-spezifischen Spleißfaktoren charakterisieren.

Übersetzt von Simon Hertlein

Table of Contents

Abstract	I
Zusammenfassung	III
Table of Contents	V
List of Figures	VII
List of Tables	IX
Abbreviations	X
Gene abbreviations	XII
1 Introduction	1
1.1 Neurogenesis during embryonic development	2
1.1.1 Formation and patterning of the neural tube	2
1.1.2 Neural progenitors in the dorsal telencephalon	6
1.1.3 Neurogenesis	8
1.1.4 Regulation of neurogenesis	10
1.1.5 A novel tool to investigate cell-fate determination in the central nervous system: the Btg2 ^{RFP} /Tubb3 ^{GFP} mouse line	13
1.2 Alternative splicing: an additional level of genomic regulation	15
1.2.1 The splicing reaction	16
1.2.2 What makes splicing alternative?	18
1.2.3 Regulation of alternative splicing	19
1.2.4 The challenge to detect splicing	23
1.2.5 New sequencing technologies reveal a high transcriptome complexity	29
1.2.6 Splicing in nervous system development	31
1.2.7 Aims of the project	36
2 Materials and methods	38
2.1 Materials	38
2.1.1 Bacteria, cells, mouse strains	38
2.1.2 Vector	38
2.1.3 Primers	38
2.1.4 Chemicals and buffers	41
2.1.5 Antibodies	42
2.1.6 Kits and enzymes	42
2.2 Methods	43
2.2.1 Animal experiments	43
2.2.2 Molecular biology	44
2.2.3 Immunohistochemistry	46
2.2.4 Bioinformatics	47
3 Results	53
3.1 Splicing factors are differentially expressed during neurogenic commitment and neurogenesis	53

3.2	Detection of alternative splicing	55
3.2.1	Isoform-switching	55
3.2.2	Exon usage and splicing events	57
3.3	Validation	62
3.3.1	The isoform switching method has a poor validation rate	62
3.3.2	Analysis at the exon level has a high rate of validation	65
3.4	Pattern and representation of splicing events	67
3.4.1	Splicing choices during neurogenic commitment define the splicing profiles of neurons	67
3.4.2	Splicing events: microexon inclusion characterizes neurogenic commitment	69
3.5	Alternative splicing changes the protein output of genes involved in neurogenesis	75
3.5.1	Spliced genes are involved in neurogenesis and signaling	75
3.5.2	Impact of alternative splicing on the proteome	77
3.6	Splicing regulation: neural exon features and splicing factor binding	79
3.6.1	Included neural exons are short and preceded by strong exon-definition features	79
3.6.2	Early included exons are enriched for nSR100 binding sites	85
3.7	The Btg2 ^{RFP} /Tubb3 ^{GFP} mouse line outperforms previous models for the study of cell-type-specific splicing in the brain	88
3.8	In vivo manipulation of splice variants	90
4	Discussion	94
4.1	The combination of different bioinformatic approaches allows an accurate identification of splicing events at the exon-level	95
4.2	Splicing choices during neurogenic commitment establish a neural signature characterized by microexon inclusion	97
4.3	Splicing during neocortical development leads to the generation of alternative protein isoforms in genes involved in neurogenesis and signaling	98
4.4	Neural exons are short and present strong features facilitating inclusion	101
4.5	Neural exons are prevalently regulated by nSR100 during neurogenic commitment	102
4.6	In vivo overexpression of splice variants highlights isoform-specific functions in neurogenic commitment	105
4.7	Concluding remarks and future perspectives	108
5	Supplementary figures	110
6	References	118
	Acknowledgments	137
	Anlage I	138
	Anlage II	139

List of Figures

Figure 1.1 Formation of the neural tube	3
Figure 1.2 Patterning of the neural tube	5
Figure 1.3 Interkinetic nuclear migration	6
Figure 1.4 Symmetric and asymmetric divisions	8
Figure 1.5 Neurogenesis in the neocortex.....	9
Figure 1.6 Btg2 ^{RFP} /Tubb3 ^{GFP} murine line	14
Figure 1.7 Splicing reaction	17
Figure 1.8 Classification of splicing events.....	19
Figure 1.9 Splicing regulation	22
Figure 1.10 Analysis workflow of differential splicing from bulk short-read RNA sequencing	25
Figure 1.11 Different ways to calculate the percent spliced in (PSI)	28
Figure 1.12 Splicing in the nervous system.....	34
Figure 3.1 Differential expression of splicing factors in cells of the neurogenic lineage....	54
Figure 3.2 Detection of Isoform Switching with Cuffdiff2	56
Figure 3.3 Exon usage	57
Figure 3.4 Comparison of the results obtained with the different exon-centric methods...	59
Figure 3.5 Combination of the DEXSeq-PSI and VAST-tools datasets.....	61
Figure 3.6 qPCR validation of isoform switching predicted by Cuffdiff2	64
Figure 3.7 Validation of exons predicted to be alternatively spliced by DEXSeq-PSI and VAST-tools.....	66
Figure 3.8 Pattern of exon inclusion/exclusion	68
Figure 3.9 Classification of splicing events with VAST-tools	70
Figure 3.10 Classification of splicing events derived by the combination of VAST-tools and DEXSeq-PSI	72
Figure 3.11 Pattern of exon inclusion in different length classes of cassette exons	74
Figure 3.12 Gene ontology analysis	76
Figure 3.13 Impact of alternative splicing on the proteome.....	78
Figure 3.14 Exon features of spliced cassette exons	80
Figure 3.15 Features of introns upstream spliced cassette exons	82
Figure 3.16 Strength of splicing regulatory features of included exons and of microexons	84
Figure 3.17 Zfp36 binding sites	86

Figure 3.18 nSR100 binding sites	87
Figure 3.19 Comparison with other dataset of splicing in the mouse neurogenic lineage	89
Figure 3.20 <i>In vivo</i> assay of Faim isoforms.....	92
Supplementary Figure 1 <i>In vivo</i> assay of genes predicted to undergo isoform switching during neurogenic commitment.....	111
Supplementary Figure 2 nSR100 binding sites in different exon length classes	112
Supplementary Figure 3 nSR100 binding sites in exons (>51 nt).....	113
Supplementary Figure 4 Nova2 binding sites	114
Supplementary Figure 5 Nova2 binding sites in exons (>51 nt).....	115
Supplementary Figure 6 Ptpb1/2 binding sites	116
Supplementary Figure 7 Splicing of Rest and Ptpb2	117

List of Tables

Table 2.1 Bacteria, cells, mouse strains.....	38
Table 2.2 Primers for Cloning.....	39
Table 2.3 Primers for RT-qPCR.....	40
Table 2.4 Primers for RT-PCR.....	40
Table 2.5 Buffers for General Use.....	41
Table 2.6 Buffers for Immunohistochemistry.....	41
Table 2.7 Culture Media.....	42
Table 2.8 Primary Antibodies for Immunohistochemistry.....	42
Table 2.9 Kits and Enzymes.....	43
Table 3.2 External contribution to the project.....	93
Supplementary Table 1 Spliced genes identified by isoform switching analysis and selected for validation and <i>in vivo</i> assay.....	110

Abbreviations

A	Adenine
aa	Amino acid
7-AAD	7-amino-actinomycin D
AMPA	α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
AP	Apical progenitor
bHLH	Basic-helix-loop-helix
bp	Base pairs
BP	Basal progenitor
BrP	Branch point
cDNA	Complementary DNA
CDS	Coding sequence
circRNA	Circular RNA
CNS	Central nervous system
CP	Cortical plate
Ct	Cycle threshold
CTD	Carboxy terminal domain
DAPI	4',6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
DP	Differentiating progenitors
E	Embryonic day
eMIC	Enhancer of microexons
ER	Exclusion reads
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
FACS	Fluorescent activated cell sorting
FC	Fold change
FDR	False discovery rate
FPKM	Fragment per kilobase of transcript per million mapped reads
GFP	Green fluorescent protein
GO	Gene ontology
H3K4	Histone 3 lysine 4
HLH	Helix-loop-helix
hnRNP	Heterogeneous nuclear ribonucleoproteins
Id	Inhibitors of differentiation or inhibitors of DNA binding
IKNM	Interkinetic nuclear migration
InR	Inclusion reads
IR	Intron retention
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
IUE	In utero electroporation

IZ	Intermediate zone
log₂FC	Log2 fold change
LTD	Long term depression
miR124	Micro RNA 124
miRNA	Micro RNA 124
mRNA	Messenger RNA
M	Mitosis
MZ	Molecular zone
N	Neuron
ncRNA	Non-coding RNA
NEC	Neuroepithelial cell
NICD	Notch intra cellular domain
nls	Nuclear localization signal
NMD	Nonsense mediated decay
nt	Nucleotide
ORF	Open reading frame
P	Postnatal day
PCR	Polymerase chain reaction
pg	Picograms
PoI II	Polymerase II
PP	Proliferating progenitors
PPT	Polypyrimidine tract
PSI	Percent spliced in
qPCR	Quantitative polymerase chain reaction
RFP	Red fluorescent protein
RG	Radial glia
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
RT	Reverse transcription
scRNA-Seq	Single-cell RNA sequencing
snRNP	Small nuclear ribonucleoproteins
SR	Seine-Arginine
SS	Splice site
SVZ	Subventricular zone
T	Thymine
TSS	Transcription start site
UTR	Untranslated region
VZ	Ventricular zone

Gene abbreviations

Gene symbol used	Official gene symbol	Description
Abi2	Abi2	Abl-interactor 2
Add1	Add1	Adducin 1 (alpha)
Apba2	Apba2	Amyloid beta (A4) precursor protein-binding, family A, member 2
Apbb1	Apbb1	Amyloid beta (A4) precursor protein-binding, family B, member 1
Aspm	Aspm	Abnormal spindle microtubule assembly
Bcl-x	Bcl2l1	BCL2-like 1
Bk	Kcnma1	Potassium calcium-activated channel subfamily M alpha 1
Blbp	Fabp7	Fatty acid binding protein 7, brain
BMP	Bmp	Bone morphogenetic protein
Btg2 (Tis21)	Btg2	BTG anti-proliferation factor 2
Cdk4	Cdk4	Cyclin-dependent kinase 4
Celf5	Celf5	CUGBP, Elav-like family member 5
Csdc2	Csdc2	Cold shock domain containing C2, RNA binding
Cux1, Cux2	Cux1, Cux2	Cut-like homeobox 1/2
CyclinD	Ccnd1	Cyclin D1
Daam1	Daam1	Dishevelled associated activator of morphogenesis 1
Dab1	Dab1	Disabled 1
Dll1	Dll1	Delta like canonical Notch ligand 1
Dscam	Dscam	Down syndrome cell adhesion molecule
E-cadherin	Cdh1	Cadherin 1
Eef1a1	Eef1a1	Eukaryotic translation elongation factor 1 alpha 1
Elavl2	Elavl2	ELAV like RNA binding protein 1
Erc1, Erc2	Erc1, Erc2	ELKS/RAB6-interacting/CAST family member 1, ELKS/RAB6-interacting/CAST family member 2
Faim	Faim	Fas apoptotic inhibitory molecule
Fas	Fas	Fas (TNF receptor superfamily member 6)
Fgf (Fgf2, Fgf8, Fgf10, Fgf15)	Fgf (Fgf2, Fgf8, Fgf10, Fgf15)	Fibroblast growth factor 2/8/10/15
Rbfox	Rbfox (Rbfox1, Rbfox2, Rbfox3)	RNA binding protein, fox-1 homolog (C. elegans) 1/2/3
Foxg1	Foxg1	Forkhead box G1
Git1	Git1	GIT ArfGAP 1

Gene symbol used	Official gene symbol	Description
Glast	Slc1a3	Solute carrier family 1 (glial high affinity glutamate transporter), member 3
Gli3	Gli3	GLI-Kruppel family member GLI3
Gopc	Gopc	Golgi associated PDZ and coiled-coil motif containing
GluA2	Gria2	Glutamate receptor, ionotropic, AMPA2 (alpha 2)
Gsx2	Gsx2	GS homeobox 2
Hes1, Hes5	Hes1, Hes5	Hes family bHLH transcription factor 1/5
Homer1	Homer1	Homer scaffolding protein 1
Igf	Igf	Insulin-like growth factor
Insm1	Insm1	Insulinoma-associated 1
Kat5/Tip60	Kat5	K(lysine) acetyltransferase 5
Lgn	Gpsm2	G-protein signaling modulator 2
Lsd1	Kdm1a	Lysine (K)-specific demethylase 1A
Med23	Med23	Mediator complex subunit 23
Mkrn1	Mkrn1	Makorin, ring finger protein, 1
N-cadherin	Cdh2	Cadherin 2
Neo1	Neo1	Neogenin
NeuroD1, NeuroD2, NeuroD6	NeuroD1, NeuroD2, NeuroD6	Neurogenic differentiation 1/2/6
NGF	NGF	Nerve growth factor
Ngn2	Neurog2	Neurogenin 2
Nin	Nin	Ninein
Nkx2-1	Nkx2-1	NK2 homeobox 1
Nmdar1	Grin1	Glutamate receptor, ionotropic, NMDA1
Nova1, Nova2	Nova1, Nova2	Neuro-oncological ventral antigen (NOVA) alternative splicing regulator 1/2
nSR100/Srrm4	Srrm4	Serine/arginine repetitive matrix 4
Numb	Numb	NUMB endocytic adaptor protein
p300	Ep300	E1A binding protein p300
Parp6	Parp6	Poly (ADP-ribose) polymerase family, member 6
Pax6	Pax6	Paired box 6
Porcn	Porcn	Porcupine O-acyltransferase
Psd95	Dlg4	Discs large MAGUK scaffold protein 4
Ptk2	Ptk2	PTK2 protein tyrosine kinase 2
Ptbp1, Ptbp2	Ptbp1, Ptbp2	Polypyrimidine tract binding protein 1/2
RBPjk	Rbpj	Recombination signal binding protein for immunoglobulin kappa J region
Rnps1	Rnps1	RNA binding protein with serine rich domain 1
Sam68	Khdrbs1	KH domain containing, RNA binding, signal transduction associated 1

Gene symbol used	Official gene symbol	Description
SF1	Sf1	Splicing factor 1
Shh	Shh	Sonic hedgehog
Six3	Six3	Sine oculis-related homeobox 3
Sox1, Sox3	Sox1, Sox3	SRY (sex determining region Y)-box 1/3
Srrm2/3/4	Srrm2/3/4	Serine/arginine repetitive matrix 2/3/4
Srsf11	Srsf11	Serine and arginine rich splicing factor 11
Svet1	Svet1	Subventricular expressed transcript 1
Svil	Svil	Supervillin
Tbr2	Eomes	Eomesodermin (T-box brain protein 2)
Tgfβ 1/2/3	Tgfβ 1/2/3	Transforming growth factor beta 1/2/3
Tmcc2	Tmcc2	Transmembrane and coiled-coil domains 2
TNFα	Tnf	Tumor necrosis factor
Tubb3	Tubb3	Tubulin, beta 3 class III
U2AF	U2af	U2 auxiliary factor
Wnt	Wnt	Wingless-type MMTV integration site family
Wnt5b	Wnt5b	Wingless-type MMTV integration site family, member 5B
XIAP	Xiap	X-linked inhibitor of apoptosis
Zfp36	Zfp36	Zinc finger protein 36
Zfp521	Zfp521	Zinc finger protein 521
Zfyve27	Zfyve27	Zinc finger, FYVE domain containing 27

1 Introduction

During development, cells with identical genomic information acquire different identities. Lineage-restricted stem cells proliferate to increase their pool and, as development proceeds, progressively divide into differentiated cells that form specialized tissues. The choice between proliferative and differentiative division is subjected to tight regulation that is essential to ensure proper size and structure of the tissues that form an organism. Despite decades of research, the mechanisms by which a stem cell switches from proliferative to differentiative division, i.e. cell-fate commitment, are still elusive. Researchers investigating cell-fate commitment have mostly focused on differential gene expression and disregarded the transcript diversity generated by alternative splicing. In this process, different genomic information can be expressed in the mature transcript of a gene, dramatically expanding the genomic coding potential. As a result, the same gene can give rise to several transcripts with distinct functions without changing the overall expression level. The development of new sequencing technologies has revealed an unexpected transcript complexity, with novel isoforms found even in well-studied model organisms. The role of alternative splicing in genomic regulation is therefore gaining more and more attention, and is likely to be crucial in development.

In this chapter, the developmental processes that are known to contribute to the formation of a complex tissue such as the neocortex of the brain will be described using *Mus musculus* as a model. Mechanisms of alternative splicing and its regulation, examples that support the role of this process in the development of the nervous system and strategies to analyze alternative splicing decisions will also be presented.

1.1 Neurogenesis during embryonic development

1.1.1 Formation and patterning of the neural tube

Formation of the neural tube

Upon fertilization, the mouse zygote undergoes a series of proliferation, migration and commitment events to form the inner cell mass, a group of pluripotent stem cells destined to form the embryo which is surrounded by the trophoblast, the precursor of extraembryonic structures. The inner cell mass further divides into the hypoblast, or primitive endoderm, and the epiblast by embryonic day 4.5 (E4.5) (Gilbert and Barresi, 2016).

During gastrulation (E6.5-8.0), cells of the epiblast migrate through a rostro-caudal medial streak, termed the primitive streak, to specify the three germ layers that form different tissues of the organism: endoderm, mesoderm and ectoderm. The endoderm, the innermost layer, will give rise to lung and digestive epithelia. The mesoderm, the middle layer, forms muscles, blood vessels and connective tissues. The outermost layer, the ectoderm, is the precursor of the epidermis and of the nervous system (Gilbert and Barresi, 2016). A thickening of cells in the anterior part of the primitive streak forms the node, the organizer of the embryo body plan, while cells that migrate through the node give rise to the notochord, a medial rod-like mesodermal structure important for embryonic tissue patterning (Balmer et al., 2016).

The specification of the ectodermal neural field, the neural plate, requires the coordination of paracrine factors as well as intrinsic signaling. BMP secreted by cells of the ectoderm drives an epidermal specification while BMP-inhibitory factors, secreted by the node and the notochord, allow the medial field of the ectoderm to commit to a neural fate (Gilbert and Barresi, 2016). In mammals, *in vitro* studies suggest that neural specification depends also on the activation of intrinsic factors such as the nuclear zinc finger protein Zfp521 (Kamiya et al., 2011). Weak Zfp521 expression appears diffusely at embryonic day E6.5-7.0, becoming stronger and restricted to the prospective neural plate as gastrulation proceeds (E7.0-8.0). Zfp521, in association with the co-activator p300, binds to the promoter of neuroectodermal specific genes (e.g. Sox1, Sox3, Ncad, Pax6) and promotes their expression, thus determining the neural specification.

Shortly after specification, the neural plate folds to form the neural tube, a three-dimensional structure which is the precursor of the central nervous system (CNS), in a process called

neurulation: the edges of the neural plate thicken and begin moving upwards, bending along the contact points with the epidermis, the dorsolateral hinge points; concomitantly, a groove starts forming along the midline of the neural plate, corresponding to the medial hinge point that anchors the neural plate to the notochord (Figure 1.1). The neural plate edges ultimately fuse together forming a hollow tube that invaginates beneath the epidermis (E8.0-8.5).

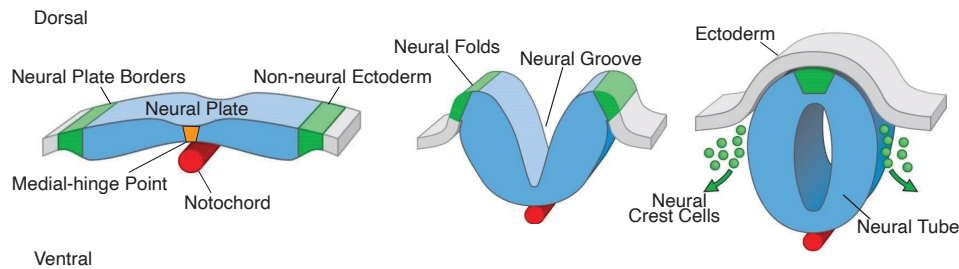


Figure 1.1 Formation of the neural tube. The neural plate (light blue) anchors to the notochord (red) through the medial-hinge point (left panel, orange) and folds forming the neural groove (middle panel). The regions at the boundary with the non-neural ectoderm become the neural folds (green) that eventually fuse together originating the neural tube (right panel). Some cells of the neural crest migrate to form part of the peripheral nervous system. (Adapted from Liu A and Niswander, 2005).

Neural crest cells from the roof of the neural plate delaminate and migrate to form components of the peripheral nervous system. A change in expression of adhesion molecules from E-cadherin in the ectoderm to N-cadherin in the neural tube determines the separation between epidermal and neural cells (Gilbert and Barresi, 2016). The hollow space in the neural tube is filled with cerebrospinal fluid and will give rise to the ventricular system of the brain and the central canal of the spinal cord (Rubenstein and Rakic, 2013).

Patterning of the neural tube

The neural tube is patterned along the rostro-caudal axis to specify different structures of the CNS. In a first step (E9.0), enlargements of the neural tube form three vesicles: the prosencephalon (forebrain), mesencephalon (midbrain) and rhombencephalon (hindbrain). In a second step (E10.5-11.0), the forebrain is further divided into telencephalon, the precursor of cerebrum and hippocampus, and diencephalon, which gives rise to thalamus, hypothalamus and the optic cup, the neural structure of the retina. Likewise, the hindbrain separates into metencephalon, the precursor of the pons and cerebellum, and myelencephalon, the precursor of the medulla oblongata (Figure 1.2 A) (Gilbert and Barresi, 2016).

The rostro-caudal patterning is regulated by differential expression of morphogens and transcription factors that act already before the closure of the neural tube. In particular, the anterior edge of the neural plate, the anterior neural ridge, is a source of factors promoting the telencephalic specification, namely *Fgf8*, *Foxg1* and *Six3*. Further, Wnt proteins show a low rostral to high caudal gradient in the neural plate and their inhibition by *Fgf8*, *Foxg1* and *Six3* in the rostral field is crucial for the initial patterning of the telencephalon (Figure 1.2 B) (Rubenstein and Rakic, 2013).

As for the rostro-caudal patterning, gradients of morphogens and transcription factors are responsible for the dorso-ventral specification. In mammals, commitment to the dorsal fate of the telencephalon requires the transcription factors *Gli3* and *Pax6* and the action of morphogens such as BMP and Fgf proteins. Ventral fates are specified by *Shh*, initially expressed by the midline of the neural plate and subsequently by the ventral midline of the neural tube. *Shh* acts by repressing *Gli3* and by promoting the expression of the transcription factor *Nkx2-1* in the ventralmost part of the neural tube, which gives rise to the medial ganglionic eminence (Figure 1.2 C, left). Wnt proteins, although inhibited during the early establishment of the telencephalic field, are required later on to sustain the expression of *Pax6*. The region at the boundary between the *Pax6* and *Nkx2-1* fields starts to express the transcription factor *Gsx2* and specifies the prospective lateral ganglionic eminence. The medial and lateral ganglionic eminences originate the basal ganglia of the brain such as the striatum and the globus pallidus and are the main sources of inhibitory interneurons (Figure 1.2 C, middle) (Rubenstein and Rakic, 2013).

After the dorsal (pallial) and ventral (subpallial) telencephalic domains have been established, the former is further patterned into the medial pallium that originates the hippocampal primordium and the hem, the dorsal pallium that gives rise to the neocortex, the lateral pallium that originates the piriform cortex and finally the ventral pallium that forms the antihem and claustramygdaloid complex (Figure 1.2 C, right).

The neocortex, site of higher cognitive functions, is the evolutionary most recent part of the brain that underwent dramatic expansion in mammals, occupying most of the dorsal telencephalon. During development, cells of the dorsal telencephalon originate multiple cell types: at early stages they give rise to the excitatory neurons that form the cortical plate of the neocortex and at later stages they form astrocytes and oligodendrocytes (Götz and Huttner, 2005). Different types of neural progenitor cells presenting various degrees of fate-commitment coexist in the developing dorsal telencephalon.

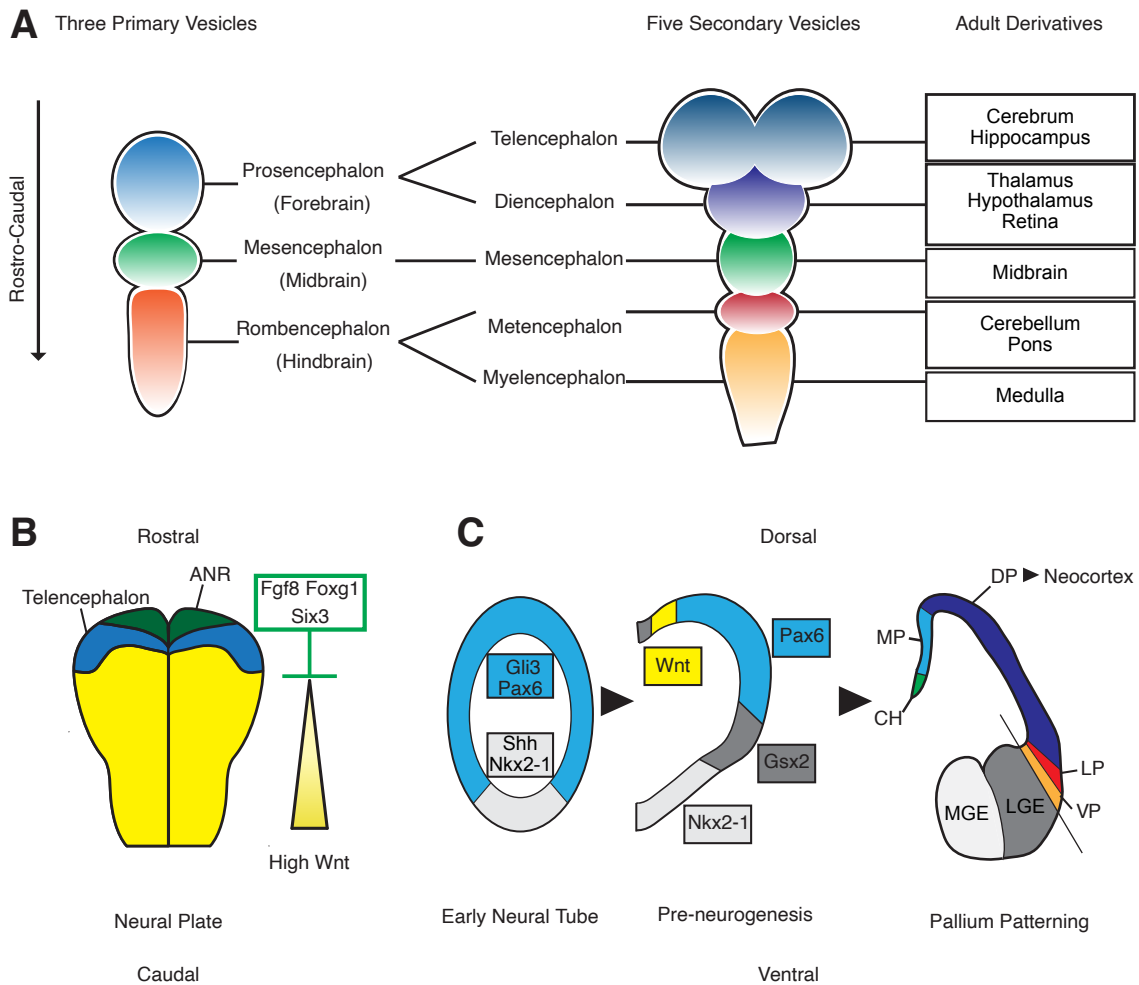


Figure 1.2 Patterning of the neural tube. A: Formation of the brain vesicles. The neural tube enlarges to give rise to three primary vesicles (left panel): prosencephalon (blue), mesencephalon (green) and rhombencephalon (orange). Subsequently the three primary vesicles generate five secondary vesicles (right panel): telencephalon (dark blue), diencephalon (purple), mesencephalon (green), metencephalon (red) and myelencephalon (orange). The five secondary vesicles are the precursors of the different structures of the CNS (right table). B: the anteriormost part of the neural plate, the anterior neural ridge (green), releases factors such as Fgf8, Foxg1 and Six3 that allow the specification of the prospective telencephalic field (blue) by counteracting the diffusion of Wnt proteins. C: after neural tube closure, Gli3 and Pax6 determine the dorsal region of the CNS while Shh and Nkx2-1 specify the ventral fate (C, left). At later stages of neural tube patterning (C, middle), Pax6 establishes the prospective pallium field, while Nkx2-1 is expressed in the prospective medial ganglionic eminence (MGE, right panel). The region at the boundary between Pax6 and Nkx2-1 starts expressing Gsx2 (C, middle) and will give rise to the lateral ganglionic eminence (LGE, right panel). Wnt proteins diffusing from dorsal regions of the neural tube sustain Pax6 expression. The pallium is further specified into distinct regions that will generate the different structures of the brain (right panel: CH=cortical hem; MP=medial pallium; DP=dorsal pallium; LP=lateral pallium; VP=ventral pallium). The largest portion of the pallium, the dorsal pallium, will give rise to the neocortex. (Adapted from Rubenstein and Rakic, 2013).

1.1.2 Neural progenitors in the dorsal telencephalon

The neural plate and neural tube are formed by neuroepithelial cells (NEC) which are elongated cells connected to the lumen on the apical side and to the basal lamina on the basal side. NEC show typical epithelial features and a high apico-basal polarization of their cellular components. The nuclei of the NEC change their position in different phases of the cell cycle in a process termed interkinetic nuclear migration (IKNM). Nuclei move basally during G1, undergo S phase in the basal compartment, then migrate towards the apical side during G2 where they undergo mitosis (Figure 1.3 A). As a consequence of this movement and asynchrony in NEC cell cycle, the neuroepithelium looks stratified even though it is formed by a single cell layer (Figure 1.3 B) (Sauer, 1936).

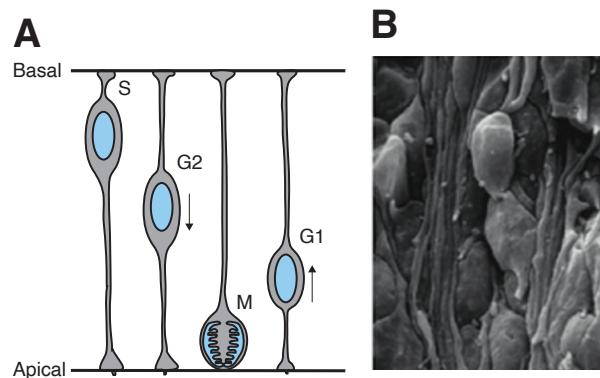


Figure 1.3 Interkinetic nuclear migration. A: schematic representation of neuroepithelial cells (NEC) undergoing interkinetic nuclear migration (IKNM). NEC are localized in the basal compartment during S-phase of the cell cycle while they move apically during G2. After mitosis (M) which takes place in the apical compartment, the nuclei start moving basally during G1. B: electron microscopical image of the pseudostratified neural epithelium (Rubenstein and Rakic, 2013).

A rapid proliferation of NEC after the closure of the neural tube forms the ventricular zone (VZ), a proliferative compartment lining the ventricle (Angevine et al., 1970). Around E10.5, NEC produce the first wave of neurons and start transforming into radial glial cells (RG) in a gradual process that is completed by E12.0 in the mouse dorsal pallium (Hartfuss et al., 2001; Noctor et al., 2002).

RG retain some features of NEC such as apico-basal polarity, IKNM, Pax6 and Nestin expression while losing some epithelial features such as tight junctions. RG are characterized by expression of astroglial markers Blbp, Glast, Vimentin and Tenascin-C (Campbell and Götz, 2002; Götz and Huttner, 2005; Kriegstein and Alvarez-Buylla, 2009).

NEC and RG constitute the main proliferative cells of the dorsal telencephalon and are referred to as apical progenitors (AP). RG generate a second population of neural progenitors, the intermediate precursor cells or basal progenitors (BP) (Haubensak et al., 2004; Noctor et al., 2004). BP do not undergo IKNM, lose their apical and basal processes and migrate basally to the VZ to form a second proliferative compartment, the subventricular zone (SVZ) around E13.0 (Haubensak et al., 2004; Miyata et al., 2004). BP are also characterized by downregulation of Pax6 and astroglial markers as well as the expression of the non-coding RNA Svet1 and the transcription factors Cux1, Cux2 and Tbr2 (also known as Eomes, the distinctive marker of BP) (Götz and Huttner, 2005; Kriegstein and Alvarez-Buylla, 2009).

BP are neural-committed cells (Miyata et al., 2004; Wu SX et al., 2005) that generate either two BP or two neurons (Attardo et al., 2008; Noctor et al., 2008). Starting from E14.5, BP represent the main source of projection neurons that migrate through the intermediate zone (IZ) and form the cortical plate (CP) (Haubensak et al., 2004; Noctor et al., 2004).

The capacity of dividing neural progenitors to expand the neural progenitor pool as well as to generate cells with more restricted potential results from the choice between different modes of cell division. Cortical progenitor cells can divide either symmetrically or asymmetrically. In a symmetric division, a mother cell generates two daughter cells that are identical to one another. The daughter cells can have the same (symmetric proliferative division) or different (symmetric consumptive division) identity as that of the mother cell. In an asymmetric division however, the two daughter cells have different identity from each other. Asymmetric division can be self-renewing if one of the daughter cells has the same identity as that of the mother cell and consumptive if the daughter cells differ from the mother cell as well as from one another (Figure 1.4 A) (Taverna et al., 2014).

The polarized nature of AP is at the basis of asymmetric divisions and the positioning of the cleavage furrow determines whether the cellular components will be inherited equally or unequally by the daughter cells (reviewed in Kriegstein and Alvarez-Buylla, 2009; Taverna et al., 2014). For example, one of the cellular components that is key for the maintenance of proliferative status is the apical membrane. This tiny fraction of membrane (only 1-2% of the whole plasma membrane) can be either dissected or bypassed by the cleavage furrow during mitosis and thus be inherited by both or only one of the daughter cells (Figure 1.4 B) (Kosodo et al., 2004). During an asymmetric division, the proliferating daughter cell inherits the primary cilium protruding into the ventricle and the basal process with the basal endfoot

contacting the basal lamina. A timely switch from proliferative to differentiative division determines the generation of the neurons that will form the neocortex.

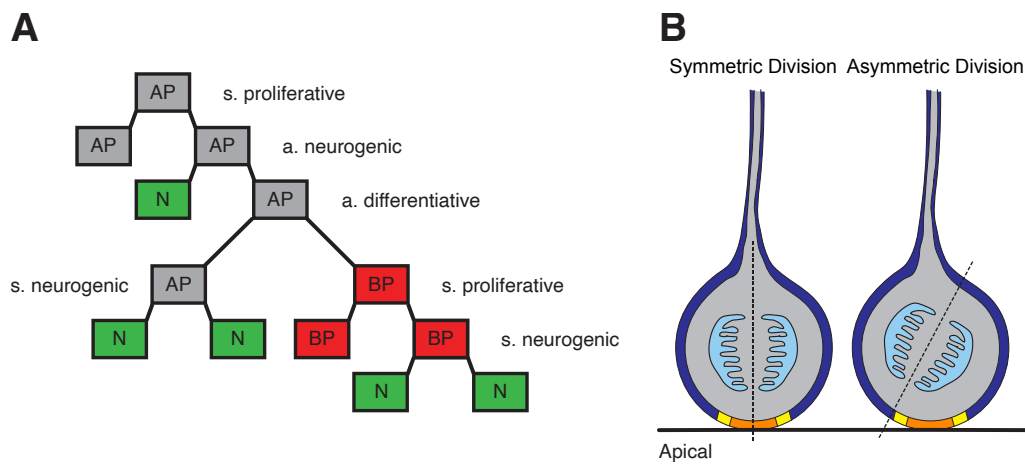


Figure 1.4 Symmetric and asymmetric divisions. A: schematic representation of outcomes of symmetric (s.) and asymmetric (a.) divisions. AP=apical progenitors; BP=basal progenitors; N=neurons. B: schematic representation of symmetric and asymmetric division in neural stem cells. Adherens junctions (yellow) separate the apical membrane (red) from the basolateral membrane (dark blue). Dashed lines represent the plane of cell division that can dissect or bypass the apical membrane in symmetric or asymmetric divisions, respectively.

1.1.3 Neurogenesis

The mammalian neocortex is responsible for higher cognitive functions and presents an extreme complexity in its structure, being composed of many functional units and layers of neurons with different properties. Neocortical layers are populated by projection excitatory neurons that derive from neural precursors of the dorsal telencephalon and are modulated by inhibitory neurons, migrated from the ventral telencephalon (Mérot et al., 2009).

Each cortical layer is formed by distinct subtypes of projection neurons, generated at specific time points during development, expressing characteristic markers and showing a precise pattern of axonal input and output with other structures of the brain (Mérot et al., 2009; Tiberi et al., 2012).

The first burst of neurons generated at E10.5 by NEC forms the preplate, while AP lining the lumen of the ventricle constitute the VZ. Subsequently generated neurons migrate into the preplate to form the CP, splitting the preplate into subplate and marginal zone (MZ). Concomitantly, the Cajal-Retzius cells, inhibitory neurons derived mainly from the hem (Rubenstein and Rakic, 2013), migrate into the MZ and secrete Reelin that, together with Notch (Rice and Curran, 2001; Gaiano, 2008), regulates the migration of newborn neurons

in the CP to form six layers in an inside-out manner. Neurons that are generated early constitute the deep layers while ones generated later form the upper layers (Figure 1.5).

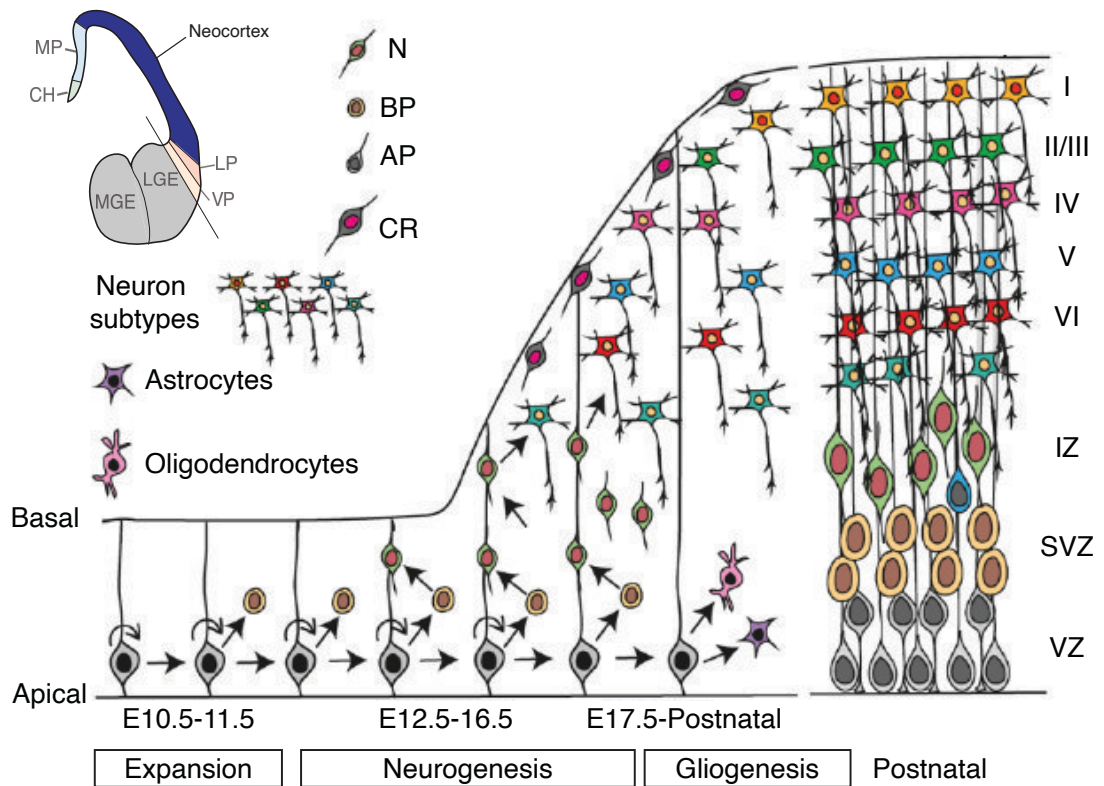


Figure 1.5 Neurogenesis in the neocortex. Different neuronal subtypes are generated in a timely regulated manner by apical progenitors (AP), mostly indirectly through the generation of basal progenitors (BP). Newborn neurons (N) migrate along the radial glial cell fibers to form the different cortical layers (I-VI) in an inside-out manner. Cajal-Retzius cells (CR), mainly generated by the hem, are positioned basally and regulate neuronal migration. At the end of the neurogenic phase, AP generate astrocytes and oligodendrocytes. IZ=intermediate zone; SVZ=subventricular zone; VZ=ventricular zone. Circular arrows indicate proliferation, straight arrows differentiation/maturation. (Adapted from Mukhtar and Taylor, 2018).

The neurogenic process progressively depletes the proliferating pool and leaves a single cell-thick ependymal layer in the VZ. The cortical SVZ disappears as well, being retained only in the lateral walls of the lateral ventricles where it forms a niche of neural precursor cells that are capable of generating olfactory neurons in adulthood (Mérot et al., 2009). It is clear that a correct formation of the neocortical layers requires a tight regulation of the neurogenic process.

1.1.4 Regulation of neurogenesis

1.1.4.1 Extrinsic factors regulating neurogenesis

Morphogens and cues from the cerebrospinal fluid, basal lamina, meninges, other progenitors and newborn neurons influence the decision of an AP to undergo a proliferative or differentiative division. The cerebrospinal fluid contains signaling molecules such as Igf, Fgf, Tgf β , Shh, BMP and Wnt proteins (Lehtinen and Walsh, 2011). Igf and Shh stimulate progenitor self-renewal (Lehtinen et al., 2011; Komada, 2012). Proteins of the Fgf family can promote proliferation (e.g Fgf8) or, on the contrary, repress it and induce differentiation (e.g Fgf15). Progenitors themselves secrete Fgf proteins such as Fgf2 and Fgf10 in the VZ. Fgf10 is expressed earlier and promotes the transition from NEC to RG while Fgf2 is involved in sustaining the proliferation of progenitors in the early phases of neurogenesis (Lehtinen and Walsh, 2011; Tiberi et al., 2012). Cajal-Retzius cells are another source of Fgf as well as Reelin, which sustains neurogenesis and AP conversion to BP by interacting with the Notch pathway (Lakoma et al., 2011). Tgf β 1/2/3 from the meninges and cerebrospinal fluid promote neurogenesis (Tiberi et al., 2012). Retinoic acid secreted by the meninges also stimulates the production of neurons from RG (Siegenthaler et al., 2009).

The role of Wnt signaling in neurogenesis is more controversial. Several studies, in which Wnt proteins and/or other Wnt pathway components were manipulated, reported a positive effect of Wnt signaling in AP self-renewal and an anti-differentiative function (Chenn and Walsh, 2002). Other studies showed instead a pro-neurogenic role of the canonical Wnt pathway (Hirabayashi et al., 2004). These apparently contradictory results have been integrated into a largely accepted model that proposes that Wnt has a proliferative effect on AP, while inhibiting the transition to BP, and a neurogenic effect on BP, promoting their differentiation into neurons.

1.1.4.2 Intrinsic factors regulating neurogenesis

In addition to extracellular factors, neurogenesis is influenced by cell-intrinsic factors such as proneural transcription factors, Notch signaling and cell-cycle regulators.

Proneural genes

A class of transcription factors from the bHLH (basic-helix-loop-helix) family, the proneural genes, is responsible for the activation of a neurogenic differentiation program, promoting the exit from the cell cycle and the inhibition of astroglial differentiation (Wilkinson et al.,

2013; Imayoshi and Kageyama, 2014). The main proneural genes are Neurogenin (Ngn) 1 and 2, expressed in the dorsal telencephalon, and Mash1, expressed in the ventral telencephalon and, at low levels, in the hippocampal primordium. The proneural genes form heterodimers with ubiquitously expressed transcription factors of the bHLH E family, recruit the p300/CREB-binding protein complex and bind to E-box sequences upstream of neurogenic genes (NeuroD1, NeuroD2 and NeuroD6), thus promoting their expression and consequently the differentiation into neurons (Wilkinson et al., 2013; Imayoshi and Kageyama, 2014). The mammalian Ngn2 acquired a tyrosine in position 241 during evolution and its phosphorylation is necessary for the pyramidal shaping of glutamatergic neurons and their correct migration in the cortical layers (Hand et al., 2005).

Other regulators of neural progenitors in the mammalian cortex are the Id (inhibitors of differentiation or inhibitors of DNA binding) proteins. Id proteins have a HLH domain but lack the basic domain and are therefore unable to bind directly to DNA (Tzeng, 2003). They compete with proneural genes for binding to the E proteins, thus preventing neurogenesis, and sustain proliferation by potentiating Notch signaling or by indirectly promoting the expression of positive regulators of the cell cycle (Ruzinova and Benezra, 2003; Imayoshi and Kageyama, 2014).

Notch signaling and lateral inhibition

A well conserved mechanism to regulate differentiation is lateral inhibition mediated by the Notch pathway (Gaiano and Fishell, 2002). When the Notch transmembrane protein is bound by its receptor Delta-like 1 (Dll1) from a neighboring cell, its intracellular domain is cleaved (Notch intracellular domain, NICD). Upon cleavage, NICD forms a complex with the DNA binding protein RBPjk, translocates to the nucleus and activates the expression of several target genes. The main targets of the NICD-RBPjk complex in the mammalian developing nervous system are the bHLH Hes genes, Hes1 and Hes5, which function as repressors of transcription of cyclin inhibitors and proneural genes (Hatakeyama et al., 2004). Therefore, cells in which the Notch pathway is activated keep a radial glial identity (Gaiano et al., 2000). Hes1 has the ability to repress its own promoter leading to an oscillatory expression of the Hes1 protein. This oscillatory behavior stimulates a complementary oscillatory expression of Ngn2 and Dll1 in AP during the proliferative phase (Shimojo et al., 2011). In this way, AP are able to induce the Notch pathway reciprocally, thus keeping the AP in a stem-like state and inducing expansion of the progenitor pool. During the neurogenic period, Ngn2 and Dll1 are expressed by BP and neurons in a sustained manner which activates the Notch pathway in AP. The AP undergoes asymmetric

division and the newborn BP and neurons activate Notch signaling in the AP while migrating along the radial fibers (Shimojo et al., 2011). Several factors modulate responsiveness to Notch signaling including Reelin (Gaiano, 2008; Keilani and Sugaya, 2008; Lakoma et al., 2011) and Fgf signaling (Rash et al., 2011), the presence of the Notch-repressor Numb (Bultje et al., 2009) or the presence of RBPjk (Mizutani et al., 2007).

Factors regulating AP to BP transition

As explained above, neurogenesis can be achieved by direct differentiation of AP into neurons or through a previous expansion of neural progenitors via the generation of BP. Some transcription factors that regulate the transition from AP to BP have been identified.

The insulinoma-associated 1 (Insm1) gene starts to be expressed in the cortex at E10.5 at moderate levels in AP and at higher levels in BP, regardless of their neurogenic or non-neurogenic fate. Insm1 is required for the BP specification and self-renewal since its ablation leads to a reduction of SVZ and CP compartments and erroneous molecular specification of BP. Consistently, transient overexpression of Insm1 in the developing mouse cortex increases the pool of self-renewing BP (Farkas et al., 2008).

Another key factor for BP generation and self-renewal is the regulation of the cell cycle. BP are characterized by a longer G1 phase than AP (Arai et al., 2011) and transient overexpression of cell-cycle regulators CyclinD and Cdk4 in the developing mouse cortex has been shown to prevent G1 lengthening, leading to an increase of BP (Lange et al., 2009).

Despite this, a definite answer to what the key factors that control the decision of a neural stem cell to proliferate or differentiate are remains elusive, mainly due to the fact that proliferating and differentiating neural progenitors coexist in time and space in the developing mouse cortex. To overcome this limitation, the Calegari group developed a double reporter mouse line that allows the isolation of self-renewing and differentiating progenitors as well as neurons, the Btg2^{RFP}/Tubb3^{GFP} mouse line.

1.1.5 A novel tool to investigate cell-fate determination in the central nervous system: the Btg2^{RFP}/Tubb3^{GFP} mouse line

As discussed in the previous sections, both AP and BP can undergo proliferative and differentiative divisions, although to a different extent. In 1999, Btg2 (also known as Tis21) was identified as a marker of differentiating progenitors being expressed in AP and BP undergoing neurogenic divisions (Iacopetti et al., 1999). While the Btg2 messenger RNA (mRNA) is rapidly degraded after mitosis, the protein is retained in newborn neurons. In 2013, the group of Federico Calegari developed a double transgenic mouse line that expresses the red fluorescent protein (RFP) under the control of Btg2 and the green fluorescent protein (GFP) under the control of Tubb3, a marker of neurons (Figure 1.6) (Aprea et al., 2013). In this way, it is possible to isolate proliferating progenitors (PP), differentiating progenitors (DP) and neurons (N) from the developing mouse cortex by fluorescent activated cell sorting (FACS) based on the combinatorial expression of RFP and GFP: PP (RFP⁻/GFP⁻), DP (RFP⁺/GFP⁻) and N (RFP⁺/GFP⁺).

At the peak of neurogenesis in the mouse neocortex (E14.5), PP comprise mostly proliferating AP (about 60% of AP in the VZ), and to a lesser extent proliferating BP, while DP include differentiating AP (40% of AP in the VZ) and mostly BP (80% of BP in SVZ). Transcriptome analysis confirmed an enrichment of proliferating genes and markers of AP in PP, differentiating genes and BP-markers in DP and neural genes in N.

The Btg2^{RFP}/Tubb3^{GFP} mouse line demonstrated itself to be a powerful tool to study factors involved in neurogenic commitment, allowing for the identification of several uncharacterized transcripts with a role in corticogenesis, such as non-coding RNAs (Aprea et al., 2013, 2015; Artegiani et al., 2015), circular RNAs (Dori et al., 2019) and micro-RNAs (Dori et al., 2020). The diversity of transcripts involved in corticogenesis highlighted by these studies prompted us to investigate, in the context of neural development, an additional source of transcript variability: alternative splicing.

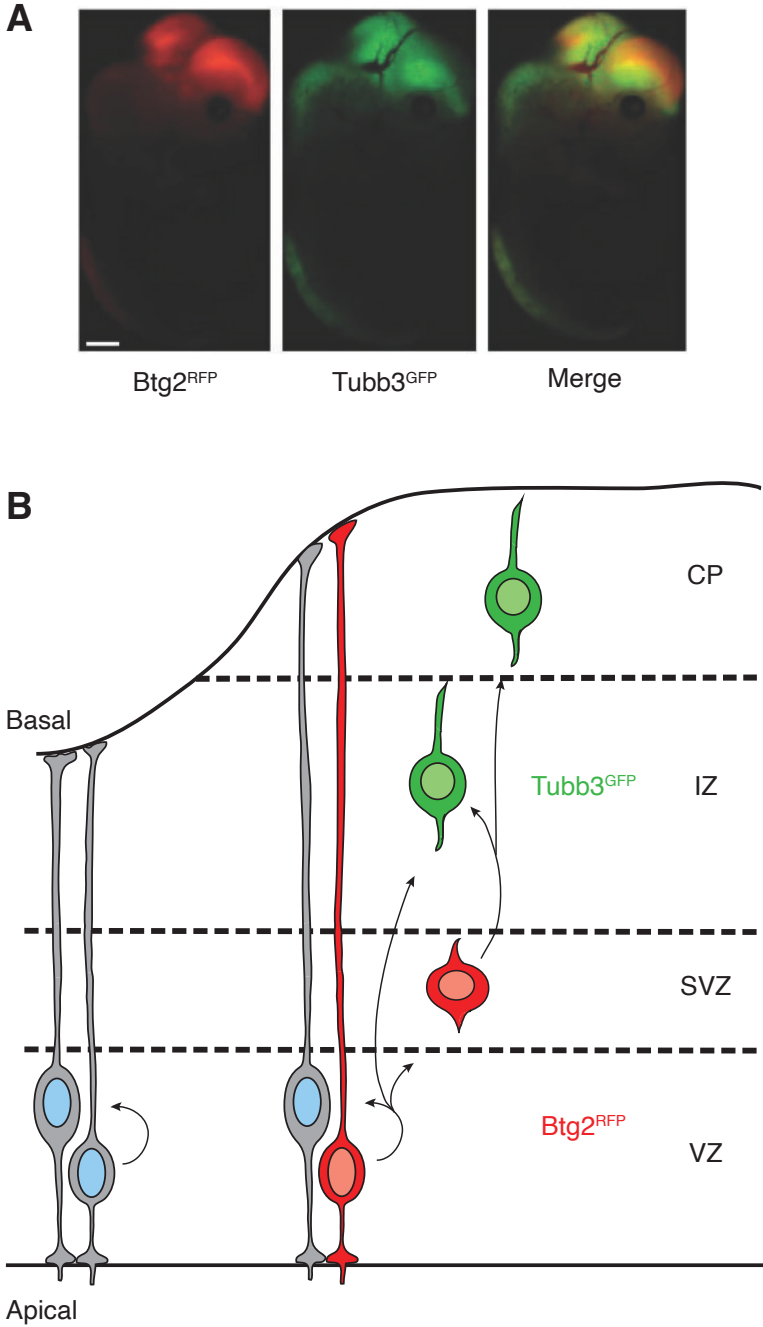


Figure 1.6 Btg2^{RFP}/Tubb3^{GFP} mouse line. A: whole-mount picture of an E14.5 Btg2^{RFP}/Tubb3^{GFP} embryo showing individual and merged RFP and GFP channels. B: schematic representation of Btg2^{RFP} in differentiating progenitors (DP, red) and Tubb3^{GFP} in newborn neurons (N, in green) in the Btg2^{RFP}/Tubb3^{GFP} mouse developing neocortex (Adapted from Aprea et al., 2013).

1.2 Alternative splicing: an additional level of genomic regulation

As exemplified in the previous section, neurogenesis is characterized by dramatic changes in gene and protein expression. Alternative splicing, the mechanism by which multiexonic genes produce mRNA encoding different genomic information, represents an additional level of regulation of gene expression. Protein isoforms resulting from alternative splicing can differ in functional domains, DNA- or RNA-binding domains, localization, sites of post-translational modifications or in protein-protein interactions (Stamm et al., 2005; Kelemen et al., 2013). Several studies have reported that splicing affecting protein regions with an undefined structure, known as disordered regions, can also have important consequences, as such regions often modulate protein-protein interactions (Dyson and Wright, 2005; Buljan et al., 2012; van der Lee et al., 2014). Alternative splicing can affect also non-coding regions of the transcripts, such as the untranslated regions (UTR) at the 5' and 3' end modulating translation efficiency and transcript stability (Mockenhaupt and Makeyev, 2015). A special case of alternative splicing is represented by the retention of an intron in the mature mRNA, which then becomes susceptible to degradation by non-sense-mediated decay (NMD) (Celik et al., 2015).

One of the best known examples of splicing-dependent modification of protein function is given by the pro-/antiapoptotic protein Bcl-x: the long isoform Bcl-xl promotes cell survival while a shorter isoform Bcl-xs triggers cell death (Boise et al., 1993). In *Drosophila melanogaster*, alternative splicing of Dscam, encoding an axon guidance receptor, produces 38,000 alternative transcripts, expressing more protein variants than all the protein-coding genes in *D.melanogaster* (Schmucker et al., 2000).

Alternative splicing is rarely observed in unicellular eukaryotes, however it is widespread in complex vertebrates (95% of protein-coding genes in humans are capable of undergoing alternative splicing) (Kim E et al., 2007; Kornblihtt et al., 2013; Pan et al., 2008; Wang ET et al., 2008). The protein diversification induced by splicing could explain why organisms with similar repertoire of protein-coding genes show high phenotypic differences. Indeed, comparative genomic studies suggest that gene structures that favor the transition of exons from being always included (constitutive splicing) to being alternatively included (alternative splicing) arose during evolution (Gelfman et al., 2012). Consistently, primates show higher levels of alternative splicing relative to other species and a comparison of splicing profiles of physiologically equivalent organs in different species reflect the identity of a species more

closely than that of the organ type (Barbosa-Morais et al., 2012). Interestingly, this is in contrast to the highly conserved organ-specific changes in gene expression across species, suggesting that alternative splicing might have a prominent role shaping species-specific differences (Barbosa-Morais et al., 2012).

On the other hand, distinct subsets of alternative exons or “exon networks” are often conserved among species, subject to tissue-specific regulation, and enriched in tissue-related functions (Calarco et al., 2011; Ellis et al., 2012). Moreover, there is evidence that alternative splicing can be a key determinant of tissue-specific cell-fate commitment (Kornblihtt et al., 2013; Fiszbein and Kornblihtt, 2017).

At the basis of alternative splicing is the discontinuous structure of gene coding information in the genome. Multiexonic genes are organized in a modular fashion, with exons encoding the sequence that will be included in the mRNA separated by introns that are removed during the splicing process.

1.2.1 The splicing reaction

In vertebrates, the length of exonic sequences is in the range of 50-250 nucleotides (nt) while introns are usually much longer, on the order of a few thousand nt (Gelfman et al., 2012). Therefore, the RNA polymerase II (Pol II) that synthesizes mRNA in eukaryotic cells must distinguish the shorter exons from the much longer introns in a process known as “exon definition” (Collins and Penny, 2006; Hollander et al., 2016).

The splicing reaction is directed by four regulatory sequences: the donor and acceptor splice sites (SS), 5' SS and 3' SS respectively, the branch point sequence (BrP), and the polypyrimidine tract (PPT) (Watson, 2014). The splicing machinery is composed of five small nuclear ribonucleoproteins (snRNP), U1, U2, U4, U5, U6 that form the spliceosome complex. The first step of the splicing reaction involves the recognition and binding of the 5' SS by U1 while splicing factor 1 (SF1), in association with the U2 auxiliary factor (U2AF), recognizes the BrP. In a second step, U2 displaces SF1 and U2AF, and associates to the BrP. Subsequently, the two SS are brought together by a complex of U4, U5 and U6. In a first phosphoryltransferase reaction, U6 displaces U1 from the 5' SS and allows the connection of the intronic 5' end to the conserved adenine (A) of the BrP. The second phosphoryltransferase reaction triggers the junction of the two exons and releases the intron in a circular form, termed lariat, which is subsequently degraded (Figure 1.7) (Watson, 2014).

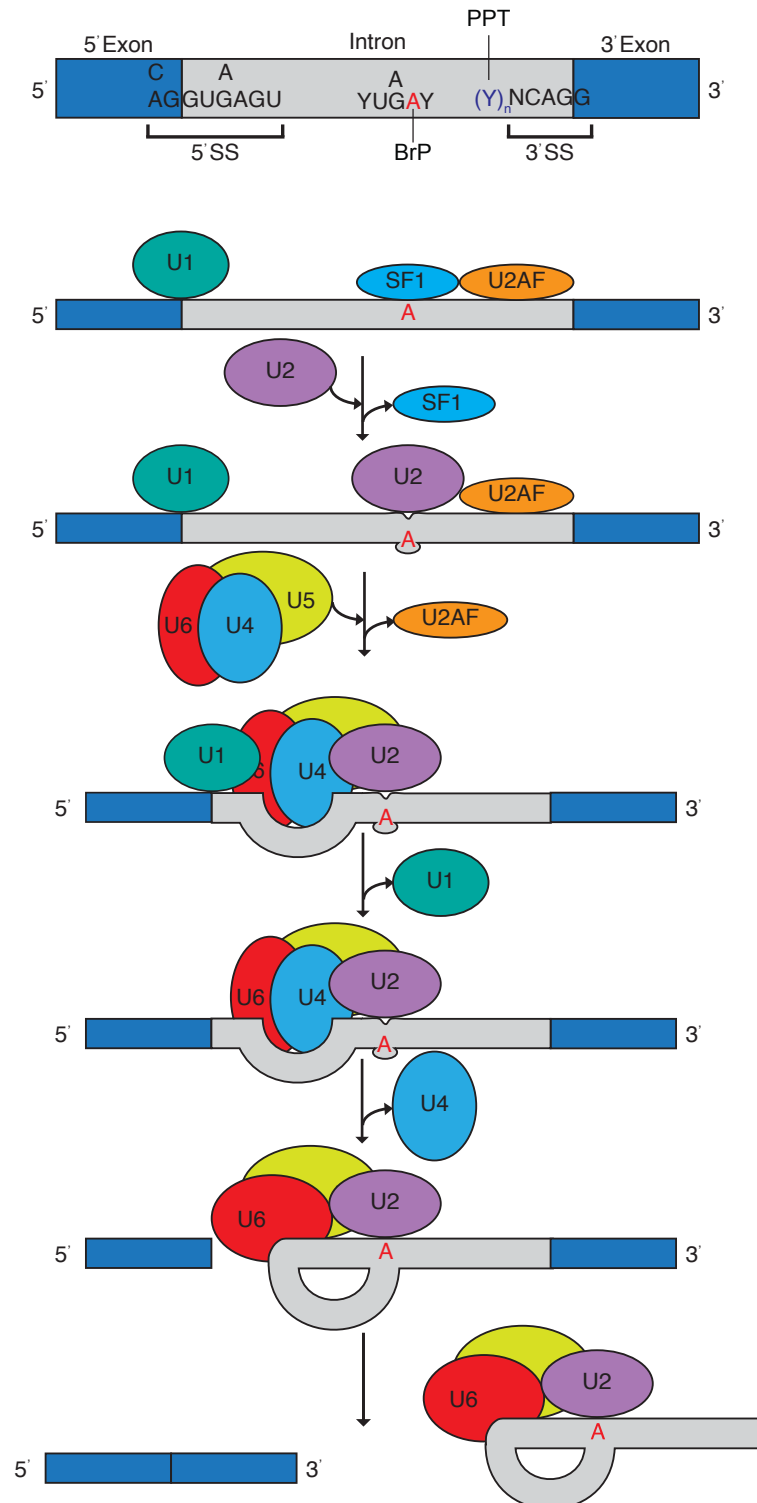


Figure 1.7 Splicing reaction. Components of the spliceosome complex recognize key sequences within the introns, the branch point (BrP) and the polypyrimidine tract (PPT), and the splice sites (SS) at the boundary between exons (blue) and introns (grey). In a multistep process the introns are removed, and the exons joined together, in the final mRNA. SF1=splicing factor 1; U2AF=U2 auxiliary factor.

1.2.2 What makes splicing alternative?

Concomitantly with the finding that gene coding regions are discontinuous in the genome, it was reported that multiple transcripts could arise from the same gene (Berget et al., 1977; Chow and Broker, 1978; Choi et al., 1980). While some exons are always included in the mature mRNA (constitutive exons), others are under a tight regulation and can be either included or excluded (alternative exons).

The splicing reaction can lead to different outcomes, conventionally referred to as “splicing events”. In addition to including/excluding entire exons within the transcript (exon skipping or cassette exons), the spliceosome can also choose between different SS of an exon (alternative 5' and 3' SS). Moreover, the splicing reaction can fail to properly remove introns, therefore retaining them in the final transcript (intron retention). Transcript diversity can also arise from alternative transcription mechanisms, such as the choice between alternative transcription start sites (alternative first exon) and alternative polyadenylation sites (alternative last exon) (Figure 1.8).

Although it is still not clear why some exons are alternatively spliced, several factors have been identified that regulate this process (Chen M and Manley, 2009). Some features such as the length of the exon and of flanking introns, in addition to the characteristics of the splicing regulatory sequences, are believed to contribute to the “alternative nature” of some exons. For instance, SS similar to the consensus sequence (“strong” SS) are more efficiently recognized than more divergent ones (“weak” SS) (Shapiro and Senapathy, 1987; Yeo and Burge, 2004). The BrP sequence is highly degenerated in mammals with only 2 nucleotides being conserved: the branch point A and a thymine (T) two positions upstream (TNA motif). More than one BrP can be present in the upstream intron, and the distance of the BrP from the 5' SS and from the PPT, as well as PPT features such as its length and pyrimidine content, all contribute to determine how efficiently an exon is recognized and, consequently, spliced in the nascent transcript (Corvelo et al., 2010; Li X et al., 2019). While a complete picture is missing, the strength of the splice regulatory sequence, kinetics of Pol II transcription (Fong et al., 2014; Saldi et al., 2016), splicing factor binding (Busch and Hertel, 2012; Kim KK et al., 2013; Weyn-Vanhentenryck et al., 2014; Saito et al., 2019), long-non coding RNAs (Hu et al., 2016; Romero-Barrios et al., 2018) and epigenetic modifications (Luco et al., 2011; Maunakea et al., 2013; Nieto Moreno et al., 2015; Cheng et al., 2017) have been found to be implicated in exon definition and alternative splicing decisions.

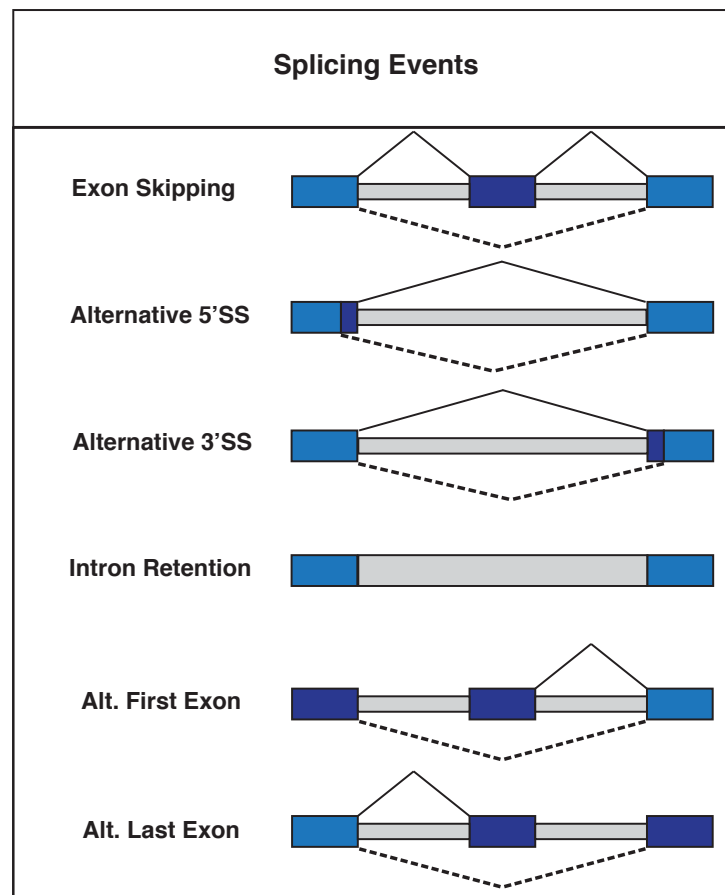


Figure 1.8 Classification of splicing events. Alternative splicing events are classified in different categories depending on the exons, partial exons or introns involved. Exons or exonic parts alternatively spliced are depicted as dark blue boxes, not spliced exons as light blue boxes, introns as grey boxes. Continued and dotted lines represent the splice choices leading to inclusion and exclusion of the alternative exonic sequence, respectively. Alt.=alternative.

1.2.3 Regulation of alternative splicing

1.2.3.1 Splicing factors

Splicing factors are proteins that facilitate or inhibit the inclusion of an exon in the mature mRNA by binding to regulatory sequences, i.e. splicing enhancers or splicing silencers, respectively.

Splicing factors recognize *cis* regulatory elements present in the gene (Busch and Hertel, 2012). Depending on their location and function, *cis* elements can be divided into four categories: exonic splicing enhancers (ESE), exonic splicing silencers (ESS), intronic splicing enhancers (ISE) and intronic splicing silencers (ISS) (Cáceres and Kornblihtt, 2002; Wang Z and Burge, 2008; Barash et al., 2010) (Figure 1.9 A).

Members of the SR (serine-arginine) family bind to the ESE and facilitate the inclusion of the exon (Tacke and Manley, 1999; Graveley, 2000; Cáceres and Kornblihtt, 2002). Silencer regulatory elements, both intronic and exonic, are usually recognized by heterogeneous nuclear ribonucleoproteins (hnRNP) that act to exclude the exon (Black, 2003; Martínez-Contreras et al., 2007). However, some hnRNP and hnRNP-like splicing factors, such as Nova and Rbfox, have been shown to bind to ISE and facilitate splicing (Ule et al., 2006; Chen M and Manley, 2009). SR proteins enhance splicing by promoting the recruitment of spliceosome components to the splice sites (Blencowe, 2000; Zhou and Fu, 2013). hnRNP may inhibit SS recognition by associating with splicing silencers in the vicinity of the SS and sterically blocking snRNP from accessing it (Martínez-Contreras et al., 2007; Chen M and Manley, 2009). Splicing inhibitors can also act by masking BrP or splicing enhancer sequences (Chen M and Manley, 2009). In some instances, silencer sequences are located 100-200 nt away from the alternative exon; in such cases, it has been proposed that hnRNP that are bound to sites spanning the alternative exon interact with each other, “looping out” the alternative exon and thus inhibiting spliceosome assembly (Martínez-Contreras et al., 2007; Chen M and Manley, 2009). Importantly, the same splicing factor binding motif can act as either silencer or enhancer, depending on its positioning. A well-known example is represented by the Nova2 target sequence, which enhances splicing when positioned in the downstream intron but leads to exon exclusion when positioned upstream, or within the exonic region (Ule et al., 2006).

1.2.3.2 Coupling with transcription

More than 30 years ago, Beyer and Osheim reported the first *in vivo* observation that splicing can occur co-transcriptionally (Beyer and Osheim, 1988). Since then, accumulating evidence supports the existence of coupling mechanisms between splicing and transcription (Kornblihtt et al., 2013; Kornblihtt, 2015; Naftelberg et al., 2015; Herzelt et al., 2017). Two main, non-mutually exclusive, models have been proposed: recruitment coupling and kinetic coupling (Kornblihtt et al., 2013; Naftelberg et al., 2015).

The recruitment coupling model suggests that splicing factors are brought to their target sites by the transcriptional complex. For example, the carboxy terminal domain (CTD) of the Pol II, a feature absent in other polymerases, is proposed to associate either directly or indirectly with splicing factors, thereby recruiting them to their target sequences and thus modulating splicing (Misteli and Spector, 1999; Muñoz et al., 2010; Nojima et al., 2018). In support of this model, truncated Pol II lacking the CTD shows aberrant splicing in fibronectin

mRNA (de la Mata and Kornblihtt, 2006). The promoter sequence also seems to play a role in directing splicing choices, as replacement of the promoter of one gene with that of another affects the splicing pattern of the nascent transcript (Cramer et al., 1997; Kornblihtt, 2005). Moreover, the mediator complex Med23, that links transcription factors bound to regulatory sequences with those that target core promoters, has been found to recruit several RNA processing proteins including splicing factors of the hnRNP family (Huang Y et al., 2012).

In the kinetic model, the elongation rate of Pol II transcription is fundamental to the decision of which splice sites are recognized and committed to be spliced (Kornblihtt et al., 2013; Naftelberg et al., 2015). In the original model of SS selection, the first SS to be synthesized is the first one to be committed to be spliced, in a “first come, first served” view (Aebi and Weissman, 1987). According to this model, a slow elongation rate would favor the recognition of weak SS upstream of stronger SS, increasing the inclusion of exons with weak SS. Slow Pol II mutants, sequences that induce Pol II pausing or drugs that reduce the elongation rate have indeed been shown to increase exon inclusion, whereas higher exon skipping was observed upon induction of more open chromatin states or overexpression of elongation-promoting transcription factors (Kornblihtt et al., 2013; Naftelberg et al., 2015). In other cases, however, slow elongation promoted exon exclusion (Dujardin et al., 2014), suggesting that the rate of Pol II elongation constitutes a “window of opportunity” for regulatory factors that either favor or suppress exon inclusion to be recruited on the nascent mRNA (Roberts et al., 1998). Recent findings have revealed an even more complex scenario in which distinct groups of exons respond differently to changes in Pol II elongation rate (Fong et al., 2014). The strength of SS, intron length and regulatory sequences flanking the alternative exons are very likely to play a role in the response to Pol II elongation rate variations.

In support of the splicing-transcription coupling, chromatin state and epigenetic modifications have been reported to influence splicing decisions. More compact chromatin states could act as roadblocks for Pol II, slowing down the elongation rate and affecting splicing decisions according to the kinetic model (Alló et al., 2010; Nieto Moreno et al., 2015). In addition to influencing the Pol II speed by changing chromatin state, histone modifications have been shown to recruit splicing factors and modulate exon inclusion (Luco et al., 2011; Kornblihtt et al., 2013; Naftelberg et al., 2015).

Nucleosome positioning is also important in directing splicing. Nucleosomes are generally associated with exons at a fixed ratio of one nucleosome per exon, while they present a

random distribution in introns. It has been thus suggested that nucleosomes influence the process of exon recognition by altering the Pol II elongation rate (Andersson et al., 2009; Tilgner et al., 2009; Herzelt et al., 2017).

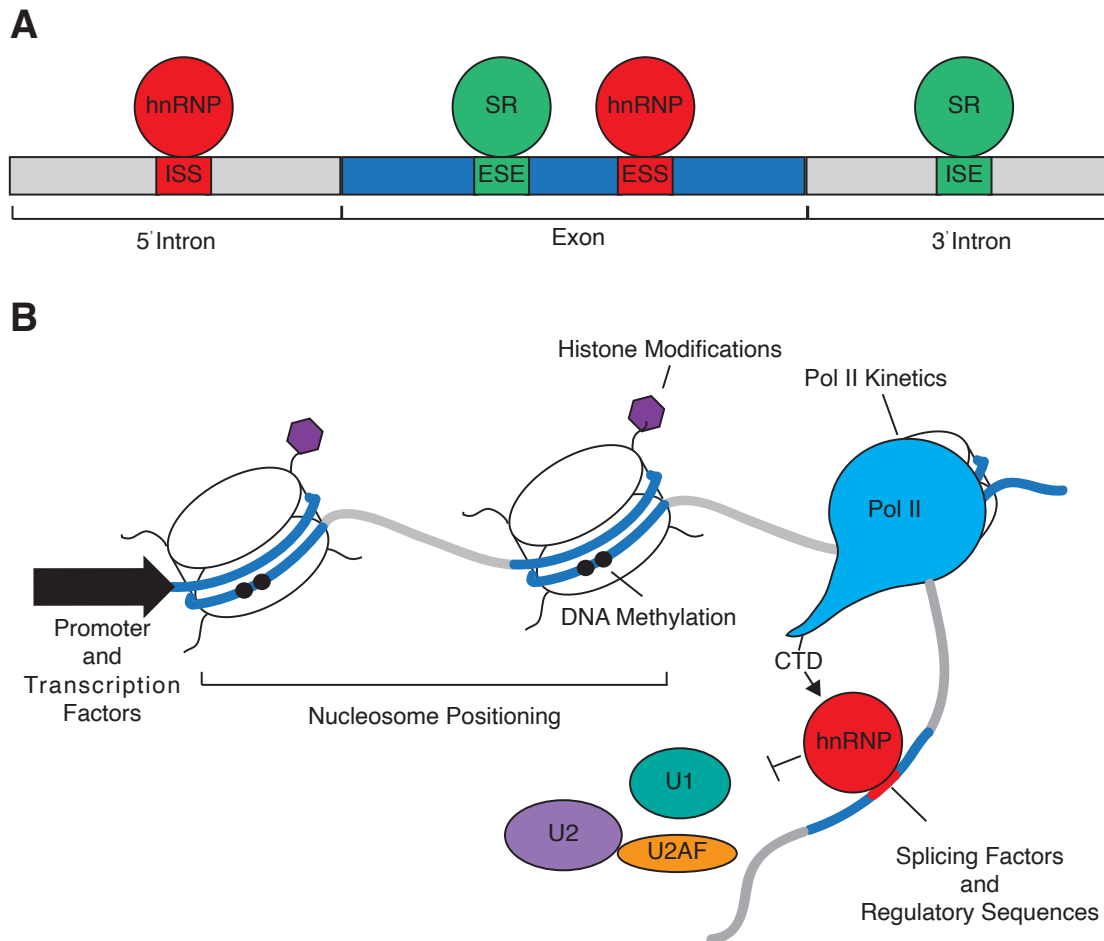


Figure 1.9 Splicing regulation. A: Splicing enhancer (green) and silencer (red) regulatory sequences (*cis* elements) bound respectively by Serine-arginine (SR, green) and heterogeneous nuclear ribonucleoproteins (hnRNP, red) splicing factors (*trans* factors). *Cis* regulatory splicing sequences are classified based on their ability to promote or inhibit splicing and on their exonic/intronic positioning. B: summary of factors potentially regulating splicing; blue and grey lines represent exons and introns, respectively. ESE=exonic splicing enhancer; ESS=exonic splicing silencer; ISE=intronic splicing enhancer; ISS=intronic splicing silencer; Pol II=RNA polymerase II; CTD=carboxy terminal domain.

Other epigenetic modifications such as DNA methylation, have been shown to regulate splicing decisions (Nieto Moreno et al., 2015; Yearim et al., 2015; Shayevitch et al., 2018), although systematic studies on the relationship between splicing and DNA methylation *in vivo* under physiological conditions are still missing. An additional level of splicing regulation is represented by epitranscriptomic modifications, such as adenosine methylation (m⁶A) on

the mRNA, however this field is still in its infancy (Adhikari et al., 2016; Hausmann et al., 2016; Lence et al., 2016).

1.2.4 The challenge to detect splicing

Most of our current knowledge about the function of splicing derives from studies on splicing factors. Conversely, the role of specific splice variants is largely unknown. This discrepancy is due to the challenge to identify and quantify the expression of individual isoforms that, by their nature, share the same sequence, albeit to a different extent. Indeed, current transcriptome sequencing approaches utilize bulk tissue extracts and fragment mRNA carrying out sequencing only on the complementary DNA (cDNA) fragments (Figure 1.10 A). Reads of short length in most of the cases could be assigned to several isoforms. A plethora of bioinformatic tools have been developed to detect alternative splicing (Alamancos et al., 2014; Hooper, 2014; Mehmood et al., 2019). The different methods can be divided in two categories, depending on the strategy adopted to investigate splicing: isoform-centric and exon-centric tools.

The first, isoform-centric, approach aims at reconstructing the expression of each isoform of a gene. Cuffdiff2 (Trapnell et al., 2012, 2013), Kallisto (Bray et al., 2016), Sailfish (Patro et al., 2014) and Salmon (Patro et al., 2017) have been developed for this purpose and offer an estimation of isoform expression from short-read sequencing. Cuffdiff2 applies statistical testing to assess the significance of changes in the proportion of isoform expression, relative to total gene expression, i.e. “isoform switching”, which is considered as a measure of alternative splicing. In addition to the identification of splicing processes, Cuffdiff2 allows for the detection of isoform variability resulting from differential promoter usage and estimates the impact of isoform switching on the protein output. Therefore, it is possible to distinguish three not mutually exclusive categories of isoform switching: change in transcription start site (TSS), change in coding sequence (CDS) and splicing proper (Splicing) between isoforms sharing the same TSS (Figure 1.10 B). As the final product of a gene is given by all the information carried by the mRNA isoforms, this approach allows for characterization of the major changes that alternative splicing induces to the information expressed by a gene. Given the high uncertainty in assigning each read to the correct isoform, bioinformatic tools that use this approach are prone to high error rates. To limit the number of false positives, isoform-centric tools are more conservative and detect a lower number of differentially spliced genes compared to bioinformatic tools that analyze splicing at the exon level (Alamancos et al., 2014; Hooper, 2014; Mehmood et al., 2019). Another

major limitation resides in the difficulty in validating the bioinformatic prediction by qPCR, as it is not always possible to design isoform-specific primers.

The second, exon-centric, method aims at calculating the differential inclusion of exons in a gene between different conditions or cell populations without any attempt to assign the differentially spliced exon to a specific isoform. It is possible to distinguish two main variations of this method: feature-based and event-based. In the feature-based approach, annotated transcripts are divided into counting features, such as exons or exon-exon junctions, and the sequencing data is used to estimate the abundance of each feature compared to all the gene's features. To take into account different exon boundaries derived from alternative splice site selection, some feature-based tools such as DEXSeq (Anders et al., 2012) use a simplified gene structure model in which each exon is divided into exon bins, depending on all the possible exon boundaries annotated in the reference transcriptome. Statistical testing is then applied to estimate changes in the representation of an exon bin compared to all the other exon bins of the gene. In analogy to the metrics used for differential gene expression analysis, DEXSeq reports a logarithm base 2 of the fold change (\log_2FC) of exon usage as a measure of differential splicing (Figure 1.10 C). Event-based tools, such as MISO (Katz et al., 2010), SUPPA2 (Trincado et al., 2018), rMATS (Shen et al., 2014) and VAST-tools (Irimia et al., 2014; Tapial et al., 2017) consider the specific splicing event involved, instead of focusing on the inclusion of an exon or exonic part. The different splicing events are annotated in libraries of known exon-exon junctions that can vary between different tools. One of the most comprehensive collection of splicing event libraries available today is the VastDB from the VAST-tools pipeline: species-specific libraries of exon-exon junctions have been assembled, based on a combination of gene annotation and sequencing data from various species and tissues, and subsequently employed as references for read alignment to identify and quantify alternative splicing events (Irimia et al., 2014).

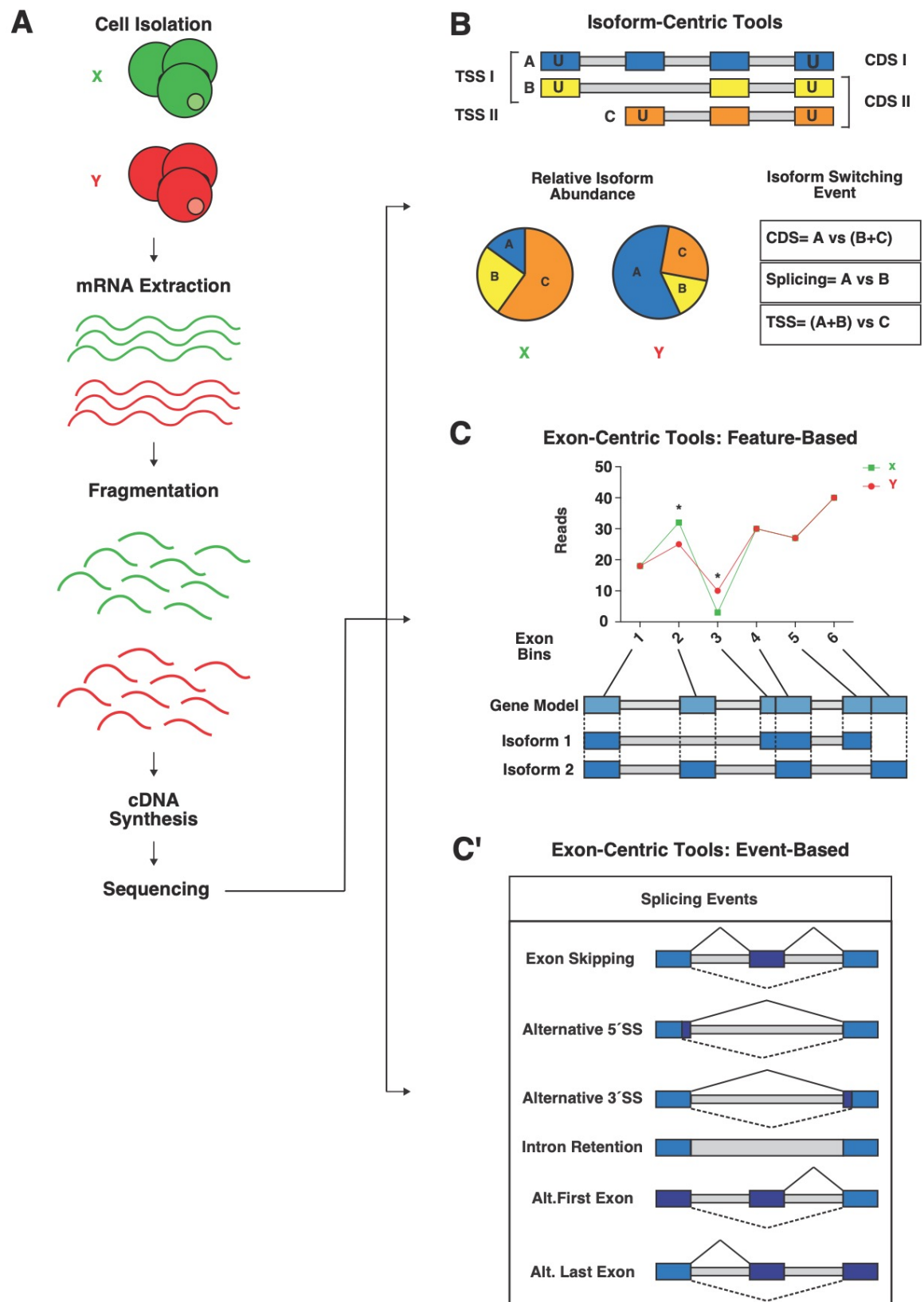


Figure 1.10 Analysis workflow of differential splicing from bulk short-read RNA sequencing (legend continued on the next page).

(Figure 1.10 legend continuation) A: messenger RNA (mRNA) is isolated from different cell populations (X and Y cell types), fragmented and converted in complementary DNA (cDNA) that is then sequenced. The data can be analyzed to detect splicing with isoform-based (B) or exon-based (C and C') tools. B, top: schematic representation of alternative isoforms of a gene. Colored blocks represent exons, grey blocks introns. U=untranslated region, CDS=coding sequences, TSS=transcription start site. Bottom: example of isoform switching event between X and Y cell populations. The total area of the pie chart represents the total gene expression while each colored fraction illustrates the corresponding isoform expression in proportion to the total gene expression. Isoform switching events are classified based on changes in the proportion of protein isoforms being expressed (CDS), different TSS usage (TSS) or Splicing. C: schematic representation of a hypothetical differential exon usage event between X and Y cell populations. The simplified gene model is depicted below the graph: light blue boxes represent exon bins, grey boxes introns. The annotated structure of the gene's isoforms is also shown: dark blue boxes represent exons, grey boxes introns. Dotted lines mark the exon boundaries that are used to define the exon bins. The graph reports the expression of each exon normalized to the expression of the whole gene. Exon bins that show differential usage are marked by asterisks. C': event-based tools use an internal library of exon-exon junctions to determine the specific splicing events. Exons or exonic parts that are alternatively spliced are depicted as dark blue boxes, not spliced exons as light blue boxes and introns as grey boxes. Continued and dotted lines represent the splice choices leading to inclusion and exclusion, respectively, of the alternative exonic sequence.

The sequencing data are then used to evaluate the inclusion/exclusion of a splicing event (Figure 1.10 C'). The level of inclusion is usually expressed as a percentage, the percent spliced in (PSI), with 0 denoting an event never being included in the analyzed transcripts and 100 denoting an event being always included. Variations in the PSI between two conditions or cell populations (Δ PSI) are indicative of changes in splicing. Conventionally, a Δ PSI of at least $\pm 10\%$ ($|\Delta$ PSI ≥ 10) is considered as a threshold for alternative splicing detection.

Different tools use various strategies to calculate the PSI. For instance, VAST-tools takes into account only junction reads, with reads spanning exon-exon boundaries between the considered exon and neighboring exons supporting the inclusion of the considered exon (inclusion reads, InR) and reads covering the junctions between exons flanking the considered exon supporting its exclusion (exclusion reads, ER) (Figure 1.11, top). Other tools also consider reads that map to exon bodies. One example is given by a variation of DEXSeq introduced by Schafer and collaborators (Schafer et al., 2015) (hereafter referred to as DEXSeq-PSI). The DEXSeq-PSI method can be regarded as a hybrid between feature-based and event-based tools. As with other event-based tools, instead of considering the exon usage with respect to all other exons of a gene, this method compares the read counts that support inclusion and exclusion of a given exon. Reads that map to the exon body, and junction reads that span exon-exon boundaries between the alternative exon and the neighboring ones, are regarded as InR. Conversely, split reads that span upstream or downstream exons, omitting the alternative exon, are counted as ER (Figure 1.11, top). As in the original DEXSeq method, exons are split into exon bins according to the exon boundaries annotated in the transcriptome, allowing for the use of a simplified

gene model as a reference instead of the junction library utilized by other event-based tools (Figure 1.11, middle). Both VAST-tools and DEXSeq-PSI calculate the PSI as the percentage of the ratio between InR and the sum of InR and ER (Figure 1.11, bottom).

Feature-based methods as well as the hybrid DEXSeq-PSI do not rely on prior knowledge of the linkage between exons and also allow for the capture of alternative transcription events, such as alternative first and last exons. Due to the way in which the PSI is calculated, DEXSeq-PSI can detect alternative first and last exon choices only when they do not coincide with the very first and very last exons annotated for a gene. On the other hand, feature-based tools are not able to classify the specific splicing events involved and the binning of the exons makes it difficult to identify the actual exons that are alternatively spliced. Event-based tools can give a more precise picture of the splicing profile of the analyzed sample and allow for an easier identification of the SS involved. However, these tools rely on known junctions annotated in their libraries, are therefore bound to a model of exon connectivity and in most cases cannot detect alternative first and last exons.

A limitation of all exon-based approaches is that it is not always possible to assign an exon to a unique isoform. Exons shared by multiple isoforms can be reported as differentially spliced due to subtle changes in the level of expression of each isoform, making it difficult to predict the impact that such splicing choices might have on the final main product of the gene.

Because of the complexity of the problem, novel bioinformatic tools for alternative splicing analysis from short-read sequencing are continuously being developed, while new sequencing technologies have been revealed to be beneficial for the improvement of transcript and splice junction annotation.

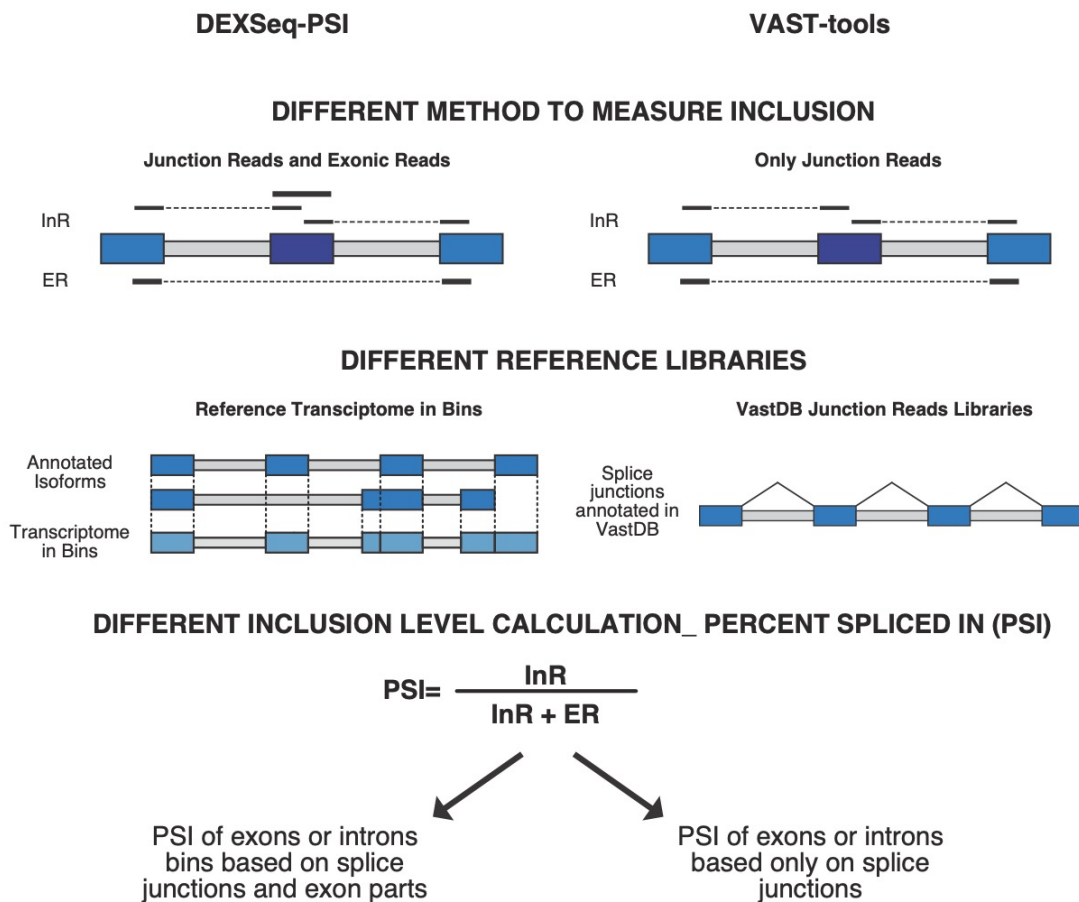


Figure 1.11 Different ways to calculate the percent spliced in (PSI). DEXSeq-PSI and VAST-tools use different methods to estimate the PSI. Top: DEXSeq-PSI takes into account reads that fall into exon bodies (continuous black lines) and junction reads (split black lines connected by dotted lines) that connect the spliced exon to its neighboring exons to calculate the exon inclusion (inclusion reads, InR). Split reads that map to positions upstream and downstream of the spliced exon but not to the exon itself are measures of exon exclusion (exclusion reads, ER). VAST-tools uses only the information derived by junction reads to measure InR and ER. Dark blue boxes represent spliced exons, blue boxes represent neighboring exons and introns are depicted as grey boxes. Middle: the two pipelines use different references. DEXSeq-PSI employs a simplified gene structure with exons divided into bins (light blue boxes) according to the exon boundaries (dotted lines) annotated for the different isoforms. VAST-tools uses internal libraries of exon-exon junctions (lines connecting the exons). Bottom: the PSI is calculated as the ratio between InR and the sum of InR and ER.

1.2.5 New sequencing technologies reveal a high transcriptome complexity

In the last decade, two major advancements in sequencing technology started to revolutionize our view of transcriptome complexity and of cell-fate determination: long-read sequencing and single-cell sequencing.

Long-read sequencing technology such as PacBio (Eid et al., 2009; Rhoads and Au, 2015) and Nanopore (Jain et al., 2018) allows for the reading of a full-length transcript and in latest versions, for the direct sequencing of RNA molecules without the need to first convert them into cDNA, enabling detection of RNA modifications. Long-reads can improve the current annotation of full-length transcripts and uncover the existence of novel isoforms and isoform features, such as alternative SS, TSS and polyadenylation sites, even in well-studied model organisms (Hardwick et al., 2019; Stark et al., 2019). However, caution must be taken when reporting new isoforms and isoform features as it has been shown that quite a considerable fraction of newly reported isoforms and isoform features are artifacts (Tardaguila et al., 2018). Moreover, the high error rate, the short-transcript bias, the limited depth of the sequencing, the high amount and quality of material required and the paucity of bioinformatic pipelines that are able to adequately handle these noisy data pose constraints to the use of such a technology, let alone its use in the quantification of splice variants (Hardwick et al., 2019; Stark et al., 2019). The combination of long-read data for a proper annotation of the actually expressed transcripts and of short-read data for the quantification of transcript features seems to be the most promising approach for faithful detection of splicing events (Tardaguila et al., 2018).

Of great importance, especially for developmental studies, is the possibility to obtain the transcriptional profile of individual cells by single-cell sequencing. Traditional bulk population sequencing can detect only the average expression signal of a group of cells, which could present different transcriptional profiles when analyzed individually. Since its first appearance in 2009 (Tang et al., 2009), single cell RNA sequencing (scRNA-Seq) has gained increasing attention and has proven to be a powerful tool to dissect tissue heterogeneity, with the possibility to identify novel cell types and to delineate lineage relationships (Wen and Tang, 2016; Kumar et al., 2017). Several methods have been developed in the single-cell field aiming to improve sensitivity and specificity of the RNA sequencing, to add spatial resolution and to integrate RNA sequencing with other molecular readouts such as DNA sequencing, detection of epigenetic modifications and single-cell

proteomics (Stuart et al., 2019; Stuart and Satija, 2019). Full-length transcript sequencing approaches for scRNA-Seq are currently available (Picelli et al., 2013; Fan et al., 2015; Sheng et al., 2017) and have been useful for the discovery of novel transcripts, although they show a bias for longer genes (Phipson et al., 2017). Despite a rapid development of single-cell technology, this method still suffers from important limitations. First of all, the availability of material is very low, considering that the average cell contains only ~10 picograms (pg) of total RNA, of which only ~0.1 pg is mRNA. Therefore, extensive amplifications of the cDNA are unavoidable and can lead to artifacts in the detected amount of different cDNA molecules, due to differences in amplification efficiencies (Wang Y and Navin, 2015). The introduction of unique molecular identifiers (UMI) into the primer oligonucleotide used for reverse transcription allows for *a posteriori* corrections of these artifacts, since all the amplicon cDNAs harboring the same UMI are derived from the same original cDNA molecule and consequently can be compressed to a single primer-mRNA hybridization event (Islam et al., 2014). The limited amount of material imposes also a restriction on the detection and quantification of the expressed transcripts. It is estimated that available scRNA-Seq technology allows for the detection of only a maximum of ~30% of the transcripts actually present, with medium-low expressed and inefficiently retrotranscribed genes usually being lost (Svensson et al., 2017; Zheng GX et al., 2017). Importantly, different cell types are isolated with variable efficiencies and the tissue dissociation and sequencing library preparation protocols can introduce additional artifacts in the cell-type representation from a tissue and in gene expression. Moreover, the highly noisy nature, amount and complexity of data to be handled, poses computational challenges and requires the development of bioinformatic strategies different from the ones adopted for analysis of bulk RNA-seq data (Chen G et al., 2019; Kulkarni et al., 2019). Particularly, bioinformatic tools for the detection of splicing from scRNA-Seq are limited to the splicing profiling of highly expressed genes, focus on few exons or splice-junctions and capture only the predominant isoforms, while less abundant isoforms are lost (Song et al., 2017; Vu et al., 2018).

A combination of different sequencing strategies seems promising in revealing cell-type-specific isoform variability. Gupta et al. (2018) achieved single-cell resolution of isoform expression in mouse cerebellar cell-types by performing scRNA-Seq short-read sequencing and bulk RNA long-read sequencing from cDNA molecules retrotranscribed with UMI-barcoded primers. This approach allowed for the detection of cell-type-specific isoforms, several of which were previously unknown, thus also improving the annotation of the mouse transcriptome. However, this method requires multiple replicates to achieve a reliable

quantification, increasing the experimental costs. Moreover, the study was limited to a few thousand cells and the authors warn that increasing the number of cells sequenced would decrease both sensitivity and specificity (Gupta et al., 2018).

In conclusion, while long-read and scRNA-sequencing can improve isoform annotation and cell-type-specific isoform detection, a more reliable quantification of alternative splicing events requires high quality deep sequencing data that, as of today, can be obtained only by short-read bulk tissue sequencing. Regardless of the sequencing technology used, studies addressing splicing and isoform detection revealed an unexpected isoform variability, identified tissue and cell-type-specific splice variants and highlighted various degrees of splicing usage in different tissues with brain, testes and liver showing the highest levels of alternative splicing in humans (Yeo et al., 2004).

1.2.6 Splicing in nervous system development

The mammalian nervous system is characterized by an extensive use of alternative splicing, (Yeo et al., 2004). Several studies have highlighted the importance of a proper splicing pattern for neuronal maturation, migration, adhesion, identity specification and synapsis formation (Grabowski, 2011; Norris et al., 2014; Raj B and Blencowe, 2015; Liu J et al., 2018; Su et al., 2018). Due to the challenging task of identifying individual splice variants by the available sequencing technologies, and since the expression of specific sets of splicing factors in different tissues is thought to be at the basis of tissue-specific splicing patterns (Grosso et al., 2008; Baralle and Giudice, 2017), previous studies have focused on the role of splicing factors, while the function of individual isoforms is largely unknown.

A neural-specific splicing factor of the SR family, nSR100 (also known as Srrm4), seems to act specifically in neuronal cell differentiation (Calarco et al., 2009). This splicing factor is expressed only in brain sub-regions and sensory organs and its depletion in a mouse neuroblastoma cell line leads to the exclusion of several neural-enriched exons (Norris and Calarco, 2012; Raj B and Blencowe, 2015), while mice lacking nSR100 show abnormal neocortical development with defective neuritogenesis (Quesnel-Vallières et al., 2015) and autistic-like behaviour (Quesnel-Vallières et al., 2016). One of the targets of nSR100 is Rest, which acts as a repressor of genes involved in neurogenesis (Schoenherr and Anderson, 1995). nSR100 mediates the inclusion of an exon in the Rest transcript which introduces a premature stop codon, resulting in the production of a truncated protein devoid of repressor activity (Raj B et al., 2011). In contrast to other SR proteins, which are

conserved across metazoans (Barbosa-Morais et al., 2006), the nSR100 gene is highly conserved in vertebrates but absent in invertebrates. The recent evolution of the nSR100 protein was suggested to have contributed to the increased complexity of the nervous system in higher vertebrates (Calarco et al., 2009). However, recent studies revealed that most invertebrate species possess a single gene orthologous to the vertebrate paralogs Srrm2, Srrm3 and nSR100/Srrm4. This Srrm2/3/4 gene gives rise to several isoforms, some of which share protein domains with the Srrm4/nSR100 and have a neural-specific expression profile (Torres-Méndez et al., 2019).

Other important splicing factors for the control of splicing regulation in the developing nervous system are the polypyrimidine tract binding proteins Ptbp1 and Ptbp2. The main role of Ptbp2 is to prevent premature inclusion of exons that characterize isoforms of mature postnatal neurons. Depletion of Ptbp2 leads to shrinkage of the neural progenitor pool, premature neurogenesis and misregulation of alternative splicing events in genes important for cytoskeletal remodeling, neurite outgrowth and synapse formation (Licatalosi et al., 2012; Li Q et al., 2014). These defects are partially rescued by overexpression of Ptbp1 (Vuong et al., 2016). Although these two splicing factors are very similar and present great redundancy in their binding sites and regulated exon sets, they modulate exon inclusion with different efficiencies, present some specificity in their targets, have different affinities for co-regulators, such as Raver1, and show opposite expression patterns (Keppetipola et al., 2012; Vuong et al., 2016). During mouse embryonic development, Ptbp1 is at first broadly expressed in non-neuronal tissues and in neural progenitors (Makeyev et al., 2007) until it is replaced by its paralogous Ptbp2, that maintains a sustained expression until postnatal day 14 (P14) (Licatalosi et al., 2012). Ptbp1 directly induces the exclusion of exon 10 in Ptbp2, in the absence of which the Ptbp2 transcript is degraded by NMD (Spellman et al., 2007). The action of Ptbp1 is counteracted by nSR100 which promotes the inclusion of some of the exons repressed by Ptbp1 (Raj B et al., 2014), among them the Ptbp2 exon 10 (Calarco et al., 2009). In developing neurons, the micro RNA miR124 silences Ptbp1 (Makeyev et al., 2007) allowing the inclusion of exons targeted by nSR100 and of exon 10 in the Ptbp2 transcript, while in non-neural tissues devoid of nSR100 and of miR124, Ptbp1 keeps repressing its target exons and Ptbp2. In this way, Ptbp1 contributes to the restriction of the neural-specific splicing program to cells of the neural lineage.

Splicing mediated by Nova proteins is important for the maintenance of neurons in the brain stem and the spinal cord, synapse formation, neuronal migration and correct cortical lamination (Norris and Calarco, 2012; Raj B and Blencowe, 2015). The Nova2 paralog

Nova1 is important for the maintenance of motoneurons and for the formation of neuromuscular junctions, since mice deficient in Nova1 show apoptosis in the brain stem and motoneurons (Jensen et al., 2000).

Different neuronal subtypes show specific splicing profiles and the same splicing factor can mediate different splicing events in distinct neuronal subpopulations. By tagging and cross-linking transcripts targeted by Nova2 in a cell-specific manner, Saito and collaborators showed that Nova2 regulates different sets of splicing events in excitatory and inhibitory neurons in mouse (Saito et al., 2019). Neuronal type-specific isoform expression was also observed in the worm *Caenorhabditis elegans* by monitoring splicing patterns with a two-color splicing reporter (Norris et al., 2014).

Synapse formation involves the expression of synaptic scaffolding proteins such as the post-synaptic density protein 95 (Psd95) (Cho et al., 1992) and adhesion proteins such as neuexins and neuroligins, located pre- and post-synaptically, respectively (Ushkaryov et al., 1992; Dean and Dresbach, 2006). Ptbp1 and 2 have been shown to regulate Psd95 splicing, leading to the degradation of its transcript by NMD. As development proceeds miR124 inhibits Ptbp1 while Ptbp2 is downregulated, thus removing the brake on Psd95 expression and allowing for the development of synapses (Zheng S et al., 2012). Neuexins, and to a lesser extent neuroligins, undergo alternative splicing. The three neuexin genes present a complex structure and have the potential to give rise to thousands of isoforms by combinatorial use of alternative SS and alternative promoters (Ullrich et al., 1995; Treutlein et al., 2014). It is likely that the various isoforms of neuexins and neuroligins have different affinities for each other and act as a “code” to mediate synaptic adhesion of different types of neurons (Boucard et al., 2005; Treutlein et al., 2014). Several RNA binding proteins including Ptbp2 and Sam68 have been identified as regulators of splicing of these two classes of adhesion molecules (Norris and Calarco, 2012).

Alternative splicing can also be used by neurons to “fine-tune” protein expression depending on environmental stimuli and thus modulate their responses in neural networks (Norris and Calarco, 2012; Hermey et al., 2017). Activity-dependent splicing events have been observed in synaptic scaffolding proteins, as well as in neurotransmitter receptors and ion channels (Norris and Calarco, 2012; Hermey et al., 2017). Depolarization of neural cells *in vitro* has been shown to lead to altered inclusion of a group of alternative exons in genes involved in signaling, transcriptional regulation and vesicle transport (Quesnel-Vallières et al., 2016). A specific splicing program mediated by Nova2 could also be important for the

maintenance of synaptic strength, as Nova2 knock-out mice show impairment in long-term potentiation, a persistent strengthening of synapses that is at the base of memory (Huang CS et al., 2005).

These examples demonstrate that alternative splicing plays a critical role in shaping correct neuronal networks. Increasing attention has been directed to the elucidation of the impact of aberrant splicing on the etiology of neurological disorders (Licatalosi and Darnell, 2006) such as schizophrenia (Takata et al., 2017), Alzheimer’s disease (Raj T et al., 2018) and autism spectrum disorder (Irimia et al., 2014; Weyn-Vanhentenryck et al., 2014; Quesnel-Vallières et al., 2016). Indeed, a widespread misregulation of splicing of neural-enriched exons, several of which are characterized by short length and therefore referred to as microexons, has been reported in neurodevelopmental disorders (Raj B and Blencowe, 2015).

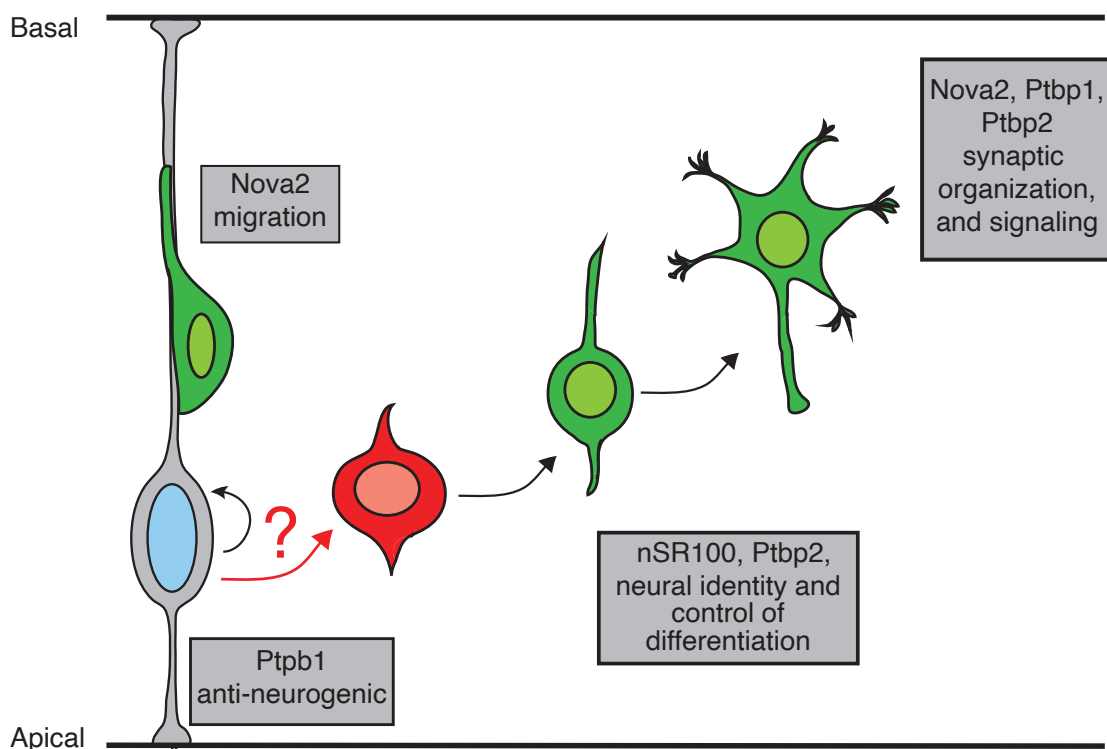


Figure 1.12 Splicing in the nervous system. Known splicing factors regulating development, migration, maturation and function of neurons (green cells). Nova2 regulates the splicing of genes involved in migration of newborn neurons along the radial glial fibers (left) and of genes important for synapse organization and synaptic transmission. nSR100 and Ptpb2 are splicing factors specific for neural tissues and are involved in neuronal maturation, whereas Ptpb1 is expressed in undifferentiated neural stem cells (grey cells) and non-neural tissues.

Microexons: when less is more

Of particular interest for the splicing pattern in neural tissues are microexons. There is no universal agreement on the definition of a microexon, with some researchers including only exons between 3 and 27 nt (Irimia et al., 2014; Quesnel-Vallières et al., 2016) and others including all exons shorter than 51 nt (Li YI et al., 2015). From a mechanistic point of view, microexons represent a peculiar case of alternative splicing since they are so short that the spliceosome cannot physically assemble at both the 3' and 5' SS (Dominski and Kole, 1991). When not constitutively included, microexons generally lack ESE (Li YI et al., 2015) being preferentially skipped when enhancing splicing factors are absent (Irimia et al., 2014; Yang and Chen LL, 2014; Quesnel-Vallières et al., 2015, 2016).

Although microexon regulation constitutes only 1% of all splicing events, microexons represent more than one third of neural enriched conserved spliced exons, with shorter microexons (3-15 nt) being more enriched than longer ones (Irimia et al., 2014). Sequencing studies have revealed the presence of an extensive neural-specific microexon program in several vertebrates as well as in the amphioxus *Branchiostoma launceolatum*, a non-vertebrate species (Torres-Méndez et al., 2019). Microexons are more often in frame and more conserved than longer neural enriched exons, even the intronic sequences flanking them have a higher degree of conservation than those flanking long neural-specific exons (Irimia et al., 2014; Li YI et al., 2015). Compared to longer exons, microexons are characterized by weaker 3' and stronger 5' SS and by stronger BrP and PPT in the upstream intron (Torres-Méndez et al., 2019). Microexon inclusion seems to depend on a dedicated exon definition mechanism, involving the splicing factors nSR100, Srsf11 and Rnps1. The intron upstream of several neural microexons contains a bi-partite ISE formed by UC repeats followed within 50 nt by a UGC motif positioned 2-20 nt from the microexon 3' SS. The presence of this ISE compensates for the lack of ESE, which would be difficult to accommodate in the short exonic sequence of microexons. Srsf11 binds the UC repeats and interacts with both Rnps1 and nSR100, the latter of which is bound to the UGC motif. This splicing factor complex interacts with components of the spliceosome, promoting its assembly (Gonatopoulos-Pournatzis et al., 2018). Recently, a dedicated domain for the catalysis of microexon splicing, the so-called enhancer of microexon domain (eMIC), has been identified in the nSR100 protein. This domain was acquired through alternative splicing events in bilaterian ancestors and subsequently lost in the nematode and platyhelminth clades. The eMIC interacts with factors involved in early spliceosomal assembly, SF1 and U2AF, to promote specific inclusion of microexons (Torres-Méndez et al., 2019). A separate class of microexons is enhanced by Rbfox, while Ptbp1 was found to

mediate microexon skipping (Irimia et al., 2014; Yang and Chen LL, 2014; Li YI et al., 2015; Raj B and Blencowe, 2015; Gonatopoulos-Pournatzis et al., 2018).

Another interesting feature of microexons is that they usually encode for unstructured regions or binding domains, thus their inclusion/exclusion is likely to change protein-protein interaction networks (Buljan et al., 2012; Irimia et al., 2014; Li YI et al., 2015). For example, the inclusion of a 6 nt microexon in the transcript of *Apbb1*, a nuclear adaptor protein, increases its binding affinity for the histone deacetylase *Kat5/Tip60* (Irimia et al., 2014). Likewise, microexon inclusion in *Zfyve27*, a membrane protein involved in vesicular trafficking in neurons, increases its interaction with partner proteins and promotes neurite outgrowth (Ohnishi et al., 2014). In the case of the histone H3K4 demethylase *Lsd1*, a 12 nt microexon is necessary for binding to the supervillin (*Svil*) protein with alternative splicing transforming *Lsd1* from a co-repressor to a co-activator that removes repressive histone marks from target genes involved in neurite morphology together with *Svil* (Laurent et al., 2015). Recent studies have shown a dynamic regulation of microexon splicing in response to neural depolarization *in vitro*, suggesting that they might also be involved in modulating the synaptic response (Hermey et al., 2017)

Taken together, this suggests a primary role for microexons splicing in shaping protein interaction networks fundamental for correct nervous system development and function.

1.2.7 Aims of the project

The formation of the mammalian neocortex, a remarkably complex tissue, requires a correct balance between proliferation of neural progenitors and differentiation into neurons during embryonic development. Several factors have been found to be implicated in this process. However, previous studies focused on genes that are activated or repressed in different cell types, ignoring transcript diversity generated by alternative splicing. The role of alternative splicing in neuron maturation and function is well established. In contrast, whether and how splicing decisions regulate neural cell-fate commitment is still unclear. This project aims to address this question by identification and manipulation of splicing events marking the neurogenic commitment. By taking advantage of the *Btg2^{RFP}/Tubb3^{GFP}* mouse line, the cell-type-specific transcriptome from PP, DP and N was previously obtained in our lab (Aprea et al., 2013, 2015). By analyzing these data with several bioinformatic tools, the splicing profiles of cells of the neurogenic lineage were characterized and, for the first time, splicing events that take place during neurogenic commitment and neurogenesis were identified.

The type of splicing events involved, their pattern of inclusion and the implication of such splicing choices in the final product of the gene were further investigated. Additionally, splicing factors and exon features potentially involved in the regulation of splicing choices in the developing murine neocortex were identified. *In vivo* assays to elucidate the role of different splice variants on the neurogenic commitment were also carried out by *in utero* electroporation of different isoforms in the developing mouse embryonic brain.

2 Materials and methods

2.1 Materials

2.1.1 Bacteria, cells, mouse strains

Bacteria, cell or mouse line	Supplier
One Shot™ Top10 E. coli	Thermo Fisher Scientific
C57BL/6J0laHsd	Biomedical Services (BMS) of the MPI-CBG Janvier Labs
Btg2 ^{RFP} /Tubb3 ^{GFP}	Biomedical Services (BMS) of the MPI-CBG

Table 2.1 Bacteria, cells, mouse strains

2.1.2 Vector

pDSV_mRFPnls (Lange et al., 2009)

T-vector pMD19 for TA-cloning (Takara Bio)

2.1.3 Primers

All primers were ordered from Eurofins Genomics

Primers for cloning

Underlined the site for the restriction enzyme. When not specified the primers were common to both isoforms of interest.

Gene	Isoform	Sequence	Name
Faim	Both Isoforms	5'-CTAAGA <u>ATTCG</u> ACTACGTCGTTGGGATCG-3'	Faim_F
		5'-CAATCTAG <u>ACTCCT</u> CGGCTCAGAACTC-3'	Faim_R
Wnt5b	ENSMUST00000117171	5'-GTAGCTAG <u>CATGG</u> TTGAGGCCGGCCATG-3'	W71_F
		5'-CTACGA <u>ATTCC</u> AGTCACTTACAGACATACTGGTCC-3'	W71_R

	ENSMUST00000118120	5'-GTACGAATTCAGTGCAGAGACCGGAGATG-3'	W20_F
		5'-GCATTCTAGATGTGGTGCAGTCACTTACAGAC-3'	W20_R
Celf5	ENSMUST00000118763	5'-CGAACTCGAGACGGCATGAAGGACCTG-3'	C63_F
		5'-GACTCTAGATCAAAGCTCACGAATCCGAAACAC-3'	C63_R
	ENSMUST00000119060	5'-GACTGAATTC CCCCATCTTCATCTACCCAAGATG -3'	C60_F
		5'-GACTCTAGATCAAAGCTCACGAATCCGAAACAC-3'	C60_R
Mkrn1	ENSMUST00000031985	5'-CTACCTCGAGTGTGTGTGGGATAAACAGTAATG-3'	M85_F
		5'-CAATCTAGACACGCAACGCTGCTATAG-3'	M85_R
	ENSMUST00000114822	5'-CGATGAATTC TAAACAGTAATGGCGGAGGCT -3'	M22_F
		5'-CAATCTAGAGACACTGAGTTCACTGTGGC-3'	M22_R
Tmcc2	ENSMUST00000045473	5'-CATCGAATTC TGCGGACCCACTTACACCATG -3'	T73_F
		5'-GAGTCTAGATCAGCTGGGCAGCAGCACAT-3'	T73_R
	ENSMUST00000142609	5'-GAACGAATTC CCAGGACTGTACAGATGAAGTCC -3'	T9_F
		5'-GTGTCTAGATCAGCTGGGCAGCAGCACAT-3'	T9_R
Porcn	ENSMUST00000077595	5'-GATGCTAGCCTGTGTGGGTCCACAATG-3'	P95_F
	ENSMUST00000082320	5'-GATGCTAGCGGATTTAAAGGGCCCGCTCG-3'	P20_F
	Both Isoforms	5'-GTATGAATTCGATGGTCATATGCCTCAGCC-3'	P_R

Table 2.2 Primers for Cloning

Primers for RT-qPCR

Gene	Isoform	Sequence	Name
Celf5	ENSMUST00000118763	5'-CCGCTGACGGAGTGAAG-3'	C63_F
		5'-CCGAACTGCTCGAACAGC-3'	C63_R
	ENSMUST00000119060	5'-CTCCACACGAGAGACCCTATG-3	C60_F
		5'-GCCGTTGAGAAGCTCGCTAC-3'	C60_R
Tmcc2	ENSMUST00000045473	5'-GAAGACGCTGCTTGCCTTCTG-3'	T73_F
		5'-CCAAACATGGAGCCCTCTGAGAG-3'	T73_R
	ENSMUST00000142609	5'-CTGCTCAGAGACAGTGGACCTAG-3'	T9_F
		5'-GCCACAAGATCTCCCTTGTCAC-3'	T9_R
	ENSMUST00000138717	5'-CAAAGAAGTGGTGTTCCTGCC-3'	T17_F
		5'-GACTCCACGGCCTCCTGGATATC-3'	T17_R
	Whole gene	5'-GACCCATCAGCCTGGATGTG-3'	Tmcc2_F
		5'-CTCCTGCTCGATCTTGATCTGC-3'	Tmcc2_R
Porcn	ENSMUST00000077595	5'-CAAAGCCAGGGGACCATG-3'	P95_F
		5'-CTGTCAGGTCCCATTCCAGGTG-3	P95_R
	ENSMUST00000082320	5'-CAAAGCCAGGTGGCTACGAG-3'	P20_F
		5'-GGTCTAGAGACTGTCAGGTCCC-3'	P20_R

	Whole gene	5'-GCTTCCTTTCTGCCAGCTCC-3'	Porcn_F
		5'-GCAGGCCGAAGCAGATGGTAAGA-3'	Porcn_R

Table 2.3 Primers for RT-qPCR

Primers for RT-PCR

Gene	Sequence	Name
Add1	5'- GGCTCTGAAGAGAACCTGGA-3'	Add1_F
	5'-TCATCAGGGGAGAGGTCAGG-3'	Add1_R
Abi2	5'- AGATGACATTGGACATGGAGTG-3'	Abi2_F
	5'- CAAGTGTCCCTTCCCTGAC-3'	Abi2_R
Apba2	5'- CATCCGAATGATGCAGGCTCA-3'	Apba2_F
	5'- AGACCTTGATCCTCTGGGTCG-3'	Apba2_R
Daam1	5'-CTCTGCCTACCAAAGACAGCA-3'	Daam1_F
	5'-TCTGAGCTCTCCGACCATCA-3'	Daam1_R
Faim	5'-CTGACTACGTCGTGGGATCG-3'	Faim_F
	5'- CTTGCCCTGATGTGGTCCCAT-3'	Faim_R
Git1	5'-ATCCCACAGATGGCTGACAG-3'	Git1_F
	5'-CGCTTGCAGCTTCTTCTTGG-3'	Git1_R
Gopc	5'-GCTTCATGCCAAGACTGGTCA-3'	Gopc_F
	5'-CCAACCTCGCCCCATAAACT-3'	Gopc_R
Neo1	5'-GGAAACAGACTGACTCACCAG-3'	Neo1_F
	5'-CAGGTCTGGTAGTCGGCTTC3'	Neo1_R
Parp6	5'-GGTCCCACATTGAGAATTGGCA-3'	Parp6_F
	5'-TGGTGTTTCATCCTGTTGTATCTCTG-3'	Parp6_R
Ptk2	5'-ACACATACACCATGCCCTCG-3'	Ptk2_A_F
	5'-GCTCAGGTACACGCCTTGAT-3'	Ptk2_A_R
Ptk2	5'-GCCAACAGTGAAAAGCAAGGC-3'	Ptk2_B_F
	5'-CGAGGGCATGGTGTATGTGTC-3'	Ptk2_B_R
Eef1a1	5'-ACAAGCGAACCATCGAAAAG-3'	Eef1a1_F
	5'-GTCTCGAATTTCCACAGGGA-3'	Eef1a1_R

Table 2.4 Primers for RT-PCR

2.1.4 Chemicals and buffers

Buffers for general use

Solution	Composition
Phosphate buffer saline (PBS)	137 mM NaCl 2.7 mM KCl 10 mM Na ₂ HPO ₄ 1.8 mM KH ₂ PO ₄ in H ₂ O – pH = 7.4
PFA 4%	1.3 M formaldehyde 100 mM Na ₂ HPO ₄ /NaH ₂ PO ₄ in H ₂ O – pH = 7.4
Sucrose solution	30% w/v sucrose in PBS
Tris-borate-EDTA (TBE) (10X)	0.89 mM Tris base 0.89 mM boric acid 20 mM EDTA in H ₂ O – pH = 8.0

Table 2.5 Buffers for General Use

For immunohistochemistry

Solution	Composition
Citrate buffer	4 mM sodium citrate 6 mM citric acid in H ₂ O – pH = 6.0
Quenching solution	0.1 M glycine in PBS – pH = 7.4
Blocking buffer	10 % donkey serum 0.3 % triton-X 100 in PBS
Incubation solution	3 % donkey serum 0.3 % triton-X 100 in PBS
DNA denaturalization solution	2 M HCl in H ₂ O
DAPI (1000X)	0.1 w/v DAPI in H ₂ O

Table 2.6 Buffers for Immunohistochemistry

Culture media

Medium	Composition
LB medium (CRTD media kitchen)	1% w/v tryptone 0.5% w/v yeast extract 171 mM NaCl in H ₂ O – pH = 7.0
LB agar	1.5% agar in LB medium
SOC medium	2% w/v tryptone 0.5% w/v yeast extract 8.56 mM NaCl 2.5 mM KCl 10 mM MgCl ₂ 20 mM glucose in H ₂ O – pH = 7.0

Table 2.7 Culture Media

2.1.5 Antibodies

Primary antibodies for immunohistochemistry

Antigen	Species	Supplier	Catalog number	Dilution
RFP	Rat	Chromotek	5F8	1:400
Tbr2	Rabbit	Abcam	ab183991	1:500

Table 2.8 Primary Antibodies for Immunohistochemistry

Secondary antibodies

For immunohistochemistry, IgG raised in donkey (against rabbit and rat) and DyLight-conjugated (Cy2 or Cy3) were used as secondary antibodies, all purchased from Jackson ImmunoResearch and used at a dilution of 1:500.

2.1.6 Kits and enzymes

Kit/Enzyme	Provider	Catalog
Q5 high-fidelity DNA polymerase	NEB	M0491S
Phusion high-fidelity DNA polymerase	NEB	M0530S
iQ™ SYBR® Green Supermix	Bio-Rad	170-8880
Restriction enzymes	NEB	

DNaseI	NEB	M0303S
SuperScript™ III Reverse Transcriptase	Invitrogen	18080-093
Antarctic Phosphatase	NEB	M0289S
T4 DNA ligase	NEB	M0202S
Quick RNA Mini Prep™	Zymo Research	R1054
QIAprep Spin Miniprep Kit	Qiagen	27106
EndoFree Plasmid Maxi Kit	Qiagen	12362
Invisorb Fragment CleanUP	STRATEC Biomedical	1020300200
Neural Tissue Dissociation Kit with Papain (P)	Miltenyi Biotec	130-092-628
7-AAD	BD Pharmigen	559925

Table 2.9 Kits and Enzymes

2.2 Methods

2.2.1 Animal experiments

All experimental procedures were performed according to local regulations and approved by the “Landesdirektion Sachsen” under the licenses 11-1-2011-41 and TVV 16/2018.

Animals for *in utero* electroporation and RNA extraction

Plugged C57BL/6J female mice were purchased from Janvier Labs and housed at the Biomedical Services Facilities (BMS) of the MPI-CGB under standard conditions: 12-hour light-dark cycles, $22 \pm 2^\circ\text{C}$ temperature, $55 \pm 10\%$ humidity, food and water supplied *ad libitum*. Mice were anesthetized with Isoflurane (Baxter) and sacrificed by cervical dislocation at E14.5 for RNA extraction from embryo cortices or at E15.5 for embryo brain histology.

Animals for RNA extraction for RT-PCR and qPCR

Btg2^{RFP}/Tubb3^{GFP} males were housed at the BMS under standard condition (see above) and time-mated with C57BL/6J females. Pregnant females were anesthetized and sacrificed as described above at E14.5. Brains of RFP/GFP double-positives embryos were dissected and the lateral cortices isolated by removal of meninges and ganglionic eminence. PP, DP and N were isolated by FAC-sorting.

Cell dissociation and FAC-sorting

Cells of the lateral cortices were dissociated using Papain-based Neural Tissue Dissociation Kit (Miltenyi Biotec) according to the manufacturer’s protocol. Cells were

resuspended in 500 μ l - 1 ml of ice-cold PBS and 7-AAD (BD Pharmingen, 1:100) or DAPI (1:1000) were added for discrimination of dead cells. Sorting was performed by BD FACSAria™ III (BD Biosciences) with previously described gating (Aprea et al., 2013).

In utero electroporation (IUE)

Purified plasmid DNA construct was resuspended in sterile PBS at a concentration of 1-4 μ g/ μ l. IUE was performed as previously described (Artegiani et al., 2012). C57BL/6J E13.5 pregnant dams received pain treatment by subcutaneous injection of 100 μ l of Carprofen (dosage of 5 mg/kg) one hour prior to surgery. Animals were then anesthetized with Isoflurane, the uterus was exposed and 1 μ l of DNA solution was injected into the embryo's left ventricle, followed by the application of 6 electric pulses (30V and 50 ms each at 1 s intervals) through platinum electrodes using a BTX-830 electroporator (Genetronics). After electroporation, the uterus was reembedded and the surgical incision was closed in two ways: absorbable suture (Vicryl Plus Ethicon) to close the inner muscle layer and surgical clips to close the outer skin layer. The wound was carefully cleaned with an antiseptic 10% iodine solution (Betadine). When applicable, pain treatment was reapplied 24 and 48 hours after surgery.

2.2.2 Molecular biology

Cell dissociation and FAC-sorting

Btg2^{RFP}/Tubb3^{GFP} E14.5 embryo brains were dissected and the lateral cortices dissociated using Papain-based Neural Tissue Dissociation Kit (Miltenyi Biotech) according to the manufacturer's protocol. Cells were resuspended in 500 μ l - 1 ml of ice-cold PBS and 7-AAD (BD Pharmingen, 1:100) or DAPI (1:1000) were added for dead cells discrimination. Sorting was performed by BD FACSAria™ III (BD Biosciences) with previously described gating (Aprea et al., 2013).

RNA extraction

For RT-PCR and RT-qPCR, total RNA was isolated using Quick RNA Mini Prep kit (Zymo Research) from Btg2^{RFP}/Tubb3^{GFP} embryo-derived PP, DP and N cells sorted as described above. RNA quality and integrity were assessed by Bioanalyzer (Agilent Genomics). Only RNA with RNA integrity values (RIN) above 8.0 was used. For cloning, total RNA was extracted from E14.5 wild type embryos whole brain or lateral cortices using RNeasy Mini Kit (Qiagen).

RT-qPCR

RT-qPCR was used to assess isoform changes in proportion. About 30 ng of total RNA from each sorted cell population were DNase-treated (NEB) and retrotranscribed using SuperScript™ III Reverse Transcriptase (Invitrogen) according to the manufacturer's protocols. To ensure that the transcripts were retrotranscribed without biases towards the 3' end, random hexamers were used for the retrotranscription reaction. For each target sequence, 1 ng of cDNA was used for qPCR amplification per technical replicate. iQ™ SYBR® Green Supermix (Bio-RAD)-based quantitative PCRs were carried out using gene- and isoform-specific primer pairs (Table 2.3) on a Stratagene MX 3005P machine with a standard qPCR program: 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec, 60°C for 30 min, 72°C for 1 min. Melting curves with 0.5 °C incremental steps from 55°C to 95°C were carried out at the end of the amplification reaction. In the case of Celf5, it was not possible to design primers that efficiently amplified a sequence common to the whole gene. The amplicon for the Celf5-C63 isoform revealed to be particularly high in CG content that rendered the amplification suboptimal and produced artifacts in the melting curve profile. The amplification required the addition of 2.5% formamide in the qPCR mixture to be efficient and the identity of the amplicon was confirmed by TA-cloning and sequencing. Results were analyzed using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) using the housekeeping gene Eef1a1 as a reference for normalization. Expressions of each splice variant (sv) relative to the parent gene and to other splice variants were calculated as following:

$$\text{Expression relative to the gene} = 2^{(Ct \text{ gene} - Ct \text{ sv1})} \text{ or } 2^{(Ct \text{ gene} - Ct \text{ sv2})}$$

$$\text{Expression relative to another isoform} = 2^{(Ct \text{ sv1} - Ct \text{ sv2})}$$

with sv1 and sv2 indicating splice variant 1 and 2, respectively.

RT-PCR

RT-PCR was used to assess changes in exon inclusion. About 15 ng of total RNA from each sorted cell population from each litter were DNase-treated (NEB) and retrotranscribed using SuperScript™ III Reverse Transcriptase (Invitrogen) according to the manufacturer's protocols. To avoid biases towards the 3' end of the transcripts, only random hexamers were used in the retrotranscription reaction. For each target sequence 1 ng of cDNA was used for PCR amplification. Amplification by PCR was carried out with Q5® High-Fidelity DNA polymerase (NEB) with standard conditions: 98 °C 30 sec, 27-33 cycles of 98°C 15 sec, 68°C 25 sec, 72°C 30 sec, 72°C 2 min. PCR products were resolved in 15% Polyacrylamide

gels in Protean® II xi cell for electrophoresis (Biorad) and visualized with SYBR Gold Nucleic Acids Gel stain (Thermofisher). For each target sequence the number of amplification cycles with highest sensitivity within the exponential amplification phase was determined experimentally.

Cloning

Splice variants were amplified using primers containing sequence recognition sites for restriction enzymes (Table 2.2). When possible, isoform-specific primers were used. In case of primers amplifying multiple isoforms, the isoform of interest was identified by size selection and sequencing. The splice variants were amplified from cDNA derived from total RNA of whole brain or lateral cortices from E14.5 wild type embryos. For all constructs, PCR products were in-column purified with Invisorb Fragment CleanUp kit (STRATEC Biomedical) and digested with the appropriate NEB restriction enzymes. Digested products were run on a 1-2% agarose gel and the band corresponding to the expected size was excised and purified using Invisorb Fragment CleanUp kit. Likewise, pDSV_mRFPnls vector backbones were digested with the appropriate restriction enzymes, dephosphorylated and gel purified. Ligations were carried out using T4 DNA ligase (NEB) with a 3:1 insert:vector molar ratio at 16°C overnight. 2-4 µl of ligation mix were transformed into Top10 E. coli competent cells (Thermo Fisher Scientific) according to the manufacturer's protocol and plated overnight on LB agar plates containing ampicillin. For each construct, colonies were inoculated in 5 mL LB Buffer supplemented with ampicillin and cultured overnight. Plasmid DNA was then purified using QIAprep Spin Miniprep Kit (Qiagen). The correct insertion of the fragment was confirmed by sequencing (Eurofins Genomics).

2.2.3 Immunohistochemistry

After dissection, brains were fixed in 4% PFA overnight at 4°C, cryoprotected in 30% sucrose and cryosectioned (10 µm thick slices). Cryosections were then permeabilized (0.5 % Triton X-100 in PBS) for 20 min, quenched for 30 min and blocked for 30 min at RT. All primary antibodies were incubated overnight at 4°C, followed by washing and incubation with secondary antibodies for 2h at RT. For Tbr2 staining, antigen-retrieval was performed in Citrate Buffer for 1 h at 70°C. Nuclei were counterstained with DAPI.

Image processing: sections were imaged using an automated microscope (ApoTome; Carl Zeiss), pictures digitally assembled using Axiovision or Zen software (Carl Zeiss) and

composites analyzed using Photoshop CS6 (Adobe). Cellular quantifications were normalized per RFP⁺ cells (electroporated population, total).

2.2.4 Bioinformatics

Sequencing data

Btg2^{RFP}/Tubb3^{GFP} transcriptome

Transcriptome sequencing of Btg2^{RFP}/Tubb3^{GFP} embryo-derived PP, DP and N cells was performed by Julieta Aprea as published (Aprea et al., 2013, 2015). Highly pure populations of PP (Btg2^{RFP}/Tubb3^{GFP}-), DP (Btg2^{RFP}/Tubb3^{GFP}-) and neurons (Tubb3^{GFP}+) were isolated by FAC-sorting at E14.5 from 3 embryos of different litters of the Btg2^{RFP}/Tubb3^{GFP} murine line. The mRNA was isolated and sequencing was performed on an Illumina HiSeq 2000, either 75 bp single-end sequencing (Aprea et al., 2013), resulting in ca 30-40 million reads per sample, or 100 bp paired-end sequencing (Aprea et al., 2015) resulting in ca 90 million reads per sample. The reverse reads of the paired-end sequencing were subsequently trimmed to 66 bp for quality reasons. Transcript assembly was performed by the sequencing facility using the “Tuxedo Suite” of Bowtie, Tophat and Cufflinks (Trapnell et al., 2010; Langmead and Salzberg, 2012; Kim D et al., 2013). The alignment was performed on the mm9 mouse genome and transcriptome assembly using the splice junction mapper Tophat 2 (v.2.2.10) (Kim D et al., 2013) which used Bowtie 2 (v.2.2.1) (Langmead and Salzberg, 2012) for mapping. Ensembl annotation v67 was used as a reference (Yates et al., 2020).

EGFP^{GFP} transcriptome

The raw sequencing data from cells of mouse E14.5 EGFP^{GFP} neocortex (Zhang et al., 2016) were downloaded from the Gene Expression Omnibus (GEO) database of the National Center for Biotechnology Information (NCBI), accession number: GSE76198. The alignment with Tophat and Bowtie resulted of bad quality; therefore the reads were mapped to the mm9 mouse transcriptome and genome using gsnap (Wu TD and Nacu, 2010) (alignment performed by Mathias Lesche, Deep Sequencing Group).

Detection of splicing

For all splicing analysis only the paired-end sequencing data were employed (Aprea et al., 2015). Analysis of the isoform expression and change in proportion was performed with Cuffdiff2 (Trapnell et al., 2012, 2013) with a false discovery rate (FDR) of 10% (Cuffdiff2 bioinformatic analysis performed by Mathias Lesche, Deep Sequencing Group). Lowly expressed genes (<1 fragment per kilobase of transcript per million mapped reads, FPKM)

were excluded from the dataset and were not further analyzed. Analysis of differential exon usage was performed with DEXSeq (v.1.14) (Anders et al., 2012) with FDR at 10% (performed by Mathias Lesche, Deep Sequencing Group). To exclude lowly expressed genes and to reduce the potential false positives, the exon usage dataset has been filtered for expression (at least 20 read counts/replicate in at least one cellular population) change in exon usage (\log_2 Fold Change $\geq |0.3|$) and FDR (5%). When the exon boundaries do not coincide among a gene's transcripts (alternative 5' or 3' splice site choice), DEXSeq divides the exon into bins. To estimate the real number and size of exons alternatively spliced, contiguous exon bins displaying the same splicing pattern in the transition from PP to DP or from DP to N have been merged using the merge function of bedtools v2.26 (Quinlan and Hall, 2010). Two other methods were employed to detect splicing at the exon level and estimate the inclusion of exons: DEXSeq-PSI (Schafer et al., 2015) and VAST-tools (v.2.2.2) (Irimia et al., 2014; Tapial et al., 2017). Both tools use the Percent Spliced In (PSI) as a metric for exon inclusion, which considers as supporting exon inclusion (Inclusion Reads, InR) the reads spanning exon-exon junctions between an exon and its possible flanking exons. The reads supporting exon exclusion (Exclusion Reads, ER) are exon-exon junction reads that map to positions upstream and downstream but never into the exon itself. DEXSeq-PSI also takes into account reads mapping to the exon bodies, while VAST-tools only uses junction reads. The Percent Spliced In (PSI) is then derived from the ratio between InR and the sum of InR and ER:

$$PSI = \frac{InR}{InR + ER} \times 100$$

Changes in PSI (delta PSI, Δ PSI) are indicative of differential splicing of an exon/exon bin. VAST-tools was run on the raw paired-end sequencing data with the *VAST-tools align* command using the mouse mm9 VastDB library (Mmu) and the version 2 of the intron retention module (--IR_version 2) (Braunschweig et al., 2014) to estimate exon and intron inclusion in each replicate. The Δ PSI was calculated with the *VAST-tools compare* command requiring an average difference of at least 10% in $|\Delta$ PSI| (--min_dPSI 10) and a minimum $|\Delta$ PSI| of 9% for each replicate (--min_range 9). Only exons with a minimum of 15 reads supporting inclusion (--noVLOW option) were considered. To include only potentially biological relevant alternative splice site choices (alternative 3' and alternative 5' splice site), these events were considered only when reference exons showed an average PSI of at least 25% in all compared samples (--min_ALT_use=25). The same parameters used for the VAST-tools analysis were applied to the DEXSeq-PSI pipeline. Contiguous exon bins with the same splicing pattern were joined together (merge function) and their coordinates

were compared to those of events derived from the Ensembl mm9 v67 transcriptome annotation and/or from VastDB libraries (intersect function) using bedtools v2.26 (a total inclusion of the merged bins into the annotated reference was required, option -F 1). The merged bins were assigned to the closest splicing event when their length could cover at least 2/3 (66.67%) of the event length. The PSI of events found to be spliced by both VAST-tools and DEXSeq-PSI methods was calculated as the mean of the PSI derived by the two tools. As for the exon usage analysis with DEXSeq a filter on gene expression (≥ 20 read counts/replicate in at least one exon in one cell population considered) was used.

Gene ontology analysis

Genes that resulted to be differentially spliced according to DEXSeq-PSI and VAST-tools were clustered based on their association to common gene ontology (GO) terms in biological processes, molecular functions, cellular components or biological pathways. The cluster module of the Database for Annotation and Integrated Discovery DAVID (v. 6.8) (Huang DW et al., 2009a, b) was used for clustering and enrichment evaluation of GO terms of spliced genes. High stringency settings were applied and all the multiexonic genes expressed in the cell populations of interest were used as a background. Only GO clusters with an enrichment score ≥ 3 were considered. Highly related GO terms were manually grouped into single GO terms to minimize redundancy. Cytoscape (Shannon et al., 2003) was employed to visualize the enriched clusters in order to highlight the number of spliced genes in each cluster and the common genes among them.

Impact of alternative splicing on coding and non-coding sequences

Genomic coordinates of differentially spliced exons in the present dataset and of coding and non-coding exons annotated in Ensembl v67 were compared with the *intersect* function of bedtools v2.26 to evaluate the impact of splicing on coding and non-coding regions of the transcripts. Overlaps of at least 1 bp were considered. Coding exons with a length not multiple of 3 bp or containing an annotated stop codon were considered to disrupt the open-reading frame (ORF) when included. Spliced exon coordinates were converted into protein coordinates and mapped to annotated protein domain coordinates via the Ensembl Perl Application Programming Interface (API). Overlaps of at least 1 bp with annotated domains were considered as domain-modifying. Specific protein domains in splice variants of interest were further analyzed using protein databases such as Uniprot, Pfam and Interproscan.

Analysis of exon features

The compare exons function (cmpr_exons) of Matt (v1.3.0) (Gohr and Irimia, 2019) was used to retrieve and compare exon features of included (In) and excluded (Ex) exons as well as exons of different length class. The analysis focused on features more directly involved in exon definition mechanisms such as the strengths of the splice sites, of the branch points and of the polypyrimidine tracts. The intronic regions to be considered for branch point prediction were set as 150 nt at the 3' end of each intron neglecting the first 20 nt at their 5' end. The results were displayed as violin plots and boxplots. The graphs were scaled in order to contain the boxplots while removing most of the outliers, i.e. values falling below the "minimum" and above the "maximum" of the boxplot where:

$$\text{minimum} = 25^{\text{th}} \text{ percentile} - [1.5 \times (75^{\text{th}} \text{ percentile} - 25^{\text{th}} \text{ percentile})]$$

$$\text{maximum} = 75^{\text{th}} \text{ percentile} + [1.5 \times (75^{\text{th}} \text{ percentile} - 25^{\text{th}} \text{ percentile})]$$

The packages ggplot2 (Wickham, 2016) of R (R Core Team, 2017) was used to generate the graphs. Heatmaps were employed as alternative representations of the exon feature differences. The values of the different features were transformed to a 0-100 scale and differences of the medians were represented with a color-coded scale from -10 to +10. To generate the heatmaps, the packages ggplot2 and ComplexHeatmap (Gu et al., 2016) of R were used.

Splicing factor differential expression and analysis of enrichment for splicing factors binding sites

Splicing factor expression was derived from the gene differential expression analysis previously performed by Julieta Aprea on the single-end Btg2^{RFP}/Tubb3^{GFP} transcriptome (Aprea et al., 2013) with DESeq (Anders and Huber, 2010). A minimum $|\log_2\text{FC}|$ of 0.58 at a FDR of 5% was required to consider a gene differentially expressed. The normalized counts of all the mouse splicing factors annotated in the Catalog of Inferred Sequence Binding Proteins of RNA (CISBP-RNA) (Ray et al., 2013) were analyzed and the differentially expressed splicing factors were grouped according to their expression patterns from PP to DP and from DP to N. Eight different expression patterns were identified: group 1) genes upregulated in both the PP-DP and the DP-N transitions (up-up), group 2) genes upregulated in the PP-DP transition but keeping a constant expression in the DP-N transition (up-constant), group 3) genes upregulated in the PP-DP transition and downregulated in N (up-down), group 4) genes downregulated in both the PP-DP and the DP-N transitions (down-down), group 5) genes downregulated in the PP-DP transition and

with constant expression in N (down-constant), group 6) genes downregulated in the PP-DP transition and upregulated in the DP-N transition (down-up), group 7) genes with constant expression in the PP-DP transition and upregulated in the DP-N transition (constant-up), group 8) genes with constant expression in the PP-DP transition and downregulated in N (constant-down). To represent the differentially expressed splicing factors in a heatmap, the counts were transformed into z-scores as follows:

$$z_i = \frac{x_i - \bar{x}}{s}$$

where z_i is the standardized value or z-score, x_i the count of a gene in a cell population, \bar{x} the mean of the gene counts in the three cell populations and s the standard deviation. Splicing factors with the same expression pattern were grouped together and clustered via hierarchical clustering within the expression pattern group. The R packages `ggplot2` and `ComplexHeatmap` were used to generate the graph. To assess enrichment for splicing factor binding sequences nearby differentially spliced cassette exons, the `rna_maps` function of `Matt` was used. Upstream and downstream exon border coordinates were derived from `VastDB` libraries when available and from the mouse annotation with the `refGenome` R package (v1.2.0) otherwise. A sliding window of 31 nt for enrichment of splicing factor binding motifs was used to scan exonic regions (35 nt) at the spliced exon extremities and both at the 5' end of the upstream exon and at the 3' end of the downstream exons. In the same way 135 nt intronic regions flanking the spliced exon and both downstream of the preceding exon and upstream of the following exon were scanned. When the enrichment for binding sites was compared among exons of different length classes, 13 nt of exonic region and a 17 nt sliding window were used. The Perl regular expression motifs annotated in the CISBP-RNA database were employed for the analysis. A subset of 2000 alternative cassette exons, not significantly spliced in the present dataset, was taken as a reference to assess the motif enrichment.

Statistical analysis

For differential exon usage (DEXSeq) and isoform relative expression (Cuffdiff) analyses, the Benjamini–Hochberg procedure was applied for multiple t-test adjustment and FDR. A chi-square test with Yates correction was used to assess the significance in the proportion of exons with an impact on the coding sequence. Exon feature differences were evaluated with the Mann-Whitney U test. A permutation test with 10,000 permutations was applied to determine the significance of enrichment for binding sites of splicing factors. A linear regression model was employed to determine the correlation between exon usage values

obtained from different methods and between predicted and experimentally measured PSI. For all other experiments, a minimum of 3 biological replicates was used, unless otherwise specified. Statistical differences of mean values were calculated by two-tailed student's t-test.

3 Results

3.1 Splicing factors are differentially expressed during neurogenic commitment and neurogenesis

In order to address the role of alternative splicing in cell-fate determination in the developing mouse neocortex, I initially analyzed available transcriptome data of PP, DP and N cell populations FAC-sorted from *Btg2^{RFP}/Tubb3^{GFP}* murine line at E14.5, the peak of neurogenesis (Aprea, et al., 2013, 2015).

Differential gene expression data obtained as normalized counts with DESeq tool (Anders and Huber, 2010), revealed that several genes are up-/downregulated during neurogenic commitment and neurogenesis (Aprea et al., 2013). I examined these data to reconstruct the expression profiles of known splicing factors and RNA-binding proteins annotated in the Catalog of Inferred Sequence Binding Proteins of RNA (CISBP-RNA) (Ray et al., 2013). As previously described (Aprea et al., 2013), a gene was considered differentially expressed if its normalized gene count changed by at least 50% ($|\log_2FC| \geq 0.58$) between 2 cell populations (p -adjusted < 0.05). Of the 325 splicing factors and RNA-binding proteins found present in the cells of the neurogenic lineage, 110 were differentially expressed during development (Figure 3.1). Of those, about 1/3 (38 genes) were regulated in the neurogenic commitment step (23 up- and 15 downregulated, groups 1 to 3 and groups 4 to 6, respectively).

The majority of the up- and downregulated genes in this stage further changed their expression levels with the same pattern in N (15 up- and 10 downregulated, groups 1 and 4, correspondingly) or kept the same expression levels during neurogenesis (7 up-constant, group 2 and 4 down-constant, group 5). Only two splicing factors, *Elavl2* and *Csdc2*, showed a switch expression pattern, up-down and down-up, respectively. The remaining 2/3 kept a constant expression in PP and DP and were up- (23 genes, group 7) or down-regulated (49 genes, group 8) in N.

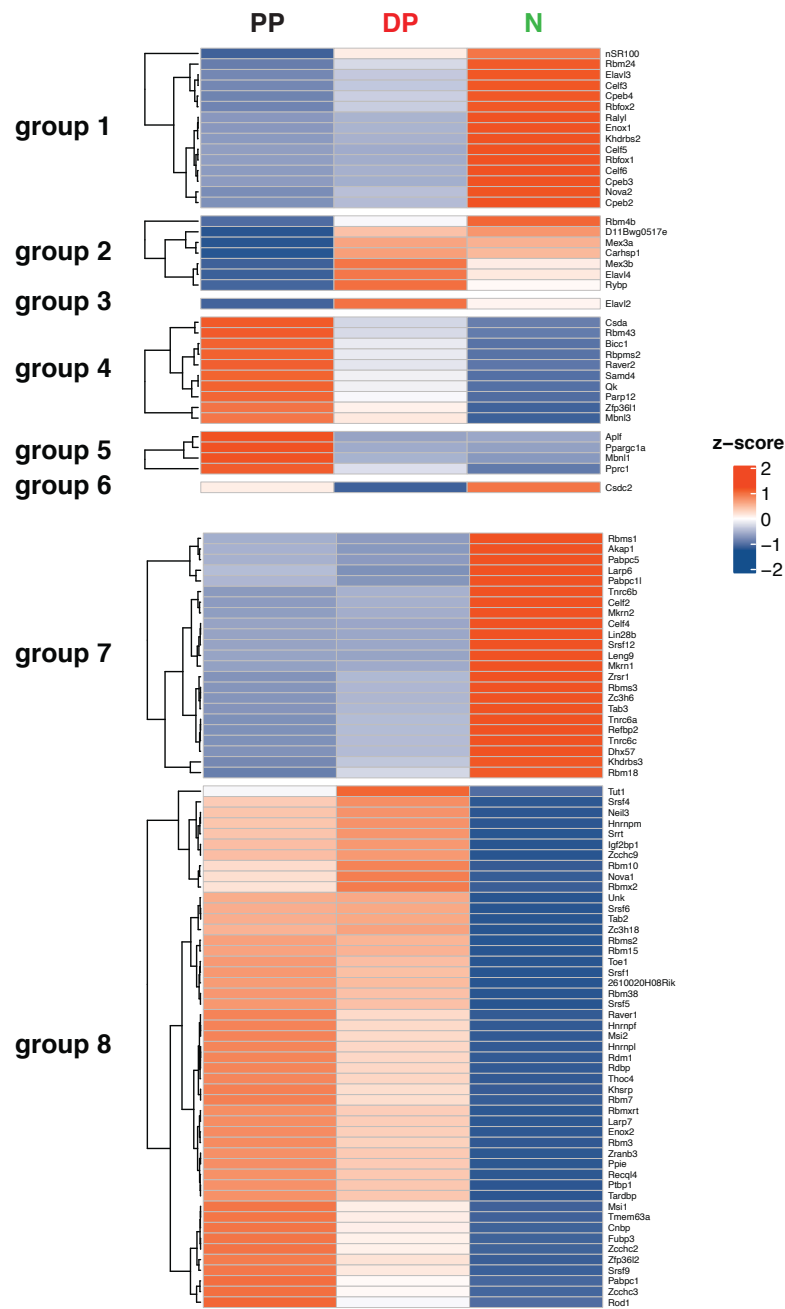


Figure 3.1 Differential expression of splicing factors in cells of the neurogenic lineage. Heatmap of the normalized gene counts in the three cell populations. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons.

Of note, the pro-neurogenic splicing factors nSR100 and Nova2, started to be upregulated in DP and were even further upregulated in N (group 1). On the other hand, Ptpb1, which counteracts neural differentiation had a constant expression in PP and DP and was

downregulated in N ($\log_2FC = -1.47$). Concomitantly, Raver1, a co-factor of Ptbp1, was downregulated and miR124, which silences Ptbp1, was upregulated (Dori et al., 2020). The brain-specific Ptbp1 homologue Ptbp2 did not appear to be differentially expressed across the three cell populations analyzed.

The fact that splicing factors with a known role in the nervous system such as nSR100 and Nova2 were shown to be differentially regulated between PP and DP suggests that specific splicing events could be involved not only in neurogenesis and neural maturation, but already in neurogenic commitment and have a crucial role for cell-fate determination. Therefore, I analyzed the splicing profiles of cells of the neurogenic lineage with special attention to the differences between PP and DP in order to identify splicing events critical for neurogenic commitment.

3.2 Detection of alternative splicing

Given the aforementioned uncertainty in splicing detection and the variety of available methods, I decided to test different bioinformatic strategies and to use the best PP, DP and N deep sequencing transcriptome data available in our lab (Aprea et al., 2015), in order to characterize cell-population-specific splicing profiles. The mRNA was isolated at the peak of neurogenesis, at E14.5, from Btg2^{RFP}/Tubb3^{GFP} mouse embryos of 3 litters and 100 bp paired-end sequencing was performed giving a high quality transcriptome suitable for detection of splicing (Aprea et al., 2015). The raw sequencing data were aligned to the mouse reference transcriptome and genome (Ensembl assembly mm9, version 67) with a splice-aware pipeline (alignment performed by Mathias Lesche, Deep Sequencing Group). I analyzed the raw and aligned reads with various bioinformatic tools at isoform- and exon-level to identify genes differentially spliced in cells of the neurogenic lineage. I then compared the results with the raw sequencing data and subjected them to validation.

3.2.1 Isoform-switching

To investigate splicing at the isoform-level, I chose to use the popular tool Cuffdiff2 (Trapnell et al., 2012, 2013; analysis performed by Mathias Lesche, Deep Sequencing Group). The isoform reconstruction method identified 249 and 881 genes undergoing isoform switching between PP-DP and DP-N, respectively (False Discovery Rate, FDR 10%). As previously mentioned, Cuffdiff2 distinguishes genes that undergo isoform switching into three non-mutually exclusive classes: change in coding sequence (CDS), in promoter usage (TSS) and in splicing proper (Splicing).

The majority of isoform switching genes detected involve a change in the coding sequence: 178 genes (CDS: 178/249, 71.49%) in PP-DP and 703 genes (CDS: 703/881, 79.80%) in DP-N (Figure 3.2 A and B, respectively). Of the 249 genes found to undergo isoform switching in the PP-DP transition, about half showed isoform variability resulting from mechanisms of splicing (Splicing: 116/249, 46.59%) while about 1/3 derived from alternative transcription promoter usage (TSS: 81/249, 32.53%) (Figure 3.2A).

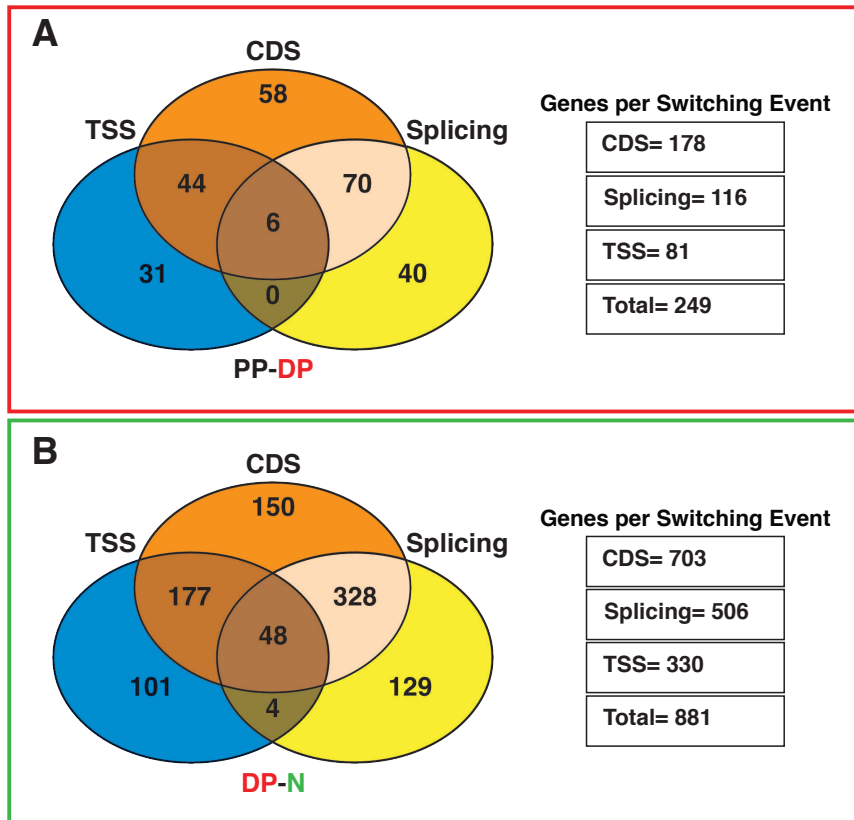


Figure 3.2 Detection of Isoform Switching with Cuffdiff2. Isoform switching events are classified based on changes in the proportion of protein isoforms being expressed (CDS), different TSS usage (TSS) or Splicing. Isoform switching between PP and DP (A, red box) and DP and N (B, green box). The number of genes changing in TSS (blue), CDS (orange) and Splicing (yellow) events are reported in the graph (left). Genes undergoing more than one type of isoform switching are reported in the area of overlap of the switching events they belong to. Total number of genes in each event is listed in the table (right). The area of the circles is not proportional to the number of genes detected. CDS=coding sequence; TSS=transcription start site; PP=proliferating progenitors; DP=differentiating progenitors; N=neurons.

Similarly, more than half of the isoform switching genes in neurogenesis (DP-N transition) arose from alternative splicing choices (Splicing: 506/881, 57.43%), whereas a smaller proportion was a consequence of alternative promoter usage (TSS: 330/881, 37.46%) (Figure 3.2 B).

This analysis shows that several genes modify their transcripts proportion during corticogenesis, especially in the neurogenic phase (DP-N) resulting mainly in a change in protein variants being expressed. Splicing choices seem to account for only about half of the genes undergoing isoform switching, pointing to a roughly equal contribution of alternative splicing and alternative transcription mechanisms in regulating isoform variability.

3.2.2 Exon usage and splicing events

Three exon-centric tools were employed to identify exons and splicing events exhibiting differential inclusion in PP, DP and N. DEXSeq (Anders et al., 2012) was used to identify exons that change their representation among all the exons of a gene, i.e. exons showing differential usage. As mentioned above, DEXSeq reports \log_2FC of exon usage as a measure of differential splicing. In the present study, exon bins with a \log_2FC of at least ± 0.3 ($|\log_2FC| \geq 0.3$) in usage were considered as differentially spliced (FDR 5%). Contiguous exon bins spliced with the same pattern were merged together and treated as a single exon (hereafter referred to as “exons”, unless specified otherwise). The exon usage analysis with DEXSeq (performed by Mathias Lesche; post-analysis filtering on gene expression and \log_2FC as well as merging of co-spliced contiguous exon bins performed by me) identified 353 genes differentially spliced between PP and DP and 1,820 between DP and N (Figure 3.3).

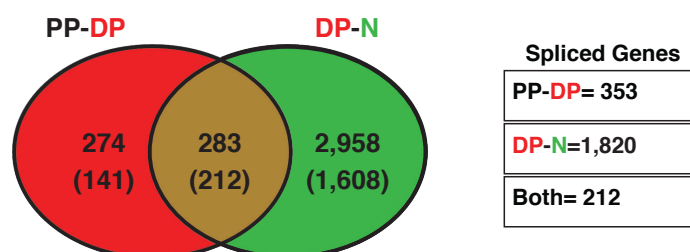


Figure 3.3 Exon usage. Changes in exon usage between PP-DP (red circle) and DP-N (green circle). The number of exons and of genes (in brackets) differentially spliced is reported. Genes and exons alternatively spliced in both transitions are reported in the area overlapping the two circles. The total number of genes differentially spliced in each comparison is listed in the table. The area of the circles is not proportional to the number of genes or exons detected. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons.

Of the 557 exons (353 genes) alternatively spliced in the neurogenic commitment switch (PP-DP), 274 exons (141 genes) kept a constant inclusion in N, while 283 exons (212

genes) resulted differentially spliced also in N. Compared to DP, N showed differential inclusion levels of 2,958 exons (1,608 genes), that were not alternatively spliced between PP and DP (Figure 3.3).

To gain a deeper insight about the inclusion levels of the different exons and about the specific splicing events involved, the PSI of exon bins and of splicing events annotated in the VastDB libraries were calculated with DEXSeq-PSI (Schafer et al., 2015) and with the event-based tool VAST-tools (Irimia et al., 2014; Tapial et al., 2017), respectively. A $|\Delta\text{PSI}| \geq 10$ was chosen as a threshold for considering an exon bin/splicing event alternatively included.

The approaches used by these different bioinformatic tools have important consequences for the detection of alternative splicing. When comparing the $\log_2\text{FC}$ of the exon bins reported to be spliced by DEXSeq with their respective ΔPSI obtained with the DEXSeq-PSI method it was possible to observe only a qualitative correlation (Figure 3.4 A, left, $R^2=0.26$, $p\text{-value}<0.0001$). About 1/3 of exon bins showing a differential usage with DEXSeq resulted alternatively spliced according to DEXSeq-PSI (209/630, 33.17% in PP-DP and 1,111/3,916, 28.37% in DP-N). The remaining differentially used exon bins showed little ($0 < |\Delta\text{PSI}| < 10$, PP-DP: 177/630, 28.10%; DP-N: 1,237/3,916, 31.59%) or no difference ($|\Delta\text{PSI}|=0$, PP-DP: 244/630, 38.73%; DP-N: 1,568/3,916, 40.04%) in inclusion (Figure 3.4 B, left). The sequencing reads of a subset of exon bins with significant differences in usage ($|\log_2\text{FC}| \geq 0.3$ according to DEXSeq) but not in inclusion ($|\Delta\text{PSI}| < 10$ according to DEXSeq-PSI or VAST-tools) was visually inspected and found to be not supporting differential splicing. Of note, a considerable number of exon bins showed extensive differences in inclusion according to DEXSeq-PSI, but little or no $\log_2\text{FC}$ in usage according to DEXSeq (498 in PP-DP and 1,578 in DP-N) (Figure 3.4 B, left).

(Figure 3.4 legend continuation) The sizes of the circles are fixed; the area of overlap is proportional to the fraction of exon bins relative to the group of exon bins differentially used detected by DEXSeq. B right: comparison of exons with $|\Delta\text{PSI}| \geq 10$ according to VAST-tools and to DEXSeq-PSI. The sizes of the circles are proportional to the number of exons (events or merged exon bins) found differentially spliced by the two tools, the area of overlap is proportional to the number of exons found by both tools. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons.

A stronger linear correlation was observed when comparing the ΔPSI of exon bins calculated with DEXSeq-PSI and the ΔPSI of their corresponding splicing events estimated by VAST-tools ($R^2=0.63$, $p\text{-value}<0.0001$) (Figure 3.4 A, right). The majority of splicing events detected by VAST-tools were identified also by DEXSeq-PSI analysis (179/303, 59.08% in PP-DP and 735/1,214, 60.54% in DP-N) (Figure 3.4 B, right).

The sequencing reads of a group of splicing events detected only by one of these tools (VAST-tools: 124 in PP-DP and 479 in DP-N; DEXSeq-PSI: 366 in PP-DP and 1,249 in DP-N) were inspected visually and found to be in agreement with the splicing prediction in the vast majority of cases.

Given the high variability in ΔPSI of exon bins with the same $\log_2\text{FC}$ and the discrepancies in exons showing differential usage as reported by DEXSeq but negligible inclusion changes according to DEXSeq-PSI and to visual inspection of the sequencing data, I decided to discard the analysis done with the original DEXSeq pipeline in favor of the DEXSeq-PSI and the VAST-tools methods.

In order to reduce the exclusion of false negatives and given the high correlation between the DEXSeq-PSI and the VAST-tools methods, I opted for a combination of the results obtained by the two splicing pipelines. To do so, the exons detected to be spliced by DEXSeq-PSI were assigned to splicing events annotated in VAST-DB or reconstructed from the Ensembl annotation. Since DEXSeq-PSI analysis takes into account not only splice junctions but also reads mapping to exon bodies, it allows for the detection of alternative transcription events when additional exons are annotated in the gene structure upstream of a transcript's first exon and downstream of a transcript's last exon. Therefore, alternative first and last exons that do not coincide with the gene's first TSS and last polyadenylation site were considered when mapping exons to splicing events. Contiguous exon bins spliced with the same pattern were joined together and assigned to the most plausible event. Only merged exon bins that could cover at least 2/3 (66.67%) of an event were considered (Figure 3.5 A).

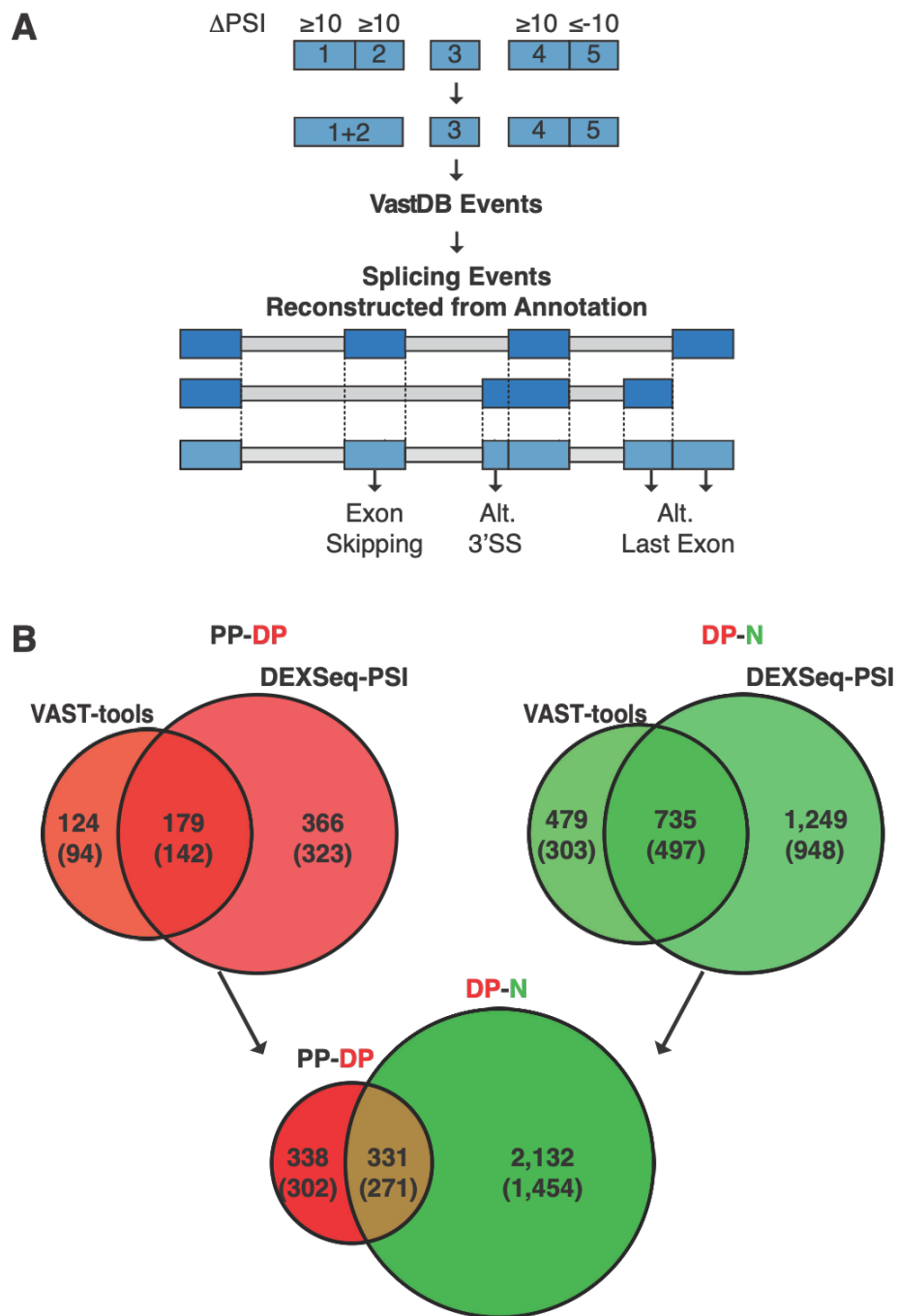


Figure 3.5 Combination of the DEXSeq-PSI and VAST-tools datasets. A: Contiguous exon bins showing $\Delta\text{PSI} \leq -10$ or $\Delta\text{PSI} \geq +10$ were merged together and their coordinates mapped to events annotated in the VastDB libraries or reconstructed from the Ensembl mm9 transcript annotation. B: all the events found by one or both tools were combined in a unique dataset. The number of events for each tool and comparison are reported in the respective circle. The number of genes alternatively spliced is reported in brackets. The area of the circles is proportional to the number of events. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; PSI=percent spliced in; Alt.=alternative; SS=splice site.

The final dataset consists of 338 events (302 genes) alternatively included only in PP-DP, 2,132 events (1,454 genes) only in DP-N and 331 events (271 genes) differentially included in PP-DP and further spliced in DP-N, with the same or opposite pattern (Figure 3.5).

Overall, analysis at the exon level detected more genes undergoing alternative splicing compared to the ones at the isoform level. Both methods revealed a more pronounced change in splicing choices during neurogenesis (DP-N) relative to neurogenic commitment (PP-DP). Among exon-based approaches, tools that measure exon inclusion vs exclusion as a PSI, rather than exon usage relative to all the exons of the gene, were found to be more accurate.

3.3 Validation

The isoform switching and the differential exon inclusion predicted by the bioinformatic tools were validated using qPCR or RT-PCR on mRNA extracted from PP, DP and N from the Btg2^{RFP}/Tubb3^{GFP} murine line.

3.3.1 The isoform switching method has a poor validation rate

The isoform switching analysis identified several genes involved in neurogenesis that changed their protein output between PP and DP. Among those, genes with the most relevant alteration in coding sequences were selected for validation by qPCR. Isoform-specific and whole-gene primers were designed and the qPCR conditions were adjusted to maximize their specificity and efficiency.

For two genes tested, Tmcc2 (Transmembrane and coiled-coil domain family 2) and Celf5 (CUGBP, Elav-like family member 5), it was not possible to design isoform-specific primers for all the splice variants of interest and a subtraction strategy was adopted.

In the case of Tmcc2, the primers detecting a short isoform that decreased its proportion during neurogenic commitment, T9, were overlapping with a non-coding isoform, T17, predicted to be very lowly expressed in our dataset. As a consequence, the T9 qPCR cycle threshold (Ct) observed would result from the expression of both T9 and T17. Therefore, T17-specific primers were designed in order to calculate the real T9 expression by subtracting T17 values. Oddly, T17 consistently showed a lower Ct value indicating higher expression than T9 in all the conditions tested, which could not be ascribed to different primer efficiency nor to genomic contamination of the mRNA used (data not shown). It was then concluded that this discrepancy was most likely resulted from improper annotation of

the transcript structures and the primers were considered to be specific for the splice variants tested.

The primers for Celf5 isoforms C60 and C63 showed a similar problem. While it was possible to design primers specific for the long isoform C63, qPCR primers for the short variant C60 were overlapping with C63. Therefore, also for Celf5 isoforms the strategy adopted was to subtract the C63 expression value from the one of C60. However, the high GC-content (75% GC) of the C63 amplicon was affecting amplification efficiency leading to very high Ct values and introducing artifacts in the qPCR dissociation curve, a measure to evaluate primer specificity. The addition of formamide to the qPCR mixture in order to reduce GC-secondary structure and improve primer specificity, dramatically decreased the C63 Ct value without altering the ones of C60. The specificity of C63 primers was further confirmed by TA-cloning and sequencing of C63 amplicons obtained with formamide-enriched qPCR mixture following the same qPCR protocol used for the validation. Primer pairs for the whole Celf5 gene revealed an extremely low amplification efficiency in every condition tested, even in the presence of formamide or other qPCR additives. Therefore, Celf5 gene primers were not taken into account and the validation was performed by comparing only C60 and C63 expressions relative to each other.

While the general tendency of change in isoform expression and proportion during neurogenesis was in agreement with the bioinformatic prediction, none of the isoform switching events tested could be validated (Figure 3.6). An isoform switching was observed for one splice variant relative to the whole gene (Porcn: P95/Porcn, Tmcc2: T9/Tmcc2) but not for the complementary splice variant tested (Porcn: P20/Porcn, Tmcc2: T73/Tmcc2) or the change in relative abundance of the isoforms showed to be of lower magnitude than predicted possibly due to inefficient amplification (Celf5: C60/C63).

The results highlighted not only a technical limitation in designing isoform-specific primers with comparable efficiencies, but also inaccuracies in the available transcript annotation that might account for the discrepancies observed. Because of such misannotations and of the intrinsic complexity in reconstructing full transcripts expression from short reads, Cuffdiff2 failed in providing an accurate picture of change in splice variants representation in our samples. Given the unsuccessful validation of isoform switching, the dataset obtained with this method was discarded and no further bioinformatic analyses was carried on those data.

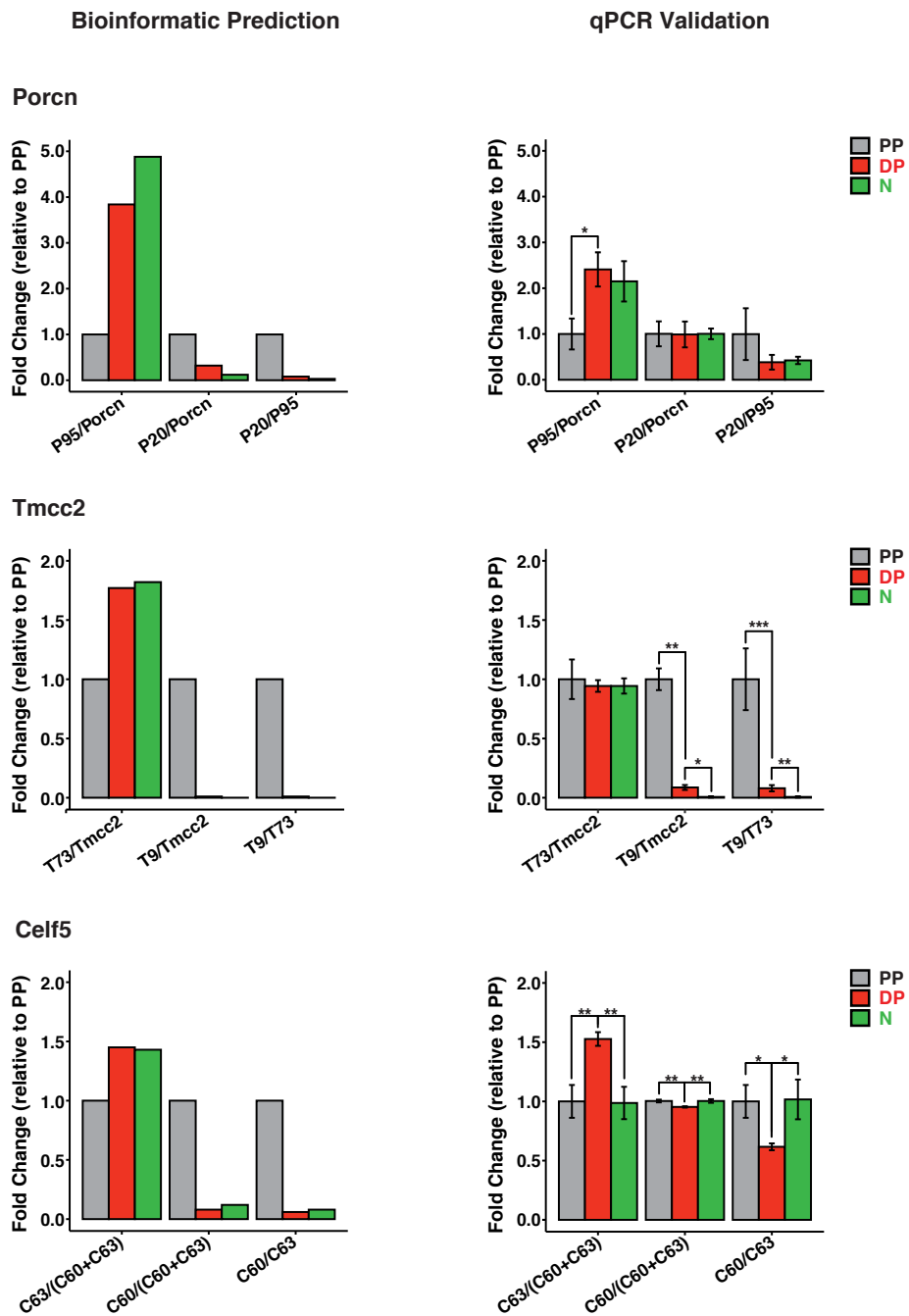


Figure 3.6 qPCR validation of isoform switching predicted by Cuffdiff2. Left: prediction of fold changes in isoform proportion relative to PP. Right: fold change in proportion of the tested isoforms relative to their proportion in PP. Three technical replicates from one biological replicate. *p-value<0.05, **p-value<0.005, ****p-value<0.0001, paired t-test. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons.

3.3.2 Analysis at the exon level has a high rate of validation

From the exon-centric analysis 11 cassette exons from 10 different genes were selected and their changes in inclusion were validated by RT-PCR. Primers were designed to flank the exons predicted to change inclusion between PP and DP, yielding two amplification products differing in size by the length of the alternative exon tested (Figure 3.7 A). This approach ensures an equal efficiency in amplification of the splice variants, since the same primer pairs are used to amplify both variants. Moreover, since the splicing event to be tested is a different inclusion of an exon, and not its relative proportion to the whole gene, the validation is carried out only by comparing the abundance of the inclusion and exclusion amplification products.

All of the 11 spliced exons tested exhibited a change in inclusion pattern in agreement with the bioinformatic prediction ($R^2=0.95$, $p\text{-value}<0.0001$), thus validating the dataset obtained with the analysis at the exon level (Figure 3.7 B). Of note, when the PCR products obtained with this technique were run in an agarose gel, an unexpected additional amplicon was consistently observed at a slightly higher molecular weight than the amplicon corresponding to the splice variant including the alternative exon. This could be indicative of a novel isoform or simply be a PCR artifact. Restriction digestion of the PCR products excluded the possibility of the unexpected amplicons being novel isoforms (data not shown). Consistently with what had been observed by other groups, the unexpected amplicons were not detected when the PCR products were run in polyacrylamide gels, confirming that they represented indeed artifacts, most likely heteroduplexes resulting from pairing of the highly complementary inclusion and exclusion amplicons (Manuel Irimia, personal communication).

The results demonstrate the suitability of exon-based tools for the detection of splicing changes with short reads data as well as the high accuracy of the PSI metrics in representing differences in exon inclusion. The validation attested the quality of the exon-based dataset that was therefore employed for further analyses.

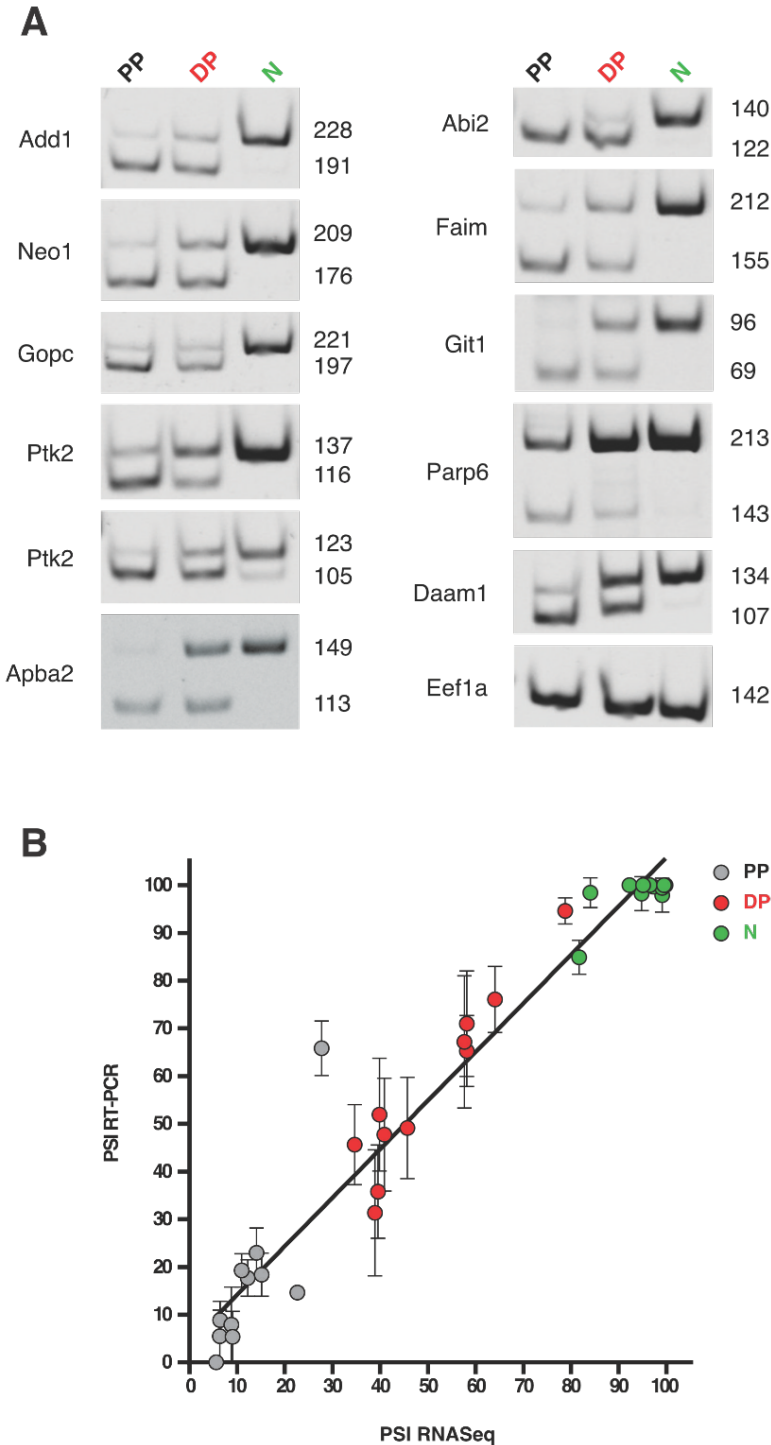


Figure 3.7 Validation of exons predicted to be alternatively spliced by DEXSeq-PSI and VAST-tools. A: for each splicing event, the upper and lower bands represent the form with exon inclusion and exclusion, respectively (bp indicated to the right of each picture). The intensity of the PCR product is proportional to the amount of splice variant amplified. Eef1a was taken as a control that equal amount of total RNA was used for each sample. B: correlation between the PSI measured by the RT-PCR and the PSI predicted bioinformatically. Mean and standard error of the mean (SEM) are reported; $n \geq 2$, from different biological replicates. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; PSI=percent spliced in.

3.4 Pattern and representation of splicing events

3.4.1 Splicing choices during neurogenic commitment define the splicing profiles of neurons

To better characterize the splicing profiles of the cells of the neurogenic lineage, I investigated the pattern and level of inclusion of the alternative splicing events detected.

Inclusion events (In) were more frequent than exclusion (Ex) ones in genes differentially spliced in PP-DP and DP-N transitions (411 In vs 258 Ex and 1,536 In vs 927 Ex in PP-DP and DP-N, respectively) (Figure 3.8 A).

The vast majority of the events spliced in or out in the transition from PP to DP were either kept with a similar inclusion level in N (170/411, 41.36% of In and 168/258, 65.12% of Ex) or were further spliced with the same pattern from DP to N (227/411, 55.23% of In and 72/258, 27.91% of Ex). Only a few events showed a “switch-like” pattern, incrementing inclusion from PP to DP and then decreasing it in N (14/411, 3.41% of In), or vice versa, first losing and then gaining inclusion (18/258, 6.98% of Ex) (Figure 3.8 A).

The distribution of PSI of splicing events in the cells of the neurogenic lineage was further analyzed. Events that gained inclusion during neurogenic commitment (PP-DP.In) had a distribution of PSI skewed towards low levels in PP with the majority of them showing less than 50% of inclusion (25th percentile PSI: 7.89; median PSI: 22.27; 75th percentile PSI: 46.80; distribution statistics hereafter reported always in this order). The distribution became more uniform in DP (PSI: 35.63; 53.30; 73.66) with more than half of the included events representing the predominant but not unique form and was skewed towards high values in N (PSI: 66.00; 86.90; 97.36) with half of the events reaching an almost total inclusion. These data indicate that the exons gained during neurogenic commitment might be important for the acquisition and maintenance of the neurogenic identity (Figure 3.8 B, top left).

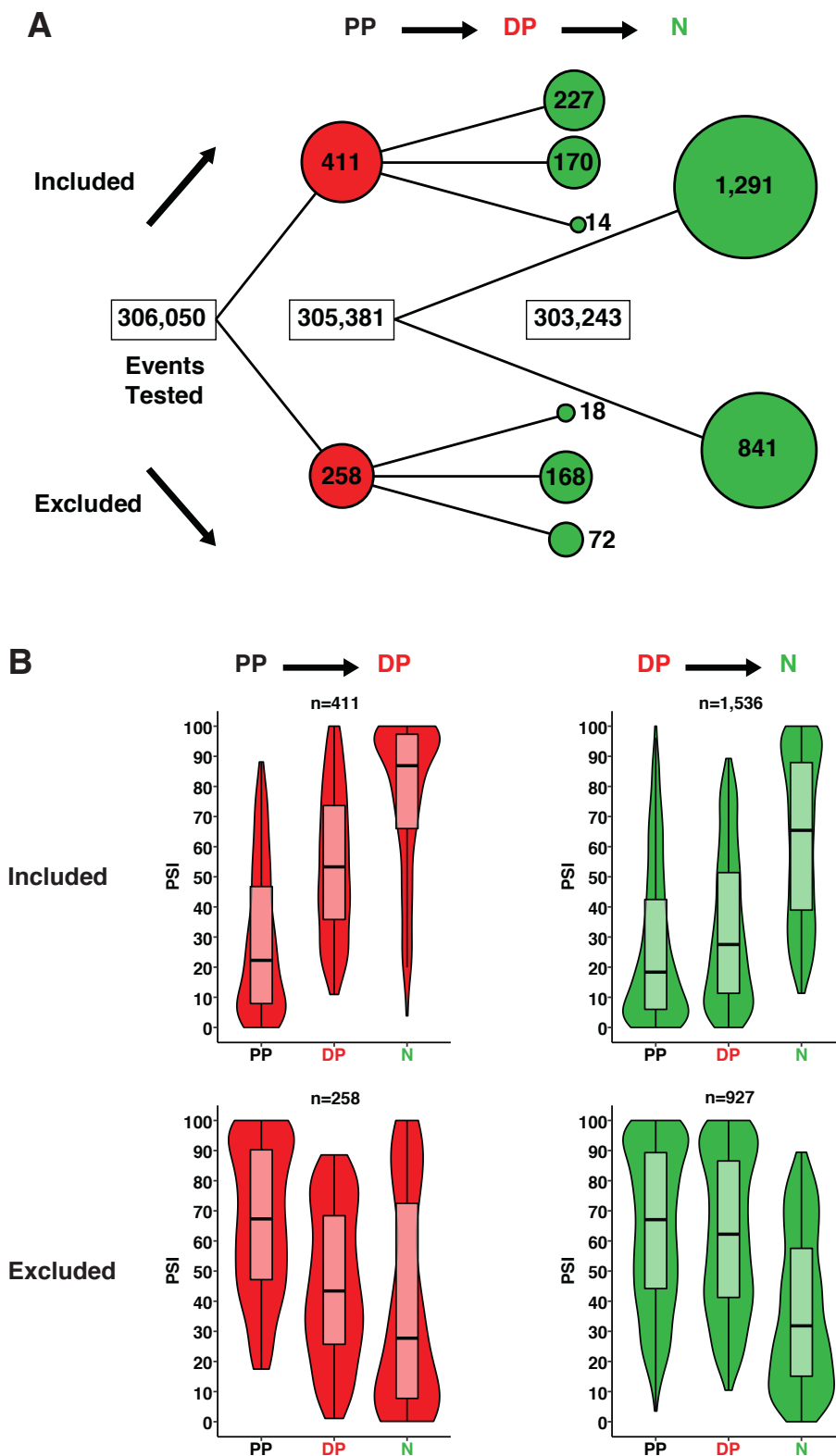


Figure 3.8 Pattern of exon inclusion/exclusion (legend continued on next page).

(Figure 3.8 legend continuation) A: change in exon inclusion during differentiation of cells of the neurogenic lineage. Number of exons differentially spliced between PP and DP (red circles) and between DP and N (green circles) are represented. The areas of the circles are proportional to the number of exons with differential inclusion. Lines pointing upwards and downwards indicate higher and lower inclusion, respectively. The number of not differentially spliced exons expressed in each cell population is reported in the middle of the figure (numbers in white boxes). B: distribution of PSI in PP, DP and N for events that increase (top) or decrease (bottom) their inclusion during neurogenic commitment (left) or during neurogenesis (right). PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; PSI=percent spliced in.

The majority of the exons which were more skipped in DP than in PP (PP-DP.Ex) represented the predominant splice form in PP (PSI: 46.96; 67.33; 90.50). In DP, the PSI distribution was centered at average levels (PSI: 25.96; 43.41; 68.51) while being skewed towards low values in N, although a subset of events showed high PSI in this cell population (PSI: 7.70; 27.72; 72.57) (Figure 3.8 B, bottom left).

Events spliced-in between DP and N (DP-N.In) were generally excluded in PP (PSI: 5.97; 18.35; 42.51) and in DP to a lesser extent (PSI: 11.32; 27.43; 51.42) while most of them became the predominant form in N (PSI: 38.96; 65.43; 87.94) (Figure 3.8 B, top right).

Events that were more excluded in N (DP-N.Ex) had PSI values distributed in medium-high ranges in PP and in DP and in more than half of the cases they represented the most frequent splice form (PP PSI: 44.16; 67.06; 89.33; DP PSI: 41.23; 62.25; 86.56). The drop in inclusion of these events in N brought the PSI distribution to low levels with only ~1/4 of them keeping a medium-high PSI value (PSI: 15.10; 31.82; 57.64) (Figure 3.8 B, bottom right)

The observed distribution patterns of PSI of spliced exons highlight a strong tendency in gained exons to reach a total inclusion in the expressed isoforms as development proceeds, particularly pronounced for exons gained at earlier stages, in the PP-DP transition. Conversely, events decreasing their representation during corticogenesis showed milder changes.

3.4.2 Splicing events: microexon inclusion characterizes neurogenic commitment

The analyses employed allowed not only the detection of transcript regions differentially spliced but also their categorization into classes of splicing events: exon skipping, alternative splice site selection and intron retention. The VAST-tools pipeline provides this classification although it is not able to detect alternative first and last exons. In addition, the tool has a dedicated module for the detection of microexons (3-27 nt).

At a $|\Delta\text{PSI}| \geq 10$, VAST-tools identified 209 events (170 genes) more included and 94 events (82 genes) more excluded in the PP-DP transition for a total of 303 alternative splicing events: 138 skipped exons (51 Ex, 87 In), 82 skipped microexons (5 Ex, 77 In), 7 alternative 5'SS (4 Ex, 3 In), 15 alternative 3'SS (11 Ex, 4 In) and 61 retained introns (23 Ex, 38 In) (Figure 3.9).

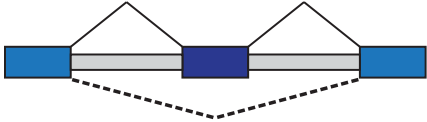
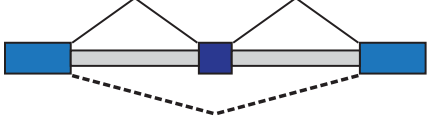
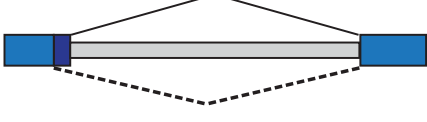
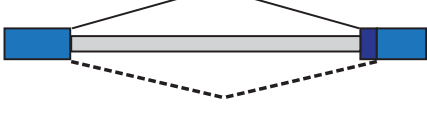
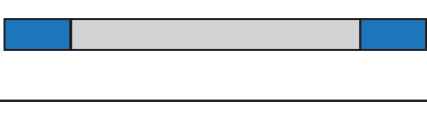
Splicing Events		$ \Delta\text{PSI} \geq 10$ PP-DP		$ \Delta\text{PSI} \geq 10$ DP-N	
		PP>DP	PP<DP	DP>N	DP<N
Exon Skipping (>27 nt)		51	87	222	440
Microexon Skipping (3-27 nt)		5	77	11	128
Alternative 5'SS		4	3	30	20
Alternative 3'SS		11	4	28	22
Intron Retention		23	38	114	199
Total Events (Genes)		94 (82)	209 (170)	405 (320)	809 (508)

Figure 3.9 Classification of splicing events with VAST-tools. On the left: schematic representation of the splicing events. Exons or exonic parts alternatively spliced are depicted as dark blue boxes, not spliced exons as light blue boxes, introns as grey boxes. Continued and dotted lines represent the splice choices leading to inclusion and exclusion of the alternative exonic sequence, respectively. Depending on their size, cassette exons can be classified into microexons (3-27 nt) and exons (>27 nt). Retained introns are depicted as grey boxes with the same thickness as exonic sequences. On the right: number of splicing events per category found at $|\Delta\text{PSI}| \geq 10$ in PP-DP and in DP-N. The spliced events are distinguished in excluded when their PSI is higher in the mother cell compared to the daughter cell (PP>DP and DP>N) and in included when the opposite is true (PP<DP and DP<N). PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; PSI=percent spliced in; nt=nucleotides.

With the same parameters, the analysis detected 809 events (508 genes) and 405 events (320 genes) more and less included, respectively, in DP to N differentiation. The total of 1,214 events were classified as follows: 662 skipped exons (222 Ex, 440 In), 139 skipped

microexons (11 Ex, 128 In), 50 alternative 5'SS (30 Ex, 20 In), 50 alternative 3'SS (28 Ex, 22 In) and 313 retained introns (114 Ex, 199 In) (Figure 3.9).

In this study, the results obtained with VAST-tools were expanded with those obtained from the DEXSeq-PSI pipeline. This approach not only increased the number of events detected for each category present in VAST-tools but also allowed the inclusion of alternative transcription events in the analysis, i.e. alternative first and last exons. Due to the uncertainty of the microexon definition (Irimia et al., 2014; Li YI et al., 2015) and to reach a better overview of the splicing events involved, skipped exons of 28-51 nt in length were classified as a second kind of microexon (28-51 nt).

The final dataset comprises 669 alternative splicing events in the PP-DP transition, 258 events (218 genes) with lower and 411 events (341 genes) with higher inclusion, respectively. The detected PP-DP spliced exons were categorized in the different events as follows: 182 skipped exons (86 Ex, 96 In), 59 skipped microexons (28-51 nt) (12 Ex, 47 In), 123 skipped microexons (3-27 nt) (8 Ex, 115 In), 20 alternative 5'SS (10 Ex, 10 In), 31 alternative 3'SS (23 Ex, 8 In), 76 retained introns (32 Ex, 44 In), 115 alternative first exons (55 Ex, 60 In) and 63 alternative last exons (32 Ex, 31 In) (Figure 3.10).

In the comparison between DP and N, DEXSeq-PSI analysis identified 1,249 additional events for a total of 2,463 events, of which 927 (714 genes) were more excluded and 1,536 (1,078 genes) more included. The identified splicing events comprised: 880 skipped exons (352 Ex, 528 In), 205 skipped microexons (28-51 nt) (39 Ex, 166 In), 232 skipped microexons (3-27 nt) (17 Ex, 215 In), 87 alternative 5'SS (51 Ex, 36 In), 97 alternative 3'SS (49 Ex, 48 In), 424 retained introns (157 Ex, 267 In), 342 alternative first exons (165 Ex, 177 In) and 196 alternative last exons (97 Ex, 99 In) (Figure 3.10).

Microexons of both categories were more frequently spliced in than out during corticogenesis, in agreement with an enrichment of such exons in neural tissues (Irimia et al., 2014; Li YI et al., 2015). Of note, microexon inclusion (115 in the 3-27 nt range and 47 in the 28-51 nt range) represents more than 1/3 of all the included events in PP-DP (162/411 events, 39.42% of PP-DP.In), with shorter microexons (3-27 nt) constituting the most frequent event among the included ones. Although microexons still represent a considerable proportion, about 1/4, of events more included in N than in DP (215 microexons in the 3-27 nt range and 166 microexons in the 28-51 nt range for a total of 381 out of 1,536 included events, 24.80% of DP-N.In), N show a proportionally higher gain of longer exons (528/1,536, 34.38% of DP-N.In).

Splicing Events		Δ PSI ≥ 10 PP-DP		Δ PSI ≥ 10 DP-N	
		PP>DP	PP<DP	DP>N	DP<N
Exon Skipping (>51 nt)		86	96	352	528
Microexon Skipping (28-51 nt)		12	47	39	166
Microexon Skipping (3-27 nt)		8	115	17	215
Alternative 5'SS		10	10	51	36
Alternative 3'SS		23	8	49	48
Intron Retention		32	44	157	267
Alt. First Exon		55	60	165	177
Alt. Last Exon		32	31	97	99
Total Events (Genes)		258 (218)	411 (341)	927 (714)	1,536 (1,078)

Figure 3.10 Classification of splicing events derived by the combination of VAST-tools and DEXSeq-PSI. On the left: schematic representation of the splicing events. Dark blue boxes, light blue boxes and grey boxes represent alternative exonic sequences, not spliced exons and introns, correspondingly. Splicing choices leading to inclusion or exclusion of alternative exonic sequences are depicted as continued and dotted lines, respectively. Depending on the size, cassette exons were separated into microexons (3-27 nt), longer microexons (28-51 nt) and exons (>51 nt). Retained introns are depicted as grey boxes with the same thickness as exonic sequences. On the right: number of splicing events per category found at a $|\Delta$ PSI| ≥ 10 in PP-DP and in DP-N. Excluded events have higher PSI in the mother cells than in the daughter ones (PP>DP and DP>N) while the opposite is true for included events (PP<DP and DP<N). PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; PSI=percent spliced in.

These data indicate a previously unappreciated tendency of neural progenitors to activate a microexon splicing program concomitantly with their neurogenic commitment. To

investigate whether different exon length classes showed a bias not only for the timing of splicing but also for the magnitude of PSI change, I analyzed the dynamic of inclusion of different classes of cassette exons gained during neocortical development.

Microexons (3-27 nt) that gained inclusion during both neurogenic commitment and neurogenesis had low PSI in PP (mean: 13.42) and increased their representation with comparable magnitudes in PP-DP (mean Δ PSI: 34.89) and DP-N (mean Δ PSI: 42.57) reaching an average PSI of 90.88 in N (Figure 3.11, top left). Microexons (3-27 nt) differentially spliced only during neurogenic commitment resulted to be already moderately included in PP (mean PSI: 45.37) and almost doubled their PSI in DP (mean Δ PSI: 36.29) becoming the predominant splice form in committed progenitors and postmitotic neurons (mean PSI: 81.67 in DP and 87.49 in N) (Figure 3.11, middle left). Later included microexons (3-27 nt) were generally absent in PP (mean PSI: 7.50) and DP (mean PSI: 10.09) and although considerably increasing their PSI during neurogenesis (mean Δ PSI: 38.52), they mostly remained the minor splice form in N (mean PSI: 48.61) (Figure 3.11, bottom left).

Similar dynamics, although with generally more gradual changes during neurogenic commitment can be observed for microexons (28-51 nt) alternatively spliced in both PP-DP and DP-N (mean PSI: 15.49 in PP, 43.15 in DP, 87.54 in N; mean Δ PSI: 27.66 in PP-DP and 44.39 in DP-N) (Figure 3.11, top center) and only in PP-DP (mean PSI: 53.83 in PP, 70.05 in DP, 74.96 in N; mean Δ PSI: 16.22 in PP-DP) (Figure 3.11, middle center). Microexons (28-51 nt) differentially spliced only during neurogenesis were also low represented in neural progenitors (mean PSI: 17.62 in PP and 19.81 in DP) while they were included in half of the cases in N (mean PSI: 50.39; mean Δ PSI: 30.58) (Figure 3.11, bottom center).

Longer exons differentially spliced at each differentiation step were usually already present, although at low levels, in PP (mean PSI: 26.06) and showed steeper PSI gains during neurogenesis (mean Δ PSI: 39.90; mean PSI in N: 89.23) than during neurogenic commitment (mean Δ PSI: 23.31; mean PSI in DP: 49.36) (Figure 3.11, top right). Exons differentially spliced only in PP-DP were frequently present already in PP (mean PSI: 55.99) and showed a modest gain in inclusion during neurogenic commitment (mean Δ PSI: 17.41) being spliced-in ~70% of the times in DP (mean PSI: 74.46) and N (mean PSI: 73.40) (Figure 3.11, middle right). Exons regulated only during neurogenesis showed a broad range of inclusion patterns with a bias for low PSI in progenitors (mean PSI: 31.08 in PP

and 32.70 in DP), giving an average increase in their representation of ~30% in the DP-N transition (mean Δ PSI: 27.52; mean PSI in N: 60.22) (Figure 3.11, bottom right).

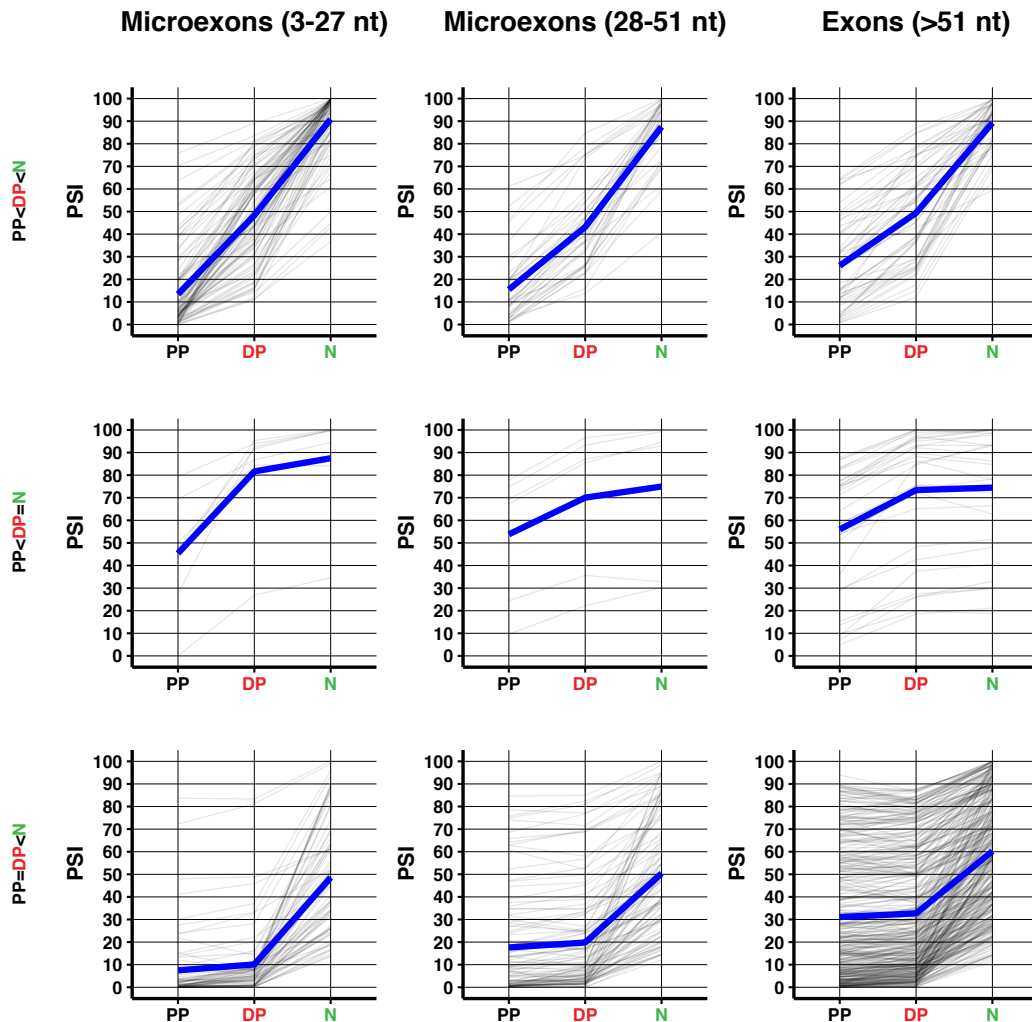


Figure 3.11 Pattern of exon inclusion in different length classes of cassette exons. Each grey line represents the PSI change dynamics of a single exon in the three cell populations. Thicker blue lines represent the mean PSI change dynamic of the group of exons per inclusion pattern and length class. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; PSI=percent spliced in; nt=nucleotides.

In summary, although various patterns could be distinguished within each exon length class, early activated microexons (3-27 nt) present, on average, a more pronounced tendency to steeply gain inclusion during neurogenic commitment (mean Δ PSI range ~35-36) compared to longer microexons (mean Δ PSI range ~16-28) and longer exons (mean Δ PSI range ~17-23).

3.5 Alternative splicing changes the protein output of genes involved in neurogenesis

3.5.1 Spliced genes are involved in neurogenesis and signaling

To examine the role of genes differentially spliced in the neurogenic lineage, a gene ontology (GO) analysis was performed using DAVID (web-based Database for Annotation Visualization and Integrated discovery) (Huang DW et al., 2009a, b) for gene ontology terms in biological and molecular function, cellular components and biological pathways from the Reactome database (Fabregat et al., 2018). The enrichment in biological terms of genes differentially spliced between two cell populations relative to multiexonic genes expressed in the corresponding cell populations was assessed with the Functional Annotation Clustering module, which groups highly related enriched biological terms in functional clusters and provides an enrichment score. Highly similar clusters with redundant GO terms were manually joined together in a singular cluster. The relationships between enriched GO clusters was visualized with Cytoscape (Shannon et al., 2003) (Figure 3.12): the clusters size is proportional to the number of genes included while the thickness of the lines linking clusters represents the number of genes shared by the connected clusters.

Differentially spliced genes in PP-DP are enriched in terms associated with neuronal development (enrichment score: 12.20). The genes belonging to this cluster are also functionally related to clusters of cell motility and regulation of cell development that show a modest enrichment (enrichment scores: 6.31 and 5.13). Other notable enriched GO clusters are GTPase activity (enrichment score: 7.16), synapse formation and function (enrichment score: 6.74), cytoskeletal components (enrichment score: 6.46) and plasma membrane organization (enrichment score: 5.64), while cell-cell junction components, vesicle membranes and protein complex assembly show a low enrichment score (3.79, 3.16 and 3.05, correspondingly) (Figure 3.12, top).

Among the GO terms of genes that change their splicing profile in the DP-N transition, the highest enriched cluster is represented by genes with functions in synaptogenesis and synaptic transmission (enrichment score: 13.78).

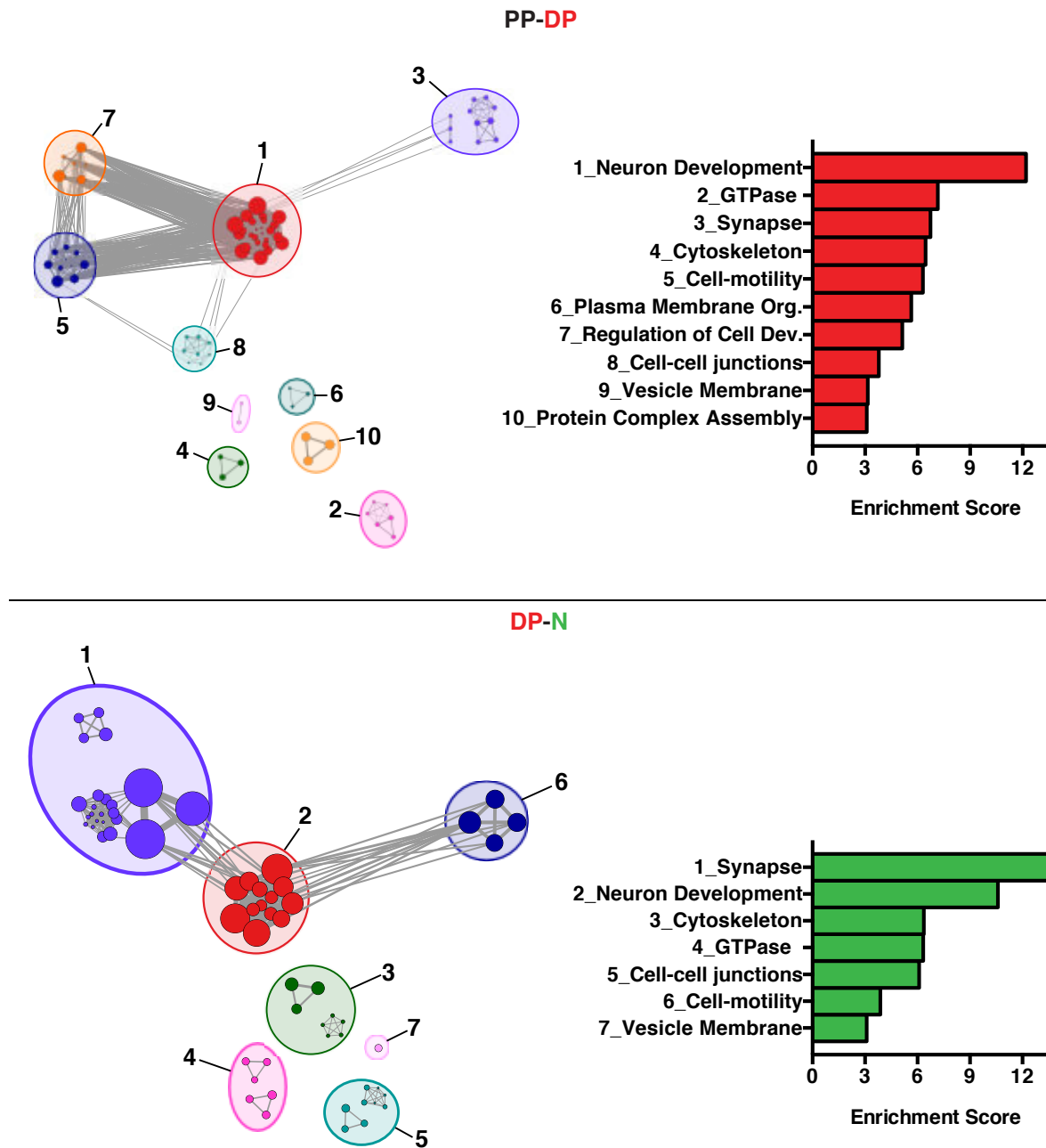


Figure 3.12 Gene ontology analysis. Enrichment map for gene ontology (GO) terms and Reactome pathways of genes differentially spliced between PP-DP (top) and between DP-N (bottom). Left: Cytoscape plot, each node represents a GO term. The node size is proportional to the genes included in the GO term and the width of the edges to the number of genes shared between GO categories. Right: enrichment score of the most relevant categories as reported by DAVID functional annotation clustering. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons.

A high enrichment score is observed also for the neuronal development cluster (enrichment score: 10.60) that shares several genes with the synapse and the cell motility clusters (enrichment score: 3.88). Cytoskeletal components, GTPase activity and cell-cell junction components clusters also show a modest enrichment score (6.36, 6.33 and 6.05,

respectively). Low enriched clusters are represented by cell motility (enrichment score: 3.88) and vesicle membranes terms (enrichment score: 3.10) (Figure 3.12, bottom).

Overall, the GO term analysis shows that alternative splicing during neocortical development has an impact on genes involved in neurogenesis, neuron maturation, as well as synaptic formation and activity.

3.5.2 Impact of alternative splicing on the proteome

As showed in the previous sections, the bioinformatic analysis detected several differentially spliced genes in cells of subsequent stages of differentiation in the neurogenic lineage and, interestingly, these genes are known to be involved in neurogenesis and neuron maturation, but what is the impact of these detected splicing events on the final gene product? To address this question, I investigated whether the differentially spliced exons were part of untranslated regions or of coding sequences. Furthermore, In the case of coding exons, I assessed their potential to change the open reading frame (ORF) by introducing a frameshift or a stop codon or to alter known protein domains.

When comparing exons alternatively spliced in neurogenic commitment (PP-DP transition), a higher proportion of protein-coding exons was found among PP-DP.In (291/411, 70.80%) than among PP-DP.Ex (159/258, 61.63%, p-value=0.018) (Figure 3.13, A-left). PP-DP.In protein-coding exons were proportionally more frame-preserving (196/291, 67.35% vs 77/159, 48.43%, p-value<0.001) but less domain-coding (132/291, 45.36% vs 89/159, 55.97%, p-value=0.040) than the PP-DP.Ex ones (Figure 3.13, B-left and C-left, correspondingly).

A similar trend was observed in the subsequent stage of neural differentiation, with DP-N.In being more often protein-coding (1,000/1,536, 65.10% vs 548/927, 59.12%, p-value=0.003) and frame-preserving (654/1,000, 65.40% vs 254/548, 46.35%, p-value<0.0001) than DP-N.Ex (Figure 3.13, A-right and B-right, respectively). A lower proportion of coding exons overlapping a known domain in DP-N.In compared to DP-N.Ex (493/1,000, 49.30% vs 313/548, 57.12%, p-value=0.004) was also observed (Figure 3.13, C-right).

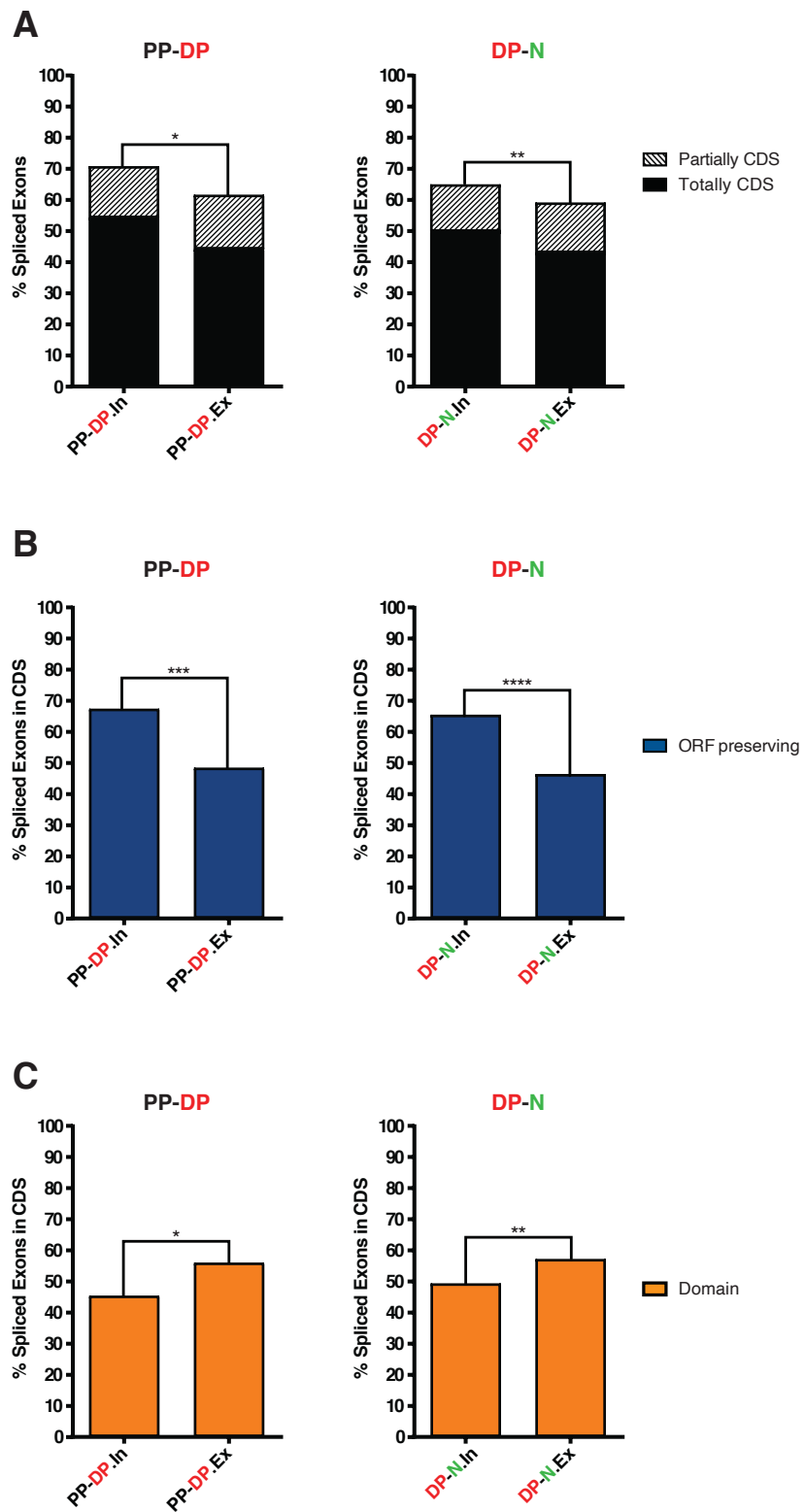


Figure 3.13 Impact of alternative splicing on the proteome. A: proportion of exons partially (Partially CDS) or totally (Totally CDS) overlapping coding sequences (CDS) among exons included vs excluded in PP-DP (left) or DP-N (right) transitions (legend continued on the next page).

(Figure 3.13 legend continuation) Fraction of exons in A preserving the open reading frame (ORF, B) or overlapping an annotated protein domain (C), left PP-DP comparison, right DP-N. *p-value<0.05, **p-value<0.005, ****p-value<0.0001, chi-square test. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons.

Taken together, splicing choices during neocortical development promote the inclusion more often than the exclusion in the nascent transcripts of exons protein-coding, mostly frame-preserving and preferentially not domain overlapping, potentially leading to the production of alternative proteins differing in regions unstructured or with an unknown function.

3.6 Splicing regulation: neural exon features and splicing factor binding

3.6.1 Included neural exons are short and preceded by strong exon-definition features

To gain insight into the mechanism regulating inclusion and exclusion decisions during the splicing of neural alternative exons I investigated several characteristics of cassette exons using the compare features function of Matt (v1.3.0, Gohr and Irimia, 2019). More precisely, exon length, SS strength (Yeo and Burge, 2004), PPT and BrP features have been considered (Corvelo et al., 2010). Polypyrimidine content and length of the predicted PPTs were combined into a score that describes the PPT strength, i.e. the likelihood that the splicing machinery could recognize such sequence. Similarly, for the predicted BrPs, a score was calculated based on the BrP sequence, the pyrimidine content between the BrP and the 3' SS, the distance to the downstream PPT and the score of the PPT itself (Corvelo et al., 2010). The affinity for SF1 binding of the intronic regions upstream of spliced exons (150 nt before the 3' SS) was also included as a proxy of the BrP strength (Corioni et al., 2011).

As shown before (Figure 3.10), microexons constitute a considerable proportion of the cassette exons which are more included during neurogenic commitment and neurogenesis, while they are almost never excluded during these developmental stages. Therefore, it was not surprising to find that included exons in PP-DP and/or in DP-N are shorter than the excluded ones on average (PP-DP median: 33 nt In vs 100 nt Ex, p-value<0.001; DP-N median: 64 nt In vs 102 nt Ex, p-value<0.001, Mann-Whitney U test) (Figure 3.14 A). The different proportion of microexons is not the only factor that influences the length distribution

of In and Ex cassette exon groups. Indeed, when restricting the comparison to exons of the same length class, a general tendency for included exons to be shorter than the excluded ones was observed (exons: 101.5 nt In vs 117 nt Ex, p-value= 0.012, in PP-DP and 101 nt In vs 111 nt Ex, p-value< 0.001, in DP-N, Figure 3.14 B; microexons (28-51 nt): 36 nt In vs 40.5 nt Ex, p-value=0.106 n.s, in PP-DP and 39 nt In vs 44 nt Ex, p-value=0.002, in DP-N, Figure 3.14 C) with the exception of microexons (3-27 nt), for which the few excluded tend to be longer in PP-DP (24 nt Ex vs 18 nt In, p-value=0.057, n.s.), but shorter in DP-N (15 nt Ex vs 18 nt In, p-value=0.352, n.s, Figure 3.14 D).

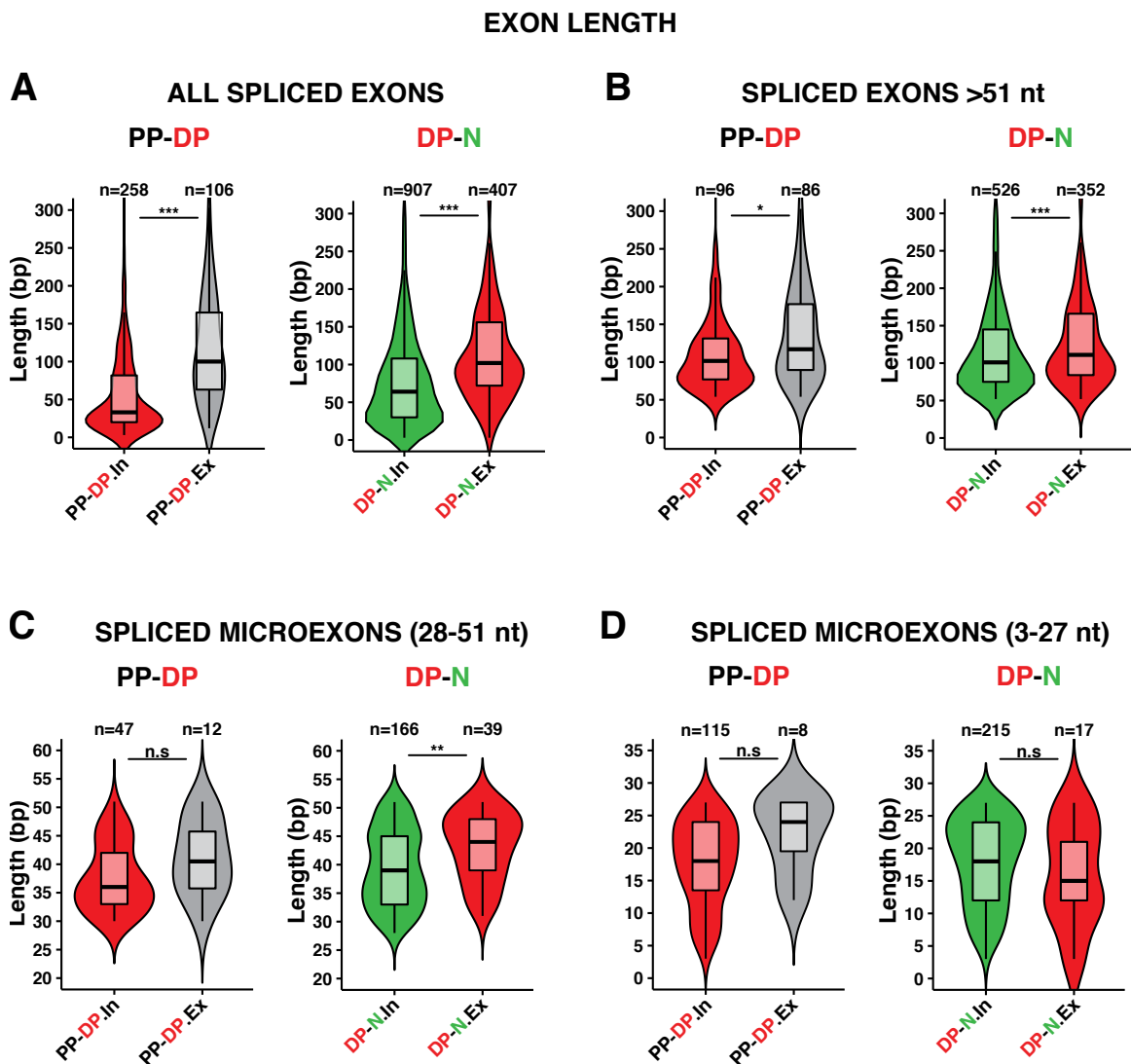


Figure 3.14 Exon features of spliced cassette exons. Violin plots and boxplots showing the distribution of exon length of cassette spliced exons in PP-DP and DP-N. Boxplot elements: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile. P-value from Mann-Whitney U tests: *p-value ≤ 0.05 , **p-value ≤ 0.01 , ***p-value ≤ 0.001 . PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; nt=nucleotides; In=included; Ex=excluded.

Exons that were more represented in cells of the neurogenic lineage as development proceeds, were also characterized by features in the upstream intron that facilitate the exon definition and, therefore favor inclusion. In particular, they showed higher BrP (PP-DP: 1.50 In vs 1.14 Ex, p -value <0.001 ; DP-N: 1.42 In vs 1.17 Ex, p -value <0.001 , Figure 3.15, A) SF1-binding (PP-DP: -5.99 In vs -6.36 Ex, p -value=0.020; DP-N: -6.08 In vs -6.37 Ex, p -value <0.001 , Figure 3.15, B) and PPT scores (PP-DP: 32 In vs 26 Ex, p -value <0.001 ; DP-N: 34 In vs 27 Ex, p -value <0.001 , Figure 3.15, C) relative to excluded exons.

The 3'SS resulted to be weaker (PP-DP: 6.30 In vs 7.53 Ex, p -value <0.001 ; DP-N: 7.07 In vs 7.75 Ex, p -value=0.006, Figure 3.15, E) and the 5'SS stronger in included than in excluded exons (PP-DP: 8.68 In vs 7.88 Ex, p -value=0.007; DP-N: 8.34 In vs 7.88 Ex, p -value=0.021, Figure 3.15, F). No significant differences were found in the BrP, PPT and SF1 binding in the downstream intron, as well as in the SS strength of the flanking exons.

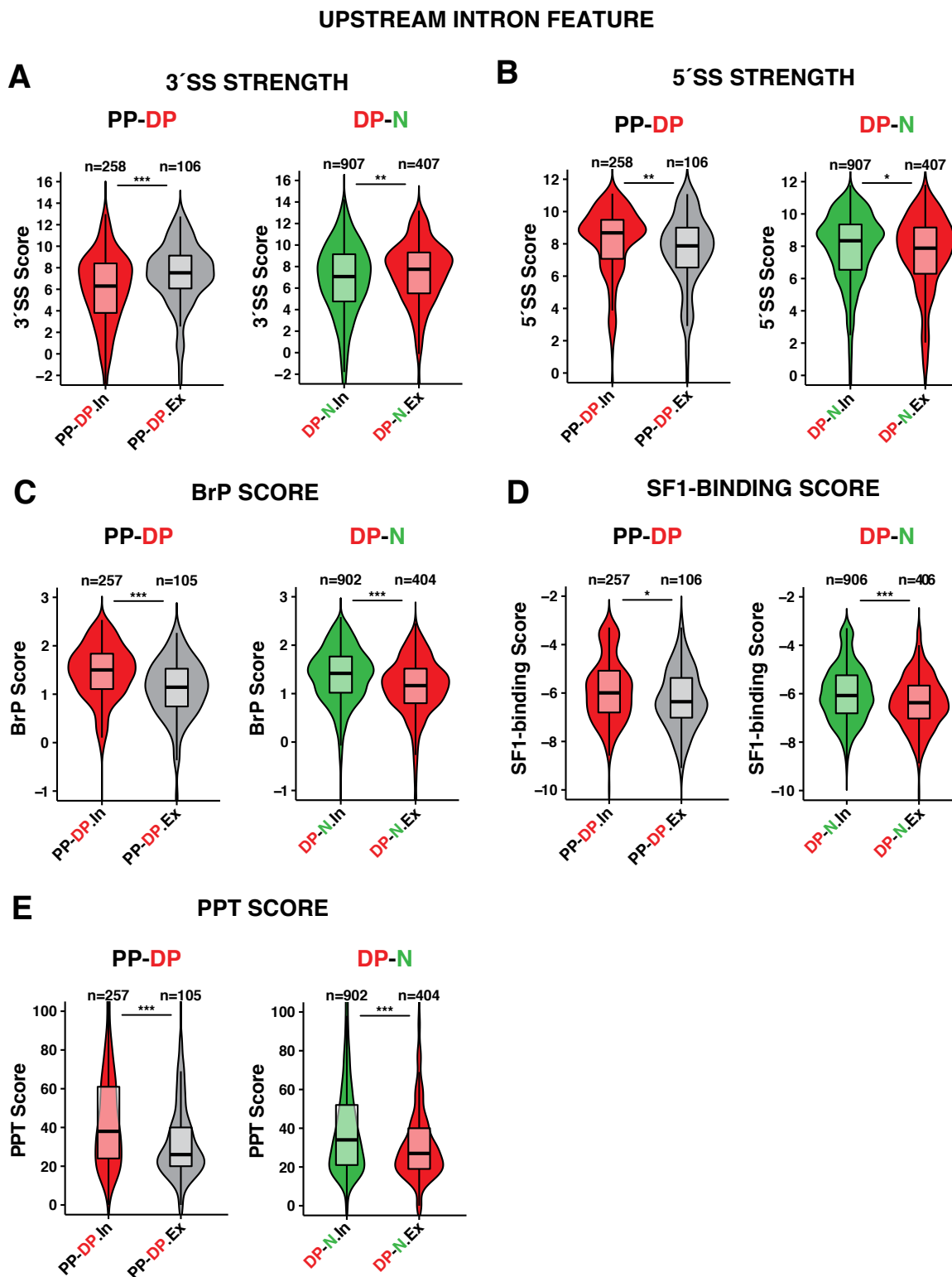


Figure 3.15 Features of introns upstream of spliced cassette exons. Violin plots and boxplots showing the strength score of alternative exon splice sites (SS), branch point (BrP), polypyrimidine tracts (PPT) and SF1 binding sites in introns upstream of cassette spliced exons in PP-DP and DP-N. Boxplot elements: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile (legend continued on the next page).

(Figure 3.15 legend continuation) P-value from Mann-Whitney U tests: *p-value ≤ 0.05 , **p-value ≤ 0.01 , ***p-value ≤ 0.001 . PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; nt=nucleotides; In=included; Ex=excluded.

Other groups (Li YI et al., 2015; Torres-Méndez et al., 2019) had reported that microexons are characterized by stronger 5'SS, BrP and PPT in the upstream intron, and by weaker 3'SS, relative to longer exons. When comparing alternative microexons (3-27 nt; n=614) and longer exons (>51 nt; n=7,932) annotated in the present dataset, regardless of their inclusion levels or splicing patterns, the presence of stronger BrPs (median score: 1.34 vs 1.07, p-value<0.001) and PPTs (median score: 34 vs 25, p-value<0.001) in the upstream intron as well as weaker 3'SS (6.09 vs 7.49, p-value<0.001) and stronger 5'SS (8.52 vs 7.65, p-value<0.001) for microexons (3-27 nt) was confirmed (Figure 3.16). The 3'SS of the downstream exon resulted to be modestly weaker than the one of alternative longer exons (downstream exon 3'SS: 8.77 vs 8.95, p-value=0.003) while the BrP of the downstream intron showed a score slightly higher than the one of longer exons (1.25 vs 1.18, p-value=0.005, Figure 3.16). The second class of microexons (28-51 nt; n=986) instead showed mildly higher scores than longer alternative exons for almost all the features considered: 5'SS strength of the upstream exon (9.14 vs 8.92, p-value=0.004), BrP strength in the upstream intron (1.14 vs 1.07, p-value<0.001), PPT strength in the upstream intron (27 vs 25, p-value<0.001), 3' and 5' SS strength (3'SS: 7.98 vs 7.49, p-value=0.001; 5'SS: 8.14 vs 7.65, p-value<0.001) (Figure 3.16).

Since exons of distinct length classes differ in their features and given their highly uneven proportions in the included and excluded groups, I restricted the comparison of exon features to alternatively spliced longer exons only (PP-DP: 96 In, 86 Ex; DP-N: 528 In, 352 Ex), in order to highlight the characteristics most likely linked to changes in inclusion levels. Even when microexons were omitted, included exons showed stronger features in their upstream introns with stronger BrP (PP-DP: 1.44 In vs 1.11 Ex, p-value=0.001; DP-N: 1.38 In vs 1.15 Ex, p-value<0.001) and PPT (PP-DP: 33 In vs 25 Ex, p-value=0.010; DP-N: 31 In vs 26 Ex, p-value<0.001) than excluded exons. Higher affinity for SF1 binding in the upstream intron was observed for exons increasing their PSI during differentiation of DP into N (DP-N: -6.12 In vs -6.38 Ex, p-value<0.001), while only a small trend was observed for exons included during neurogenic commitment (PP-DP: -6.05 In vs -6.30 Ex, p=0.179 n.s) (Figure 3.16). No significant differences were observed for SS strength and features in the downstream intron.

3.6.2 Early included exons are enriched for nSR100 binding sites

Splicing factors are the main players in the determination of tissue-specific splicing profiles. As previously shown, differential gene expression analysis revealed several splicing factors and RNA-binding proteins as being differentially regulated in PP, DP and N. To obtain a deeper understanding of the role of different splicing factors in regulating the inclusion/exclusion of neural alternative cassette exons, I tested the enrichment for splicing factor motifs nearby the spliced exons compared to not spliced exons. Splicing factor binding motifs, annotated in the Catalog of Inferred Sequence Binding Proteins of RNA (CISBP-RNA) (Ray et al., 2013) were searched in the exonic regions (35 nt) at the spliced exon extremities and both at the 5' end of the upstream exon and at the 3' end of the downstream exons. In the same way, intronic regions (135 nt) flanking the spliced exon and both downstream of the preceding exon and upstream of the following exon were scanned.

The highest enrichment was found for binding motifs of the zinc finger protein Zfp36 in the intronic region, proximal to the 5' end of exons which increased their inclusion during differentiation in PP-DP and/or in DP-N transitions. A peak of the Zfp36 binding motif was observed for each group of exons tested irrespective of their change in PSI and in proximity of their downstream exon, suggesting a role for this RNA-binding protein in exon definition. Nonetheless, exons with increasing PSI showed a higher enrichment (~10% more) for Zfp36 binding motif compared to exons less included and to exons not differentially spliced (Figure 3.17). Zfp36 expression showed general low levels and only a modest, not significant increase from PP to N (\log_2FC DP/PP=0.25 and \log_2FC N/DP=0.42). Therefore, it seems unlikely that Zfp36 alone is a major factor in influencing change in splicing choices in PP, DP and N.

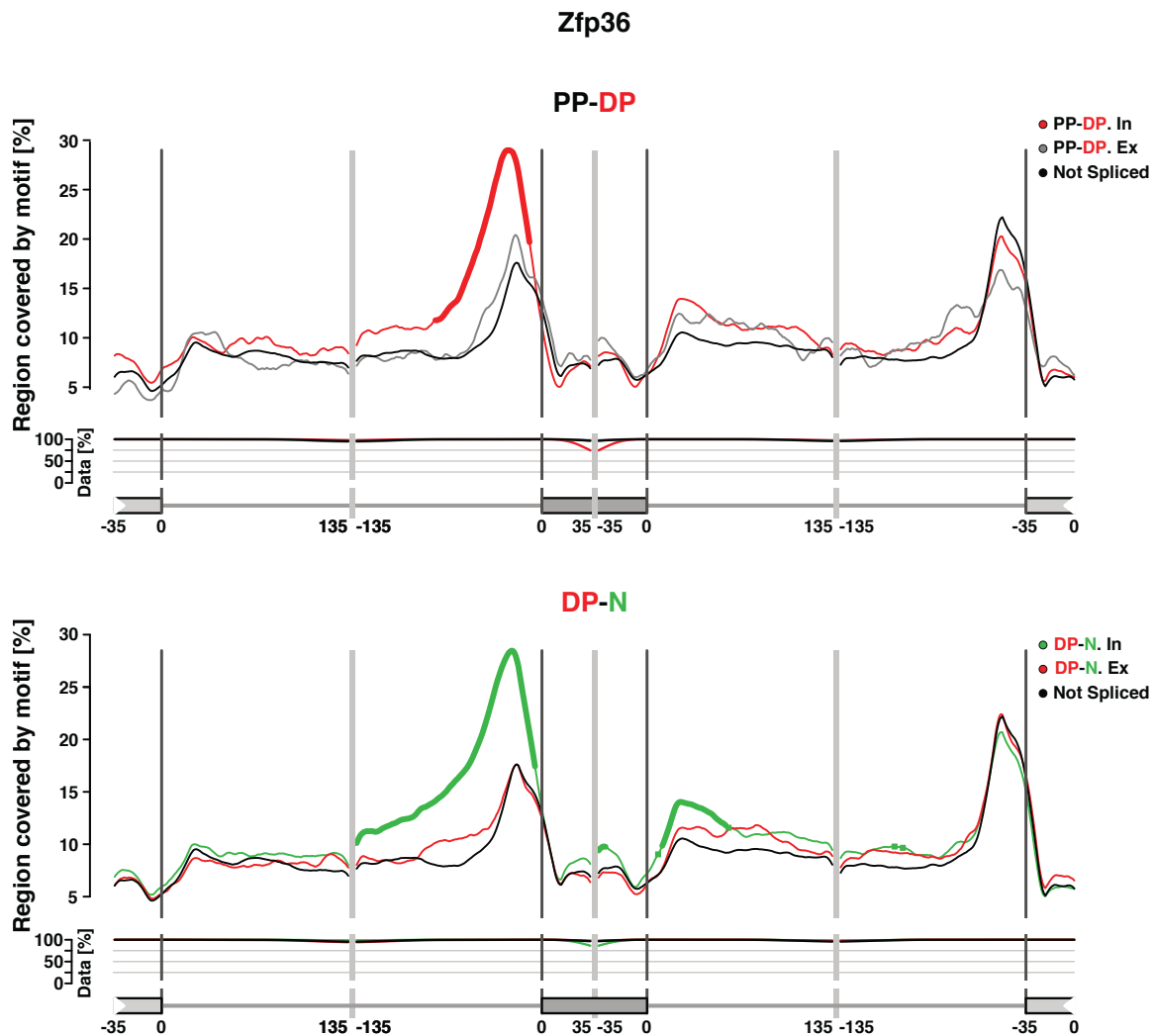


Figure 3.17 Zfp36 binding sites. Enrichment for Zfp36 binding motif nearby spliced exons compared to not spliced exons in neurogenic commitment (top) or neurogenesis (bottom). Length of sliding window: 31 nt. The “Data” subpanel refers to the proportion of total sequences used for the RNA map at each position, as implemented by Matt. Thicker regions are positions showing significant enrichment. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; In=included; Ex=excluded.

When comparing exons alternatively spliced during neurogenic commitment, the second highest enrichment was found for nSR100 binding motifs in the upstream intron in proximity to the 3’SS in gained exons (~6% more in respect to Ex and to not spliced exons). A similar, although less pronounced, pattern (~3% and ~4% more than Ex and not spliced exons, respectively) was observed also when comparing exons differentially spliced in DP-N. Interestingly, an enrichment for nSR100 binding motifs was observed also in the downstream intron of both included and excluded exons in the DP-N transition compared to not spliced exons (Figure 3.18).

Among included exons, microexons (3-27 nt) and microexons (28-51 nt) showed higher enrichment values for nSR100 compared to longer exons, in agreement with the predominant role of this splicing factor in microexon regulation (Supplementary Figure 2)

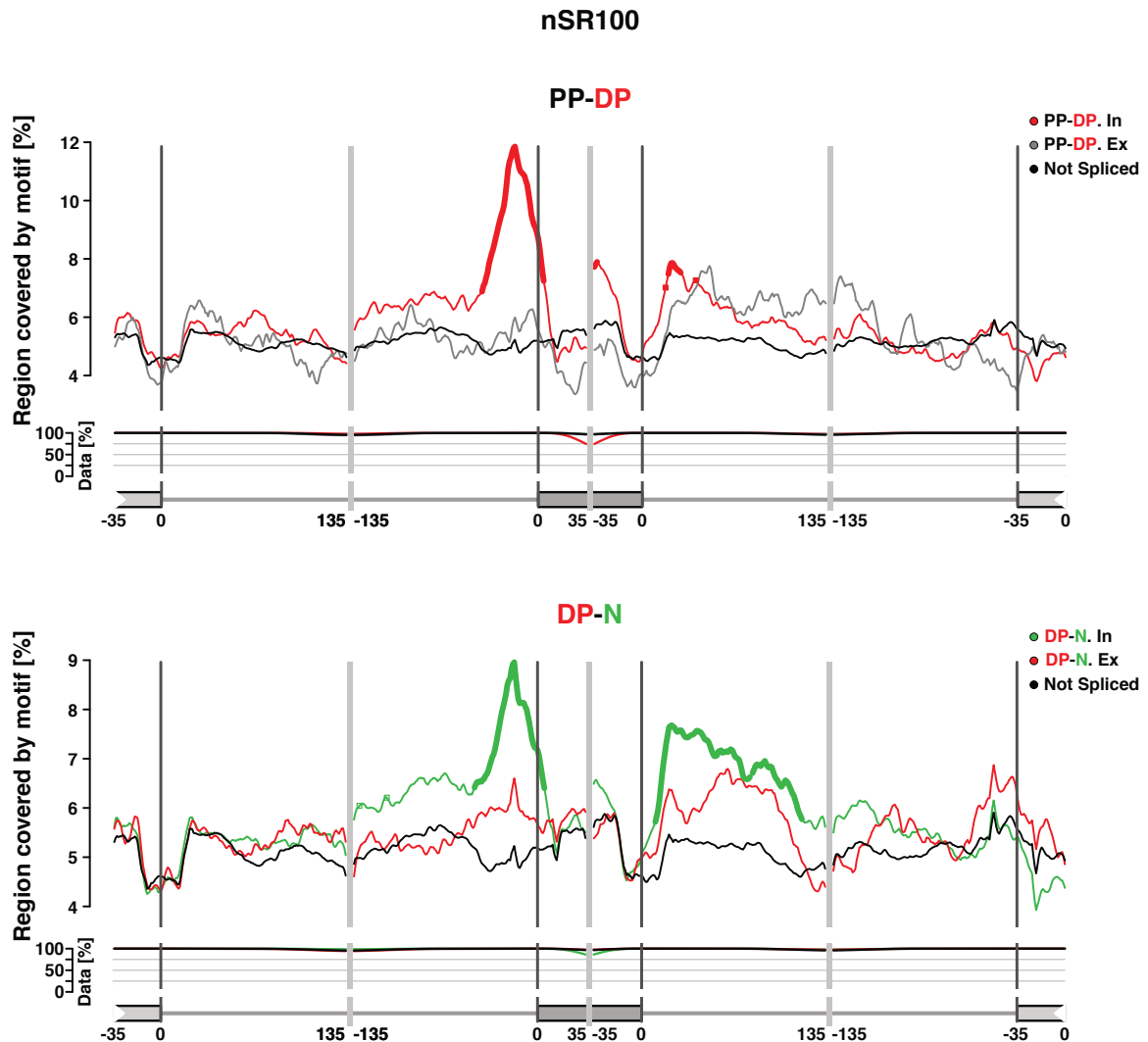


Figure 3.18 nSR100 binding sites. Enrichment for nSR100 binding motif nearby spliced exons compared to not spliced exons in neurogenic commitment (top) or neurogenesis (bottom). Length of sliding window: 31 nt. The “Data” subpanel refers to the proportion of total sequences used for the RNA map at each position, as implemented by Matt. Thicker regions are positions showing significant enrichment. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; In=included; Ex=excluded.

Other enrichments were found for binding sites of Ptbp proteins with included exons in PP-DP and/or DP-N transition showing higher percentages of regions covered by Ptbp-binding sites (Supplementary Figure 6). However, the implication of such enrichments in splicing decision is not obvious since Ptbp1 and Ptbp2 exhibit affinity for the same binding sites (Vuong et al., 2016), but generally different repressive strength (Keppetipola et al., 2012)

and their target exons can be regulated also by nSR100, which promotes exon inclusion (Raj B et al., 2014).

No striking enrichments were observed for other differentially expressed splicing factors, suggesting that, even though they might be important for the regulation of specific subsets of neural exons, they do not represent main directors of a neural-specific splicing program.

The data place nSR100 as the major splicing factor mediating inclusion of exons marking the neurogenic commitment while other RNA-binding proteins seem to have a more marginal role.

3.7 The Btg2^{RFP}/Tubb3^{GFP} mouse line outperforms previous models for the study of cell-type-specific splicing in the brain

Other groups have previously addressed the question of the role of alternative splicing in neurogenic commitment in the mammalian neocortex. Notably, in 2016 the Walsh group reported the first study on cell-specific alternative splicing in the developing mouse neocortex (Zhang et al., 2016). Using a Tbr2^{EGFP} reporter mouse, Zhang et al. analyzed the transcriptome of the apical progenitors in the VZ (VZ cells) compared to a mixture of basal progenitors and neurons (nonVZ cells) at E14.5. The study identified several genes differentially spliced during development, with enrichment in cytoskeletal-related biological terms, and proposed a mechanism by which a splicing event could lead to a change in the plane of cell division, therefore determining neurogenic commitment. As a proof of the quality of our dataset, I analyzed the data of Zhang et al. with the same tool and filters used in my study and compared the differential exon usage of the two datasets (alignment done by Mathias Lesche). Due to the short length of the sequencing data of the Zhang dataset, it was not possible to apply the VAST-tool pipeline that requires read length >50 nt. The comparison was therefore carried on with the exon usage analysis performed with DEXSeq.

The majority of splicing events between PP-N in our dataset and the ones between VZ-nonVZ cells published by Zhang et al. was overlapping (600/906, 66.23% of events reported by Zhang et al. were present also in our dataset) and furthermore, our sequencing revealed to be of higher quality and depth allowing the identification of a greater number of splicing events (6,057 vs 906, 6.69 fold more) (Figure 3.19 B).

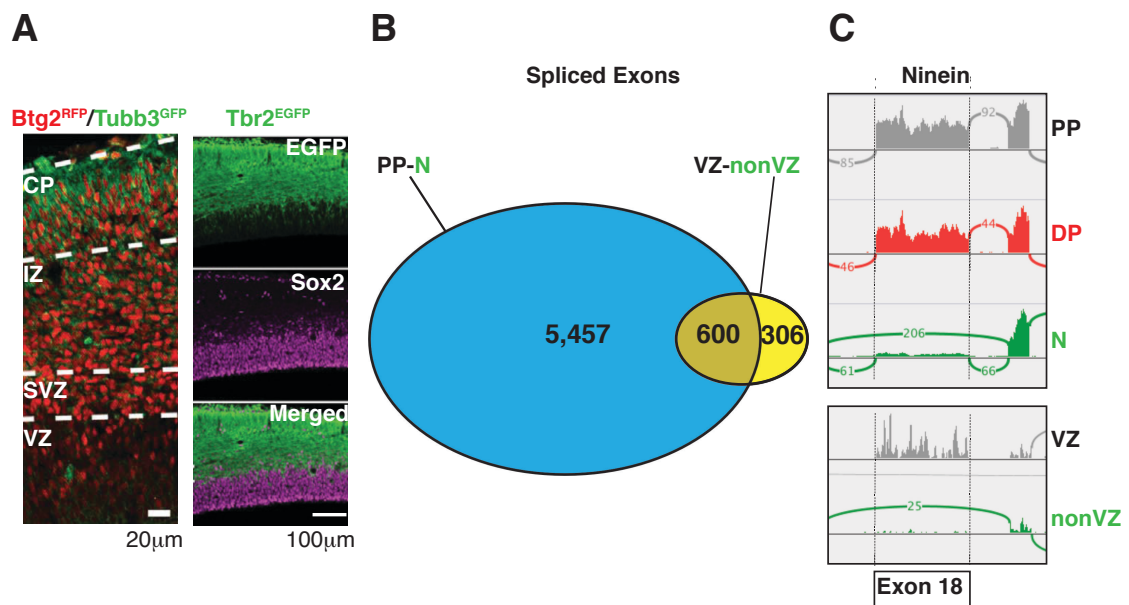


Figure 3.19 Comparison with other dataset of splicing in the mouse neurogenic lineage. A: Fluorescence picture of a cryosection through the lateral cortex of a $Btg2^{RFP}/Tubb3^{GFP}$ (left) and of a $Tbr2^{EGFP}$ (right; immunostained for Sox2 and EGFP) embryo at E14.5 (Aprea et al., 2013; Zhang et al., 2016). In the $Btg2^{RFP}/Tubb3^{GFP}$ mouse line, it is possible to distinguish VZ (RFP^{-}/GFP^{-}), SVZ/IZ (RFP^{+}/GFP^{-}) and CP (GFP^{+}). The $Tbr2^{EGFP}$ reporter mouse shows a clear distinction only between VZ ($EGFP^{-}$) and other regions of the neocortex ($EGFP^{+}$). B: Venn diagram for splicing events detected in the $Btg2^{RFP}/Tubb3^{GFP}$ (PP-N, light blue) and $Tbr2^{EGFP}$ (VZ-nonVZ, yellow) transcriptome analysis. Common splicing events detected are represented in the overlapping area. C: sequencing data of exon 18 of *Ninein*; the peaks correspond to reads mapping to the exons, lines connecting the peaks are reads spanning the exon-exon junction. Exon 18 is included in PP and DP and excluded in N. VZ=ventricular zone; SVZ=subventricular zone; IZ=intermediate zone; CP=cortical plate; PP=proliferating progenitors; DP=differentiating progenitors; N=neurons.

A key difference between the two datasets resides in the advantage of being able to discriminate DP from PP and N with the $Btg2^{RFP}/Tubb3^{GFP}$ mouse line, whereas DP are grouped with N in the nonVZ population in the model of Zhang et al. (Figure 3.19 A). This is critical since the neurogenic commitment, per se, is defined as the switch between PP and DP and therefore, the key players of this process should be identified in the comparison between those two populations. The lack of distinction between DP and N in the Zhang work had an important repercussion. Indeed, to prove that splicing has a role in neurogenic commitment, Zhang and colleagues manipulated *in vivo* in the mouse embryo brain two isoforms of *Ninein* (*Nin*), a centrosomal protein required for the positioning and anchorage of the minus end of microtubule which has been shown to be critical for the maintenance of IKNM of apical progenitors (Shinohara et al., 2013). When overexpressed, the *Nin* nonVZ-specific isoform dramatically decreased the proportion of manipulated cells in the VZ, while the *Nin* VZ-specific isoform did not show any phenotype. The main difference between *Nin*-VZ and *Nin*-nonVZ is the alternative exon 18 (>2,000 nt) which is included in the first isoform and excluded in the second one. Our data confirmed those findings, with PP showing a

strong inclusion of exon 18, which is instead totally absent in N. However, exon 18 shows an inclusion level in DP that is comparable to the one in PP indicating that the splicing event is happening between DP and N and not during the neurogenic commitment (Figure 3.19 C).

In reason of the higher cell-specificity, the higher data quality and the combination of different bioinformatic approaches, the present study constitutes a more accurate and complete dissection of the alternative splicing regulation in the mammalian developing neocortex compared to previous studies. Moreover, the ability to discern between proliferating and differentiating neural progenitors offered by the *Btg2^{RFP}/Tubb3^{GFP}* mouse model is critical for the identification of splicing choices involved in commitment to a neurogenic fate.

3.8 *In vivo* manipulation of splice variants

To assess whether alternative splicing could have a role in neurogenic commitment, I induced a transient overexpression of different isoforms derived from the same gene *in vivo*, in the developing mouse neocortex. As previously shown, alternative splicing during neurogenic commitment and neurogenesis has a major effect in changing the gene protein output. Therefore, I selected genes differentially spliced in the CDS between PP and DP, thus generating alternative protein isoforms, for *in vivo* manipulation. The isoforms of interest were cloned into an RFPnls-expressing vector and overexpressed in the lateral wall of wild type E13.5 mouse embryonic brains by in utero electroporation. The distribution of electroporated cells in the lateral cortex was assessed two days later at E15.5.

Transient overexpression of switching isoforms identified by the splicing analysis at the transcript level highlighted isoform-specific effects on corticogenesis (Supplementary Table 1 and Supplementary Figure 1). However, the switching events tested could not be validated questioning, therefore, the biological relevance of the manipulated isoforms for neurogenic commitment.

Conversely, splicing analysis at the exon level showed a high rate of validation. To investigate the role of alternatively included exons, gene isoforms differing only for the spliced exons were electroporated in the developing neocortex.

As a model case I chose to investigate the splicing of the Fas apoptotic inhibitory molecule (Faim) that showed a gain in the inclusion of one exon of 57 nt (exon 2) in both PP-DP and

DP-N transitions. In the Ensembl mm9 mouse database only 2 isoforms for Faim are annotated. Both are protein coding and differing only for exon 2, which was detected as differentially spliced in our dataset. Moreover, although a modest increase in expression was observed, the gene did not result to be differentially expressed (PP-DP $\log_2FC=0.18$; DP-N $\log_2FC=0.53$), indicating alternative splicing with the switching of the two coding isoforms as the sole form of transcriptional regulation of Faim in the developing mouse neocortex.

The alternative spliced exon 2 is placed upstream the canonical TSS and contains an alternative translation start codon. The inclusion of this exon, mediated by nSR100 (Coccia et al., 2017), results in the insertion of 66 coding nt (44 nt of the spliced exon and 22 nt of the alternative 5'UTR) adding 22 aa with an undefined structure to the N-terminus of the protein, while preserving the remaining protein sequence (Figure 3.20). The two resulting isoforms, Faim short (Faim-S, exon excluded) and Faim long (Faim-L, exon included) differ in their expression pattern and function. Faim-S is ubiquitously expressed whereas Faim-L is neuron-specific (Zhong et al., 2001; Segura et al., 2007).

Consistently, in the data presented here, the alternative exon is mostly absent in PP (PSI 22.64), indicating Faim-S as the major isoform in this cell population. Its PSI increases to 57.67 in DP with Faim-S and Faim-L being roughly equally represented, whereas it is virtually always included in N (PSI 96.55), thus, confirming the neural-specificity of Faim-L.

Preliminary data highlight distinct roles of Faim splice variants in corticogenesis. Overexpression of Faim-S showed a marked decrease in the proportion of neurons reaching the CP, while increasing the population of electroporated cells in the other cortical areas. Faim-L overexpression led only to a modest decrease of cells in the CP, while increasing neurons residing in the IZ (Figure 3.20).

The data support the notion that different splice variants exert specific functions and that the switch in splicing pattern contributes in regulating neocortical development under physiological conditions.

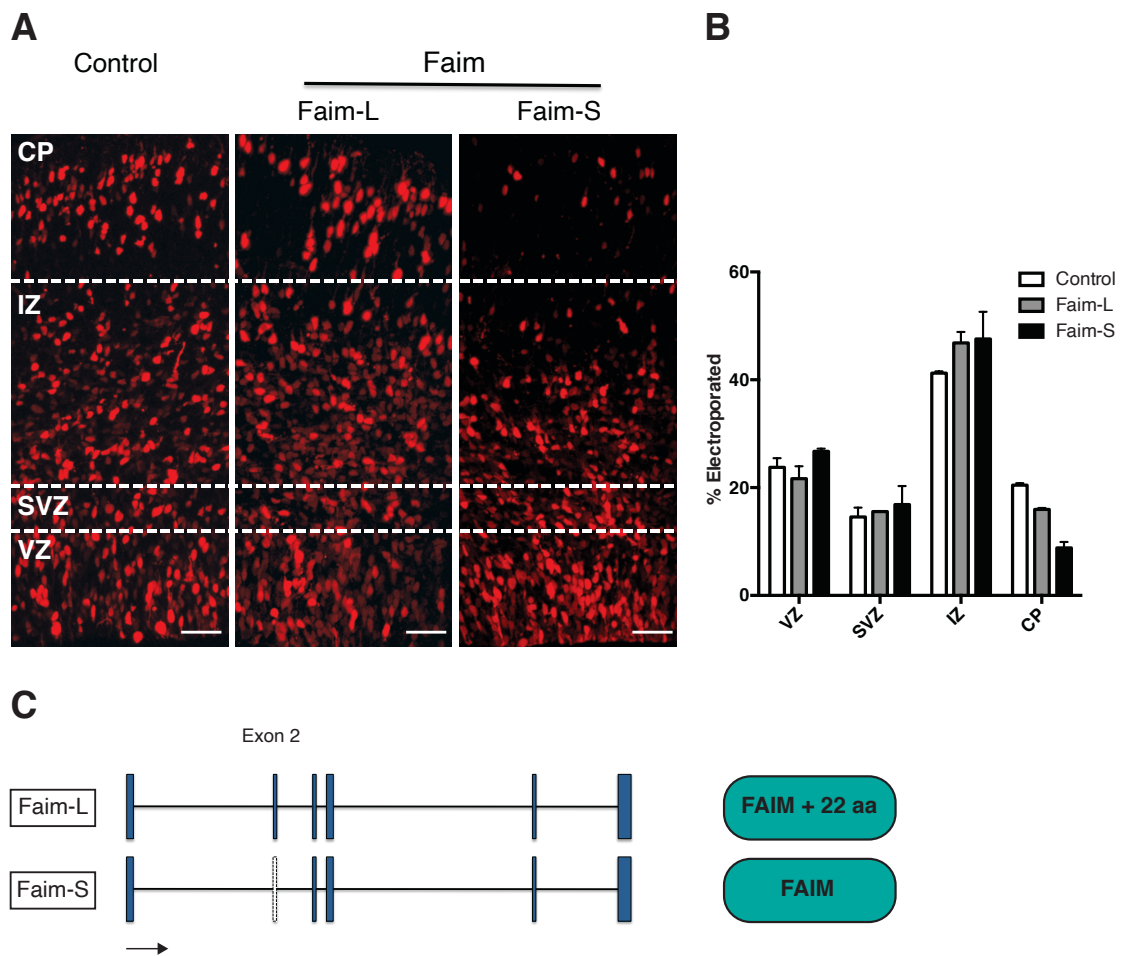


Figure 3.20 *In vivo* assay of Faim isoforms. A: fluorescence images of cryosections through the E15.5 lateral cortex of mouse embryo brains electroporated with an empty RFPnls vector (control) or a splice variant-expressing RFPnls plasmid (Faim-L or Faim-S). B: distribution of electroporated cells in the different regions of the neocortex. Percentages are calculated based on the total of electroporated cells. Mean and standard error of the mean (SEM) are reported (n=2). C: transcript structures of the manipulated long (Faim-L) and short (Faim-S) Faim isoforms. Blue boxes represent exons, in scale, that are included in the transcripts, white boxes show exons excluded compared to the complementary isoform. Lines represent introns. The arrow indicates the direction of transcription. VZ=ventricular zone; SVZ=subventricular zone; IZ=intermediate zone; CP=cortical plate; aa=amino acids. Scale bars in A =50 μ m.

Contribution	Author
Sequencing of transcriptome of PP, DP and N derived from Btg2 ^{RFP} /Tubb3 ^{GFP} murine line	Julieta Aprea (75 bp single-end Aprea et al., 2013; 100 bp paired-end Aprea et al., 2015)
Alignment of Btg2 ^{RFP} /Tubb3 ^{GFP} sequencing data and gene expression analysis with DESeq	Deep Sequencing Group
Cuffdiff2 pipeline for isoform switching analysis	Mathias Lesche, Deep Sequencing Group
DEXSeq pipeline for exon usage	Mathias Lesche, Deep Sequencing Group
Alignment of transcriptome of VZ and nonVZ cells derived from Tbr2 ^{EGFP} reporter mouse from (Zhang et al., 2016)	Mathias Lesche, Deep Sequencing Group

Table 3.1 External contribution to the project.

4 Discussion

How cells with equal genomic information acquire specific identities during embryonic development to form a complete and fully functional organism is still an unanswered question in developmental biology. With the advent of next generation sequencing researchers started to investigate changes in transcription that associate with cell-fate commitment and identified several genes that are activated or repressed during differentiation. However, most of these studies did not take into account the variety of isoforms derived from alternative splicing and transcription mechanisms. While the fact that a gene could potentially give rise to several products was known since the late '70s, it was generally assumed to be a marginal phenomenon and that most genes would express prevalently only one isoform. In the last decades this view has been completely overturned and it is now accepted that alternative splicing and transcription are widespread processes, which produce several functional isoforms with specific roles in different biological aspects. Alternative isoforms could explain the staggering discrepancies observed between organisms with very similar number of genes but highly different phenotype. However, the function of specific isoforms is largely unclear and their potential role in defining cell-identities has not yet been clarified because of a still persistent bias considering each gene as encoding mainly one product and due to the technical limitations in detecting alternative splicing changes and in manipulating single splice variants.

In this project, I addressed these issues investigating the role of alternative splicing in cell-fate commitment in a complex tissue such as the developing mammalian neocortex. For this purpose, I analyzed the transcriptome of cells of the neurogenic lineage at subsequent stages of differentiation, isolated from the $Btg2^{RFP}/Tubb3^{GFP}$ mouse line. This mouse line demonstrated to be a powerful tool for the identification of novel transcripts involved in neurogenic commitment and neurogenesis, such as ncRNA (Aprea et al., 2013, 2015), circRNA (Dori et al., 2019) and miRNA (Dori et al., 2020). By employing several bioinformatic tools, I characterized for the first time the splicing profiles of PP, DP and N, identified specific splicing events that marked the neurogenic commitment, predicted their

functional impact on the protein output of the gene and shed some light on the regulation of such splicing choices. Moreover, *in vivo* manipulation highlighted isoform-specific effects on neurogenic commitment, strongly suggesting a causal relationship between splicing regulations and cell-fate commitment.

4.1 The combination of different bioinformatic approaches allows an accurate identification of splicing events at the exon-level

The differential gene expression analysis of PP, DP and N transcriptomes revealed that several splicing factors and RNA-binding proteins show different expression levels in cells of the neurogenic lineage during embryonic development. Notably, well-known factors dedicated to splicing of neural-enriched events such as nSR100 and Nova2 started to be up-regulated already during neurogenic commitment. Such differential expression pattern is a strong indicator that cells of the neurogenic lineage might be characterized by different splicing profiles and that alternative splicing choices that take place at subsequent steps of neurogenic differentiation might be important for the acquisition of a neuronal identity. To test this hypothesis and to identify which genes are differentially spliced during neocortical development, I applied several bioinformatic tools.

I first tried to achieve an isoform-centric view of the splicing profile of cells of the neurogenic lineage, by performing an isoform-switching analysis with Cuffdiff2 (Trapnell et al., 2012, 2013). With this approach, several genes with a known function in neurogenesis and cell-fate determination, such as components of the Wnt pathway, were found to switch protein variant expression as a result of differential splicing during neurogenic commitment (Figure 3.2 and Supplementary Table 1). Validation by qPCR confirmed a trend of change in isoform expression and proportion generally in agreement with the bioinformatic prediction, but of lower magnitude than expected (Figure 3.6).

Analysis at the exon level with DEXSeq (Anders et al., 2012), DEXSeq-PSI (Schafer et al., 2015) and VAST-tools (Irimia et al., 2014; Tapial et al., 2017) detected a higher number of genes undergoing alternative splicing in the developing mouse neocortex than isoform-based tools (Figure 3.3 and Figure 3.5). Although all these tools investigate splicing from an exon-centric point of view, the differences in these methods caused a discrepancy in the number and magnitude of splicing changes detected (Figure 3.4). Visual inspection of the distribution of aligned reads on different exons supported splicing events detected with

methods adopting the PSI metric, while they were often inconsistent with the exon usage changes reported by DEXSeq as \log_2FC . Therefore, I opted for a combination of DEXSeq-PSI and VAST-tools using the PSI as a metric for differential splicing. The dataset obtained in this way achieved a 100% validation rate with a high correlation between bioinformatically predicted and experimentally measured magnitudes of ΔPSI (Figure 3.7).

The dissimilarity in results obtained with various bioinformatic tools shows that detecting splicing is still a very tricky task. First of all, splicing can be looked at from two different points of view: at the isoform and at the exon level. Some bioinformatic packages, such as MISO (Katz et al., 2010) and SUPPA2 (Trincado et al., 2018), provide the option to run the analysis at both levels. Even in such cases, only a subset of genes would result differentially spliced in both modalities. Indeed, the transcripts of a gene could significantly change their proportion in different conditions or cell populations, but they might share exons in such a way that not a single one by itself results differentially included. On the other hand, an exon shared by multiple isoforms all changing their proportion modestly but in the same direction would result alternatively spliced although the overall proportion of a gene isoform is not significantly altered.

When considering splicing from the isoform point of view, additional problems are apparent. With traditional RNA-Seq technology, the isoform expression must be reconstructed from short reads that cannot, in most of the cases, be uniquely assigned to a splice variant. Several mathematical models have been developed to estimate the expression of single transcript from short read sequencing, with or without the aid of a previous annotation. Even though they provide a quite good approximation, they still show a high degree of error in both the false positive and false negatives components (Alamancos et al., 2014; Hooper, 2014; Mehmood et al., 2019).

Pipelines that allow the sequencing of isoforms in their full length such as PacBio (Eid et al., 2009; Rhoads and Au, 2015) and Nanopore (Jain et al., 2018) represent undoubtedly a great advancement in the field. However, the error rate, the bias for transcript length, the high amount and quality of material required preclude their application for quantification purposes. Nonetheless, PacBio and Nanopore technologies are useful for the annotation of novel genomes and for the improvement of transcriptome annotation of model organisms, including human and mouse. The combination of long- and short reads sequencing data seems promising to better quantify splice variants (Gupta et al., 2018; Tardaguila et al., 2018; Hardwick et al., 2019; Stark et al., 2019).

In spite of the uncertainty in detecting differential splicing, all the analysis employed in this study revealed the existence of a cell-specific splicing program in the cells of the neurogenic lineage at different stages of differentiation. The combination of multiple exon-centric pipelines allowed the identification of alternatively spliced exons and a reliable quantification of their inclusion changes during neocortical development.

4.2 Splicing choices during neurogenic commitment establish a neural signature characterized by microexon inclusion

A high proportion of inclusion events was detected at both the neurogenic commitment and the neurogenesis steps, while fewer exons showed a loss in inclusion during neocortical development (Figure 3.8 A). The events detected in the switch from proliferation to differentiation were mostly kept with the same inclusion level or further spliced with the same pattern in neurons indicating that the acquisition or loss of these sets of exons are related to the definition of a neuronal identity. In this regard, taking into account the magnitude of the Δ PSI, it is interesting to observe an inversion of the overall exon representation from PP to N for the majority of splicing events detected during the neurogenic commitment step: gained exon, while poorly represented in PP, became the predominant splice variant in N. The opposite pattern was observed for lost exons. This was particularly striking in the PP-DP.In group, that reached a PSI close to 100 in N (Figure 3.8 B, top-left). In other words, there is a subset of alternative splicing events mostly absent in PP that starts to be included when a neural progenitor commits to a neural fate in DP and becomes always included in N.

Splicing events belonging to this group comprise mostly cassette exons (62.77%) and especially microexons (3-27 nt) (27.98%) (Figure 3.10). This particular class of exons, previously disregarded as an annotation anomaly, has been recently re-evaluated as a main signature of the nervous system. Neural-specific microexon programs have been found in several species with highly conserved microexons sequences and flanking intronic regions (Irimia et al., 2014; Li Yi et al., 2015; Torres-Méndez et al., 2019). *In vitro* studies to monitor microexon inclusion during mouse embryonic stem cell differentiation into glutamatergic neurons showed an activation of microexon programs at later stages of neural development, with most microexons displaying a sharp increase in PSI in differentiating post-mitotic neurons. Although higher PSI values were observed at later stages of neuronal maturation,

several microexons started to be included with modest PSI in the transition from NEC-like state to RG-like state and showed an almost linear increase in PSI at subsequent stages of post-mitotic neuronal maturation (Irimia et al., 2014) .

In the data presented here, I observed an activation of a microexon program before the generation of postmitotic neurons, at the neurogenic commitment stage, with microexons (3-27 nt) as the most represented class among PP-DP.In (Figure 3.10). This group of microexons is generally absent or scarcely represented in PP, it gains inclusion in DP and N, with changes of comparable Δ PSI magnitude from mother to daughter cells, and reaches, on average, an almost total inclusion in N. Microexons (3-27 nt) that are activated at later stages, in DP-N transition, are also scarcely represented in PP and DP and reach a mean PSI of ~50% in newly generated neurons. Similar patterns could be observed for included longer microexons (28-51 nt) and longer exons (>51 nt), although these classes of cassette exons show generally milder inclusion increments during neurogenic commitment (Figure 3.11). Consistently with the literature, very few microexons of the 3-27 nt or 28-51 nt group were excluded during neocortical development. Although different clusters of inclusion patterns could be distinguished among microexons as well as among other classes of cassette exons, these data do not support the notion that microexon program activation is a feature of differentiating post-mitotic neurons. Instead, their acquisition at early stages of development suggests a primary role of a group of microexons in the determination and maintenance of a neurogenic identity.

4.3 Splicing during neocortical development leads to the generation of alternative protein isoforms in genes involved in neurogenesis and signaling

Overall, splicing choices that take place during neurogenic commitment are kept or even exacerbated during neurogenesis and are characterized by the gain of transcript regions with a prevalence of microexons, a known neural-enriched class of exons. This pattern of inclusion and the kind of splicing events involved suggest that the definition of a specific splicing profile is important for the acquisition of a neurogenic identity. Consistently, when compared to multiexonic genes expressed in the cells of the neurogenic lineage, genes that undergo alternative splicing during neurogenic commitment are enriched for terms related to neural development (Figure 3.12). A high enrichment in neural development processes was also present in genes differentially spliced between DP and N, although the most

enriched term in this group was synapse formation and function, in agreement with a critical role of splicing in neuronal maturation (Grabowski, 2011; Norris et al., 2014; Raj B and Blencowe, 2015; Liu J et al., 2018; Su et al., 2018). Other enriched terms observed in both PP-DP and DP-N alternatively spliced genes were GTPase activity and cytoskeletal components. Similar terms were previously reported to be enriched also in genes showing neural-regulated splicing in the coding region (Irimia et al., 2014). In the developing mouse neocortex, other researchers reported a strong enrichment for cytoskeletal component terms and more modestly for GTPase activity and synapse formation in genes that changed their splicing profile from proliferating progenitors to more differentiated cells (Zhang et al., 2016). In addition to cytoskeletal terms, Liu and collaborators noticed an enrichment also in genes involved in neural development, chromatin modifications, retrograde transport from the endosome to the Golgi, cell motility and neurite outgrowth in genes showing alternative exon skipping in neural progenitors and neurons (Liu J et al., 2018). Taking into account the differences among these studies with regard to the model used, the bioinformatic pipelines employed and the group of spliced genes considered, these similarities in enriched GO terms are quite remarkable. Compared to similar studies, the stronger enrichment for neural development terms observed in the presented data could be explained not only by the more comprehensive splicing analysis conducted but also by the finer dissection of cells populating the developing mouse neocortex. Indeed, the isolation of PP, DP and N allowed the identification of splicing events specifically associated with neurogenic commitment, separating them from those related to neurogenesis and neuronal maturation processes.

Notably, compared to transcript regions lost during differentiation, exons gained in PP-DP and DP-N transitions were proportionally more protein coding and frame-preserving (Figure 3.13 A and B), thus they are predicted to lead to the production of alternative protein isoforms. The finding that the majority of protein coding exons excluded are able to disrupt the ORF is also indicative of an increase in alternative protein isoforms, since the loss of ORF-disrupting regions such as premature stop codons and frame-shifting exons would most likely restore a functional transcript encoding a full-length protein.

Gained exons encoding for a protein were overlapping known domains or annotated signatures less frequently than protein coding lost exons (Figure 3.13 C). The majority of annotated features in the available protein databases are domains, i.e. protein sequences able to fold independently into a three-dimensional structure that defines their function. Protein sequences devoid of annotation are usually unable to adopt a defined structure and

are therefore considered disordered regions (van der Lee et al., 2014). Such disordered regions have been generally considered simple linkers between domains, without any function other than allowing a proper conformation of the protein and the correct positioning of its key domains. However, it is now established that disordered regions often harbor sites for important post translational modifications, can bind other proteins, thus mediating protein-protein interactions and are able to adopt a structured conformation upon binding to specific ligands. The flexibility and lack of steric hindrance of the disordered regions represent advantages for binding to enzymes for post-translational modifications and to other protein partners to form large protein assemblies, as exemplified by the fact that hubs in protein-protein interaction networks are characterized by intrinsic disorder (van der Lee et al., 2014; Latysheva et al., 2015). Another remarkable property of proteins with high disorder is their ability to undergo phase transition forming membraneless compartments in a cell, concentrating components with shared functions (van der Lee et al., 2014; Uversky, 2017; Alberti et al., 2019).

The fact that exons gained during neocortical development encode preferentially protein segments devoid of annotation seems to indicate that they mainly add disordered regions to the final protein, potentially increasing its ability to interact with binding partners. This observation is in line with previous reports that tissue-specific exon networks usually encode disordered regions (Buljan et al., 2012). However, caution must be taken when making such conclusions. Indeed, a sharp distinction between ordered and disordered proteins is probably not reflecting the reality and it has been proposed that proteins are present in a conformational continuum from totally disordered to fully structured. In this scenario, intrinsically disordered proteins would move more dynamically from one state to the other sampling different conformations (Uversky, 2017).

Moreover, it is important to consider that protein sequences positioned nearby domains can alter the properties of a domain, for example by inducing conformational changes that can mask or expose the functional domain. To alter protein function, it is not necessary the differential splicing of large parts of the coding sequence: insertion or removal of few amino acids in key positions can induce dramatic changes in the protein properties. A renowned example is the splicing of Piccolo. This large protein (>500kDa) of the presynaptic active zone undergoes alternative splicing in one of its two calcium-binding domains, in which the insertion of exon 15 encoding just 9 aa induces a conformational change that leads to a reduced Ca^{2+} affinity (Garcia et al., 2004).

The observation that exons differentially spliced in the developing mouse neocortex are generally encoding for peptides that might alter the protein function or its ability to bind specific partners reshaping the protein-protein interaction network suggests a coordinated change in splicing decision of several exons under a tight regulation.

4.4 Neural exons are short and present strong features facilitating inclusion

Exons that gain inclusion as neocortical development proceeds are characterized by a generally short length and are preceded by introns with features that favor exon definition and thus inclusion (Figure 3.14 and Figure 3.15). As already discussed, a high proportion of included exons is represented by microexons, whereas these exon classes and especially the microexons (3-27 nt) are seldom excluded in DP and N.

As the name suggests, microexons are defined primarily by their length and ironically the value of this criteria is still debated, with groups considering microexons only exons of 3-27 nt (Irimia et al., 2014) and others accepting a broader range up to 51 nt (Li YI et al., 2015). In this study, I adopted both definitions to classify microexons, distinguishing cassette exons into microexons of 3-27 nt, 28-51 nt and longer exons above 51 nt. The short length poses a constrain to microexon recognition by the Pol II and splicing machinery, impairing their inclusion. Most likely to compensate for such impairment, microexons present stronger features in their upstream intron as reported by other groups (Li YI et al., 2015; Torres-Méndez et al., 2019) and confirmed by this study (Figure 3.16). In agreement with previous reports (Torres-Méndez et al., 2019), alternative microexons (3-27 nt) in the presented dataset are preceded by stronger BrP and PPTs and have weaker 3'SS and stronger 5'SS than longer (>51 nt) alternative exons. In addition, the feature analysis highlighted previously unreported differences in the downstream sequences, specifically a moderately stronger BrP in the downstream intron and slightly weaker 3'SS of the following exon.

On the other hand, microexons (28-51 nt) showed less marked differences compared to longer exons with slightly higher scores in their SS, in their upstream BrP and PPT and in the 5'SS of the preceding exons. Such distinct features as well as the different propensities to be included or excluded during neural development support the choice to treat microexons (28-51 nt) as a separate class, while they have been previously grouped either with longer exons (Irimia et al., 2014) or with shorter microexons (Li YI et al., 2015) .

Noteworthy, even when microexons are not taken into account, included alternative exons have an average length significantly shorter than excluded ones (Figure 3.14) and are characterized by stronger BrPs and PPTs in their upstream introns, which facilitate recognition and inclusion, in both PP-DP and DP-N transitions (Figure 3.16).

Overall, the data suggest that neural-enriched transcripts preferentially include short exons while losing longer ones and the gained exons have features that favor inclusion once a neuronal splicing program is activated.

4.5 Neural exons are prevalently regulated by nSR100 during neurogenic commitment

In agreement with previous studies, nSR100 seems to act as a main regulator of a neural splicing program (Calarco et al., 2009; Raj B et al., 2011; Irimia et al., 2014; Quesnel-Vallières et al., 2015). Neural exons gaining inclusion during neurogenic commitment and/or neurogenesis were enriched for nSR100 binding site in the upstream intron compared to excluded and not spliced alternative exons (Figure 3.18).

Among included exons, more pronounced enrichments were found for microexons (3-27 nt) and, to a lesser extent, for microexons (28-51 nt) rather than for longer exons (Supplementary Figure 2), confirming that the majority of microexons depends on nSR100 for their inclusion (Irimia et al., 2014; Gonatopoulos-Pournatzis et al., 2018; Torres-Méndez et al., 2019). While nSR100 has a prominent role in mediating microexon inclusion, having even evolved the eMIC domain specialized for this task (Torres-Méndez et al., 2019), its function is not restricted to microexon regulation (Quesnel-Vallières et al., 2015). Consistently, when microexons were removed from the dataset, a small but significant enrichment for the nSR100 binding site was still present in the intron upstream of longer included exons compared to excluded and not spliced ones in the PP-DP transition (Supplementary Figure 3, top). Instead, when analyzing long exons alternatively spliced during neurogenesis, such enrichment was considerably reduced with included and excluded exons showing a similar profile (Supplementary Figure 3). In summary, the data show that nSR100 mediates the inclusion of a group of neural exons, mostly but not only microexons, already during neurogenic commitment. nSR100 acts during neurogenesis by sustaining or triggering the inclusion of almost exclusively neural microexons, while other splicing factors seem to be more relevant for splicing regulation of longer neural exons during neurogenesis. For example, Nova2 showed its typical enrichment pattern upstream

of excluded and downstream of included exons only when comparing cassette exons with or without microexons spliced in the DP-N transition (Supplementary Figure 4 and Supplementary Figure 5)

nSR100 is upregulated early during corticogenesis (Figure 3.1) and mediates the inclusion of cassette exons that mark the generation of DP with a predominant representation of neural-enriched microexons. Therefore, this splicing factor and the splicing events it controls are primary regulators of a neurogenic splicing profile that defines a neuronal identity. Instead, other splicing factors would mainly regulate events important for neurogenesis and neuronal maturation. Consistently, transient silencing of nSR100 in the developing mouse brain affected neurogenesis, while increasing the fraction of cells residing in the VZ/SVZ (Raj B et al., 2011). More precisely, nSR100 seems to be critical for the generation of DP and for the correct development of neurons instead of neurogenesis *per se*. Indeed, mice deficient for nSR100 show an incorrect cortical layering in the forebrain and a thinning of the SVZ, among several defects in the central and peripheral nervous system. Neurogenesis is not impaired, although it occurs prematurely and the generated neurons have defective neurites (Quesnel-Vallières et al., 2015).

While these studies highlighted the role of nSR100 and microexons in neocortical development, they could not distinguish between events controlling neurogenic commitment and events regulating neurogenesis. The present study overcomes these limitations allowing a better understanding of the temporal regulation of nSR100 targets. For example, two splicing events under nSR100 control have been proposed to play a major role in mouse neocortical development: the splicing of Ptbp2 exon 10 and the inclusion of a neural-specific exon between exons 3 and 4 of the Rest transcript (Calarco et al., 2009; Raj B et al., 2011).

In the absence of exon 10, the transcript of the neural splicing factor Ptbp2 undergoes NMD. This alternative splicing event is promoted by the splicing factor Ptbp1, while nSR100 sustains exon 10 inclusion. In the present dataset, Ptbp2 is neither differentially expressed nor differentially spliced and exon 10 is always included with a PSI of at least 95 in all three cell populations (Supplementary Figure 7, right). Therefore, the low levels of nSR100 expressed in PP are sufficient to inhibit Ptbp1 repressive effects on Ptbp2 exon 10. These data are in agreement with previous reports of a dominant effect of nSR100 over Ptbp1 on a subset of co-regulated exons (Raj B et al., 2014). In addition, the present data allow a more defined picture of the relationships among these splicing factors: Ptbp2 was always

highly expressed in all the cells of the neurogenic lineage, coexisting in neural progenitors with high levels of Ptbp1 and low but increasing levels of nSR100, which are sufficient to counteract Ptbp1 repression on specific neural exons already in PP (Figure 3.1). In neurons, the upregulation of nSR100 concomitantly with the silencing of Ptbp1 by miR-124 removes the brakes on neural exons, allowing the definitive acquisition of a neuronal identity. At the same time, the repressive action of Ptbp2 on “adult” neural exons is maintained, ensuring that such exons get included at the right stage of neuronal maturation. I observed a small enrichment for Ptbp-binding motifs in included exons over non-spliced ones (Supplementary Figure 6). This probably reflects exon inclusion mediated by nSR100 and co-regulated by Ptbp1 and Ptbp2. The gain in inclusion might also derive from the substitution of Ptbp1 with Ptbp2 repression, which is generally a weaker repressor of exon inclusion (Keppetipola et al., 2012).

The acquisition of a neural-specific 16 nt microexon in the mouse transcript of *Rest* leads to the production of a truncated protein devoid of the ability to bind to the repressor element 1 motifs in the regulatory region of neuronal genes. Our sequencing data show that this exon is absent in PP and DP but strongly included in N along with the retention of an intron, also reported as a neural-specific event (Supplementary Figure 7, left). This shows that *Rest* is still active as a neural repressor in DP while its splicing in neurons would lead to the expression of neuronal genes.

These two examples highlight the importance of a cell type-specific characterization of the splicing profiles of cells at distinct stages of differentiation. Indeed, although both events were already identified as being under nSR100 control, Ptbp2 exon 10 inclusion is probably the default choice for cells of the neurogenic lineage and nSR100 sustains such inclusion counteracting the action of Ptbp1 in neural progenitors. Similarly, as nSR100 increases its expression while development proceeds, it triggers the inclusion of the *Rest* microexon leading to the generation of the truncated *Rest* isoform and allowing the expression of neural genes in post-mitotic neurons. It is thus clear that nSR100 mediates a broad spectrum of events important for different stages of neocortical development but its role in regulating the neurogenic commitment ensuring a timely generation of DP and preventing a premature neurogenesis, is most likely due to the inclusion of a specific subset of cassette exons, most of which are neural-enriched microexons.

As for the aforementioned Ptbp2 and *Rest* splicing, the cell-specificity offered by the *Btg2^{RFP}/Tubb3^{GFP}* mouse line was critical for the timely distinction of the splicing choices

marking the progression of cells of the neurogenic lineage towards differentiation. Above everything else, the distinction of PP, DP and N with this system allows the identification of splicing events specifically linked to the neurogenic commitment.

Indeed, previous studies that addressed the question of how splicing could regulate cell-fate commitment in the nervous system could not isolate committed neural progenitors from neurons (Zhang et al., 2016) or from proliferating progenitors (Liu J et al., 2018), leading to inexact conclusions. In particular, the Walsh group proposed an elegant mechanism by which the alternative inclusion of the Nin exon 18 would change the plane of division of the proliferating progenitors thus leading to neurogenic commitment (Zhang et al., 2016). However, our data show that such splicing event takes place during neurogenesis with Nin exon 18 being equally represented in PP and DP transcripts and lost only in N. Therefore, has no role in the switch from a proliferative to a differentiative division.

4.6 *In vivo* overexpression of splice variants highlights isoform-specific functions in neurogenic commitment

To gain insight into isoform-specific functions in the developing mouse neocortex and to test whether the splicing choices that mark the transition from proliferation to differentiation have a causal relationship with this process, *in vivo* assays were performed. The strategy employed consisted in the cloning and transient overexpression of specific isoforms in the proliferating NSC pool lining the mouse embryonic brain lateral ventricles, by *in utero* electroporation. This technique allows a rapid assessment of the role of candidate genes in the developing mouse brain. However, when considering the investigation of alternative splicing, this strategy is less straightforward. Indeed, it is necessary to first determine which isoforms change their relative abundance as a consequence of differential splicing, in order to clone and overexpress them in the system under study. For this purpose, the only available option is to apply isoforms-reconstructing bioinformatic pipelines for short-read sequencing data. Unfortunately, as discussed above, such pipelines are highly error prone and at the same time conservative, i.e. show a high rate of both false positives and false negatives. Long-read sequencing would potentially overcome the limitation of such approaches in the future providing a direct estimation of the isoform abundance in a sample. Yet, as of today, this technology cannot be quantitative but only qualitative.

Exon-based tools are more accurate and sensitive in detecting differential splicing. However, while it is possible to assign one exon to a single isoform, these cases are a

minority and an alternative exon is usually shared among several isoforms, posing a constraint on the selection of specific isoforms to clone and manipulate *in vivo*.

In this study, I first used isoform-reconstruction methods to identify splice variants with major changes in their relative proportion during the switch from proliferation to differentiation. Although such a change in proportion could not be validated, transient overexpression of the identified spliced variants altered the distribution of manipulated cells in the developing mouse neocortex in an isoform-specific manner. This effect was particularly evident for *Celf5* and *Tmcc2* isoform manipulation, where the overexpression of one splice variant caused a retention of proliferating progenitors in the VZ and a consequent reduction of N in the CP, whereas the overexpression of another splice variant with an opposite change in proportion did not cause any alteration of the physiological neocortical development (Supplementary Figure 1).

A similar scenario was observed in the case of *Faim* isoform manipulation, where the overexpression of the ubiquitous *Faim-S* promoted a marked retention of proliferating cells in the VZ and a decrease of N in the CP, while the overexpression of the neural-enriched isoform *Faim-L* had little effect on the distribution of electroporated cells (Figure 3.20). This corroborates the hypothesis of isoform-specific functions in directing neural fate commitment. *Faim* has been identified as an antagonist of apoptosis mediated by the death receptor Fas in B-lymphocytes, expressing the *Faim-S* variant (Schneider et al., 1999). Apoptosis is a programmed cell death fundamental for development and maintenance of tissue homeostasis throughout the lifespan of an organism. Death receptors such as Fas as TNF α sense extracellular cues that trigger the apoptotic response (Fuchs and Steller, 2011). In the nervous system, death receptors have been shown to be involved also in non-apoptotic processes like neurite branching and axonal growth (Lambert et al., 2003; Planells-Ferrer et al., 2016). Consistently, *in vitro* studies from the Comella lab have shown that *Faim* transcripts have diverse functions in neural cells. While *Faim-S* has a clear anti-apoptotic effect in the immune system (Huo et al., 2019) and it is necessary for neuritogenesis induced by nerve-growth factor (NGF) (Sole et al., 2004; Segura et al., 2007), it failed to protect neural cells from axonal degeneration caused by NGF withdrawal or from apoptosis induced by stimulation of Fas or TNF α (Segura et al., 2007). On the contrary, *Faim-L* had a clear protective effect on axonal degeneration and apoptosis caused by NGF withdrawal or death receptor stimulation (Segura et al., 2007). These survival effects were mediated by a specific interaction of *Faim-L* with the x-linked inhibitor of apoptosis protein (XIAP). The inclusion of exon 2 to *Faim-L* but not *Faim-S* introduces 22

aa to the protein, with a conserved alanine that allows the interaction of Faim-L with XIAP. This prevents ubiquitylation and degradation of XIAP, thus exerting its antiapoptotic function by inhibiting caspases (Moubarak et al., 2013). In addition, data revealed that the interaction of Faim-L with XIAP also prevented the internalization of AMPA receptor subunit GluA2 and synaptic long-term-depression (LTD) *in vitro* (Martínez-Mármol et al., 2016).

According to the Ensembl mm9 annotation employed for this study, the *Faim* gene has only two splice variants that differ for only one exon. Such extremely simplified structure made it particularly easy to detect by both isoform- and exon-based analysis, as well as to clone and manipulate by *in utero* electroporation. On the other hand, other isoforms have been detected for this gene in the last years and Ensembl mm10 reports five isoforms for mouse and nine for human in its latest release, version 99. While the alternative exon reported in the present study is specific for Faim-L according to the Ensembl annotations, there is evidence that at least in human it is expressed also in another variant, Faim-L_2a (Coccia et al., 2017). Therefore, while the present data show a differential splicing for exon 2 and confirm its enrichment in neurons, it is not possible to exclude that Faim splice variants other than the ones manipulated are involved in such splicing switch.

The uncertainties in isoform annotations and the discrepancies among the different annotation databases are undoubtedly a limitation for splicing analysis at the isoform level and for subsequent isoform-based manipulation. Of note, current annotations may be inaccurate also for single exons. One remarkable case is the annotation of the Rest neural-specific exon in mouse. While this microexon splicing has been observed since 1998 (Palm et al., 1998) and is supported by further studies (Raj B et al., 2011), it has not been annotated in Ensembl nor in VastDB. As a result, in the present dataset such event has been mislabeled as a 32 nt alternative last exon according to Ensembl mm9 transcript annotation. The sequencing data clearly support the presence of a microexon in this position, with a read peak of 16 nt in length and not corresponding to the end of the transcript, as strongly shown by the presence of junction reads connecting this peak with both the previous and the following exons (Supplementary Figure 7, left). These observations highlight the importance of a proper annotation at exon, splice site and transcript levels, in order to correctly detect and categorize splicing events. In addition, since microexons are usually even more problematic to detect than other exons, our current estimation of the number of microexons differentially spliced during neurogenic commitment and neurogenesis is probably underrated.

Cell-type-specific long read sequencing could be a valuable resource to better locate the exon splice sites and reconstruct exons connectivity, ultimately providing a transcript model of the isoforms actually expressed in each tissue. With the aid of short-reads sequencing data long-read sequencing could provide a quantitative estimation of isoform proportional change in different cell populations. Such strategy would be helpful for the identification of cell-type-specific splice variants and for the characterization of their function *in vivo* by *in utero* electroporation with the approach adopted in this study.

It has to be kept in mind though that such experimental design relies on the assumption that changes in splicing choices are significant when they cause an increase in a specific splice variant representation while proportionally decreasing the representation of another splice variant. Alternatively, the change in inclusion of certain exons and the presence of multiple isoforms at a specific ratio, rather than the increase of one key isoform, might be relevant for driving cell-fate commitment. In this scenario where the ratio of different isoforms and/or the global representation of certain exons encoded by multiple transcripts are the critical factors driving cell-fate commitment, a simple overexpression of one isoform would probably not be the best approach to investigate the role of alternative splicing choices *in vivo*, while a technique directed to manipulate single splicing events would be more appropriate. Artificial splicing factors such as RNA-binding Pumilio domains engineered with splicing regulatory domains (Wang Y and Wang Z, 2016) and the RNA-directed CasRx systems (Koneremann et al., 2018), as well as short antisense oligonucleotides targeting splice sites (Havens and Hastings, 2016), have been proven to be effective in transiently modulating splicing choices *in vitro* and in cell cultures. Provided their effectiveness *in vivo*, they could represent a valuable tool for understanding the functional role of specific splicing events under physiological conditions.

4.7 Concluding remarks and future perspectives

This study characterizes for the first time the splicing profiles of cells of the neurogenic lineage at subsequent stages of differentiation, allowing the identification of specific splicing events linked to cell-fate determination towards a neural identity, i.e. the neurogenic commitment. The data clearly point to the inclusion of a subset of cassette exons, mostly microexons, under the control of the neural splicing factor nSR100 as the main trait of a proliferative to differentiative switch. Since those inclusions prevalently code for unstructured protein domains of genes involved in nervous system development, they

would likely lead to a change in binding properties of such proteins and thus to a remodeling of protein-protein interaction networks.

Although nSR100 manipulations performed by others (Quesnel-Vallières et al., 2015) indicate a causal relationship between the splicing choices mediated by this splicing factors and neocortical development, it is still an open question how individual nSR100-dependent events could lead to a neurogenic commitment. *In vivo* manipulation of different splice variants, amongst them the nSR100-regulated Faim transcripts, has demonstrated isoform-specific effects on neocortical development, although their role under physiological condition is still uncertain. To overcome the limitations of isoform-centric tools in reconstructing transcripts expression from short-reads, we are currently strengthening our data with a Btg2^{RFP}/Tubb3^{GFP} cell-type-specific long read transcriptome (part of the PhD project of Beatriz Cardoso de Toledo). These data will be valuable not only to improve our analysis but also to identify splice variants to manipulate *in vivo*, thus determining isoform-specific function under physiological conditions. In parallel, we are investigating the feasibility to employ other methods to modulate splicing choices at the exon level in the developing neocortex, such as the RNA-directed CasRx systems (Konermann et al., 2018) and the splicing regulatory short antisense oligonucleotides (Havens and Hastings, 2016). These resources would help to determine a causality between alternative splicing and cell-fate commitment, as well as to characterize isoform-specific functions in the nervous system.

In the future, it would be interesting to gain a global view of the different proteome landscapes and protein-protein network remodeling triggered by tissue-specific splicing programs and to determine how other gene regulatory mechanisms such as epigenetic modifications could impact alternative exon inclusion. Lastly, it would be of great advantage for cell-identity studies to investigate alternative splicing at the single-cell level, which remains an important computational challenge as of today.

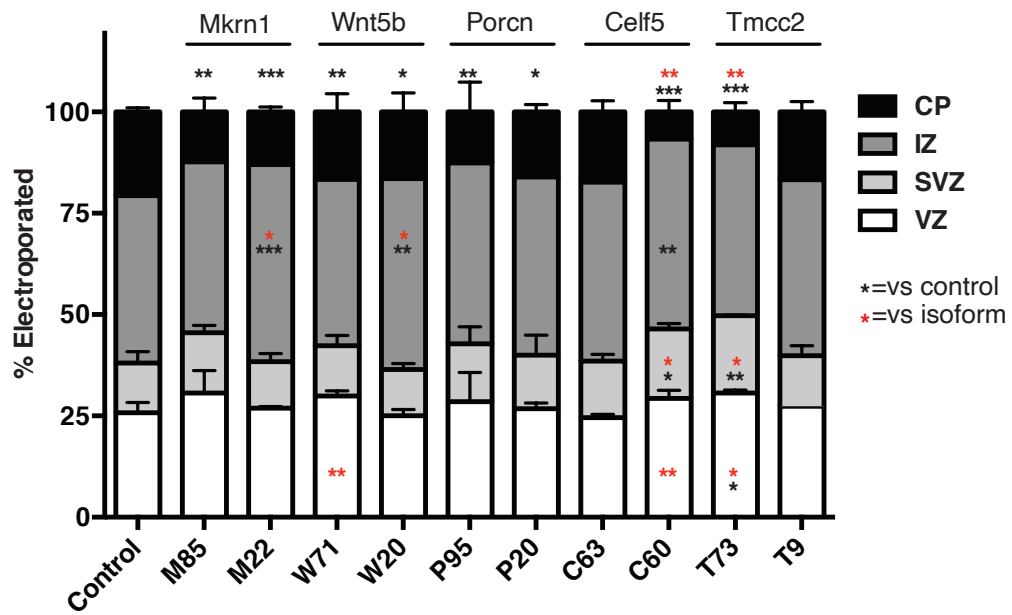
Dissecting tissue- and cell-specific splicing profiles, their interplay with other gene regulatory mechanisms and how such transcript variability could shape tissue-specific proteomes would be of great interest not only to uncover a role of alternative splicing and transcriptional regulatory processes in cell-fate determination but also to shed light on how aberrant splicing could be involved in disease etiology such as neurological disorders.

5 Supplementary figures

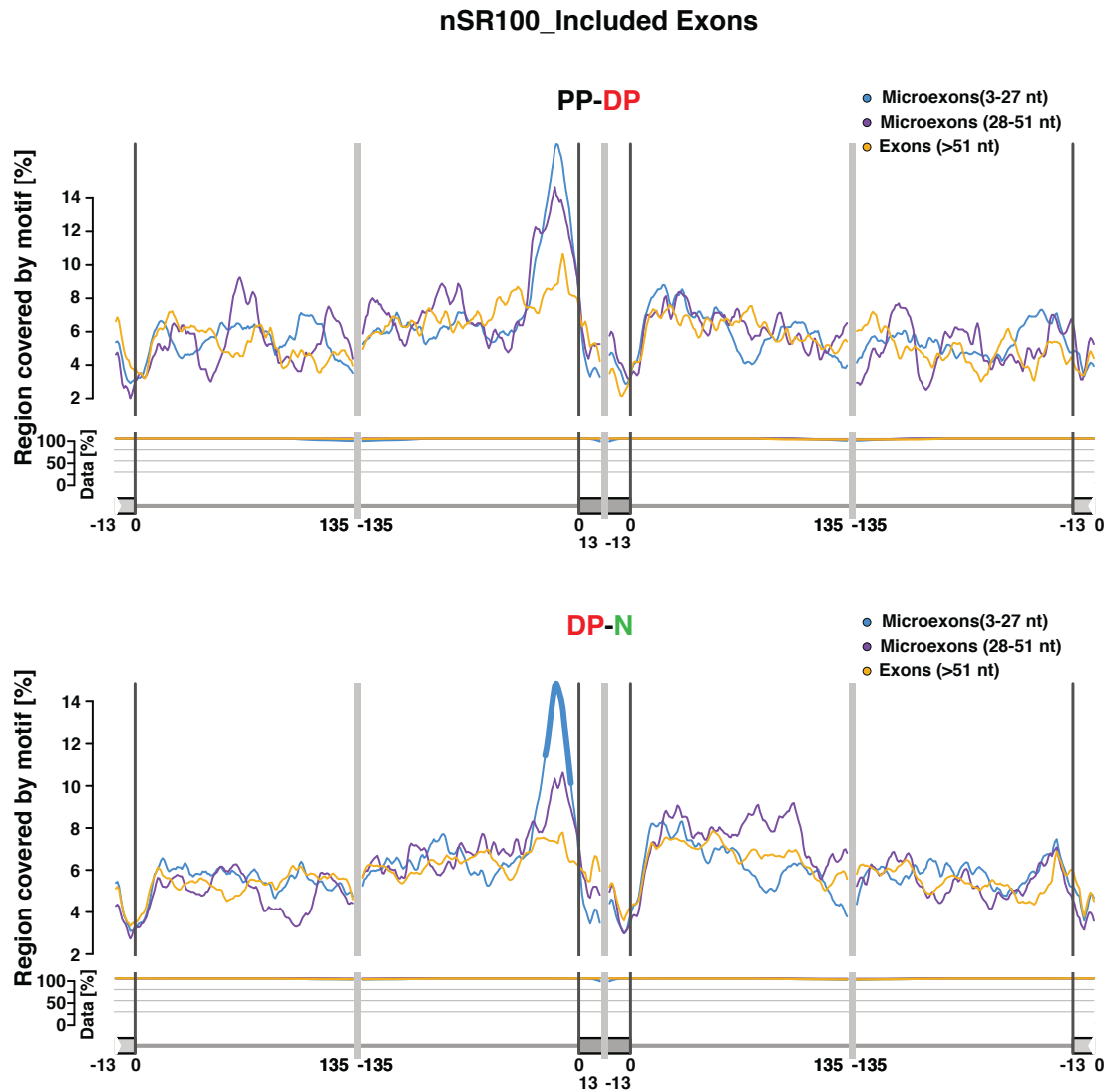
Genes with isoform switching resulting in changes in coding sequences according to Cuffdiff2 selected for *in vivo* manipulation.

Gene name	Function	CDS	$\Delta\%$	Protein differences
Wnt5b	Cell fate decision and patterning during neurogenesis	W71	+ 41.47%	Signal peptide present
		W20	- 41.29%	Signal peptide absent
Porcn	Palmitoylation of Wnt proteins, necessary for their secretion	P95	+ 53.34%	All functional domains present
		P20	- 54.10%	Lack of 5 aa near the palmitoylation domain
Celf5	Regulation of splicing, mRNA editing and translation	C63	+ 26.35%	2 RNA-binding domains present
		C60	- 24.37%	Lack of 1 RNA binding domain
Mkx1	Zinc-finger protein. Regulation of transcription, ubiquitination, cell cycle arrest and apoptosis	M85	+19.56%	All 5 constitutive zinc-finger domains present
		M22	-19.10%	Lack of 3 zinc-finger domains
Tmcc2	Interacts with apolipoprotein E and amyloid precursor protein	T73	+ 41.16%	All constitutive domains present
		T9	- 39.41%	Lack of a disorganized region

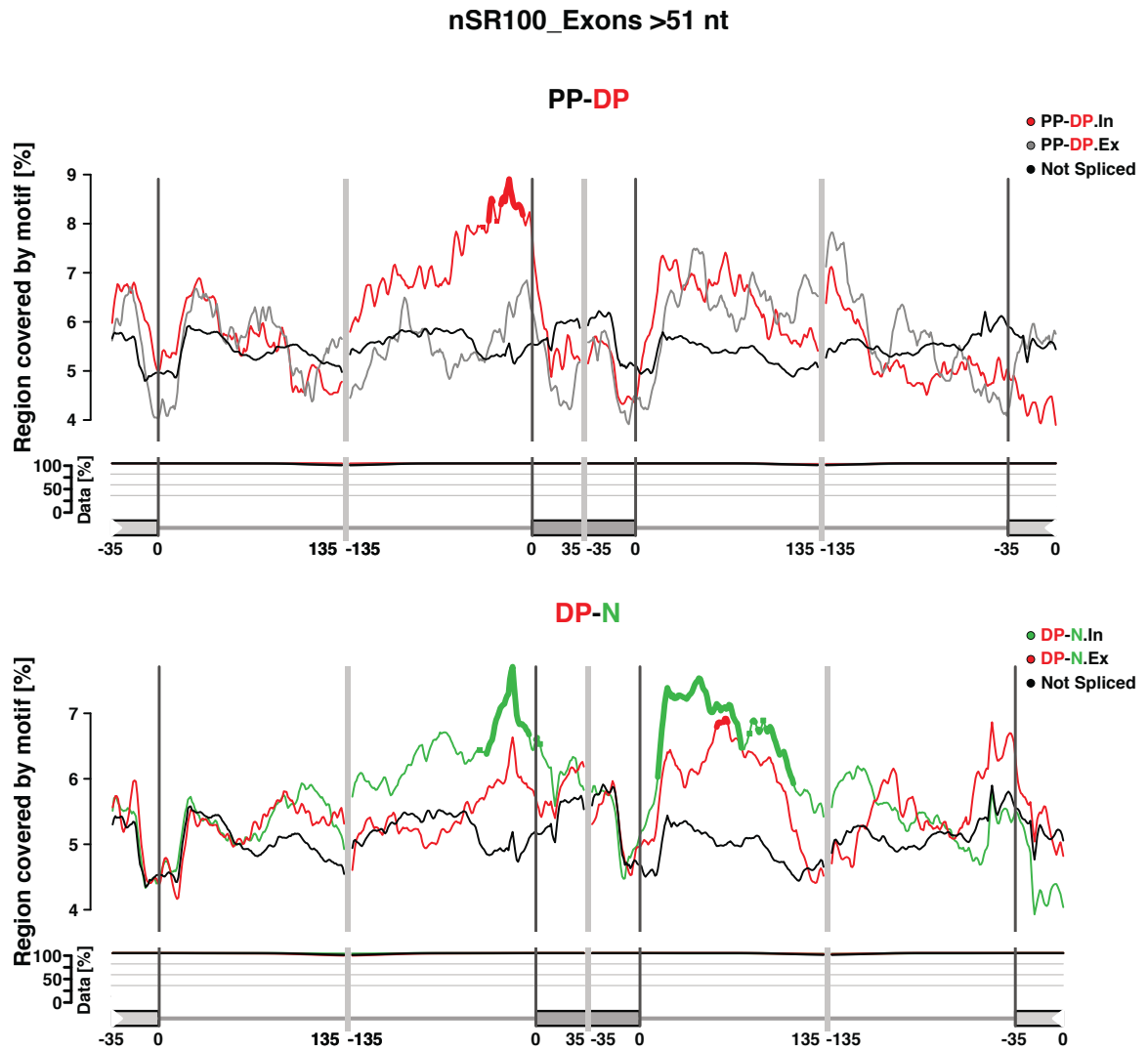
Supplementary Table 1 Spliced genes identified by isoform switching analysis and selected for validation and *in vivo* assay. The name of the gene is indicated with the official gene symbol while the name of the isoforms is shown with the first letter of the gene and the last two digits of the Ensemble identifier. Predicted changes in isoform abundance relative to the whole gene expression in PP-DP are reported ($\Delta\%$). The differences in protein isoforms are indicated in the last column.



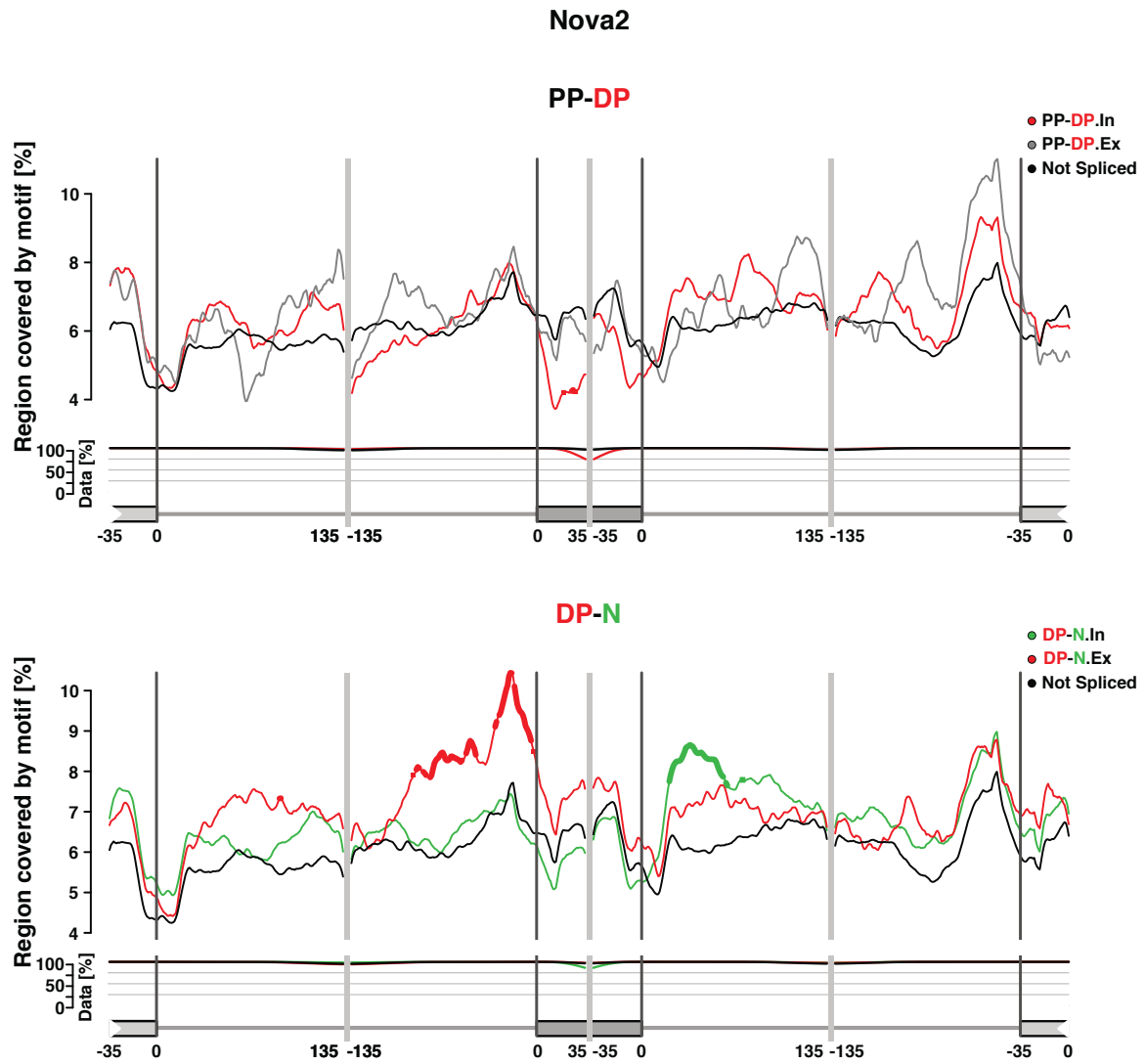
Supplementary Figure 1 *In vivo* assay of genes predicted to undergo isoform switching during neurogenic commitment. Distribution of electroporated cells in the different regions of the neocortex. Percentages are calculated on the total of electroporated cells. Mean and standard deviation (SD) are reported (n≥3). Significant differences in RFP distribution respective to the control (black *) and to the relative isoform (red *) are indicated. *p-value<0.05, **p-value<0.005, ***p-value<0.0001, paired t-test.



Supplementary Figure 2 nSR100 binding sites in different exon length classes. Enrichment for nSR100 binding motif nearby included microexons (3-27 nt) and microexons (28-51 nt) compared to included longer exons (>51 nt) in neurogenic commitment (top) or neurogenesis (bottom). Length of sliding window: 17 nt. The “Data” subpanel refers to the proportion of total sequences used for the RNA map at each position, as implemented by Matt. Thicker regions are position showing significant enrichment. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; nt=nucleotides.

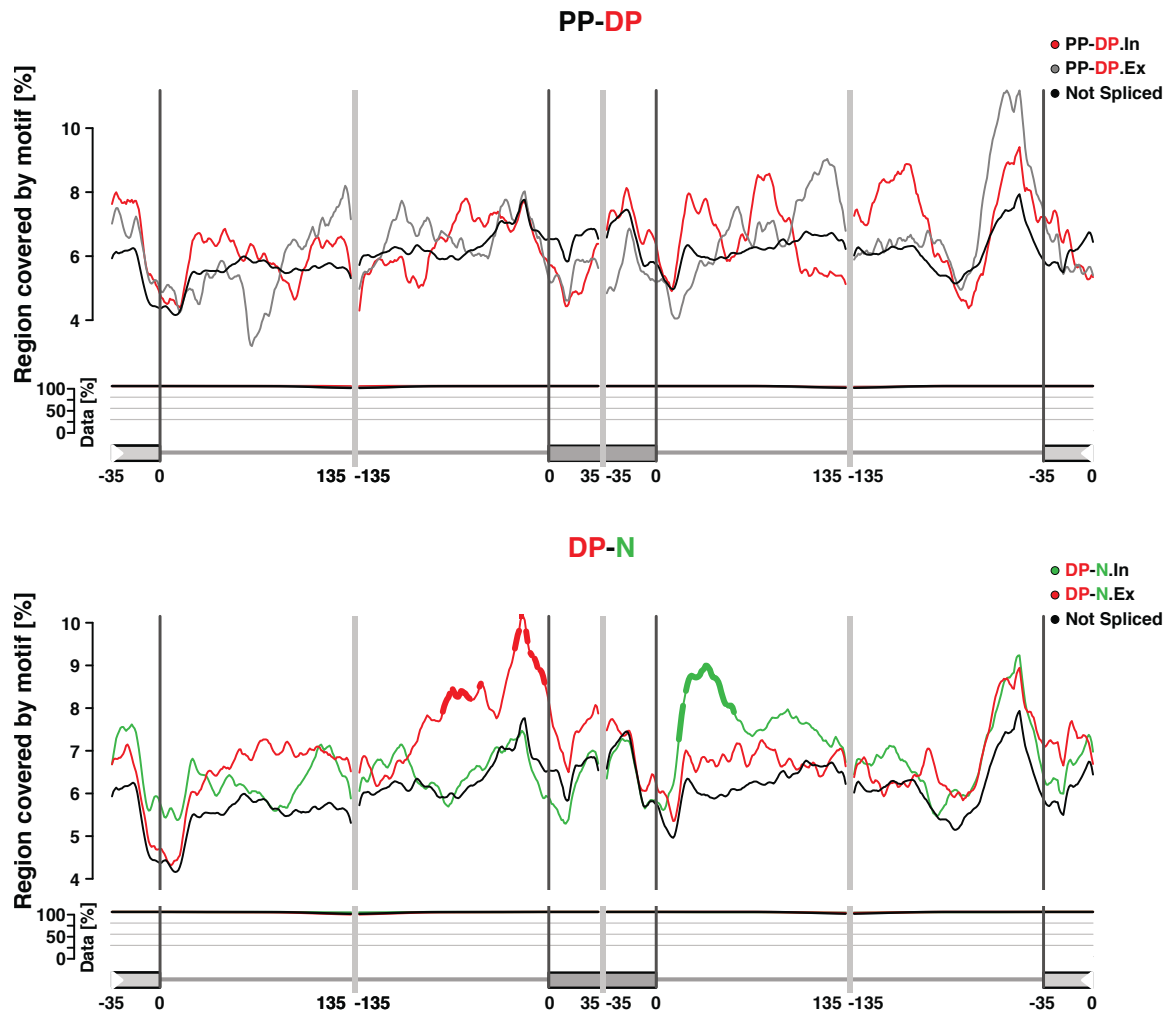


Supplementary Figure 3 nSR100 binding sites in exons (>51 nt). Enrichment for nSR100 binding motif nearby spliced longer exons (>51 nt) compared to not spliced longer exons (>51 nt) in neurogenic commitment (top) or neurogenesis (bottom). Length of sliding window: 31 nt. The “Data” subpanel refers to the proportion of total sequences used for the RNA map at each position, as implemented by Matt. Thicker regions are position showing significant enrichment. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; In=included; Ex=excluded.

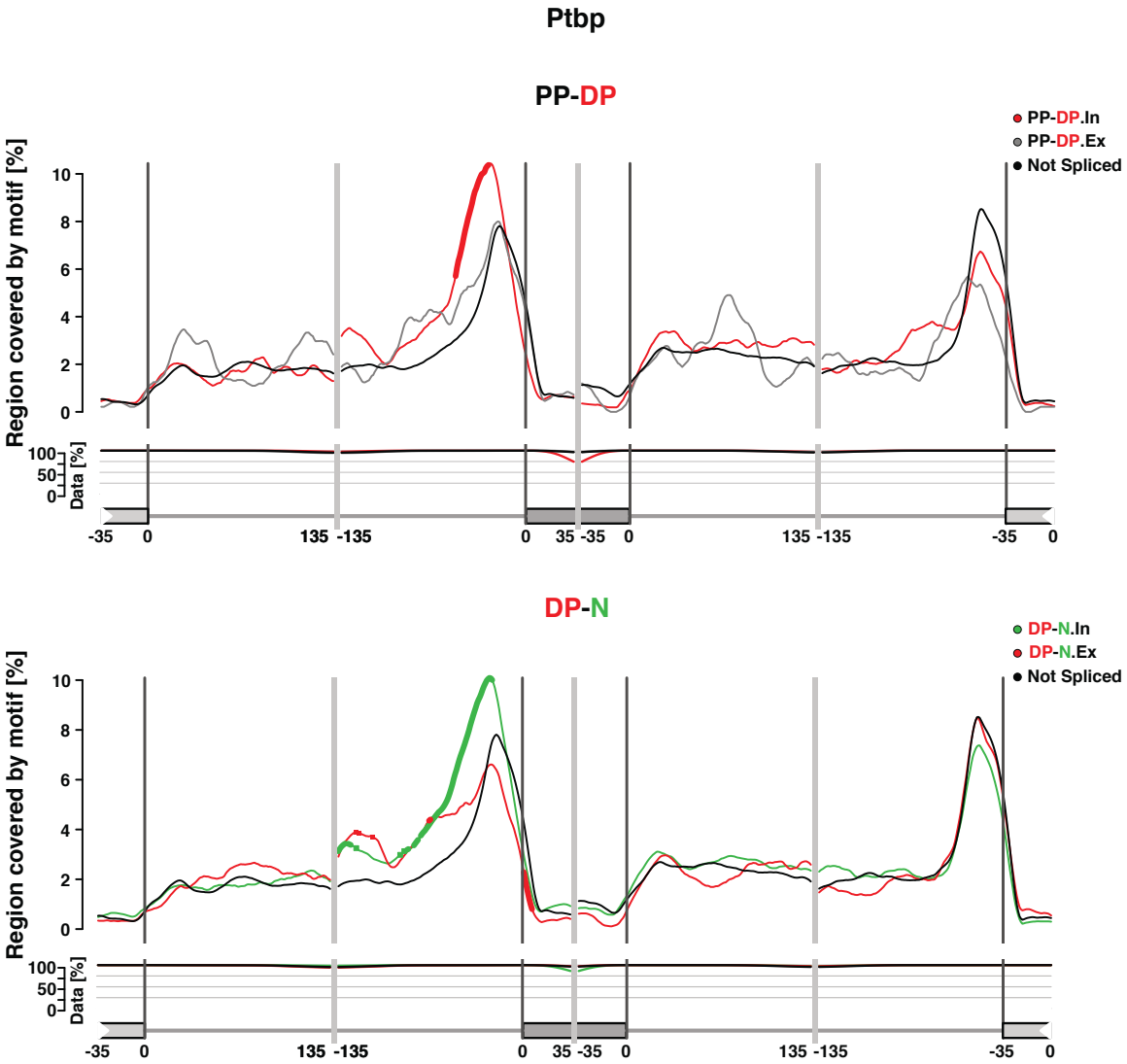


Supplementary Figure 4 Nova2 binding sites. Enrichment for Nova2 binding motif nearby spliced exons compared to not spliced exons in neurogenic commitment (top) or neurogenesis (bottom). Length of sliding window: 31 nt. The “Data” subpanel refers to the proportion of total sequences used for the RNA map at each position, as implemented by Matt. Thicker regions are position showing significant enrichment. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; In=included; Ex=excluded.

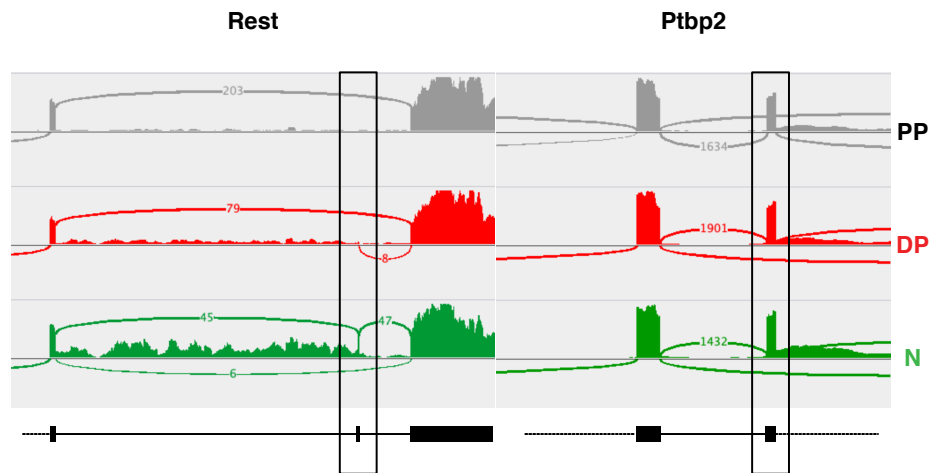
Nova2_Exons >51 nt



Supplementary Figure 5 Nova2 binding sites in exons (>51 nt). Enrichment for Nova2 binding motif nearby spliced longer exons (>51 nt) compared to not spliced longer exons (>51 nt) in neurogenic commitment (top) or neurogenesis (bottom). Length of sliding window: 31 nt. The “Data” subpanel refers to the proportion of total sequences used for the RNA map at each position, as implemented by Matt. Thicker regions are position showing significant enrichment. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; In=included; Ex=excluded.



Supplementary Figure 6 Ptbp1/2 binding sites. Enrichment for Ptbp1/2 binding motif nearby spliced exons compared to not spliced exons in neurogenic commitment (top) or neurogenesis (bottom). Length of sliding window: 31 nt. The “Data” subpanel refers to the proportion of total sequences used for the RNA map at each position, as implemented by Matt. Thicker regions are position showing significant enrichment. PP=proliferating progenitors; DP=differentiating progenitors; N=neurons; In=included; Ex=excluded.



Supplementary Figure 7 Splicing of Rest and Ptbp2. Sequencing data of the neural exon of Rest (left) and of exon 10 of Ptbp2 (right) in proliferating progenitors (PP), differentiating progenitors (DP) and neurons (N): the peaks correspond to reads mapping to the exons, lines connecting the peaks are reads spanning the exon-exon junction.

6 References

- Adhikari S, Xiao W, Zhao YL, Yang YG. 2016. m6A: Signaling for mRNA splicing. *RNA Biol*, 13(9):756–759
- Aebi M, Weissman C. 1987. Precision and orderliness in splicing. *Trends Genet*, 3:102–107
- Alamancos GP, Agirre E, Eyraas E. 2014. Methods to study splicing from high-throughput rna sequencing data. In: Hertel KJ (ed) *Spliceosomal Pre-mRNA Splicing. Methods in Molecular Biology (Methods and Protocols)* vol 1126 pp. 357–397. Humana Press, Totowa, NJ
- Alberti S, Gladfelter A, Mittag T. 2019. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell*, 176(3):419–434
- Alló M, Schor IE, Muñoz MJ, de la Mata M, Agirre E, Valcárcel J, Eyraas E, Kornblihtt AR. 2010. Chromatin and alternative splicing. *Cold Spring Harb Symp Quant Biol*, 75:103–111
- Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol*, 11(10):R106
- Anders S, Reyes A, Huber W. 2012. Detecting differential usage of exons from RNA-seq data. *Genome Res*, 22(10):2008–2017
- Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J. 2009. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res*, 19(10):1732–1741
- Angevine JB, Bodian D, Coulombre AJ, Edds MV, Hamburger V, Jacobson M, Lyser KM, Prestige MC, Sidman RL, Varon S, Weiss PA. 1970. Embryonic vertebrate central nervous system: Revised terminology. *Anat Rec*, 166(2):257–261
- Apra J, Lesche M, Massalini S, Prenninger S, Alexopoulou D, Dahl A, Hiller M, Calegari F. 2015. Identification and expression patterns of novel long non-coding RNAs in neural progenitors of the developing mammalian cortex. *Neurogenesis (Austin)*, 2(1):e995524.
- Apra J, Prenninger S, Dori M, Ghosh T, Monasor LS, Wessendorf E, Zocher S, Massalini S, Alexopoulou D, Lesche M, Dahl A, Groszer M, Hiller M, Calegari F. 2013. Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. *EMBO J*, 32(24):3145–3160

- Arai Y, Pulvers JN, Haffner C, Schilling B, Nüsslein I, Calegari F, Huttner WB. 2011. Neural stem and progenitor cells shorten S-phase on commitment to neuron production. *Nat Commun*, 2:154
- Artegiani B, de Jesus Domingues AM, Bragado Alonso S, Brandl E, Massalini S, Dahl A, Calegari F. 2015. Tox: a multifunctional transcription factor and novel regulator of mammalian corticogenesis. *EMBO J*, 34(7):896–910
- Artegiani B, Lange C, Calegari F. 2012. Expansion of embryonic and adult neural stem cells by in utero electroporation or viral stereotaxic injection. *J Vis Exp*, (68):4093
- Attardo A, Calegari F, Haubensak W, Wilsch-Bräuninger M, Huttner WB. 2008. Live imaging at the onset of cortical neurogenesis reveals differential appearance of the neuronal phenotype in apical versus basal progenitor progeny. *PLoS One*, 3(6):e2388
- Balmer S, Nowotschin S, Hadjantonakis AK. 2016. Notochord morphogenesis in mice: Current understanding & open questions. *Dev Dyn*, 245(5):547–557
- Baralle FE, Giudice J. 2017. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol*, 18(7):437–451
- Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ. 2010. Deciphering the splicing code. *Nature*, 465(7294):53–59
- Barbosa-Morais NL, Carmo-Fonseca M, Aparício S. 2006. Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion. *Genome Res*, 16(1):66–77
- Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, Kim T, Misquitta-Ali CM, Wilson MD, Kim PM, Odom DT, Frey BJ, Blencowe BJ. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science*, 338(6114):1587–1593
- Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA*, 74(8):3171–3175
- Beyer AL, Osheim YN. 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev*, 2(6):754–765
- Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, 72(1):291–336
- Blencowe BJ. 2000. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci*, 25(3):106-110
- Boise HL, Gonzalez-Garcia M, Postema CE, Ding L, Lindsten T, Turka LA, Mao X, Nunez G, Thompson CB. 1993. bcl-x, a bcl-2-Related Gene That Functions as a Dominant Regulator of Apoptotic Cell Death. *Cell*, 74:597-608 .
- Boucard AA, Chubykin AA, Comoletti D, Taylor P, Südhof TC. 2005. A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to α - and β -neurexins. *Neuron*, 48(2):229–236

- Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis T, Frey B, Irimia M, Blencowe BJ. 2014. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res*, 24(11):1774–1786
- Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*, 34(5):525–527
- Buljan M, Chalancon G, Eustermann S, Wagner GP, Fuxreiter M, Bateman A, Babu MM. 2012. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol Cell*, 46(6):871–883
- Bultje RS, Castaneda-Castellanos DR, Jan LY, Jan YN, Kriegstein AR, Shi SH. 2009. Mammalian Par3 regulates progenitor cell asymmetric division via notch signaling in the developing neocortex. *Neuron*, 63(2):189–202
- Busch A, Hertel KJ. 2012. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip Rev RNA*, 3(1):1–12
- Cáceres JF, Kornblihtt AR. 2002. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet*, 18(4):186–193
- Calarco JA, Superina S, O'Hanlon D, Gabut M, Raj B, Pan Q, Skalska U, Clarke L, Gelinis D, van der Kooy D, Zhen M, Ciruna B, Blencowe BJ. 2009. Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. *Cell*, 138(5):898–910
- Calarco JA, Zhen M, Blencowe BJ. 2011. Networking in a global world: establishing functional connections between neural splicing regulators and their target transcripts. *RNA*, 17(5):775–791
- Campbell K, Götz M. 2002. Radial glia: multi-purpose cells for vertebrate brain development. *Trends Neurosci*, 25(5):235–238
- Celik A, Kervestin S, Jacobson A. 2015. NMD: At the crossroads between translation termination and ribosome recycling. *Biochimie*, 114:2–9
- Chen G, Ning B, Shi T. 2019. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front Genet*, 10:317
- Chen M, Manley JL. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol*, 10(11):741–754
- Cheng TL, Chen J, Wan H, Tang B, Tian W, Liao L, Qiu Z. 2017. Regulation of mRNA splicing by MeCP2 via epigenetic modifications in the brain. *Sci Rep*, 7:42790
- Chenn A, Walsh CA. 2002. Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. *Science*, 297(5580):365–369
- Cho KO, Hunt CA, Kennedy MB. 1992. The rat brain postsynaptic density fraction contains a homolog of the *Drosophila* discs-large tumor suppressor protein. *Neuron*, 9(5):929–942

- Choi E, Kuehlt M, Wall R. 1980. RNA splicing generates a variant light chain from an aberrantly rearranged kappa gene. *Nature*, 286(5775):776-779
- Chow LT, Broker TR. 1978. The spliced structures of adenovirus 2 fiber message and the other late mRNAs. *Cell*, 15(2):497-510.
- Coccia E, Calleja-Yagüe I, Planells-Ferrer L, Sanuy B, Sanz B, López-Soriano J, Moubarak RS, Munell F, Barneda-Zahonero B, Comella JX, Pérez-García MJ. 2017. Identification and characterization of new isoforms of human fas apoptotic inhibitory molecule (FAIM). *PLoS One*, 12(10):e0185327
- Collins L, Penny D. 2006. Investigating the Intron Recognition Mechanism in Eukaryotes. *Mol Biol Evol*, 23(5):901-910
- Corioni M, Antih N, Tanackovic G, Zavolan M, Krämer A. 2011. Analysis of in situ pre-mRNA targets of human splicing factor SF1 reveals a function in alternative splicing. *Nucleic Acids Res*, 39(5):1868-1879
- Corvelo A, Hallegger M, Smith CW, Eyras E. 2010. Genome-wide association between branch point properties and alternative splicing. *PLoS Comput Biol*, 6(11):e1001016
- Cramer P, Pesce CG, Baralle FE, Kornblihtt AR. 1997. Functional association between promoter structure and transcript alternative splicing. *Proc Natl Acad Sci USA*, 94(21):11456-11460
- Dean C, Dresbach T. 2006. Neuroligins and neuroligins: linking cell adhesion, synapse formation and cognitive function. *Trends Neurosci*, 29(1):21-29
- Dominski Z, Kole R. 1991. Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol*, 11(12):6075-6083
- Dori M, Cavalli D, Lesche M, Massalini S, Alieh LHA, de Toledo BC, Khudayberdiev S, Schrott G, Dahl A, Calegari F. 2020. MicroRNA profiling of mouse cortical progenitors and neurons reveals miR-486-5p as a regulator of neurogenesis. *Development*, 147(9):dev190520
- Dori M, Haj Abdullah Alieh L, Cavalli D, Massalini S, Lesche M, Dahl A, Calegari F. 2019. Sequence and expression levels of circular RNAs in progenitor cell types during mouse corticogenesis. *Life Sci Alliance*, 2(2):e201900354
- Dujardin G, Lafaille C, de la Mata M, Marasco LE, Muñoz MJ, Le Jossic-Corcós C, Corcos L, Kornblihtt AR. 2014. How slow RNA polymerase II elongation favors alternative exon skipping. *Mol Cell*, 54(4):683-690
- Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*, 6(3):197-208
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Veceli J, Wegener J, Wu D, Yang

- A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science*, 323(5910):133–138.
- Ellis JD, Barrios-Rodiles M, Çolak R, Irimia M, Kim T, Calarco JA, Wang X, Pan Q, O’Hanlon D, Kim PM, Wrana JL, Blencowe BJ. 2012. Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol Cell*, 46(6):884–892
- Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, Milacic M, Roca CD, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Viteri G, Weiser J, Wu G, Stein L, Hermjakob H, D’Eustachio P. 2018. The Reactome Pathway Knowledgebase. *Nucleic Acids Res*, 46(D1):D649–D655
- Fan X, Zhang X, Wu X, Guo H, Hu Y, Tang F, Huang Y. 2015. Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. *Genome Biol*, 16(1):148
- Farkas LM, Haffner C, Giger T, Khaitovich P, Nowick K, Birchmeier C, Pääbo S, Huttner WB. 2008. Insulinoma-associated 1 has a panneurogenic role and promotes the generation and expansion of basal progenitors in the developing mouse neocortex. *Neuron*, 60(1):40–55
- Fiszbein A, Kornblihtt AR. 2017. Alternative splicing switches: Important players in cell differentiation. *Bioessays*, 39(6)
- Fong N, Kim H, Zhou Y, Ji X, Qiu J, Saldi T, Diener K, Jones K, Fu XD, Bentley DL. 2014. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev*, 28(23):2663–2676
- Fuchs Y, Steller H. 2011. Programmed cell death in animal development and disease. *Cell*, 147(4):742–758
- Gaiano N. 2008. Strange bedfellows: Reelin and Notch signaling interact to regulate cell migration in the developing neocortex. *Neuron*, 60(2):189–191
- Gaiano N, Fishell G. 2002. The Role of Notch in promoting glial and neural stem cell fates. *Annu Rev Neurosci*, 25:471–490
- Gaiano N, Nye JS, Fishell G. 2000. Radial glial identity is promoted by Notch1 signaling in the murine forebrain. *Neuron*, 26(2):395–404
- Garcia J, Gerber SH, Sugita S, Südhof TC, Rizo J. 2004. A conformational switch in the Piccolo C2A domain regulated by alternative splicing. *Nat Struct Mol Biol*, 11(1):45–53
- Gelfman S, Burstein D, Penn O, Savchenko A, Amit M, Schwartz S, Pupko T, Ast G. 2012. Changes in exon-intron structure during vertebrate evolution affect the splicing pattern of exons. *Genome Res*, 22(1):35–50
- Gilbert SF, Barresi MJF. 2016. *Developmental Biology*. Eleventh ed. Sunderland, Massachusetts : Sinauer Associates, Inc., Publishers.

- Gohr A, Irimia M. 2019. Matt: Unix tools for alternative splicing analysis. *Bioinformatics*, 35(1):130–132
- Gonatopoulos-Pournatzis T, Wu M, Braunschweig U, Roth J, Han H, Best AJ, Raj B, Aregger M, O’Hanlon D, Ellis JD, Calarco JA, Moffat J, Gingras AC, Blencowe BJ. 2018. Genome-wide CRISPR-Cas9 Interrogation of Splicing Networks Reveals a Mechanism for Recognition of Autism-Misregulated Neuronal Microexons. *Mol Cell*, 72(3):510-524.e12
- Götz M, Huttner WB. 2005. The cell biology of neurogenesis. *Nat Rev Mol Cell Biol*, 6(10):777–788
- Grabowski P. 2011. Alternative splicing takes shape during neuronal development. *Curr Opin Genet Dev*, 21(4):388–394
- Graveley BR. 2000. Sorting out the complexity of SR protein functions. *RNA*, 6(9):1197–1211
- Grosso AR, Gomes AQ, Barbosa-Morais NL, Caldeira S, Thorne NP, Grech G, von Lindern M, Carmo-Fonseca M. 2008. Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Res*, 36(15):4823–4832
- Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18):2847–2849
- Gupta I, Collier PG, Haase B, Mahfouz A, Joglekar A, Floyd T, Koopmans F, Barres B, Smit AB, Sloan SA, Luo W, Fedrigo O, Ross ME, Tilgner HU. 2018. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat Biotechnol*, 36(12):1197–1202
- Hand R, Bortone D, Mattar P, Nguyen L, Heng JI, Guerrier S, Boutt E, Peters E, Barnes AP, Parras C, Schuurmans C, Guillemot F, Polleux F. 2005. Phosphorylation of Neurogenin2 specifies the migration properties and the dendritic morphology of pyramidal neurons in the neocortex. *Neuron*, 48(1):45–62
- Hardwick SA, Joglekar A, Flicek P, Frankish A, Tilgner HU. 2019. Getting the Entire Message: Progress in Isoform Sequencing. *Front Genet*, 10:709
- Hartfuss E, Galli R, Heins N, Götz M. 2001. Characterization of CNS precursor subtypes and radial glia. *Dev Biol*, 229(1):15–30
- Hatakeyama J, Bessho Y, Katoh K, Ookawara S, Fujioka M, Guillemot F, Kageyama R. 2004. Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. *Development*, 131(22):5539–5550
- Haubensak W, Attardo A, Denk W, Huttner WB. 2004. Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis. *Proc Natl Acad Sci USA*, 101(9):3196–3201
- Hausmann IU, Bodi Z, Sanchez-Moran E, Mongan NP, Archer N, Fray RG, Soller M. 2016. m⁶A potentiates Sxl alternative pre-mRNA splicing for robust *Drosophila* sex determination. *Nature*, 540(7632):301–304

- Havens MA, Hastings ML. 2016. Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Res*, 44(14):6549–6563
- Hermey G, Blüthgen N, Kuhl D. 2017. Neuronal activity-regulated alternative mRNA splicing. *Int J Biochem Cell Biol*, 91:184–193
- Herzel L, Ottoz DSM, Alpert T, Neugebauer KM. 2017. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol*, 18(10):637–650
- Hirabayashi Y, Itoh Y, Tabata H, Nakajima K, Akiyama T, Masuyama N, Gotoh Y. 2004. The Wnt/beta-catenin pathway directs neuronal differentiation of cortical neural precursor cells. *Development*, 131(12):2791–2801
- Hollander D, Naftelberg S, Lev-Maor G, Kornblihtt AR, Ast G. 2016. How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing? *Trends Genet*, 32(10):596–606
- Hooper JE. 2014. A survey of software for genome-wide discovery of differential splicing in RNA-Seq data. *Hum Genomics*, 8(1):3.
- Hu S, Wang X, Shan G. 2016. Insertion of an Alu element in a lncRNA leads to primate-specific modulation of alternative splicing. *Nat Struct Mol Biol*, 23(11):1011-1019
- Huang CS, Shi SH, Ule J, Ruggiu M, Barker LA, Darnell RB, Jan YN, Jan LY. 2005. Common molecular pathways mediate long-term potentiation of synaptic excitation and slow synaptic inhibition. *Cell*, 123(1):105–118
- Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37(1):1–13
- Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1):44–57.
- Huang Y, Li W, Yao X, Lin QJ, Yin JW, Liang Y, Heiner M, Tian B, Hui J, Wang G. 2012. Mediator complex regulates alternative mRNA processing via the MED23 subunit. *Mol Cell*, 45(4):459–469
- Huo J, Xu S, Lam KP. 2019. FAIM: An Antagonist of Fas-Killing and Beyond. *Cells*, 8(6):541
- Iacopetti P, Michelini M, Stuckmann I, Oback B, Aaku-Saraste E, Huttner WB. 1999. Expression of the antiproliferative gene TIS21 at the onset of neurogenesis identifies single neuroepithelial cells that switch from proliferative to neuron-generating division. *Proc Natl Acad Sci USA*, 96(8):4639–4644
- Imayoshi I, Kageyama R. 2014. bHLH factors in self-renewal, multipotency, and fate choice of neural progenitor cells. *Neuron*, 82(1):9–23
- Irimia M, Weatheritt RJ, Ellis JD, Parikhshak NN, Gonatopoulos-Pournatzis T, Babor M, Quesnel-Vallièrès M, Tapial J, Raj B, O’Hanlon D, Barrios-Rodiles M, Sternberg MJ, Cordes SP, Roth FP, Wrana JL, Geschwind DH, Blencowe BJ. 2014. A highly

- conserved program of neuronal microexons is misregulated in autistic brains. *Cell*, 159(7):1511–1523
- Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lönnerberg P, Linnarsson S. 2014. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods*, 11(2):163–166
- Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, Malla S, Marriott H, Nieto T, O’Grady J, Olsen HE, Pedersen BS, Rhie A, Richardson H, Quinlan AR, Snutch TP, Tee L, Paten B, Phillippy AM, Simpson JT, Loman NJ, Loose M. 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol*, 36(4):338–345
- Jensen KB, Dredge BK, Stefani G, Zhong R, Buckanovich RJ, Okano HJ, Yang YYL, Darnell RB. 2000. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron*, 25(2):359–371
- Kamiya D, Banno S, Sasai N, Ohgushi M, Inomata H, Watanabe K, Kawada M, Yakura R, Kiyonari H, Nakao K, Jakt LM, Nishikawa S, Sasai Y. 2011. Intrinsic transition of embryonic stem-cell differentiation into neural progenitors. *Nature*, 470(7335):503–509
- Katz Y, Wang ET, Airoidi EM, Burge CB. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods*, 7(12):1009–1015
- Keilani S, Sugaya K. 2008. Reelin induces a radial glial phenotype in human neural progenitor cells by activation of Notch-1. *BMC Dev Biol*, 8(1):69
- Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, Stamm S. 2013. Function of alternative splicing. *Gene*, 514(1):1–30
- Keppetipola N, Sharma S, Li Q, Black DL. 2012. Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2. *Crit Rev Biochem Mol Biol*, 47(4):360–378
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, 14(4):R36
- Kim E, Magen A, Ast G. 2007. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res*, 35(1):125–131
- Kim KK, Nam J, Mukoyama YS, Kawamoto S. 2013. Rbfox3-regulated alternative splicing of Numb promotes neuronal differentiation during development. *J Cell Biol*, 200(4):443–458
- Komada M. 2012. Sonic hedgehog signaling coordinates the proliferation and differentiation of neural stem/progenitor cells by regulating cell cycle kinetics during development of the neocortex. *Congenit Anom (Kyoto)*, 52(2):72–77
- Konermann S, Lotfy P, Brideau NJ, Oki J, Shokhirev MN, Hsu PD. 2018. Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell*, 173(3):665–676.e14

- Kornblihtt AR. 2005. Promoter usage and alternative splicing. *Curr Opin Cell Biol*, 17(3):262–268
- Kornblihtt AR. 2015. Transcriptional control of alternative splicing along time: ideas change, experiments remain. *RNA*, 21(4):670–672
- Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol*, 14(3):153–165
- Kosodo Y, Röper K, Haubensak W, Marzesco AM, Corbeil D, Huttner WB. 2004. Asymmetric distribution of the apical plasma membrane during neurogenic divisions of mammalian neuroepithelial cells. *EMBO J*, 23(11):2314–2324
- Kriegstein A, Alvarez-Buylla A. 2009. The glial nature of embryonic and adult neural stem cells. *Annu Rev Neurosci*, 32:149–184
- Kulkarni A, Anderson AG, Merullo DP, Konopka G. 2019. Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Curr Opin Biotechnol*, 58:129–136
- Kumar P, Tan Y, Cahan P. 2017. Understanding development and stem cells using single cell-based analyses of gene expression. *Development*, 144(1):17–32
- Lakomá J, Garcia-Alonso L, Luque JM. 2011. Reelin sets the pace of neocortical neurogenesis. *Development*, 138(23):5223–5234
- Lambert C, Landau AM, Desbarats J. 2003. Fas—Beyond death: a regenerative role for Fas in the nervous system. *Apoptosis*, 8(6):551–562
- Lange C, Huttner WB, Calegari F. 2009. Cdk4/CyclinD1 overexpression in neural stem cells shortens G1, delays neurogenesis, and promotes the generation and expansion of basal progenitors. *Cell Stem Cell*, 5(3):320–331
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4):357–359
- Latysheva NS, Flock T, Weatheritt RJ, Chavali S, Babu MM. 2015. How do disordered regions achieve comparable functions to structured domains? *Protein Sci*, 24(6):909–922
- Laurent B, Ruitu L, Murn J, Hempel K, Ferrao R, Xiang Y, Liu S, Garcia BA, Wu H, Wu F, Steen H, Shi Y. 2015. A Specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. *Mol Cell*, 57(6):957–970
- van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM. 2014. Classification of intrinsically disordered regions and proteins. *Chem Rev*, 114(13):6589–6631
- Lehtinen MK, Walsh CA. 2011. Neurogenesis at the brain–cerebrospinal fluid interface. *Annu Rev Cell Dev Biol*, 27(1):653–679

- Lehtinen MK, Zappaterra MW, Chen X, Yang YJ, Hill AD, Lun M, Maynard T, Gonzalez D, Kim S, Ye P, D'Ercole AJ, Wong ET, LaMantia AS, Walsh CA. 2011. The cerebrospinal fluid provides a proliferative niche for neural progenitor cells. *Neuron*, 69(5):893–905
- Lence T, Akhtar J, Bayer M, Schmid K, Spindler L, Ho CH, Kreim N, Andrade-Navarro MA, Poeck B, Helm M, Roignant JY. 2016. m⁶A modulates neuronal functions and sex determination in *Drosophila*. *Nature*, 540(7632):242–247
- Li Q, Zheng S, Han A, Lin CH, Stoilov P, Fu XD, Black DL. 2014. The splicing regulator PTBP2 controls a program of embryonic splicing required for neuronal maturation. *ELife*, 3: e01201
- Li X, Liu S, Zhang L, Issaian A, Hill RC, Espinosa S, Shi S, Cui Y, Kappel K, Das R, Hansen KC, Zhou ZH, Zhao R. 2019. A unified mechanism for intron and exon definition and back-splicing. *Nature*, 573(7774):375–380
- Li YI, Sanchez-Pulido L, Haerty W, Ponting CP. 2015. RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Res*, 25(1):1–13
- Licatalosi DD, Darnell RB. 2006. Splicing regulation in neurologic disease. *Neuron*, 52(1):93–101
- Licatalosi DD, Yano M, Fak JJ, Mele A, Grabinski SE, Zhang C, Darnell RB. 2012. Ptpb2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain. *Genes Dev*, 26(14):1626–1642
- Liu A, Niswander LA. 2005. Bone morphogenetic protein signalling and vertebrate nervous system development. *Nat Rev Neurosci*, 6(12):945–954
- Liu J, Geng A, Wu X, Lin RJ, Lu Q. 2018. Alternative RNA Splicing Associated With Mammalian Neuronal Differentiation. *Cereb Cortex*, 28(8):2810–2816
- Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods*, 25(4):402–408
- Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. 2011. Epigenetics in alternative pre-mRNA splicing. *Cell*, 144(1):16–26
- Makeyev EV, Zhang J, Carrasco MA, Maniatis T. 2007. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol Cell*, 27(3):435–448
- Martinez-Contreras R, Cloutier P, Shkreta L, Fiset JF, Revil T, Chabot B. 2007. hnRNP proteins and splicing control. In: Blencowe BJ, Graveley BR (eds) *Alternative Splicing in the Postgenomic Era*. Springer New York, New York, NY, pp. 123–147
- Martínez-Mármol R, Barneda-Zahonero B, Soto D, Andrés RM, Coccia E, Gasull X, Planells-Ferrer L, Moubarak RS, Soriano E, Comella JX. 2016. FAIM-L regulation of XIAP degradation modulates Synaptic Long-Term Depression and Axon Degeneration. *Sci Rep*, 6(1):35775

- de la Mata M, Kornblihtt AR. 2006. RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat Struct Mol Biol*, 13(11):973–980
- Maunakea AK, Chepelev I, Cui K, Zhao K. 2013. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res*, 23(11):1256–1269.
- Mehmood A, Laiho A, Venäläinen MS, McGlinchey AJ, Wang N, Elo LL. 2019. Systematic evaluation of differential splicing tools for RNA-seq studies. *Brief Bioinform*, 5:bbz126
- Mérot Y, Rétaux S, Heng JI. 2009. Molecular mechanisms of projection neuron production and maturation in the developing cerebral cortex. *Semin Cell Dev Biol*, 20(6):726–734
- Misteli T, Spector DL. 1999. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol Cell*, 3(6):697–705
- Miyata T, Kawaguchi A, Saito K, Kawano M, Muto T, Ogawa M. 2004. Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells. *Development*, 131(13):3133–3145
- Mizutani K, Yoon K, Dang L, Tokunaga A, Gaiano N. 2007. Differential Notch signalling distinguishes neural stem cells from intermediate progenitors. *Nature*, 449(7160):351–355
- Mockenhaupt S, Makeyev EV. 2015. Non-coding functions of alternative pre-mRNA splicing in development. *Semin Cell Dev Biol*, 47–48:32–39
- Moubarak RS, Planells-Ferrer L, Urresti J, Reix S, Segura MF, Carriba P, Marqués-Fernández F, Sole C, Llecha-Cano N, Lopez-Soriano J, Sanchis D, Yuste VJ, Comella JX. 2013. FAIM-L is an IAP-binding protein that inhibits XIAP ubiquitinylation and protects from Fas-induced apoptosis. *J Neurosci*, 33(49):19262–19275
- Mukhtar T, Taylor V. 2018. Untangling Cortical Complexity During Development. *J Exp Neurosci*, 12:1179069518759332
- Muñoz MJ, de la Mata M, Kornblihtt AR. 2010. The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends Biochem Sci*, 35(9):497–504
- Naftelberg S, Schor IE, Ast G, Kornblihtt AR. 2015. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu Rev Biochem*, 84(1):165–198
- Nieto Moreno N, Giono LE, Cambindo Botto AE, Muñoz MJ, Kornblihtt AR. 2015. Chromatin, DNA structure and alternative splicing. *FEBS Lett*, 589(22):3370–3378
- Noctor SC, Flint AC, Weissman TA, Wong WS, Clinton BK, Kriegstein AR. 2002. Dividing precursor cells of the embryonic cortical ventricular zone have morphological and molecular characteristics of radial glia. *J Neurosci*, 22(8):3161–3173

- Noctor SC, Martínez-Cerdeño V, Ivic L, Kriegstein AR. 2004. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat Neurosci*, 7(2):136–144
- Noctor SC, Martínez-Cerdeño V, Kriegstein AR. 2008. Distinct behaviors of neural stem and progenitor cells underlie cortical neurogenesis. *J Comp Neurol*, 508(1):28–44
- Nojima T, Rebelo K, Gomes T, Grosso AR, Proudfoot NJ, Carmo-Fonseca M. 2018. RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol Cell*, 72(2):369-379.e4
- Norris AD, Calarco JA. 2012. Emerging Roles of Alternative Pre-mRNA Splicing Regulation in Neuronal Development and Function. *Front Neurosci*, 6:122
- Norris AD, Gao S, Norris ML, Ray D, Ramani AK, Fraser AG, Morris Q, Hughes TR, Zhen M, Calarco JA. 2014. A Pair of RNA-binding proteins controls networks of splicing events contributing to specialization of neural cell types. *Mol Cell*, 54(6):946–959
- Ohnishi T, Shirane M, Hashimoto Y, Saita S, Nakayama KI. 2014. Identification and characterization of a neuron-specific isoform of protrudin. *Genes Cells*, 19(2):97–111
- Palm K, Belluardo N, Metsis M, Timmusk T. 1998. Neuronal expression of zinc finger transcription factor REST/NRSF/XBR gene. *J Neurosci*, 18(4):1280–1296
- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, 40(12):1413–1415
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*, 14(4):417–419
- Patro R, Mount SM, Kingsford C. 2014. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol*, 32(5):462–464
- Phipson B, Zappia L, Oshlack A. 2017. Gene length and detection bias in single cell RNA sequencing protocols. *F1000Res*, 6:595
- Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods*, 10(11):1096–1098
- Planells-Ferrer L, Urresti J, Coccia E, Galenkamp KM, Calleja-Yagüe I, López-Soriano J, Carriba P, Barneda-Zahonero B, Segura MF, Comella JX. 2016. Fas apoptosis inhibitory molecules: more than death-receptor antagonists in the nervous system. *J Neurochem*, 139(1):11–21
- Quesnel-Vallières M, Dargaie Z, Irimia M, Gonatopoulos-Pournatzis T, Ip JY, Wu M, Sterne-Weiler T, Nakagawa S, Woodin MA, Blencowe BJ, Cordes SP. 2016. Misregulation of an Activity-Dependent Splicing Network as a Common Mechanism Underlying Autism Spectrum Disorders. *Mol Cell*, 64(6):1023–1034

- Quesnel-Vallières M, Irimia M, Cordes SP, Blencowe BJ. 2015. Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes Dev*, 29(7):746–759
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842
- R Core Team. 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria URL: <https://www.R-project.org/>
- Raj B, Blencowe BJ. 2015. Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron*, 87(1):14–27
- Raj B, Irimia M, Braunschweig U, Sterne-Weiler T, O’Hanlon D, Lin ZY, Chen GI, Easton LE, Ule J, Gingras AC, Eyraas E, Blencowe BJ. 2014. A global regulatory mechanism for activating an exon network required for neurogenesis. *Mol Cell*, 56(1):90–103
- Raj B, O’Hanlon D, Vessey JP, Pan Q, Ray D, Buckley NJ, Miller FD, Blencowe BJ. 2011. Cross-regulation between an alternative splicing activator and a transcription repressor controls neurogenesis. *Mol Cell*, 43(5):843–850
- Raj T, Li YI, Wong G, Humphrey J, Wang M, Ramdhani S, Wang YC, Ng B, Gupta I, Haroutunian V, Schadt EE, Young-Pearse T, Mostafavi S, Zhang B, Sklar P, Bennett DA, De Jager PL. 2018. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer’s disease susceptibility. *Nat Genet*, 50(11):1584–1592
- Rash BG, Lim HD, Breunig JJ, Vaccarino FM. 2011. FGF signaling expands embryonic cortical surface area by regulating Notch-dependent neurogenesis. *J Neurosci*, 31(43):15604–15617
- Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, Na H, Irimia M, Matzat LH, Dale RK, Smith SA, Yarosh CA, Kelly SM, Nabet B, Mecnas D, Li W, Laishram RS, Qiao M, Lipshitz HD, Piano F, Corbett AH, Carstens RP, Frey BJ, Anderson RA, Lynch KW, Penalva LO, Lei EP, Fraser AG, Blencowe BJ, Morris QD, Hughes TR. 2013. A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, 499(7457):172–177
- Rhoads A, Au KF. 2015. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*, 13(5):278–289
- Rice DS, Curran T. 2001. Role of the Reelin signaling pathway in central nervous system development. *Annu Rev Neurosci*, 24:1005–1039
- Roberts GC, Gooding C, Mak HY, Proudfoot NJ, Smith CW. 1998. Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Res*, 26(24):5568–5572
- Romero-Barrios N, Legascue MF, Benhamed M, Ariel F, Crespi M. 2018. Splicing regulation by long noncoding RNAs. *Nucleic Acids Res*, 46(5):2169–2184

- Rubenstein J, Rakic P. 2013. Patterning and Cell Type Specification in the Developing CNS and PNS: Comprehensive Developmental Neuroscience. Academic Press
- Ruzinova MB, Benezra R. 2003. Id proteins in development, cell cycle and cancer. *Trends Cell Biol*, 13(8):410–418
- Saito Y, Yuan Y, Zucker-Scharff I, Fak JJ, Jereb S, Tajima Y, Licatalosi DD, Darnell RB. 2019. Differential NOVA2-Mediated Splicing in Excitatory and Inhibitory Neurons Regulates Cortical Development and Cerebellar Function. *Neuron*, 101(4):707-720.e5
- Saldi T, Cortazar MA, Sheridan RM, Bentley DL. 2016. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J Mol Biol*, 428(12):2623–2635
- Sauer FC. 1936. The interkinetic migration of embryonic epithelial nuclei. *J Morphol*, 60(1):1–11
- Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N. 2015. Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI). *Curr Protoc Hum Genet*, 87:11.16.1-11.16.14
- Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL. 2000. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell*, 101(6):671–684
- Schneider TJ, Fischer GM, Donohoe TJ, Colarusso TP, Rothstein TL. 1999. A novel gene coding for a Fas apoptosis inhibitory molecule (FAIM) isolated from inducibly Fas-resistant B lymphocytes. *J Exp Med*, 189(6):949–956
- Schoenherr CJ, Anderson DJ. 1995. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science*, 267(5202):1360-3.
- Segura MF, Sole C, Pascual M, Moubarak RS, Perez-Garcia MJ, Gozzelino R, Iglesias V, Badiola N, Bayascas JR, Llecha N, Rodriguez-Alvarez J, Soriano E, Yuste VJ, Comella JX. 2007. The long form of Fas apoptotic inhibitory molecule is expressed specifically in neurons and protects them against death receptor-triggered apoptosis. *J Neurosci*, 27(42):11228–11241
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11):2498–2504
- Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res*, 15(17):7155–7174
- Shayevitch R, Askayo D, Keydar I, Ast G. 2018. The importance of DNA methylation of exons on alternative splicing. *RNA*, 24(10):1351–1362
- Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. 2014. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci USA*, 111(51):E5593–E5601

- Sheng K, Cao W, Niu Y, Deng Q, Zong C. 2017. Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat Methods*, 14(3):267–270
- Shimojo H, Ohtsuka T, Kageyama R. 2011. Dynamic expression of Notch signaling genes in neural stem/progenitor cells. *Front Neurosci*, 5:78
- Shinohara H, Sakayori N, Takahashi M, Osumi N. 2013. Ninein is essential for the maintenance of the cortical progenitor character by anchoring the centrosome to microtubules. *Biol Open*, 2(7):739–749
- Siegenthaler JA, Ashique AM, Zarbali K, Patterson KP, Hecht JH, Kane MA, Folias AE, Choe Y, May SR, Kume T, Napoli JL, Peterson AS, Pleasure SJ. 2009. Retinoic acid from the meninges regulates cortical neuron generation. *Cell*, 139(3):597–609
- Sole C, Dolcet X, Segura MF, Gutierrez H, Diaz-Meco MT, Gozzelino R, Sanchis D, Bayascas JR, Gallego C, Moscat J, Davies AM, Comella JX. 2004. The death receptor antagonist FAIM promotes neurite outgrowth by a mechanism that depends on ERK and NF- κ B signaling. *J Cell Biol*, 167(3):479–492
- Song Y, Botvinnik OB, Lovci MT, Kakaradov B, Liu P, Xu JL, Yeo GW. 2017. Single-Cell Alternative Splicing Analysis with Expedition Reveals Splicing Dynamics during Neuron Differentiation. *Mol Cell*, 67(1):148-161.e5
- Spellman R, Llorian M, Smith CW. 2007. Crossregulation and functional redundancy between the splicing regulator PTB and its Paralogs nPTB and ROD1. *Mol Cell*, 27(3):420–434
- Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H. 2005. Function of alternative splicing. *Gene*, 344:1–20
- Stark R, Grzelak M, Hadfield J. 2019. RNA sequencing: the teenage years. *Nat Rev Genet*, 20(11):631–656
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell*, 177(7):1888-1902.e21
- Stuart T, Satija R. 2019. Integrative single-cell analysis. *Nat Rev Genet*, 20(5):257–272
- Su CH, D D, Tarn WY. 2018. Alternative Splicing in Neurogenesis and Brain Development. *Front Mol Biosci*, 5:12
- Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC, Cvejic A, Teichmann SA. 2017. Power analysis of single-cell RNA-sequencing experiments. *Nat Methods*, 14(4):381–387
- Tacke R, Manley JL. 1999. Determinants of SR protein specificity. *Curr Opin Cell Biol*, 11(3):358–362
- Takata A, Matsumoto N, Kato T. 2017. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat Commun*, 8:14519

- Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*, 6(5):377–382
- Tapial J, Ha KCH, Sterne-Weiler T, Gohr A, Braunschweig U, Hermoso-Pulido A, Quesnel-Vallières M, Permanyer J, Sodaei R, Marquez Y, Cozzuto L, Wang X, Gómez-Velázquez M, Rayon T, Manzanares M, Ponomarenko J, Blencowe BJ, Irimia M. 2017. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res*, 27(10):1759–1768
- Tardaguila M, de la Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, Del Risco H, Ferrell M, Mellado M, Macchietto M, Verheggen K, Edelmann M, Ezkurdia I, Vazquez J, Tress M, Mortazavi A, Martens L, Rodriguez-Navarro S, Moreno-Manzano V, Conesa A. 2018. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res*, 28(3):396–411
- Taverna E, Götz M, Huttner WB. 2014. The cell biology of neurogenesis: toward an understanding of the development and evolution of the neocortex. *Annu Rev Cell Dev Biol*, 30(1):465–502
- Tiberi L, Vanderhaeghen P, van den Aemele J. 2012. Cortical neurogenesis and morphogens: diversity of cues, sources and functions. *Curr Opin Cell Biol*, 24(2):269–276
- Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcárcel J, Guigó R. 2009. Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol*, 16(9):996–1001
- Torres-Méndez A, Bonnal S, Marquez Y, Roth J, Iglesias M, Permanyer J, Almudí I, O'Hanlon D, Guitart T, Soller M, Gingras AC, Gebauer F, Rentzsch F, Blencowe BJ, Valcárcel J, Irimia M. 2019. A novel protein domain in an ancestral splicing factor drove the evolution of neural microexons. *Nat Ecol Evol*, 3(4):691–701
- Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*, 31(1):46–53
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 7(3):562–578
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*, 28(5):511–515
- Treutlein B, Gokce O, Quake SR, Südhof TC. 2014. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proc Natl Acad Sci USA*, 111(13):E1291–E1299

- Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, Eyraas E. 2018. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol*, 19(1):40
- Tzeng SF. 2003. Inhibitors of DNA binding in neural cell proliferation and differentiation. *Neurochem Res*, 28(1):45-52
- Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature*, 444(7119):580–586
- Ullrich B, Ushkaryov YA, Südhof TC. 1995. Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. *Neuron*, 14(3):497–507
- Ushkaryov YA, Petrenko AG, Geppert M, Südhof TC. 1992. Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor and laminin. *Science*, 257(5066):50-56.
- Uversky VN. 2017. Protein intrinsic disorder-based liquid–liquid phase transitions in biological systems: Complex coacervates and membrane-less organelles. *Adv Colloid Interface Sci*, 239:97–114
- Vu TN, Wills QF, Kalari KR, Niu N, Wang L, Pawitan Y, Rantalainen M. 2018. Isoform-level gene expression patterns in single-cell RNA-sequencing data. *Bioinformatics*, 34(14):2392–2400
- Vuong JK, Lin CH, Zhang M, Chen L, Black DL, Zheng S. 2016. PTBP1 and PTBP2 Serve Both Specific and Redundant Functions in Neuronal Pre-mRNA Splicing. *Cell Rep*, 17(10):2766–2775
- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221):470–476
- Wang Y, Navin NE. 2015. Advances and applications of single-cell sequencing technologies. *Mol Cell*, 58(4):598–609
- Wang Y, Wang Z. 2016. Design of RNA-Binding Proteins: Manipulate Alternative Splicing in Human Cells with Artificial Splicing Factors. In: Lin RJ. (eds) *RNA-Protein Complexes and Interactions*. *Methods in Molecular Biology*, 1421. Humana Press, New York, NY
- Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, 14(5):802–813
- Watson JD. 2014. *Molecular biology of the gene*. Seventh edition. Pearson, Boston.
- Wen L, Tang F. 2016. Single-cell sequencing in stem cell biology. *Genome Biol*, 17:71
- Weyn-Vanhenyck SM, Mele A, Yan Q, Sun S, Farny N, Zhang Z, Xue C, Herre M, Silver PA, Zhang MQ, Krainer AR, Darnell RB, Zhang C. 2014. HITS-CLIP and Integrative

- modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep*, 6(6):1139–1152
- Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York
- Wilkinson G, Dennis D, Schuurmans C. 2013. Proneural genes in neocortical development. *Neuroscience*, 253:256–273
- Wu SX, Goebbels S, Nakamura K, Nakamura K, Kometani K, Minato N, Kaneko T, Nave KA, Tamamaki N. 2005. Pyramidal neurons of upper cortical layers generated by NEX-positive progenitor cells in the subventricular zone. *Proc Natl Acad Sci USA*, 102(47):17172–17177
- Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, 26(7):873–881
- Yang L, Chen LL. 2014. Microexons go big. *Cell*, 159(7):1488–1489
- Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Marugán JC, Cummins C, Davidson C, Dodiya K, Fatima R, Gall A, Giron CG, Gil L, Grego T, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, Kay M, Lavidas I, Le T, Lemos D, Martinez JG, Maurel T, McDowall M, McMahon A, Mohanan S, Moore B, Nuhn M, Oheh DN, Parker A, Parton A, Patricio M, Sakthivel MP, Abdul Salam AI, Schmitt BM, Schuilenburg H, Sheppard D, Sycheva M, Szuba M, Taylor K, Thormann A, Threadgold G, Vullo A, Walts B, Winterbottom A, Zadissa A, Chakiachvili M, Flint B, Frankish A, Hunt SE, Ilesley G, Kostadima M, Langridge N, Loveland JE, Martin FJ, Morales J, Mudge JM, Muffato M, Perry E, Ruffier M, Trevanion SJ, Cunningham F, Howe KL, Zerbino DR, Flicek P. 2020. Ensembl 2020. *Nucleic Acids Res*, 48(D1):D682-D688
- Yearim A, Gelfman S, Shayevitch R, Melcer S, Glaich O, Mallm JP, Nissim-Rafinia M, Cohen AH, Rippe K, Meshorer E, Ast G. 2015. HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell Rep*, 10(7):1122–1134
- Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*, 11(2–3):377-394
- Yeo G, Holste D, Kreiman G, Burge CB. 2004. Variation in alternative splicing across human tissues. *Genome Biol*, 5(10):R74
- Zhang X, Chen MH, Wu X, Kodani A, Fan J, Doan R, Ozawa M, Ma J, Yoshida N, Reiter JF, Black DL, Kharchenko PV, Sharp PA, Walsh CA. 2016. Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. *Cell*, 166(5):1147-1162.e15
- Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R, Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente WJ, Ericson NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ, Bielas JH. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*, 8(1):14049

- Zheng S, Gray EE, Chawla G, Porse BT, O'Dell TJ, Black DL. 2012. PSD-95 is post-transcriptionally repressed during early neural development by PTBP1 and PTBP2. *Nat Neurosci*, 15(3):381–388
- Zhong X, Schneider TJ, Cabral DS, Donohoe TJ, Rothstein TL. 2001. An alternatively spliced long form of Fas apoptosis inhibitory molecule (FAIM) with tissue-specific expression in the brain. *Mol Immunol*, 38(1):65–72
- Zhou Z, Fu XD. 2013. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*, 122(3):191–207

Acknowledgments

I want to thank Fede, for the opportunity to join his lab and to always challenge me to give my best, and my TAC members, Nadine Vastenhow and Michael Hiller for helpful suggestions for my project.

I am grateful to all the CRTD staff, especially the Deep Sequencing facility for their help in analyzing the data. Special thanks to Mathias who introduced me to the tricky world of bioinformatics.

A big thank to my lab members, in particular Juli, who started the work on the Btg2^{RFP}/Tubb3^{GFP} mouse line that opened several interesting projects, including mine. To Bea, who joint me in this project and is now providing valuable data for a deeper understanding of alternative splicing in neocortical development. To Sara, Ivan, Gaby, Simon, Frank and Daniel for the fruitful discussions, the constructive criticisms, the nice atmosphere and fun we had in and outside the lab. I am very grateful to Simon for his patience in correcting my thesis and listening to my complains while pushing me to my limits during our running sessions: thank you my friend and all the best for your next chapter!

To Chiara Ceriotti, Alice Dimitra Cezanne and Frank Darmis for feedbacks and additional revisions of this thesis: your suggestions really helped me a lot!

To all the people that I met in Dresden and contributed to make this experience unique and make me grow as a person.

Last, but not least, I would like to thank my family, for their love and for always supporting me in my choice to pursue a scientific path. Grazie, per essere sempre stati al mio fianco in questi anni e avermi sempre spinto ad andare avanti anche nei momenti piu' difficili: vi voglio bene!

THANK YOU ALL!

Anlange I

Technische Universität Dresden

Medizinische Fakultät Carl Gustav Carus

Promotionsordnung vom 24. Juli 2011

Erklärungen zur Eröffnung des Promotionsverfahrens

1. Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.
2. Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts habe ich Unterstützungsleistungen von folgenden Personen erhalten: nicht zutreffend.
3. Weitere Personen waren an der geistigen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.
4. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.
5. Die Inhalte dieser Dissertation wurden in folgender Form veröffentlicht: nicht zutreffend
6. Ich bestätige, dass es keine zurückliegenden erfolglosen Promotionsverfahren gab.
7. Ich bestätige, dass ich die Promotionsordnung der Medizinischen Fakultät der Technischen Universität Dresden anerkenne.
8. Ich habe die Zitierrichtlinien für Dissertationen an der Medizinischen Fakultät der Technischen Universität Dresden zur Kenntnis genommen und befolgt.

Dresden, 23.10.2020

Leila Haj Abdullah Alieh

Anlange II

Hiermit bestätige ich die Einhaltung der folgenden aktuellen gesetzlichen Vorgaben im Rahmen meiner Dissertation

- das zustimmende Votum der Ethikkommission bei Klinischen Studien, epidemiologischen Untersuchungen mit Personenbezug oder Sachverhalten, die das Medizinproduktegesetz betreffen

Aktenzeichen der zuständigen Ethikkommission

DD 24-9168.11-1/2011-11, HD 35-9185.81/G-61/15

- die Einhaltung der Bestimmungen des Tierschutzgesetzes

Aktenzeichen der Genehmigungsbehörde zum Vorhaben/zur Mitwirkung

LDS Achtenzeichen: 11-1-2011-41, TVV 16-2018

- die Einhaltung des Gentechnikgesetzes

Projektnummer

LDS Achtenzeichen: 55-8811.71/210

- die Einhaltung von Datenschutzbestimmungen der Medizinischen Fakultät und des Universitätsklinikums Carl Gustav Carus.

Dresden, 23.10.2020

Leila Haj Abdullah Alieh

