The Role of N1-Src regulated splicing in neuronal differentiation

Sofia Kudasheva

MSc by Research

University of York

March 2021

Abstract

Alternative splicing (AS) is one of the main contributors to transcriptome diversity and functional complexity involved in the process of neuronal development. Evidence suggests that many splicing regulators and alternative splicing events are neuron-specific and aberrations in the regulation of these events have been linked to various neurodevelopmental disorders. N1-Src is an evolutionarily conserved neuronal splice variant of the ubiquitous tyrosine kinase c-Src. It has been implicated in neural development and as a prognostic indicator in neuroblastoma, a childhood cancer that is caused by failure of neural crest cells to differentiate. Results from knockdown experiments where N1 exon inclusion was prevented with splice-blocking antisense morpholino oligos revealed that N1-Src is a key regulator of primary neurogenesis in *Xenopus*. Preliminary short and long read RNA-Seq data from Xenopus embryos suggest a role for N1-Src in regulation of an alternative splicing programme during early neurogenesis, with transcripts encoding the splicing/RNA processing machinery themselves being the most spliced targets. This study aimed to further describe the N1-Src-regulated splicing network in the developing *Xenopus* nervous system using bioinformatic analysis of various publicly available and Evans/Isaacs lab RNASeq datasets. A differential splicing (DS) analysis pipeline was developed to detect and quantify alternative splicing events that occur during early stages of *Xenopus* embryo development relevant to neurogenesis. By correlating alternative splicing quantifications with RNA-binding protein motif enrichment analysis, this project proposed mechanisms for Src regulation of alternative splicing.

List of contents

List of figures

List of tables

Acknowledgements

I would like to thank Gareth Evans for his excellent supervision and continual support. Thank you to my TAP panel, Pegine and Harv for their support and insightful comments. I wish to thank Katherine Newling and the whole York Genomics and Bioinformatics lab for their continual support and their expertise in the area of bioinformatics.

Author's declaration

I declare that this thesis is my original work and I am the sole author. This work has not previously been presented for an award. All external sources of information are acknowledged as references.

Chapter 1: Introduction

1.1: Alternative splicing

RNA splicing is an mRNA processing mechanism occurring in eukaryotic organisms whereby the core splicing machinery, the spliceosome, binds the conserved splice sites, removes introns and ligates exons together to generate mature mRNA (Jurica and Roybal, 2013). The spliceosome complex comprises five small nuclear ribonucleoprotein particles (snRNPs) and a large number of auxiliary proteins. A conserved set of *cis*-acting elements known as the core splicing signals (5′ and 3′ splice sites, branch site, and polypyrimidine tract) guides the interactions between spliceosomal components and pre-mRNA (Singh, 2002). The spliceosome assembly begins with the recognition of the 5′ splice site by the snRNP U1 and the binding of splicing factor 1 (SF1) to the branch point and of the U2 auxiliary factor (U2AF) to the polypyrimidine tract and 3′ terminal AG (Chen and Manley, 2009; Wang et al., 2015). Splice site selection is controlled by many separate components, including non-spliceosomal splicing regulatory factors (Jangi et al., 2014; Matera and Wang, 2014), RNA secondary structures (McManus and Graveley, 2011), RNA polymerase elongation speed (Fong et al., 2014), and epigenetic regulation (Luco et al., 2010).

Alternative splicing (AS) is a highly regulated process by which different pairs of splice sites are selected to produce multiple RNA and protein isoforms from a single gene. *Trans*-acting splicing factors (SFs) bind to enhancer or silencer motifs close to 5′ and 3′ splice acceptor sites to promote or prevent the usage of a particular splice site by the spliceosome (Figure 1). The serine/arginine-rich (SR) protein family generally play a role in promoting spliceosome formation and binding of splicing machinery to the new RNA transcript, whilst the heterogeneous nuclear ribonucleoproteins (hnRNPs) often function as antagonists to SR-protein-regulated alternative splicing events (Figure 1). hnRNPs bind to exon splicing silencers and inhibit the inclusion of exons (Wang et al., 2015; Dvinge, 2018).

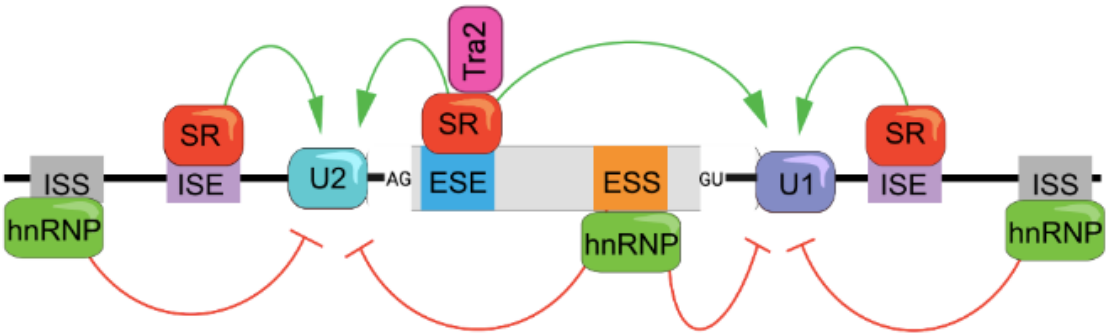


Figure 1. An overview of the regulation of alternative splicing.
SR proteins bind to exonic splicing enhancers (ESEs) and U-rich intronic splicing enhancers (ISEs) to stimulate the binding of U2 small nuclear ribonucleoprotein (snRNP) to the upstream 3′ splice site (ss) or the binding of the U1 snRNP to the downstream 5′ ss. SR proteins function with other splicing co-activators, such as transformer 2 (TRA2). hnRNPs bind ESSs and ISSs and prevent binding of the snRNPs.

In addition to generating functionally distinct protein isoforms, AS may lead to changes in localization of proteins, their post-translational modifications or binding affinities to ligands. AS can also control gene expression levels. This is accomplished through different mechanisms, for example, by intron retention (IR) or inclusion of alternative exons containing premature termination codons (PTC) into the mRNA. PTC-containing transcripts are exported to the cytoplasm and targeted to the nonsense-mediated decay (NMD) pathway (Lareau et al., 2007). Many vertebrate introns contain PTCs, however IR transcripts can also be degraded by a mechanism independent of NMD, which requires components of the nuclear RNA surveillance machinery, including the nuclear pore-associated protein Tpr and the exosome complex (Yap et al., 2012). Many splicing factors, including polypyrimidine-tract-binding protein 1 (PTBP1), autoregulate expression levels in this way by binding their own pre-mRNAs and promoting unproductive splicing events (Wollerton et al., 2004; Pervouchine et al., 2019; Ding et al., 2020). Moreover, AS can generate alternative 5′ and 3′ untranslated regions (UTRs), which impact translation efficiency, mRNA stability and localization in the cytoplasm (Hughes, 2006).

1.2: Splicing and neural development

AS role in generating complex proteomic diversity explains the correlation between AS and the complexity of vertebrate central nervous systems. Vertebrate neural development involves dramatic morphological and functional changes in individual cells as they differentiate from neural progenitors (NPCs) to neurons. The neuronal cell fate depends on its positioning along the anterior-posterior and dorsal-ventral neural axis of the vertebrate embryo. Patterning of this axis is shaped through position-dependent gradients of signalling molecules. Alternative splicing and RNA-binding proteins along with other various signalling pathways and transcription factors that respond to these gradients are responsible for the development of

correct functional neural cell types with specific transcript and protein expression profiles at the correct place and time (Yeo et al., 2004; Raj and Blencowe, 2015; Weyn-Vanhentenryck et al., 2018). A large body of evidence suggests that neurons have developed unique systems for RNA processing, with many RNA binding proteins (RBPs) being specifically expressed in neurons. Mutations in neuron-specific RBPs and aberrations in neural AS patterns have been linked to neurological disorders (Licatalosi and Darnell, 2006). Misregulation of splicing has been repeatedly implicated in autism spectrum disorder (Irimia et al., 2014; Gonatopoulos-Pournatzis et al., 2020).

While most tissue differential splicing (DS) patterns are species-specific in vertebrates, the alternative exon inclusion events in vertebrate brains are highly conserved (Madgwick et al., 2015). This suggests the existence of a core set of conserved functions for AS across vertebrate nervous systems. However, little is known about the *in vivo* functions of the SFs that are responsible for these conserved splicing events or the functions of the individual AS events that are controlled by these factors.

Data from RNASeq and cross-linking and immunoprecipitation (CLIP) experiments have been combined to create RNA splicing maps. These maps correlate the RBP binding sites with their effect on AS regulation (Witten and Ule, 2011). These genome-wide maps suggested that PTBP1/2 and RBFOX proteins antagonistically modulate the NPC-to-neuron transition by regulating neuron-specific exon inclusion. RBFOX binding in the intron downstream from alternatively spliced exons increases exon inclusion, whilst upstream PTBP1 binding has been shown to prevent exon inclusion (see below, Weyn-Vanhentenryck et al., 2014; Li et al., 2015). PTBP1 expression is high in nonneuronal and neural stem cells, where it promotes exon skipping in its neuronal paralogue PTBP2/nPTB. The resulting introduction of PTC into *Ptbp2* transcripts results in its degradation via NMD (Spellman et al., 2007). During differentiation Ptbp1 is downregulated by neuron-specific microRNA miR-124, which allows for Ptbp2 upregulation at later stages in development (Makeyev et al., 2007; Zhang et al., 2016b)). Rbfox1/2/3 in turn reach highest expression levels in differentiated neurons ((Makeyev et al., 2007; Zhang et al., 2016b). Ptbp2 is also subject to autoregulation and cross-regulation by RBFOX2 via NMD (Jangi et al., 2014).

RBFOX proteins are exceptional in their ability to recognize a long well-defined motif (U)GCAUG (Jin et al., 2003), while most other splicing factors recognize short (~3–7 nt) and degenerate sequence motifs, which occur frequently in pre-mRNAs. This limits the ability of motif-based bioinformatic target prediction tools to achieve both high specificity and sensitivity.

RBPs have developed different mechanisms to improve their binding specificity. Many RBPs utilise their modular structures to bind RNA with multiple RNA-binding domains (RBDs). For example, neuron-specific Nova binds to clusters of YCAY (Y = C/T). Its binding to exonic clusters blocks U1 snRNP binding and prevents exon inclusion, whereas Nova binding to an intronic YCAY cluster downstream of the regulated exon enhances spliceosome assembly and exon inclusion (Ule et al., 2006; Zhang et al., 2010).

Similarly to RBPs, many short exons of 3 to 27 nt, known as microexons, are predominantly or exclusively expressed in the nervous system. Differential inclusion of neuron-specific microexons is the most highly conserved splicing event during neural development, with inclusion increasing as differentiation progresses (Irimia et al., 2014; Torres-Méndez et al., 2019). The neuron-specific splicing factor nSR100/SRRM4 regulates inclusion of most mammalian neural microexons (Irimia et al., 2014). While the functional significance for the majority of these microexons has yet to be demonstrated, the ones that have been described, including the N1-Src microexon, play critical roles in various aspects of neuronal development, such as neurite elongation, axon morphogenesis and guidance, and neurogenesis (Kotani et al., 2007; Leung et al., 2015; Keenan et al., 2017; Lewis et al., 2017).

1.3: The Src family of non-receptor kinases

N1-src is a neuronal specific isoform of the Src proto-oncogene, the founding member of Src family kinases (SFKs) (Mustelin, 1994). The 11 members of the SFK family include Src, Fyn, Yes, Blk, Yrk, Frk, Fgr, Hck, Lck, Srm, and Lyn (Parsons and Parsons, 2004). SFKs have roles in the regulation of key signaling pathways involved in cell fate specification, cancer and development. Most cells in vertebrate organisms express at least one SFK, with some expressing multiple isoforms of the same protein (Thomas & Brugge, 1997). SFKs interact with many cellular cytosolic, nuclear and membrane proteins, modifying these proteins by phosphorylation of tyrosine residues. This multifunctionality stems from the conserved modular structure that is shared by all family members (Figure 2). SFKs comprise six distinct functional regions: the myristoylated N-terminal membrane-targeting Src homology 4 (SH4) domain, the SH3, SH2 and tyrosine kinase (SH1) domains and a short C-terminal tail (Engen et al., 2008). SH3 and SH2 domains bind proteins at polyproline peptide (PXXP) motifs and phosphotyrosine motifs respectively (Arold et al., n.d.). Potential substrates containing these motifs may bind, relocate and activate Src. The C-terminal tail of Src contains a conserved tyrosine residue that acts as an autoinhibitory phosphorylation site (Y527), bound by its SH2 domain when phosphorylated. The adjacent kinase domain contains an activating autophosphorylation site (Y416).

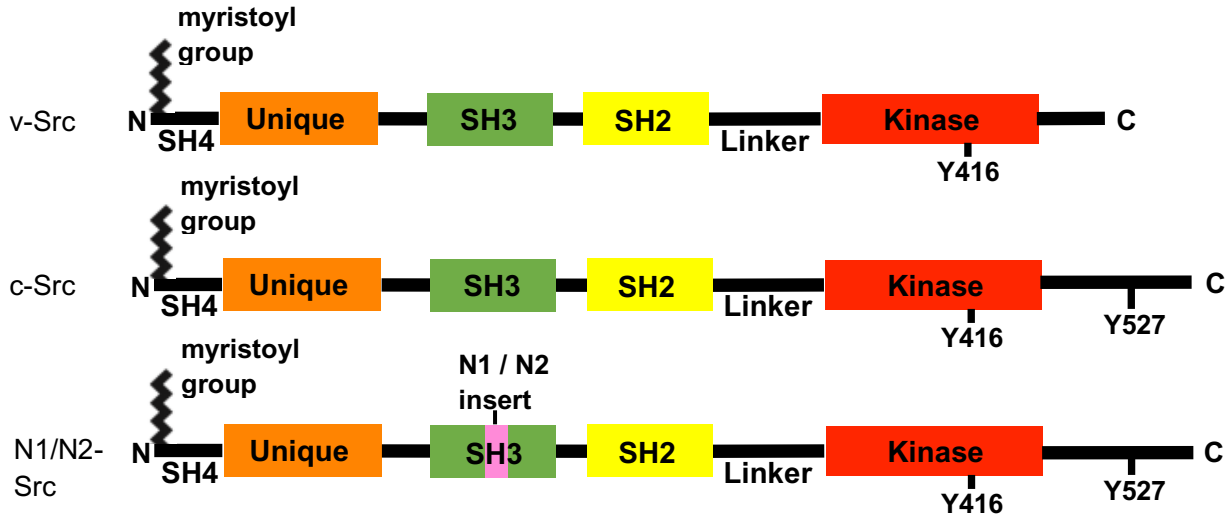


Figure 2. Domain structure of v-Src, c-Src, and Neuronal Src proteins.
The conserved Src-homology domains of Src family kinases are depicted, together with the positions of regulatory phosphorylation sites. v-Src has a shorter C-terminal domain that lacks the autoinhibitory phosphorylation site (Y527). Adapted from (Wetherill, 2016)

The ligand-binding surfaces of SH3 and SH2 not only confer substrate specificity, but are also involved in low-affinity intramolecular interactions that regulate the transitions of the protein between active and inactive states (Xu et al., 1999). Situated between the SH4 and SH3 domains is the 'unique' domain (UD), an intrinsically disordered region. Recent evidence shows that a fuzzy interaction between a UD of one Src partner in its open active conformation with a kinase domain of a second partner is required for Src dimerisation (Spassov et al., 2018). Dimerisation allows Src to bind to membranes much stronger by simultaneously inserting two myristoyl chains. Moreover, i enhances autophosphorylation of Y416 and consequently Src kinase activity. Therefore, shifting between monomeric and self-associated states provides another level of regulation of Src signalling (Le Roux et al., 2016).

1.4: Cellular Functions of C-Src

Due to functional redundancy between Src family members and their multiple isoforms, identification of the specific role of each Src family kinase is very difficult. However, C-Src is the most studied SFK with many known functions in different cellular processes including gene transcription, cell adhesion and migration, cell cycle progression and apoptosis. The involvement of C-Src in such a variety of pathways reflects a requirement for its intricate spatial and temporal regulation. Src's flexible protein domains in conjunction with myristoylation can determine its subcellular localization by mediating attachment to different cellular membranes. In its resting state, C-Src is associated with endosomal membranes near the perinuclear microtubule-organizing center (Kaplan et al., 1992), and is delivered to the cytoplasmic side of the plasma membrane upon stimulation (Sandilands et al., 2004). At the plasma membrane,

Src is involved in receptor-induced signal transduction pathways. Receptor clustering or dimerization leads to activation and recruitment of Src to the receptor complexes where it phosphorylates the tyrosine residues within the receptor cytoplasmic domains. For example, phosphorylation of epidermal growth factor (EGF) receptor clathrin heavy chain is required for its internalisation (Wilde et al., 1999). Subcellular localisation of Src can affect its function. C-Src activity causes RhoA inhibition at focal adhesion sites (Thomas and Brugge, 1997), but activation of the same protein at podosomes (Berdeaux et al., 2004). Within the nucleus, Src is thought to help regulate the cell cycle and cell division by its interactions with other proteins. For example, Src phosphorylates Sam68 (Src-Associated substrate in Mitosis of 68 kDa) during mitosis thereby changing binding ability and specificity of Sam 68-RNA interactions (Taylor et al., 1995). This prepares Sam68 for interaction with G1-specific messages and promotes exit from mitosis (Pillay and Nakano, 1996). The activation of C-Src leads to the promotion of survival, proliferation, adhesion and invasion pathways, like the RAS-MAPK and PIK3-Akt signalling pathways. Since uncontrolled growth is a necessary step for the tumour development and progression; mutations that result in increased activity or overexpression of C-Src have been implicated in a number of human cancers (Dehm and Bonham, 2004; Ishizawar and Parsons, 2004). *c-Src* is a proto-oncogene, a normal gene that becomes an oncogene due to aberrant expression, whereas its viral homologue encoded by Rous sarcoma virus, *v-Src* was the first discovered oncogene (Martin 2001). It lacks the C-terminal inhibitory phosphorylation site (Y527), and is therefore constitutively active as opposed to C-Src, which is only activated under certain circumstances where it is required (Figure 2, Smart et al., 1981). Before the discovery of the *Src* gene in chickens, cancer was thought to only be caused by foreign agents, viruses.

1.5: C-Src Functions in the Brain

Although expressed ubiquitously, maximal c-Src expression levels are 5–200-fold higher in platelets, neurons and osteoclasts than in other tissues. Early studies of Src expression in the developing rat brain revealed that its peak expression and activity correspond to neurogenesis and neuronal growth (Cartwright et al., 1988). Since then C-Src has been implicated in multiple processes in neuronal development. Discovery of alpha and beta-tubulin as major Src substrates in growth cone membranes provided evidence for its role in neurite outgrowth (Matten et al., 1990). C-Src function in the brain is also regulated by receptor-induced signal transduction. Cell adhesion receptors, in particular L1-CAM, control C-Src activity during neurite outgrowth (Beggs et al., 1994; Ignelzi et al., 1994)

1.6: Neuronal Srcs

Studies on the expression of C-Src in rat and chick embryos had indicated that neural tissues contain elevated levels of a form of a Src protein that was structurally distinct from C-Src expressed in non-neuronal cell cultures (Brugge et al., 1987). It was later revealed that two neuronal splice variants of Src, N1- and N2-Src, exist in mammals and birds, which contain a 6 and 17 amino acid insert in their SH3 domains respectively (Levy and Brugge 1989; Pyper and Bolen, 1990). The 6 aa insert (RKVDVR) in the N1-Src protein results from a single microexon inclusion between exons 3 and 4 of *c-Src*, which is repressed in nonneuronal cells by PTBP1 binding to the repressor elements in the N1 microexon 3′ splice site and in the downstream intron (Figure 3, Chan and Black, 1997; Chou et al., 2000). PTBP1 blocks the interaction of the U1 snRNP with U2AF and, thus, prevents assembly of U2AF at the downstream 3′ splice site (Sharma et al. 2005). In addition to the PTBP-binding sites that surround the N1 exon, there is a strong UGCAUG enhancer in the downstream intron that is recognised by RBFOX proteins and promotes N1 inclusion. There are multiple additional *cis*-elements that affect N1 splicing, including binding motifs for hnRNP A1 in the exon and for hnRNP H in the downstream intron (Chou et al., 1999; Rooke et al., 2003).

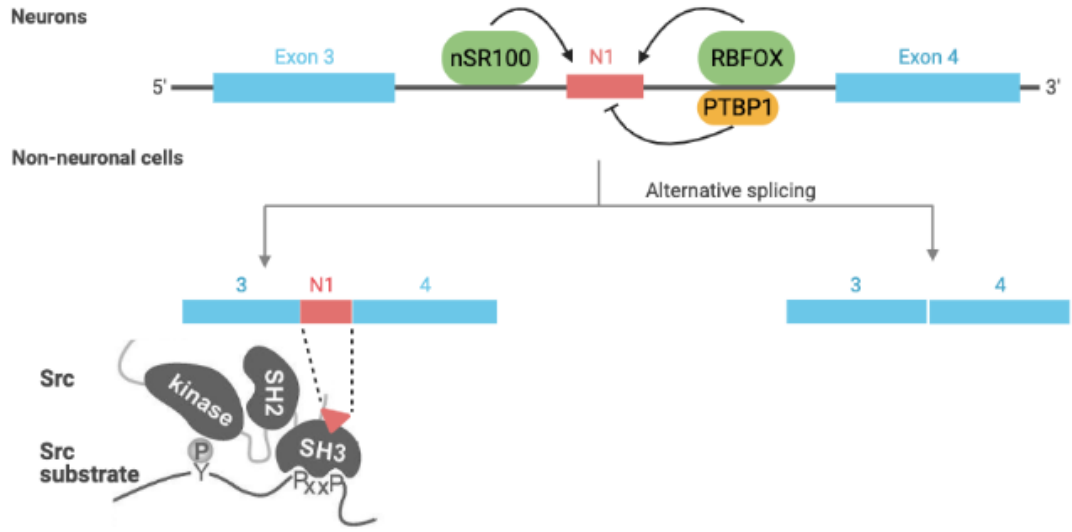


Figure 3. Regulation of alternative splicing of the Src N1 microexon.
The N1 microexon is repressed in non-neuronal cells by polypyrimidine tract-binding protein (PTB), which binds to repressor elements in the N1 3′ splice site and in the downstream intron, blocking the assembly of the spliceosomal complex. RBFOX proteins bind to enhancer elements in the intron downstream from N1 to stimulate its splicing. In neurons the N1 exon codes for a six amino acid insert in the n-Src loop of the substrate binding SH3 domain. Adapted from (Keenan et al., 2015).

The residues encoded by the N1 microexon affect SH3 domain substrate specificity of N1-Src (Keenan et al 2015). Furthermore, N1-Src has an enhanced constitutive kinase activity due to a loss of an intramolecular interaction of the SH3 domain with the SH2-kinase linker

(Brugge et al. 1987; Martinez et al. 1987; Keenan et al. 2015). N1-Src is highly active during neural development with its levels showing a 8-to 20 fold increase in an embryonal carcinoma (EC) cell line after treatment with retinoic acid (RA), which drives differentiation of EC cells into neuron-like cells that grow neurite-like processes and express neuronal markers (Lynch et al., 1986). Expression of constitutively active N1-Src in a transgenic mouse under the control of the L7 Purkinje cell promoter, affected the assembly of microtubules, causing aberrant dendritic morphogenesis (Kotani et al., 2007). Interestingly, both knockdown and overexpression of N1-Src inhibited neurite elongation in cultured neurons (Keenan et al., 2017). The potential mechanism underlying this effect could be due to N1-Src increased catalytic activity which could lead to aberrant substrate phosphorylation and formation of promiscuous ligand interactions upon overexpression. Investigations by (Wetherill, 2016; Keenan et al., 2017) into the role of N1-Src in neurite outgrowth suggested that it acts in L1-CAM and RhoA signalling pathways. The data implied that constitutive activation of RhoA prevented N1-Src mediated process elongation, however N1-Src did not promote process outgrowth via the inhibition of RhoA. (Wetherill, 2016) proposed that N1-Src overexpression enhances RhoA activation, therefore self-regulating via a negative feedback loop.

1.7: N1-Src and neuroblastoma

Despite Src having known roles in promoting cancer development and progression, N1-Src has been implicated as a positive prognostic indicator in neuroblastoma, a childhood cancer caused by the failure of certain neural crest cells to differentiate along the parasympathetic lineage (Tomolonis et al., 2018). It exhibits a high level of clinical and genetic heterogeneity with the course of disease ranging from spontaneous regression without any required medical intervention to treatment-resistant tumour progression, metastasis and death (Brodeur, 2003). Despite extensive sequencing efforts, only a few recurrent somatic mutations have been identified (Chen et al., 2015).

All patients with metastatic/Stage 4 disease diagnosed after one year of age or those with an amplification of the *MYCN* gene are classified as high-risk. High-risk neuroblastoma patients have only a 40% likelihood of survival (Maris and Matthay, 1999). They are typically treated with multimodal therapy, including high-dose chemotherapy, surgical resection, radiation therapy, bone marrow transplantation, and retinoic acid (RA). RA has been shown to induce growth arrest and differentiation of human neuroblastoma cells (Hämmerle et al., 2013; Janesick et al., 2015). It is used to prevent relapse, however the responsiveness to RA therapy is variable and unpredictable with many cases of neuroblastoma recurrence (Matthay et al., 2009).

A high N-Src to C-Src ratio is a positive prognostic marker in neuroblastoma with N-Src being highly expressed in infant cases, which tend to have a good prognosis (Bjelfman et al., 1990). Additionally, *N1-Src* mRNA is highly expressed in neuroblastoma cell lines with the ability to differentiate but not in the cell lines lacking the capacity to mature in response to RA, irrespective of *MYCN* gene amplification and overexpression (Matsunaga et al., 1993). Exon array and CHIP-Seq analysis identified an alternative splicing programme in Stage 4 neuroblastoma tumours mediated by MYCN binding to promoter regions of the splicing factors PTBP1 and HNRNPA1, leading to their overexpression and poor survival in high risk patients (Guo et al., 2011; Zhang et al., 2016a). Splicing inhibitors that are being developed as anti-cancer agents might represent an excellent therapeutic for neuroblastoma by mimicking the events that occur in normal differentiation of the neural crest cells. Interestingly, data from (Lewis et al., 2017) suggests that N1-Src is expressed at low levels in adult *Xenopus* heart tissue. N1-Src has also been found in back muscle cells of chick embryos at early stages of development.(Atsumi et al., 1993) This might be due to neural crest cells that migrate from the neural tube to the heart and back to form cardiac and striated muscle tissue there.

1.8: A role for N1-Src in *Xenopus* primary neurogenesis

Xenopus is one of the major model systems for the study of vertebrate embryogenesis. There are multiple advantages to the use of *Xenopus* as an experimental system, such as the availability of large abundant eggs that are easily manipulated and the striking synteny between the frog and human genomes (Amin et al. 2014). Even though neuron- specific splicing appears to be a feature of higher animals, frogs also have a neuronal splice variant of C-Src but it has only a 5 amino acid insert instead of the 6 amino acids seen in other organisms. The distribution of charged and hydrophobic residues contained within the insert is retained and it has been shown that the activity of mammalian and *Xenopus* n1-src is conserved.

In frogs, as in other lower vertebrates, an initial wave of neurogenesis gives rise to a simple pattern of primary neurons (Henningfeld et al. 2007). This process of primary neurogenesis is used as a model for the study of mechanisms involved in neural cell fate decisions. During that process the balance in activity of proneural and neurogenic genes selects individual cells in the dorsal ectoderm for differentiation into neurons (Bertrand et al. 2002). Neurogenesis can be tracked by monitoring the expression of a neuronal-specific tubulin (*tubb2a*). *N1-Src* knockdown by splice site-blocking morpholinos in *X.tropicalis* embryos maintains C-Src expression but inhibits primary neurogenesis, as indicated by a reduction in *tubb2a*-positive neurons (Lewis et al., 2017). This suggests that apart from regulating the activity of proteins involved in neuronal architecture, adhesion and morphogenesis, N1-Src has an earlier role in regulating the expression of genes essential for the transition of vertebrate neural progenitors to differentiated neurons. This is supported by the fact that *n1-src* expression is highly

regulated throughout early frog development, unlike C-src, whose expression is relatively constant during that time (Collett and Steele 1993). Unpublished proteomic and phosphoproteomic data from the Evans lab (Lewis et al., 2017) indicates that N1-Src interacts with and phosphorylates a subset of splicing factor proteins.

1.9: Src and splicing during neurogenesis

Early functional studies indicated that Src regulates pre-mRNA processing by tyrosine phosphorylation (Gondran and Dautry, 1999). Phosphoproteomic mass spectrometry and transcriptomic RNASeq data obtained in the Evans and Isaacs labs support those findings and suggest that N1-Src regulates a programme of alternative splicing during neuronal differentiation. Interestingly, preliminary *in silico* analysis suggests that putative tyrosine phosphorylation sites are specifically enriched within *trans*-regulatory factors as compared to the core splicing machinery. Taken together, Src phosphorylation is likely to impact the recruitment of the spliceosome and splice site selection, rather than the general splicing process.

1.10: Aims

As stated above normal neural differentiation is subject to a tightly controlled splicing programme, which if disrupted can lead to neurodevelopmental disorders and cancer. Preliminary data from the Evans and Isaacs labs have led to the hypothesis that N1-Src regulates splicing in neurogenesis and this is at the level of the splicing of splicing factor genes. The overarching aim of this study was to test this hypothesis through the bioinformatic analysis of publicly available human and frog transcriptomic datasets. These datasets were subjected to a variety of approaches, including analysis of alternative splicing, functional clustering of gene sets, classification of splicing events and seeking mechanistic insight by searching for RNA binding protein motifs at splice junctions. The specific aims were:

i) Discover if Src is a conserved regulator of the splicing of splicing factors. This was addressed by analysing differential alternative splicing in an RNAseq dataset from an inducible v-Src expressing cell line.

ii) Establish if the splicing of splicing factor genes is a feature of normal neuronal development. Here, various RNAseq time-series from dissected tissues during normal *Xenopus* embryo development were subjected to alternative splicing analysis.

iii) Gain insight into how N1-Src regulates splicing during neurogenesis. Finally, our preliminary RNAseq dataset from N1-Src knockdown *Xenopus* embryos was analysed to discover if perturbed splicing events correlate with those identified in aims i) and ii).

Chapter 2: Methods

2.1: RNA-seq datasets and genome assemblies

All publicly available RNA-Seq datasets used in the current study are listed in Table 1.
Datasets used to address aims:
i) Ji et al. 2019 described the gene regulatory network and the transcriptional changes during transformation in a MCF10-ER-Src cell line model of breast cancer. A single-end library was constructed followed by RNA-Seq using the Illumina HiSeq 2000 system.
ii) Session et al. have used high throughput sequencing to study the evolution of gene expression in X.laevis. mRNA for sequencing was extracted at 14 developmental stages and poly(A) enriched for mRNA in duplicate. For each timepoint, high throughput sequencing was performed using Illumina HiSeq 2000 with paired-end library preparation. Peshkin et al. 2015 measured mRNA levels across 18 developmental stages to characterise the relationship between coding RNA and protein dynamics during X.laevis differentiation. Two separate methods of mRNA extractions were used: poly(A) enrichment and ribosomal RNA (rRNA) depletion, therefore producing no technical replicates. This study used HiSeq 1000 for sequencing. Similarly to Session et al, Tan et al. 2013 used poly(A) selected mRNA to prepare a paired-end library followed by sequencing with HiSeq 2000 to characterize the transcriptome dynamics during *X.tropicalis* development. Using a paired-end mRNA library and the Illumina HiSeq 2000, Plouhinec et al., 2017 provided RNA sequencing data profiling the transcriptomes of *X.laevis* ectodermal domains at 2 stages of embryo differentiation.
iii) A.Pizzey of the Evans and Isaacs labs used short-read RNASeq of mRNA from N1-Src knockdown and control *X.tropicalis* embryos to investigate the role of N1-Src in vertebrate neurogenesis. For the Illumina RNA-Seq library, the NEBNext Poly(A) mRNA Magnetic Isolation Module was used to isolate Poly(A) mRNA from total RNA. After a paired-end cDNA library construction, transcripts were sequenced on an Illumina HiSeq 3000 machine.

GENCODE hg38 transcript annotations were used for transcript quantification in the MCF10-ER-Src cell line. Ensembl Xenopus_tropicalis_v9.1 and UCSC Xenopus_laevis_v2 assemblies and gene annotations were used for the Xenopus dataset analysis.

The list of putative c-Src and N1-Src phosphorylation substrates was collated from the results of bioinformatic analysis by James Ormond (unpublished MSc dissertation) and LC-MS/MS phosphoproteomic data (West, 2019).

2.2: Data preprocessing

All fastqc files downloaded from Gene Expression Omnibus (GEO) using axel were quality checked with FastQC. TrimGalore (Martin, 2011) was used to remove adaptors and low-quality reads. Adapter contamination could cause the reads that share maximal exact matches to fail to align within the required score threshold. Trimmed reads were then either aligned to the genome using HISAT2 (Pertea et al., 2016) or passed straight to Salmon (Patro et al., 2017) for transcript quantification using the following code.:

```
for i in *_1.fastq.gz
do
  prefix=$(basename $i _1.fastq.gz)
  salmon quant -i salmon_index --libType A -1 ${prefix}_1.fastq.gz -2 ${prefix}_2.fastq.gz -o quant/${prefix} -p
15 --validateMappings;
done
```

The --validateMappings flag is used for the selective alignment approach. The whole reference transcriptome and genome were used as a decoy sequence for the index used in quantification in mapping-based mode. This was done to avoid reads coming from a novel locus that are similar with annotated transcripts from being false mapped to the reference. k-mers of length 31 were used for all datasets, except the MCF10-ER-Src, where k-mers of length 29 were used, due to reads being 50bp long instead of 100.

For the *X.laevis* datasets, the trimmed and quality checked reads were aligned to the genome with HISAT2 and then the output SAM files were converted to BAM, sorted by their genomic location and indexed with SAMtools (Li et al. 2009). These BAM files were visualised using IGV (Robinson et al. 2011) and passed to StringTie, which assembles and quantifies the transcripts in each sample. The assembled transcripts and the reference gene annotation GTF files were merged together by the StringTie --merge module, which created a non-redundant uniform set of transcripts for all samples. The newly assembled transcriptome was then used as a reference for Salmon transcript quantification. In the case of human and *X.tropicalis* files, the alignment and transcriptome assembly steps were skipped as the transcriptome is well-annotated.

2.3: Differential splicing analysis

All the datasets were analysed with SUPPA (Trincado et al. 2018) psiPerEvent module to obtain the percent-spliced-in (PSI) value for each AS event. PSI represents the fraction of transcripts that have the exon inclusion. It is calculated as the ratio between the reads supporting the inclusion of the exon and the total number of reads attributed to that event, that is the sum of reads supporting inclusion and reads supporting exclusion.

$$PSI = \frac{inclusion\ reads}{exclusion\ reads + inclusion\ reads} \times 100$$

Each type of event was defined by a unique set of genomic coordinates around the splice site. For exon skipping events, the start and end coordinates for the different exonic regions involved in the event; the external coordinates of the event were only used for the intron retention and alternative first exon events.

The SUPPA diffSplice module was then used for DS analysis. The chromosome coordinates of differential AS events considered significant at p <= 0.1, |dPSI < 0.1| were then extracted with R.

VAST-TOOLS (Tapial et al. 2017) was then used for the human and X.tropicalis datasets. Following genome alignment with the vast-tools align module, the output files from the MCF10-ER-Src dataset were pulled together with the vast-tools merge module to compensate for low read coverage. The diff module was then used for differential AS analysis. VAST-TOOLS uses bowtie to align reads and provides both cRPKM and PSI values. Vertebrate Alternative Splicing and Transcription Database (VastDB) that is used for DS quantification does not offer annotations for X.laevis so this tool could not be used for this species data analysis.

The heatmaps were constructed using Heatmapper, an interactive web-based tool (Babicki et al., 2016). Z-scores (the number of standard deviations by which the value of a raw PSI is above or below the mean PSI) were plotted instead of normalized PSI values and are calculated with:

$$\mathrm{Z} = \frac{\mathrm{x} - \mu}{\sigma}$$

Z	=	Z-score
x	=	observed value
μ	=	mean of the sample
σ	=	standard deviation of the sample

The Z-scores are computed after the clustering, so that it only affects the graphical aesthetics. For the RNA processing heatmaps annotated genes within the GO term class “RNA processing” (GO:0006396) were considered as “regulators of RNA processing”. “RNA processing” contains the “RNA splicing” class within itself. It was chosen due to some RBPs having proposed but unconfirmed roles in RNA splicing.

2.4: Motif enrichment analysis

The SUPPA significant splicing change file was split by event type. Chromosome coordinates of sequences that were 150 nt upstream and downstream of splice sites involved in each type of splicing event were extracted using one of the scripts provided as part of the MoSEA

pipeline (Trincado et al., 2018). bedtools (Quinlan and Hall, 2010) was used to convert those coordinates to a FASTA file. Motif enrichment analysis of the sequences was then performed using AME (McLeay and Bailey 2010; MEME Suite), with statistical analysis by one-tailed Fisher's exact test. AME identifies known user-provided motifs that are enriched in the sequences of interest compared with control sequences. A motif was considered significantly enriched at E-value < 10, p <= 0.05. Shuffled input sequences were used as a control as it preserves the GC content and k-mer frequency. The MEME feature 'fasta-unique-names' was used to append the duplicate number "_i" after any duplicate sequence names that arose, to account for two separate events of the same class occurring in one gene. Position weight matrices of the motifs from the CISBP-RNA (Catalog of Inferred Sequence Binding Preferences of RNA binding proteins) database (Ray et al., 2013) were used for this project. Matt (Gohr and Irimia, 2019) was used to produce motif RNA maps for all CISBP-RNA IUPAC binding motifs with a sliding window of length 31 nt, which slides up to position 35 nt into exons and up to position 135 nt into introns. Events were discarded if they did not pass the VAST-TOOLS VLOW threshold of ≥20/15/10 actual reads mapping to the sum of exclusion splice junctions. Below is the code used for implementing Matt:

```
matt get_vast $vts_file -complex IR,IR-S,IR-C -a Tam -b Ctrl -minqab VLOW -minqglob N -gtf $gtf -f gene_id > gencode_IR.tab
matt get_vast $vts_file -complex S,C1,C2,C3,MIC -a Tam -b Ctrl -minqab VLOW -minqglob N -gtf $gtf -f gene_id > gencode_SE.tab
matt def_cats gencode_SE.tab GROUP 'silenced=DPSI_GRPA_MINUS_GRPB[15,100]' 'enhanced=DPSI_GRPA_MINUS_GRPB[-100,-15]' 'unregulated=DPSI_GRPA_MINUS_GRPB[-1,1] PSI_Ctrl[10,90]' | matt add_cols gencode_SE.tab -
matt rna_maps_cisbp gencode_SE.tab UPSTRM_EX_BORDER START END DOSTRM_EX_BORDER SCAFFOLD STRAND GROUP[silenced,enhanced,unregulated] 31 35 135 $fasta cisbprna_regexps -d $output_dir
```

The Src-regulated TRA2A intron retention event chromosome coordinates pulled from the VAST-TOOLS diff output were inserted into oRNAment (Benoit Bouvrette et al., 2020) for interactive visualisation of RBP binding motif instances.

2.5: Functional enrichment analysis

Functional enrichment analysis was performed according to Reimand et al., 2016 using the functional enrichment analysis tools g:Profiler (https://biit.cs.ut.ee/gprofiler) and GSEA (https://www.gsea-msigdb.org/gsea/). Structured controlled vocabularies from Gene Ontology, as well as information from the curated KEGG and Reactome databases were included in the analysis. Only functional categories with more than three members and fewer than 800 members were included in the analysis. Significance was assessed using the hypergeometric test with multiple testing correction by the Benjamini and Hochberg method. The Cytoscape plug-in Enrichment map (http://baderlab.org/Software/EnrichmentMap)

(Isserlin et al., 2014) was used to visualize and arrange functional data. Genes selected for Gene Ontology enrichment analysis were those appearing to be showing significant alternative splicing changes according to at least two of the four tools used.

2.6: Event clustering

The SUPPA clustering module was used to calculate the clusters of events according to PSI values across conditions. SUPPA uses density-based spatial clustering (DBSCAN; (Pedregosa et al., 2011) which clusters events that might not have similar PSI values within the same samples, but behave similarly across conditions.

2.7: Differential gene expression analysis

Differential gene expression was quantified using Sleuth version 0.29.0. Estimates of transcript abundances (Salmon) were normalized to gene length for gene level analysis.

2.8: Summary of software and algorithms used in this study

AME	(McLeay and Bailey, 2010)	https://meme-suite.org/meme/tools/ame
TrimGalore	-	http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
g:Profiler	(Reimand et al., 2016)	https://biit.cs.ut.ee/gprofiler
Cytoscape – Enrichment map	(Isserlin et al., 2014)	http://baderlab.org/Software/EnrichmentMap
HISAT2	(Kim et al., 2019)	http://daehwankimlab.github.io/hisat2/
FastQC	(Andrews and Others, 2010)	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Salmon	(Patro et al., 2017)	https://combine-lab.github.io/salmon/
Sleuth	(Pimentel et al., 2016a)	https://pachterlab.github.io/sleuth/
SUPPA2	(Trincado et al., 2018)	https://github.com/comprna/SUPPA
3DRNASeq	(Guo et al., 2019)	https://3drnaseq.hutton.ac.uk/app_direct/3DRNAseq/
GSEA	(Subramanian et al., 2005)	https://www.gsea-msigdb.org/gsea/
axel	-	https://github.com/axel-download-accelerator/axel
VAST-TOOLS	(Tapial et al., 2017)	https://github.com/vastgroup/vast-tools
Matt	(Gohr and Irimia, 2019)	http://matt.crg.eu/

Heatmapper	(Babicki et al., 2016)	http://www.heatmapper.ca/
IGV	(Robinson et al. 2011)	

Chapter 3: Results

The central hypothesis of this study arose from unpublished work by Alastair Pizzey in the Evans and Isaacs labs in which short and long-read RNA-Seq analysis was performed on Stage 14 *X. Tropicalis* embryos injected with control or N1-Src antisense morpholinos (described in Pizzey's PhD thesis, unpublished). Preliminary differential splicing analysis using DEXseq (Anders et al. 2012) identified perturbed splicing in splicing factor genes in N1-Src knockdown *X.tropicalis* embryos. Bioinformatic analysis by James Ormond (described in Ormond's MSc dissertation, unpublished) suggests that multiple splicing factors contain putative Src tyrosine phosphorylation sites, which interestingly are specifically enriched within *trans*-regulatory factors as compared to core splicing machinery. Furthermore, proteomic and phosphoproteomic analyses in the Evans lab identified splicing factors as SH3 domain binding partners and substrates of N1-Src (West 2019).

To gain insight into whether N1-Src regulates splicing factor gene splicing in normal neuronal development and the mechanism responsible, several relevant publicly available RNA-seq datasets were identified. These datasets (Table 1) were selected to address the role of Src in regulating splicing and the role of splicing factor gene splicing in neuronal development and differentiation.

Table 1: Questions addressed and the datasets used in this thesis

Aim	Model system	RNASeq Data	Source
1. Is Src a conserved regulator of alternative splicing?	Human cancer cell line (MCF10A-ER-Src) where v-Src expression is induced	Control vs Tamoxifen (4 replicates)	GEO accession: GSE115598 (Ji et al., 2019)
2. Is splicing an important aspect of normal *Xenopus* neural development and differentiation?	Whole *X. laevis* embryos at different developmental stages	Stages 10, 13-14, 20 (3 technical and biological replicates)	GEO accession: GSE73430; GSE73905 (Session et al., 2016); (Peshkin et al., 2015)
	Whole *X.tropicalis* embryos at different developmental stages	Stages 10, 14-15 (2 biological replicates)	GEO accession: GSE37452 (Tan et al., 2013)
	X. laevis ectodermal sections at different developmental stages	Stages 12.5, 14 and 17 (3 biological replicates)	GEO accession: GSE103240 (Plouhinec et al., 2017)

3. How does N1-Src regulate splicing during neurogenesis?	*X.tropicalis* Stage 14 embryos where N1-Src is knocked down	Control MO vs N1-Src AMO (Stage 16)	(Pizzey, PhD thesis, unpublished) (Lewis et al., 2017)

3.1 Developing a data analysis pipeline to process RNA-Seq datasets for alternative splicing

RNA-Seq data analysis started with the development of a pipeline (Figure 4) that could provide useful insights into differential splicing changes happening between stages in a developmental time series or between conditions in a human cell line. The pipeline included quality control and pre-processing of the raw data, followed by transcript quantification and finally, DS analysis and RBP motif enrichment.

As mentioned above, this work has been motivated by findings from experiments in *X.tropicalis*. However, in order to provide enough replicates for the statistical analysis that is carried out as part of multiple steps of this pipeline, data from a related tetraploid species, X.laevis, was used to address some of the questions. Due to *X.laevis* being a less well-annotated organism, additional steps involving genome alignments and guided transcriptome assembly were added. Aligning to a reference transcriptome rather than to a genome is usually faster and requires less computational power. However, it fails to recognise unannotated novel transcripts or intron retention, which is relevant in the case of studying alternative splicing in an organism with a less well-annotated transcriptome. Therefore, trimmed and quality checked *X.laevis* reads were aligned to the genome with a fast splice-aware aligner HISAT2. Prior to processing these newly aligned reads with Stringtie, they were indexed and sorted with SAMtools and the resulting BAM files were loaded onto IGV to be visualised. Stringtie was then used to produce a reference guided transcriptome assembly. Gene abundance files from Stringtie turned out to not be compatible with most of the downstream AS and gene expression analysis tools. Therefore, to avoid an additional format adjusting step, Stringtie was only used for *X.laevis* transcriptome assembly and Salmon was used for all transcript quantification. This also led to the use of a consistent transcript quantification method for all datasets.

3.1.1. Differential splicing analysis

Multiple DS tools were considered and then selected for the analysis after confirming previous observations that showed that different tools perform significantly differently across different datasets or numbers of samples (Mehmood et al., 2019). There are currently three

major categories of methods used by splicing analysis tools: exon-based, isoform-based or event-based. Exon- and event-based methods fall into one overarching category of event-based methods, that calculate the number of sequencing reads falling on each counting unit, which can be exons or exon junctions. The isoform-based methods reconstruct the full-length transcripts before AS analysis and are useful for biological interpretation when the focus is on specific genes with known functional implications of changes in expression of particular isoforms. However, after running preliminary analyses using different types of tools, event-based methods were favoured. Their output is easier to interpret as they report on types of AS event and the aim of this project was to look at general trends in splicing events rather than effects on specific transcripts.

The first event-based method selected was SUPPA (Trincado et al., 2018), which quantifies splicing events themselves by calculating the PSI values for each event, which measures the fraction of mRNAs expressed from a gene that contains that event. SUPPA uses transcript abundances determined with Salmon to estimate the PSI values for each of the four standard types of splicing events: alternative 3' splice site, alternative 5′ splice site; intron retention, and exon skipping, in addition to two other event types, alternative first exon (AF) and alternative last exon (AL). It can only perform pairwise differential AS analysis, however it can quantify the PSI values in multiple conditions, which is useful for time-series data. It also has a much lower false positive rate than other DS methods, especially at shorter read lengths.

The second tool selected was 3DRNA-seq (Guo et al., 2019), an R Shiny app selected for its user-friendly interface and use of Salmon transcript abundance files as input. 3D RNASeq performs differential AS analysis with *limma* (Ritchie et al., 2015), an exon-based method that fits a linear model to the exon-level counts from Salmon and then tests for differential exon usage. Finally, the exon-level statistics are converted to gene-level test statistics to identify DS genes. The limitation of this approach is that unlike SUPPA, it does not infer the type of the splicing event occurring in a gene but only identifies the differentially expressed exons/ transcripts between experimental conditions. limma can also be used for differential gene expression analysis. A custom Python script was used to extract gene and transcript IDs from gene annotation gtf files for 3DRNASeq analysis.

A few other DS tools were considered, including DICESeq (Huang and Sanguinetti, 2016) which uses HISAT BAM alignment files to estimate the changes in isoform proportions from time series RNA-seq experiments. This tool was the perfect candidate for developmental stages analysis, however potentially due to the code being inefficient, the script would run for days only to be terminated for an unknown reason.

Finally, VAST-tools (Tapial et al., 2017) that was initially disregarded for only being compatible with data from a small number of model organisms. It was re-evaluated due to its ability to detect microexon splicing and compatibility of its output with Matt, a toolkit for

downstream analyses of AS which includes creation of RBP maps. The VAST-tools microexon module uses exon-exon and exon-microexon-exon junctions to specifically quantify short exons (3-15 nucleotides). The VAST-tools *diff* module (Tapial et al., 2017) allows testing for differential AS based on replicates and read depth for each event. It therefore allowed the use of the similarly disregarded *X.tropicalis* duplicate developmental series dataset. It also allowed for a more sensitive detection of AS events in the MCF10-ER-Src dataset, where reads could be merged together to compensate for low read depth. None of the tools identified Src transcripts as being significantly spliced in the N1 exon knockdown dataset, potentially due to the exon being only 15 nucleotides long. However, the single splicing event SUPPA2 recognised for Src after knockdown was exon skipping, but the p-value was 0.0519, making it a non-significant event. Therefore, a less stringent arbitrary p-value of 0.1 was selected for use with the less sensitive tools based on N1 exon skipping being a confirmed event.

3.1.2. Motif enrichment analysis

To explore the regulatory mechanisms involved in the developmentally and Src-regulated AS events, motif enrichment analysis was included in the analysis pipeline. Trincado et al. use their own motif scan and enrichment analysis (MoSEA) tool for the analysis of SUPPA2 output. Sections of scripts used by MoSEA for extraction of relevant chromosome coordinates were applied to SUPPA output split by event type, to then be used as input for AME motif enrichment analysis.

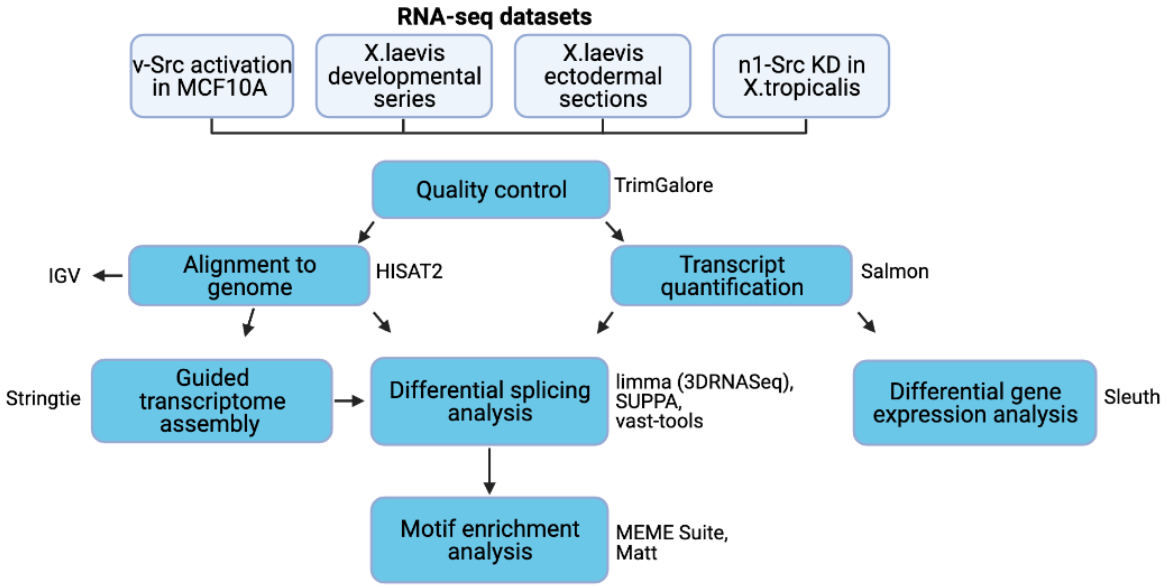


Figure 4. Data analysis pipeline. Flowchart of the optimised pipeline used to analyse RNA-Seq datasets in this study. Each blue box describes the type of analysis performed and the relevant software packages are listed beside them.

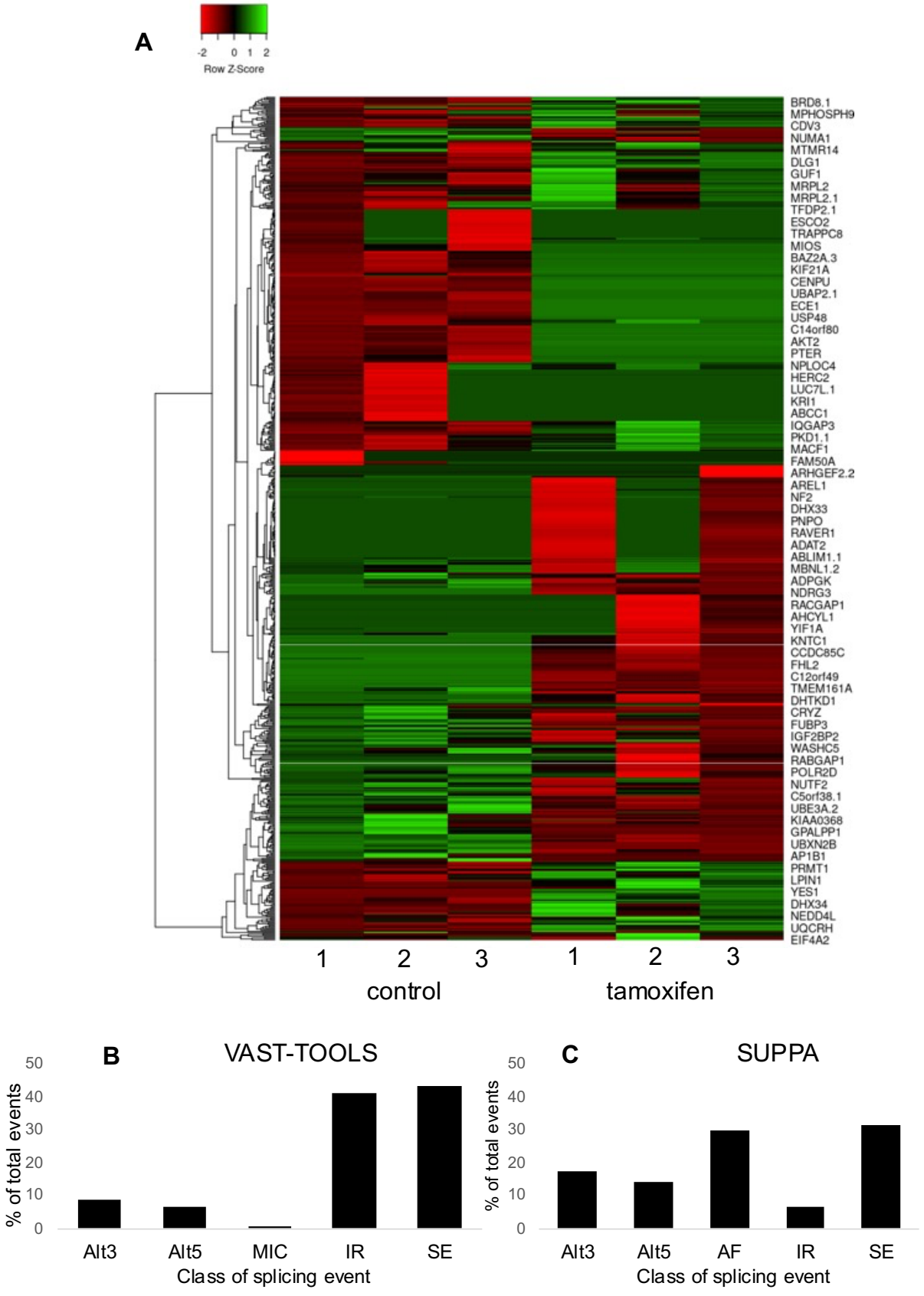
3.2: v-Src expression regulates a programme of alternative splicing in a human MCF10-ER-Src cell line

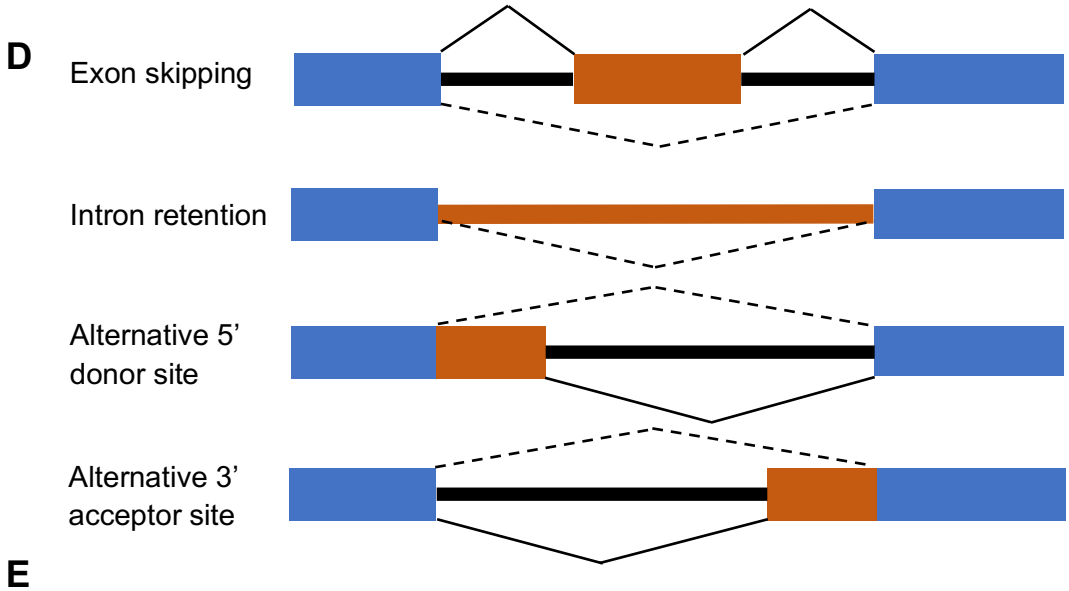
To determine if Src function in regulating alternative splicing is conserved in human cells, we selected expression data from a study where transient activation of viral Src (v-Src) was used to model oncogenesis (Ji et al., 2019). In this model, a non-tumorigenic human epithelial cell line (MCF10A) containing ER-Src, a derivative of v-Src that is fused to the ligand-binding domain of the estrogen receptor (ER), is treated with a partial ER agonist tamoxifen. Tamoxifen rapidly induces the transcriptional expression of the v-Src-encoding transgene, which in turn triggers a pro-inflammatory positive feedback loop that stably transforms the cell line and maintains it even after tamoxifen removal (Iliopoulos et al., 2009). The transforming abilities of v-Src come from it being constitutively active. The C-terminal domain of v-Src is truncated and lacks the regulatory Tyr527, the residue on the C-terminal tail that promotes autoinhibition upon its phosphorylation (Figure 2, Sefton and Hunter, 1986). Similarly to viral forms, neuronal splice variants of Src demonstrate increased catalytic activity compared to c-Src. Previous studies suggest that this is due to the neuronal exon insertion disrupting the interactions between the SH3 domain and the SH2 catalytic domain linker, which are required for Src autoinhibition (Arold et al., 2001; Brugge et al., 1987). Additionally, N1-Src activity is not affected by Tyr-527 phosphorylation (Brugge et al., 1987; Xu et al., 1999). This makes v-Src activation a good model for N1-Src kinase activity.

The RNA-Seq data was produced using mRNA extracted from MCF-10A-ER-Src cells 24 h after treatment with 1 μM 4-hydroxy-tamoxifen or its vehicle, ethanol, as a control. The authors discuss the large scale effect that v-Src activation has on the activity of pro-inflammatory transcription factor networks, chromatin structure and gene expression, however, they do not mention splicing. Unfortunately, during analysis of the data I discovered that the read depth was only really suitable for differential gene expression analysis and not splicing, as it would only be sensitive to changes in splicing of highly expressed genes. However, analysis with two AS algorithms vast-tools (Figure 5B) and SUPPA (Figure 5C) still revealed a range of splicing changes occurring as a result of Src activation. The vast-tools *diff* module is able to provide estimates if replicates are not available and therefore sample replicates were combined to compensate for low read depth. Four main AS patterns (Figure 5D), including alternative 3′ and 5′ splice sites (Alt3 and Alt5, respectively), exon skipping (SE) and intron retention (IR), were examined, with SUPPA also quantifying alternative first exons (AF) and VAST-TOOLS - microexons (MIC). Alternative 5’ splice sites, also known as alternative donor sites, change the 3' boundary of the upstream exon, whilst alternative 3' splice sites change the boundary of the downstream exon. Splicing events with a cut-off of ΔPSI (change in average PSI) > 10 and $p < 0.1$ for SUPPA or $p < 0.05$ for VAST-TOOLS

were considered statistically significant (Figure 5A). Previous findings suggested that N1-Src regulates alternative splicing of splicing regulators themselves. Similarly functional enrichment analysis of differentially spliced genes revealed that along with gene ontology terms relating to regulation of cell cycle, DNA repair and tumour suppressor protein 53 (TP53) regulated transcription which are relevant to cell transformation in cancer, RNA splicing and spliceosome were also two of most enriched terms (Figure 5E). The author attempted to investigate the correlation between the GO, KEGG, Reactome terms and the different clusters of exon skipping/ inclusion events in tamoxifen treated compared to control samples. However, due to being unable to extract useable lists of genes within the clusters, I was only able to produce a heatmap of differentially spliced exon events in genes falling in the GO term class “RNA processing” (GO:0006396, Supplementary figure1).
As mentioned above, *in silico* screening by J. Ormond of phosphotyrosine sites within splicing factors revealed that many of them are SFK phosphorylation sites. This allowed us to suggest which SF proteins could be putative substrates of Src phosphorylation that would subsequently alter the splice pattern of transcripts of these splice factors and other transcripts (Supplementary Table 1). From that list, *CLK2, DHX38, GEMIN5, KHSRP,*

PLRG1, PRPF18, RBM25 and SRSF11 were differentially spliced after v-Src expression was induced in MCF10-ER-Src cells.





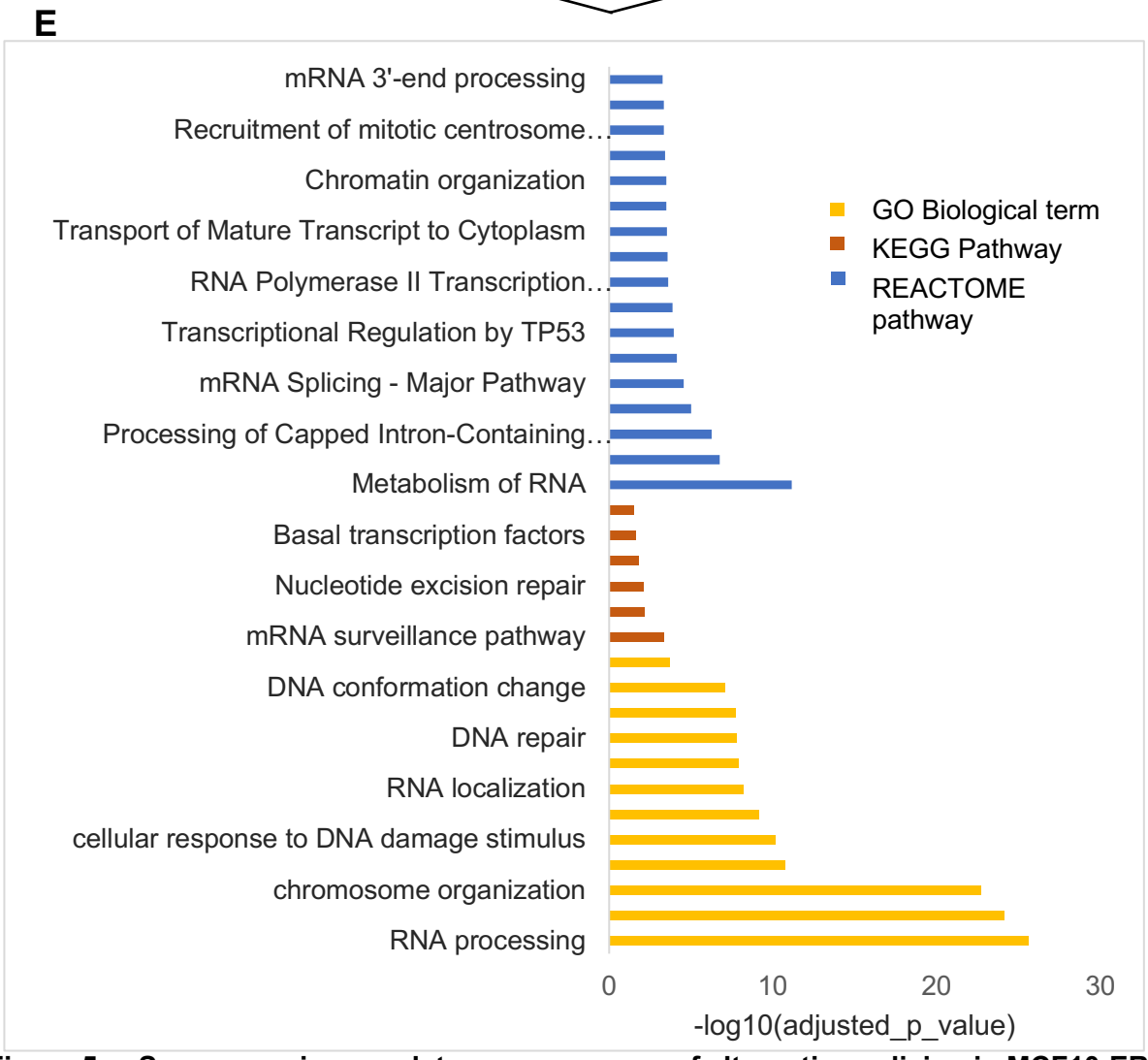


Figure 5. v-Src expression regulates a programme of alternative splicing in MCF10-ER-Src cells.

(**A)** Heatmap of differentially spliced exon events occurring between control and tamoxifen treated MCF10-ER-Src cells represented as Z-scores of percent-spliced-in(PSI) values, detected by VAST-TOOLS, with the corresponding hierarchical clustering. Due to low sequence coverage, reads from the GEO GSE115598 dataset were combined for differential

splicing analysis. **B & C.** Summary of classes of splicing events plotted as a % of total events according to VAST-TOOLS (**B**; 1761 total events) or SUPPA (**C**; 251 total events). Abbreviations: Alt3 - alternative 3' splice site, Alt5 - alternative 5′ splice site; IR – intron retention, SE – exon skipping, AF - alternative first exon, MIC - microexon inclusion. (**D**) Diagrams of the four main classes of AS events. Blue boxes, flanking constitutive exons; orange boxes, alternative spliced exons/regions; solid lines, splice junctions supporting the inclusion isoform; dotted lines, splice junctions supporting the exclusion isoform. (**E**) Functional enrichment analysis of genes differentially spliced upon v-Src activation in MCF10A-ER-Src cells.

One of the main aims of this project has been to describe the general trends in alternative splicing changes regulated by Src expression and make a link between the classes of regulated splicing events and the SFs that regulate them. Intron retention levels were hardest to evaluate due to different event-based tools having different approaches leading to varying strengths in identifying IR. Whilst some tools consider an intron retention event as significant as long as it is detected in one of the transcripts, SUPPA calculates the ratio of the abundance of transcripts that include the IR over the abundance of the transcripts that contain other forms of the splicing event. According to (Mehmood et al., 2019) despite SUPPA being able to detect the highest proportion of the qPCR-validated DS genes across the datasets compared to other tools, it only identified 8% of the qPCR-validated IR events, which is supported by the comparison with VAST-TOOLS performed in this study (Figure 5B,C). VAST-TOOLS reported intron retention to be the second most prevalent event type (722) in MCF10A-ER-Src cells after exon skipping (763), whilst SUPPA could only detect 8 intron retention events. According to VAST-TOOLS, transcripts of 6 of the aforementioned putative Src substrates underwent intron retention, including *CLK2, DHX38, KHSRP, PLRG1, RBM25* and *SRSF11*. VAST-TOOLS also discovered 3 microexon inclusion events.

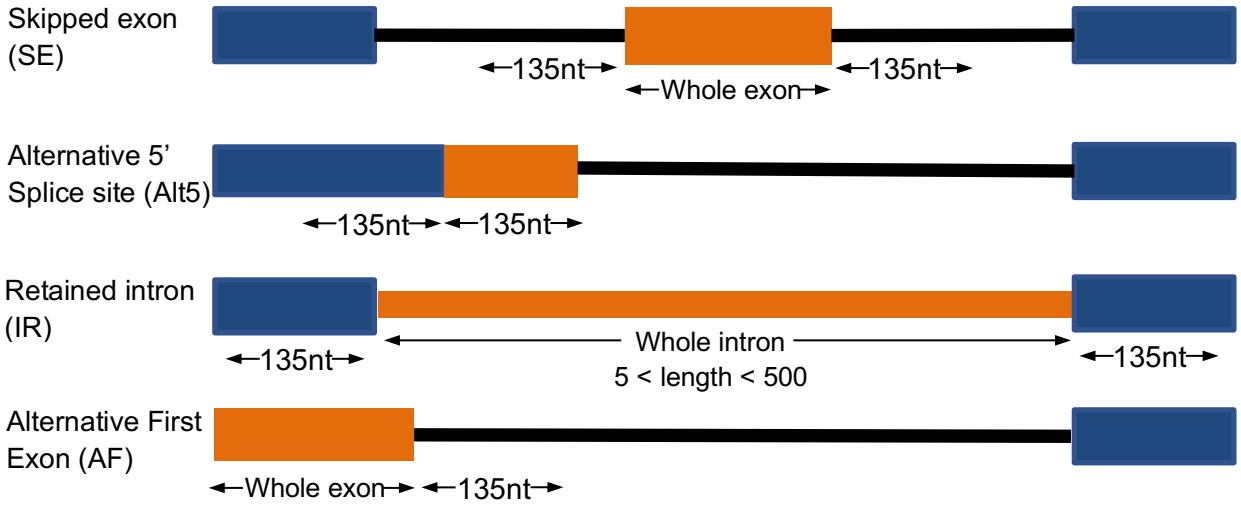


Figure 6. Diagram of splicing events highlights the sequence windows selected for the motif analysis with AME.

Exons are shown as boxes and introns as lines. The retained/skipped section is shown in orange (Adapted from https://github.com/comprna/MoSEA)

To investigate which potential SFs were regulated by v-Src expression, motif enrichment analysis of sequences surrounding the differential splicing events was carried out (Figure 6). To determine the factors that may preferentially regulate exon skipping events, motif enrichment analysis was carried out on the sequences surrounding the skipped exon junctions taking a 135 nt window into the neighbouring intron for MEME and Matt analysis (Table 2). MoSEA sequence selection script extracts the whole regulated exon sequence, whilst Matt only uses 35 nt on each end of the exon for its analysis. According to AME there were no significantly enriched RBP motifs in sequences relevant to intron retention and alternative 5' splice site events at E-value < 10 and p <= 0.05. This can be explained by low numbers of events discovered for these classes by SUPPA. The only putative Src substrate enriched in all events was SRSF1. Unfortunately, this SR protein is necessary for all splicing reactions to occur and so is likely to be enriched at all exon boundaries. However, SRSF1 can also influence splice selection and its regulatory impact of SRSF1 on splicing depends on its interaction with other splicing factors and its position in relation to the splice site (Anczuków et al. 2015). The results of the Matt analysis suggested that SRSF1 and SRSF9 binding to regions in the exon leads to an upregulation of exon silencing or skipping. Whereas, binding of NONO, another putative Src substrate, in the upstream intron leads to exon inclusion or (Figure 7).

Table 2. Consensus motifs of putative Src substrate SRSF1 are enriched at v-Src regulated splice junctions in a human cell line. RNA binding protein motif enrichment within a 135 nt window around the splice sites of differential splicing events occurring upon v-Src activation in MCF10A cells was calculated using AME (MEME Suite) one-tailed Fisher's exact test (E-value < 10, p <= 0.05). Putative Src substrates are indicated in bold. The adj-p-value is the optimal enrichment p-value of the motif according to Fisher's exact test, adjusted for multiple tests using a Bonferroni correction. The E-value is the adjusted p-value multiplied by the number of motifs in the motif file and represents the expected number of motifs that would be as enriched in the submitted sequences as this one. A total of 97 RBP binding motifs from the CISBP-RNA (DNA-encoded) database were examined.

Event	Motif RBP	Consensus	adj_p-value	E-value	motif_ID
SE	ELAVL2	HTYMTTTWTWTTY	2.03E-04	1.97E-02	M328_0.6
SE	IGF2BP3	AMAHWCA	9.74E-04	9.45E-02	M163_0.6
SE	HNRNPCL1	ATTTTTT	6.09E-03	5.91E-01	M158_0.6
SE	U2AF2	TTTTTYC	6.33E-03	6.14E-01	M077_0.6

SE	RBM6	MATCCAR	1.54E-02	1.49E+00	M161_0.6
SE	ELAVL1	TTWTTTT	1.80E-02	1.75E+00	M031_0.6
SE	RALY	TTTTTTG	3.05E-02	2.96E+00	M150_0.6
SE	CPEB4	YTTTTTT	3.25E-02	3.15E+00	M149_0.6
SE	PABPC4	AAAAAAA	3.38E-02	3.27E+00	M042_0.6
SE	TIA1	TTTTTTG	3.40E-02	3.30E+00	M075_0.6
SE	PABPC1	ARAAAAA	3.69E-02	3.58E+00	M146_0.6
SE	SART3	ARAAAAA	3.69E-02	3.58E+00	M062_0.6
SE	HNRNPC	ATTTTTK	3.99E-02	3.87E+00	M025_0.6
SE	KHDRBS2	RATAAAM	4.71E-02	4.57E+00	M176_0.6
SE	EIF4B	GTHGGAA	6.87E-02	6.66E+00	M290_0.6
SE	IGF2BP2	AMAWACA	6.93E-02	6.73E+00	M032_0.6
SE	FXR1	AAYGACRA	8.69E-02	8.43E+00	M152_0.6
AF	HNRNPL	ACACACA	9.32E-06	9.04E-04	M027_0.6
AF	IGF2BP3	AMAHWCA	9.37E-06	9.09E-04	M163_0.6
AF	IGF2BP2	AMAWACA	3.40E-05	3.30E-03	M032_0.6
AF	ELAVL2	HTYMTTTWTWTTY	5.55E-04	5.38E-02	M328_0.6
AF	SRSF4 (*D.melanogaster*)	GGAGGGV	8.65E-03	8.39E-01	M126_0.6
AF	**SRSF1**	GGAGGAV	1.11E-02	1.08E+00	M102_0.6
AF	Fusip1	AGAGAAM	1.39E-02	1.35E+00	M019_0.6
AF	PABPC4	AAAAAAA	1.43E-02	1.38E+00	M042_0.6
AF	KHDRBS3	GATAAACV	3.06E-02	2.96E+00	M033_0.6
AF	BRUNOL5	TGTGTGT	3.45E-02	3.34E+00	M157_0.6
AF	RBM3 (*M.musculus*)	GTGTGTG	3.98E-02	3.86E+00	M049_0.6
AF	HNRNPH2	GGGAGGG	5.97E-02	5.79E+00	M151_0.6
AF	FXR1	AAYGACRA	8.80E-02	8.54E+00	M152_0.6
Alt3	**SRSF1**	GGAGGAM	3.17E-02	3.07E+00	M102_0.6
Alt3	RBM5	GAAGGAA	4.52E-02	4.39E+00	M053_0.6

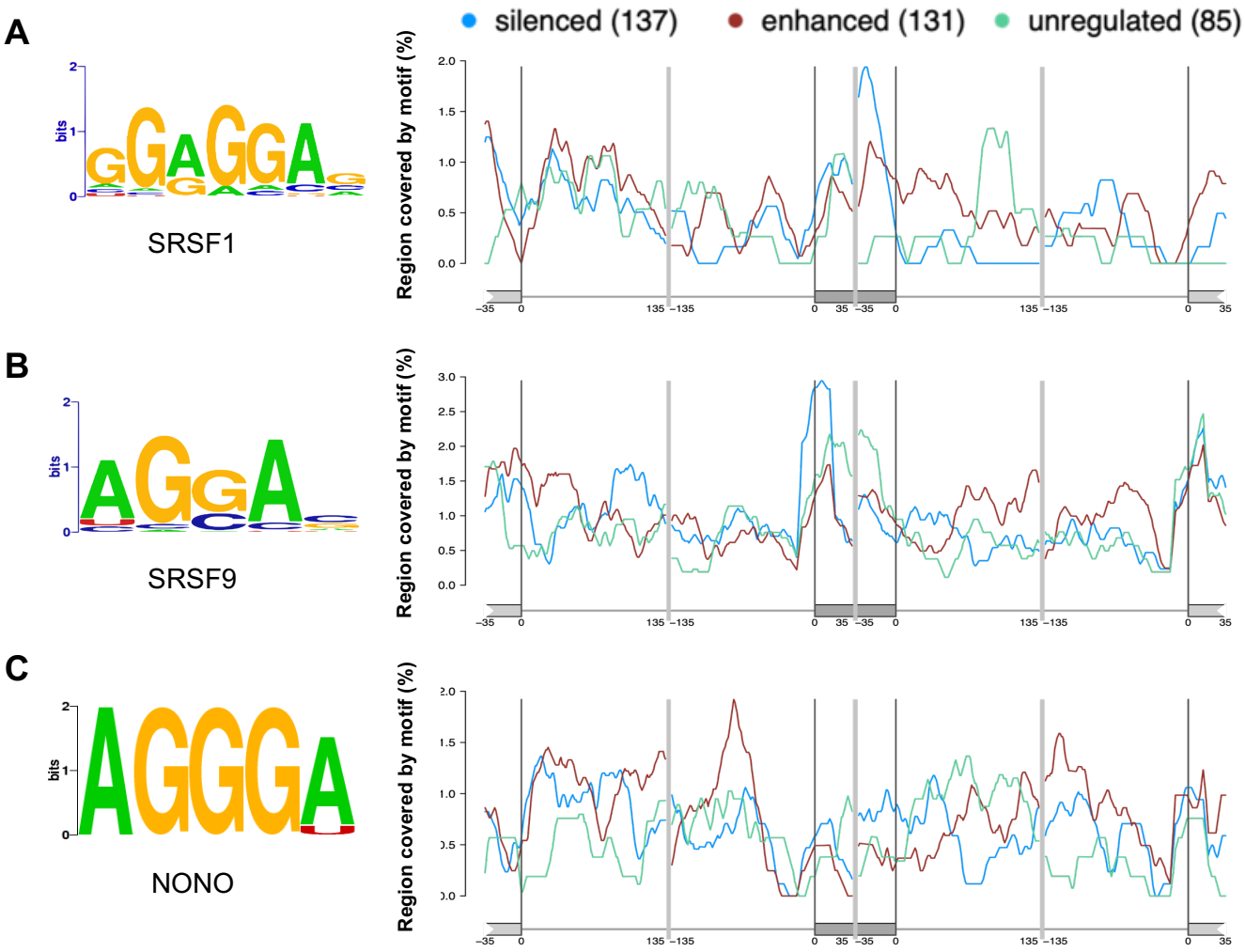


Figure 7. RNA splicing mapping reveals spatial enrichment of RNA binding motifs for SRSF1, SRSF9 and NONO across all exon skipping events.
Right: Motif RNA-maps represented as percent of region covered by the motif within a sliding window of length 31 nt, which slides up to position 35 nt into exons and up to position 135 nt into introns for SRSF1 (A), SRSF9 (B) and NONO (C) with corresponding binding motif sequence logos where the y-axis represents bits of information (left). Exons are indicated as dark grey boxes. A total of 353 exons that passed the vast-tools VLOW threshold of ≥20/15/10 actual reads mapping to the sum of exclusion splice junctions were considered, 167629 events were discarded. Exon events fall into categories: silenced (downregulated exon inclusion, ΔPSI ≥15), enhanced (upregulated exon inclusion, ΔPSI ≤−15) and unregulated exons (|ΔPSI| ≤1).

3.4: Splicing is an important feature of *Xenopus* early development.

As previously described in Section 1.2, changes in splicing profiles of cells across developmental stages play an important role at different points in vertebrate neural development. To confirm that these specific splicing changes are an important aspect of *Xenopus* primary neurogenesis, trends in splicing in whole *X.laevis* and *X.tropicalis* embryos

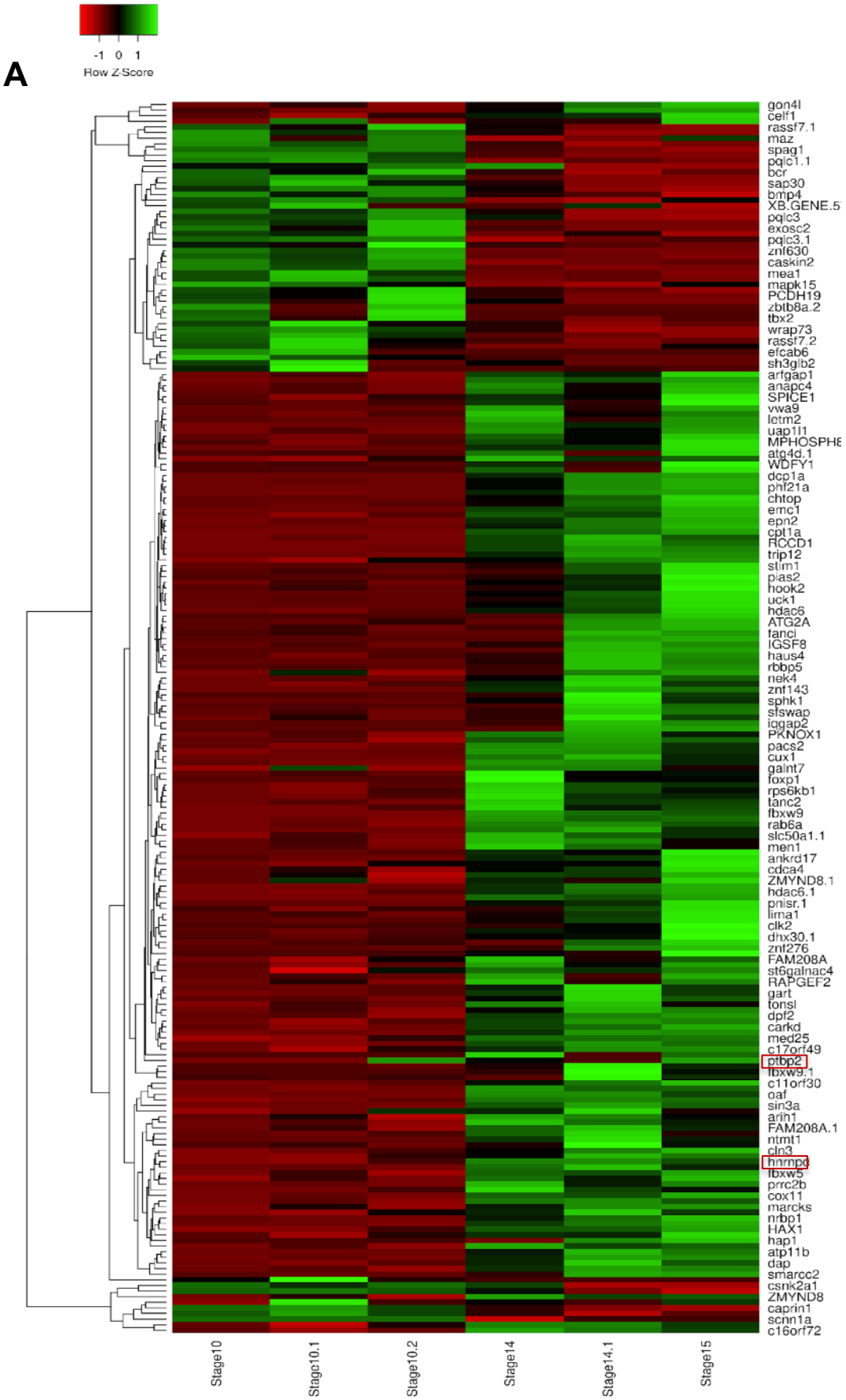
across three stages of development covering neurogenesis were investigated. The X.tropicalis developmental time-series (GSE37452) was selected initially due to previous research on N1-Src in the lab being carried out using this organism. However, the dataset only included two biological replicates per stage, which was not enough for the statistical analysis carried out by splicing tools to define significant differences. Hence, the project shifted towards *X.laevis*, which is a related species with a sequenced genome. Developmental time series datasets of the *X.laevis* RNA-Seq short reads were available in NCBI GEO database under accession numbers GSE73430 and GSE73905 (Peshkin et al., 2015; Session et al., 2016). The two datasets were combined to fulfil the need for triple repeats. In both datasets total RNA was collected from *X.laevis* J strain embryos according to the Nieuwkoop and Faber staging system. To allow for mRNA quantification, highly abundant ribosomal RNAs (rRNAs) must be removed from total RNA before sequencing. Peshkin et al., 2015 performed two distinct techniques to deal with rRNA: the first using poly(A) enrichment with oligo (dT) primers and the second using rRNA depletion, whilst Session et al., 2016 only used the poly(A) capture method. Sequencing was performed on Illumina HiSeq instruments producing 100-bp paired-end reads in both sets. In order to reduce technical variance between replicates, only the poly(A) set from GSE73905 was used. Stages 10, 15 and 25 were selected for analysis as they correspond to developmental events relevant to neurogenesis: gastrulation, start and end of neurulation. During neurula stages in *Xenopus* embryos, the ectodermal tissue undergoes its first differentiation into neural tissue, separating the neural plate from the ectoderm. This process of primary neurogenesis is regulated by N1-Src signalling (Lewis et al. 2017). Additionally, the datasets had the required number of RNA-Seq data replicates for these stages. Despite *X.laevis* being a model organism, it has a very poorly annotated transcriptome. This led me to create a new assembly using the genome assembly provided by NCBI, the splice-aware genomic alignment tool HISAT and the Stringtie assembler module. Upon learning that the *X.tropicalis* alternative splicing profile was included in the VastDB database, the *X.tropicalis* developmental data was analysed using VAST-tools.

Expectedly, results showed large scale changes in splicing with exon skipping and alternative first exon usage being the most commonly occurring types. Only 10 differential splicing events occurring between stages 10 and 14-15 were shared between X.laevis and X.tropicalis as reported by vast-tools and SUPPA. This might be due to variability in splicing event detection numbers between different splicing tools, but also due to whole embryo bulk RNA-Seq data being too noisy for the tools to be able to pick out the significant splicing events.

As outlined in section 1.1, intron retention is an important contributor to post-transcriptional regulation of gene expression and transcript diversity during normal differentiation and development (Pimentel et al., 2016b; Middleton et al., 2017). Splicing isoform diversity is highest in undifferentiated stem cells and decreases upon neural commitment and differentiation (Wu et al. 2010), and IR has been suggested to promote the elimination of non-functional or physiologically irrelevant transcripts and isoforms (Braunschweig et al., 2014). One consistent pattern that has emerged across many biological contexts is that regulatory IR particularly affects spliceosome components and splicing factors (Pimentel et al., 2016b; Jacob and Smith, 2017); Boutz *et al*. 2015). In most cases IR events in splicing factor pre-mRNAs act to down-regulate expression of those SFs. Splicing analysis of RNA-Seq data from whole X.tropicalis embryos suggests an increase in intron retention levels as X.tropicalis neurogenesis progresses, including in transcripts of splicing regulators like Hnrnpd and Rbfox2 (Figure 8A).

SUPPA was used to see whether IR events would cluster across *X.laevis* developmental stages 10,15,25 and whether the genes within those clusters would have roles in similar biological processes and pathways. For a maximum reachability distance of 0.15, I obtained three well-differentiated clusters (silhouette score = 0.387; Figure 8B; Supplementary table 2). Given the variability of events across stages, it is hard to find compact core clusters, therefore the minimum number of events per cluster was set to 5. Once the first cluster definition is overcome, it is easier to identify events in the neighbourhood. Functional enrichment analysis of the largest cluster (Cluster 1, 361 genes) revealed it to be enriched for GO terms relating to RNA splicing(GO:0008380), localization (GO:0006403) and transport(GO:0050658). Levels of intron retention in that cluster rose between stages 10 and 15 and fell again at stage 25. The other two clusters showed an opposite pattern wherein the levels of intron retention were lowest at stage 15, the bigger of the two clusters showed

enrichment for genes involved in stem cell differentiation (GO:0048863) and the PI3K/AKT signalling pathway (KEGG:04151).



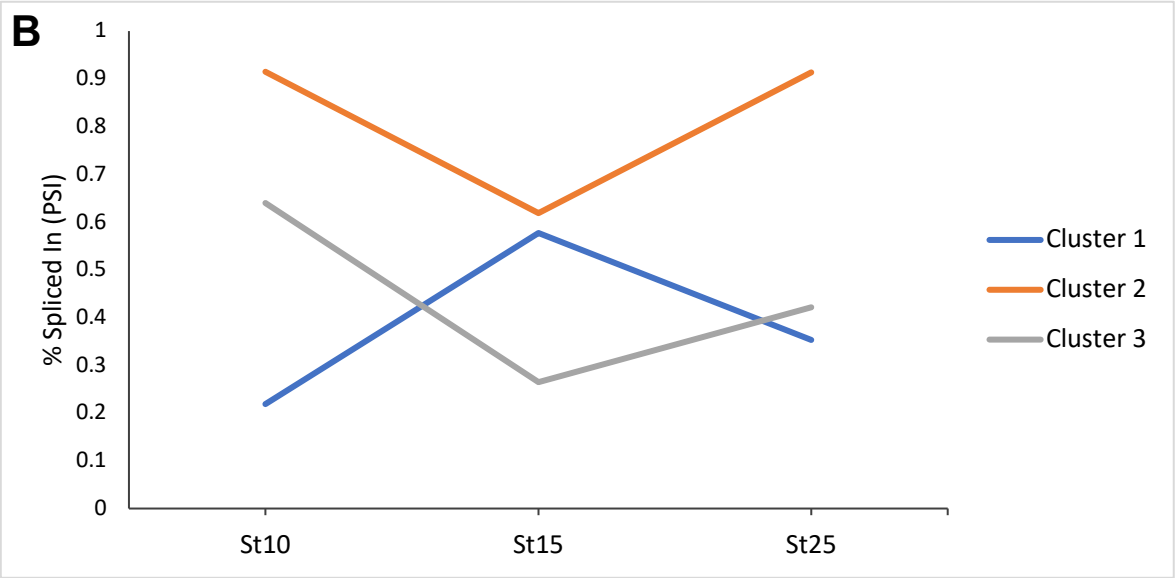


Figure 8. Intron retention levels in differentially spliced transcripts increase during X.tropicalis developmental stages relevant to neurogenesis.
(**A**) Heatmap of intron retention events represented as Z-scores of percent-spliced in (PSI) values, reported by VAST-TOOLS ($p < 0.1$, $|dPSI| > 10$), across X. tropicalis developmental stages 10,14 and 15 (GSE37452), with the corresponding hierarchical clustering. (**B**) The average PSI (*y-axis,* where 1 indicates 100%) per stage (*x-axis*) of the intron retention events in the three clusters obtained. Density-based clustering performed on the 510 regulated intron retention events that change splicing significantly in at least one comparison between adjacent steps across three differentiation stages (10,15,25).

3.3: Xenopus neural plate and crest development are regulated by alternative splicing

As mentioned previously, low read depth and heterogeneity of whole organism cell populations lead to variations in transcript expression between samples and conditions to be dominated by technical or biological noise. To minimise the effects of expression noise, as that observed in aforementioned whole embryo datasets, and focus on regions relevant to neurogenesis, transcriptomic data from defined cell populations of developing *X.laevis* ectoderm was selected for analysis (GSE103240; (Plouhinec et al., 2017). *X.laevis* embryos were dissected at Nieuwkoop and Faber stages 12.5, 14 and 17 which correspond respectively with the transition from gastrulation to neurulation, mid- and late neurula. 3 biological replicates were collected per dissected region for all three stages. RNA-Seq data from regions of lineage progenitors for neural plate (NP), neural crest (NC), which emerges from neural plate border (NB) cells, and non-neural ectoderm (NNE) was selected for this project. The structures were subdivided into anterior and posterior regions. The anterior NP forms the forebrain and the midbrain, whilst the posterior NP forms the hindbrain and the spinal cord. Only premigratory NC was dissected at stage 17 as it is the only specified region

at that stage. RNA-Seq libraries were sequenced on an Illumina HiSeq 2000 machine with a target of 15–20 million 100 bp paired reads per sample.

As discussed previously, precise spatio-temporal regulation of spliceosome machinery and SF expression is essential during neural development. The authors reported that the “mRNA splicing” GO term was enriched in 6 of the weighted gene co-expression network analysis groups. These are groups of genes that share similar temporal and spatial expression profiles.

Despite N1-Src knockdown causing failure of all neuronal types in the Xenopus primary nervous system to differentiate, NB and NC were regions of particular interest due to them being the supposed origin of neuroblastoma. As mentioned in section 1.7, overexpression of N-Srcs has been shown to differentiate neuroblastoma cell lines derived from metastatic high risk tumours and therefore our hypothesis is that N-Src regulated splicing might be a good target for neuroblastoma therapy.

Motif enrichment analysis of the 56 DS junction sequences suggests that putative Src-substrates SRSF1 and RBM47 are amongst regulators of IR during the development of neural crest in Xenopus between stages 14 and 17 (Table 3). 7 out of the 34 annotated genes that underwent IR are RNA-binding proteins and regulators of translation.

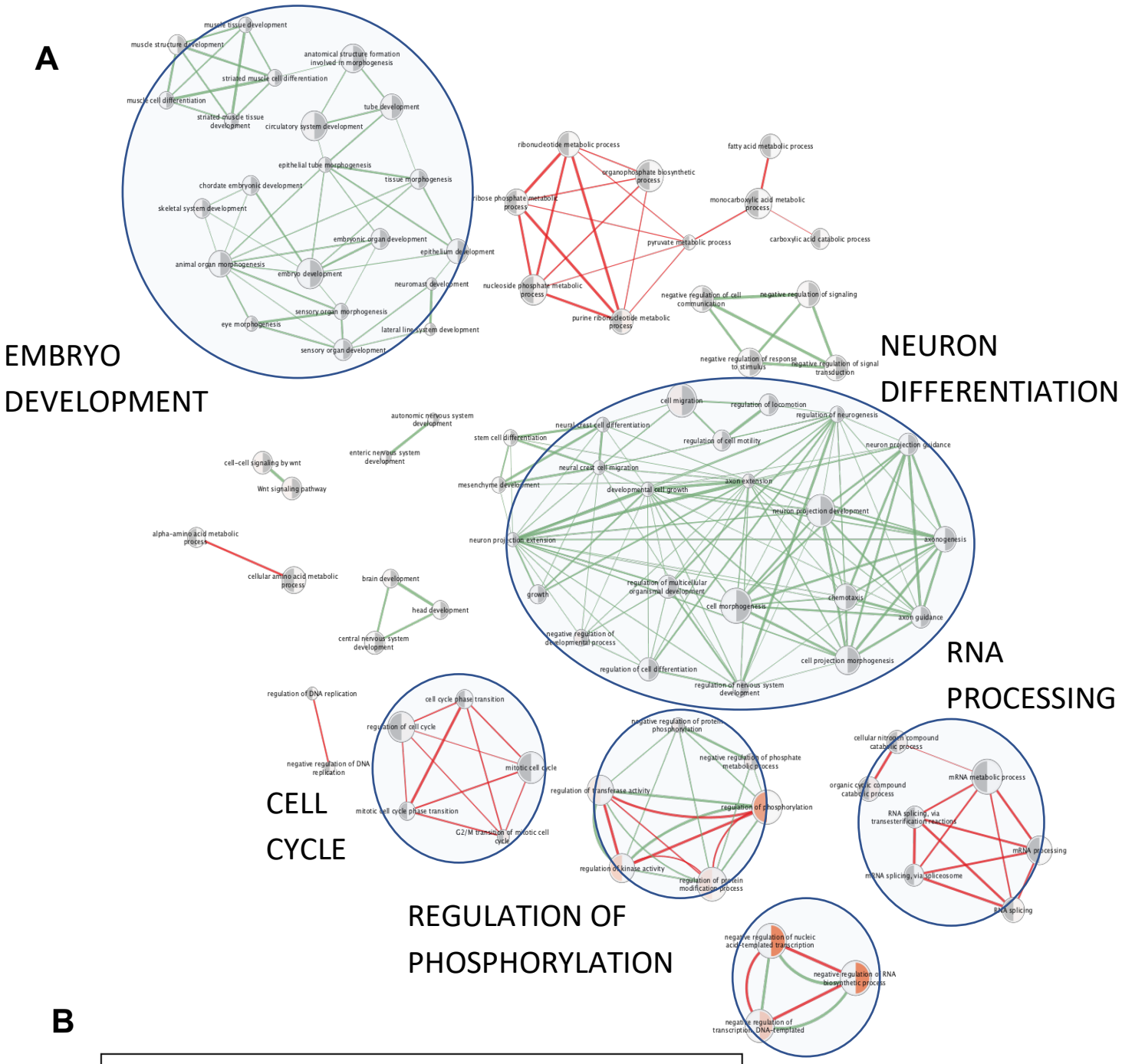
Table 3. Consensus motifs of putative Src substrates SRSF1 and RBM47 are enriched at intron retention splice junctions in developing *X.laevis* neural crest. RNA binding protein motif enrichment within a 135bp window around the splice sites of differential splicing events occurring during the transition of neural border at st.14 to neural crest at st17 was calculated using AME one-tailed Fisher’s exact test (E-value < 10, p <= 0.05). Putative Src substrates are indicated in bold. The adj-p-value represents the optimal enrichment p-value of the motif, adjusted for multiple tests using a Bonferroni correction. The E-value is the adj-p-value multiplied by the number of motifs in the motif file and represents the expected number of motifs that would be as enriched in the submitted sequences as this one.

Event	Motif RBP	consensus	adj_p-value	E-value
IR	CNOT4_00156	GACAGAN	9.20E-04	9.85E-02
IR	ENOX1_00149	MAGACAG	4.33E-03	4.63E-01
IR	**SRSF1**_00163	GGAGGAG	1.30E-02	1.39E+00
IR	PCBP3_00215	HTTTCCCT	2.09E-02	2.23E+00
IR	TRA2_00078	GAAGAAG	2.66E-02	2.85E+00

IR	**RBM47**_00279	GATGAWN	3.05E-02	3.26E+00
IR	RBM45_00241	GACGACM	3.81E-02	4.07E+00

Next transcriptomes and splicing profiles of anterior neural plate sections were analysed. Differential gene expression analysis with Sleuth was performed to confirm mRNA splicing as an important mechanism regulating central nervous system development. GO term enrichment analysis showed that genes that are differentially expressed in Xenopus neural plate development are linked to neuron differentiation, regulation of phosphorylation, transcription and RNA processing. Most of the differentially expressed genes involved in splicing were downregulated (Figure 9). Interestingly, 31 of the 153 downregulated SF genes are putative Src substrates, including SRSF1, SRSF9 and NONO. The downregulation of SF expression supports the observation that transcript diversity declines during differentiation and therefore splicing regulator expression falls allowing only the physiologically relevant transcripts to be successfully spliced.

SUPPA reported various differential splicing events occurring during *X.laevis* NP development, however many of them occurred in unannotated transcripts and genes of unknown function. Out of the 201 annotated transcripts, only 76 belonged to a gene with a described function. Two of these genes are RNA processing regulators Ptbp1 and Igf2bp3 (Figure 9C).


A
EMBRYO DEVELOPMENT
NEURON DIFFERENTIATION
RNA PROCESSING
CELL CYCLE
REGULATION OF PHOSPHORYLATION
REGULATION OF TRANSCRIPTION
muscle tissue development
muscle structure development
anatomical structure formation involved in morphogenesis
striated muscle cell differentiation
muscle cell differentiation
tube development
striated muscle tissue development
circulatory system development
epithelial tube morphogenesis
tissue morphogenesis
chordate embryonic development
skeletal system development
embryonic organ development
epithelium development
animal organ morphogenesis
embryo development
neuromast development
sensory organ morphogenesis
eye morphogenesis
lateral line system development
sensory organ development
ribonucleotide metabolic process
fatty acid metabolic process
organophosphate biosynthetic process
ribose phosphate metabolic process
monocarboxylic acid metabolic process
pyruvate metabolic process
carboxylic acid catabolic process
nucleoside phosphate metabolic process
purine ribonucleotide metabolic process
negative regulation of cell communication
negative regulation of signaling
negative regulation of response to stimulus
negative regulation of signal transduction
cell migration
regulation of locomotion
regulation of neurogenesis
autonomic nervous system development
neural crest cell differentiation
stem cell differentiation
regulation of cell motility
neuron projection guidance
enteric nervous system development
cell–cell signaling by wnt
neural crest cell migration
Wnt signaling pathway
mesenchyme development
axon extension
developmental cell growth
neuron projection development
alpha-amino acid metabolic process
neuron projection extension
axonogenesis
cellular amino acid metabolic process
brain development
regulation of multicellular organismal development
growth
chemotaxis
head development
cell morphogenesis
axon guidance
central nervous system development
negative regulation of developmental process
cell projection morphogenesis
regulation of cell differentiation
regulation of nervous system development
regulation of DNA replication
cell cycle phase transition
regulation of cell cycle
negative regulation of DNA replication
mitotic cell cycle
mitotic cell cycle phase transition
G2/M transition of mitotic cell cycle
negative regulation of protein phosphorylation
negative regulation of phosphate metabolic process
regulation of transferase activity
regulation of phosphorylation
regulation of kinase activity
regulation of protein modification process
cellular nitrogen compound catabolic process
mRNA metabolic process
organic cyclic compound catabolic process
RNA splicing, via transesterification reactions
mRNA processing
mRNA splicing, via spliceosome
RNA splicing
negative regulation of nucleic acid-templated transcription
negative regulation of RNA biosynthetic process
negative regulation of transcription, DNA-templated
B
-log10(p)
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
dPSI

C

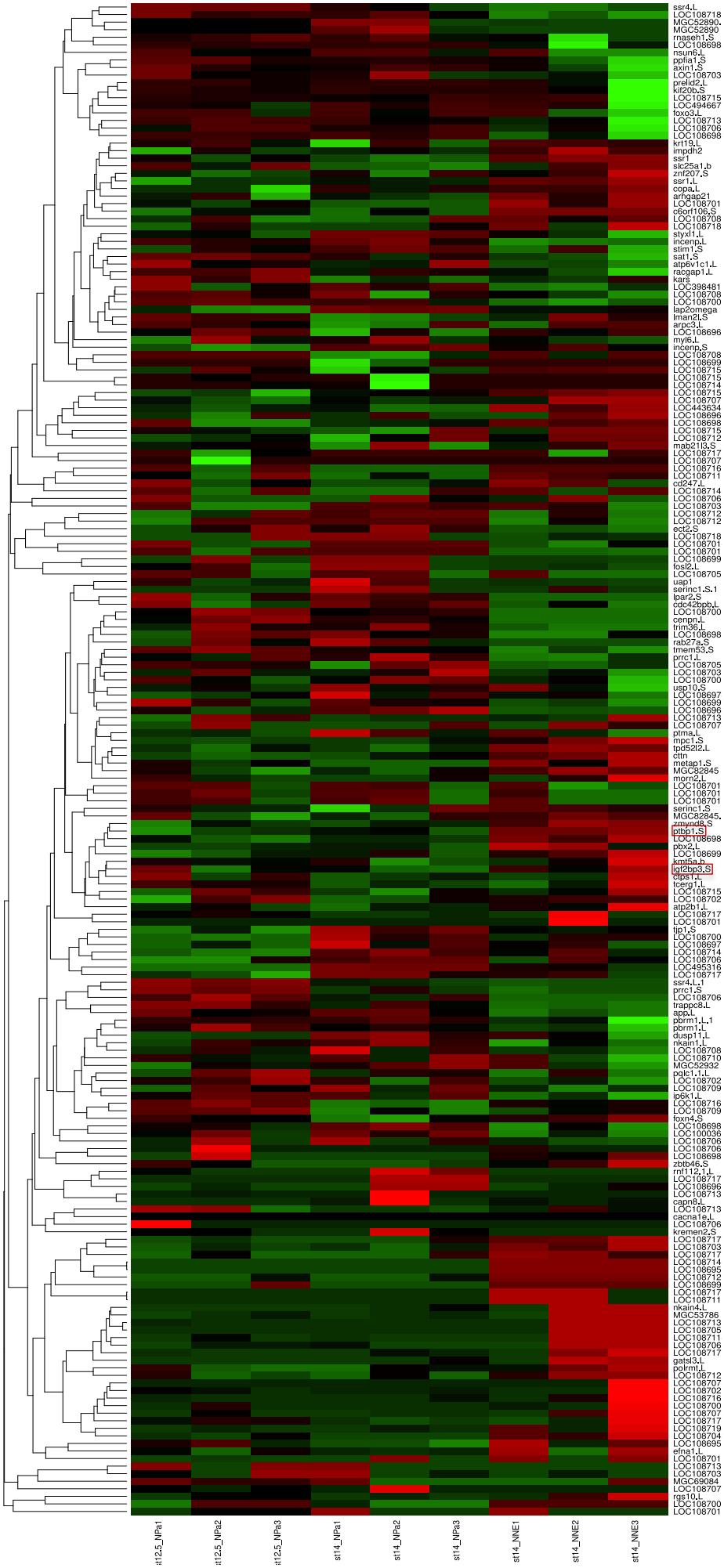

ssr4.L
LOC108718
MGC52890.
MGC52890
rnaseh1.S
LOC108698
nsun6.L
ppfia1.S
aixin1.S
LOC108703
prelid2.L
kif20b.S
LOC108715
LOC494667
foxo3.L
LOC108713
LOC108706
LOC108698
krt19.L
impdh2
ssr1
slc25a1.b
znf207.S
ssr1.L
copa.L
arhgap21
LOC108701
c6orf106.S
LOC108708
LOC108718
styxl1.L
incenp.L
stim1.S
sat1.S
atp6v1c1.L
racgap1.L
kars
LOC398481
LOC108708
LOC108700
lap2omega
lman2l.S
arpc3.L
LOC108696
myl6.L
incenp.S
LOC108708
LOC108699
LOC108715
LOC108715
LOC108714
LOC108715
LOC108707
LOC443634
LOC108696
LOC108698
LOC108715
LOC108712
mab21l3.S
LOC108717
LOC108707
LOC108716
LOC108711
cd247.L
LOC108714
LOC108706
LOC108703
LOC108712
LOC108712
ect2.S
LOC108718
LOC108701
LOC108701
LOC108699
fosl2.L
LOC108705
uap1
serinc1.S.1
lpar2.S
cdc42bpb.L
LOC108700
cenpn.L
trim36.L
LOC108698
rab27a.S
tmem53.S
prrc1.L
LOC108705
LOC108703
LOC108700
usp10.S
LOC108697
LOC108699
LOC108696
LOC108713
LOC108707
ptma.L
mpc1.S
tpd52l2.L
ctln
metap1.S
MGC82845
morn2.L
LOC108701
LOC108701
LOC108701
serinc1.S
MGC82845.
zmynd8.S
ptbp1.S
LOC108698
pbx2.L
LOC108699
kmt5a.b
igf2bp3.S
dbp5.L
tceg1.L
LOC108715
LOC108702
atp2b1.L
LOC108717
LOC108701
tjp1.S
LOC108700
LOC108697
LOC108714
LOC108706
LOC495316
LOC108717
ssr4.L.1
prrc1.S
LOC108706
trappc8.L
app.L
pbrm1.L.1
pbrm1.L
dusp11.L
nkain1.L
LOC108708
LOC108710
MGC52932
pqlc1.1.L
LOC108702
LOC108709
ip6k1.L
LOC108716
LOC108709
foxn4.S
LOC108698
LOC100036
LOC108706
LOC108706
LOC108698
zbtb46.S
rnf112.1.L
LOC108717
LOC108696
LOC108713
capn8.L
LOC108713
cacna1e.L
LOC108706
kremen2.S
LOC108717
LOC108703
LOC108717
LOC108714
LOC108695
LOC108712
LOC108699
LOC108717
LOC108711
nkain4.L
MGC53786
LOC108713
LOC108705
LOC108711
LOC108706
LOC108717
gata3.L
polrmt.L
LOC108712
LOC108707
LOC108702
LOC108716
LOC108700
LOC108707
LOC108717
LOC108719
LOC108704
LOC108695
efna1.L
LOC108701
LOC108713
LOC108703
MGC69084
LOC108707
rgs10.L
LOC108700
LOC108701
st12.5_NPa1
st12.5_NPa2
st12.5_NPa3
st14_NPa1
st14_NPa2
st14_NPa3
st14_NNE1
st14_NNE2
st14_NNE3

Figure 9: Xenopus neural plate development involves differential expression and splicing of genes coding for regulators of RNA processing.
(**A**) Analysis of Gene Ontology term enrichment among genes whose expression changes between stages 12.5 and 14-15 of *X.laevis* neural plate development. GO clusters enriched for downregulated genes are indicated in red, for upregulated genes in green. (**B**) Volcano plot highlighting differential exon skipping events occurring in X.laevis neural plate between stages 12.5 and 14-15, with events considered significant at $p<0.1$, $|dPSI| > 0.1$. (**C**) Heatmap of exon skipping events represented as Z-scores percent-spliced in (PSI) values, reported by VAST-TOOLS ($p < 0.1$, $|dPSI| > 10$), across *X. tropicalis* ectodermal tissues: anterior neural plate (NPa) and nonneural ectoderm (NNE) at stages 12.5 and 14 (GSE37452), with the corresponding hierarchical clustering.

3.4: N1-Src knockdown and v-Src expression cause changes in splicing of splicing factors TRA2A and HNRNPA1

Insights gained from the unpublished N1-Src knockdown data produced in the Evans lab have been the foundation of the current project. The knockdown in Xenopus embryos was achieved through the use of antisense morpholino oligos (MOs) that specifically bind to their selected target site to block access of cell components to that site. X.tropicalis embryos at the 1-2-cell stage were injected bilaterally with 20ng non-overlapping antisense MOs targeted to the splice acceptor and donor sites of the n1-src microexon or 20ng standard control MOs. This way c-Src expression was unaffected by the antisense MO injection. Four biological repeats were carried out and mRNA for short-read sequencing was harvested at stage 16. Bioinformatic analysis and experimental validation with PCR previously done by the Isaacs and Evans labs revealed that N1-Src knockdown causes intron retention in transcripts of key splicing factors TRA2A and HNRNPA1 (Fig 10). It provided more evidence for the hypothesis that N1-Src regulates the splicing of splicing factors during primary neurogenesis. These two PCR-validated splicing events have been used in the current project as guidance during choice of differential splicing analysis tools. According to analysis with VAST-TOOLS, v-Src expression in MCF10-ER-Src human cell line promotes alternative splicing in transcripts of *TRA2A* and *HNRNPA1*, further validating Src's role in their regulation. However, instead of having an opposite effect on intron retention, v-Src activation caused the same intron retention in *TRA2A* but an alternative 3' acceptor site in *HNRNPA1* transcript. The exon junctions at either side of the intron retention event in *TRA2A* contain SRSF1 and SRSF9 binding sites, along with a TRA2A binding site in the downstream exon (Figure 10C). This only implied that all of these putative Src substrates are key regulators and so most transcripts contain many of their binding consensus motifs at the exon junctions. A binding motif of tra2, a TRA2A orthologue in Drosophila, was enriched at exon junctions involved in intron retention events regulated during specification of neural crest in X.laevis embryos. This suggests an extra level to N1-Src

regulation of alternative splicing, whereby phosphorylation of one set of splicing factors promotes changes in splicing of a new set of SFs which then leads to global changes in alternative splicing.

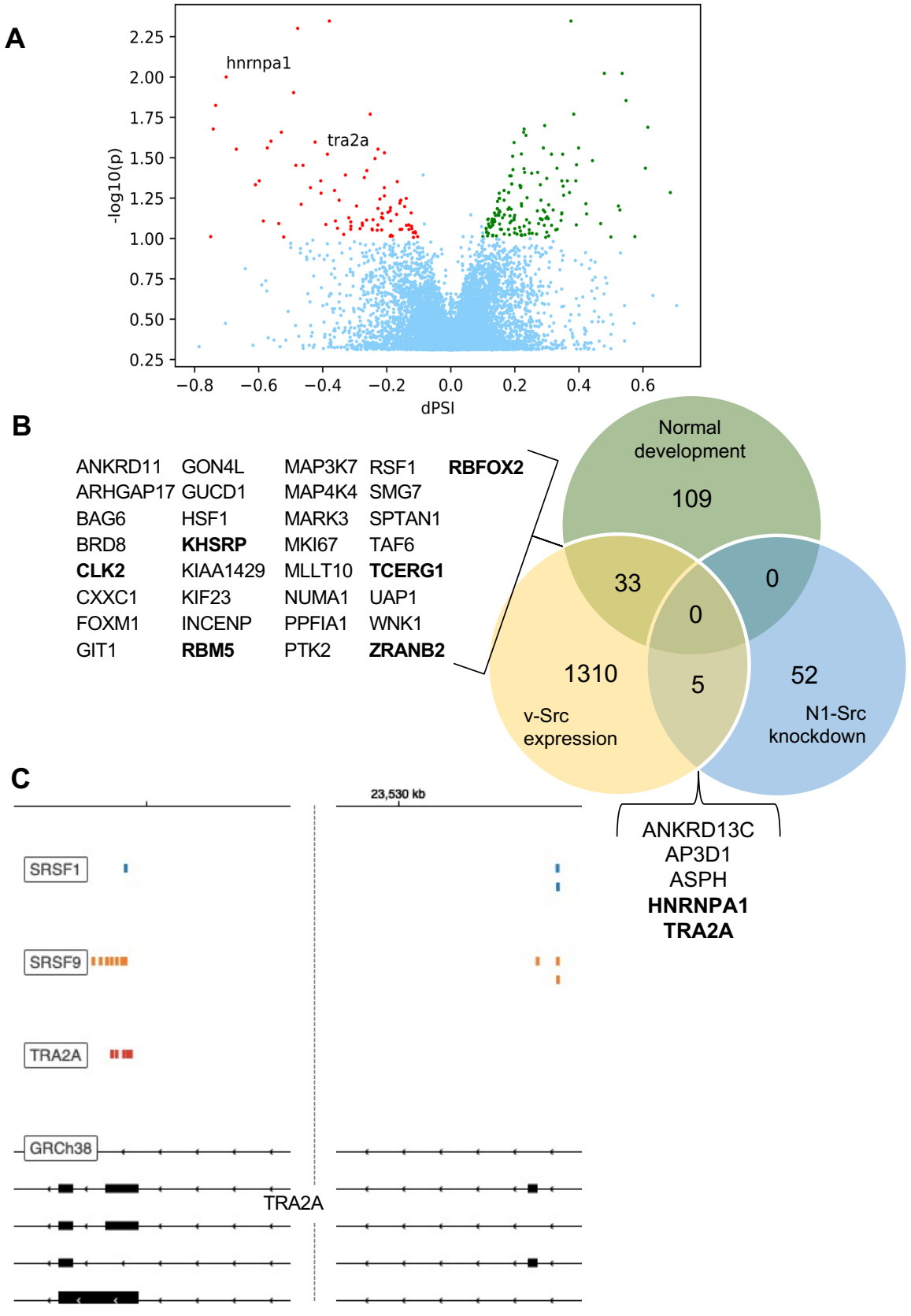


Figure 10. N1-Src knockdown in frog embryos and v-Src expression in a human MCF10-ER-Src cell line cause alternative splicing in transcripts of splicing factors HNRNPA1 and TRA2A.
(A) Volcano plot highlighting significant differential exon skipping events ($p<0.1$, $|dPSI| > 0.1$) that occur upon N1-Src knockdown. (B) Venn diagram showing the overlap among the differentially spliced genes following N1-Src knockdown in X.tropicalis embryos, v-Src activation in human cell line and normal development of X.tropicalis embryo. Splicing factors are indicated in bold. (C) Instances of SRSF1, SRSF9 and TRA2A binding motifs within the chromosome sequence (chr7:23521841-23531788) surrounding the TRA2A intron that is retained after v-Src expression is activated in MCF10-ER-Src cells. The dotted line down the middle acts as a 'broken axis' for the intron.

Chapter 4: Discussion

N1-Src is a tyrosine kinase known to be essential in the normal development of the nervous system both during early neurogenesis and later stages of neuron morphogenesis (Keenan et al., 2017; Lewis et al., 2017). Its high expression has also been shown to correlate with the differentiation of neuroblastoma cells (Matsunaga et al.,1993). However, its mechanism of action remains less understood. Preliminary transcriptomic and phosphoproteomic data suggest N1-Src to be a regulator of splicing. In this study, we have further investigated the implication of specific alternative splicing events and factors in N1-Src induced neurogenesis.

4.1: Src regulates the splicing of splicing factors

By analysing data from human and *Xenopus* models, we show that Src is a conserved regulator of the splicing of splicing factors. Evolutionary conservation provides evidence for functional significance of Src's role in modulation AS patterns. Similarly to N1-Src in *X.tropicalis* embryos, v-Src expression in human MCF10-ER-Src cells regulates alternative splicing of SFs hnRNPA1 and TRA2A. Interestingly, instead of having opposite effects, both the N1-Src knockdown and induction of v-Src expression cause an intron retention event in the TRA2A transcripts to occur. We also confirm that splicing of splicing factors occurs during normal *Xenopus* neurula. As mentioned in section 1.6, hnRNPA1 represses N1-Src splicing *in vitro* (Rooke et al., 2003), therefore N1-Src regulation of *HNRNPA1* splicing might be part of a feedback loop.
Signalling-activated kinases have previously been shown to mediate SF localization and activity through phosphorylation (Naro and Sette 2013). For example, SR protein kinase (SRPK)-mediated phosphorylation of SRSF1 RS domain is required for its nuclear import and for its interaction with U1snRNP, thus promoting spliceosome assembly (Long et al. 2019;

Gonçalves et al. 2014). This suggests a mechanism to N1-Src regulation of neurogenesis, whereby phosphorylation of one set of splicing factors promotes changes in splicing of a new set of SFs which then leads to global changes in alternative splicing patterns.

The primary aim of this work has been to look for correlations between DS events, splicing regulator expression and motif enrichment across different biological settings. This would help us identify potential splicing factor targets of Src regulation that play roles in vertebrate nervous system development. Unfortunately, little overlap was not only seen between results of analysis of different datasets but also analysis of the same datasets with different bioinformatics tools. This issue can be explained by the differences in algorithms and approaches used by these tools but also by features of the datasets themselves.

As mentioned in section 1.2, most RBP binding motif sequences are short and occur frequently in differentially spliced transcripts. This limited the ability of AME to achieve both high specificity and sensitivity in reporting the enriched motifs from the CISBP-RNA database as these motifs tend to appear frequently in the genome. As with the case of SRSF1, the effect of many splicing regulators' binding depends on its location in regard to the exon junction. Therefore, the Matt RNA map function was a much more useful tool as it provided information on spatial distribution of the enriched motifs. Unfortunately, this option was only available for the human dataset as it did not cover Xenopus RBPs and therefore could not determine the motif distribution for the developmental series datasets.

Additionally, the current advice for experiments looking for information on alternative splicing is that they require a sequencing depth of 20–60 million reads per sample. The read coverage for the *X.tropicalis* and MCF10A-vSrc datasets was under 10 million reads, which is only really enough for differential gene expression analysis. As justified previously, the use of the vast-tools diff module enabled the correction for biases caused by low numbers of reads for these datasets, however combining files and reducing the number of replicates has an inevitable impact on statistical analysis. Therefore, future DS analysis should only be undertaken in cases where relevant deep sequencing datasets are available.

One putative Src substrate that consistently appeared as a regulator of DS events across different datasets was the essential splicing regulator SRSF1. It was identified as a regulator of Alt3' splice site events in the human cell line and as a regulator of intron retention during normal *X.laevis* neural crest development. Additionally, the preliminary motif enrichment analysis done by the York Genomics and Bioinformatics facility upon the N1-Src knockdown *X.tropicalis* embryo RNA-Seq data indicated SRSF9 to be a regulator of splice site selection

(Katherine Newling). Its binding motifs are enriched at the DS junctions of both IR and exon skipping events. RBP maps produced by Matt suggested that in cells where v-Src is expressed SRSF1 and SRSF9 binding to regions in the regulated exons leads to exon skipping. According to James Ormond's data analysis (unpublished MSc dissertation), the phosphorylation motifs recognised by Src in the RRM1 domains of SRSF1 and SRSF9 are very similar and therefore potentially closely related.

4.2 : Regulation of splicing as a therapeutic strategy

Disruptions in alternative splicing patterns have been associated with a range of neurodevelopmental and neurodegenerative disorders and cancer (Sanders et al. 2020; Porter et al. 2018). High-risk neuroblastoma has been associated with high expression levels of spliceosomal components and regulatory splicing factors, including HNRNPA1. Additionally, recent experiments showed that selective spliceosome inhibition with pladienolide B reduces tumour growth *in vivo* (Shi et al., 2020) In the study presented here more evidence to support the previous observations that some splicing factors are downregulated during normal neuronal differentiation and that one of the proposed mechanisms that regulate this process is IR, which is increased during neurogenesis. High levels of N1-Src are a positive prognostic predictor in neuroblastoma. In section 3.3, it was shown that consensus RNA binding motifs of the putative Src-substrates SRSF1 and RBM47 are enriched at junctions of IR events regulated during normal neural crest development in *X.laevis*. This presented an idea for a mechanism in which the disruption of normal Src-mediated intron retention events would lead to over- or aberrant expression of splicing factors in the developing neural crest. Potentially, due to the SUPPA algorithm being less sensitive to IR events, the few genes identified were enriched for RNA binding but not splicing. Therefore, more transcriptomic data would be needed to confirm this hypothesis. Interestingly, N1-Src knockdown and overexpression seem to have similar, instead of opposite, effects on the processes it regulates (Kotani et al. 2007; Keenan et al. 2017). Therefore, simply increasing the levels of N1-Src would not be a suitable therapeutic strategy for neuroblastoma as its expression has a global effect on a complex splicing programme. However, antisense oligonucleotide (ASO) drugs that module the AS of specific genes are being developed to treat nervous system disorders. Nusinersen is a clinically approved ASO medication used in treating spinal muscular atrophy via altering exon inclusion in the *SMN1* transcript and promoting the production of the full-length SMN protein (Chiriboga 2017). Similar therapies could be explored and developed specifically for each splicing variant affected by aberrant N1-Src expression.

The results of this project so far provide a starting point for further laboratory research. They further describe the key players in the splicing cascade regulated by Src phosphorylation during neuronal development. In the future, protein-centric methods of studying RNA-protein interaction, such as HITS-CLIP can be used to investigate which splicing events and isoform switches are regulated by the splicing factors phosphorylated by Src (Licatalosi et al., 2008; Ramanathan et al., 2019).

Definitions and abbreviations

Alt3 Alternative 3’ site
Alt5 Alternative 5’ site
AS Alternative splicing
DNA Deoxyribonucleic acid
ESE Exonic splicing enhancer
ESS Exonic splicing silencer
hnRNP heterogeneous nuclear ribonucleoprotein
IR intron retention
ISE Intronic splicing enhancer
ISS Intronic splicing silencer
MO Morpholino oligo
mRNA Messenger ribonucleic acid
NB Neural plate border
NC Neural crest
NMD Nonsense mediated decay
NNE Non-neural ectoderm
NP Neural plate
NPC Neural progenitor cell
nt Nucleotide
PTB Polypyrimidine-tract-binding protein
PTC Premature termination codon
P-value Probability value
RA Retinoic acid
RBP RNA binding protein
RNA Ribonucleic acid
SE Exon skipping
SF Splicing factor
SFK Src family kinase
snRNP small nuclear ribonucleoprotein particle
UTR Untranslated region

References

Andrews S, Others (2010) FastQC: a quality control tool for high throughput sequence data.

Arold ST, Ulmer TS, Mulhern TD, Werner JM, Ladbury JE, Campbell ID, Noble MEM (2001) The role of the SH3-SH2 interface in the regulation of Src kinases. Available at: http://www.jbc.org/content/early/2001/02/02/jbc.M011185200.full.pdf.

Atsumi S, Wakabayashi K, Titani K, Fujii Y, Kawate T (1993) Neuronal pp60c-src(+) in the developing chick spinal cord as revealed with anti-hexapeptide antibody. J Neurocytol 22:244–258.

Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, Wishart DS (2016) Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res 44:W147–W153.

Beggs HE, Soriano P, Maness PF (1994) NCAM-dependent neurite outgrowth is inhibited in neurons from Fyn-minus mice. J Cell Biol 127:825–833.

Benoit Bouvrette LP, Bovaird S, Blanchette M, Lécuyer E (2020) oRNAment: a database of putative RNA binding protein target sites in the transcriptomes of model species. Nucleic Acids Res 48:D166–D173.

Berdeaux RL, Díaz B, Kim L, Martin GS (2004) Active Rho is localized to podosomes induced by oncogenic Src and is required for their assembly and function. J Cell Biol 166:317–323.

Bjelfman C, Hedborg F, Johansson I, Nordenskjöld M, Påhlman S (1990) Expression of the neuronal form of pp60c-src in neuroblastoma in relation to clinical stage and prognosis. Cancer Res 50:6908–6914.

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 24:1774–1786.

Brodeur GM (2003) Neuroblastoma: biological insights into a clinical enigma. Nat Rev Cancer 3:203–216.

Brugge J, Cotton P, Lustig A, Yonemoto W, Lipsich L, Coussens P, Barrett JN, Nonner D, Keane RW (1987) Characterization of the altered form of the c-src gene product in neuronal cells. Genes Dev 1:287–296.

Cartwright CA, Simantov R, Cowan WM, Hunter T, Eckhart W (1988) pp60c-src expression in the developing rat brain. Proc Natl Acad Sci U S A 85:3348–3352.

Chan RC, Black DL (1997) The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream. Mol Cell Biol 17:4667–4676.

Chen J, Hackett CS, Zhang S, Song YK, Bell RJA, Molinaro AM, Quigley DA, Balmain A, Song JS, Costello JF, Gustafson WC, Van Dyke T, Kwok P-Y, Khan J, Weiss WA (2015) The genetics of splicing in neuroblastoma. Cancer Discov 5:380–395.

Chen M, Manley JL (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol 10:741–754.

Chou MY, Rooke N, Turck CW, Black DL (1999) hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. Mol Cell Biol

19:69–77.

Chou M-Y, Underwood JG, Nikolic J, Luu MHT, Black DL (2000) Multisite RNA Binding and Release of Polypyrimidine Tract Binding Protein during the Regulation of c-src Neural-Specific Splicing. Mol Cell 5:949–957.

Dehm SM, Bonham K (2004) SRC gene expression in human cancer: the role of transcriptional activation. Biochem Cell Biol 82:263–274.

Ding F, Su C, Chow K-HK, Elowitz MB (2020) Dynamics and functional roles of splicing factor autoregulation. Cold Spring Harbor Laboratory:2020.07.22.216887 Available at: https://www.biorxiv.org/content/10.1101/2020.07.22.216887v1.full [Accessed January 10, 2021].

Dvinge H (2018) Regulation of alternative mRNA splicing: old players and new perspectives. FEBS Lett 592:2987–3006.

Engen JR, Wales TE, Hochrein JM, Meyn MA 3rd, Banu Ozkan S, Bahar I, Smithgall TE (2008) Structure and dynamic regulation of Src-family kinases. Cell Mol Life Sci 65:3058–3073.

Fong N, Kim H, Zhou Y, Ji X, Qiu J, Saldi T, Diener K, Jones K, Fu X-D, Bentley DL (2014) Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes Dev 28:2663–2676.

Gohr A, Irimia M (2019) Matt: Unix tools for alternative splicing analysis. Bioinformatics 35:130–132.

Gonatopoulos-Pournatzis T et al. (2020) Autism-Misregulated eIF4G Microexons Control Synaptic Translation and Higher Order Cognitive Functions. Mol Cell 77:1176–1192.e16.

Gondran P, Dautry F (1999) Regulation of mRNA splicing and transport by the tyrosine kinase activity of src. Oncogene 18:2547–2555.

Guo W, Tzioutziou N, Stephen G, Milne I, Calixto C, Waugh R, Brown JWS, Zhang R (2019) 3D RNA-seq - a powerful and flexible tool for rapid and accurate differential expression and alternative splicing analysis of RNA-seq data for biologists. bioRxiv:656686 Available at: https://www.biorxiv.org/content/10.1101/656686v1.full [Accessed December 10, 2019].

Guo X, Chen Q-R, Song YK, Wei JS, Khan J (2011) Exon array analysis reveals neuroblastoma tumors have distinct alternative splicing patterns according to stage and MYCN amplification status. BMC Med Genomics 4:35.

Hämmerle B, Yañez Y, Palanca S, Cañete A, Burks DJ, Castel V, Font de Mora J (2013) Targeting neuroblastoma stem cells with retinoic acid and proteasome inhibitor. PLoS One 8:e76761.

Han H et al. (2017) Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. Mol Cell 65:539-553.e7.

Huang Y, Sanguinetti G (2016) Statistical modeling of isoform splicing dynamics from RNA-seq time series data. Bioinformatics 32:2965–2972.

Hughes TA (2006) Regulation of gene expression by alternative untranslated regions. Trends Genet 22:119–122.

Ignelzi MA, Miller DR, Soriano P, Maness PF (1994) Impaired neurite outgrowth of src-minus cerebellar neurons on the cell adhesion molecule L1. Neuron 12:873–884.

Iliopoulos D, Hirsch HA, Struhl K (2009) An epigenetic switch involving NF-kappaB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. Cell 139:693–706.

Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M, Quesnel-Vallières M, Tapial J, Raj B, O'Hanlon D, Barrios-Rodiles M, Sternberg MJE, Cordes SP, Roth FP, Wrana JL, Geschwind DH, Blencowe BJ (2014) A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell 159:1511–1523.

Ishizawar R, Parsons SJ (2004) c-Src and cooperating partners in human cancer. Cancer Cell 6:209–214.

Jacob AG, Smith CWJ (2017) Intron retention as a component of regulated gene expression programs. Hum Genet 136:1043–1057.

Janesick A, Wu SC, Blumberg B (2015) Retinoic acid signaling and neuronal differentiation. Cell Mol Life Sci 72:1559–1576.

Jangi M, Boutz PL, Paul P, Sharp PA (2014) Rbfox2 controls autoregulation in RNA-binding protein networks. Genes Dev 28:637–651.

Jin Y, Suzuki H, Maegawa S, Endo H, Sugano S, Hashimoto K, Yasuda K, Inoue K (2003) A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. EMBO J 22:905–912.

Ji Z, He L, Regev A, Struhl K (2019) Inflammatory regulatory network mediated by the joint action of NF-kB, STAT3, and AP-1 factors is involved in many human cancers. Proc Natl Acad Sci U S A 116:9453–9462.

Jurica MS, Roybal GA (2013) RNA Splicing. In: Encyclopedia of Biological Chemistry (Second Edition) (Lennarz WJ, Lane MD, eds), pp 185–190. Waltham: Academic Press.

Kaplan KB, Swedlow JR, Varmus HE, Morgan DO (1992) Association of p60c-src with endosomal membranes in mammalian fibroblasts. J Cell Biol 118:321–333.

Karimi K, Fortriede JD, Lotay VS, Burns KA, Wang DZ, Fisher ME, Pells TJ, James-Zorn C, Wang Y, Ponferrada VG, Chu S, Chaturvedi P, Zorn AM, Vize PD (2018) Xenbase: a genomic, epigenomic and transcriptomic model organism database. Nucleic Acids Res 46:D861–D868.

Keenan S, Lewis PA, Wetherill SJ, Dunning CJR, Evans GJO (2015) The N2-Src neuronal splice variant of C-Src has altered SH3 domain ligand specificity and a higher constitutive activity than N1-Src. FEBS Lett 589:1995–2000.

Keenan S, Wetherill SJ, Ugbode CI, Chawla S, Brackenbury WJ, Evans GJO (2017) Inhibition of N1-Src kinase by a specific SH3 peptide ligand reveals a role for N1-Src in neurite elongation by L1-CAM. Sci Rep 7:43106.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37:907–915.

Kotani T, Morone N, Yuasa S, Nada S, Okada M (2007) Constitutive activation of neuronal Src causes aberrant dendritic morphogenesis in mouse cerebellar Purkinje cells. Neurosci Res 57:210–219.

Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 446:926–929.

Le Roux A-L, Busquets MA, Sagués F, Pons M (2016) Kinetics characterization of c-Src binding to lipid membranes: Switching from labile to persistent binding. Colloids Surf B Biointerfaces 138:17–25.

Leung LC, Harris WA, Holt CE, Piper M (2015) NF-Protocadherin Regulates Retinal Ganglion Cell Axon Behaviour in the Developing Visual System. PLoS One 10:e0141290.

Lewis PA, Bradley IC, Pizzey AR, Isaacs HV, Evans GJO (2017) N1-Src Kinase Is Required for Primary Neurogenesis in Xenopus tropicalis. J Neurosci 37:8477–8485.

Licatalosi DD, Darnell RB (2006) Splicing regulation in neurologic disease. Neuron 52:93–101.

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, Darnell JC, Darnell RB (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456:464–469.

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 25:1–13.

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T (2010) Regulation of alternative splicing by histone modifications. Science 327:996–1000.

Lynch SA, Brugge JS, Levine JM (1986) Induction of altered c-src product during neural differentiation of embryonal carcinoma cells. Science 234:873–876.

Madgwick A, Fort P, Hanson PS, Thibault P, Gaudreau M-C, Lutfalla G, Möröy T, Abou Elela S, Chaudhry B, Elliott DJ, Morris CM, Venables JP (2015) Neural differentiation modulates the vertebrate brain specific splicing program. PLoS One 10:e0125998.

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 27:435–448.

Maris JM, Matthay KK (1999) Molecular biology of neuroblastoma. J Clin Oncol 17:2264–2279.

Martinez R, Mathey-Prevot B, Bernards A, Baltimore D (1987) Neuronal pp60c-src contains a six-amino acid insertion relative to its non-neuronal counterpart. Science 237:411–415.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10–12.

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

Matsunaga T, Shirasawa H, Tanabe M, Ohnuma N, Takahashi H, Simizu B (1993) Expression of alternatively spliced src messenger RNAs related to neuronal differentiation in human neuroblastomas. Cancer Res 53:3179–3185.

Matten WT, Aubry M, West J, Maness PF (1990) Tubulin is phosphorylated at tyrosine by pp60c-src in nerve growth cone membranes. J Cell Biol 111:1959–1970.

Matthay KK, Reynolds CP, Seeger RC, Shimada H, Adkins ES, Haas-Kogan D, Gerbing RB,

London WB, Villablanca JG (2009) Long-term results for children with high-risk neuroblastoma treated on a randomized trial of myeloablative therapy followed by 13-cis-retinoic acid: a children's oncology group study. J Clin Oncol 27:1007–1013.

McLeay RC, Bailey TL (2010) Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics 11:165.

McManus CJ, Graveley BR (2011) RNA structure and the mechanisms of alternative splicing. Curr Opin Genet Dev 21:373–379.

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 Available at: http://dx.doi.org/10.1093/bib/bbz126.

Middleton R, Gao D, Thomas A, Singh B, Au A, Wong JJ-L, Bomane A, Cosson B, Eyras E, Rasko JEJ, Ritchie W (2017) IRFinder: assessing the impact of intron retention on mammalian gene expression. Genome Biol 18:51.

Parsons SJ, Parsons JT (2004) Src family kinases, key regulators of signal transduction. Oncogene 23:7906–7909.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417–419.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É (2011) Scikit-learn: Machine Learning in Python. J Mach Learn Res 12:2825–2830.

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11:1650–1667.

Pervouchine D, Popov Y, Berry A, Borsari B, Frankish A, Guigó R (2019) Integrative transcriptomic analysis suggests new autoregulatory splicing events coupled with nonsense-mediated mRNA decay. Nucleic Acids Res 47:5293–5306.

Peshkin L, Wühr M, Pearl E, Haas W, Freeman RM Jr, Gerhart JC, Klein AM, Horb M, Gygi SP, Kirschner MW (2015) On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development. Dev Cell 35:383–394.

Pillay I, Nakano H (1996) Radicicol Inhibits Tyrosine Phosphorylation of the Mitotic Src Substrate Sam68 and Retards Subsequent Exit from Mitosis of Src-transformed Cells1. Cell Cycle Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1001.3470&rep=rep1&type=pdf.

Pimentel H, L BN, Puente S, Melsted P, Pachter L (2016a) Differential analysis of RNA-Seq incorporating quantification uncertainty. Cold Spring Harbor Laboratory:058164 Available at: https://www.biorxiv.org/content/10.1101/058164v1 [Accessed February 25, 2021].

Pimentel H, Parra M, Gee SL, Mohandas N, Pachter L, Conboy JG (2016b) A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis. Nucleic Acids Res 44:838–851.

Plouhinec J-L, Medina-Ruiz S, Borday C, Bernard E, Vert J-P, Eisen MB, Harland RM, Monsoro-Burq AH (2017) A molecular atlas of the developing ectoderm defines neural,

neural crest, placode, and nonneural progenitor identity in vertebrates. PLoS Biol 15:e2004045.

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Raj B, Blencowe BJ (2015) Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. Neuron 87:14–27.

Ramanathan M, Porter DF, Khavari PA (2019) Methods to study RNA-protein interactions. Nat Methods 16:225–234.

Ray D et al. (2013) A compendium of RNA-binding motifs for decoding gene regulation. Nature 499:172–177.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47.

Rooke N, Markovtsov V, Cagavi E, Black DL (2003) Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1. Mol Cell Biol 23:1874–1884.

Sandilands E, Cans C, Fincham VJ, Brunton VG, Mellor H, Prendergast GC, Norman JC, Superti-Furga G, Frame MC (2004) RhoB and actin polymerization coordinate Src activation with endosome-mediated delivery to the membrane. Dev Cell 7:855–869.

Sefton BM, Hunter T (1986) From c-src to v-src, or the case of the missing C terminus. Cancer Surv 5:159–172.

Session AM et al. (2016) Genome evolution in the allotetraploid frog Xenopus laevis. Nature 538:336–343.

Shi Y, Rraklli V, Maxymovitz E, Li S, Westerlund I, Bedoya-Reina OC, Yuan J, Bullova P, Christofer Juhlin C, Stenman A, Larsson C, Kogner P, O'Sullivan MJ, Schlisio S, Holmberg J (2020) Alternative splicing in neuroblastoma generates RNA-fusion transcripts and is associated with vulnerability to spliceosome inhibitors. Cold Spring Harbor Laboratory:851238 Available at: https://www.biorxiv.org/content/10.1101/851238v2.full [Accessed March 30, 2021].

Singh R (2002) RNA-protein interactions that regulate pre-mRNA splicing. Gene Expr 10:79–92.

Smart JE, Oppermann H, Czernilofsky AP, Purchio AF, Erikson RL, Bishop JM (1981) Characterization of sites for tyrosine phosphorylation in the transforming protein of Rous sarcoma virus (pp60v-src) and its normal cellular homologue (pp60c-src). Proc Natl Acad Sci U S A 78:6013–6017.

Spassov DS, Ruiz-Saenz A, Piple A, Moasser MM (2018) A Dimerization Function in the Intrinsically Disordered N-Terminal Region of Src. Cell Rep 25:449–463.e4.

Spellman R, Llorian M, Smith CWJ (2007) Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. Mol Cell 27:420–434.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl

Acad Sci U S A 102:15545–15550.

Tan MH, Au KF, Yablonovitch AL, Wills AE, Chuang J, Baker JC, Wong WH, Li JB (2013) RNA sequencing reveals a diverse and dynamic repertoire of the Xenopus tropicalis transcriptome over development. Genome Res 23:201–216.

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 27:1759–1768.

Taylor SJ, Anafi M, Pawson T, Shalloway D (1995) Functional Interaction between c-Src and Its Mitotic Target, Sam 68. J Biol Chem 270:10120–10124.

Thomas SM, Brugge JS (1997) CELLULAR FUNCTIONS REGULATED BY SRC FAMILY KINASES. Available at: https://www.annualreviews.org/doi/abs/10.1146/annurev.cellbio.13.1.513 [Accessed February 5, 2019].

Tomolonis JA, Agarwal S, Shohet JM (2018) Neuroblastoma pathogenesis: deregulation of embryonic neural crest development. Cell Tissue Res 372:245–262.

Torres-Méndez A, Bonnal S, Marquez Y, Roth J, Iglesias M, Permanyer J, Almudí I, O'Hanlon D, Guitart T, Soller M, Gingras A-C, 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 3:691–701.

Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, Eyras E (2018) SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biol 19:40.

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 444:580–586.

Wang Y, Liu J, Huang BO, Xu Y-M, Li J, Huang L-F, Lin J, Zhang J, Min Q-H, Yang W-M, Wang X-Z (2015) Mechanism of alternative splicing and its regulation. Biomed Rep 3:152–158.

Wetherill SJ (2016) The Role of N1-Src in Neuronal Development. Available at: http://etheses.whiterose.ac.uk/15573/ [Accessed February 6, 2019].

Weyn-Vanhentenryck SM, Feng H, Ustianenko D, Duffié R, Yan Q, Jacko M, Martinez JC, Goodwin M, Zhang X, Hengst U, Lomvardas S, Swanson MS, Zhang C (2018) Precise temporal regulation of alternative splicing during neural development. Nat Commun 9:2189.

Weyn-Vanhentenryck 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 6:1139–1152.

Wilde A, Beattie EC, Lem L, Riethof DA, Liu S-H, Mobley WC, Soriano P, Brodsky FM (1999) EGF Receptor Signaling Stimulates SRC Kinase Phosphorylation of Clathrin, Influencing Clathrin Redistribution and EGF Uptake. Cell 96:677–687.

Witten JT, Ule J (2011) Understanding splicing regulation through RNA splicing maps. Trends Genet 27:89–97.

Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CWJ (2004) Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol Cell 13:91–100.

Xu W, Doshi A, Lei M, Eck MJ, Harrison SC (1999) Crystal Structures of c-Src Reveal Features of Its Autoinhibitory Mechanism. Mol Cell 3:629–638.

Yap K, Lim ZQ, Khandelia P, Friedman B, Makeyev EV (2012) Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. Genes Dev 26:1209–1223.

Yeo G, Holste D, Kreiman G, Burge CB (2004) Variation in alternative splicing across human tissues. Genome Biol 5:R74.

Zhang C, Frias MA, Mele A, Ruggiu M, Eom T, Marney CB, Wang H, Licatalosi DD, Fak JJ, Darnell RB (2010) Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. Science 329:439–443.

Zhang S et al. (2016a) MYCN controls an alternative RNA splicing program in high-risk metastatic neuroblastoma. Cancer Lett 371:214–224.

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 (2016b) Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. Cell 166:1147–1162.e15.