UNIVERSITAT POLITÈCNICA DE VALÈNCIA

DEPARTAMENTO DE BIOTECNOLOGÍA

Experimental validation of tappAS: a computational framework to assess the functional impact of alternative splicing in a neural differentiation context

TRABAJO FIN DE MÁSTER EN BIOTECNOLOGÍA BIOMÉDICA

ALUMNO/A: Pablo Bonilla Villamil

TUTOR/A: Victoria Moreno Manzano

Curso Académico: 2018-2019

VALENCIA, 03 de Julio de 2019

1. Index

1. Resumen

El *Splicing* alternativo (SA) es un mecanismo regulador de la expresión génica que contribuye a la diversidad proteómica al aumentar el número de especies de ARNm que se transcriben a partir de un solo gen. Como resultado, las variedades o isoformas de la misma proteína pueden presentar una estructura, ubicación y función diferentes. La variabilidad en la expresión de las isoformas transcritas podría jugar un papel importante en los procesos de diferenciación y en la determinación del destino celular, en la aparición de enfermedades y se puede usar para la caracterización de diferentes poblaciones dentro de un solo tipo celular. TappAS Tool es una aplicación Java desarrollada para el análisis de datos, obtenidos mediante RNAseq, a nivel de gen e isoforma. Proporciona un conjunto de herramientas para estudiar el *EA*, pero también su implicación funcional, es decir, si este *EA* causa algún efecto en la funcionalidad de la proteína. El objetivo de este proyecto es la validación experimental de TappAS para revelar su efectividad, especificidad y sensibilidad para detectar y evaluar las tasas de expresión en el nivel de isoformas. De acuerdo con la literatura, los niveles de *AA* son particularmente altos en el sistema nervioso, por lo tanto, el *pipeline* se desarrolló utilizando los datos obtenidos mediante PacBio RNAseq de tres linajes de células neuronales de ratón: Células Progenitoras Neurales (NPC), Células Progenitoras de Oligodendrocitos (OPC) y Células Progenitoras de Motoneurona (MTNs). El proyecto se dividió en la elección de los genes candidatos, para lo que se aprovechaba el análisis *in silico* de TappAS, y en la validación experimental, que consistió en el análisis de los niveles de expresión y la presencia de las isoformas en los cultivos mediante la reacción en cadena de la polimerasa en tiempo real (qPCR) y Western Blot (WB).

Palabras clave: Alternative splicing, TappAS, RNAseq, isoformas, NPCs, OPCs, MTNs,

2. Abstract

Alternative splicing (*AS*) is a regulatory mechanism of gene expression that contributes to proteomic diversity by increasing the number of messenger RNA (mRNA) species that are transcribed from a single gene. As a result, varieties or isoforms of the same protein may present different structure, location and function. Variability in the expression of the transcript isoforms might play an important role in differentiation processes and in the determination of cell fate, in the appearance of diseases and may be used for the characterization of different cellular populations within a single cell type. TappAS Tool is a Java application developed for the analysis of RNAseqdata at the gene and isoform level. It provides a set of tools to study the *AS* but also its functional implication, i.e. if this *AS* cause any effect in the functionality of the protein. The aim of this project is the experimental validation of TappAS to reveal its effectiveness, specificity and sensitivity to detect and evaluate expression rates at the isoform level. According to the literature, *AS* levels are particularly high in the nervous system, therefore the pipeline was developed using PacBio RNAseqdata from three mouse neural cell lineages, Neural Progenitors Cells (NPCs), Oligodendrocyte Progenitor Cells (OPCs) and Motoneuron Progenitors Cells (MTNs). The project was divided in the election of gene candidates leveraging TappAS *in silico* analysis and in the experimental validation, which consisted in the analysis of the expression rates and presence of the isoforms in the cultures by means of Real time polymerase chain reaction (qPCR) and Western Blot (WB).

Keywords: Alternative splicing (*AS*), TappAS, RNAseq, isoforms, NPCs, OPCs, MTNs,

3. Objectives

The present project aims two objectives. One of objectives of the project is the analysis of TappAs in order to disclosure the effectiveness, specificity and sensibility of the application to detect and assess expression rates at the isoform level by single gene expression analysis. The second goal is taking advantage of the information obtained by TappAS to find genes involved in the differentiation of two neural lineages that could be determinant for the cell-fate determination. For that purpose, the strategy is divided in:

1. Determination of the transcript isoforms presence and expression levels by means of real-time polymerase chain reaction (qPCR) to analyse whether information provided by TappAS is reliable.
2. Identification of the predicted isoforms at the protein level by Western Blotting in order to provide a functional role of the Alternative splicing (*AS*).

4. Introduction

4.1. The mechanism of Alternative Splicing

The central dogma of molecular biology explains the flow of the gene expression in a two-step process. It is normally referred to as “DNA makes RNA makes protein”. A molecule of DNA, which contains the original information, is transcribed to a temporary copy of RNA that finally is translated into a protein which compounds the machinery of the cell. The transcriptome is subjected to regulatory processes referred to as Alternative Transcript Processing (AltTP), which includes Alternative splicing (*AS*), Alternative Polyadenylation (APA) and Alternative Transcription Start Sites SATSS), provides a mechanism to increase diversity and fine-tune transcriptomes and proteomes (Schaefke et al. 2018). These three processes determine which transcripts (*i.e.* transcripts isoforms) a gene may use and contribute to the adaptability of eukaryote transcriptome. *AS* is able to produce transcripts composed by different exons, *APA* has been proposed as a mechanism to scape microRNA regulation by shortening 3’ UTR regions (Fu et al. 2018) and *ATSS* are believed to regulate the inclusion of Upstream Open Reading Frames (uORFs) that control translational rates (Wang et al. 2016).

Constitutive splicing is considered the process of intron removal and exon ligation when they appear in the transcript in the same order as they appear in the gene. However, *AS* is a deviation from this canonical sequence where some exons are skipped resulting in various forms of mature messenger RNA (mRNA) (Zheng et al. 2005). Gilbert first proposed the concept of *AS* in 1978, as the mechanism underlying the discrepancy between the number of protein-coding genes in humans (25.000) and the >90.000 different proteins that are generated (Gilbert 1978). The comparison of mRNA with genomic sequences in the late 1970s showed that prior to the export into the cytosol, viral sequences are removed from the pre-mRNA and the remaining sequences are joined together (Berget et al. 1977; Chow et al. 1977) During the last decades, genome-wide studies have shown that ~95% of multi-exon human genes undergo splicing in a developmental, tissue-specific or signal-transduction dependent manner (Pan et al. 2008; Wang et al. 2008). Although *AS* is a significantly important process in vertebrates, several transcriptome studies reveal that *AS* is also found in

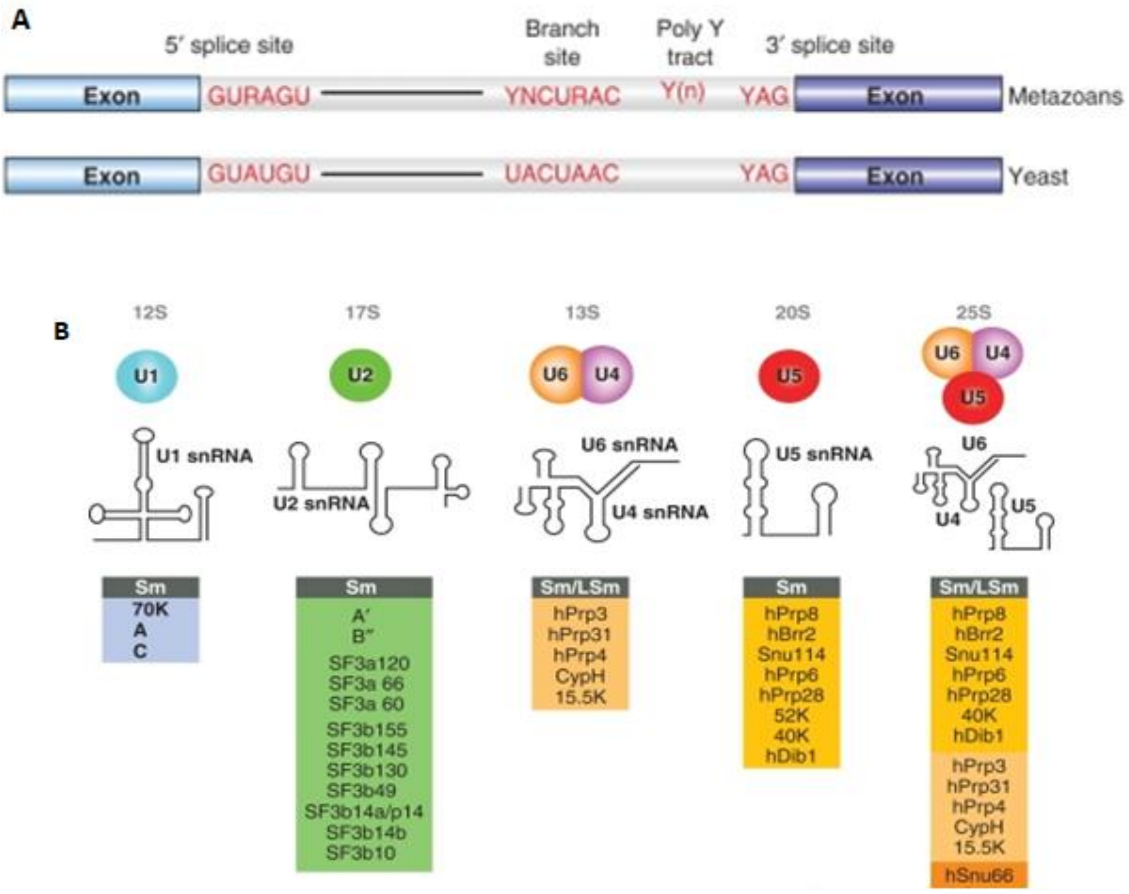
invertebrates (*Drosophila* ~40%) (Gibilisco et al. 2016), plants (*Arabidopsis thaliana* 60%) (Zhang et al. 2017) and fungi (*Verticillium dahliae* 50%) (Jin et al. 2017). The species of higher eukaryotes exhibit a higher proportion of alternatively spliced genes, which is an indication of a prominent role for the mechanism in evolution.

AS is a ubiquitous regulatory mechanism of gene expression, considered one of the major mechanisms of generating transcriptome complexity, which regulates the processing of RNA from pre- to post-transcriptional events providing a powerful cellular mechanism to increase the diversity of transcripts and proteins from a limited source of genes, thereby expanding the regulatory and functional repertoire of eukaryotic organisms. *AS* mediates diverse biological processes over the entire life span of organisms, and its patterns constantly change under physiological conditions, allowing an organism to respond to changes in the environment by switching which exons it expresses and, therefore changing the translated proteins encoded in the mRNAs causing profound functional implications. Thus, *AS* has a role in almost every aspect of protein function, including binding between proteins and ligands, nucleic acids or membranes, localization and enzymatic properties (Kelemen et al. 2013). In a broader vision, it plays a significant functional role in species differentiation and genome evolution (Blencowe 2006) as well as in the development of functionally from simple to complex tissues and is involved in cell differentiation and development, tissue identity and stress (Wang et al. 2016; Furlanis and Scheiffele 2018; Baralle and Giudice 2017; Pleiss et al. 2007).

4.2. Regulation of the Alternative Splicing

RNA cleavage and ligation reactions necessary for intron removal are regulated in a cell-type and developmental-stage specific manner through an extensive protein-RNA interaction network involving *cis*-elements and *trans*-acting factors which play an essential role in defining exon and intron identity. *Cis*-elements which are contained or closely linked to the gene they affect, comprise conserved sequences at the 5′ and 3′ splice sites (GU-AG dinucleotides) that define the boundary of an intron with its upstream and downstream exon, the polypyrimidine tract (YAG) and the branch site (BS) both upstream to the 3’ ss (Will and Lührmann 2011) (Figure 1A). In addition, *cis*-acting pre-mRNA elements include exonic and intronic splicing enhancers (ESEs and ISEs) or silencers (ESSs and ISSs), which modulate splicing by binding regulatory

proteins that stimulate or repress the assembly of spliceosomal complexes (Goren et al. 2006) (Wang et al. 2008). *Trans*-acting factors, which are diffusible and can influence unlinked genes by binding to the *cis*-elements (Schaefke et al. 2018), particularly, RNA-binding proteins (RBPs), whose combinatorial repertoire determines the splicing-site choice and whose coordinated and close regulation is essential to generate context-specific splicing program (Park et al. 2018). All these elements are recognized by a dynamic, flexible and large ribonucleoprotein (RNP) macromolecular machinery called the spliceosome, which is composed of different proteins. The spliceosome is composed by five small nucleoproteins (snRNP) complexes (U1, U2, U3, U4/U6 AND U5) (**Figure 1B**) associated to ~100 core proteins in yeast (Fabrizio et al. 2009) and >300 different proteins in the major human spliceosome (Jurica and Moore 2003).Humans even possess a second spliceosome called the minor spliceosome which contains the U11, U12, U4atac, U5 and U6atac snRNAs (Steitz et al. 2008).Thus, the spliceosome is a macromolecular complex which works in a constant assembly and disassembly cycle on each intron to form a catalytically active complex (**Figure 1C**).



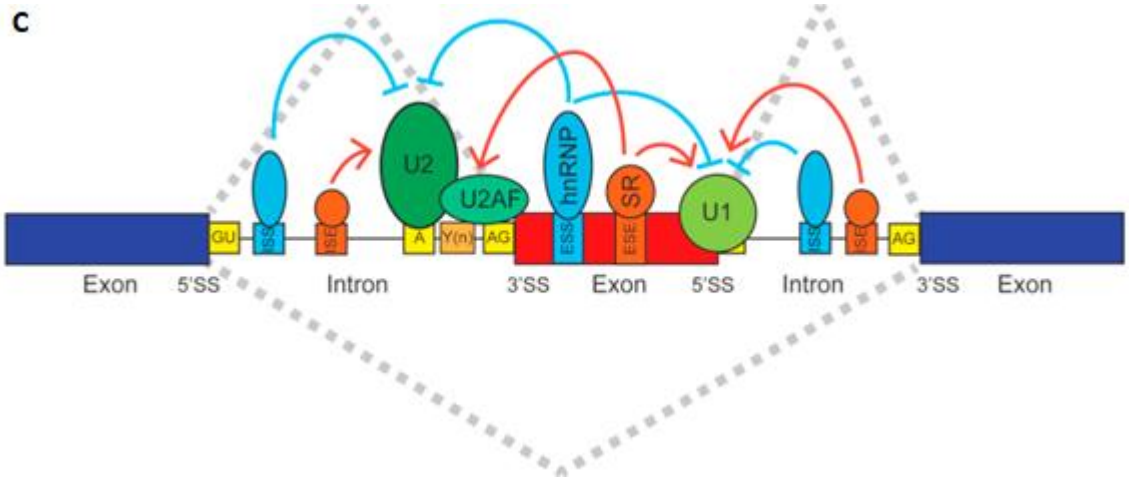


Figure 1. *Cis*-elementes and small ribonucleoproteins (snRNPs) and the spliceosome. Adapted from (Will and Lührmann 2011;Park et al. 2018). (A) Conserved sequences found at the 5′ and 3′ splice sites and branch site of U2-type pre-mRNA introns in metazoans and budding yeast (*S. cerevisiae*). Here, two exons (blue) are separated by an intron (grey). The consensus sequences in metazoans and yeast at the 5' splice site (SS), branch point sequence (BPS), and 3' splice site (SS) are as indicated, where N is any nucleotide, R is a purine, and Y is a pyrimidine. The polypyrimidine tract is a pyrimidine-rich stretch located between the BPS and 3' SS.(B) Protein composition and snRNA secondary structures of the major human spliceosomal snRNPs. (C) Alternative splicing is regulated by an extensive protein-RNA interaction network involving cis elements within the pre-mRNA and trans-acting factors that bind to these cis elements.

Insights into the global complexity of *AS* in mammals have revealed seven main types of alternative splicing (Blencowe 2006). The basic patterns include, cassette-type alternative exons (exon skipping), which is the most prevalent in vertebrates and invertebrates (~30%); intron retention (the most prevalent in lower metazoans), alternative selection of 5' or 3' splice sites (~25%), which is capable of inducing fine changes in the coding sequence (as little as a single codon); mutually exclusive alternative exons; and alternative splicing coupled with alternative first or last exons (**Figure 2A**). Beyond these basic patterns involving binary choices, many complex AS pattern exists in the transcriptome (Vaquero-Garcia et al. 2016) (**Figure 2B**).

Thus, through these mechanisms *AS* can generate mRNAs that differ in their untranslated regions (UTRs) or coding sequences, which may generate mRNA isoforms with distinct functions, stability or subcellular localization, as well as introduce termination codons leading to mRNA downregulation via nonsense-mediated decay

(NMD) (Lareau et al. 2007). On the other hand, some *AS* events may not result in the production of functional proteins, since the transcript may be non-coding, RNA stability may be affected, and localization could prevent the correct function of the transcript or protein. Therefore, *AS* is implicated in the spatiotemporal regulation of mRNAs by modifying the stability and localization of the splice isoforms (Iijima et al. 2016), therefore regulating gene function and adapting it to the needs of the organism.

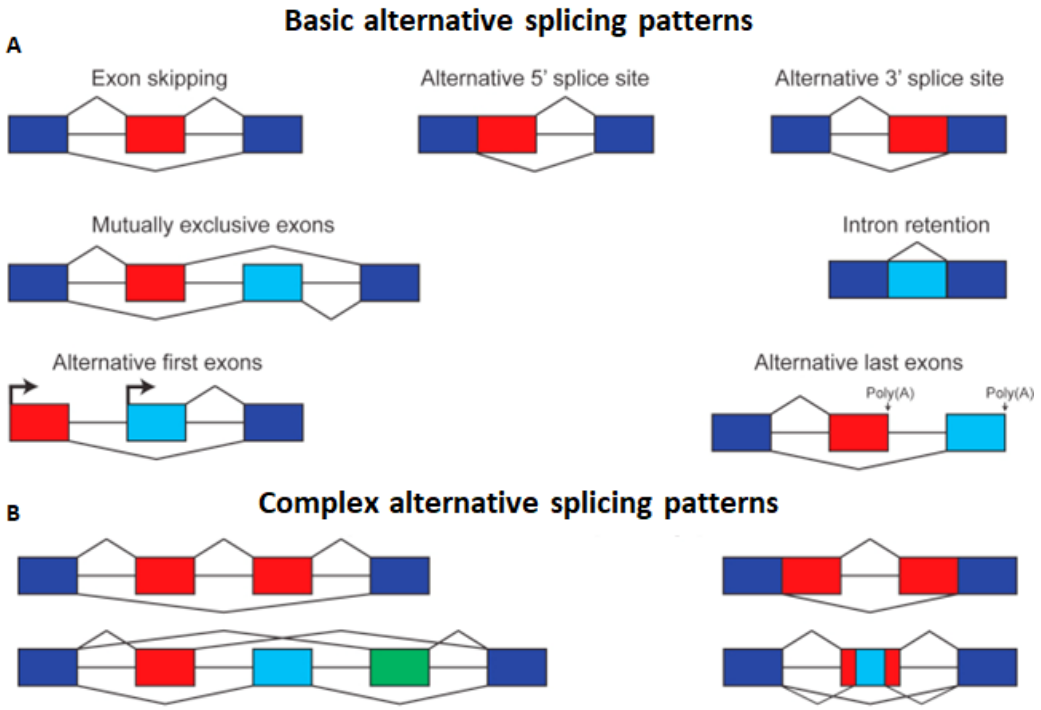


Figure 2. Alternative splicing patterns. Adapted from (Park et al. 2018). Basic (A) and complex (B) patterns of alternative splicing. Dark-blue boxes represent constitutively spliced exons. Red, light blue, and green boxes represent alternatively spliced exons.

4.3. Impact of the Alternative splicing in the nervous system

Neurogenesis is characterized by global changes in the transcriptomes and proteomes of differentiating cells. Previous expressed sequence tagged (ESTs) mapping studies revealed that the brain presents the highest number of *AS* events compared to other organs (Yeo et al. 2004), and different brain regions are associated with complex patterns of alternative splicing (Johnson et al. 2009). Although neurons has been the centre of attention, glia (astrocytes, oligodendrocytes, and microglia) and vascular cells (endothelial cells and pericytes) have been found to be also essential for the proper development and function of the nervous system (Allen and Barres 2009;

Molofsky et al. 2012). Further, an important RNAseq transcriptome study from the cortex revealed cell-type specific *AS* genes among these cell types and even within a single group, which reflects the significant functional heterogeneity found in the nervous system. Thus, this work led to the publication of the first brain cell-type specific transcriptomic database, which even included *AS* events (Zhang et al. 2014).

While most spliceosome components are constitutively expressed, tissue-specific RNA-binding proteins (RBPs) direct spliceosome machinery to specific splice sites in order to generate tissue-specific splicing patterns. Neuron-specific *AS* is one example controlled by the coordinated action of many brain-specific RBPs, and their mechanisms and roles have been reviewed in several articles ((Raj and Blencowe 2015;Lara-Pezzi et al. 2017;Vuong et al. 2016). For instance, the NOVA family of RBPs play a key role in neuronal migration, axon outgrowth, and axon guidance. It has been described that Nova deficiency disrupts the *AS* of DCC and the production of DCC splice variants controlled by NOVA has a crucial function during many stages of commissural neuronal development (Leggere et al. 2016). Moreover, *AS* of specific genes has also been shown to play important roles in the development and function of the nervous system, such as the mechanisms by which differentially spliced isoforms of neuroligin and neurexin mediate synaptic adhesion, which offer a well characterized example of how cell-specific splicing can have considerable functional consequences (Chih et al. 2006).

Besides, dysregulation of the *AS* programmes has been related to the onset of neurological disorders. Genetic association studies of neurodevelopmental disorders including autism, schizophrenia, and intellectual disability syndromes, have identified numerous mutations in transcriptional regulators (Najmabadi et al. 2011;Ronan et al. 2013;De Rubeis et al. 2014;Iossifov et al. 2014;McCarthy et al. 2014), which comprise a large fraction of nuclear proteins such as DNA-binding transcription factors (TFs), histone modification enzymes and their target proteins ((Ronan et al. 2013;De Rubeis et al. 2014;Iossifov et al. 2014). Nevertheless, it is not clear the association between the mutations found in the splicing regulators and the development of neurological disorders.

4.4. TappAS

Despite many computational approaches have been developed to elucidate the dynamics of isoform regulation, the genome-wide study of the functional impact caused by AltTP was a hard task due to the lack of tools and methods integrating contextual isoform levels data and isoform-resolved functional annotation. Traditionally, transcriptome studies have focused either in the characterization of specific isoforms from single genes (Kelemen et al. 2013) or in computational approaches *in silico* (Pan et al. 2009). The lack of functional information in bioinformatic tools dedicated to the splicing is due to the inability to correctly quantify isoform expression from RNAseq data (Steijger et al. 2013) and define the real diversity of transcripts. However, third generation sequencing technologies have solved the issue enabling the study of diversity and expression levels of transcript isoforms combining Illumina sequencing (Weirather et al. 2017) and PacBio Iso-Seq sequencing.

TappAS Tool (http://tappAS.org) is a comprehensive computational framework for the RNAseq data analysis at the gene and isoform levels implemented in an interactive and dynamic Java GUI application which combines statistical and graphical tools for the analysis of AltTP and the functional implication of differential isoform usage. TappAS takes advantage of extensive isoform-resolved functional annotation including structural, coding and non-coding features from different databases which are combined to functionally classify transcript isoforms. Besides, TappAS integrates these features at isoform resolution with isoform expression data, providing a set of statistical methods and graphical tools (normalisation, PCA analysis, venn diagrams, a visualisation engine...) that enable the study of the biological role of isoform regulation.

Since *AS* is particularly active in the Central Nervous System (CNS), the development of Tappas pipeline has been conducted leveraging the transcriptome analysis of two mouse neural cell types, Neural Precursor Cells (NPCs), Oligodendrocyte Precursor Cells (OPCs) and Motoneuron Precursor Cells (MTNs), whose transcriptomes have been defined using PacBio long read RNA sequencing and isoform expression quantified using Illumina sequencing. The intention behind the selection of these *in vitro* models is assessing the differences between two cell lineages that diverge from

the same initial point in order to find genetic differences with biological importance in the differentiations process and in the cell-fate determination. TappAS has demonstrated its effectivity and sensibility to recapitulate a great deal of existing knowledge on isoform function and even reveal new functional insights. The functional effect analysis of *AS* regulation, which is known as the Functional Analysis of Alternative Isoform Usage (**Figure 3**), is divided into three main modules (**Figure 3**). Module 1 defines and measures the functional divergence among gene isoforms within the overall transcriptome. Module 2 focuses on the evaluation of expression levels to understand transcriptome dynamics. Finally, Module 3 addresses the integration of isoform functional, structural and expression information to reveal the potential effect of differential isoform usage on gene properties (**Figure 3**).

Module 1: Isoform functional diversity

To understand functional and regulatory variability between isoforms derived from the same gene The Functional Diversity (FD) analysis identifies the nature and measures the magnitude of changes triggered by alternative processing of transcripts. For that purpose, FD analysis systematically evaluates the genome-wide level of functional (positional functional features), regulatory (positional regulatory features) and structural (CDS, UTRs and PolyA sites) diversity across isoforms.

Module 2: Transcriptome dynamics

Module 2 provides tools for studying transcriptome dynamics, including changes in isoform usage and modulation of their absolute levels in order to quantify and estimate switching events in a case-control or along a time course analysis. Differential Expression (DE) analysis performs low-count isoform filtering and normalisation, provides expression values at the transcript, gene and CoDing sequence (CDS) levels. In addition, estimates post-transcriptional regulation by testing the Differential Usage of Isoforms (DIU), i.e. the changes in the relative abundance of isoforms derived from the same gene.

Module 3: Functional impact triggered by isoform regulation

Module 3 studies the contextual modelling of the functional effect triggered by differential isoform usage. To link functional diversity with isoform usage dynamics tappAS includes the Differential Feature Inclusion (DFI) analysis that profiles the dynamic change in positionally-annotated functional features content of full-length

isoforms, that are significantly altered or disrupted across isoforms in time and which modulate the functional and regulatory outcome of the gene in different experimental conditions. Additionally, alternative polyadenylation and UTR shortening/lengthening analysis were coupled to study the impact of isoform regulation on UTR modulation, key for transcript fate regulation. Finally, joint visualization of isoform expression levels and functional and structural elements allows the mapping of functional differences at transcript models.

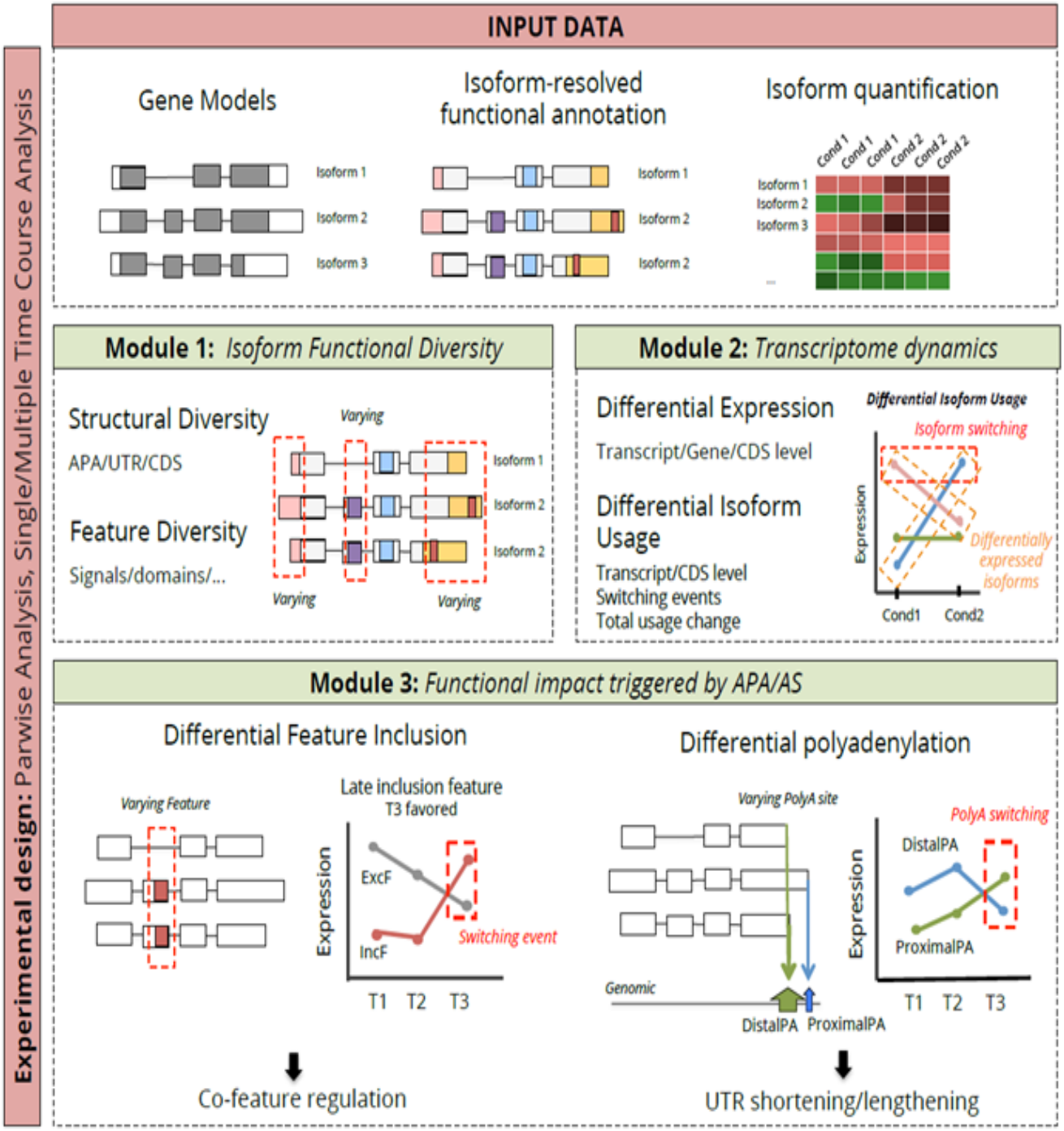


Figure 3. **Overview of tappAS modules for Functional Transcriptomics Analysis**. Three main pieces of input data are required: gene models, isoform expression and functional annotation at the isoform-resolution level. Methods included in the three analysis modules were adapted to work with both pairwise and time course experimental designs. Module 1 contains a novel qualitative approach to evaluate functional diversity of alternative isoforms. Module 2 implements Differential Expression and Differential Isoform Usage analyses. Module 3 measures the functional impact as changes in the inclusion of functional features, polyA site usage and UTR length.

5. Methodology

5.1. In silico analysis for the selection of gene candidates

In order to select candidate genes to perform the validation experiments, *in silico* analysis was conducted leveraging the analysis tools and information provided by Tappas.

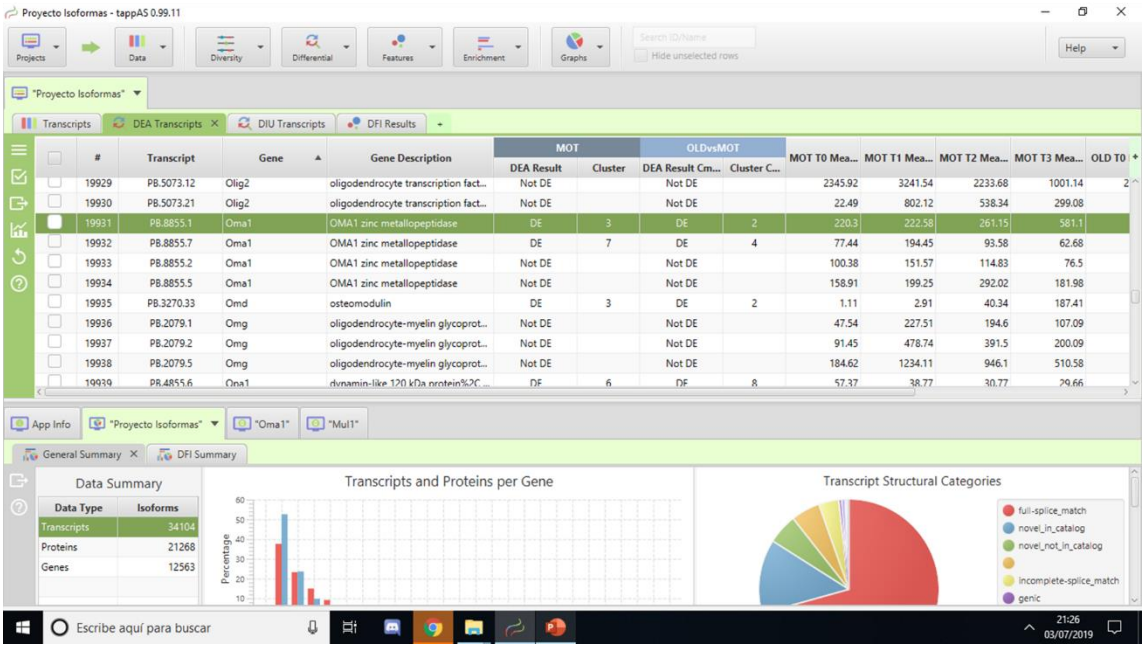
When transcript level expression matrix is introduced and the project created, Tappas provide a general view of the genes, transcripts and proteins identified and their expression level. First of all, the differential expression analysis (DEA) is run in order to find differentially expressed genes between the OPCs and the MTNs lineage (Figure 4 A). Thus, DEA performs statistical testing at the gene, transcript, or DCS level to determine if a given difference in read counts between both lineages is significant, making easier finding genes importantly involved in the differentiation process.

Then Differential Isoform Usage (DIU) is conducted to check for differential splicing at the gene transcript or protein level. In this case, DIU gives information about what genes presents transcripts expression variation along the differentiation process, enabling the identification of transcripts specially involved in one linage.

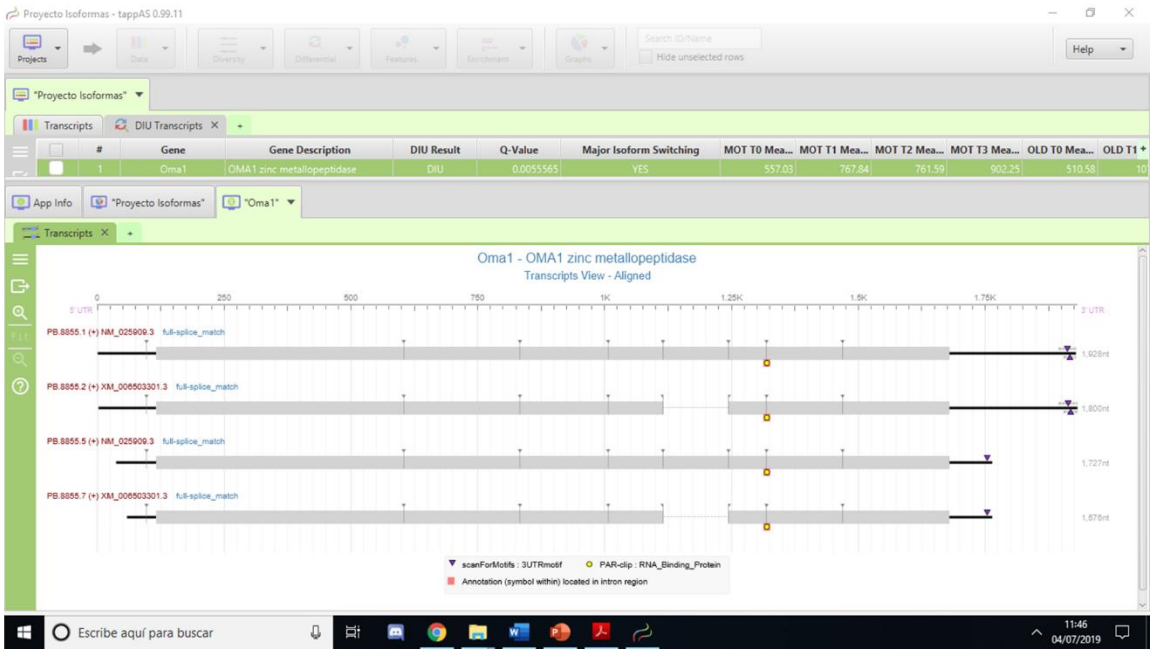
Finally, Differential Feature Inclusion Analysis (DFI) performs feature-level differentially splicing analysis with te considered features such as, Nuclear Location Signal (NLS) between isoforms. Thus, allows to find genes which present differential functional features among its isoforms.

Thus, with this analytical process, Tappas permits finding genes differentially expressed between both differentiations, whose isoforms present differential structural features and which expression vary along the differentiation process. Further, Tappas provides structural information at the genomic, transcript and protein level (Figure 4 B-C). Therefore, permits to check *AS* processes among the isoform transcripts and structural differences among the proteins encoded by those transcripts. Furthermore, since TappAS obtain resolved annotation from distinct databases, permits to differentiate between novel and already described variants.

A



B



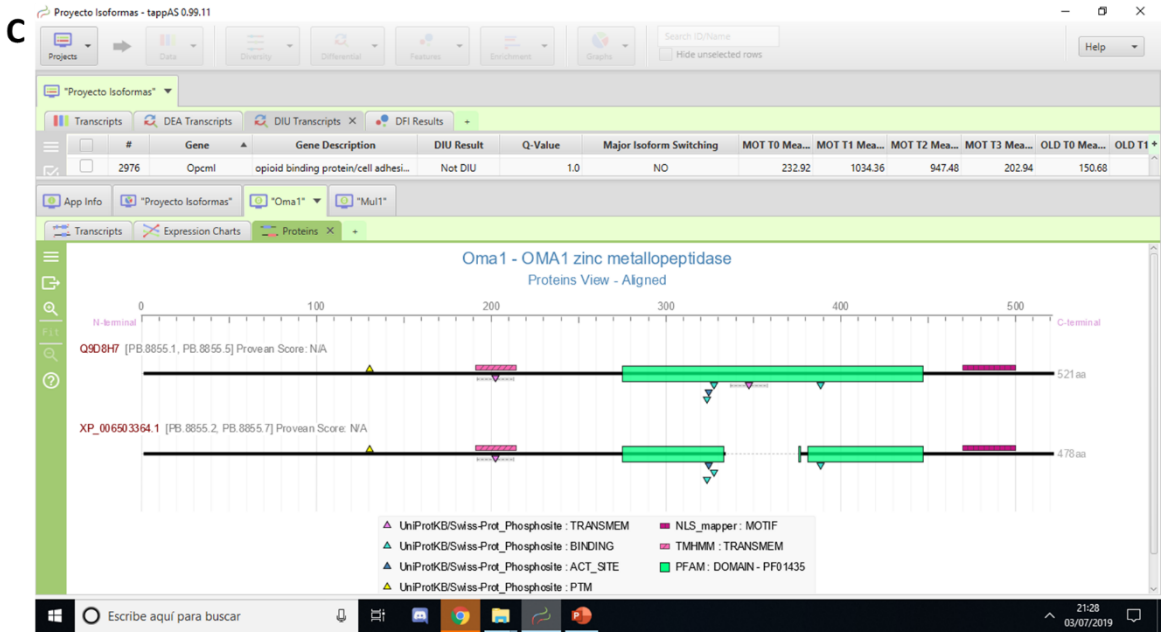


Figure 4. TappAS software Screenshots. A) DEA analysis provides statistical information about differentially expressed isoforms as well as expression values of genes isoforms in both differentiations. B) Transcripts View of OMA1. TappAS provides a view of the transcripts aligned and information about the skipped exons, the UTR lengths and microRNA and RNA-binding protein sites. C) Proteins are showed aligned with its functional fatures.

5.2. Establishment of the primary culture from neonatal mice and Neural Progenitor Cells (NPCs) differentiation induction to Oligodendrocyte Progenitor Cells (OPCs) and Motoneuron Progenitor Cells (MTNs)

Neonatal mice (4-7 days) were sacrificed by decapitation, the spinal cords were extracted as completely as possible and placed in washing medium (Gibco™ Dulbecco's Modified Eagle Medium (DMEM / F12 (Invitrogen) supplemented with 100 units/ml penicillin, 100 µg/ml streptomycin, 5 mM Hepes buffer, 0.125% NaHCO3, 0.09% glucose) on ice, under sterile conditions. The tissue is mechanically dissected and cleaned in successive centrifugation and washing steps. Finally, the cells homogenate are transferred to P60 plates treated with Neurocult™ Basal Medium (STEMCELL Technologies Inc.) supplemented with 100 units/ml penicillin, 100 µg/ml streptomycin and fibroblast growth factor 2 (bFGF, 20 ng/ul) and epidermal growth factor (EGF, 20 ng/ml) (Sigma), which allows the selection of Neural precursors cells (NPCs) that remain in suspension. After a few days the cells begin to aggregate forming neurospheres, indicating that they are proliferating. At this time, the NPCs are

transferred to P6 Ultra-Low Attach plates (ULA) and after some washes and centrifugations the debris is eliminated, and the NPCs proliferate. The cultures are maintained at 37°C in an incubator with 100% humidity and 5% O2. Cell differentiation is initiated following the protocols indicated below when an adequate number of starting cells is achieved (Figure 5). Cell cultures were examined every day in order to check their growth and exclude the presence of contamination. In addition, the culture media was changed twice a week with fresh factor.

5.2.1. Differentiation to oligodendrocyte precursors (OPCs)

The OPCs differentiation was induced following a previously described protocol (Keirstead et al. 2005). The NSCs were cultured with glial restriction medium (GRM) consisting of Gibco™ Dulbecco's Modified Eagle Medium (DMEM / F12, B27) (Invitrogen) supplemented with 25 μg/ml insulin, 6.3 ng/ml progesterone, 10 μg/ml putrescine, 50 ng/ml sodium selenite, 50 μg/ml holotransferrin, 40 ng/ml tri-iodothyroidine, supplemented with 6 ng/ml bFGF and 30 ng/ml EGF (Sigma, St. Louis, MO) for one day. Subsequently, they were incubated with 30 ng/ml of EGF, 4ng/ml FGF and 10 μM of trans retinoic acid (ATRA) for one week. Afterwards, the ATRA was removed and the cells were grown in GRM medium containing 30 ng/ml of EGF and 4 ng/ml of bFGF. On day 28 neurospheres were transferred to plates coated with Matrigel 1:20 for one week with GRM medium supplemented with 20 ng/ml of EGF and 4 ng/ml of bFGF where they were maintained until day 35. Cells were harvested at days 0 (T0), 9 (T1), 28 (T2) and 40-44 (T3) and stored at -80ºC in order to extract RNA and protein samples.

5.2.2. Differentiation to Motoneurons (MTNs)

The NSCs were incubated in flotation in Motoneuron Induction Medium (MIM) containing DMEM/F12 (Invitrogen), N2 supplement, heparin 2 μg/ml and P/S supplemented with bFGF (20 ng/ml) and ATRA (0.1 μM) for one week. The cells were induced to adhere by transferring them to ornithine/laminin-coated plates in Motoneuron medium (MM) containing Gibco™ Medium Neurobasal™ Medium (Thermo Fisher Scientific), N2 supplement, 1 μM cAMP, 0.1 μM ATRA and P/S supplemented with 200 ng/ml of human recombinant Sonic Hedgehog (rhSHH). Finally, the cells were incubated in MM medium supplemented with 10 ng/ml rh brain-derived neurotrophic factor (BDNF), rh glial cell-derived neurotrophic factor (GDNF),

rh insulin like growth factor (IGF-I) and 50 ng/ml of rhSHH where they remained until the end of the differentiation. Cell pellets were harvested at days 0 (T0), 9 (T1), 15 (T2) and 30-35 (T3) and stored at -80ºC in order to extract RNA and protein samples.

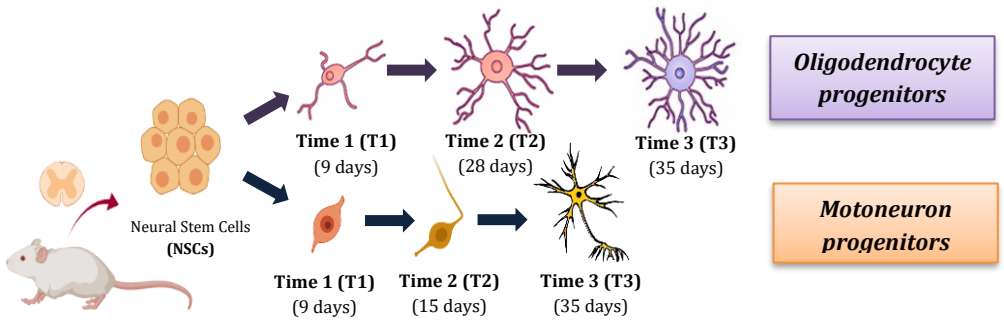


Figure 5. Scheme of the differentiations. EpSPCs extracted from mice spinal cords are isolated in order to obtain NSCs. Then the NSCs are committed to differentiate into two cell lineages, OPCs and MTNs.

5.3. Extraction of RNA, whole cell protein extract and cytoplasmic and nuclear protein extract.

Cell cultures were performed in order to obtain RNA and proteins sample for the experimental validation. RNA was extracted using the Nucleic acid and protein purification kit (Macherey-Nagel™) following the instructions provided by the fabricant. Cell samples stored at -80 ºC, were immediately immersed in liquid nitrogen and the buffer was added when they were still frozen to avoid RNA degradation. Afterwards, the RNA concentration was determined by the NanoDropTM 2000 (Thermo Fisher Scientific) and the integrity by the RNA screentape® (Agilent). RNA samples were stored at -80ºC until the RNA retrotranscription was performed in order to obtain the complementary DNA (cDNA).

Total protein fraction was extracted from cell cultures incubating the pellets on ice with lysis buffer containing 50 mM Tris-HCl, pH 7.5,150 mM NaCl, 0.02% NaN3, 0.1 SDS, 1% NP40, 1 mM EDTA, 2 mg/mL leupeptin, 2 mg/mL aprotinin, 1 mM PMSF, 1x Protease Inhibitor Cocktail (Roche Diagnostics, San Diego, CA, USA) for 30 minutes. Besides, vortex was performed every 10 minutes to prevent the deposition of the cells and ensure the proper mix with the lysis buffer. Then, 3 sonication cycles of 10 seconds were performed to ensure cell lysis. Finally, pellets were centrifuged at 12000 g and 4ºC for 15 minutes and the supernatants were collected.

Cytosolic and nuclear fractions were extracted the day when samples were collected, avoiding cell pellet congelation. Samples are washed in several washing cycles with PBS1X and centrifugations at 100 g for 5 minutes to remove the remaining cell culture medium. For the cytoplasmic protein extraction, cells were incubated in hypotonic buffer containing HEPES 10 mM pH 7.9, KCl 10 mM, EDTA 1 mM, EGTA 1 mM, DTT 1 mM, B-glycerophospate 10 mM and 1x Protease Inhibitor Cocktail (Roche Diagnostics, San Diego, CA, USA) for 10 minutes. During the incubation, the cells were vortexed every 5 minutes to avoid pellet precipitation. Afterwards, IGEPAL (CA-630) 0.4% was then added and samples were vigorously vortexed and centrifuged at 12000g at 4ºC for 5 minutes. Then the supernatants were collected as the cytoplasmic protein extracts and the remaining pellets were incubated in hypertonic extract buffer containing 10 mM TRIS pH 7.4, NaCl 400 mM, IGEPAL (CA-630) 0.5%, EDTA 1 mM, EGTA 1 mM,DTT 1 mM, B-glycerophospate 10 mM and 1x Protease Inhibitor Cocktail (Roche Diagnostics, San Diego, CA, USA) for 30 minutes and vortexed every 10 minutes, in order to destabilize the nuclear membrane and release the nuclear proteins. The mixes obtained were the nuclear protein extracts.

Finally, the concentration of protein in the samples was quantified by bicinchoninic acid technique (Pierce® BCA protein assay; Thermo Fisher Scientific) and the extracted proteins were used to assess the presence of the protein isoforms by WB.

5.4. Retrotranscription (RT) and Real time polymerase chain reaction (qPCR) development

5.4.1. Retrotranscription (RT) reaction

RT was performed using the High Capacity cDNA Reverse Transcription Kit ® (Applied Biosystems™). At the beginning cDNA was synthesized using 500 ng of total RNA but the amount of RNA was adapted to optimize the qPCR reaction and finally cDNA was synthesized from 150 ng, as it is explained later (Section 4.4.2). Once the reaction mixture (dNTPs, primers, reverse transcriptase and buffer) was prepared, the RT was carried out in an Applied Biosystems Veriti™ Thermal Cycler (Applied Biosystems™). The programme consisted in a preincubation step to activate the enzyme, a retrotranscription and a final step to denature the enzyme. When finished, the

complementary DNA (cDNA) was stored at -20ºC. The reaction volume and thermocycler programme were used as follow:

RT MIX	Volume (1x)
10X RT Buffer	2 (ul)
dNTP Mix (100 mM)	0.8
10X RT Random Primers	2
MultiScribe™ Reverse Transcriptase	1
Nuclease-free H2O	Up to 20 µl
RNA	*
Final Volume	20 µl

Settings	Step1	Step 2	Step 3	Step 4
Temperature	**25 ºC**	**37ºC**	**85ºC**	**4ºC**
Time	**10 minutes**	**120 minutes**	**5 minutes**	**Hold**

Table 1. A) RT mix. **The volume of RNA varied depending on the concentration of each sample. Nevertheless, the amount of initial RNA used was always the same (150 ng).* B) RT steps and temperatures settings.

5.4.2. Real time polymerase chain reaction (qPCR) development

Since the RNA sample was scarce and the isoforms were in a low proportion of the total NSCs RNA tests were performed to determine the adequate concentration of initial RNA to obtain evaluable Crossing threshold (Ct) values to determine the expression of the isoforms and save RNA in case it was needed in the future. For that purpose, qPCR test was performed with 500 ng of initial cDNA which was diluted (1/5, 1/10, 1/25 and 1/50), and was determined that a dilution 1/25 (~ 20 ng cDNA) was enough to obtain reliable Ct values. Furthermore, we tested different primer

concentrations to prevent non-specific amplification products. After the tests, we concluded the following working conditions for our primers:

- RNA concentration: 150 ng for the RT
- 1/5 dilution of the RT cDNA
- Primer concentration: 0.2 µM
- annealing temperature: 63ºC

Secondly, the amplification efficiency and specificity of the primer pairs were evaluated against NSCs RNA. The RT was performed from 150 ng of initial RNA and then serial dilutions of cDNA were carried out (1/5, 1/10, 1/25 and 1/50). Assay efficiency was automatically calculated from the slope of the regression line by plotting log cDNA concentrations against Ct values using the LightCycler® 480 software (ROCHE ©).

Finally, the reaction was carried out with the primers designed and the SYBR Green qPCR Green Master MIX (NZYTech, Lda. ©) in a LightCycler® 480 Instrument (ROCHE ©) using the mix and temperature programme indicated in table 3. Besides, all reactions were performed in triplicate, Negative controls were used to rule out the presence of contamination and PPiA as a housekeeping gene, since its expression remains considerably constant in the developmental CNS (Xu et al. 2018) and permits the correction of the deviations due to different sample concentrations.

qPCR MIX	**Volume (X1)**
NZY qPCR Green Master Mix (2x)	5 µl
0.2 µM forward primer	0.2 µl
0.2 µM reverse primer	0.2 µl
Template (cDNA + H2O)	1 µl
Nuclease-free H2O	Up to 10 µl
Final Volume	10 µl

Settings		Temperature	Time
Stage 1		94ºC	5 minutes
Stage 2	**Step1**	94ºC	30 seconds
	Step 2	63ºC	30 seconds
	Step 3	72ºC	30 seconds
Stage 3		4ºC	Hold

Table 2. A) qPCR Sybr green mix.it shows its components and the volume of each one for the reaction. B) qPCR steps and temperatures settings.

5.4.2.1. Analysis of the qPCR results

Instead of assuming the efficiency of the primers, amplification efficiency of the qPCR reaction was analysed by using the E-method (Tellmann 2006). This algorithm analyze the amplification efficiency of our primers by using serial dilutions of a single sample (undiluted, 1/5, 1/10, 1/25, 1/50). From this dilution the programme generates a standard curve and determine the efficiency values of either the target genes or the reference gene (housekeeping). The efficiency value represents the amount of DNA that is amplify in each amplification cycle. Thus, an efficiency of the hundred percent would mean that the amplification process is perfect, and DNA is doubling its amount each reaction cycle. In that case the efficiency of the process would be a value of 2. Assay efficiency was calculated from the slope of the regression line (Efficiency = 10(−1/slope) −1) by plotting log DNA concentrations against crossing threshold (Ct) value using the LightCycler® 480 software (ROCHE ©).

Analysis of the gene expression was performed using the $2^{(-DDCt)}$ method (Livak and Schmittgen 2001). Relative quantification describes the change in expression of the target gene relative to some reference group in this case cell culture mRNA expression at time zero (T0) of differentiation. The Ct values obtained for each target gene are normalized to an endogenous reference gene (the PPIA housekeeping gene) in order to correct results for differing amounts of input cDNA. Then the $2^{(-DDCt)}$ values from the target genes, are normalized to one of the target isoforms in order to obtain the relative

expression of all of them and compare the expression patterns between the different isoforms of the gene. OMA1 Cts values are normalized to OMA1-1 targets and in case of MUL the isoforms Cts values are normalized to MUL1-12 targets.

Experimental qPCR data was compared to the reference expression patterns obtained previously by the RNAseqsequencing provided by tappAS. For that purpose, since isoforms were not amplified independently, first the expected expression levels of the isoforms, provided by tappAS, were added together to compare them with the expression obtained with our primers. And the isoforms expressions were normalized to one of the isoforms to obtain their relative expression, as in our qPCR analysis.

5.6. Western Blotting

Protein expression was detected using the WB or “immunioblotting”. The equal amount of protein (17, 20 or 30 µg/well, depending on the gel) were mixed with loading buffer (7.5 µl). Then samples were boiled at 95ºC for 5 minutes and separated in 10% SDS-polyacrylamide gels. Afterwards, proteins were wet transferred with a transfer buffer containing Tris 14.4 g Glicine 3.03 g and 20% methanol to PVDF membranes (Thermo Fisher scientific Inc) in a Mini Trans-Blot® Cell (Bio-Rad Laboratories, Inc). Membranes were blocked with 5% non-fat milk in TBS with 0.1% Tween-20 for 1 hour at room temperature then incubated overnight at 4ºC with specific antibodies against OMA1 (Santa Cruz Biotechnology Inc, https://www.scbt.com), CAPER (Rbm 39) (Santa Cruz Biotechnology Inc, https://www.scbt.com) and RUFY (RIPX) (Thermo Fisher scientific Inc, https://www.thermofisher.com) at 1:500 dilution in 5% Bovine Serum Albumin (BSA) in TBS with 0.1% Tween-20. Subsequently, membranes were incubated with anti-mouse or anti-rabbit horseradish peroxidase-conjugated secondary antibodies (1:2500 in BSA 5%) for 1 hour at room temperature. Blots were visualized with the ECL (Thermo Fisher scientific Inc) detection system in an Alliance Q9 Advanced (Uvitec Cambridge Inc). The relative protein expression was quantified by Image Studio Lite software.

5.7. Design of specific primers for qPCR validation

5.7.1. Design of specific primers for OMA1 isoforms

Once the candidates were chosen, the next step was the design of primer pairs that hybridized specifically with the desired isoforms so that the expression could be measured and quantified by qPCR. First of all, an alignment was performed with the transcripts isoforms sequences obtained by PacBio making use of online softwares for the multiple alignment of protein and nucleotide sequences such as MUSCLE (Multiple Sequence Comparison by Log-Expectation, EMBL-EBI) (EMBL-EBI) and Multalign (Multalin interface page). The alignment enabled the visualization of common and differential sequences among the isoforms, where primers would be targeted in order to amplify specific isoforms (**Figure 6 A**). Primers were designed manually and then the sequences were checked in the Basic Local Alignment Tool (BLAST NCBI-NIH) in order to design the most specific primers as possible preventing non-specific targets and taking the melting temperature and the GC content into account, so that our primers worked properly in consonance with the SYBR Green. In addition, the formation of primer dimer and hairpins was checked using NetPrimer (PREMIER Biosoft-NetPrimer). Although many control processes were conducted, designing primers for specific isoforms was such a tough duty and sometimes it was impossible to design specific primers to one isoform. Therefore, instead of primers specific for one isoform, finally primers were designed for transcripts isoforms which shared common features and codified for the same proteins.

OMA1 primers were designed leveraging the exon 6 skipping (**Figure 6 C**). Thus, OMA1 first primer pair (OMA1-1) was designed against transcripts PB8855.1 and PB8855.5, which conserved the exon 6 and encoded for the complete protein Q9D8H7. On the other hand, the second primer pair (OMA1-2) against PB8855.2 and PB8855.8, which lacked the exon 6 and encoded for a novel protein (XP_006503354.1) that loses a transmembrane domain. In this case the objective was to amplify separately those isoforms which kept the exon 6 and those which lacked that exon.

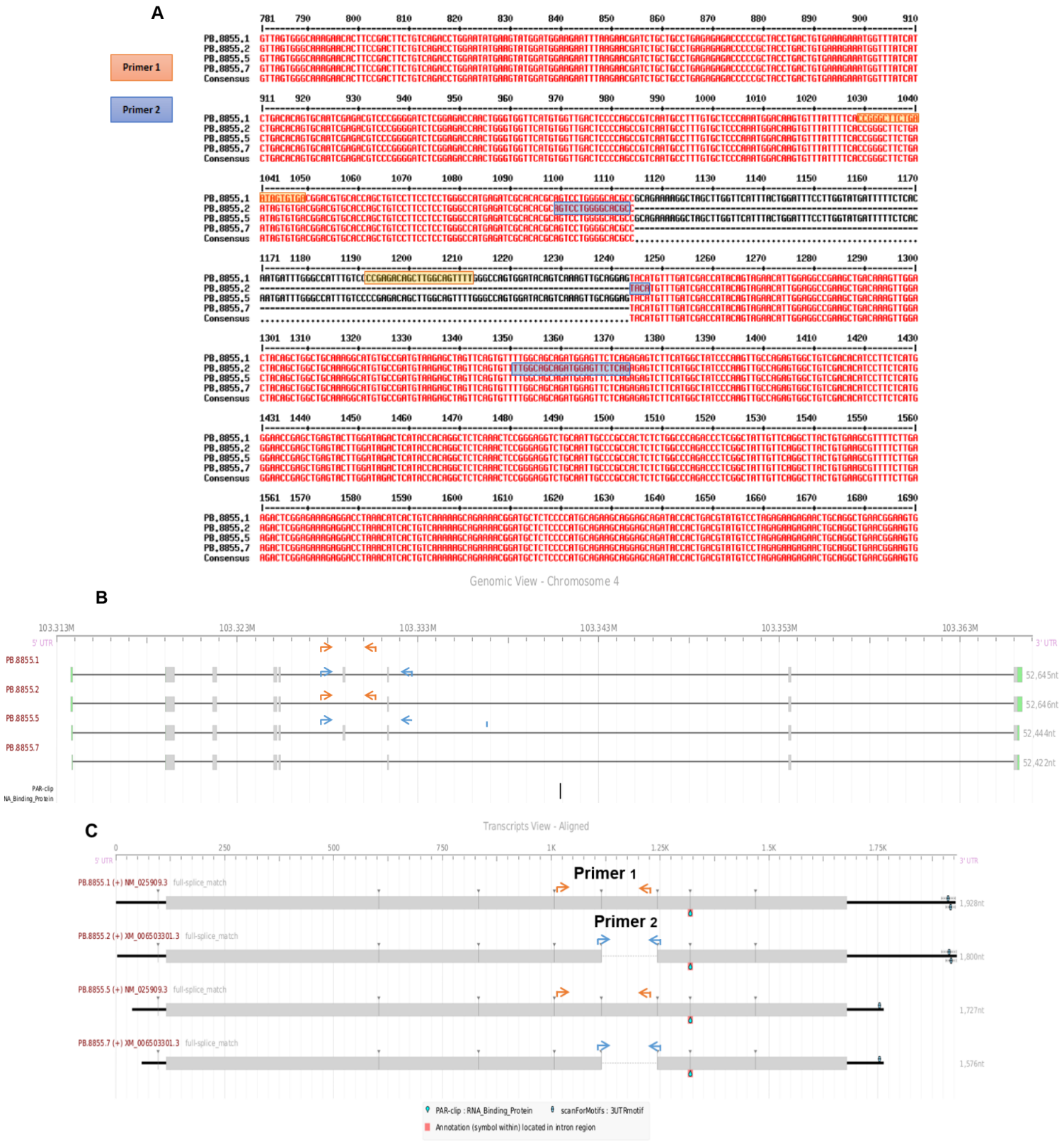
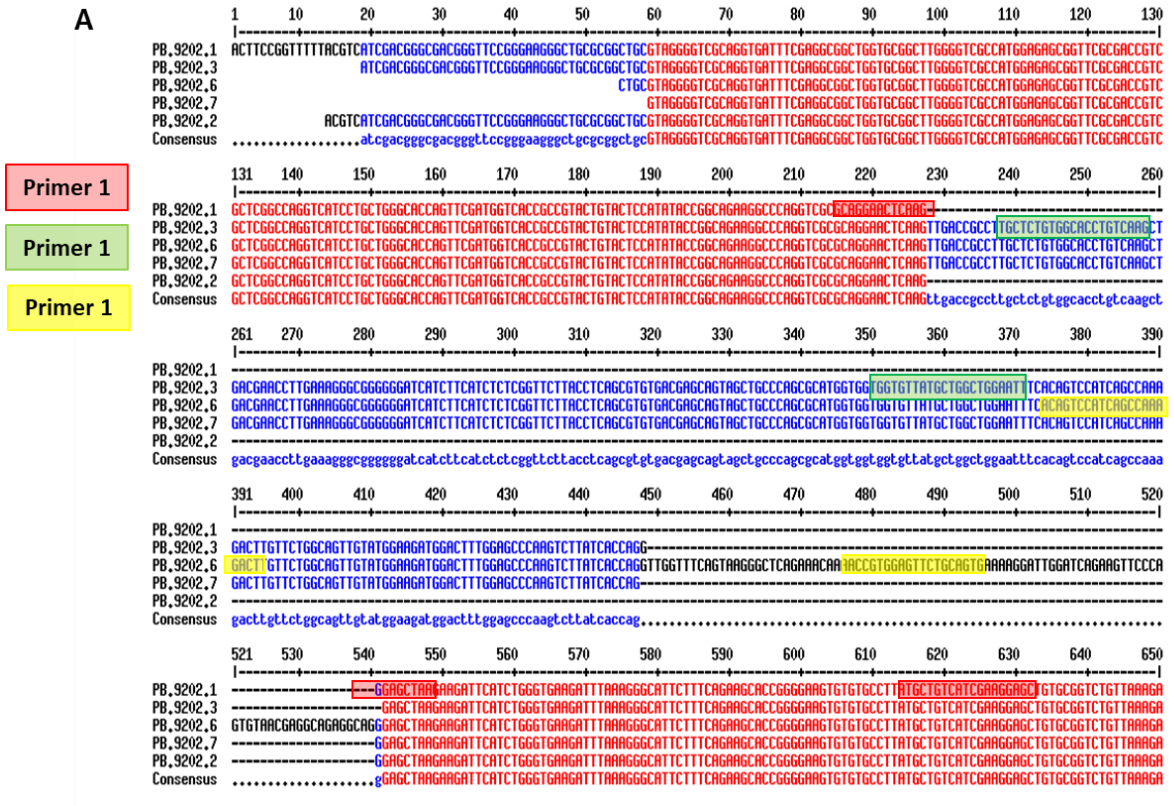


Figure 6. OMA1 visualization. A) Alignment of OMA1 isoforms with MUSCLE. In red are marked the common sequence to all the isoforms and in black those differentially expressed. The primers are highlighted where they targeted the transcripts. The positions where the primers hybridize are also marked with arrows in the genomic view B) of the gene and in the transcripts view C).

5.7.2. Design of specific primers for MUL1 isoforms

Mul1 primers design was very similar (Figure 7). The first primer pair (MUL1-12) was targeted against transcripts PB.9202.1 and PB.9202.2, that presented exon 2 skipping and encoded for the complete protein Q8VCM5. The third pair (MUL1-6) targeted specifically the novel not in catalogue transcript PB.9202.6 which according to our database conserved exon 2 and codified for a novel protein (novelProt313) which lacked the transmembrane domain at the N-terminal domain, as Q8VCM5-3. The second pair (MUL1-367) was against PB.9202.7, PB.9202.3 which presented incomplete exon 2 skipping and encoded a protein that lacked the transmembrane domain Q8VCM5-3 and against PB.9202.6 because it was not possible to avoid its amplification. Nevertheless, in this case as with OMA1, primers amplified set of transcripts which encoded for the same protein.



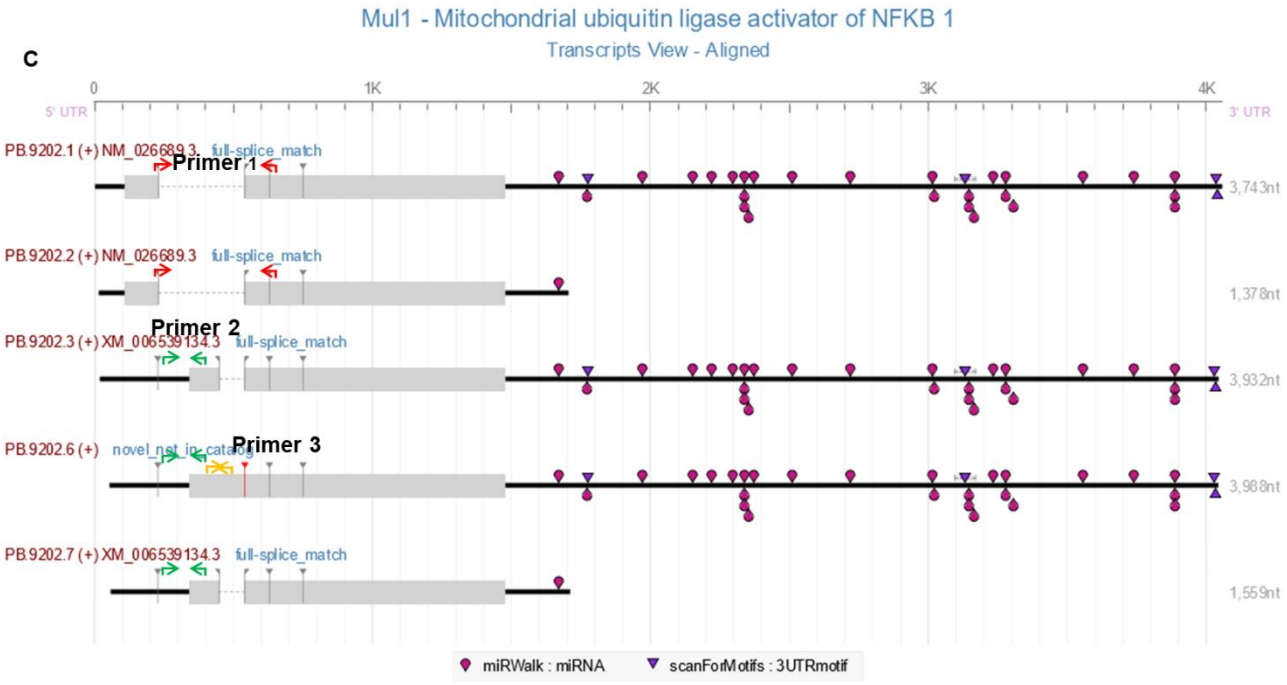


Figure 7. MUL1 visualization. A) Alignment of MUL1 isoforms with MUSCLE. in red are marked the common sequence to all the isoforms and in black those differentially expressed. The primers are highlighted where they target the transcript. B) The positions where the primers hybridize are also marked with arrows in the genomic view and in the transcripts view of the gene.

Gene	***Target Isoforms***	***Primer Forward***	***Primer Reverse***	***Product lenght (bp)***
OMA1	IS-1 IS-5	GGCTTCTGAATAGTGTGACGGA	AAACTGCCAAGCTGTCTCGG	180
	IS-2 IS-7	GTCCTGGGGCACGCCTACAT	CTGAGAACTCCATCTGCTGCCAA	146
MUL1	IS-1 IS-2	GCAGGAACTCAAGGGAGCTAA	AGCTCCTTCGATGACAGCAT	106
	IS-3 IS-7 IS-6	TGCTCTGTGGCACCTGTCAAG	AATTCCAGCCAGCATAACACCA	134
	IS-6	ACAGTCCATCAGCCAAAGACT	CACTGCAGAACTCCACGGTT	122

Table 3. Designed primers against OMA1 and MUL1 for qPCR analysis. The table shows the target isoforms for each primer pair, the Forward and Reverse primers sequences, and the predicted length of the qPCR product.

5.8. Statistical analysis

The statistical analysis of the expression values was performed by the 2way ANOVA. All tests were performed with prism Graph Pad. All data are presented as mean ± SD of at least three replicates

6. Results

6.1. Candidates selection

Taking advantage of the analysis previously showed in section 5.1, the next candidate genes were selected for its experimental validation.

6.1.1. OMA1

OMA1 is a metalloprotease located in the inner membrane of the mitochondria, where it binds through a transmembrane domain. OMA1 is involved in mitochondrial dynamics and fission-fusion processes, in order to adapt the metabolism rate to the needs. OMA1 is translated as an inactive pre-pro protein of 60 kDa which is processed to produce another form (pro-OMA1) of 40 kDa (Baker et al. 2014). Under stress conditions such as dissipation of Δφ, ROS production or decreased ATP level, pro-OMA1 activates and controls fusion-fission mitochondrial processes and is involved in the metabolic capability of the cells (Consolato et al. 2018).

According to our TappAS, OMA1 gene encodes for 4 transcript isoforms (Figure 8B), which transcription starts at ATTS. However, the most interesting thing is that due to a due to an *AS* process of exon skipping (exon 6 is skipped in some isoforms), two transcript isoforms (PB. 8855.1 and PB. 8855.5) present the exon 6 and another two (PB. 8855.2 and PB. 8855.7) present exon 6 skipping affecting the structure of the translated protein. Thus, the complete isoforms synthesize a protein with a transmembrane domain (Q9D8H7), and the isoforms which lack exon 6 lead to the synthesis of a predicted protein lacking the transmembrane domain (XP_006503364.1) (Figure 8 C).

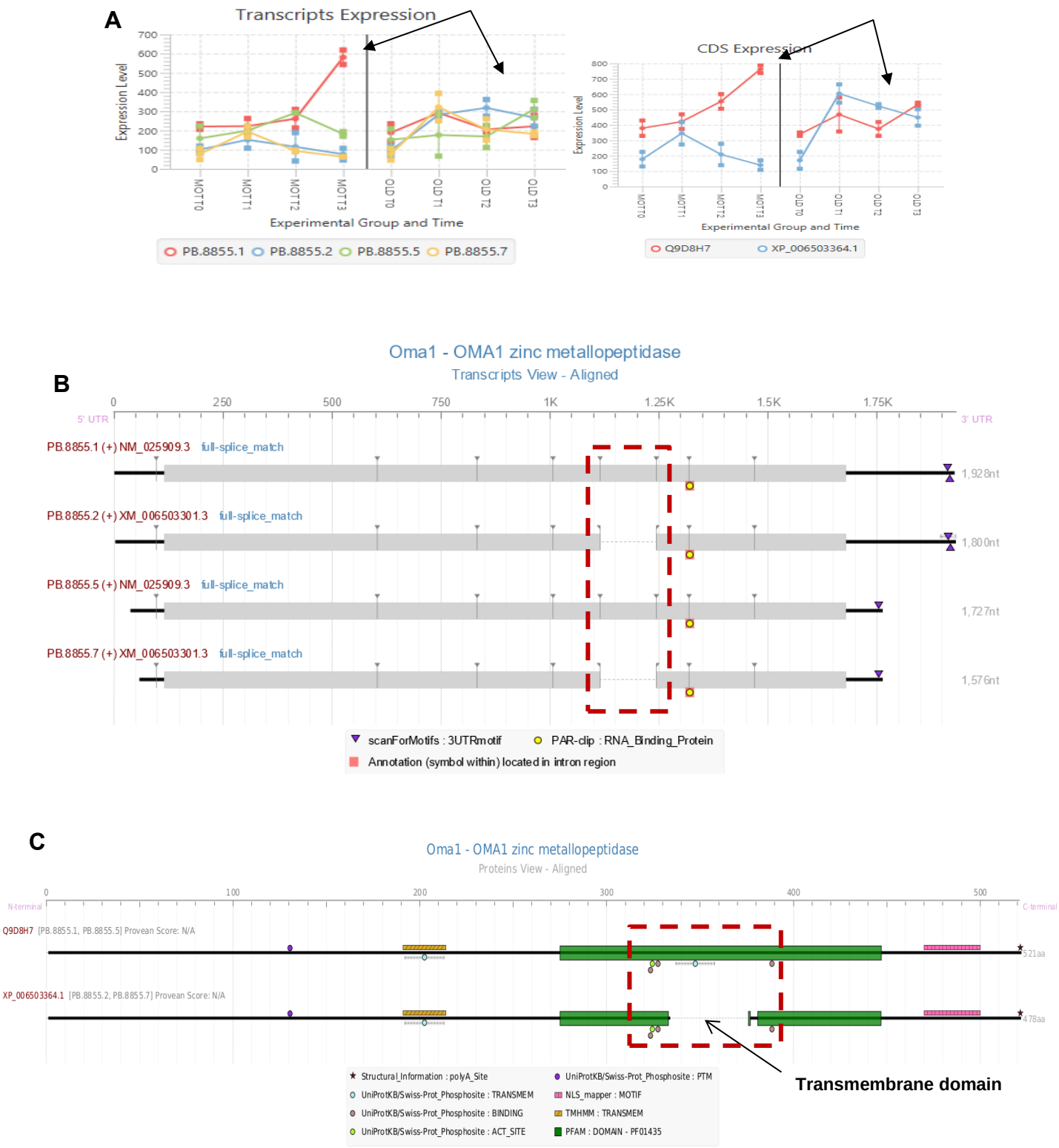


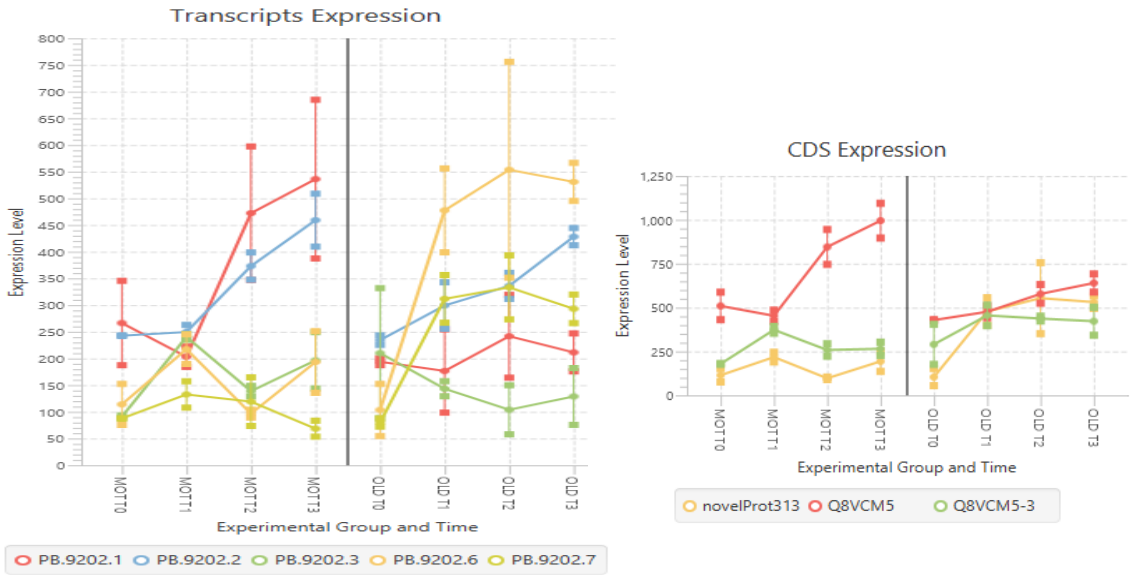
Figure 8. OMA1 information provided by TappAS. A) Show the expression levels of the transcripts and proteins isoforms. The arrows point out the differential expression patters observed in OPCs and MTN lineages. B) Transcripts structure. The square remark the exon 6 skipping C) In the proteins the square highlights the difference in the transmembrane domain.

6.1.2. MUL1

Mitochondrial ubiquitin ligase 1 (MUL1) present E3 ubiquitin ligase activity (Neuspiel et al. 2008), although in mammals it shows a ubiquitin-like modifier (SUMO) ligase activity (Braschi et al. 2009). Plays a role in the control of mitochondrial morphology and influences mitochondrial localization promoting mitochondrial fission either regulating negatively fusion promoters or activating fission proteins (Yun et al. 2014; Neuspiel et al. 2008).

MUL1 isoforms were differentially expressed in the lineages (Figure 9 A) and given the exon 2 skipping, these transcripts encode for 3 different proteins (Q8VCM5, Q8VCM5-3 and novelProt313) which vary in their structure. PB.9202.1 and PB.9202.2 present complete skipping of the exon 2 and encode for the canonical protein, Q8VCM5 that presents two transmembrane domains. PB.9202.3 and PB.9202.7 present partial skipping, but still transcribing part of the exon 2, encode for Q8VCM5-3. Finally, the novel isoform not in catalogue PB.9202.6 which present the exon 2 completely, encodes for the novelProt313. Both Q8VCM5-3 and novelProt313 lose the transmembrane domain at the N-terminal domain, and just conserve the transmembrane domain of the C-terminal domain.

A



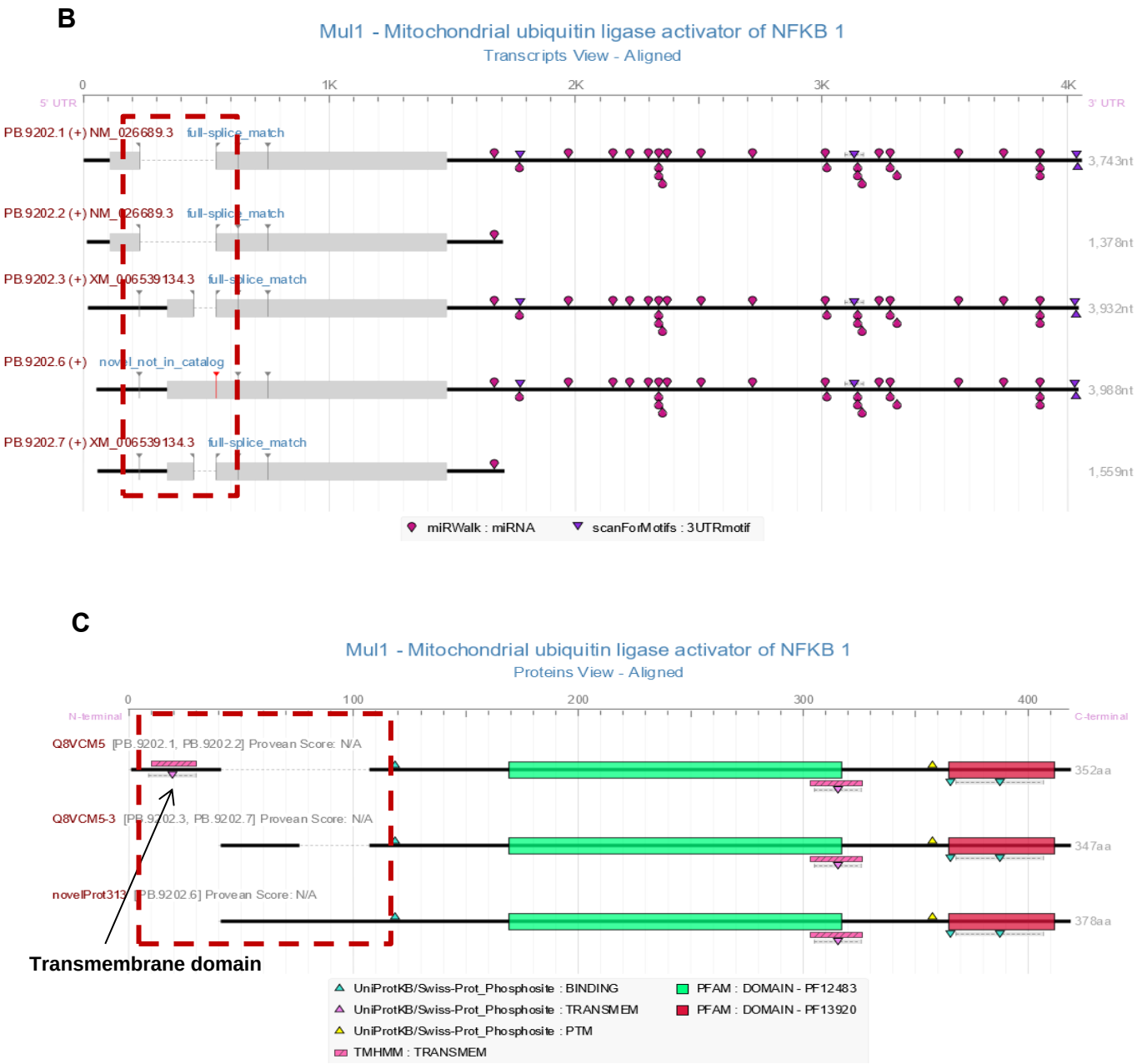


Figure 9. MUL1 information provided by TappAS. A) Expression levels of the transcripts and proteins isoforms at different points of the differentiation. B) Transcripts isoforms. The square highlights the exon 2 skipping. C) Proteins encoded by Mul1. The square pints out the featural differences in the N-terminal domain.

6.1.3. RUFY3 (RIPX)

RUN and FYVE domain–containing 3 (RUFY3), also known as RIPX, is an adapter for the union of GTPases that play roles in important cellular processes. It has been described that RUFY3 binds to Rap2 (a GTP protease) and form a stable complex which seems to have relevant functions in the formation of the cone growth and

therefore in neurite outgrowth (Honda et al. 2017). In addition, RUFY3 interacts with actin filament binding proteins influencing cell polarization and neurite outgrowth (Honda et al. 2017; Wei et al. 2014). According to tappAS, RUFY gene encode for four structurally similar proteins, three of them previously described Q9D394, Q9D394-4 and A0A0G2JFT8 and XP-006535175.1 (Figure 10). Due to distinct *AS* processing, the isoforms present skipped exons and differential UTRs, as well as ATTS (*Supplementary material*). Therefore, synthetized proteins differ in the structure of the N-terminal and C-terminal domain, which are annotated as disorder domains, what means that present alternative three-dimensional structures that can be useful for the interaction with other proteins.

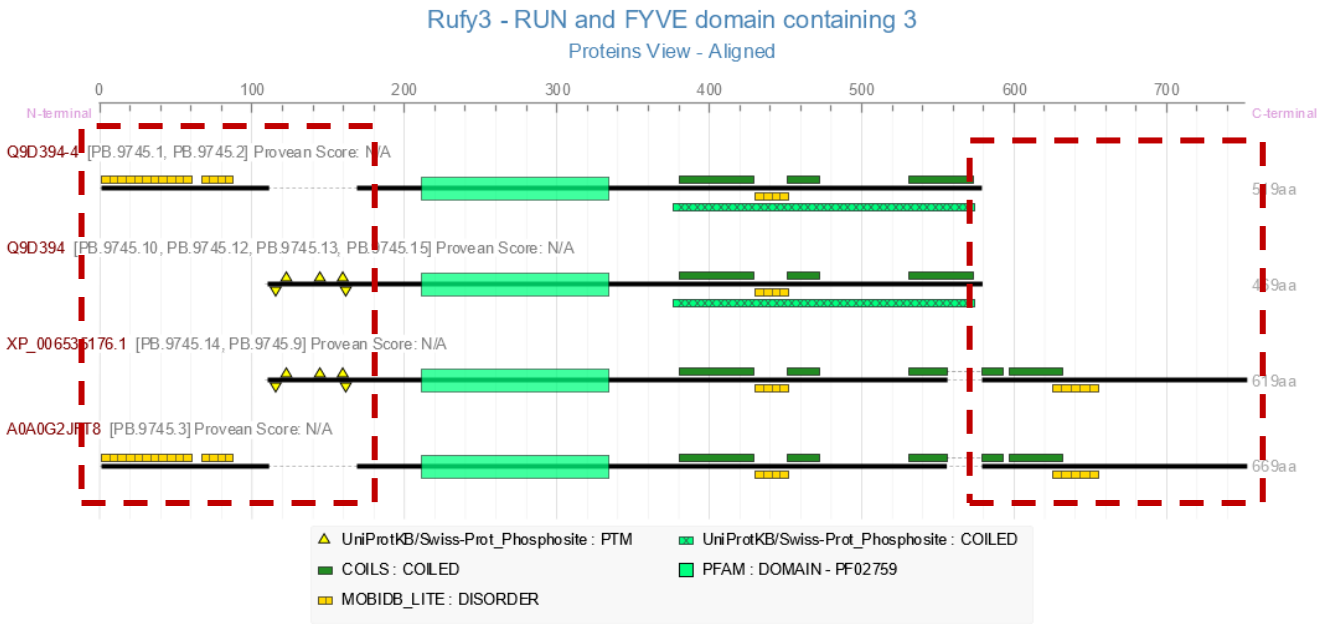


Figure 10. RUFY3 protein information. The squares highlights N-terminal and C-terminal domains, where the proteins differ in the presence of disorder domains.

6.1.4. RBM39 (CAPER)

RBM39 (CAPER) is a serine/arginine-rich (SR-rich) RNA-binding protein, involved in the *AS* and therefore acting as a splicing factor to generate proteome complexity (Mai et al. 2016). According to our data, RMB39 encodes for thirteen different transcripts and four different proteins (Figure 11). Among these proteins two of them present a similar structure (Q8VH51and Q8VH51-2), weight and both conserve a Nuclear Localisation Signal (NLS) and a Disorder motive. Since *AS* produce transcripts which

differ in the exon composition and in the UTR, the other two proteins (Q8BH51-3 XP_011237624.1) lack that N-terminal domain and therefore the NLS domain. The structural differences between the isoforms, specially the differential features (NLS) found in these proteins, made this candidate useful to identify the presence of the isoforms by WB analysis. However, it was not an ideal candidate for the analysis by qPCR, because it presented too transcripts and was difficult to design specific primers for all of them.

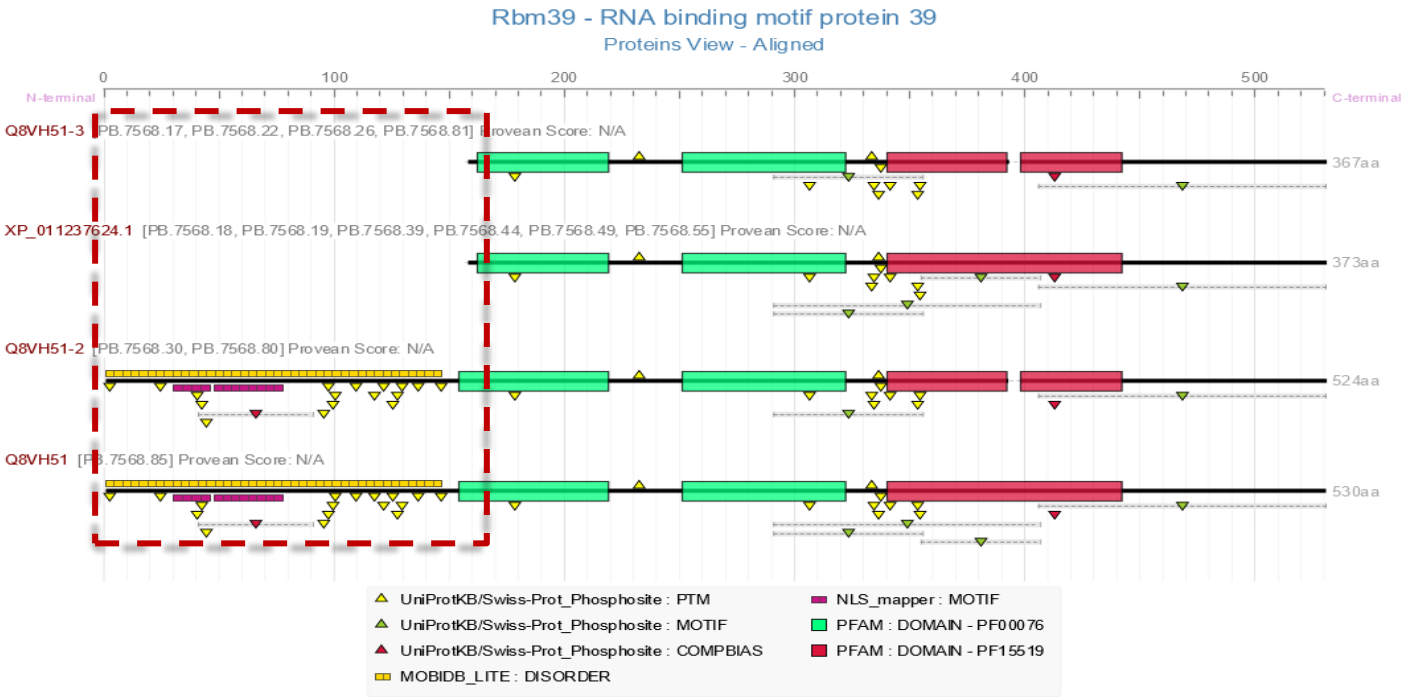


Figure 11. MUL1 information provided by TappAS Tool. A) Expression levels of the transcripts and proteins isoforms at different points of the differentiation. The arrows point out the differential expression patters observed in OPCs and MTN lineages.

6.2. Gene expression validation

6.2.1. Expression analysis of OMA1 isoforms in OPCs and MTNs lineages

As it is previously commented for the amplification of OMA1 isoforms, primer OMA1-1 was targeted against PB8855.1 and PB8855.5 transcripts and primer OMA1-2 against PB. 8855.2 and PB. 8855.7 transcripts, which presented skipping of exon 6.

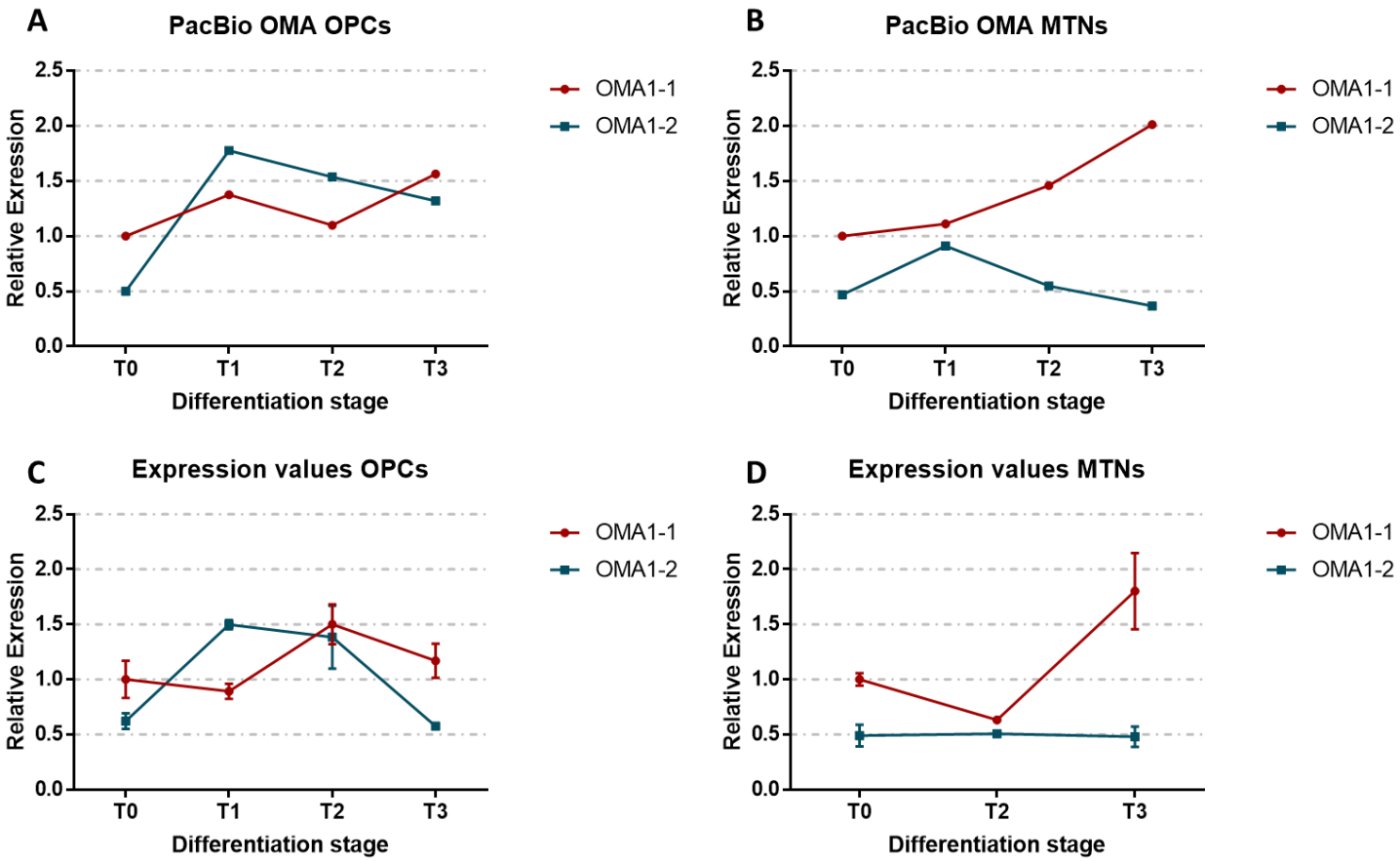


Figure 12. OMA1 isoforms expression analysis in the OPCs lineage. Left panels (A-C) show the OPCs expression values, and the right panels the MTNs expression values (B-D). The upper panels (A-B) show the expected expression values provided by tappAS and lower panels (C-D) show the experimental values. According to our data, OMA1 isoforms present the expected expression patterns along the differentiations. Although there is no significant difference, there is a clear differential pattern between OPCs and MTNs in the expression of each pair of isoforms. Relative expressions have been obtained by the $2^{(-DDCt)}$ method and expression levels have been normalised to PPIA and then to OMA1-1. Values are represented as mean ± SD. qPCR was performed in triplicate.

As the panel shows data obtained experimentally both in OPCs and MTNs lineages (**Figure 12 C-D**), apart from minor variations, present similar expression patterns distribution compared with the expected data obtained by PacBio sequencing (**Figure 12 A-B**). Further, major and minor isoforms conserve their position, i.e. there is no unexpected switching at the beginning and final stages of differentiation.

Interestingly, OPCs expression values (**Figure12 C**) present different distribution in comparison with the MTNs lineage (**Figure 12 D**) along the differentiation. In the case of the OPCs, isoforms PB8855.1 and PB8855.5 (primer OMA1-1) are expressed preferentially at T0, T1 and T3. Nevertheless, expression values of PB8855.2 and PB8855.7 are quite superior at T1 and very similar at T2. Therefore, although PB8855.1 and PB8855.5 seem to be predominant at the initial and final stages, it is likely that the other variants (PB. 8855.2 and PB. 8855.7) might be involved, at least partially in the cell fate determination into OPCs. In MTNs lineage PB8855.1 and PB8855.5 expression is superior along the whole differentiation, and the difference of expression even increase at T3.

6.2.2. Expression values of MUL1 isoforms in OPCs and MTNs lineages

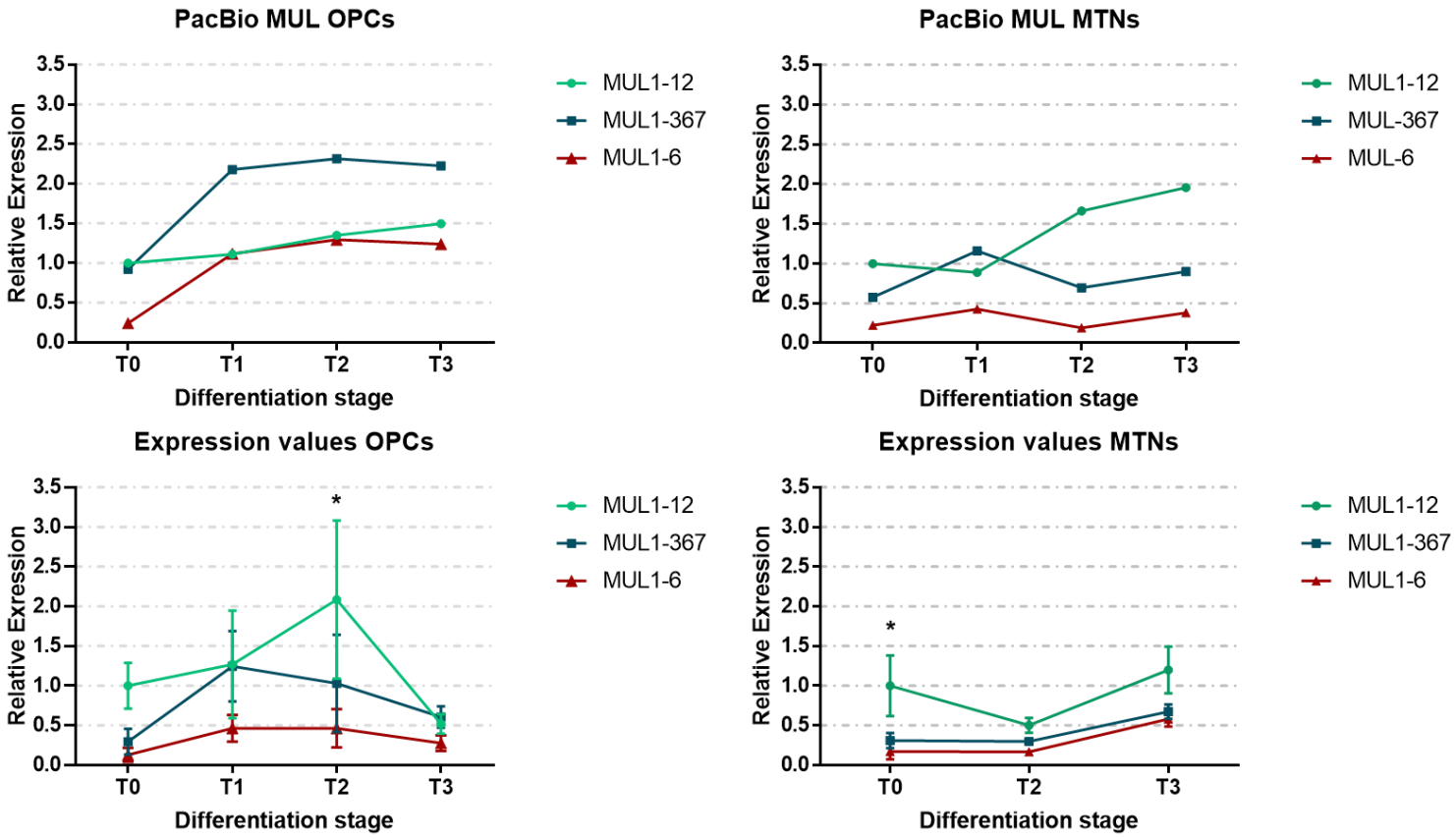


Figure 13. MUL1 expression analysis in OPCs differentiation lineage. Left panels show the OPCs expression values, and the right panels the MTNs expression values. The upper panels show the expected expression values provided by tappAS and lower panels show the experimental values. MUL1 present different isoform expression patterns along the OPCs lineage, but remains relatively constant in the MTNs differentiation, compared with the PacBio

data. Relative expressions have been obtained by the $2^{(-DDCt)}$ method and expression levels have been normalised to PPIA and then to either OMA1-1 and values are represented as mean ± SD. qPCR was performed in triplicate.

Primer MUL1-12 was targeted against transcripts PB.9202.1 and PB.9202.2, that present exon 2 skipping; MUL1-367 against PB.9202.7, PB.9202.3 which present incomplete exon 2 skipping and PB.9202.6, which conserves exon 2; and MUL1-6 against PB.9202.6.

On the one hand, MUL1 isoforms expression present an increase variation in the OPCs differentiation (**Figure 13 C**), compared to the expected data (**Figure 13 A**), even finding significant differences between MUL-12 and MUL1-6 expression values (**Figure 13 C**). Nevertheless, although there is a switch in the predominant isoforms at T1 and T2 of the OPCs lineage, MUL1-12 isoforms remain being predominant at T0 and MUL1-367 at T3. Therefore, variations obtained are observed only at the middle points, but the expected expression patterns are conserved at the first and final stage.

On the other hand, expression patterns are conserved along the whole differentiation stages of the MTNs lineage (**Figure 13 B-D**). In addition, significant differences are found between MUL-12 and MUL1-6 expression values at T0 (**Figure 13 D**), supporting the expression expected.

Remarkably, beyond the expression patterns, PB.9202.6 has been detected and amplified with the designed primer pair (MUL1-6). It should be highlighted, since it is a novel isoform not included previously in the catalogue that proofs the potential of the new method.

6.3. Analysis of the protein expression

6.3.1. Protein expression analysis of OMA1 in OPCs differentiation

To analyse the changes in the expression of OMA1 along the differentiation blotting was performed with OPCs lineage protein samples and a spinal cord protein homogenate as positive control (C+). For that purpose, and since OMA1 presented an NLS it was performed with the total protein fraction and the nuclear and cytoplasmic

faction separately, in case there was any difference in the distribution of the proteins in the cell.

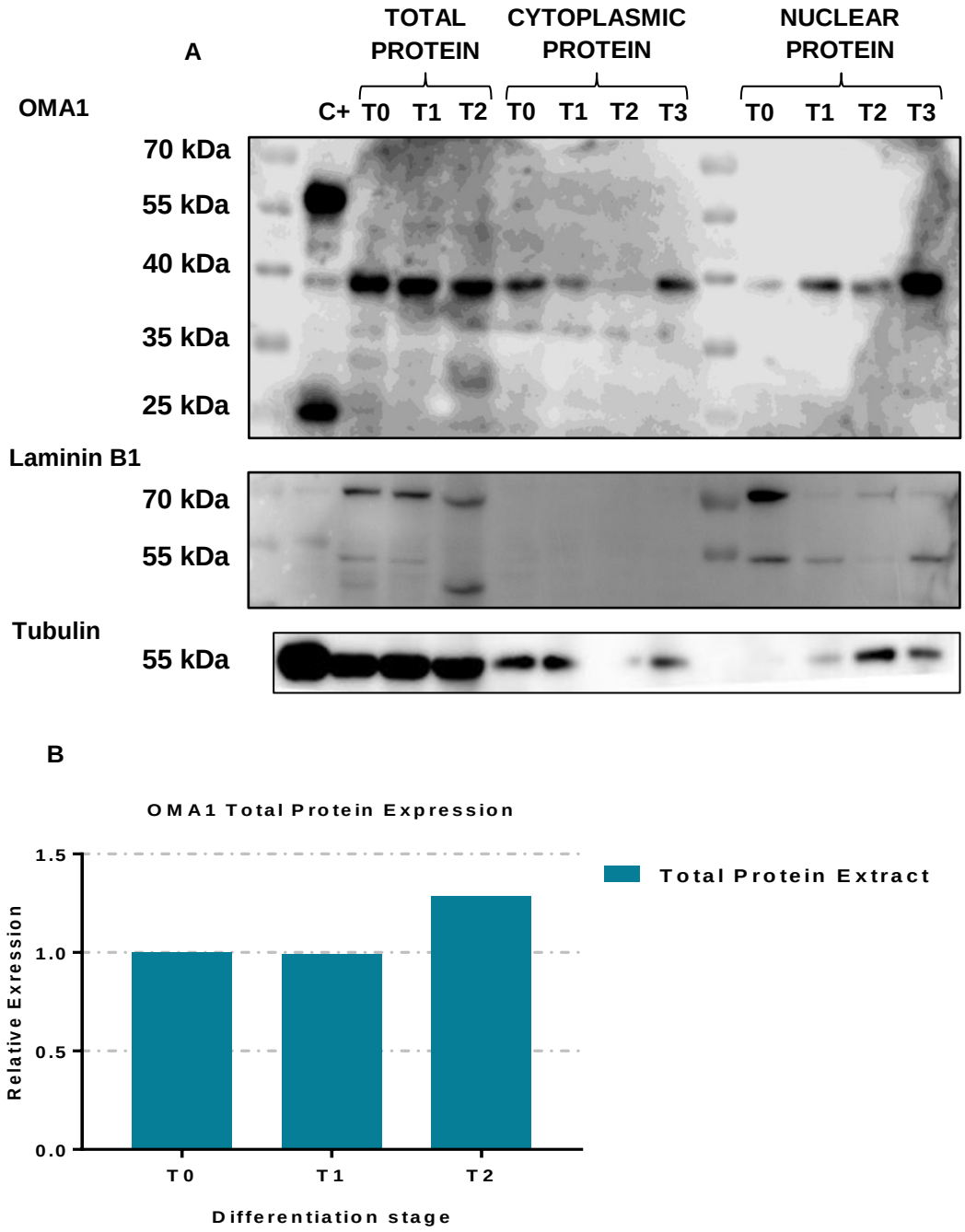


Figure 14. OMA1 protein expression in OPCs lineage. The left part of the membrane corresponds to the total protein fraction, the middle part to the cytoplasmic fraction and the right to the nuclear fraction. Spinal cord homogenate was used as C+. A) The inferior image shows the membrane with OMA1 signals, the middle image shows Laminin B1 (70-55 kDa) bands and the lower correspond to Tubulin (55 kDa). B) Quantification of OMA1 protein expression in the total protein fraction along the OPCs lineage. C) Quantification of OMA1 protein expression in the cytoplasmic and nuclear fraction of proteins.

According to the blotting results (**Figure 14 A**), in the C+ line is found the pre-pro-OMA1 inactive protein Q9D8H7 (~ 60 kiloDalton (kDa)), the active form (~ 40 kDa), and what could be the cleaved part (~ 25 kDa). Nevertheless, either in the cytoplasmic or the nuclear fraction only the active form is detected (~ 40 kDa). Further, there is no detected signal of the expected novel protein XP_006503364.1 (~ 55 kDa), maybe because the antibody does not recognise this novel protein.

In the cytoplasmic fraction at T2 there is no apparent signal neither of OMA1 (**Figure 14 A**), nor tubulin (**Figure 14 A**). Additionally, cytoplasmic fraction is free of nuclear contamination, but the nuclear fraction extracts present contamination of the cytoplasm fraction, as the tubulin shows (**Figure 14 A**). Total protein fraction was quantified to obtain relative expression of OMA1 along the OPCs differentiation. As the plot shows, the relative expression of OMA1 in the total protein fraction show a reduction at T1 and an increment at T2 (**Figure 14 B**).

Since loading control of the cytoplasmic (Tubulin) and nuclear fraction (Laminin B1) are not at the same proportion, relative quantification was not performed (**Figure 14 A**). Besides, the nuclear fraction is positive to tubulin, indicating that presents cytoplasmic contamination, so quantification would not be reliable. However, in the cytoplasmic fraction OMA1 expression seems to be higher at T0 and T3, while a weak band is detected at T1, while in the nuclear fraction signal at T0 is barley detected and then rises at T1 and T3. Although it cannot be confirmed, it seems that the nuclear OMA1 protein increases when the cytoplasmic fraction is reduced at T1. In addition, it seems that OMA1 presents a tendency to increase, since the strongest signal is detected at T3 both in the cytoplasmic and nuclear fractions. Since, this experiment is not sufficient, another differentiation is being conducted in order to make further OMA1 expression experiments in the OPC lineage.

6.3.2. Protein expression analysis of RUFY3 in NSCs

According to tappAS, RUFY gene encode for four proteins, three of them previously described (Q9D394, Q9D394-4 and A0A0G2JFT8) and a novel protein (XP-006535175.1) that structurally differ, especially in their C-terminal sequence. To assess the presence of the isoforms in our model, WB was performed with total protein extract from NSCs at T0 and a spinal cord protein homogenate as positive control

(**Figure 15**). Our results showed that in the NSCs lysate at T0 we could detect a band slightly lower than 55 (kDa) that corresponds to Q9D394, another slightly higher than 55 kDa that corresponds to Q9D394-4, one at 70 kDa that coincides with the novel protein XP-006535175.1 and higher than 70 kDa that coincides with A0A0G2JFT8. According to the expected sizes, band signals obtained in the WB confirm that all the isoforms are present in our system, even the new protein (XP-006535175.1) supporting the effectiveness of Tappas predictions.

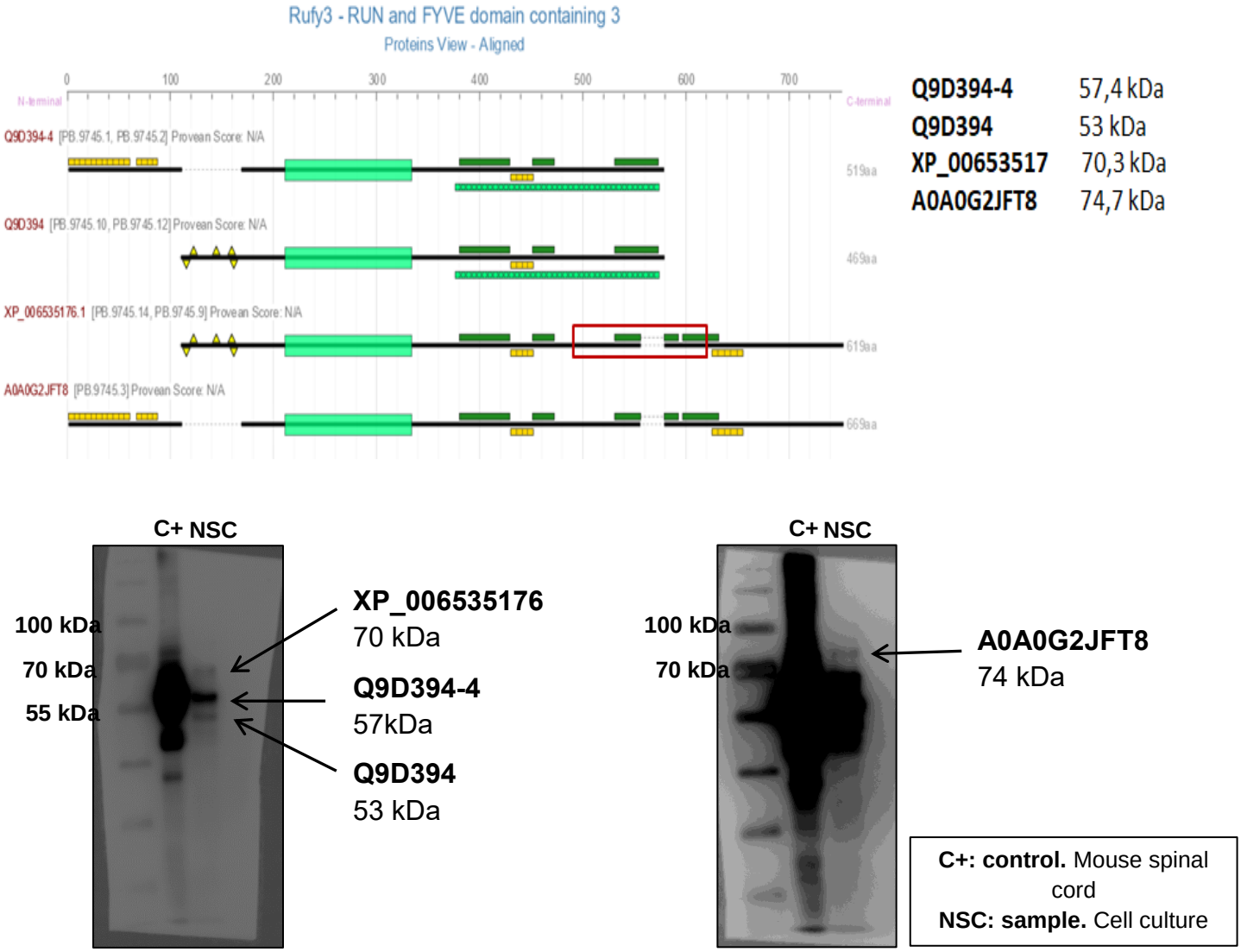


Figure 15. RUFY3 isoforms are detected in NSCs. The superior panel show the four isoforms predicted and their molecular weight in kDa. The red square highlights the region of the proteins targeted by the antibody. The lower panels show the RUFY isoforms detected in the blotting. Lane 1 correspond to spinal cord protein homogenate, used as control (C+) and lane 2 to NSCs sample at T0.

6.3.3. Protein expression analysis of RMB39 (CAPER) in NSCs

According to tappAS, RBM39 gene encode for four proteins, three of them previously described (Q8VH51, Q8BH51-2 and Q8BH51-3) and a novel protein (XP_011237624.1) that differ in their N-terminal sequence. WB was performed with total protein extract from NSCs at T0 and a spinal cord protein homogenate as positive control (**Figure 16**). In the line of the NSCs sample, our results show three clear bands in the expected weights of RBM39 isoforms. One band slightly lower than 40 kDa that matches with Q8BH51-3 and another slightly higher that coincides with XP_011237624.1. Furthermore, there is another band upper to 55 kDA that coincide with the expected size of Q8VH51 and Q8BH51-2. According to the bands obtained in the WB all the signals expected are present in the NSCs sample, even the novel protein (XP_011237624.1). since, the size of Q8VH51 and Q8BH51-2 is very similar, it is not possible to conclude if both or just one of them is present in our system.

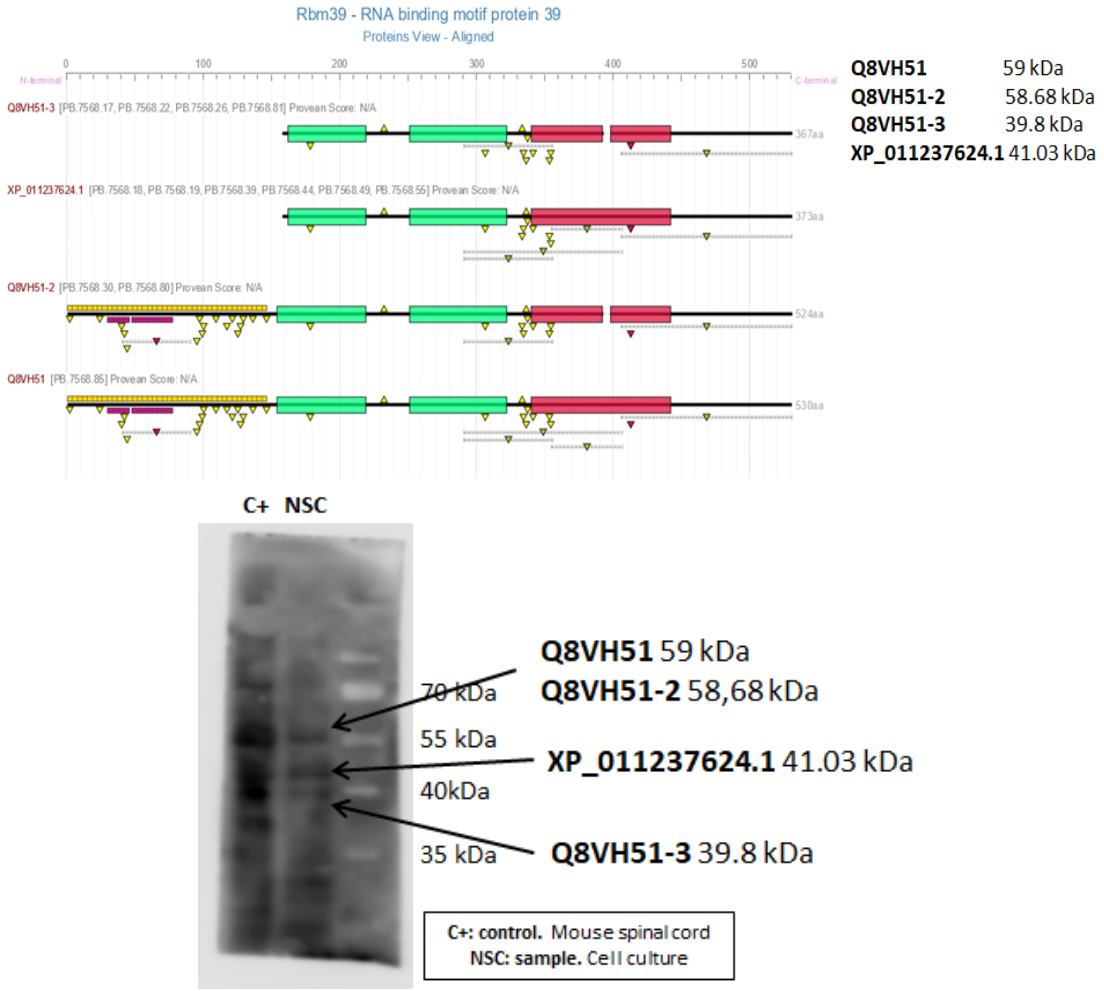


Figure 16. RBM39 (CAPER) protein expression analysis in NSCs. The superior panel show the structure and molecular weight of RBM39 isoforms. The inferior panel shows the

blotting with the isoforms detected. Lane 1 correspond to spinal cord protein homogenate used as control (C+) and lane 2 correspond to NSCs sample at T0.

7.Discussion

The present project aimed two objectives, on the one hand, one of objectives of the project was the validation of TappAs in order to confirm its efficacy analysing and predicting gene and protein data at the isoform level. On the other hand, the second main objective was leveraging the information provided by TappAS to find genes specially involved in the differentiation from NSCs to OPCs and MTNs, that might be importantly involved in the differentiation process and be differential for the cell fate determination.

In spite of the limitations of our experimental design which do not enable to conclude that all the four OMA1 isoforms are present in the sample, experimental results support the presence of transcripts isoforms encoding for the complete and the incomplete protein, supporting the theoretical data provided by TappAS. Isoforms expression analysis of OMA1 show similar patterns in our differentiation lineages compared to the RNAseq data (**Figure 12**). Although, expression values present minor variations in the expression levels at some stages of the differentiation process (OPC T2), obtaining these similar expression rates patterns represents a promising result, since it means that the theoretical data quantification provides reliable and reproducible information.

According to the functional analysis OMA1 presents an NLS, suggesting that it may play a role or stay trapped in the nucleus. However, protein expression analysis has identified the inactive pre-pro protein (Q9D8H7) and the cleaved active protein in our system but not the novel predicted protein (XP_006503364.1) (**Figure 14**). According to TappAS, XP_006503364.1 loses a transmembrane domain (**Figure 8**) that might be used in order to anchor into the mitochondrial membrane, where OMA1 plays its function. Nevertheless, it may be possibly a misfolded non-functional variant which could be rapidly degraded, in order to ensure that OMA1 is not overworking. Thus, it could be an isoform variant used to regulate and control OMA1 function in order to adapt it to the metabolic need of the organism, instead of being a classical protein with a determined function. Additionally, literature describes OMA1 as a mitochondrial metalloprotease, but non-nuclear function has been addressed.

Protein expression analysis (**Figure 14**) reveals incoherencies with the protein levels detected in TappAS in the time-course OPCs experiment (**Figure 8**). Protein abundance is a direct determinant of cellular function and is highly controlled by a number of post-transcriptional, translational and post-translational regulatory processes (Vogel and Marcotte 2012). Nevertheless, the quantitative contribution of mRNA abundance to protein abundance is controversial (Liu et al. 2016; Cheng et al. 2016) and there is still gap in the knowledge of those transcripts that are not identified at the protein level, since some of them may be expressed in small quantities, in limited tissues or under specific circumstances, and some may be under regulatory pathways (Lykke-Andersen and Bennett 2014).

MUL1 expression patterns (**Figure 13**) are quite preserved in the MTNs lineage but present more differences in OPCs compared to the theoretical data, and at stage T1 and T2 the major isoforms are swapped. However, it must be taken into account that the differentiation is not a homogeneous process and cells may be at different stages and we could find undifferentiated phenotypes and cell-fate committed populations (Raouf et al., 2008). Besides although cells are collected at the same time, is possible that the differentiation process would be running at a different pace, and it is possible that variation in gene expression may change or be slightly different from one experiment to another at those points of sample harvesting. Furthermore, the cell culture it is not homogeneous and there is a compendium of cells enriched in the cell type of interes (Teschendorff and Enver, 2017). Therefore, isoform expression values obtained may be produced from the combination of more than a unique cell lineage.
Nevertheless, it is a significant fact that inconsistencies appear at the middle stages of the differentiations, when the culture is more heterogeneous, and variabilities are more likely to appear than at the first and final stages when cell cultures more defined and homogeneous.

Apart from these limitations and according to the obtained experimental results, TappAS have enabled the experimental identification of either previous described or novel not predicted isoforms, such as MUL1 PB.9202.6 transcript. In addition, the expression rates are pretty conserved in both lineages, and generally the major and minor isoforms conserved their position along the process.

According to the expression values obtained of OMA1 isoforms (**Figure 12 C**), PB. 8855.2 and PB. 8855.7, which encodes for the novel protein XP_006503364.1, are increased at T1 in the OPC lineage. Furthermore, PB8855.1 and PB8855.5 transcripts, that encode for the functional protein Q9D8H7, decrease at T1 (**Figure 12 C**), what also coincides with the lowest values of OMA1 protein expression (**Figure 14 B**). However, those transcripts expression increase again at T2 (**Figure 12 C**) together with the increase in the total protein fraction (**Figure 14 B**). Taking these results together, the increase in the expression of the "functional transcripts", is correlative with the expression of "functional" OMA1 protein.

Differentiation is a high-energy demanding process. OPCs use energy in order to synthetize proteins and lipids to raise myelin sheaths and rely on mitochondrial respiration during differentiation and myelination (Rinholm et al. 2016). Besides, are very sensitive to energy deprivation, and it has been demonstrated that oligodendroglia differentiation induces the expression of mitochondrial genes in order to support such a demanding process and inhibition of the mitochondrial activity impairs OPCs differentiation (Schoenfeld et al. 2010). So, it is likely that the increase of OMA1 expression in the total protein fraction coincides with an important step of the differentiation as the proliferation of the oligodendrocyte precursors (Barateiro and Fernandes 2014). Besides, oligodendrocytes mitochondria are able to enter and move through the myelin sheaths (Pantoni et al. 1996) and provide metabolic support to the axons (Fünfschilling et al. 2012), providing not only a physical frame but also energy supply to the neurons, so metabolic genes must have a vital importance for the differentiation.

Additionally, it has been reported that mitochondrial mutations might affect NPCs proliferation and impair neuronal differentiation delaying neurogenesis (Lorenz et al. 2017), probably due to lack of energy. In addition, mutation in mitochondrial proteases and fission-fusion proteins are associated with neuronal loss and neurodegenerative disorders (Patron et al. 2018) and defective brain development (Smirnova et al. 2001), respectively. Since, previous literature showed that OMA1 and MUL1, both act as regulators of the mitochondrial activity, they must be playing an important role in cell differentiation to OPCs and MTNs.

In the case of RUFY3 all the variants previously described (Q9D394, Q9D394-4 and A0A0G2JFT8) and the novel predicted isoform (XP-006535175.1) have been detected in the NSCs (T0) samples (**Figure 15**). This previous experiment makes RUFY3 a promising candidate in order to asses isoforms variability in the time-course experiment either in OPCs or MTNs lineage. A miRNA study performed in a neural differentiation context has shown that RUFY3 expression is upregulated in those cell lines which present *AS* specific of the nervous system. (Makeyev et al. 2007). Furthermore, it has been described that RUFY3 interacts with actin binding proteins in the neurons and is needed for the correct axon formation and elongation (Wei et al. 2014), which is a vital process during neurogenesis in order to establish proper neural connectivity.

In addition, recent transcriptome studies has address the importance of RNA binding proteins (RNABPs) in neuronal development (Yano et al. 2015) and a number of RNABPs have been identified in neurons (Darnell 2013). Since they contribute to the transcriptome diversity, they might have a major role in the regulation of the differentiation process (Kanemitsu et al. 2017). Interestingly, RBM39 stimulates energy metabolism inducing nuclear genes expression and controlling redox levels in *C. elegans*. Moreover, maintains mitochondrial metabolism through direct transcriptional control of the mitochondria transcriptional machinery (Kang et al., 2015). Therefore, RBM39 acts regulating downstream metabolic pathways in order to optimize ATP levels according to the metabolic demand. In the case of RBM39, all the signals detected are at the expected molecular weights (**Figure 16**). Since Q8VH51 and Q8BH51-2 present very similar molecular weights, it only appears a single band and it is not possible to conclude if both are in our sample. However, the most likely is that both are present since both have been previously. The most interesting band correspond to the novel isoforms of CAPER (XP_011237624.1) and RUFY3 (XP-006535175.1). Considering the results obtained with the other candidates and given the role of RBM39 as metabolic regulator, it might be regulating mitochondrial pathways in our differentiation process. Therefore, it is a promising candidate for further analysis in the differentiation lineages in order to assess its implication in the process.

Taken all the results together, TappAS recapitulated already existing annotation, since protein isoforms already predicted have been detected. What is more, in the case of MUL1, TappAS enabled the detection of the novel not in catalogue transcript PB.9202.6 as well as predicted proteins in case of RUFY3 and RBM39. Thus, our experiments confirm the accuracy and sensibility of TappAS Tool, demonstrating that it is not only a useful tool for transcriptome analysis because of its capability to correctly address existing information, but also is an exciting and promising tool in order to find new differences in transcriptomic and proteomic studies at the isoform level. Additionally, has allowed us to infer the importance of OMA1 and MUL isoforms in the Neural differentiation process. Nevertheless, further wok must be done in order to complete elucidate their importance along the neural differentiation process, and their role in the cell-fate determination.

8. Conclusions

Regarding the experimental validation of tappAS:

1. OMA1 isoforms expression patterns reproduce those provided by TappAS Tool in the Motoneurons lineage and Oligodendrocyte linage. Further, it has been detected both the complete isoforms and those with exon 6 skipping.
2. MUL1 expression patterns reproduce the results in the Motoneuron lineage but present more variability in the oligodendrocyte lineage. Nevertheless, isoform switching has not been addressed in the experimental results at the first and final stages of the differentiation.
3. In the case of MUL1, a new isoform not in catalogue have been detected, confirming the accuracy and specificity of TappAS to provide new information at the isoform level.
4. At the protein level, OMA1 previous described isoform but not the novel predicted have been detected in the Oligodendrocyte lineage However, all the protein isoforms of RUFY3 and RBM39 have been detected in Neural Stem Cells samples, supporting the effectivity of TappAS also at the protein level.

Regarding the functional impact in the differentiation process:

1. OMA1 Isoforms present differential expression between the Motoneuron and the Oligodendrocyte lineage, implying a functional role in the cell fate determination through regulation of mitochondrial dynamics. However, further work must be done at the protein level in order to disclosure their real implication.
2. MUL1 isoforms also present expression variability at the final stage, suggesting that MUL1 might be important at the end of the differentiation process.

9. References

Allen, N. J., & Barres, B. A. (2009). Glia — more than just brain glue. *Nature*, Vol. 457, pp. 675–677. https://doi.org/10.1038/457675a

Baker, M. J., Lampe, P. A., Stojanovski, D., Korwitz, A., Anand, R., Tatsuta, T., & Langer, T. (2014). Stress-induced OMA1 activation and autocatalytic turnover regulate OPA1-dependent mitochondrial dynamics. *The EMBO Journal*, *33*(6), 578–593.

Baralle, F. E., & Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nature Reviews. Molecular Cell Biology*, *18*(7), 437–451.

Barateiro, A., & Fernandes, A. (2014). Temporal oligodendrocyte lineage progression: in vitro models of proliferation, differentiation and myelination. *Biochimica et Biophysica Acta*, *1843*(9), 1917–1929.

Berget, S. M., Moore, C., & Sharp, P. A. (1977). Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences*, Vol. 74, pp. 3171–3175. https://doi.org/10.1073/pnas.74.8.3171

BLAST: Basic Local Alignment Search Tool. (n.d.). Retrieved June 13, 2019, from https://blast.ncbi.nlm.nih.gov/Blast.cgi

Blencowe, B. J. (2006). Alternative splicing: new insights from global analyses. *Cell*, *126*(1), 37–47.

Braschi, E., Zunino, R., & McBride, H. M. (2009). MAPL is a new mitochondrial SUMO E3 ligase that regulates mitochondrial fission. *EMBO Reports*, *10*(7), 748–754.

Cheng, Z., Teo, G., Krueger, S., Rock, T. M., Koh, H. W. L., Choi, H., & Vogel, C. (2016). Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. *Molecular Systems Biology*, *12*(1), 855.

Chih, B., Gollan, L., & Scheiffele, P. (2006). Alternative splicing controls selective trans-synaptic interactions of the neuroligin-neurexin complex. *Neuron*, *51*(2), 171–178.

Chow, L. T., Gelinas, R. E., Broker, T. R., & Roberts, R. J. (1977). An amazing sequence arrangement at the 5’ ends of adenovirus 2 messenger RNA. *Cell*, *12*(1), 1–8.

Consolato, F., Maltecca, F., Tulli, S., Sambri, I., & Casari, G. (2018). m-AAA and i-AAA complexes coordinate to regulate OMA1, the stress-activated supervisor of mitochondrial dynamics. *Journal of Cell Science*, *131*(7). https://doi.org/10.1242/jcs.213546

Darnell, R. B. (2013). RNA protein interaction in neurons. *Annual Review of Neuroscience*, *36*, 243–270.

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., … Buxbaum, J. D. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, *515*(7526), 209–215.

EMBL-EBI. (n.d.). MUSCLE < Multiple Sequence Alignment < EMBL-EBI. Retrieved June 13, 2019, from https://www.ebi.ac.uk/Tools/msa/muscle/

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., & Lührmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Molecular Cell*, *36*(4), 593–608.

Fünfschilling, U., Supplie, L. M., Mahad, D., Boretius, S., Saab, A. S., Edgar, J., … Nave, K.-A. (2012). Glycolytic oligodendrocytes maintain myelin and long-term axonal integrity. *Nature*, *485*(7399), 517–521.

Furlanis, E., & Scheiffele, P. (2018). Regulation of Neuronal Differentiation, Function, and Plasticity by Alternative Splicing. *Annual Review of Cell and Developmental Biology*, *34*, 451–469.

Fu, Y., Chen, L., Chen, C., Ge, Y., Kang, M., Song, Z., … Xu, A. (2018). Crosstalk between alternative polyadenylation and miRNAs in the regulation of protein translational efficiency. *Genome Research*, *28*(11), 1656–1663.

Gibilisco, L., Zhou, Q., Mahajan, S., & Bachtrog, D. (2016). Alternative Splicing within and between Drosophila Species, Sexes, Tissues, and Developmental Stages. *PLoS Genetics*, *12*(12), e1006464.

Gilbert, W. (1978). Why genes in pieces? *Nature*, *271*(5645), 501.

Goren, A., Ram, O., Amit, M., Keren, H., Lev-Maor, G., Vig, I., … Ast, G. (2006). Comparative Analysis Identifies Exonic Splicing Regulatory Sequences—The Complex Definition of Enhancers and Silencers. *Molecular Cell*, Vol. 22, pp. 769–781. https://doi.org/10.1016/j.molcel.2006.05.008

Honda, A., Usui, H., Sakimura, K., & Igarashi, M. (2017). Rufy3 is an adapter protein for small GTPases that activates a Rac guanine nucleotide exchange factor to control neuronal polarity. *The Journal of Biological Chemistry*, *292*(51), 20936–20946.

Iijima, T., Hidaka, C., & Iijima, Y. (2016). Spatio-temporal regulations and functions of neuronal alternative RNA splicing in developing and adult brains. *Neuroscience Research*, *109*, 1–8.

Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., … Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, Vol. 515, pp. 216–221. https://doi.org/10.1038/nature13908

Jin, L., Li, G., Yu, D., Huang, W., Cheng, C., Liao, S., … Zhang, Y. (2017). Transcriptome analysis reveals the complexity of alternative splicing regulation in the fungus Verticillium dahliae. *BMC Genomics*, *18*(1), 130.

Johnson, M. B., Kawasawa, Y. I., Mason, C. E., Krsnik, Z., Coppola, G., Bogdanović, D., … Sestan, N. (2009). Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron*, *62*(4), 494–509.

Jurica, M. S., & Moore, M. J. (2003). Pre-mRNA Splicing. *Molecular Cell*, *12*(1), 5–14.

Kanemitsu, Y., Fujitani, M., Fujita, Y., Zhang, S., Su, Y.-Q., Kawahara, Y., & Yamashita, T. (2017). The RNA-binding protein MARF1 promotes cortical neurogenesis through its RNase activity domain. *Scientific Reports*, *7*(1), 1155.

Kang, Y. K., Putluri, N., Maity, S., Tsimelzon, A., Ilkayeva, O., Mo, Q., … O'Malley, B. W. (2015). CAPER is vital for energy and redox homeostasis by integrating glucose-induced mitochondrial functions via ERR-α-Gabpa and stress-induced adaptive responses via NF-κB-cMYC. *PLoS Genetics*, *11*(4), e1005116.

Keirstead, H. S., Nistor, G., Bernal, G., Totoiu, M., Cloutier, F., Sharp, K., & Steward, O. (2005). Human embryonic stem cell-derived oligodendrocyte progenitor cell transplants remyelinate and restore locomotion after spinal cord injury. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(19), 4694–4705.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., & Stamm, S. (2013). Function of alternative splicing. *Gene*, Vol. 514, pp. 1–30. https://doi.org/10.1016/j.gene.2012.07.083

Lara-Pezzi, E., Desco, M., Gatto, A., & Gómez-Gaviro, M. V. (2017). Neurogenesis: Regulation by Alternative Splicing and Related Posttranscriptional Processes. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *23*(5), 466–477.

Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., & Brenner, S. E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, *446*(7138), 926–929.

Lavie, J., De Belvalet, H., Sonon, S., Ion, A. M., Dumon, E., Melser, S., … Bénard, G. (2018). Ubiquitin-Dependent Degradation of Mitochondrial Proteins Regulates Energy Metabolism. *Cell Reports*, *23*(10), 2852–2863.

Leggere, J. C., Saito, Y., Darnell, R. B., Tessier-Lavigne, M., Junge, H. J., & Chen, Z. (2016). NOVA regulates Dcc alternative splicing during neuronal migration and axon guidance in the spinal cord. *eLife*, *5*. https://doi.org/10.7554/eLife.14264

Liu, Y., Beyer, A., & Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*, *165*(3), 535–550.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* , *25*(4), 402–408.

Lorenz, C., Lesimple, P., Bukowiecki, R., Zink, A., Inak, G., Mlody, B., … Prigione, A. (2017). Human iPSC-Derived Neural Progenitors Are an Effective Drug Discovery Model for Neurological mtDNA Disorders. *Cell Stem Cell*, *20*(5), 659–674.e9.

Lykke-Andersen, J., & Bennett, E. J. (2014). Protecting the proteome: Eukaryotic cotranslational quality control pathways. *The Journal of Cell Biology*, *204*(4), 467–476.

Mai, S., Qu, X., Li, P., Ma, Q., Cao, C., & Liu, X. (2016). Global regulation of alternative RNA splicing by the SR-rich protein RBM39. *Biochimica et Biophysica Acta*, *1859*(8), 1014–1024.

Makeyev, E. V., Zhang, J., Carrasco, M. A., & Maniatis, T. (2007). The MicroRNA miR-124 Promotes Neuronal Differentiation by Triggering Brain-Specific Alternative Pre-mRNA Splicing. *Molecular Cell*, Vol. 27, pp. 435–448. https://doi.org/10.1016/j.molcel.2007.07.015

McCarthy, S. E., Gillis, J., Kramer, M., Lihm, J., Yoon, S., Berstein, Y., … Corvin, A. (2014). De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Molecular Psychiatry*, Vol. 19, pp. 652–658. https://doi.org/10.1038/mp.2014.29

Molofsky, A. V., Krencik, R., Ullian, E. M., Tsai, H.-H., Deneen, B., Richardson, W. D., … Rowitch, D. H. (2012). Astrocytes and disease: a neurodevelopmental perspective. *Genes & Development*, *26*(9), 891–907.

Multalin interface page. (n.d.). Retrieved June 13, 2019, from http://multalin.toulouse.inra.fr/multalin/

Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S. S., Chen, W., … Ropers, H. H. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature*, *478*(7367), 57–63.

Neuspiel, M., Schauss, A. C., Braschi, E., Zunino, R., Rippstein, P., Rachubinski, R. A., … McBride, H. M. (2008). Cargo-selected transport from the mitochondria to peroxisomes is mediated by vesicular carriers. *Current Biology: CB*, *18*(2), 102–108.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, *40*(12), 1413–1415.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2009). Addendum: Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, Vol. 41, pp. 762–762. https://doi.org/10.1038/ng0609-762d

Pantoni, L., Garcia, J. H., & Gutierrez, J. A. (1996). Cerebral white matter is highly vulnerable to ischemia. *Stroke; a Journal of Cerebral Circulation*, *27*(9), 1641–1646; discussion 1647.

Park, E., Pan, Z., Zhang, Z., Lin, L., & Xing, Y. (2018). The Expanding Landscape of Alternative Splicing Variation in Human Populations. *American Journal of Human Genetics*, *102*(1), 11–26.

Patron, M., Sprenger, H.-G., & Langer, T. (2018). m-AAA proteases, mitochondrial calcium homeostasis and neurodegeneration. *Cell Research*, *28*(3), 296–306.

Pleiss, J. A., Whitworth, G. B., Bergkessel, M., & Guthrie, C. (2007). Rapid, transcript-specific changes in splicing in response to environmental stress. *Molecular Cell*, *27*(6), 928–937.

PREMIER Biosoft: User Login. (n.d.). Retrieved June 13, 2019, from http://www.premierbiosoft.com/NetPrimer/

Raj, B., & Blencowe, B. J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron*, *87*(1), 14–27.

Raouf, A., Zhao, Y., To, K., Stingl, J., Delaney, A., Barbara, M., … Eaves, C. (2008). Transcriptome analysis of the normal human mammary cell commitment and differentiation process. *Cell Stem Cell*, *3*(1), 109–118.

Rinholm, J. E., Vervaeke, K., Tadross, M. R., Tkachuk, A. N., Kopek, B. G., Brown, T. A., … Clayton, D. A. (2016). Movement and structure of mitochondria in oligodendrocytes and their myelin sheaths. *Glia*, *64*(5), 810–825.

Ronan, J. L., Wu, W., & Crabtree, G. R. (2013). From neural development to cognition: unexpected roles for chromatin. *Nature Reviews. Genetics*, *14*(5), 347–359.

Schaefke, B., Sun, W., Li, Y.-S., Fang, L., & Chen, W. (2018). The evolution of posttranscriptional regulation. *Wiley Interdisciplinary Reviews. RNA*, e1485.

Schoenfeld, R., Wong, A., Silva, J., Li, M., Itoh, A., Horiuchi, M., … Cortopassi, G. (2010). Oligodendroglial differentiation induces mitochondrial genes and inhibition of mitochondrial function represses oligodendroglial differentiation. *Mitochondrion*, *10*(2), 143–150.

Smirnova, E., Griparic, L., Shurland, D.-L., & van der Bliek, A. M. (2001). Dynamin-related Protein Drp1 Is Required for Mitochondrial Division in Mammalian Cells. *Molecular Biology of the Cell*, Vol. 12, pp. 2245–2256. https://doi.org/10.1091/mbc.12.8.2245

Steijger, T., Abril, J. F., Engström, P. G., Kokocinski, F., RGASP Consortium, Hubbard, T. J., … Bertone, P. (2013). Assessment of transcript reconstruction methods for RNA-seq. *Nature Methods*, *10*(12), 1177–1184.

Steitz, J. A., Dreyfuss, G., Krainer, A. R., Lamond, A. I., Matera, A. G., & Padgett, R. A. (2008). Where in the cell is the minor spliceosome? *Proceedings of the National Academy of Sciences of the United States of America*, *105*(25), 8485–8486.

tappAS – tappAS is a Java GUI application for the analysis of RNA-Seq data down to the isoform level. (n.d.-a). Retrieved June 9, 2019, from http://tappas.org

tappAS – tappAS is a Java GUI application for the analysis of RNA-Seq data down to the isoform level. (n.d.-b). Retrieved June 30, 2019, from http://tappAS.org

Tellmann, G. (2006). The E-Method: a highly accurate technique for gene-expression analysis. *Nature Methods*, Vol. 3, pp. i – ii. https://doi.org/10.1038/nmeth894

Teschendorff, A. E., & Enver, T. (2017). Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nature Communications*, Vol. 8. https://doi.org/10.1038/ncomms15599

Vaquero-Garcia, J., Barrera, A., Gazzara, M. R., González-Vallinas, J., Lahens, N. F., Hogenesch, J. B., … Barash, Y. (2016). A new view of transcriptome complexity and regulation through the lens of local splicing variations. *eLife*, Vol. 5. https://doi.org/10.7554/elife.11752

Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews. Genetics*, *13*(4), 227–232.

Vuong, C. K., Black, D. L., & Zheng, S. (2016). The neurogenetics of alternative splicing. *Nature Reviews. Neuroscience*, *17*(5), 265–281.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., … Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*(7221), 470–476.

Wang, X., Hou, J., Quedenau, C., & Chen, W. (2016). Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals. *Molecular Systems Biology*, *12*(7), 875.

Weirather, J. L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., … Au, K. F. (2017). Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research*, *6*, 100.

Wei, Z., Sun, M., Liu, X., Zhang, J., & Jin, Y. (2014). Rufy3, a protein specifically expressed in neurons, interacts with actin-bundling protein Fascin to control the growth of axons. *Journal of Neurochemistry*, *130*(5), 678–692.

Will, C. L., & Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology*, *3*(7). https://doi.org/10.1101/cshperspect.a003707

Xu, D., Liu, A., Wang, X., Zhang, M., Zhang, Z., Tan, Z., & Qiu, M. (2018). Identifying suitable reference genes for developing and injured mouse CNS tissues. *Developmental Neurobiology*, *78*(1), 39–50.

Yano, M., Ohtsuka, T., & Okano, H. (2015). RNA-binding protein research with transcriptome-wide technologies in neural development. *Cell and Tissue Research*, *359*(1), 135–144.

Yeo, G., Holste, D., Kreiman, G., & Burge, C. B. (2004). Variation in alternative splicing across human tissues. *Genome Biology*, *5*(10), R74.

Yun, J., Puri, R., Yang, H., Lizzio, M. A., Wu, C., Sheng, Z.-H., & Guo, M. (2014). MUL1 acts in parallel to the PINK1/parkin pathway in regulating mitofusin and compensates for loss of PINK1/parkin. *eLife*, *3*, e01958.

Zhang, R., Calixto, C. P. G., Marquez, Y., Venhuizen, P., Tzioutziou, N. A., Guo, W., … Brown, J. W. S. (2017). A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Research*, *45*(9), 5061–5073.

Zhang, Y., Chen, K., Sloan, S. A., Bennett, M. L., Scholze, A. R., O'Keeffe, S., … Wu, J. Q. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(36), 11929–11947.

Zheng, C. L., Fu, X.-D., & Gribskov, M. (2005). Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. *RNA* , *11*(12), 1777–1787.