Alternative splicing during neuronal development

Mariana Ascensão Ferreira¹,²
marianaferreira@tecnico.ulisboa.pt

Abstract

Alternative splicing (AS) is one of the mechanisms expanding transcriptome diversity and is known to occur in 95% of human multi-exon genes. AS is more frequent and more conserved, across vertebrate species, in nervous tissues and there are relatively more neuronal-specific RNA-binding proteins, suggesting that neuronal-specific regulated AS may be one major contributor to functional complexity of neurons. Importantly, links between alternative splicing misregulations and neurodevelopmental disorders have already been established.

Widespread usage of high-throughput sequencing techniques allowed the current availability of transcriptomic data from previous studies, which combined with bioinformatics approaches potentiates the accurate quantification of gene expression and exon inclusion levels. Through gene expression and AS quantification, the patterns of exon inclusion during neuronal differentiation enabled the identification of AS events changing as cells commit to the neuronal fate, as well as characterization of a neural progenitor cell line used to model neurogenesis in vitro (NPC 46C).

The NOVA1 RNA-binding protein (RBP) is known to promote exon skipping upon binding to the last nucleotides of the upstream intron. Interestingly, neuronal function AS events with switch-like changes during differentiation showed inclusion levels significantly anti-correlated with the expression of the Nova1 gene. Moreover, similar association between nSR100 RBP and neuronal-required exons is observed. These results provide a relevant contribution for profiling the role of alternative splicing in fine-tuning neuronal function.

Keywords

Alternative splicing — neuronal development — transcriptomics — neuroprogenitor cells

1 Instituto Superior Técnico, Universidade de Lisboa
2 Computational Biology Lab, Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa

1. Introduction

Alternative splicing (AS) is one of the mechanisms contributing for the expansion of the transcriptome and proteome in species with similar number of genes, and is known to occur in 95% of multi-exonic genes. By being alternatively spliced, the same gene can produce different transcripts, possibly leading to different protein products, altering the gene’s function. In addition to its correlation with species complexity, AS regulation has been associated to several mechanisms of neuronal function. This work focused on the exploration of alternative splicing events showing switch-like changes along neuronal differentiation and their regulatory mechanisms.

1.1 Molecular biology of alternative splicing

The genome is maintained relatively constant during the cell’s life cycle but genomic information must be decoded in order to synthesize genes’s products, accordingly to the environment the cell is exposed to. This process occurs with transcription, when genomic DNA sequences are copied as sequences of RNA. RNA polymerase, the enzyme catalysing transcription, assembles with accessory proteins – transcription factors – and recognizes
the transcription start site (TSS) of a gene. Unwinding of the DNA double strand is promoted by RNA polymerase as it slides through the gene sequence in the 5' → 3' direction, forming a complementary sequence of ribonucleotides. Transcription ends when RNA polymerase reaches the terminator region, containing specific nucleotide sequences that are recognized by the enzyme, originating an RNA sequence – the precursor messenger RNA (pre-mRNA) – which contains a copy of the gene. After being synthesized in the nucleus, the pre-mRNA transcripts undergo processing in order to increase their stability to leave the nucleus. The 5' end of the transcription is modified, with the inclusion of a methylated guanine nucleotide – RNA capping – and the 3' is extended with a sequence of hundreds of adenosine nucleotides, forming the poly-A tail [1].

One of the most unanticipated discoveries in biology was the fact that genes’ coding regions – exons – are intercalated with longer, non-coding regions – introns. Likewise, the pre-mRNA produced by transcription must be further processed, in order to remove intronic regions and join exons together, the splicing reaction. This often co-transcriptional reaction is catalysed by the spliceosome, the largest ribonucleoprotein complex in the nucleus of eukaryotic cells, in two sequential transesterification reactions that enable the removal of the intron. Interestingly, splicing may occur in different manners, since regions defining exons can be alternately promoted for selection by trans-acting factors. This selection is regulated by RNA-binding proteins (RBPs) that bind to regulatory regions within the transcript sequence (cis-regulatory regions), either acting as enhancers or inhibitors for splice site selection (Figure 1). This strongly regulated mechanism of exon inclusion results in the existence of complex forms of alternative splicing, depicted in Figure 2.

Alternative splicing events lead to changes in proteins’ isoforms that have impacts on their function. Some studies report isoform changes causing alterations on protein’s catalytic activity, molecular interactions, structure, subcellular localization or stability, having a great impact in cellular processes [3, 4]. One of the most widely studied consequences of alternative splicing is in the regulation of gene expression, namely by the introduction of a premature termination codon by AS, triggering the nonsense-mediated decay pathway that causes transcripts to be degraded. Importantly, links between alternative splicing and diseases, such as cancer, muscular dystrophies and developmental and degenerative disorders have already been suggested [3, 4].

The ability of cells to react to exogenous environments is based on decision defining which genes are to be expressed at any moment. By coupling dynamic regulation of transcription with alternative splicing, a process largely expanding the number of possible transcripts, the cell is empowered with a biologically efficient and elegant mechanism for fine-tuning gene expression without being fully limited by the number of genes. In fact, higher complexity animals have similar number of genes, when compared to functionally less complex animals, but are able to greatly expand the number of isoforms by alternative splicing of pre-mRNA transcripts [2, 3].

1.2 Neuronal development

The vertebrate embryo dorsal ectoderm gives rise to multipotent stem cells, which through differentiation originate the lineages of neural cells: neurons and glial cells (astrocytes and oligodendrocytes). Cell signalling controls spatially and temporally coordinated cell-fate decisions that result in timely proliferation and differentiation of cell types. In the first stages of neuronal development, neural stem cells undergo symmetrical divisions, in order to expand the number of cells. At a certain point in gestation – the neurogenic stage –, these cells start dividing asymmetrically, originating one neural
stem/progenitor cell and one neuron. Further on in development, asymmetrical divisions ensure production of one stem/progenitor cell and one glial cell. One of the players in the regulation of neural differentiation is the Notch signalling pathway, which regulates cell’s transitions between neurogenesis and gliogenesis processes, by inhibiting progression of neural progenitors to neurons and of glial progenitors to oligodendrocytes, but promoting glial differentiation into astrocytes [6, 7].

The cellular and molecular characterization of neural commitment to differentiation relies on in vitro experiments where pluripotent stem cells undergo neural induction. One of the commonly used cell lines to study neuronal differentiation in vitro are neural progenitor cells obtained through a knock-in of Sox1-GFP (green fluorescent protein) – NPC 46C –, which emits green fluorescence when Sox1 is expressed, marking the entering of cells into the neuronal phenotype. When cultured in serum-free medium, NPC 46C cells express pluripotency and embryonic stem cell markers in the first days in colony. However, pluripotency genes expression decreases as soon as after one day in colony, upon neural induction, while ectoderm markers expression increases and other lineages markers decrease, confirming the commitment of cells to the ectodermal fate. Further on, pro-neural genes expression increases from day six in culture on, consistently with the commitment and differentiation of cells to acquire the neural phenotype [8, 9, 7].

1.3 Alternative splicing during neuronal development

The recent widespread of genomic sequencing has provided evidence that the number of genes of a species does not correlate with organismal complexity. In fact, gene expression and alternative splicing complexity was compared between different tissues from vertebrate species separated by 350 million years of evolutionary distance. Strikingly, results showed that while gene expression patterns group similar tissues together, AS patterns are more similar within species than between the same tissues of different species [10]. This AS species-specific signature suggests that, in vertebrates, evolutionary pressure may have forced AS mechanisms to evolve rapidly, resulting in a more relevant contribution to species phenotypic diversity than gene expression evolutionary changes. However, AS events resulting in important functional consequences for a tissue are relatively conserved. In conclusion, AS seems to be a mechanism through which species greatly expand the transcriptome, although tight regulation seems to be coherent with events that have a strong and conserved physiological impact. Interestingly, it has been suggested that AS would be the most versatile mechanism to account for the amplification on the number of cell adhesion molecules which are fundamental for the establishment of synapses [11].

Several RBPs acting as alternative splicing regulators have been reported to have a brain-specific activity. NOVA proteins have already been suggested to regulate a network of 700 exons in mouse brains [12]. RBFOX1/2/3 proteins are a family of neural-enriched splicing factors, whose interactions with RNA have been mapped in mouse brains, showing a differential spatial distribution, suggesting that, for instance, RBFOX1 and RBFOX2 have distinct but complementary roles in AS regulation [13]. Moreover, PTBP1 and PTBP2 are splicing regulators acting in a mutually exclusive manner. PTBP1 controls the inclusion rates of a neural-specific exon in Ptbp2, which is required for the Ptbp2 action in promoting expression of pro-neural genes. During in vitro neuronal differentiation, expression levels of Ptbp1 decrease while Ptbp2 expression levels increase. This balance between Ptbp1 and Ptbp2 is depicted in Figure 3. Moreover, PTBP1 is also known to be influencing inclusion levels of an exon in Pbx1 gene, originating a change on isoform ratios for this protein, that is responsible for differences in promoting transcription of pro-neural genes between non-neuronal and neuronal cell stages [14]. This effect is shown in Figure 3. The serine/arginine-rich protein of 100kDa, nSR100, has also been reported as a player in inclusion levels of neuronal-specific exons, that act in remodelling protein-protein interactions in neurogenesis. One of such roles, illustrated in Figure 3, is in controlling the inclusion of an exon in the repressor element 1 silencing transcription factor (REST), that acts as a silencer for pro-neuronal genes, when the protein is fully functional. As cells progress to the neuronal fate, nSR100 promotes the inclusion of one exon, leading to a truncated version of the transcription factor – REST4 –, which is not functional to exert its inhibitory role (Figure 3).

Correct differentiation and consequent patterning and maturing of neurons and glial cells is responsible for the connective network of roughly 10\(^{11}\) neurons eventually interacting with the 10\(^{12}\) glial cells [16]. However, defects in synaptic maturation or connectivity may lead to developmental disorders, such as autism spectrum disorders or Rett syndrome, drastically impairing cognitive functions, locomotion and communication [17, 18]. Remarkably, AS alterations have been associated to some aspects of neurodevelopmental disorders. Microexons, shorter exons from 3 to 27 nucleotides in length, are usually more included in later stages of neuronal development, leading to transcripts related to synaptic function and neurite outgrowth, and the inclusion of many are thought to be regulated by nSR100. Individuals with autism spectrum disorder have misregulations in neural microexons, coherent with reduced levels of nSR100 [19]. Moreover, a mouse model of Rett syndrome revealed misregulation of hundreds of alternative splicing events caused by the loss of a transcription factor interacting with splicing modulators [20].
Alternative splicing highly regulated specificity in the brain supports the idea that this evolutionary advantageous strategy used by higher complexity organisms for expansion of the transcriptomic potential may be related to function complexity of mammals nervous systems. Prediction of functional consequences for AS events changing during neuronal development and elucidation of their regulatory mechanisms is required and valuable for eventual development of therapeutic targets. The goal of this work was to analyse AS and gene expression quantifications and correlations between those values by using bioinformatics tools, from which regulatory associations may be inferred. Also, the project course led to the development of a methodology to be used in AS levels calculations and the transcriptional characterization of the commonly used NPC 46C cell line.

2. Materials and Methods
The current work made use of publicly available transcriptomic datasets from human and mouse neuronal developmental cells, pre-processed in order to quantify alternative splicing and gene expression. Using these types of information, statistical hypotheses, such as of evidence for correlation between variables, were tested.

2.1 Gene expression and splicing quantification
High-throughput RNA sequencing (RNA-Seq) directly identifies sequences of nucleotides of a representative population of transcripts in a biological sample. RNA extraction and fragmentation, from which a set of RNA fragments – the library – is obtained, is followed by the conversion to a more stable form of complementary DNA and by the binding of known adapter sequences to the end of the fragments. Afterwards, sequencing occurs in a massive parallel process that identifies transcript sequences by consecutive cycles of reactions with labelled nucleotides, that can be identified by robust imaging software [21, 22, 23].

RNA-Seq data are, in general terms, lists of millions of short read sequences from which gene expression can be inferred by mapping of reads to a reference genome or transcriptome. However, for unbiased quantification of gene expression, some factors need to be taken into consideration to avoid biases. One of those factors is the number of fragments in the library used during the sequencing process. The higher the library size, the higher will be the number of reads mapping to each gene. Moreover, gene length may also be a source for biases, since longer genes will tend to accumulate more reads [24, 25, 26]. These biases may confound both within and between-sample comparisons, and therefore need to be corrected. Reads per kilobase of exon model per million mapped reads (RPKMs) is one widely used metric for normalized gene expression. One extra factor of correction is obtained when considering that only positions of a gene that enable unique mapping (i.e. sequences that are not
repeated in any other regions of the reference) should contribute for attributing reads to a gene and therefore gene length is reduced to reflect only its uniquely mappable positions. Then, normalized gene expression levels are obtained as RPKMs corrected for mappability (cRPKMs), as in Equation 1, comparable both within and between samples [26, 27].

\[
\text{cRPKM}_{\text{gene A}} = \frac{\text{number of reads gene A} \times 10^6 \times \text{uniquely mappable positions gene A}}{\text{total number of mapped reads}}
\]  

Furthermore, reads obtained from RNA-Seq may also be used to quantify splicing levels, by aligning them to splice junctions. The definition of an alternative splicing event is based on the coordinates of all the splice sites involved in the exon inclusion decision. For each AS event, the percent-spliced-in is obtained as the ratio (Equation 2) between the reads supporting the inclusion of the exon and the total number of reads, inclusive and exclusive, attributed to that event (see Figure 4).

\[
\text{PSI} = \frac{\text{Mean(C1-Alt;Alt-C2)}}{(\text{C1-C2}) + \text{Mean(C1-Alt;Alt-C2)}} \times 100
\]  

Figure 4. RNA-Seq reads used in quantification of exon inclusion. Percent-spliced-in (PSI) quantification (in Equation 2) is obtained by comparing the number of reads that align to each exon/exon junction involved in an alternative splicing event. Reads supporting inclusion of the alternative exon are reads spanning the constitutive (C1 or C2) and the alternative exon (Alt), while reads supporting exclusion are reads aligned to junctions from both C1 and C2 [10, 19].

### 2.2 Datasets
This work made use of publicly available and collaborators’ RNA-Seq datasets of neuronal differentiation both in human and mouse. Another dataset—MultiTissue, part of lab’s resources and provided by collaborator Manuel Irimia, has been used to identify the orthologous alternative splicing events between human and mouse. These datasets specificities are summarized in Table 1.

### 2.3 Statistical analysis
This project was developed by conceiving biological questions whose underlying hypothesis was statistically tested. The goal of statistical tests is to determine how different a dataset is from what is expected under the null hypothesis that *nothing is happening* and thus they serve as support for the decision of rejecting the null hypothesis or concluding that there is not enough evidence to do so [30].

**Correlation tests** Tests for correlation were used whenever the question addressed was related to the strength of the relationship between two numeric variables, for instance, PSI and gene expression values across samples. Correlation tests calculate a coefficient associated to the relation between variables, in the interval [−1, 1], where positive values correspond to positively related variables, meaning that the increase of one is reflected by the increase of the other, while negative values are associated to inverse relationships. Two types of correlation tests were used, based on different assumptions about the variables’ distributions and therefore the relationship between them: Pearson’s correlation reflects monotonic, linear relations between variables, while Spearman’s correlation is obtained by processing variables as ranks and relations are inferred on the basis of coherent variations between both variables, without necessarily corresponding to linear relations [30, 31, 32].

**Correlation tests** also provide an associated P-value, capturing the probability of having the same or a more extreme result, by random sampling, if the null hypothesis

---

is true. Importantly, all statistical analyses must be interpreted considering a threshold for acceptable P-values – or significance –, which reflects the probability from which one decides to risk that obtained results are very unlikely to have been generated by chance [32]. In this project, significance thresholds used were 0.05.

**Filtering** Gene expression and AS quantifications contain missing data which should not be taken into consideration. A maximal percentage of NA values per variable considered was defined based on exploratory analysis and all variables containing more than that percentage were removed. Moreover, in the case of exploring AS events, low variance events are not interest in a context of comparison between differentiation time points, since similar values would reflect AS events that occur in the same way across samples. By exploration of the variance and delta PSI for all AS events that passed previous filtering criteria, low variance events were removed from the analysis.

**Principal component analysis** Interpretation of gene expression and AS levels and their relations in different samples is hampered by the multidimensionality of these types of data. To account for this, principal component analysis (PCA) was performed, in order to extract the most of variance within a dataset that contains gene expression or AS events quantifications. This multivariate technique uses single value decomposition to extract linear combinations of the variables, reflecting an orthogonal projection of the data in which the variance of projected data is maximized [33, 30].

**Beta distribution** The Beta distribution is a form of probability density function used to model phenomena with values constrained to the interval [0, 1], being widely applicable in modelling probabilities or proportions. The shape of the distribution is controlled by two shape parameters, \( \alpha \) and \( \beta \). The higher the magnitude of \( \alpha \) and \( \beta \) the more narrowed is the distribution, for a given ratio between them. Moreover, when calculating PSI values, the higher the number of reads, the more evidence there is for that PSI value, which is consistent with a narrower distribution. On the other hand, the mean value of the distribution is obtained by the ratio between \( \alpha \) and \( \alpha + \beta \), which is directly comparable with the ratio used to calculate PSI values (Equation 2), if \( \alpha, \beta \) are related to inclusive and exclusive reads, respectively. Taken together, these factors make the Beta distribution a good modelling function for quantification of PSI values [34].

This modelling function was used to obtain a global PSI value for a group of biological replicates, each of them with a PSI supported by different number of reads. By fitting a global Beta function to a set of points belonging to a distribution modulating each replicate, a global distribution would reflect both the differences between inclusive and exclusive reads, and the different levels of evidence supporting each replicate’s PSI, providing a more accurate PSI quantification. This method was used to summarize the PSI values for each AS event in each tissue of human and mouse.

### 3. Results and Discussion

#### 3.1 Alternative splicing quantification

The application of the Beta distribution, as explained before, was used to obtain the global PSI value for each event in a set of samples considered to be biological replicates. By taking the number of reads supporting each individual PSI, a Beta distribution of 500 points was generated. The \( n \) samples’ distributions were then concatenated and the global set of points were fitted as a Beta distribution, from whose parameters \( \alpha \) and \( \beta \) the global PSI (which is the mean value of the global distribution) was calculated.

![Betadistribution](image)

**Figure 5. Beta distribution modelling of exon inclusion levels**

Beta distribution as a model for PSI quantification aims to correct for differences in the number of reads supporting either inclusion or skipping of exons when taking several replicates. Top panel corresponds to a specific event PSI quantification using five human brain samples, while bottom panel corresponds to the PSI of another event, using two muscle samples.

Global PSI values obtained with the described Beta distribution method were used to compare PSI values for alternative splicing events described to be orthologs between human and mice. By correlation analyses of each
3.2 AS patterns during neuronal development and their regulatory mechanisms

Alternative splicing quantifications were used to search for events showing progressively more inclusion or skipping along the neuronal differentiation timecourse. Hubbard’s dataset was used for this analysis, after filtering out events with very low variances and PSI ranges across samples, correspondent to almost constant inclusion levels during development. Following, Spearman’s correlation was tested between each of the selected events’ PSI values and time in days after differentiation. Examples of significantly (FDR < 0.01) time-correlated events are depicted in Figure 6.

![Figure 6. Time-correlated alternative splicing events](image)

Examples of exons progressively more included or excluded along differentiation, selected after testing for Spearman’s correlation across several AS events in Hubbard’s dataset.

Interestingly, some of the events whose PSI evolution is depicted in Figure 6 refer to genes with relevant functions in neuronal development. The inclusion of the exon 4 in the \textit{Tmem63b} gene, a calcium channel transmembrane protein\(^3\), is close to 100% in undifferentiated stages but drops down around day zero, when neural induction occurred, and approaches zero in the later developmental stages. Although further exploration is needed, the neuronal-specific skipping of this exon may be of potential interest in terms of functional consequences, since calcium channels are important for synaptic function. Another interesting event is the progressive inclusion of exon 44 in the \textit{Cacna1a} gene. This gene encodes for calcium voltage-dependent channels that triggers neurotransmitter release in the synaptic cleft. Interestingly, the \textit{Cacna} gene family is known to generate tens or thousands of different proteins through AS, resulting in functionally tuned calcium channels efficiently performing membrane depolarization [35]. Moreover, the pool of \textit{Cacna} transcripts is changed between developmental stages and different neuronal types. The progressive inclusion of \textit{Cacna1a’s} exon 44, shown in Figure 6, may therefore be associated with specific functions in membrane depolarization.

Moreover, for the set of significantly time-correlated events, tests for correlation with RNA-binding proteins expression were performed. Interestingly, some of the time-correlated events are anti-correlated with the

\(^3\)http://www.uniprot.org/uniprot/ (Accessed in September, 2016)
expression of NOVA1 protein, reported to be a regulator of neuronal-specific AS. Three of those events, depicted in Figure 7, have their cognate genes associated to some aspects of neuronal function. Axin1 is reported to be involved in controlling the balance between neural progenitor proliferation and neuronal differentiation, as well as in the correct formation of neurite polarity and consequent neural connectivity [36]. Exon 9 of the Axin1 gene is progressively less included as cells proceed to the neuronal lineage, which may be related to this gene function in regulating timely production of neurons.

Also, Dip2a gene encodes for a receptor for signals involved in axon path finding and patterning and that has also been proposed to have a role in synapses specification through ligand/receptor interactions [37]. The timely decrease in Dip2a exon 5 inclusion levels may be associated with these functions. Finally, Mprip exon 9 is progressively less included as cells differentiate into neurons, as shown by decreasing PSIs in Figure 7. This protein is involved in neurite outgrowth and the same AS event shows differential splicing patterns in post-mortem brains of myotonic dystrophy type I, a genetic disease in the CNS in which patients show visual spatial and attention deficits or avoidant behaviour [38]. The regulatory mechanism of this AS events is to be further explored by analysis of enrichment in RBP’s preferred RNA-binding motifs in the regions within the vicinity of the exons.

3.3 Neuroprogenitor cell line 46C characterization

The Sox1-GFP reporter neuroprogenitor cell line (NPC 46C) is used to model neuronal differentiation in vitro. However, the transcriptomic profile of its differentiation had never been compared with those of other types of neuro-differentiating cells. In order to fill this gap, principal component analysis was performed on gene expression values from different neuronal differentiation datasets.

Principal components 1 to 3 accumulate most of the variance within the merged dataset (roughly 60%) and clearly distinguish neurodevelopmental stages. Moreover, three major tendencies show up: PC1 accumulates genes that show varying levels of expression between pluripotent stages and neuronal stages, while PC2 seems to distinguish fully differentiated tissues from developing ones. PC3 separates samples labelled as radial glia and neurons at medium developmental stages (DS12 labelling corresponds to developmental stages I/II) from other groups, and likely accumulates contributions from genes important in the function of early maturing neurons.

PCA based on AS levels was also performed and a similar neuronal differentiation coherent distribution is notorious, as can be seen in Figure 9. Principal component 1 clearly separates mature tissues and later developmental stages of neurons from earlier stages of development. This means that AS events also distinguish cells by their developmental stages.

![Figure 8. Principal component analysis of gene expression in neurogenesis/nervous system samples](image_url)

Sample distribution along the principal components is coherent with the developmental stage.

![Figure 9. Principal component analysis of exon skipping events in neurogenesis/nervous system samples](image_url)

Principal component 1 clearly distinguishes neurodevelopmental stages. Separation between pluripotent samples and mature tissues is clear.

4. Conclusions

This project focused on discovering switch-like AS events associated to specific neuronal development stages and exploration of their regulatory mechanisms. The Beta distribution was applied to the quantification of AS, since it
matches some PSI properties, namely its dependence on the number of reads supporting each event. This methodology also enabled comparison of orthologous alternative splicing events between human and mouse. Moreover, analysis of AS events shifting during neuronal differentiation provided promising results, serving as a strong motivation for future work. NOVA1 is a known regulator for neuronal-specific AS and its expression was found significantly anti-correlated with the inclusion levels of genes with functions in the shaping of neuronal networks. Preliminary characterization of gene expression and alternative splicing patterns of the NPC 46C cell line in those samples was also achieved.

**Future work** The implementation of a robust statistical method to compare two or more Beta distributions is important, in order to expand the applicability of the method to differential AS analysis. Moreover, RNA-binding motif enrichment statistical analysis is already being used in the lab, to explore the evidence for a regulatory link between RBPs and their potential target exons. This methodology should be applied to explore the regulatory mechanisms underlying PSI switches during neurodevelopment.

**Acknowledgments**

To this project’s supervisor Dr. Nuno Barbosa Morais, collaborators Dr. Manuel Irimia, Dr. Domingos Henrique, Dr. Ana Pombo and Dr. Sérgio Almeida, and all the colleagues from NMorais Lab (Computational Biology) at Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa.

**References**


