University of Crete

MASTER'S THESIS

Genetic variation and expression spliceoform study of a human SLE cohort

Author:

Sofia Papanikolaou

Supervisors:

Christoforos Nikolaou

George Bertsias

MSc Program in Bioinformatics

Medical School

September 2020

Table of Contents

Acknowledgements

This Master's thesis was conducted at the Computational Genomics Laboratory at the Biology Department of the Univeristy of Crete, Greece, in collaboration with the Laboratory of Rheumatology, Autoimmunity and Inflammation at the Faculty of Medicine, under the guidance of Dr. Christoforos Nikolaou, who is the leader of Computational Genomics Group and Dr. George Bertsias, who is Assistant Professor in Rheumatology - Clinical Immunology and group leader of Laboratory of Rheumatology, Autoimmunity and Inflammation. I would like to express my sincere gratitude to them for their support and for giving me the opportunity to conduct my thesis.

I warlmy thank my mentor, Dr. Christoforos Nikolaou, who has introduced me in the field of computational genomics and always inspires me to be a better student with his trust and encouragement. I would like to express my deep appreciation to Dr. George Bertsias, whose assistance and willingness was invaluable for me. I would also like to thank my tutor and committee member Pavlos Pavlidis for his wise counsel and positivity.

This thesis would have not been completed without my co-workers and friends Eleni Lianoudaki and Ilia Varamogianni who always listen, support and understand me, as well as Lydia, Antonis and Emilios, who created a wonderful atmosphere in lab and made me feel comfortable. Special thanks to Stella for her company and support. I also thank Despoina, Chrysa and Giwta for our beautiful cooperation during this year.

I am extremely grateful to my family and my friends for their love, caring and understanding.

Abstract

The splicing of pre-mRNA is a key step in gene expression. Alternative splicing increases the complexity of the transcriptional output but in many cases it may also lead to the reduction of mRNA translation or to the production of non-functional or malfunctional proteins. The aim of this study is to detect and characterize the alternative splicing events that occur in patients suffering from Systemic Lupus Erythematosus, an autoimmune disease characterized by the production of auto-antibodies in the blood. To address the connection between SLE and alternative splicing, we used a bioinformatics tool to identify and quantify alternative splicing events based on RNA-Seq data from whole blood samples of 142 patients with SLE and 58 healthy individuals. We also performed differential alternative expression analysis between the two groups and determined the statistically significant alternative splicing events. Through computational approaches, the introduction of premature termination codons in the splice variant was examined in order to find out its possible effect to the mRNA fate. The affected genes were found to be implicated in immune processes such as antigen presentation and lymphocyte proliferation, while several genes have been associated with autoimmune diseases. This study supports a potential contribution of the alternatively spliced genes in SLE pathological processes, suggesting several biomarkers.

Περίληψη

Η συρραφή του πρώιμου μορίου mRNA διαδραματίζει καθοριστικό ρόλο στη γονιδιακή έκφραση. Η εναλλακτική συρραφή, μέσω διαφορετικών συνδυασμών εξονίων αυξάνει τη μεταγραφική πολυπλοκότητα, ωστόσο σε κάποιες περιπτώσεις μπορεί να οδηγήσει στη μείωση της πρωτεϊνοσύνθεσης ή και στην απώλεια της λειτουργικότητας των παραγόμενων πρωτεϊνών. Σκοπό της μελέτης αποτελεί η ανίχνευση και ο χαρακτηρισμός των γεγονότων εναλλακτικού ματίσματος που συμβαίνουν σε ασθενείς με Συστηματικό Ερυθηματώδη Λύκο (ΣΕΛ), ένα αυτοάνοσο νόσημα που χαρακτηρίζεται από την παραγωγή αυτοαντισωμάτων. Για την διερεύνηση της σύνδεσης μεταξύ του ΣΕΛ και της διαδικασίας του εναλλακτικού ματίσματος, χρησιμοποιήθηκε ένα εργαλείο βιοπληροφορικής για να ταυτοποιήσει και να ποσοτικοποιήσει τα γεγονότα εναλλακτικού ματίσματος σε δεδομένα μεταγραφώματος (RNA-Seq) που προέρχονται από δείγμα ολικού αίματος υγιών ατόμων και ασθενών με ΣΕΛ. Επιπλέον, πραγματοποιήθηκε διαφορική ανάλυση μεταξύ των δύο καταστάσεων και προσδιορίστηκαν τα στατιστικά σημαντικά γεγονότα. Μέσω υπολογιστικών προσεγγίσεων εξετάστηκε η εισαγωγή πρόωρου κωδικονίου λήξης και η πιθανή επίδραση του στην τελική έκφραση του mRNA. Τα γονίδια που υπόκεινται σε διαφορική εναλλακτική συρραφή βρέθηκε ότι εμπλέκονται σε διαδικασίες του ανοσοποιητικού συστήματος όπως η αντιγονοπαρουσίαση και ο πολλαπλασιασμός των λεμφοκυττάρων, ενώ κάποια γονίδια έχουν συσχετιστεί με αυτοάνοσα νοσήματα. Συνεπώς, η μελέτη αυτή υποστηρίζει τη δυνητική συνεισφορά των εναλλακτικά συρραφόμενων μεταγράφων στην παθογένεση του ΣΕΛ, προτείνοντας γονίδια δείκτες της νόσου.

1. INTRODUCTION

1.1. Splicing

Pre-mRNA splicing is a post-transcriptional mechanism by which introns are removed and exons are joined together leading to mature mRNA molecule formation. Different coding regions can be retained or not in the final transcript, a process called alternative splicing that generates novel mRNA isoforms. This procedure results in transcriptome and proteome diversity, as a gene can produce multiple transcripts. It is estimated that almost 95% of genes with multiple exons are subject to alternative splicing [1] and a gene can express approximately 10-12 unique isoforms [2].

1.1.1. Mechanism of splicing

The spliceosome, the basal splicing macromolecular machinery, is made up of five small nuclear ribonucleoproteins (snRNPs), U1, U2, U4, U5, and U6, and other additional proteins [3], catalysing a two-step transesterification reaction. Spliceosome recruitment on mRNA is achieved by recognition of several elements that define the exon-intron boundaries of pre-mRNA (**Figure 1.1**). These elements are the 5' and 3' splice sites (SS) which are most commonly conserved dinucleotide sequences located at 5' and 3' end of introns, a "GU" at the 5' end and an "AG" at the 3' end respectively [4]. Upstream from the 3' SS there is the polypyrimidine tract (Y(n)), a region high in pyrimidines (C and U) which is preceded by a branch site that typically includes an adenosine. In the first step, the 5′ splice site is recognized by the U1 snRNP complex. The U2 auxiliary factor (U2AF), composed of a large and a small subunit, binds to Py and 3' SS, while the large subunit of U2AF, binding to Py, interacts with Branchpoint Binding Protein (BBP) to help it bind to the branch point [5]. These molecules form the spliceosomal E complex. Subsequently, U2 snRNP displaces BBP and binds to branch site (A complex) and after the recruitment of U4/U6.U5 tri-snRNPs [6] and the release of the U1 and U4 snRNP, spliceosome enters its catalytically active conformation. The first transesterification reaction generates a free 5′ exon and an intron

lariat-3′ exon and the second leads to ligation of exons joining and removal of the lariat intron (**Figure 1.2**).

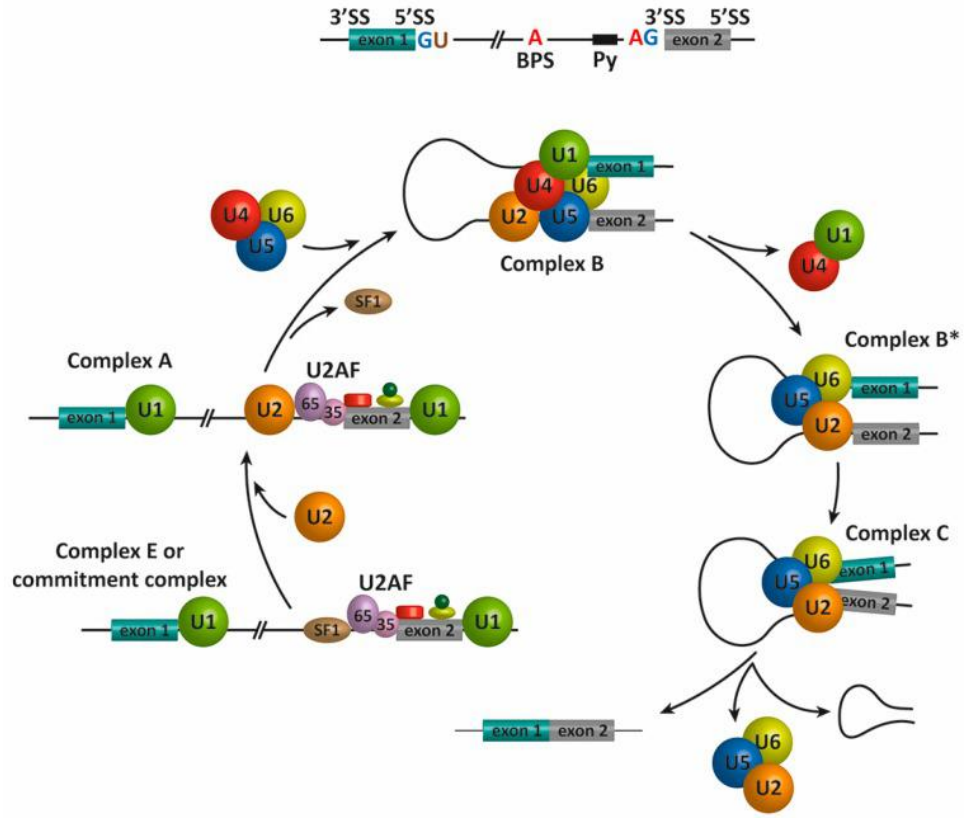


Figure 1.1 | Spliceosome assembly. Firstly, the 5′ splice site is recognized by U1 snRNP and the branch point by the splicing factor SF1/BBP, resulting in Complex E. Complex A is formed after displacement of SF1/BBP by the U2 snRNP. U4/U5–U6 tri-snRNPs bound to pre-mRNA, forming Complex B. The release of U1,U4 snRNPs leads to Complex B*. The final formation Complex C catalyses the two trans-esterification reactions [7].

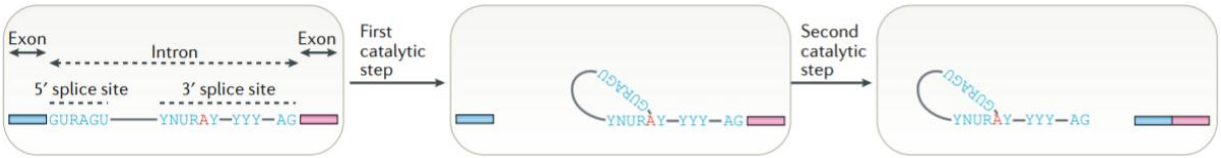


Figure 1.2 | Transesterification reaction [8].

1.1.2. Regulation of alternative splicing

The splice site composition and therefore the complementarity between splice sites and snRNAs of the spliceosome determine the alternative splice sites' selection. In other words,

stronger binding of pre-mRNA to spliceosomal components leads to retention of this region in the final mRNA molecule [9]. Additional signals that mediate spliceosome assembly are pre-mRNA cis regulatory motifs, known as Splicing Regulatory Elements (SREs) and are classified according to their position and their splicing effect into Exonic Splicing Enhancers (ESEs), Exonic Splicing Silencers (ESSs), Intronic Splicing Enhancers (ISEs) and Intronic Splicing Silencers (ISSs) (**Figure 1.3**). These elements contribute to regulation of alternative splicing, as they are recognized from auxiliary splicing factors, particularly RNA-binding proteins, like specific serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) [10] that modulate the efficiency of splice site recognition. Specifically, ESEs bind SR proteins and promote recruitment of spliceosome and inclusion of the exon they reside in. ISSs along with ESSs usually bind heterogeneous nuclear ribonucleoproteins, to repress the usage of a splice site [11] [12]. It has been reviewed that splicing factor gene expression signatures are involved in tissue-specific patterns of alternative splicing [13] [14], while splicing transitions take place during development, providing remodelling of transcriptome and diversity of proteome [15]. The formation of RNA secondary structures can result in the masking of splice sites and subsequent inhibition of regulatory elements that are normally recognised from splicing factors to modulate splicing [16].

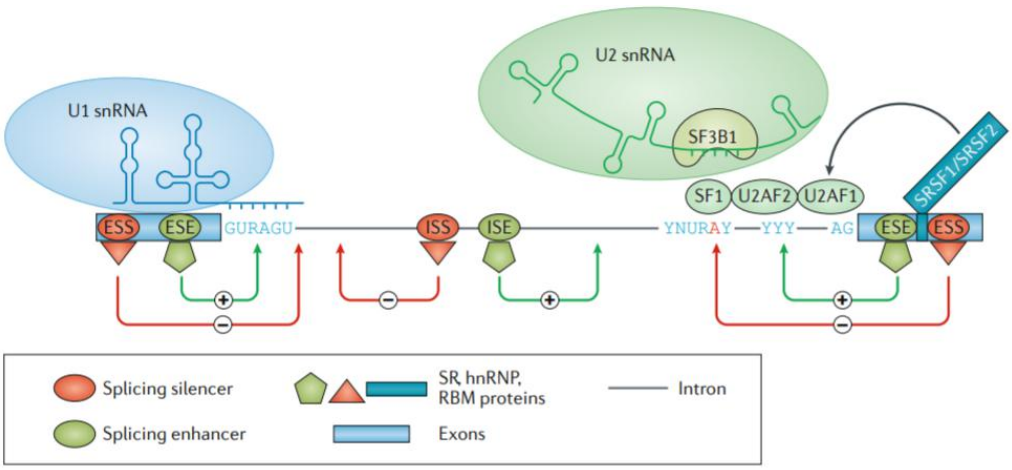


Figure 1.3 | The trans-acting factors harbour cis-regulatory elements that promote alternative splicing. The exonic splicing enhancers (ESE) and intronic splicing enhancers (ISE) promote exon inclusion, while exon repression requires the exonic splicing silencer (ESS) and intronic splicing silencer (ISS) elements [8].

The transcriptional environment contributes to the regulation of alternative splicing. Recent studies have revealed that transcription and splicing can occur simultaneously and assembly of spliceosome is preceded by termination of transcription [17] [18]. Based on the evidence of co-transcriptional splicing, it is assessed that RNA polymerase elongation rates can promote or prevent recognition of competing splice sites or regulatory cis-acting motifs [19]. The kinetic model of coupling transcription elongation and splicing suggests that high transcription elongation time leads to generation of competing exons and subsequent inclusion of downstream exons and exon skipping of upstream. On the other hand, slow transcription elongation favours the retention of alternative exons, as it prevents competition between strong and weak splice sites [20].

Except from elongation rate, the C-terminal domain of RNA polymerase II (CTD) particularly influences splicing via altering the trans-acting factor recruitment to pre-mRNA [21]. Specifically, the presence of CTD facilitates the binding of snRNPs and regulatory factors to pre-mRNA and several studies provide evidence that modifications of the CTD result in less efficient splicing [22].

Another contributing factor that affects different splicing outcomes is chromatin structure. Nucleosome positioning defines exon - intron boundaries, as it has been shown that nucleosome occupancy is higher in exons and especially exons flanked by weak splice sites – compared to introns [23].

1.1.3. Patterns of alternative splicing events

The basic patterns of alternative splicing events are depicted in **Figure 1.4**. It has been revealed that skipping of a cassette exon is the most frequent event in metazoans, usually containing long neighbouring introns [24]. The second type is intron retention, when introns are included in the final transcript [25]. Another two instances are alternative usages of 5′ donor splice sites or 3′ acceptor splice sites within exons. It is estimated that the distance between constitutive and alternative splice site is about 2-12 nucleotides [26]. In addition, there are mutually exclusive exons, where the inclusion of an exon prevents the inclusion of another [27].

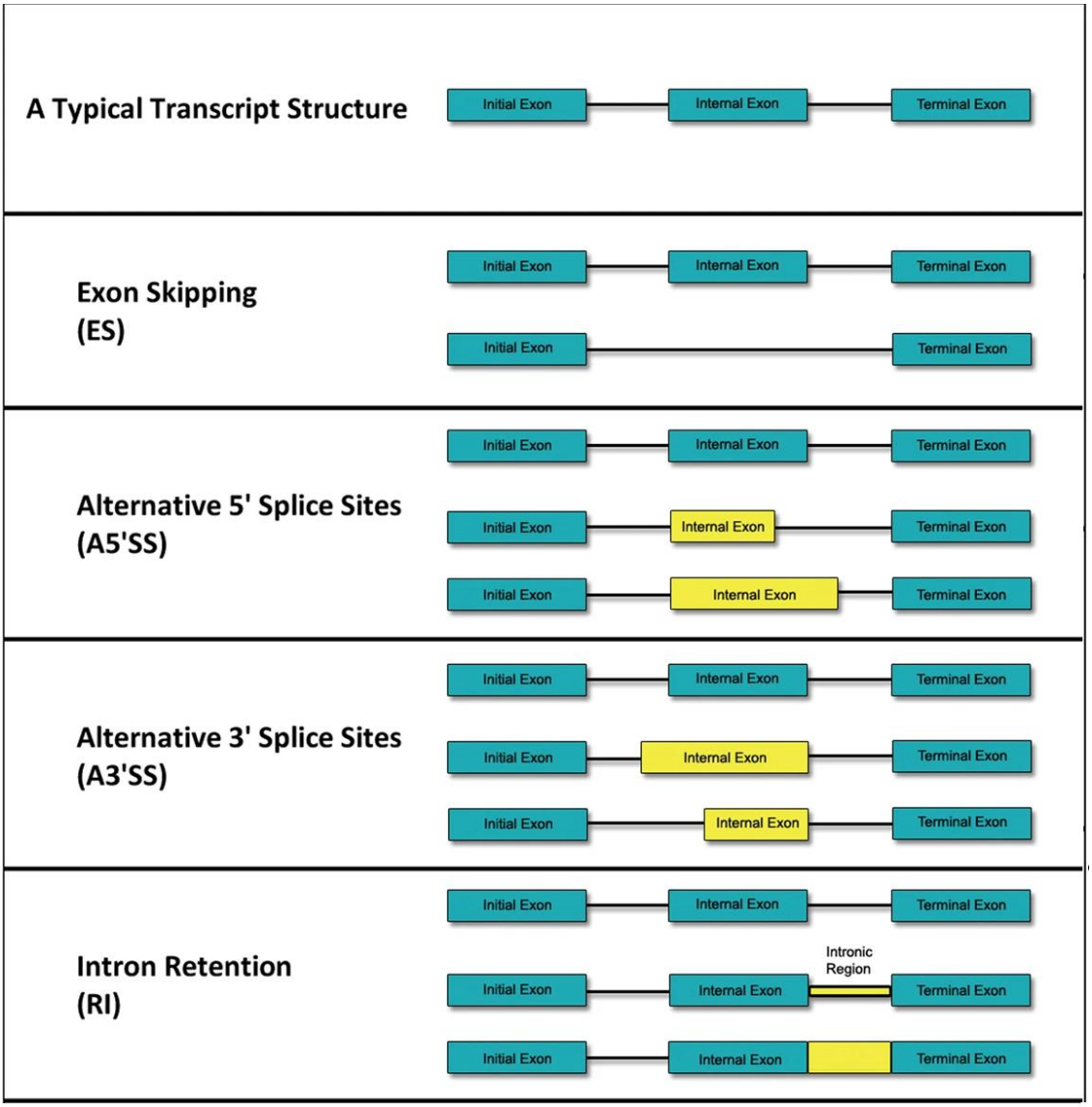


Figure 1.4 | A typical transcript structure consists of internal exons between an initial and a terminal exon and internal introns that are going to be removed. Basic types of alternative splicing events: a) In case of exon skipping, an internal exon is not included in the final transcript. When the retention of an exon prevents this of another, these exons are called mutually exclusive. b) The selection of an alternative 5' splice site alters the 3' end of the upstream exon and c) an alternative 3' splice site leads to change of downstream exon's 5' end respectively. d) Intron retention occurs when an intronic region is not excluded from the final transcript. (Modified from [28])

1.1.4. Impact on gene expression

The vast majority of the affected sites of mRNA belongs to Open Reading Frames (ORFs) resulting in disruption of it or introduction of premature termination codon (PTC) and, thus, these changes may influence the structure and the function of the produced proteins. Splice variants containing a premature termination codon can be subject to elimination via nonsense mediated mRNA decay (NMD), a mechanism that reduces errors arisen during transcription

by preventing translation of aberrant mRNA molecules [29] . The decision as to whether a splice variant that carries a PTC will have this fate depends on the existence of an exon-exon junction over 50 nucleotides downstream of the PTC (**Figure 1.5**) [30] , [31] , [32] . This procedure leads to mRNA destruction. Alternatively, mRNA traslation will take place and be completed after ribosome reaches the premature termination codon. Alternative splicing can also create variability in non-coding regions, regulating mRNA stability, subcellular localization and translation efficiency.

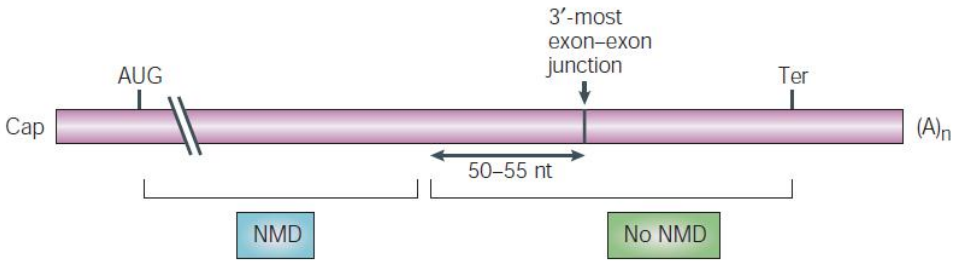


Figure 1.5 | Definition of PTCs that prompt NMD. Premature termination codons positioned in distance of 50-55 nucleotides from downstream exon-exon junctions can elicit nonsense-mediated mRNA decay [31].

1.2. Splicing and diseases

1.2.1. Splicing and immunity

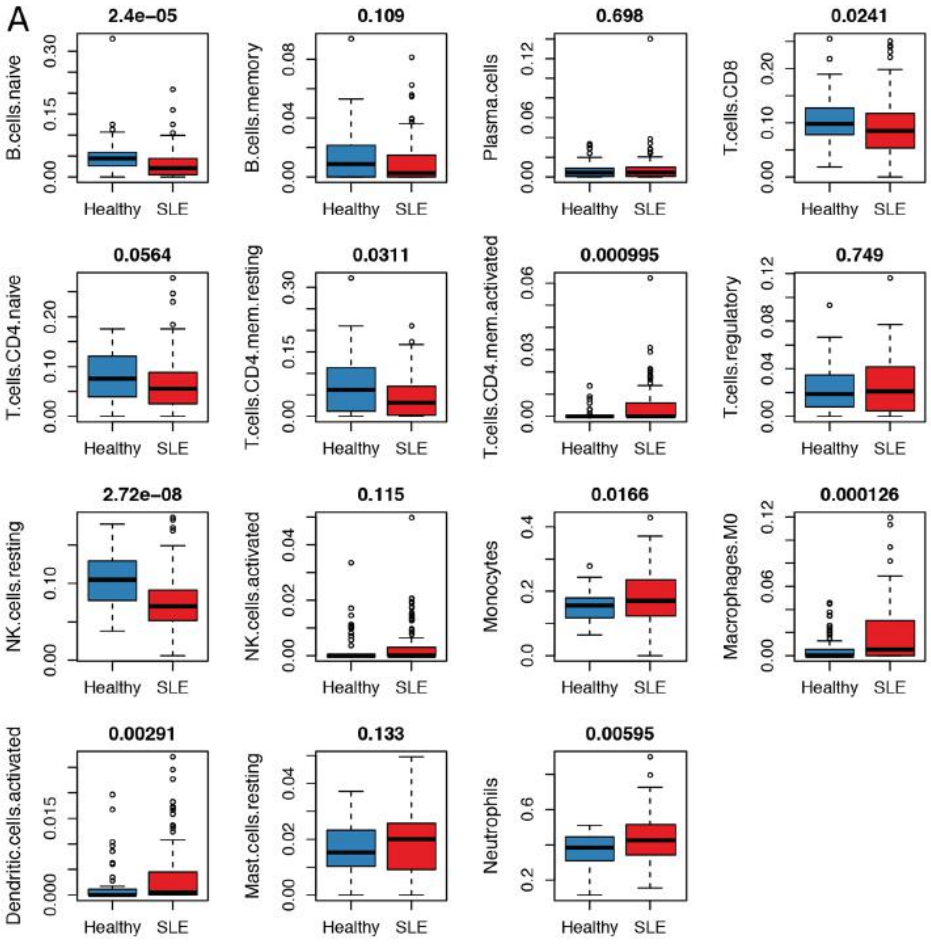
The immune system requires a complex regulatory mechanism as it has to accomplish self - nonself discrimination and, therefore, alternative splicing, which generates extensive protein diversity, is a key contributor to these. Genome-wide studies have described how the immune system utilizes the mechanism of alternative splicing in adaptive immune cell migration, differentiation and regulation of apoptosis [33]. With regard to immunoglobulin genes, it has been found that different isoforms of *VH* gene resulted from alternative splicing are involved in class switch from IgM to IgD [34]. Another example of splicing event takes place during B cell differentiation, where the secretory form of IgM is generated through splicing and cleavage-polyadenylation at the alternative poly(A) site of the transcript [35]. Also, different splicing variants of IgE transcripts are expressed from B cells, according to the stimuli they receive from their microenvironment [36] . Previous studies have revealed that T cell activation is linked to alternative splicing events related to variable exon usage [37], as in the

case of Mucosa-associated lymphoid tissue lymphoma translocation protein 1 (*MALT1)* gene that promotes CD4+ T cell activation [38] as well as Protein tyrosine phosphatase receptor type C (*PTPRC* or *CD45*) splicing patterns that lead to regulation of activation in different levels [39]. Additionally, splicing events can control innate immune responses. It has been described that upon lipopolysaccharide (LPS) treatment an alternative isoform of TLR4 is generated leading to the inhibition of TNF and NF-κB signalling and a reduction of immune response in mouse macrophages [40]. In another aspect of the effects of alternative splicing on the immune system, the generation of novel epitopes and abnormal autoantigen expression may result in modulation of immunogenicity and subsequent autoimmunity [41].

The aforementioned cases underline the significance of regulation of the splicing process in the balance between immunity and tolerance and the perturbation of splicing has been implicated in several diseases. A recent review in alterations of splicing in Multiple Sclerosis has summarized and reassessed findings of previous studies, suggesting potential biomarkers [42]. Genes investigated the most are usually derived from transcriptomic analyses of whole blood samples and are involved in inflammatory pathways, like Interleukin 7 Receptor (*IL7R*), whose soluble isoform generated through an exon skip is increased in disease state [43], or Myelin oligodendrocyte glycoprotein (*MOG*), which is a candidate autoantigen implicated in the pathogenesis of Multiple Sclerosis too [44] . Another autoimmune disease, Type 1 Diabetes, is characterised by alternative exon usage of a specific gene set among lymphocytes (CD4+, CD8+, CD19+) and a group of genes involved in autoimmunity that had been examined for splicing, it was found to be alternative spliced [45].

In the context of Systemic Lupus Erythematosus (SLE), an autoimmune disorder characterized by production of autoantibodies against nuclear and cytoplasmic antigens and multisystem inflammation, previous studies have revealed that genes associated with SLE pathogenesis are subject to alternative splicing. Specifically, several events of alternative splicing have been reported to affect genes enriched in immune signalling pathways, like exon skipping in the gene for the B-cell scaffold protein with ankyrin repeats (*BANK1*) and an alternative 3' splice site of leukocyte immunoglobulin like receptor A1 (*LILRA2*). Consequences of both splicing events are the production of a novel isoform that may lack a binding region [46], [47] . Moreover, Odhams CA et al. attempted to identify sQTLs, i.e. DNA sequences linked with variation in splicing isoform abundance, that are related to SLE [48]. Among the genes having sQTLs are *TCF7*, *SKP1*, *BLK*, *NADSYN1*, *IKZF2*, *WDFY4* and *IRF5*.

Panousis et al. in 2019 conducted a transcriptome analysis of RNA sequencing data of samples isolated from 142 SLE patients and 58 healthy individuals' whole blood [49]. Each sample had different cell type frequencies that had to be assessed for the purpose of determining the differentially expressed genes (**Figure 1.6 A**). After correction for cell variation, a functional assessment of differentially expressed genes showed enrichment in immune signalling pathways like p53, IFN and NOD-like receptor and molecular processes like RNA transport (**Figure 1.6 B**). Furthermore, quantification of alternative splicing events revealed perturbed mRNA splicing of SLE patients and splicing QTL analysis indicated a set of spliced and differentially expressed genes having sQTL too. The alternative spliced genes mediate immune system and type I interferon signalling pathways.



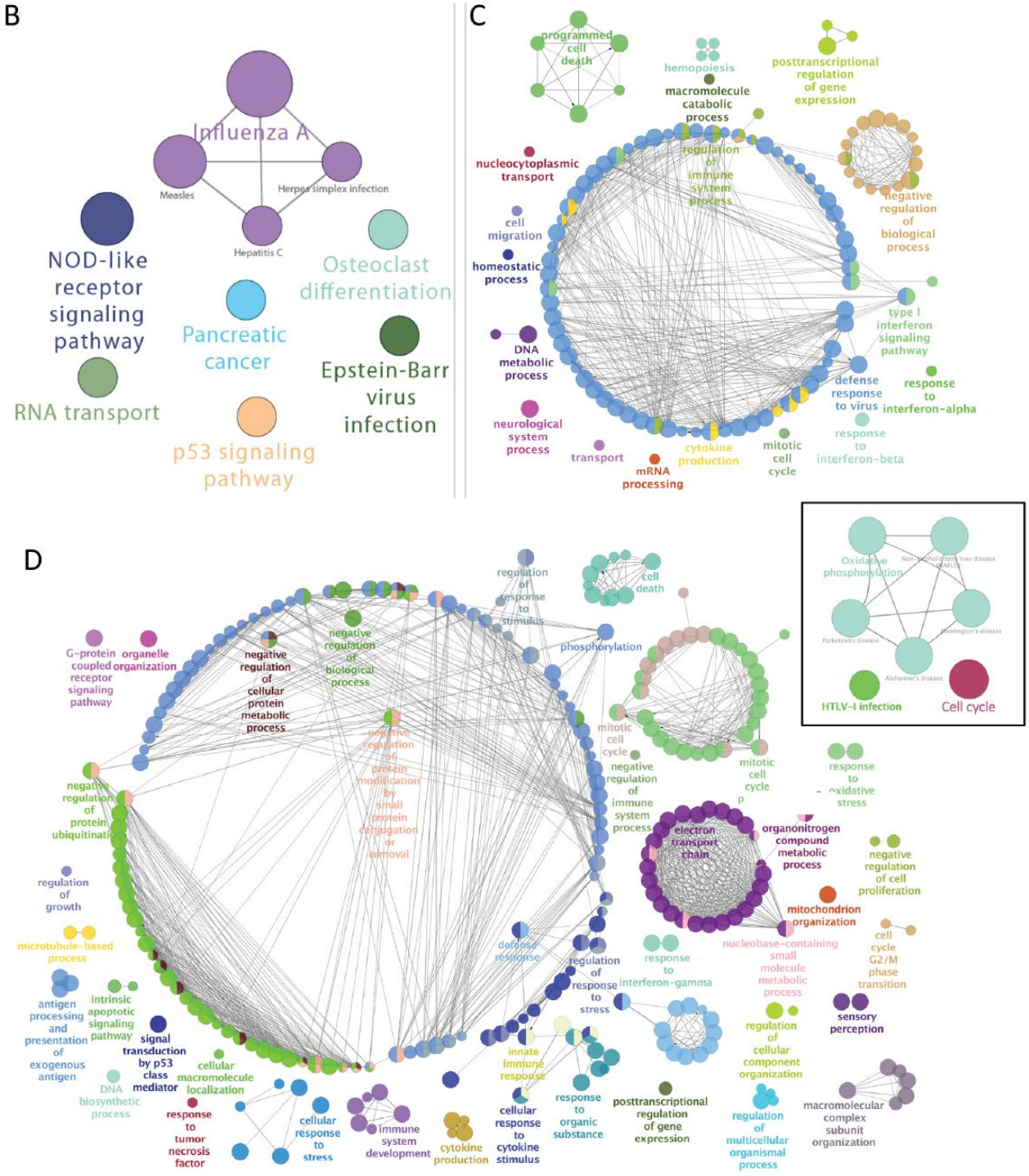


Figure 1.6 | Transcriptome analysis of Panousis et al dataset. **A.** Proportion of cells contained in Healthy and SLE samples. **B.-C.** Functional analysis of DEGs and Mapping of top enriched gene ontology terms, respectively, after estimation of cell variation. **D.** KEGG pathways of DEGs correlated with clinical activation.

1.3. Computational approaches for splicing analyses

Identification of alternative splicing can be achieved through analysis of transcriptomic data generated from RNA sequencing or microarray technologies. RNA sequencing has the ability to recognise novel events, although its requirements for computing power are much higher. Splicing analysis is accomplished either by full-length transcript reconstruction or by direct event detection, where reads that span exon-exon junctions are essential [50], [51] . Since isoform reconstruction needs long read length and high sequencing accuracy and is not computationally efficient, straightforward event detection is mainly used [52] . The methods supported for the detection can be de-novo assembly or reference-based and some of the criteria that need to be considered in selecting them are their ability to classify the splicing events and to provide statistics. For the quantification of splicing events the "percent spliced-in" is proposed as the most common metric. This estimates the efficiency of a region to be spliced into the transcripts of a gene and is equal to the percentage of the transcripts containing a spliced region of interest compared to the total transcript population [53].

2. PURPOSE OF THE RESEARCH

This project focuses on the identification of alternative splicing events and their differential analysis between healthy individuals and patients suffering from SLE and explores the introduction of premature termination codon and the shift in the translational reading frame of spliced genes. Moreover, it is also examined whether the generation of a premature termination codon would elicit an abnormal transcript leading to a truncated protein, or if non-sense mediated decay would occur and prevent early termination of translation. These changes could cause loss of function of the affected genes or decreased protein expression and therefore these genes may be candidate genes markers implicated in the pathogenesis of disease.

3. METHODS

3.1. SplAdder Tool

SplAdder is a tool for analysis of alternative splicing of data coming from high-throughput sequencing of RNA (RNA-Seq) [54]. In order to detect and quantify splicing events of input samples, SplAdder performs a splicing graph construction from reference genome and enriches this graph, taking advantage of alignment files information. Preferentially, it supports a test for differential analysis between groups of samples.

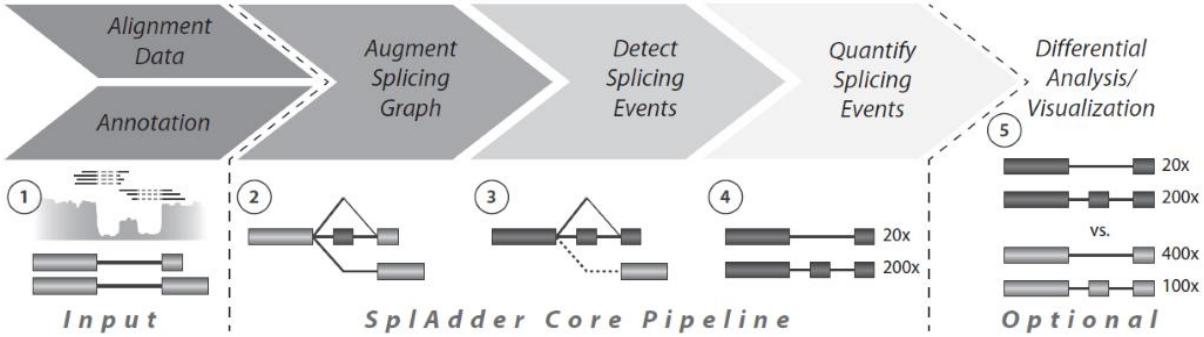


Figure 3.1 | SplAdder Pipeline.

Algorithm

1. Initial graph construction

The preliminary step of the analysis is the construction of a splicing graph, given the annotation information. The splicing graph represents a gene with transcripts as nodes and exons as edges, where a single node can describe multiple transcripts that possess a common exon.

2. Graph augmentation

Secondly, the splicing graph of each gene is augmented with a combination of events identified from input files. As a result, the final graph contains novel splice connections that may not be annotated.

3. Event detection

4. Graph quantification

Event types that can be extracted from the augmented graph consist of exon skipping, intron retention, alternative 3' and 5' splice site usage, multiple exons skipping and mutually exclusive exons. Each event is illustrated as a mini gene with two isoforms, the constitutive one and its alternative and is quantified on the basis of the number of alignments supporting the event and the mean coverage of each exon. These calculations are performed on the segment graph, a transformed presentation of splicing graph, where each exon is split into non-overlapping segments (**Figure 3.2**). Therefore, each exon is considered once, improving computational efficiency. Additional filters are used for confirmation, depending on each event type.

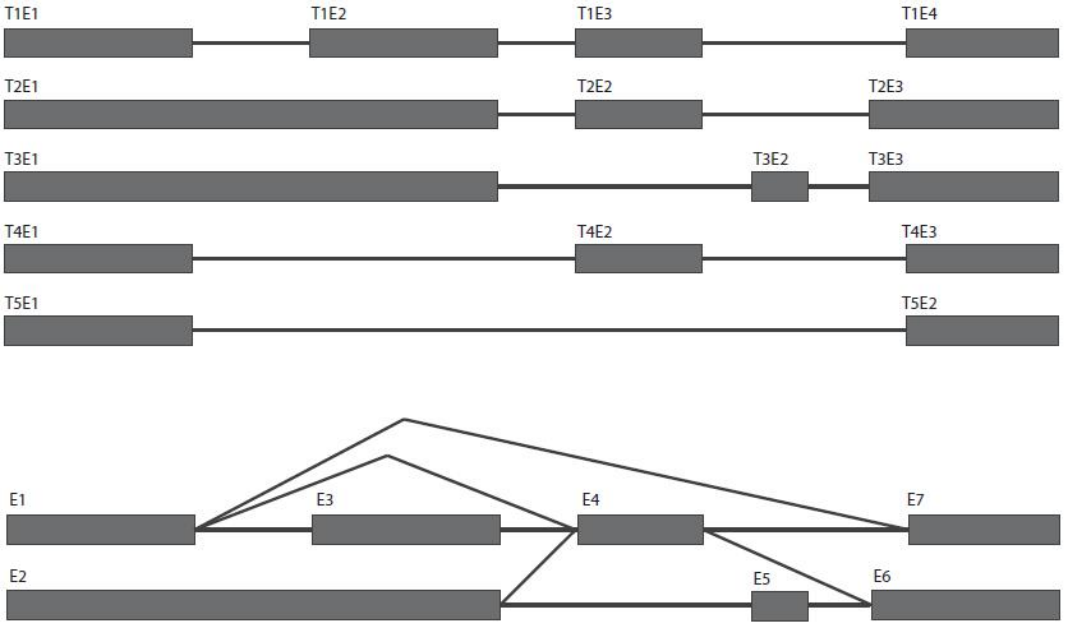


Figure 3.2 | Illustration of conversion of a transcript graph (A) into a splicing graph (B).

5. Differential analysis

Examination of the differential usage of alternative splicing events under specific conditions can be conducted with two approaches, using another tool or directly implementing tests provided by SplAdder. In the last approach, a generalized linear model is used for each splicing event i, in order to estimate the expected counts μ given the observed read counts y, assuming to be generated from negative binomial distribution.

Expression or splicing read count y for each splicing event i:

$$y^i \sim NB(\mu^i, \kappa^i)$$

where μ^i is the expected count and κ^i is the estimated dispersion across samples.

Null hypothesis H_0: $log(\mu^i) = \beta^i_0 + \beta^i_{0xpr} + \beta^i_{\Delta\ expr}$

Alternative hypothesis H_1: $log(\mu^i) = \beta^i_0 + \beta^i_{0xpr} + \beta^i_{\Delta\ expr} + \beta^i_{\Delta\ spl}$

where β^i_0 is the coefficient denoting the intercept, β^i_{0xpr} is the contribution of the observed count to the μ^i due to gene expression, $\beta^i_{\Delta\ expr}$ represents the distinction between two conditions at the expression level and $\beta^i_{\Delta\ spl}$ represents the splicing difference (alternative splicing) is included in the alternative hypothesis model.

After fitting the count data into H_0 and H_1, difference of deviances of the two GLM are examined in a x2-test. Benjamini-Hochberg adjustment is applied for multiple testing correction.

Limitations

It is worth noting that Spladder requires both alternatives of an event to be supported by reads of samples, even if they do not come from the same sample. Also, SplAdder does not support the detection of events spanning transcript start or end sites, so detection of alternative transcription start sites cannot be achieved. Additionally, events that share the same inner coordinates but different outer, are merged into one. For example, in **Figure 3.3**, although there are two exon skip events affecting two different transcripts (same skipped exon but different flanking exons), they are merged into one event. Under these circumstances, information of flanking exons coordinates cannot be exploited for recognition of their specific transcripts.

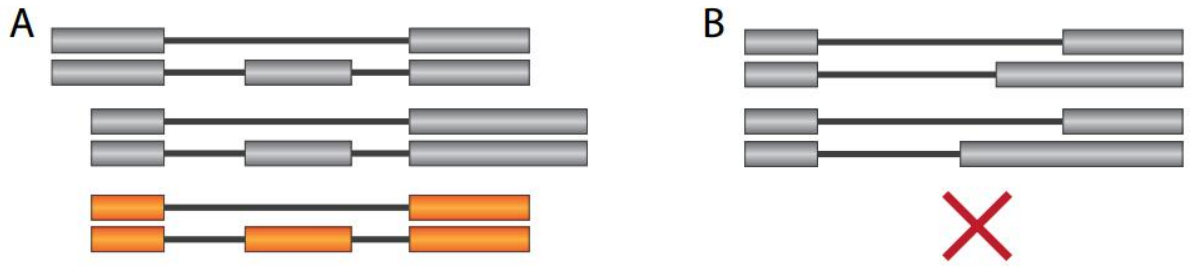


Figure 3.3 | Merging of overlapping events. A. Both events share the same inner coordinates of flanking exons but different outer, so they are merged into one event. **B.** In this case both events have only one intron in common for isoform 1, whereas the introns of isoform 2 disagree, so events cannot be merged.

3.1.1. Installation

SplAdder tool is implemented in python and requires a few packages like numpy, matplotlib, scipy, intervaltree, h5py, pysam and statsmodels (**Table 1**). As these packages and their dependencies may have conflicts between different environments and affect other installed tools, we create a conda environment, which is a package manager that keeps working environments isolated from one another, containing the specific packages SplAdder requires.

Table 1 | Packages required for Spladder

Dependencies	Version of installed packages
numpy>=1.14.6	1.17.4
matplotlib>=2.2	3.1.2
scipy>=1.3	1.3.2
intervaltree>=3.0.0	3.0.2
h5py>=2.2	2.9.0
pysam>=0.15.3	0.15.3
statsmodels>=0.9.0	0.11.0

3.1.2. Analysis pipeline

SplAdder supports three run modes. The first one is “build” mode, which is a prerequisite for the next. Construction of splicing graphs from RNA-Seq data and extraction of alternative events take place through different steps. These involve the transformation of an annotation file into a specific format, providing transcript information and splicing graphs per gene (initial graph construction) and, subsequently, the combination of different alignment files into the splicing graph, adding new edges and nodes to the existing graph, due to the novel splicing events (graph augmentation). After this, edge and node weights are quantified based on relative expression levels and different types of splicing events are extracted (graph quantification and event detection).

SplAdder accepts GTF or GFF format for annotation file, which has to refer to the same genome version that has been used for generating the alignment files. In this analysis annotation file was obtained from GENCODE project, a part of ENCylcopedia Of DNA

Elements (ENCODE) project that aims to identify and map of all gene features in the human and mouse genomes at a high accuracy [55]. Specifically, we used the main annotation file of Gencode release 19 of genome assembly version GRCh37 [56] , containing the comprehensive gene annotation on the reference chromosomes only. It should be mentioned that the default human gene annotation in the Ensembl browser is the current version of GENCODE. To display Gencode release 19 in Ensembl browser, Archive EnsEMBL release 75 - February 2014 can be used [57].

Alignment files were in bam format and, as mentioned above, were produced after mRNA sequencing in blood of 142 SLE patients and 58 healthy individuals by an already performed procedure that included mapping of the paired-end reads of 49 bp length to the GRCh37 reference human genome, using a splice aware mapper, called GEM.

The detected events: exon skipping (*exon_skip*), intron retention (*intron_retention*), alternative 3' splice site (*alt_3prime*), alternative 5' splice site (*alt_5prime*), mutually exclusive exons (*mutex_exons*), multiple (coordinated) exons skip (*mult_exon_skip*) can be further examined. A mode for differential analysis is provided in order to test alternative isoform usage between healthy and SLE group.

3.1.3. Output format

- **Build mode**

A set of result files is generated for each type of event. These include a distinct pattern of output filenames: "**merge_graphs__***event_name*_**C***confidence_level*.*file_format*".

- Confidence level

  There are four confidence levels controlling the quality of input alignment files. Level 3, which is the default level, corresponds to the highest level of filtering. Maximum number of mismatches, minimum number of alignments, minimum anchor length and maximum intron length are determinant parameters of input filtering.

- File Formats
  - HDF5 Format

    These files contain all relevant event information and are stored in the hierarchical data format HDF5, presenting complex data objects.
  - TXT Format

Event files in “txt” format contain essentially the same information as the HDF5 files for each confirmed event (**Table 2**).

Table 2 | Entries per line in .txt files.

Contents
Chromosome
Strand
Unique id
Gene
Start
End
Features

- GFF3 format

“Gff3” annotation files contain information for each event, depictured as a mini gene consisting of two different isoforms, constitutive and alternative. In particular, for (multiple) exon(s) skip events, isoform 1 consists of the previous and the next exon of skipped exon(s) and isoform 2 is the constitutive isoform, containing all involved exons (skipped exon(s) and the closest of them). Intron retention events are described as two isoforms where isoform 1 consists of the previous and the next exon of the retained intron and isoform 2 has only a novel exon, that includes retained intron and its closest exons. Alternative 3’ or 5’ splice sites events are also defined as two isoforms. For genes on the plus strand, the upstream nodes are considered as isoform 2 in the case of 3’ ss and as isoform 1 in 5’ ss cases. Respectively, for genes on the minus strand, the upstream nodes are considered as isoform 1 in the case of 3’ ss and as isoform 2 in 5’ ss cases.

Table 3 | Entries per line in ‘gff3’ files.

Contents
Chromosome
Event Type
Feature
Start
End
Strand
Event id
Gene name

- **Test mode**

A set of result files is generated for each type of event, including a distinct pattern of output filenames:

test_results_C*confidence_level_event_name***.gene_unique.tsv**
test_results_C*confidence_level_event_name***.tsv**
test_results_extended_C*confidence_level_event_name***.tsv**

As its name implies, "test_result" file includes statistical data obtained by differential analysis of each alternative splicing event between two conditions. In our analysis, condition A corresponds to healthy state and condition B to disease state. Gene name, p-value, adjusted p-value, average of event counts for each condition, fold change of event counts between the conditions in terms of logarithm to base 2 ratio, average of gene expression counts for each condition and fold change of gene expression counts between the conditions in terms of logarithm to base 2 ratios are given for each event (**Table 4**). Multiple alternative splicing events may affect the same gene, so the "gene_unique" file contains only one event for each gene, the most statistically significant one. "test_results_extended" file additionally includes event and gene counts for each sample.

Table 4 | Entries per line in .tsv files.

Contents
Event id
Gene name
P value
P value adjusted
Mean event count for condition A
Mean event count for condition B
log_2(Fold Change of mean event counts between conditions) *
Mean expression count for condition A
Mean expression count for condition B
log_2(Fold Change of mean expression counts between conditions) *

*** log_2(Fold Change) = log_2(mean count of condition A / mean count of condition B)*

- **Viz mode**

This mode provides several types of plots, like coverage, splicegraphs, event structure and transcript structure for a specific or multiple events of a gene.

3.2. Data analysis

In the first step of the analysis, statistically significant splicing events were extracted. To determine whether an event was statistically significant, p-value or adjusted p-value was used as statistical confidence measure and then compared to a confidence threshold. Quantile-Quantile plots compare distributions of p-values (or adjusted p-values) by plotting their quantiles. If points lie on a straight line, there is no evidence for deviation of observed p values from expected ones. In our analysis, QQplots suggested that there was a deviation of adjusted p-values between observed and expected values (**Figure 3.4-5**). Additionally, there was a little amount of statistically significant alternative splicing events using adjusted p-value as a measure of statistical significance. Therefore, we defined statistically significant events according to p-values instead of adjusted ones, using an alpha level of 0.01 as the threshold of significance. Analyses of statistically significant events and graphical representations were implemented in R programming language **(Script 1.1**).

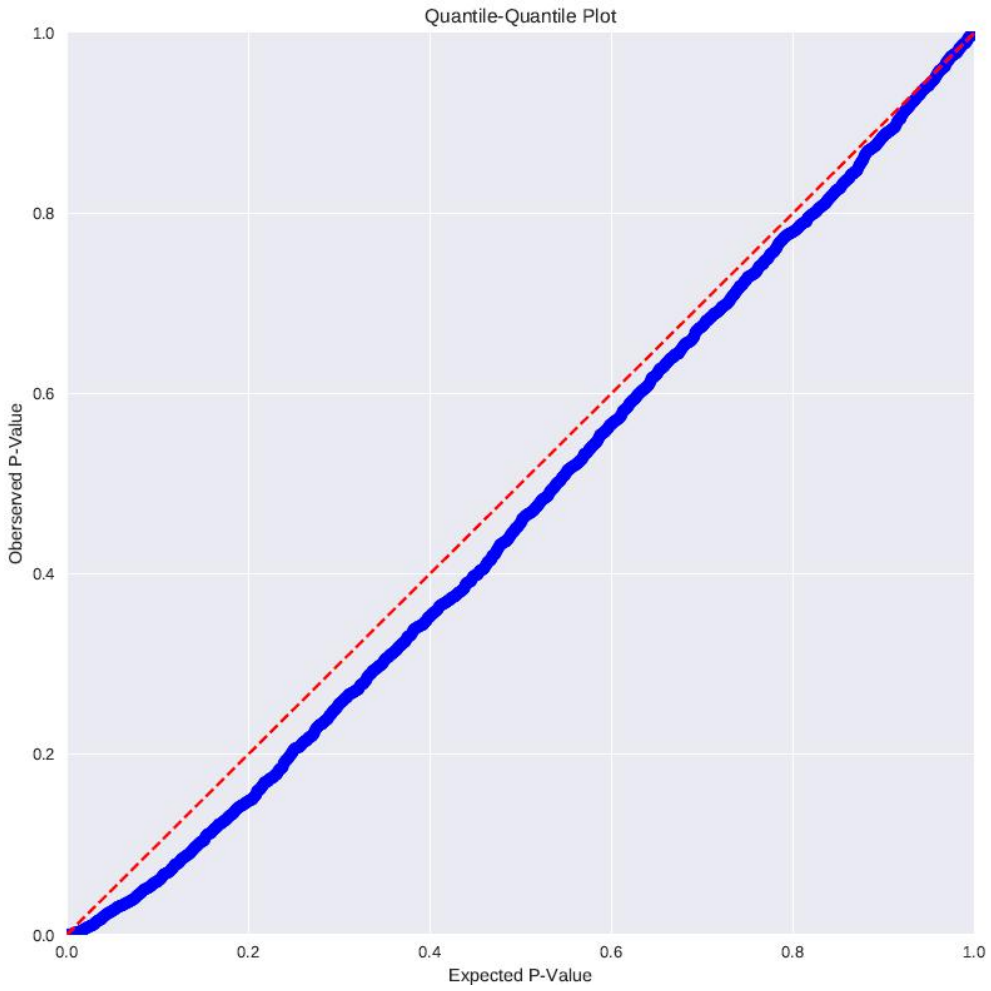


Figure 3.4 | Quantile-quantile plot of p-values of exon skip events.

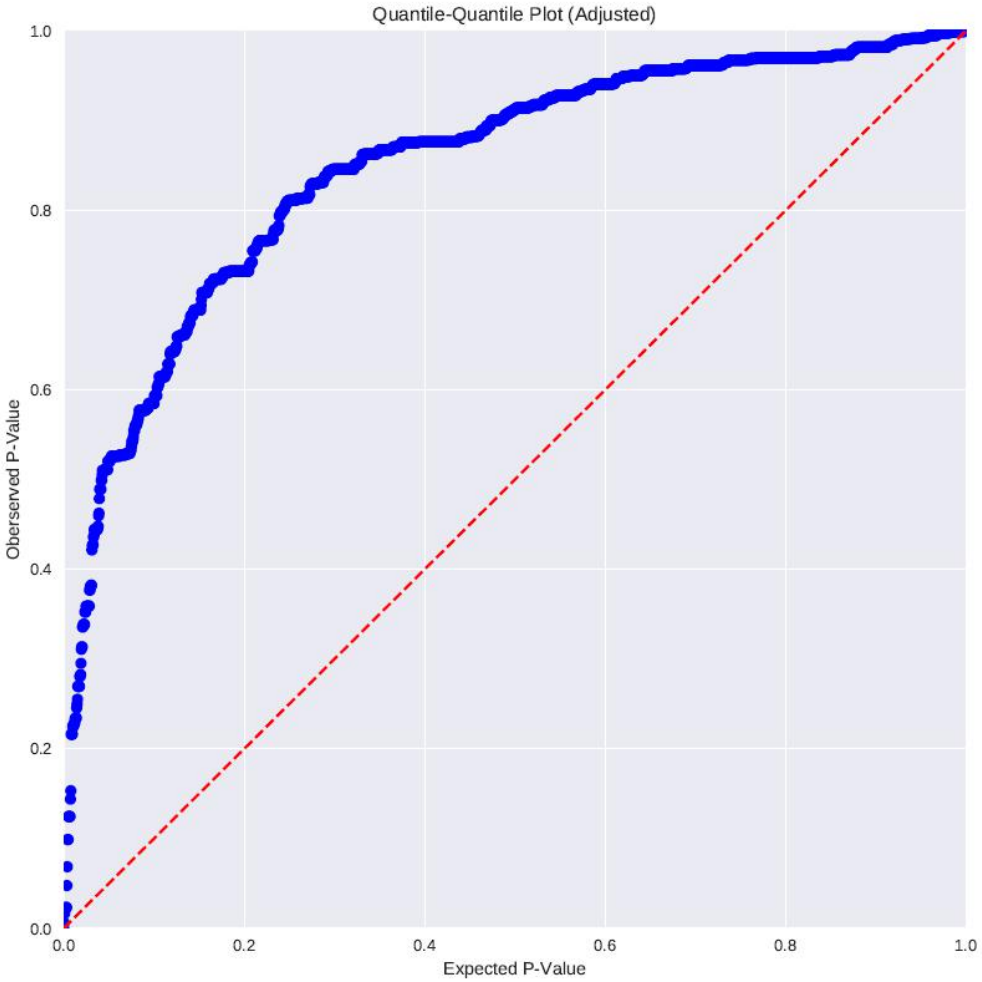


Figure 3.5 | Quantile-quantile plot of adjusted pvalues of exon skip events.

3.2.1. Intron retention events

In the first step, statistically significant intron retention events were extracted (**Script 1.2**). A new file was created, containing only retained introns' coordinates in gff3 format (**Script 1.3**). Names of the spliced genes were utilized for sequence obtainment, using UCSC Table Browser, which is an online program that provides access to genome sequence data [58]. To retrieve mRNA sequences for particular genes, several parameters have to be selected **(Table 6)**. The output file was in "fasta" format and contained all transcripts of the given genes, consisting of exons and introns. To remove all intron sequences except from retained introns, **Script 1.4** was implemented. Since a particular intron was not found to all transcripts of a gene, transcripts that did not comprise retained introns, were excluded from the set of the

examined transcripts. Introduction of premature termination codon and frameshift was checked via **Script 1.5**.

3.2.2. Exon skip events

In the case of exon skip events, statistically significant events were extracted too (**Script 1.6**). Although a similar procedure of retained intron detection could be followed again, adding some alterations related to skipped exon removal, it was preferred to find out which particular transcripts of each gene were affected and retrieve their sequences straightforwardly. Therefore, a new file was constructed, containing only skipped exons' coordinates in gff3 format and subsequently it was merged with the annotation file in order to find out which transcripts consist of the skipped exons (**Script 1.7**). The transcript names were used for sequence retrieval via UCSC Table Browser in accordance with the following parameters (**Table 5**).

Next, mRNA sequence was modified and skipped exons were removed from it (**Script 1.8**). To examine the introduction of premature termination codon that could lead either to non-sense mediated decay or termination of translation, **Script 1.9** was implemented, recognising stop codon before an exon-exon junction and computing the distance between them. Stop codons positioned in distance of at least 50 bases from downstream exon-exon junctions could elicit nonsense mediated decay [31], while stop codons positioned no less than 50 bases could terminate translation. Exon skip events causing frameshift were also detected via the same script.

Biomart was used to extract information about transcripts subjected to NMD or termination of translation, like gene symbol and transcript type **(Table 6)**.

Table 5 | Options selected in Table Browser to retrieve sequence of transcripts with retained introns and skipped exons

Categories	**Selected Parameters**	
	Intron retention event	**Exon skip event**
Clade	Mammal	Mammal
Genome	Human	Human
Assembly	Feb. 2009	Feb. 2009

	(GRCh37/hg19)	(GRCh37/hg19)
Group	Genes and Gene Predictions	Genes and Gene Predictions
Track	GENCODE Genes V19	GENCODE Genes V19
Table	Comprehensive (wgEncodeGencodeCompV19)	Comprehensive (wgEncodeGencodeCompV19)
Region	Genome	Genome
Identifiers (names/accessions)	List of gene names	List of transcript names
Output Format	Sequence	Sequence
Sequence Retrieval Region Options		
CDS exons	☑	☑
Introns	☑	
One FASTA record per region	☑	☑
Sequence Formatting Options		
Exons in upper case, everything else in lower case	☑	☑

Table 6 | Query for retrieval transcripts' information.

Dataset	Human Genes		
Filters	GENE	Input external references ID list	Transcript stable ID
Attributes	1. Gene stable ID 2. Transcript stable ID 3. Gene name 4. Transcript type		

3.2.3. Alternative 3' or 5' splice site events.

Similarly, statistically significant events of alternative splice sites were extracted (**Script 1.10**). For each type of event (3' and 5' ss) a new file was generated in gff3 format, containing both isoforms of each event (**Script 1.11**), while it was not known which was the alternative. Exons of the two isoforms of each event were merged with modified annotation file. For each event, if both isoforms were annotated, this particular event was thought to be false positive, as it was possible that the exon of the alternative isoform was a part of another transcript. If one of the two isoforms was annotated, this isoform was considered as the constitutive one.

3.2.4. Functional analysis

Finally, functional profiling of alternatively spliced genes was performed via gProfiler and ClusterProfiler tool (**Script 1.12**). Both tools implement methods for over-representation analysis, considering sources like Gene Ontology terms, biological pathways, regulatory motifs of transcription factors and microRNAs, human disease annotations and protein-protein interactions. Specifically, after the extraction of statistically significant events and their classification based on their potential effect (NMD or termination of translation), we used the affected genes for the enrichment analysis.

4. RESULTS

4.1. Identification of alternative spliced genes and transcripts

Firstly, statistically significant splicing events were found using adjusted p-value as statistical confidence measure, with a=0.05 as the cutoff. The total number of events was 47, but the affected genes were 35 (**Table 7**).

Table 7 | Number of events with adjusted p value less than 0.05.

Event Type	All	Unique genes
Alternative 3prime	1	1
Alternative 5prime	9	8
Exon skip	9	8
Intron retention	28	18
Multiple exon skip	0	0
Mutually exclusive exons	0	0
		35

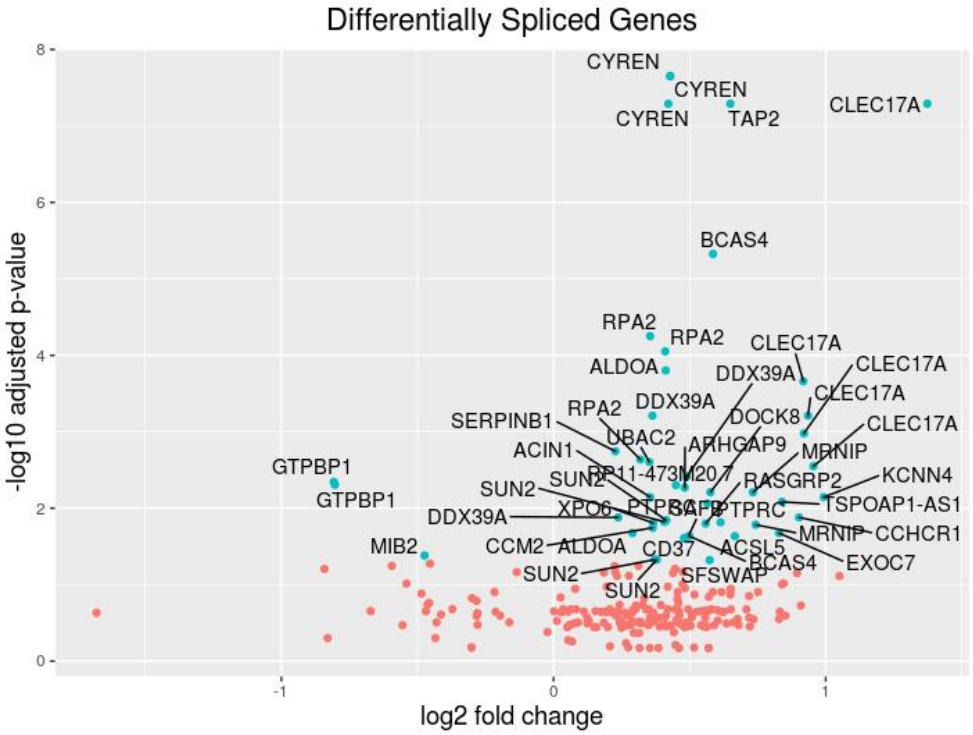


Figure 4.1 | Volcano Plot of Differentially Spliced Genes with 0.05 as cutoff for adjusted p value. Log2 Fold Change is expressed as the base-2-logarithm of the ratio of mean count of healthy samples to mean count of SLE samples.

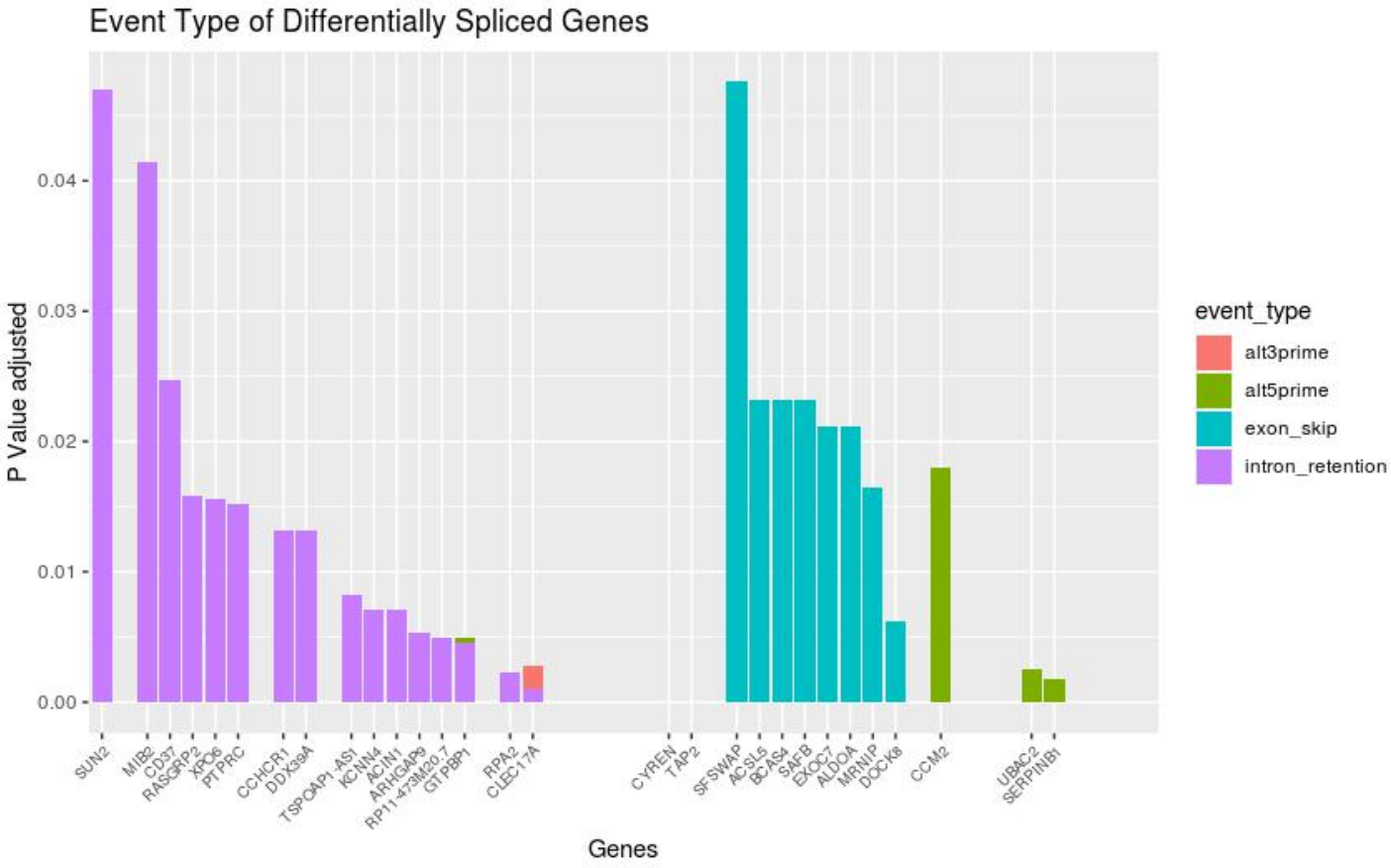


Figure 4.2 | Event type of Differentially Spliced Genes with 0.05 as a cutoff for adjusted p value. Some genes may be affected by multiple event types.

For the following exploratory analysis of the splicing events, the statistical significance measure was p-value, with an alpha of 0.01 as the cutoff. The total number of events was 236, but the affected genes were 200 (**Table 8**). The most frequent event type was intron retention (**Figure 4.3**).

Table 8 | Events with p value less than 0.01.

Event Type	All	Unique genes
Alternative 3prime	31	29
Alternative 5prime	29	27
Exon skip	61	56
Intron retention	106	80
Multiple exon skip	9	8
Mutually exclusive exons	0	0

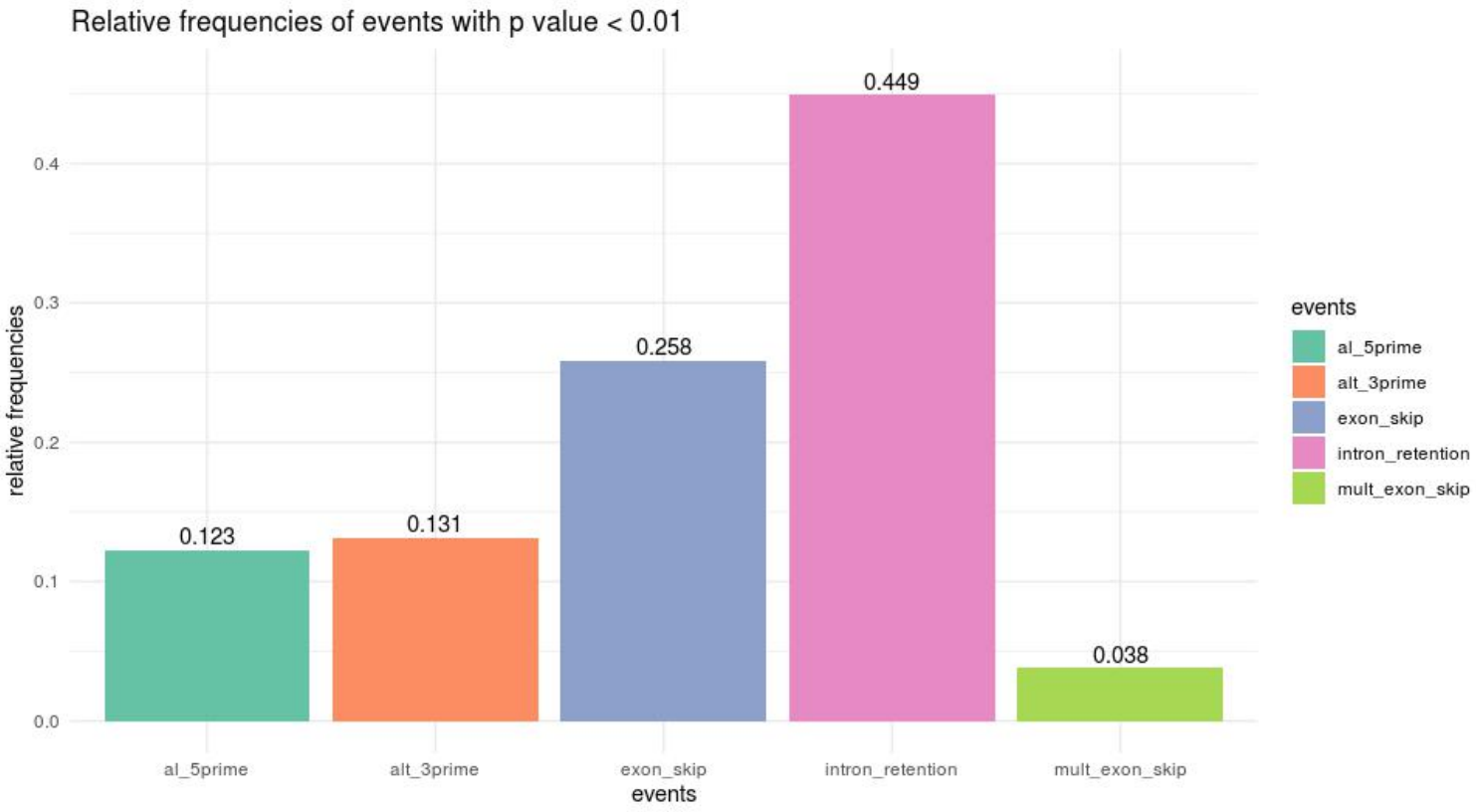


Figure 4.3 | Frequency of statistically significant alternative splicing events.

The median length of retained introns was 383bp and the median length of skipped exons was 111 bp. The largest retained intron was 4020 bp long, while the largest skipped exons were found in the proline-rich coiled-coil 2B *(PRRC2B)* gene with a length of 2082 bp and the RNA-binding motif protein (*RBM6*) gene with a length of 1279bp. The splicing graphs of these exon skip events are depicted in **Figures 4.6-7**. Both events occurred more frequently in the healthy than SLE samples. The third track of each figure represents the mean coverage of each group of samples (healthy and SLE). Although there is not great difference in the context of the coverage between the two groups, the large sample size provides statistical evidence.

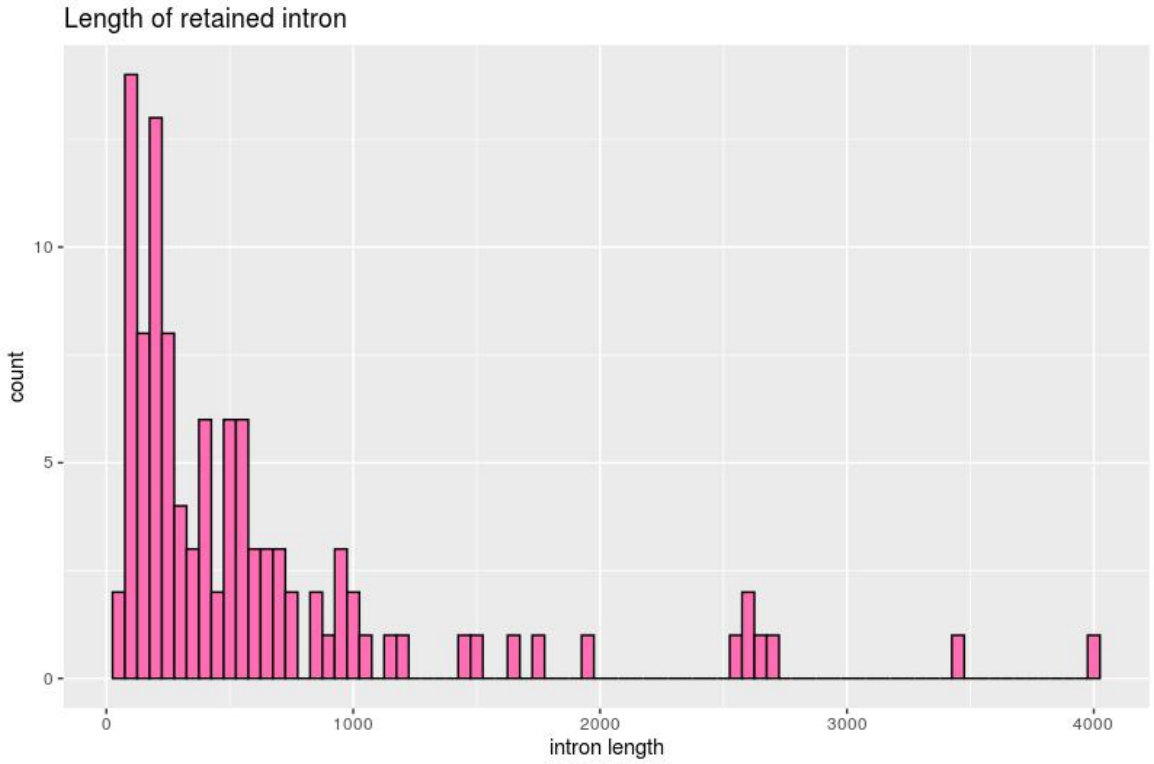


Figure 4.4 | Length of retained introns of statistically significant intron retentions.

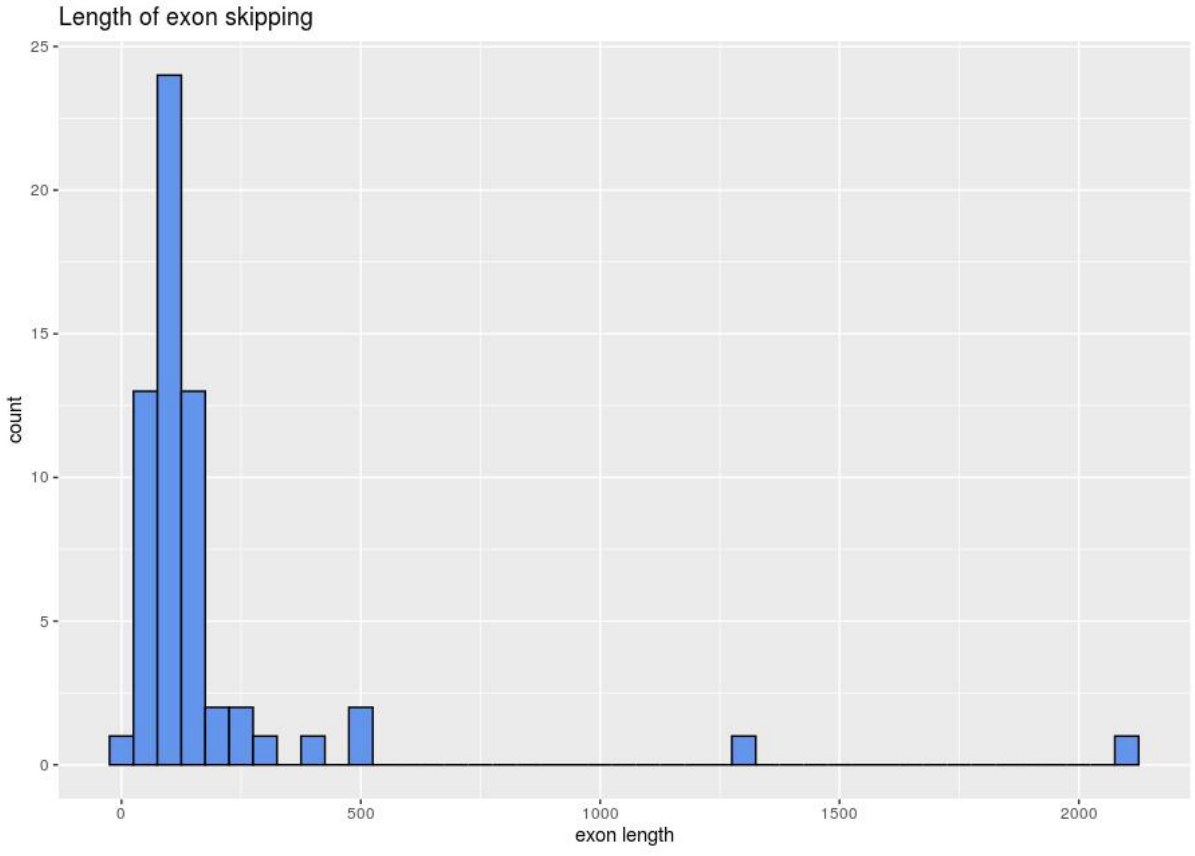


Figure 4.5 | Length of skipped exons of statistically significant exon skips.

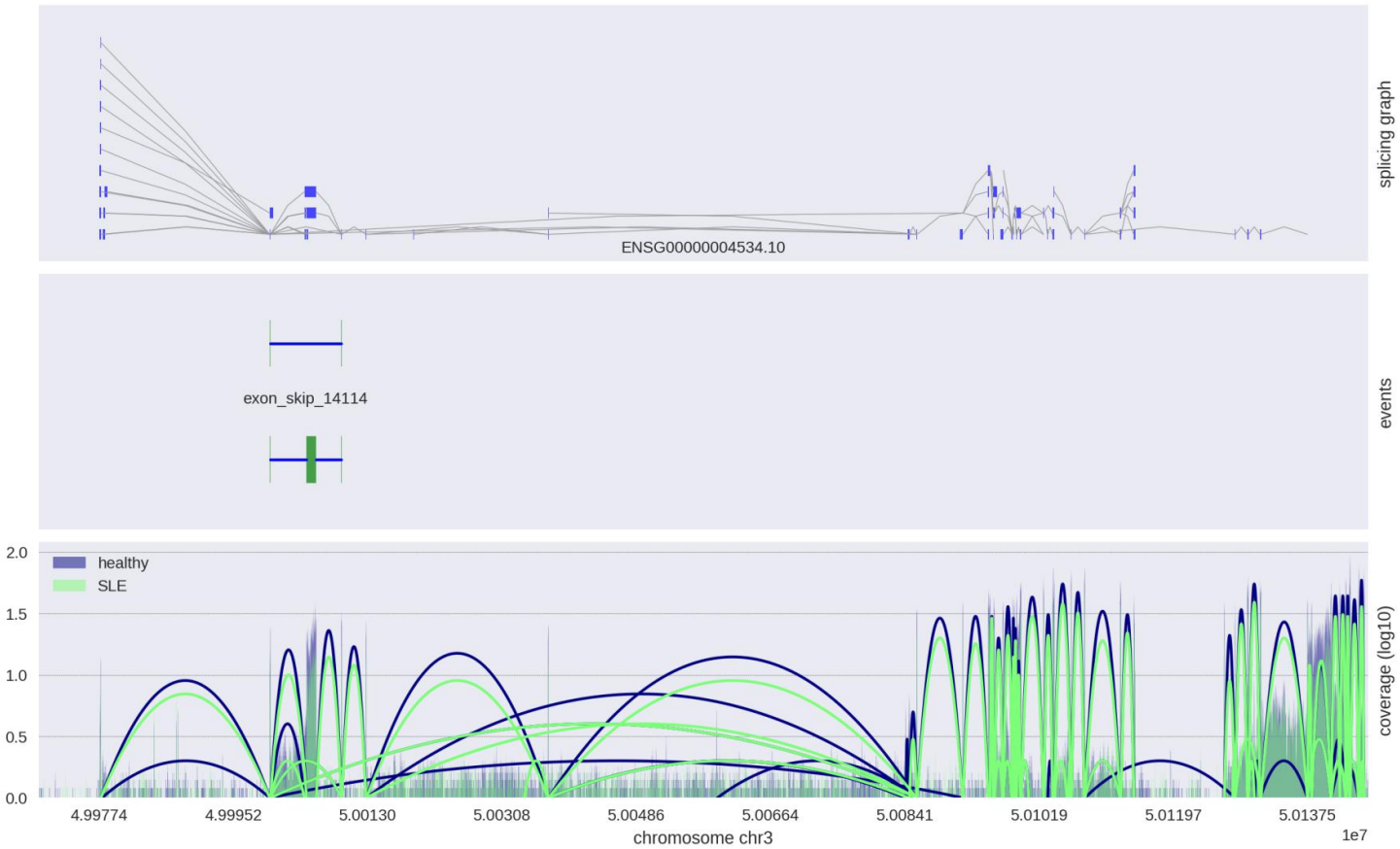


Figure 4.6 | Splicing graph of RBM6 gene. The skipped exon is the third exon of the ENST00000266022 transcript and is located in chromosome 3 in the region between 50,004,853 and 50,006,231 bp. The coverage of SLE samples is depicted in green color whereas the coverage of healthy samples is depicted in blue color.

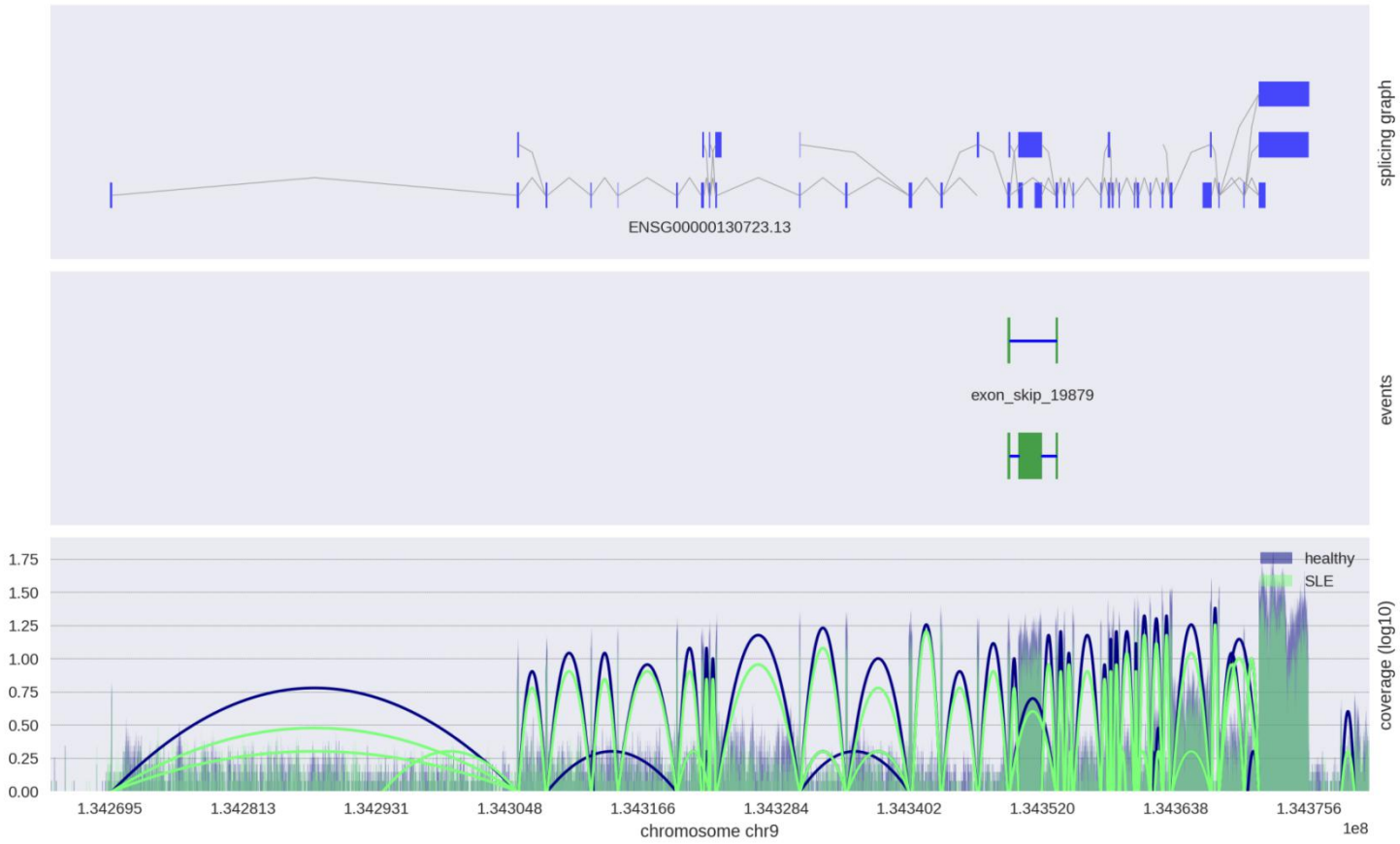


Figure 4.7 | Splicing graph of PRRC2B gene. The skipped exon is the fifteenth of ENST00000357304 transcript and the ninth of the ENST00000372249 transcript and is located on chromosome 9 in the region between 134,349,841 and 134,351,922 bp. The coverage of SLE samples is depicted in green color whereas the coverage of healthy samples is depicted in blue color.

4.2. Detection of premature termination codons

- Statistically significant intron retentions

Subsequently, the possible introduction of a premature termination codon was checked for genes affected by intron retentions and exon skips. In intron retention cases, it was found that the 59 spliced genes correspond to 227 transcripts. Introduction of a premature termination codon occurred in 55 genes and potentially 216 transcripts **(Figure 4.8)**. The TAP2 gene was the top statistically significant alternatively spliced gene, where a specific intron retention occurred more frequently in the healthy samples than SLE **(Figure 4.9)**. The next significant gene was CLEC17A, whose intron was found retained more often in the final transcript of the healthy than SLE samples **(Figure 4.10)**. The SCO2 gene was identified as alternatively spliced in the study of Panousis et al. and its splicing graph is shown in **Figure 4.11,** where

intron retention occured more frequently in the SLE samples. In the occasion of genes where no stop codon formed, it was found that no frameshift was introduced too **(Figure 4.12)**.

Table 9 | Detection of premature termination codons introduced in the alternative spliced transcripts.

Event Type	Affected Annotated genes	Probably affected transcripts	Stop Codons of Affected annotated genes	Stop Codons of Probably affected transcripts
Exon skip	56	216	26	60
Intron retention	59	227	55	216

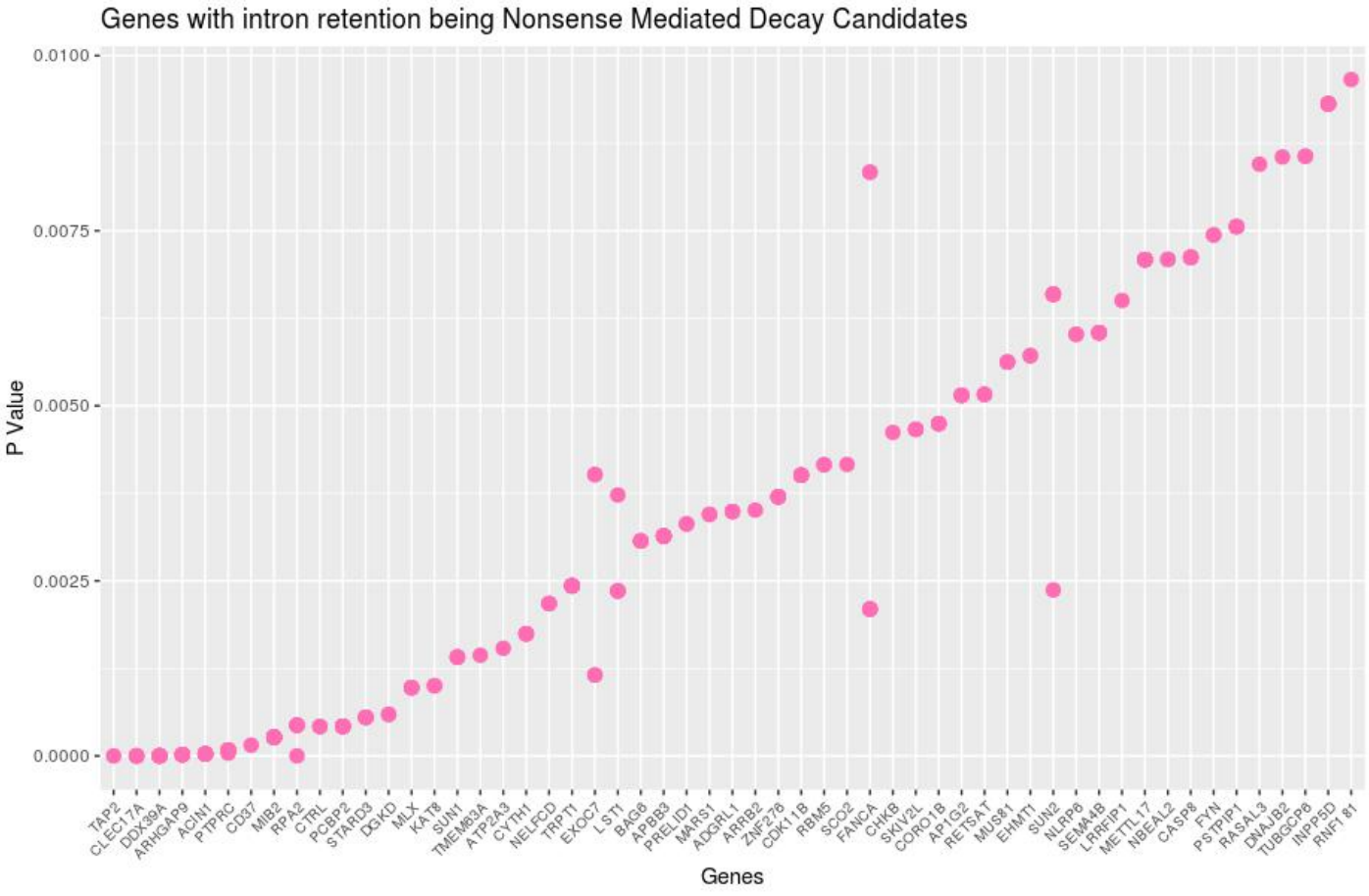


Figure 4.8 | Alternative spliced genes with intron retention leading to nonsense mediated decay.

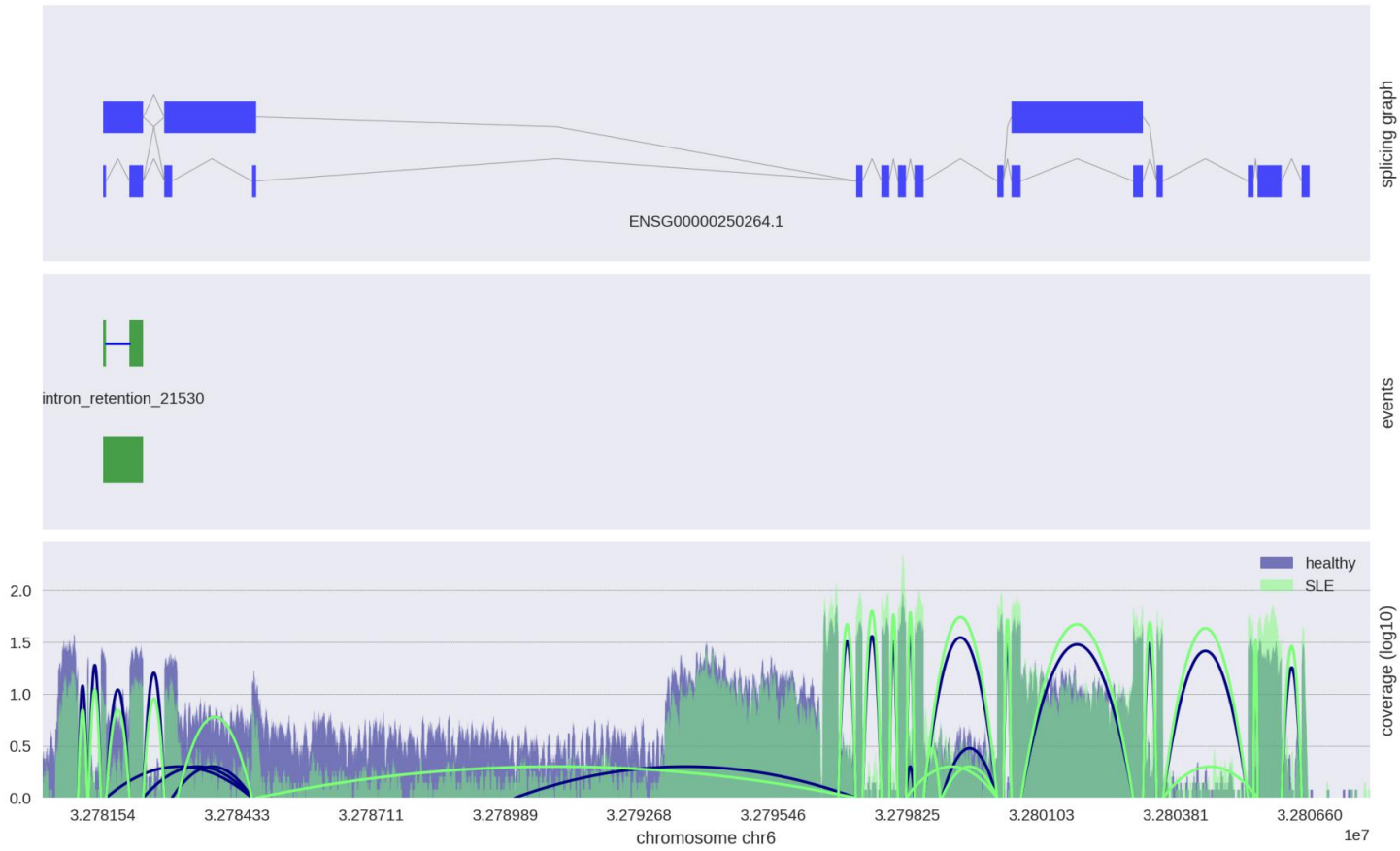


Figure 4.9 | Splicing graph of TAP2 gene. The coverage of SLE samples is depicted in green color whereas the coverage of healthy samples is depicted in blue color.

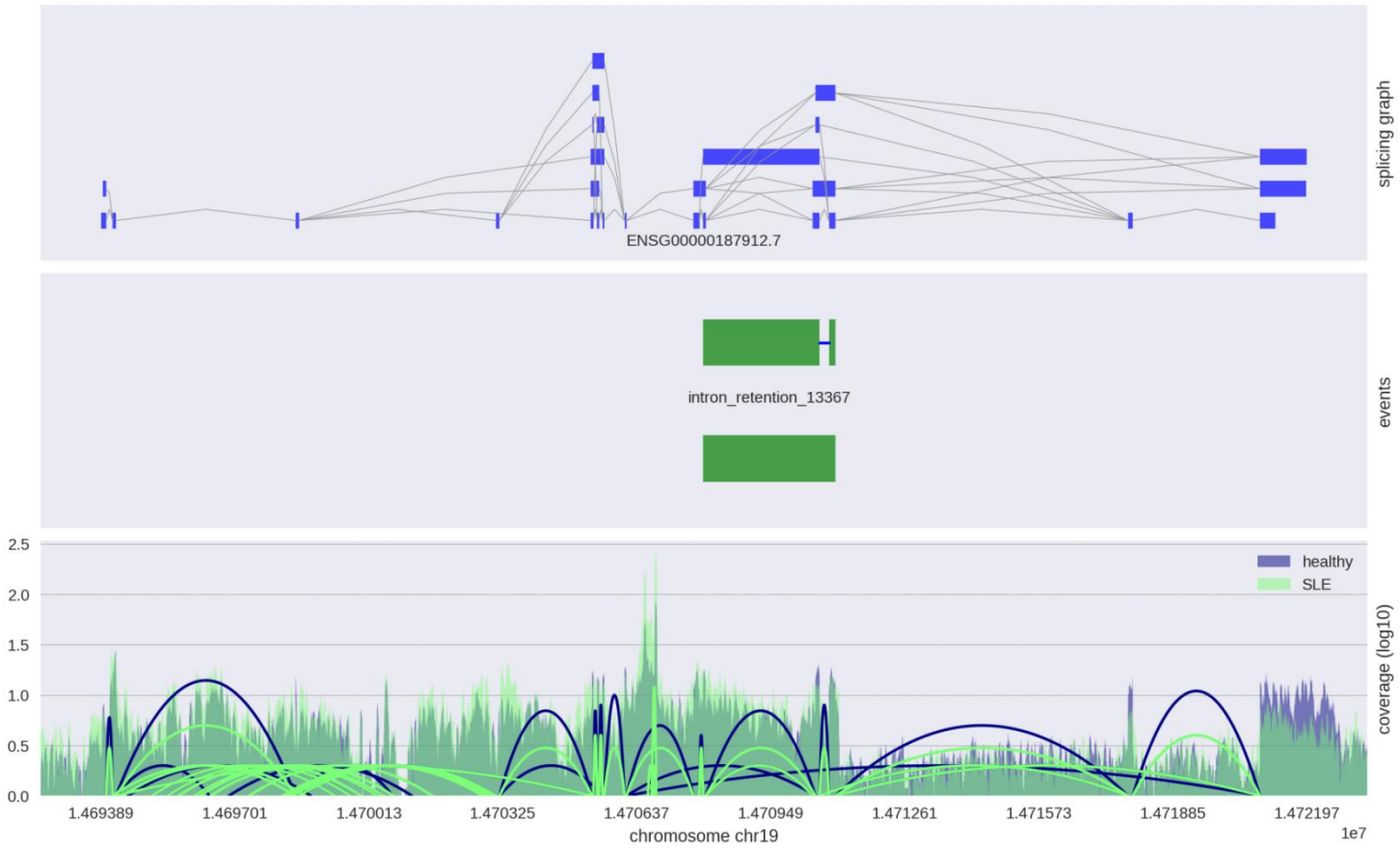


Figure 4.10 | Splicing graph of CLEC17A gene. The retained intron is downstream from the eleventh exon of the ENST00000547437 transcript and is located in chromosome 19 in the region between 14,710,625 and 14,710,842 bp. The coverage of SLE samples is depicted in green color whereas the coverage of healthy samples is depicted in blue color.

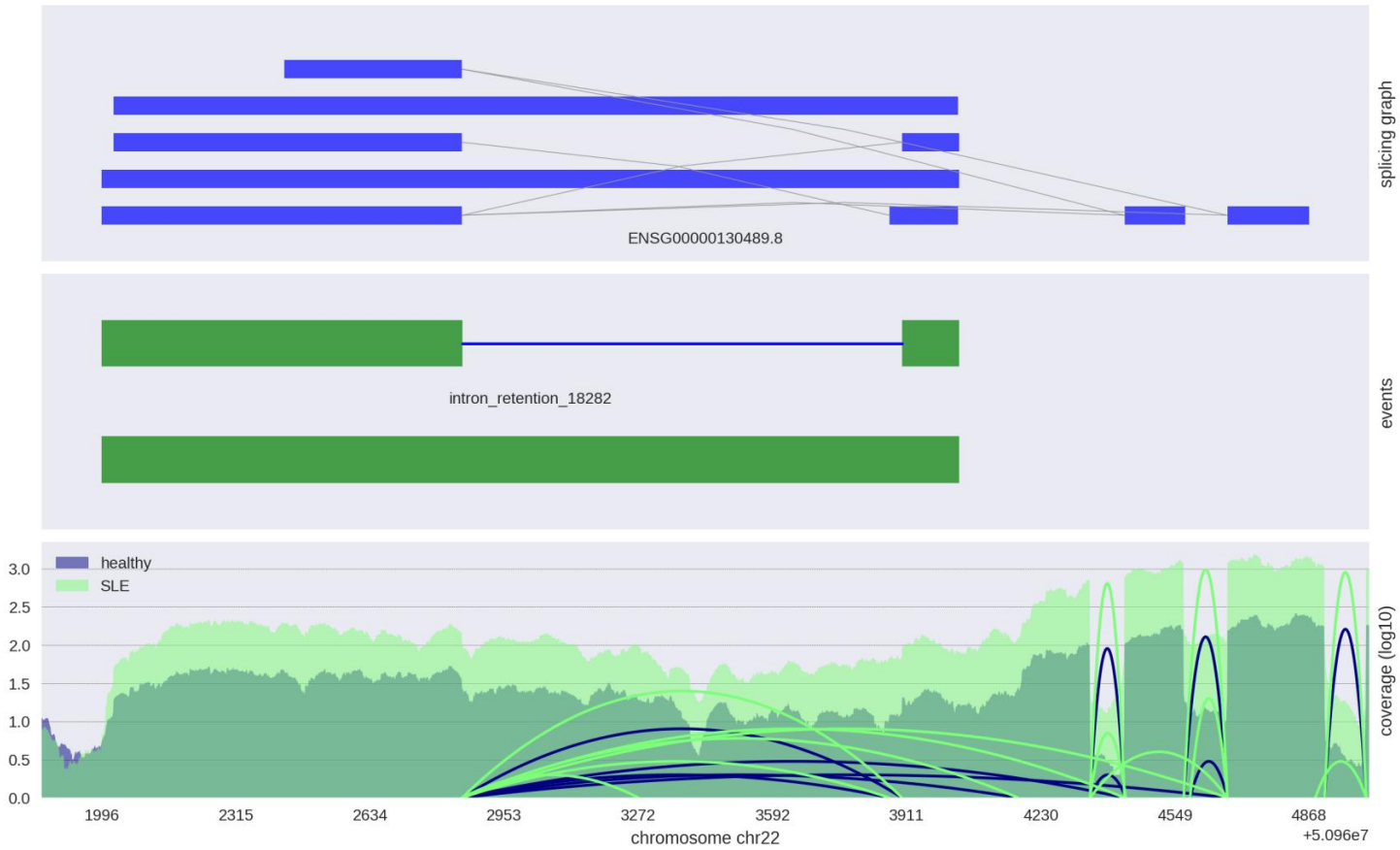


Figure 4.11 | Splicing graph of SCO2 gene. The retained intron is downstream from the first exon of the transcript and is located in chromosome 22. The coverage of SLE samples is depicted in green color whereas the coverage of healthy samples is depicted in blue color.

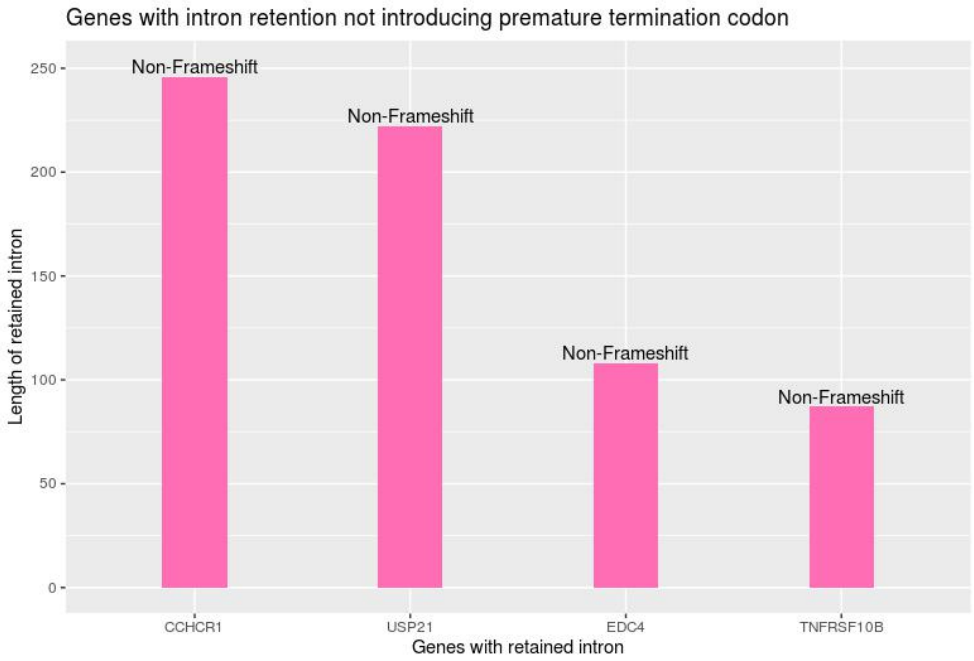


Figure 4.12 | Alternative spliced genes without formation of stop codon, after intron retention.

- Statistically significant exon skipping events

Genes affected by exon skips were 56, corresponding to 216 transcripts. Introduction of premature termination codon occurred in 22 genes and potentially 60 transcripts. In transcripts of 21 genes could be caused nonsense-mediated mRNA decay (NMD) that prevents translation (**Figure 4.13**) and in transcripts of 3 genes, the *LST1, RBM5* and *MRNIP*, a termination codon was formed probably leading to termination of translation. The *LST1* gene and *MRNIP* gene produced transcripts targeted for NMD as well as transcripts with stop codon. Transcripts of 34 genes did not contain premature termination codon after exon skip. Exon skips of 14 of them cause frameshift (**Figure 4.14**).

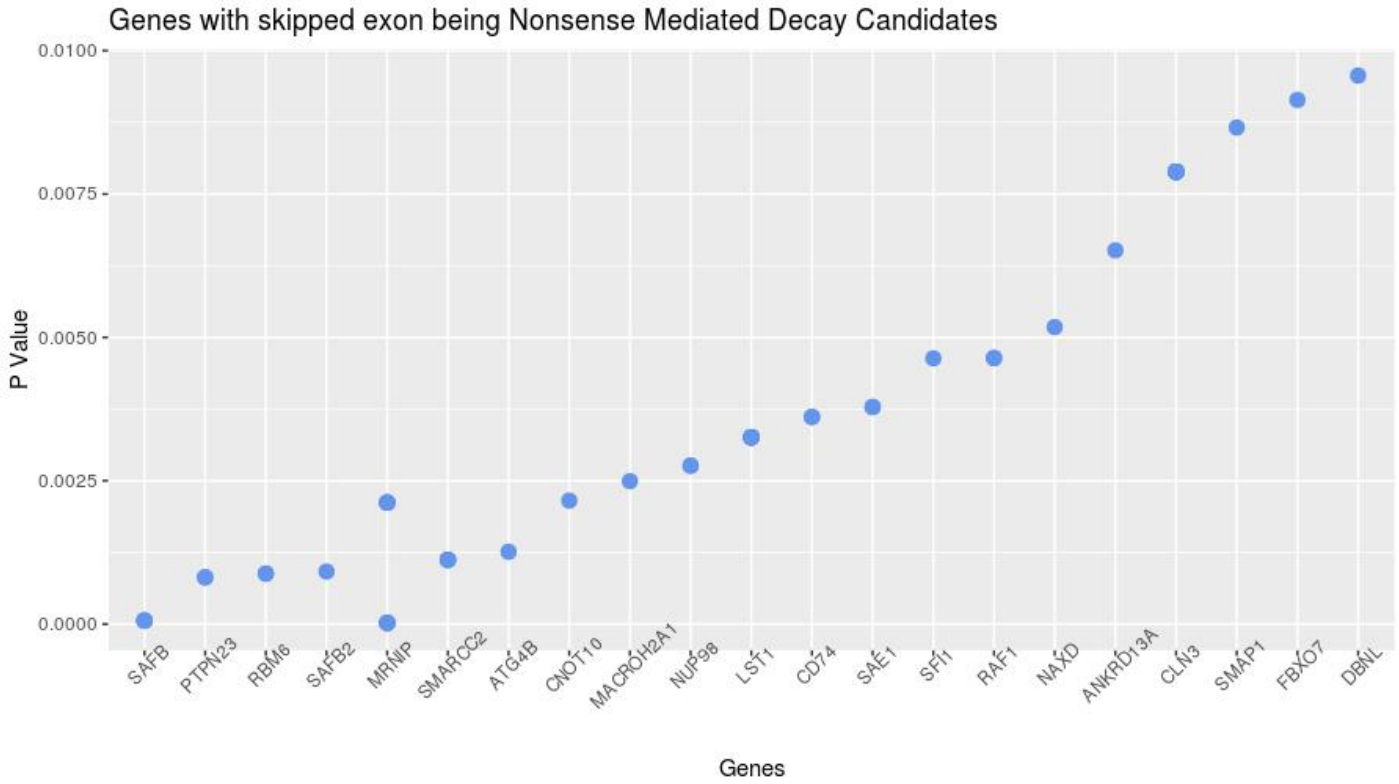


Figure 4.13 | Alternative spliced genes with exon skip leading to nonsense mediated decay.

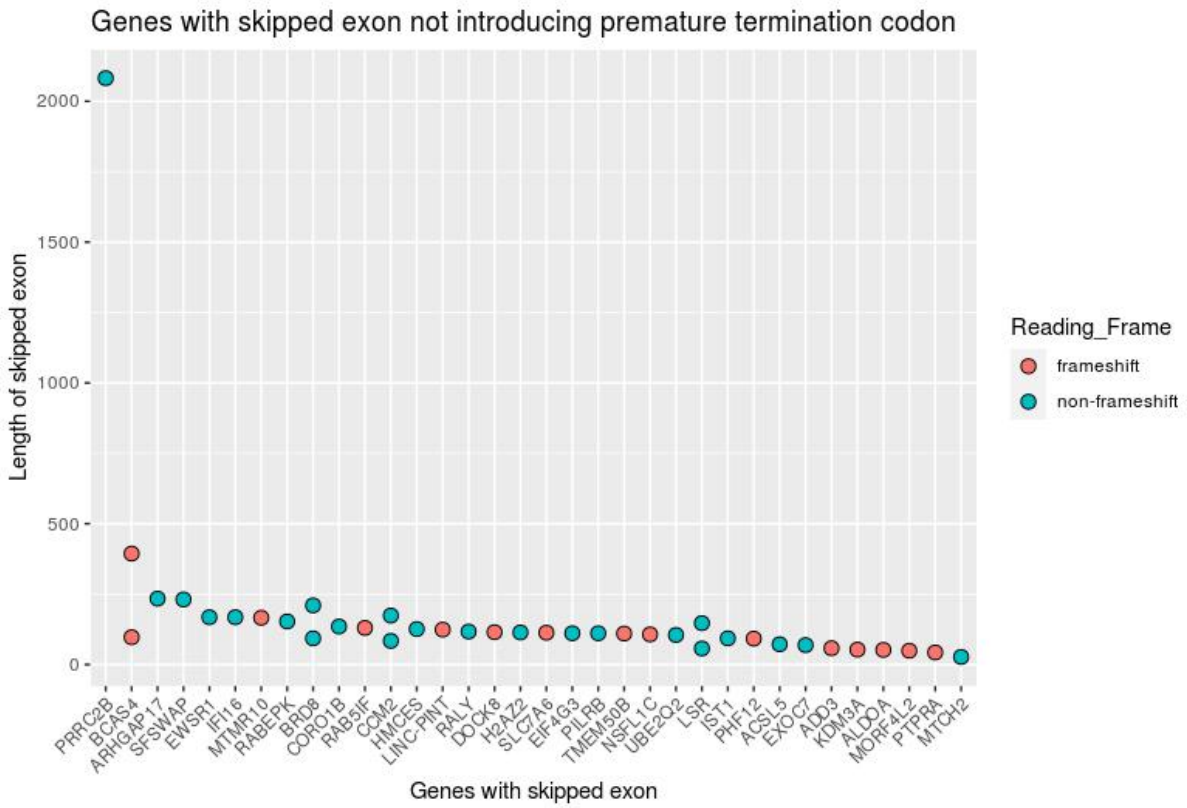


Figure 4.14 | Alternative spliced genes with exon skip where no PTC is formed.

4.3. Discrimination of annotated from non-annotated statistically significant alternative 3’ or 5’ splice site events

Alternative 3’ or 5’ splice site events, whose isoform 1 was annotated and isoform 2 was not, were characterised as true positive events. Affected genes are shown in **Figure 4.15-16**.

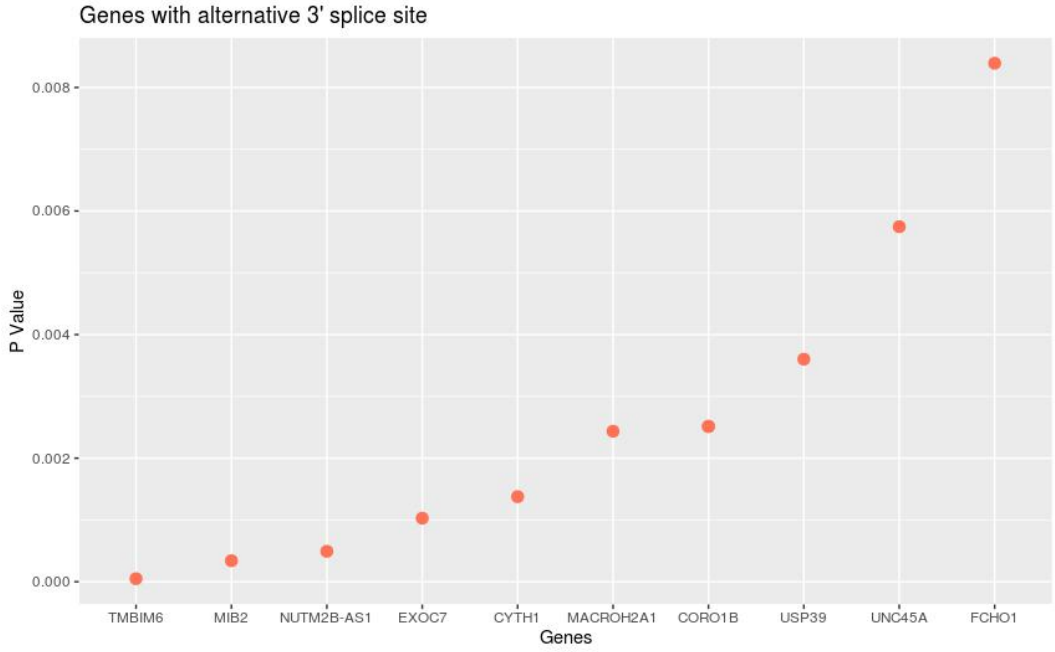


Figure 4.15 | Genes with alternative 3’ splice site, characterised as true positive events.

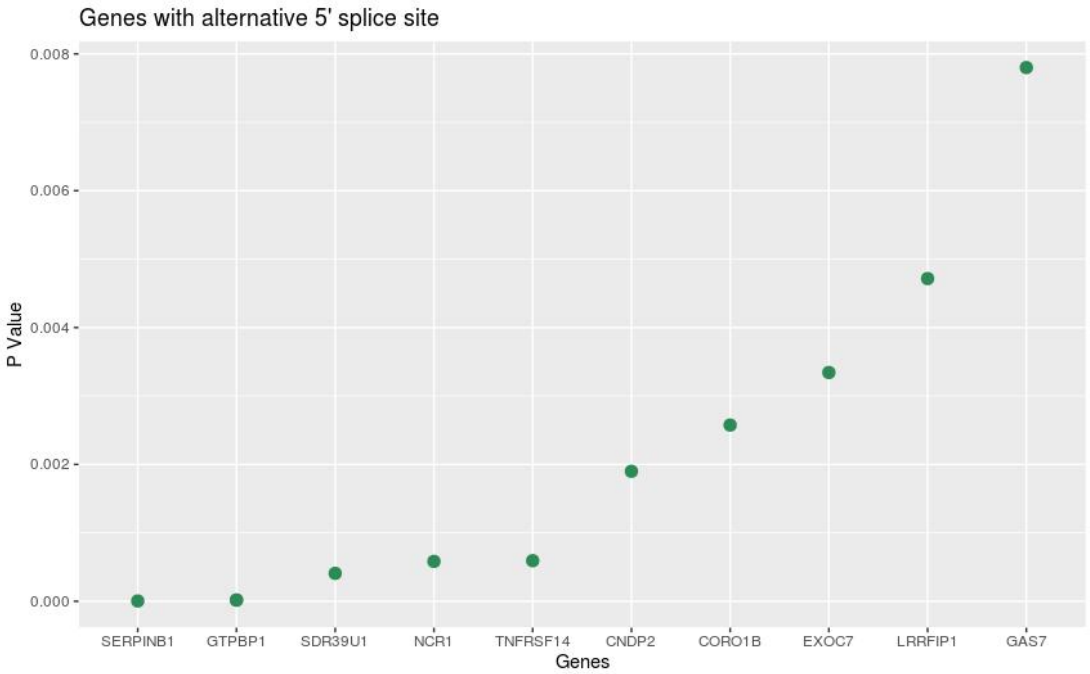


Figure 4.16 | Genes with alternative 5’ splice site, characterised as true positive events.

4.4. Functions of affected genes

Splicing events, whose statistical significance measure was p-value, with an alpha of 0.01 as the cutoff, were used for functional analysis. Overrepresentation analysis via gProfiler tool indicated a correlation between these genes and transcription factors Erm, ERG and Fli-1 (**Figure 4.17**). Also, enrichment analysis showed enriched biological processes like ubiquitin dependent protein binding and molecular functions like regulation of mRNA processing **(Figure 4.18)**.

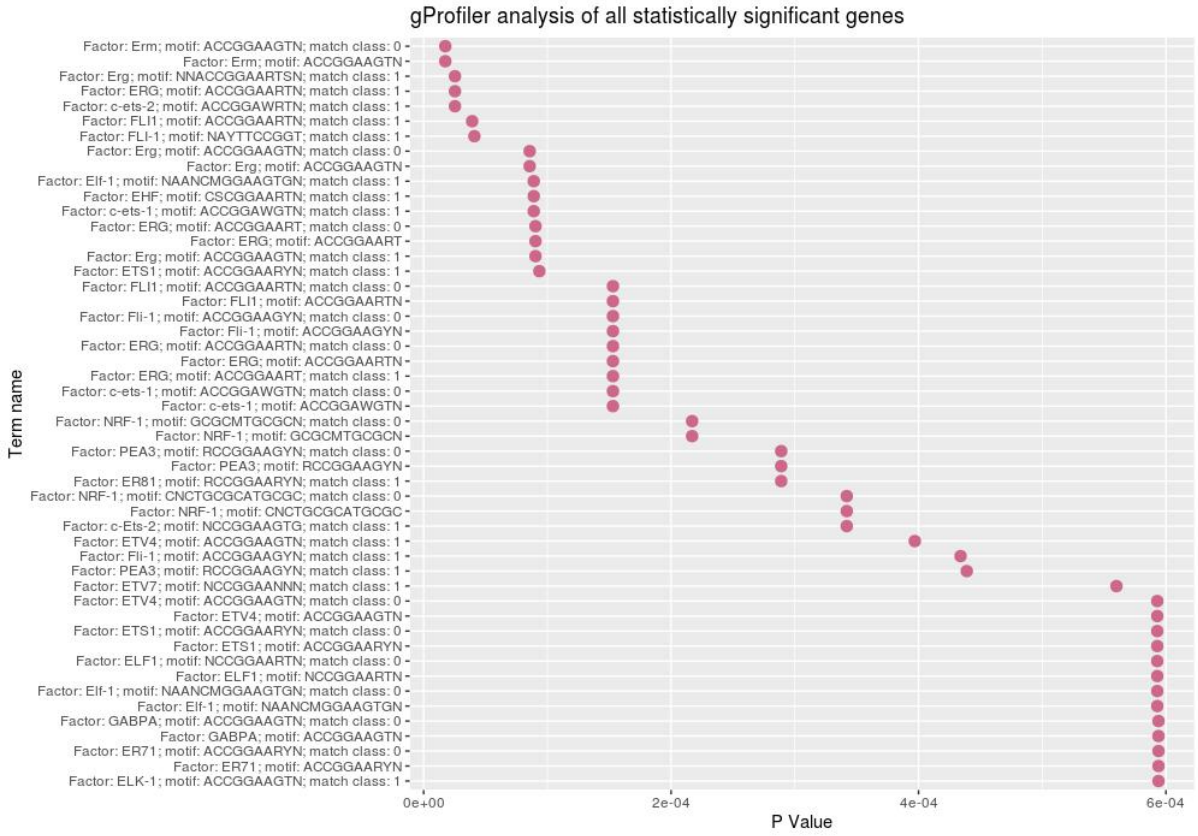


Figure 4.17 | Overrepresentation analysis of statistically signifcant spliced genes.

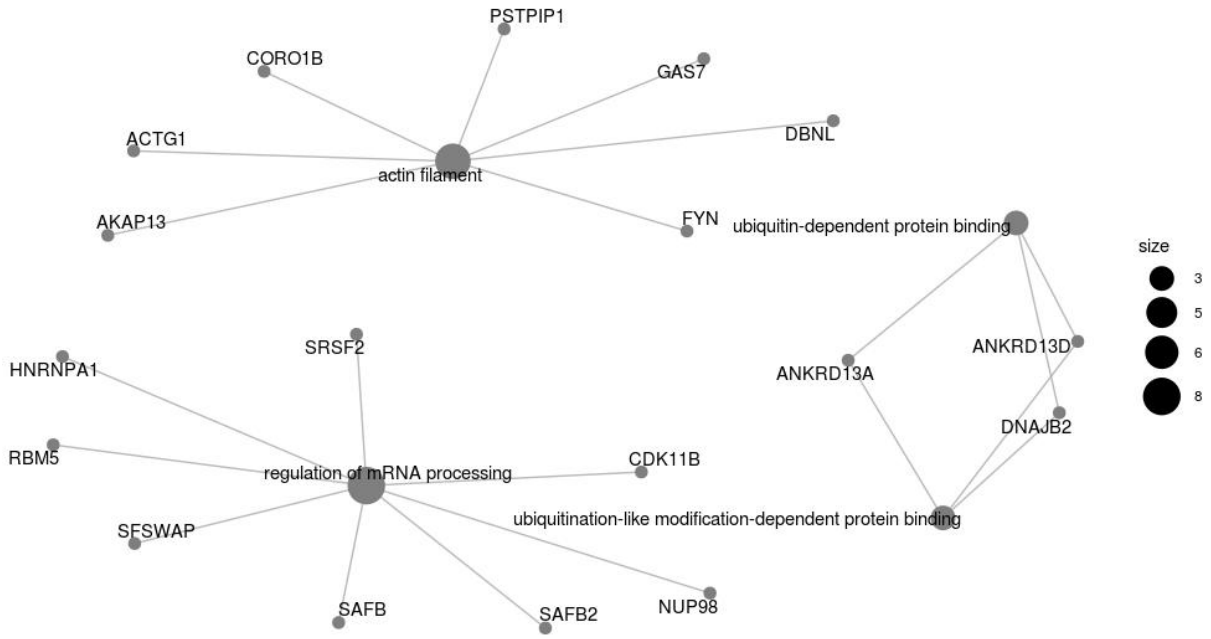


Figure 4.18 | Enrichment analysis of statistically significant spliced genes.

Genes with introduction of termination codon, that lead to either NMD or termination of translation, were also analysed for functional profiling. Binding in several transcription factors was found again, as well as molecular functions related to immune cell differentiation **(Figure 4.19-20)**.

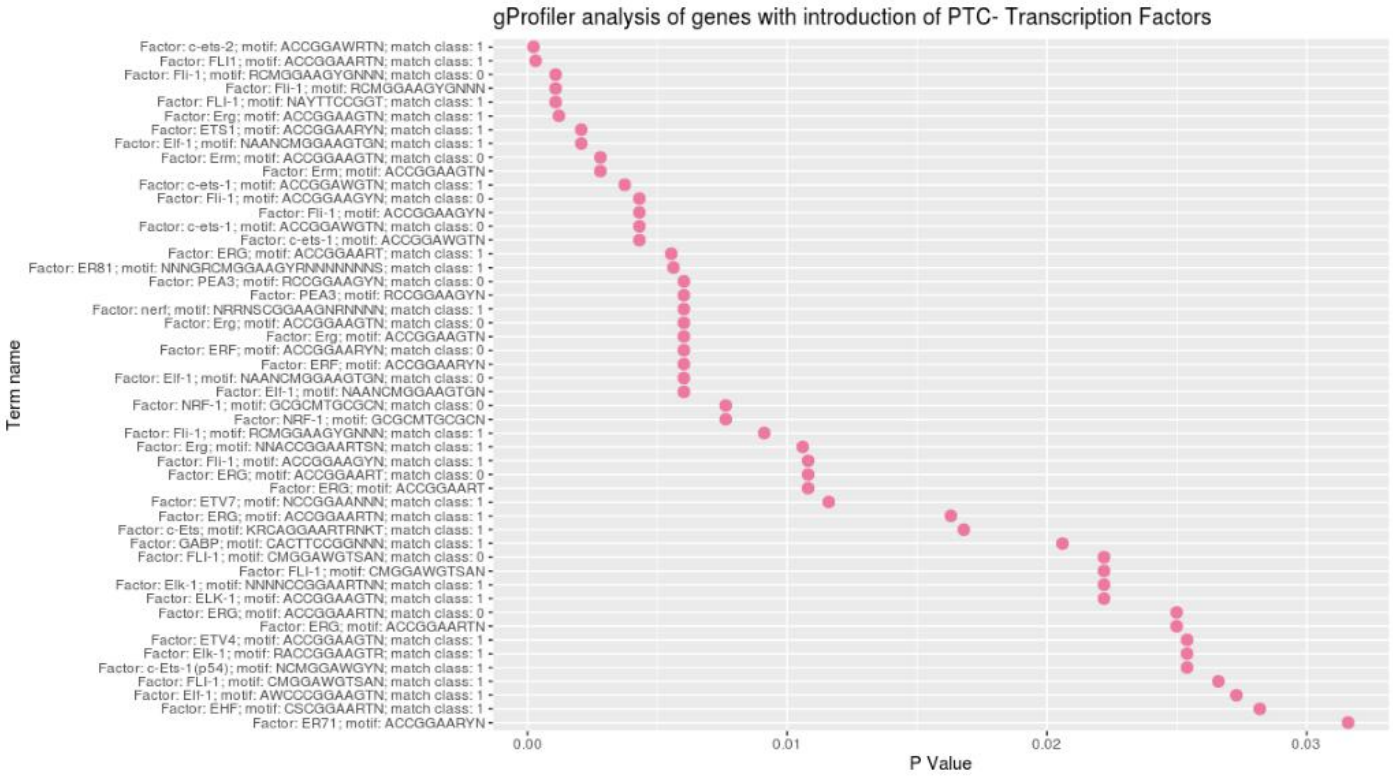


Figure 4.19 | Overrepresentation of transcription factor families in statistically significant spliced genes with introduction of premature termination codon.

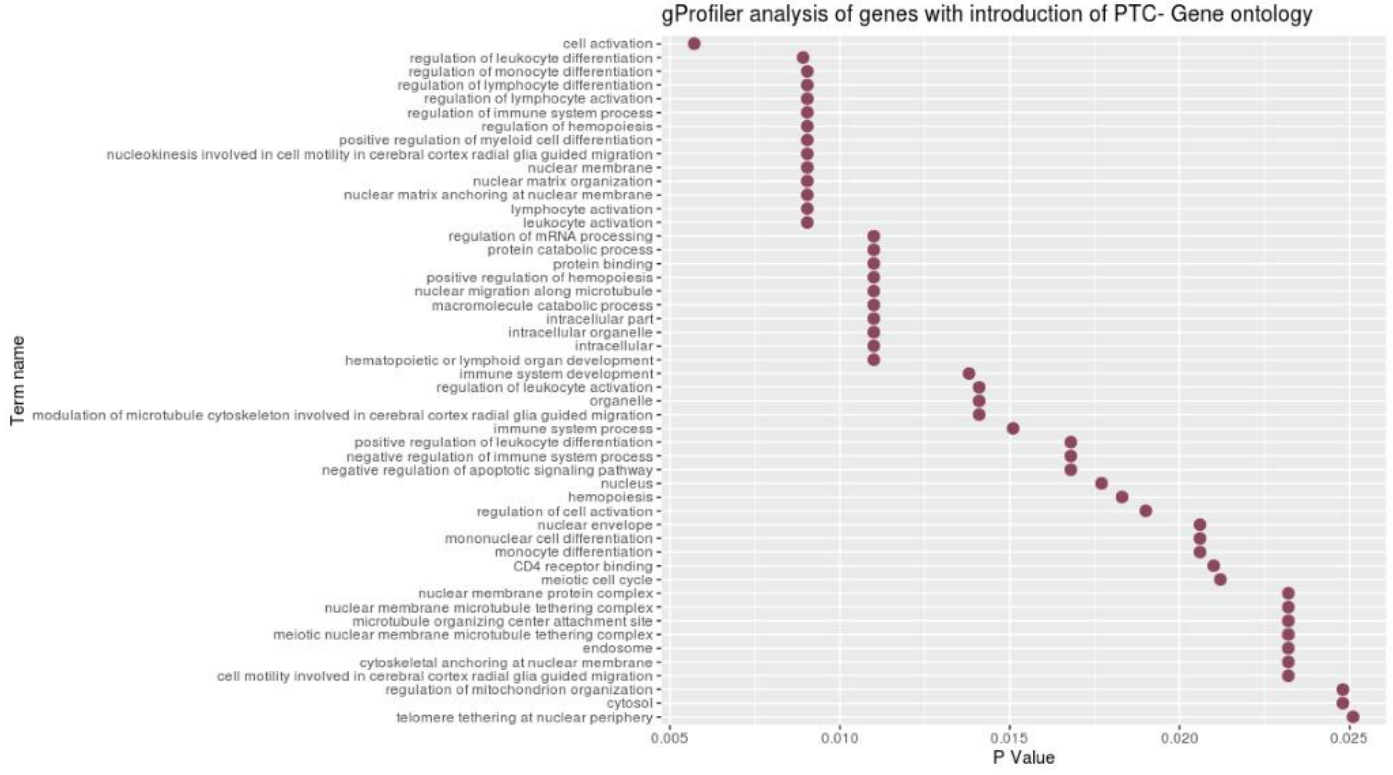


Figure 4.20 | Gene ontology overrepresentation analysis of statistically significant spliced genes with introduction of premature termination codon.

5. DISCUSSION

Alternative splicing is a major regulator of gene expression and several findings support a connection between splicing and SLE. In this project we attempted to discover splicing events of a large dataset, consisting of 200 volunteers, through the analysis of RNA sequencing data. Although the large sample size of the dataset increases statistical power, the limitation has arisen from the fact that the samples contain a heterogeneous cell population, so we expect to retrieve a global signature of splicing regulation that may be more tissue specific than cell-type specific. This is a very important feature of our work that warrants the cautious interpretation of our results. Alternative splicing events occurring in healthy individuals, or SLE patients are most likely reflecting the dynamics of the heterogeneous cell population. This dataset has already been used for a preliminary splicing analysis, but most of the genes detected as being alternatively spliced between the two conditions differ from the genes that are identified in our analysis. This difference could be derived from the fact that the method implemented in the previous study is focused on intron excisions [59],[49].

The statistically significant detected alternative splicing events included exon skipping, intron retention, usage of alternative 3' or 5' splice site and multiple exons skipping. The median length of retained introns was 383bp and the largest retained intron had 4020 bp length, belonging to the Cytohesin 1 (*CYTH1*) gene. This retention was found to introduce a premature termination codon that potentially elicits NMD in healthy samples in contrast to SLE. As *CYTH1* is thought to be involved in SLE [60] , this retention could be a normal process that reduces CYTH1 protein levels in healthy individuals, that is perturbed in SLE patients. Respectively, the median length of skipped exons was 111 bp and the largest skipped exons were found in the proline-rich coiled-coil 2B *(PRRC2B)* gene with length of 2082 bp and the RNA-binding motif protein (*RBM6*) gene with length of 1279bp. *PRRC2B'* skipped exon did not cause frameshift, so the produced protein may be lacking a significant part in some healthy samples.

In order to recover the sequence of splicing events and specify the PTC-containing isoforms we used annotation files from the UCSC Table Browser. Therefore, several splicing events generate non annotated exon junctions, as in the cases of intron retention, where an intron is not supported from the reference genome. These events' sequences cannot be retrieved.

Despite of the fact that these events are characterized as novel, they actually evolve from transcriptome complexity, which cannot be fully covered by the available annotation. Our analysis detected 22 non annotated intron retentions, so these were not used for sequence retrieval. 22 of the 56 genes containing a skipped exon and 55 of the 59 genes containing a retained intron were found to introduce premature termination codon. This could be due to the fact that introns insert a sequence in the final transcript, where the larger the sequence, the greater the possibility to include a premature termination codon.

As mentioned above, PTCs, located more than 50bp upstream of the last exon-exon junction, prompt NMD, mediating down regulation of the specific gene. In this study all of the intron retention events were classified as NMD candidates. The same was true for 21 of the 22 PTC-containing genes with skipped exon. Among **NMD candidate genes**, we identified *CD74,* the gene which encodes a chaperon molecule that mediates MHC-II antigen-presentation and has been linked to lupus pathology in mouse models [61] . Transporter 2, ATP Binding Cassette Subfamily B Member (*TAP2*) is also a risk factor for SLE [62] , while its splicing isoforms have been associated with different manifestations in Type 1 Diabetes [63] . Moreover, the C-type lectin domain containing 17A (*CLEC17A*) gene, which acts as a receptor in B-cells in germinal centers, has been included in the list of the top 20 genes with the largest number of epigenomic elements enriched in the promoter region, among SLE associated genes [64].

Regarding the genes **translated into truncated proteins** due to the premature termination codon, *LST1* gene expresses a membrane protein that impedes lymphocyte proliferation. It has been shown that decreased *LST1* expression levels and generation of a splice variant were correlated with arthritis severity in rats [65]. Additionally, it is suggested that almost 10% of the transcripts that are targets for NMD, eventually escape NMD and are translated into truncated proteins [66].

The **functional analysis** of the differential spliced genes revealed enrichment of two protein families. Firstly, the Erythroblast Transformation Specific (Ets) family, consisting of proteins like Erg, Fli and Ets, regulates cell proliferation, differentiation, migration and angiogenesis [67] and includes the highly conserved DNA binding domain ETS. Although it has been reviewed that the transcription factor Ets1 contributes to SLE development [68] , [69] , it remains unclear if there is association between splicing and Ets regulation. An abstract

published from O. Saulnier et al. in 2018 suggests interaction of the ERG transcription factor with several spliceosome components [70].

The Ezrin/Radixin/Moesin (Erm) protein family crosslink plasma membrane proteins with actin filaments and interacts with CD44 domain, promoting cell adhesion and migration [71]. The phosphorylated ERM causes increased adhesion and migration of T cells in SLE patients [72] . Nevertheless, there is currently no evidence supporting the connection between the splicing machinery and ERM.

Furthermore, enrichment analysis indicated enriched terms of mRNA regulation processes and ubiquitin-dependent protein binding, which were also enriched in DEGs pathway analysis (**Figure 1.6**).

In conclusion, the examination of differentially alternative spliced genes between SLE and healthy individuals could yield relevant insights into monitoring SLE and provide novel potential biomarkers. It would be interesting to continue with experimental validation of the key spliced isoforms through RT-qPCR and their protein products through western blot. The overrepresented transcription factors binding indicate a possible mechanism that induce splicing, but this mechanism remains elusive and further study is required. Also, as splicing regulatory proteins regulate splicing by binding to cis-acting splicing regulatory elements (SREs), it would be useful to recognise these sites and detect special patterns into them. Genotyping data could be used for mutation analysis to detect splicing-causing mutations affect the recognition of the splice sites or the SREs.

6. REFERENCES

[1] Q. Pan, O. Shai, L. J. Lee, B. J. Frey, and B. J. Blencowe, "Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing," *Nature Genetics*, vol. 40, no. 12, pp. 1413–1415, Dec. 2008, doi: 10.1038/ng.259.

[2] S. Djebali *et al.*, "Landscape of transcription in human cells," *Nature*, vol. 489, no. 7414, pp. 101–108, Sep. 2012, doi: 10.1038/nature11233.

[3] M. C. Wahl, C. L. Will, and R. Lührmann, "The Spliceosome: Design Principles of a Dynamic RNP Machine," *Cell*, vol. 136, no. 4. pp. 701–718, Feb. 20, 2009, doi: 10.1016/j.cell.2009.02.009.

[4] R. Breathnach, C. Benoist, K. O'Hare, F. Gannon, and P. Chambon, "Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 75, no. 10, pp. 4853–4857, 1978, doi: 10.1073/pnas.75.10.4853.

[5] J. A. Berglund, K. Chua, N. Abovich, R. Reed, and M. Rosbash, "The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC," *Cell*, vol. 89, no. 5, pp. 781–787, May 1997, doi: 10.1016/S0092-8674(00)80261-5.

[6] J. S. Sun and J. L. Manley, "A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing," *Genes and Development*, vol. 9, no. 7, pp. 843–854, Apr. 1995, doi: 10.1101/gad.9.7.843.

[7] M. Suñé-Pou *et al.*, "Targeting splicing in the treatment of human disease," *Genes*, vol. 8, no. 3. MDPI AG, Mar. 01, 2017, doi: 10.3390/genes8030087.

[8] S. C. Bonnal, I. López-Oreja, and J. Valcárcel, "Roles and mechanisms of alternative splicing in cancer — implications for care," *Nature Reviews Clinical Oncology*. Nature Research, Apr. 17, 2020, doi: 10.1038/s41571-020-0350-x.

[9] P. J. Shepard, E. A. Choi, A. Busch, and K. J. Hertel, "Efficient internal exon recognition depends on near equal contributions from the 3′ and 5′ splice sites," *Nucleic Acids Research*, vol. 39, no. 20, pp. 8928–8937, Nov. 2011, doi: 10.1093/nar/gkr481.

[10] H. X. Liu, M. Zhang, and A. R. Krainer, "Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins," *Genes and Development*, vol. 12, no. 13, pp. 1998–2012, Jul. 1998, doi: 10.1101/gad.12.13.1998.

[11] L. Cartegni, S. L. Chew, and A. R. Krainer, "Listening to silence and understanding nonsense: Exonic mutations that affect splicing," *Nature Reviews Genetics*, vol. 3, no. 4. Nature Publishing Group, pp. 285–298, 2002, doi: 10.1038/nrg775.

[12] A. J. Matlin, F. Clark, and C. W. J. Smith, "Understanding alternative splicing: Towards a cellular code," *Nature Reviews Molecular Cell Biology*, vol. 6, no. 5. pp. 386–398, May 2005, doi: 10.1038/nrm1645.

[13] E. T. Wang *et al.*, “Alternative isoform regulation in human tissue transcriptomes,” *Nature*, vol. 456, no. 7221, pp. 470–476, Nov. 2008, doi: 10.1038/nature07509.

[14] A. R. Grosso *et al.*, “Tissue-specific splicing factor gene expression signatures,” *Nucleic Acids Research*, vol. 36, no. 15, pp. 4823–4832, 2008, doi: 10.1093/nar/gkn463.

[15] A. Kalsotra and T. A. Cooper, “Functional consequences of developmentally regulated alternative splicing,” *Nature Reviews Genetics*, vol. 12, no. 10. Nature Publishing Group, pp. 715–729, Oct. 16, 2011, doi: 10.1038/nrg3052.

[16] Y. Jin, Y. Yang, and P. Zhang, “New insights into RNA secondary structure in the alternative splicing of pre-mRNAs,” *RNA Biology*, vol. 8, no. 3. Taylor and Francis Inc., pp. 450–457, 2011, doi: 10.4161/rna.8.3.15388.

[17] I. Listerman, A. K. Sapra, and K. M. Neugebauer, “Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells,” *Nature Structural and Molecular Biology*, vol. 13, no. 9, pp. 815–822, Sep. 2006, doi: 10.1038/nsmb1135.

[18] A. Pandya-Jones and D. L. Black, “Co-transcriptional splicing of constitutive and alternative exons,” *RNA*, vol. 15, no. 10, pp. 1896–1908, Oct. 2009, doi: 10.1261/rna.1714509.

[19] M. de La Mata *et al.*, “A slow RNA polymerase II affects alternative splicing in vivo,” *Molecular Cell*, vol. 12, no. 2, pp. 525–532, Aug. 2003, doi: 10.1016/j.molcel.2003.08.001.

[20] H. Shenasa and K. J. Hertel, “Combinatorial regulation of alternative splicing,” *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, vol. 1862, no. 11–12. Elsevier B.V., p. 194392, Nov. 01, 2019, doi: 10.1016/j.bbagrm.2019.06.003.

[21] S. McCracken *et al.*, “The C-terminal domain of RNA polymerase II couples mRNA processing to transcription,” *Nature*, vol. 385, no. 6614, pp. 357–360, Jan. 1997, doi: 10.1038/385357a0.

[22] T. Misteli and D. L. Spector, “RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo,” *Molecular Cell*, vol. 3, no. 6, pp. 697–705, Jun. 1999, doi: 10.1016/S1097-2765(01)80002-2.

[23] H. Tilgner *et al.*, “Nucleosome positioning as a determinant of exon recognition,” *Nature Structural and Molecular Biology*, vol. 16, no. 9, pp. 996–1001, Sep. 2009, doi: 10.1038/nsmb.1658.

[24] E. Kim, A. Magen, and G. Ast, “Different levels of alternative splicing among eukaryotes,” *Nucleic Acids Research*, vol. 35, no. 1, pp. 125–131, Jan. 2007, doi: 10.1093/nar/gkl924.

[25] P. A. F. Galante, N. J. Sakabe, N. Kirschbaum-Slager, and S. J. de Souza, “Detection and evaluation of intron retention events in the human transcriptome,” *RNA*, vol. 10, no. 5, pp. 757–765, May 2004, doi: 10.1261/rna.5123504.

[26] M. Hiller and M. Platzer, “Widespread and subtle: alternative splicing at short-distance tandem sites,” *Trends in Genetics*, vol. 24, no. 5. Elsevier, pp. 246–255, May 01, 2008, doi: 10.1016/j.tig.2008.03.003.

[27] T. v. Ramanouskaya and V. v. Grinev, “The determinants of alternative RNA splicing in human cells,” *Molecular Genetics and Genomics*, vol. 292, no. 6. Springer Verlag, pp. 1175–1195, Dec. 01, 2017, doi: 10.1007/s00438-017-1350-0.

[28] M. Baharlou Houreh, P. Ghorbani Kalkhajeh, A. Niazi, F. Ebrahimi, and E. Ebrahimie, "SpliceDetector: A software for detection of alternative splicing events in human and model organisms directly from transcript IDs," *Scientific Reports*, vol. 8, no. 1, p. 5063, Dec. 2018, doi: 10.1038/s41598-018-23245-1.

[29] L. E. Maquat, "When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells.," *RNA (New York, N.Y.)*, vol. 1, no. 5. Cold Spring Harbor Laboratory Press, pp. 453–465, 1995, Accessed: Jun. 08, 2020. [Online].

[30] E. Nagy and L. E. Maquat, "A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance," *Trends in Biochemical Sciences*, vol. 23, no. 6, pp. 198–199, Jun. 1998, doi: 10.1016/S0968-0004(98)01208-0.

[31] L. E. Maquat, "Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics," *Nature Reviews Molecular Cell Biology*, vol. 5, no. 2. pp. 89–99, Feb. 2004, doi: 10.1038/nrm1310.

[32] J. P. B. Lloyd, "The evolution and diversity of the nonsense-mediated mRNA decay pathway," *F1000Research*, vol. 7, p. 1299, Nov. 2018, doi: 10.12688/f1000research.15872.2.

[33] A. Schaub and E. Glasmacher, "Splicing in immune cells-mechanistic insights and emerging topics," *International Immunology*, vol. 29, no. 4, pp. 173–181, Apr. 2017, doi: 10.1093/intimm/dxx026.

[34] R. Maki *et al.*, "The role of DNA rearrangement and alternative RNA processing in the expression of immunoglobulin delta genes," *Cell*, vol. 24, no. 2, pp. 353–365, May 1981, doi: 10.1016/0092-8674(81)90325-1.

[35] G. Lamson and M. E. Koshland, "Changes in J chain and μ chain rna expression as a function of B cell differentiation," *Journal of Experimental Medicine*, vol. 160, no. 3, pp. 877–892, Sep. 1984, doi: 10.1084/jem.160.3.877.

[36] A. Saxon, D. Diaz-Sanchez, and K. Zhang, "Regulation of the expression of distinct human secreted IgE proteins produced by alternative RNA splicing," in *Biochemical Society Transactions*, 1997, vol. 25, no. 2, pp. 383–387, doi: 10.1042/bst0250383.

[37] J. Y. Ip, A. Tong, Q. Pan, J. D. Topp, B. J. Blencowe, and K. W. Lynch, "Global analysis of alternative splicing during T-cell activation," *RNA*, vol. 13, no. 4, pp. 563–572, Apr. 2007, doi: 10.1261/rna.457207.

[38] I. Meininger *et al.*, "Alternative splicing of MALT1 controls signalling and activation of CD4 + T cells," *Nature Communications*, vol. 7, Apr. 2016, doi: 10.1038/ncomms11292.

[39] I. S. Trowbridge and M. L. Thomas, "CD45: An emerging role as a protein tyrosine phosphatase required for lymphocyte activation and development," *Annual Review of Immunology*, vol. 12, no. 1. Annual Reviews Inc., pp. 85–116, Apr. 1994, doi: 10.1146/annurev.iy.12.040194.000505.

[40] K. Iwami, T. Matsuguchi, A. Masuda, T. Kikuchi, T. Musikacharoen, and Y. Yoshikai, "Cutting Edge: Naturally Occurring Soluble Form of Mouse Toll-Like Receptor 4 Inhibits Lipopolysaccharide Signaling," *The Journal of Immunology*, vol. 165, no. 12, pp. 6682–6686, Dec. 2000, doi: 10.4049/jimmunol.165.12.6682.

[41] B. Ng *et al.*, “Increased noncanonical splicing of autoantigen transcripts provides the structural basis for expression of untolerized epitopes,” *Journal of Allergy and Clinical Immunology*, vol. 114, no. 6, pp. 1463–1470, Dec. 2004, doi: 10.1016/j.jaci.2004.09.006.

[42] M. Hecker *et al.*, “Aberrant expression of alternative splicing variants in multiple sclerosis – A systematic review,” *Autoimmunity Reviews*, vol. 18, no. 7. Elsevier B.V., pp. 721–732, Jul. 01, 2019, doi: 10.1016/j.autrev.2019.05.010.

[43] S. G. Gregory *et al.*, “Interleukin 7 receptor α chain (IL7R) shows allelic and functional association with multiple sclerosis,” *Nature Genetics*, vol. 39, no. 9, pp. 1083–1091, Sep. 2007, doi: 10.1038/ng2103.

[44] C. J. Jensen, J. Stankovich, H. Butzkueven, B. J. Oldfield, and J. P. Rubio, “Common variation in the MOG gene influences transcript splicing in humans,” *Journal of Neuroimmunology*, vol. 229, no. 1–2, pp. 225–231, Dec. 2010, doi: 10.1016/j.jneuroim.2010.07.027.

[45] J. R. B. Newman *et al.*, “Disease-specific biases in alternative splicing and tissue-specific dysregulation revealed by multitissue profiling of lymphocyte gene expression in type 1 diabetes,” *Genome Research*, vol. 27, no. 11, pp. 1807–1815, Nov. 2017, doi: 10.1101/gr.217984.116.

[46] E. M. Dam *et al.*, “The BANK1 SLE-risk variants are associated with alterations in peripheral B cell signaling and development in humans,” *Clinical Immunology*, vol. 173, pp. 171–180, Dec. 2016, doi: 10.1016/j.clim.2016.10.018.

[47] K. Mamegano *et al.*, “Association of LILRA2 (ILT1, LIR7) splice site polymorphism with systemic lupus erythematosus and microscopic polyangiitis,” *Genes and Immunity*, vol. 9, no. 3, pp. 214–223, Feb. 2008, doi: 10.1038/gene.2008.5.

[48] C. A. Odhams *et al.*, “Mapping eQTLs with RNA-seq reveals novel susceptibility genes, non-coding RNAs and alternative-splicing events in systemic lupus erythematosus,” *Human Molecular Genetics*, vol. 26, no. 5, pp. 1003–1017, Mar. 2017, doi: 10.1093/hmg/ddw417.

[49] N. I. Panousis *et al.*, “Combined genetic and transcriptome analysis of patients with SLE: distinct, targetable signatures for susceptibility and severity,” *Annals of the Rheumatic Diseases*, vol. 78, no. 8, pp. 1079–1089, Aug. 2019, doi: 10.1136/annrheumdis-2018-214379.

[50] H. Wai, A. G. L. Douglas, and D. Baralle, “RNA splicing analysis in genomic medicine,” *International Journal of Biochemistry and Cell Biology*, vol. 108, pp. 61–71, Mar. 2019, doi: 10.1016/j.biocel.2018.12.009.

[51] S. Chhangawala, G. Rudy, C. E. Mason, and J. A. Rosenfeld, “The impact of read length on quantification of differentially expressed genes and splice junction detection,” *Genome Biology*, vol. 16, no. 1, p. 131, Jun. 2015, doi: 10.1186/s13059-015-0697-y.

[52] L. Denti, R. Rizzi, S. Beretta, G. della Vedova, M. Previtali, and P. Bonizzoni, “ASGAL: Aligning RNA-Seq data to a splicing graph to detect novel alternative splicing events,” *BMC Bioinformatics*, vol. 19, no. 1, p. 444, Nov. 2018, doi: 10.1186/s12859-018-2436-3.

[53] E. T. Wang *et al.*, “Alternative isoform regulation in human tissue transcriptomes,” *Nature*, vol. 456, no. 7221, pp. 470–476, Nov. 2008, doi: 10.1038/nature07509.

[54] “SplAdder: identification, quantification and testing of alternative splicing events from RNA-Seq data - PubMed.” https://pubmed.ncbi.nlm.nih.gov/26873928/ (accessed Jul. 28, 2020).

[55] J. Harrow *et al.*, "GENCODE: The reference human genome annotation for the ENCODE project," *Genome Research*, vol. 22, no. 9, pp. 1760–1774, Sep. 2012, doi: 10.1101/gr.135350.111.

[56] "Gencode Annotation File." ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_19/gencode.v19.annotation.gtf.gz.

[57] "Ensembl Browser for Gencode release." http://feb2014.archive.ensembl.org/index.html.

[58] P. A. Fujita *et al.*, "The UCSC genome browser database: Update 2011," *Nucleic Acids Research*, vol. 39, no. SUPPL. 1, pp. D876–D882, Jan. 2011, doi: 10.1093/nar/gkq963.

[59] Y. I. Li *et al.*, "Annotation-free quantification of RNA splicing using LeafCutter," *Nature Genetics*, vol. 50, no. 1, pp. 151–158, Jan. 2018, doi: 10.1038/s41588-017-0004-9.

[60] J. R. Doedens *et al.*, "Blood-Borne RNA Correlates with Disease Activity and IFN-Stimulated Gene Expression in Systemic Lupus Erythematosus," *The Journal of Immunology*, vol. 197, no. 7, pp. 2854–2863, Oct. 2016, doi: 10.4049/jimmunol.1601142.

[61] Y. Zhou *et al.*, "CD74 Deficiency Mitigates Systemic Lupus Erythematosus–like Autoimmunity and Pathological Findings in Mice," *The Journal of Immunology*, vol. 198, no. 7, pp. 2568–2577, Apr. 2017, doi: 10.4049/jimmunol.1600028.

[62] P. S. Ramos, E. E. Brown, R. P. Kimberly, and C. D. Langefeld, "Genetic factors predisposing to systemic lupus erythematosus and lupus nephritis," *Seminars in Nephrology*, vol. 30, no. 2, pp. 164–176, Mar. 2010, doi: 10.1016/j.semnephrol.2010.01.007.

[63] H.-Q. Qu *et al.*, "Genetic Control of Alternative Splicing in the TAP2 Gene: Possible Implication in the Genetics of Type 1 Diabetes," *Diabetes*, vol. 56, no. 1, pp. 270–275, Jan. 2007, doi: 10.2337/db06-0865.

[64] M. G. Dozmorov, J. D. Wren, and M. E. Alarcón-Riquelme, "Epigenomic elements enriched in the promoters of autoimmunity susceptibility genes," *Epigenetics*, vol. 9, no. 2, p. 276, 2014, doi: 10.4161/epi.27021.

[65] A. C. Y. Yau *et al.*, "Conserved 33-kb haplotype in the MHC class III region regulates chronic arthritis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 26, pp. E3716–E3724, Jun. 2016, doi: 10.1073/pnas.1600567113.

[66] D. A. de Lima Morais and P. M. Harrison, "Large-Scale Evidence for Conservation of NMD Candidature Across Mammals," *PLoS ONE*, vol. 5, no. 7, p. e11695, Jul. 2010, doi: 10.1371/journal.pone.0011695.

[67] S. Oh, S. Shin, and R. Janknecht, "ETV1, 4 and 5: An oncogenic subfamily of ETS transcription factors," *Biochimica et Biophysica Acta - Reviews on Cancer*, vol. 1826, no. 1. Biochim Biophys Acta, pp. 1–12, Aug. 2012, doi: 10.1016/j.bbcan.2012.02.002.

[68] L. A. Garrett-Sinha, A. Kearly, and A. B. Satterthwaite, "The role of the transcription factor ets1 in lupus and other autoimmune diseases," *Critical Reviews in Immunology*, vol. 36, no. 6. Begell House Inc., pp. 485–509, 2016, doi: 10.1615/CritRevImmunol.2017020284.

[69] L. Jin *et al.*, “The potential role of Ets-1 and miR-326 in CD19 + B cells in the pathogenesis of patients with systemic lupus erythematosus,” *Clinical Rheumatology*, vol. 38, no. 4, pp. 1031–1038, Apr. 2019, doi: 10.1007/s10067-018-4371-0.

[70] O. Saulnier *et al.*, “Abstract LB-323: A novel function of ETS transcription factors in alternative splicing is altered in oncogenic FET-ETS fusions,” in *Cancer Research*, Jul. 2018, vol. 78, no. 13 Supplement, p. LB-323-LB-323, doi: 10.1158/1538-7445.am2018-lb-323.

[71] T. Mori, K. Kitano, S. I. Terawaki, R. Maesaki, Y. Fukami, and T. Hakoshima, “Structural basis for CD44 recognition by ERM proteins,” *Journal of Biological Chemistry*, vol. 283, no. 43, pp. 29602–29612, Oct. 2008, doi: 10.1074/jbc.M803606200.

[72] Y. Li *et al.*, “Phosphorylated ERM Is Responsible for Increased T Cell Polarization, Adhesion, and Migration in Patients with Systemic Lupus Erythematosus,” *The Journal of Immunology*, vol. 178, no. 3, pp. 1938–1947, Feb. 2007, doi: 10.4049/jimmunol.178.3.1938.

7.SUPPLEMENTARY

7.1. Spladder Tool

7.1.1. Installation

Creation of conda environment is achieved via:

```
~$ conda create –name spladder_env python=3.6 pip=19.3 h5py=2.9.0 htslib=1.9 scipy=1.3.2 intervaltree=3.0.2 matplotlib=3.1.2 pysam=0.15.3 statsmodels=0.11.0
```

```
~$ pip install spladder=2.4.2
```

7.1.2. Analysis Pipeline

```
~$ spladder build –b filename.txt -a ~/gencode.v19.annotation.gtf -o ~/build_all
```

Parameters:

`-a`: annotation file

`-b`: list of alignment files (for all conditions)

`-o`: output directory

*filename.txt contains bam file names.

```
~$ spladder test -o ~/build_all  -a ~/healthy.txt -b ~/SLE.txt -M merge_graphs
```

Parameters:

-a: list of alignment files for condition A

-b: list of alignment files for condition B

-M: merge strategy. It supports a unique splicing graph construction for each file and a following combination into a single graph. This procedure is conducted for each gene respectively.

-o: output directory

*healthy.txt contains bam file names of healthy samples.

*SLE.txt contains bam file names of SLE samples.

*All modes operate on the same output directory.

*Parameters were set as default.

```
~$ spladder test -o ~/Documents/sofia/spladder_analysis/build_all/ -
a healthy.txt -b SLE.txt -M merge_graphs --diagnose-plots --plot-
format pdf
```

7.2. Data Analysis

```
~$ cat *unique.tsv | perl -ane 'print if $F[2]<0.01' | grep -v event
| cut -f1-1 > events_pval01.txt
```

```
~$ awk '{if ($1 ~"event_name") print $2}' events_pval01.txt >
ENSG_ids_event_name.txt
```

Script 1.1 | "analysis_of_statistically_significant_events.r":

This script computes the relative frequencies of each event type and the percentage of statistically significant events.

Script 1.12 | "*calculation_intron_retention_length.py*"

It calculates the length of spliced regions for each splicing event.

```
~$ cat test_results_C3_event_name.tsv | perl -ane 'print if
$F[2]<0.01' | grep -v event | cut -f1-10 > event_name_stat.txt
```

```
(not unique)
```

```
~$ awk '{print $2}' event_name_stat.txt | sort –u | wc –l
```

7.2.1. Strategy Intron retention events

Script 1.2 | "find_intron_retentions_pval_01_from_gff3.py":

It matches statistically significant events with their sequence features, according to event ids. As event ids are written like: "intron_retention_id" in intron_retention_stat.txt file and like "intron_retention**.**id" in gff3 file, there is a need for some modifications.

Creation of “intron_retentions_pval_01.gff3”.

Script 1.3 | “find_only_retentions_pval_01_from_gff3.py”:

This script produces a file that contains only intron retentions, with their coordinates and their gene name, from “intron_retentions_pval_01.gff3”.

Creation of "only_retentions_pval_01.gff3”.

File “ENSG_ids_intron_retention.txt” is used as input in UCSC Table Browser so that mRNA sequences (exons and introns) of the given genes are (or to be) extracted.

Output file: intron_retentions_CDS.fa

```
~$ perl -ne '{if ($_=~/>/) {print "\n",$_}; if($_!~/^>/) {chomp;
print $_}}' intron_retentions_CDS.fa | tail -n+2 > intron_ret_CDS.fa
```

Script 1.4 | “make_mrna_from_intron_analysis.py” :

A new file is made from intron_ret_CDS.fa, containing only exons and retained introns. It keeps upper case entries (exons) and entries containing same coordinates with retained introns.

Creation of “transcripts_with_intron_retentions_mrna.fa”.

Script 1.5 | “find_stop_or_ptc_retained_intron_transcripts.py”:

This script detects stop codons that are introduced after the intron retention.

Creation of “retained_intron_transcripts_PTC”.

7.2.2. Strategy for exon skip events

Script 1.6 |“find_exon_skips_pval_01_from_gff3.py”:

It matches statistically significant events with their sequence features, according to event ids. As event ids are written like: “exon_skip_id” in exon_skip_stat.txt file and like “exon_skip.id” in gff3 file, there was a need for some modifications.

Creation of “exon_skips_pval_01.gff3”.

Script 1.7 | “find_only_exon_skips.py”:

This script produces a file containing only skipped exons, with their coordinates and their gene name, from “exon_skips_pval_01.gff3”.

Creation of “only_exon_skips_pval_01.gff3”.

```
~$ awk '{print $4":"$5,$0}' only_exon_skips_pval_01.gff3 | sort > exon_mod
~$ awk '{print $4":"$5,$0}' gencode.v19.annotation.gtf | sort > gencode_mod
~$ join -o 1.2,1.4,1.5,1.6,1.8,2.11,2.13,2.15 exon_mod gencode_mod > exon-gencode
~$ awk '{gsub(/["]/,"\t",$0)}1' exon-gencode | awk '{print $8}' | sort -u > transcript_names_of_skipped_exons
```

File “transcript_names_of_skipped_exons” is used as input in UCSC Table Browser so that mRNA sequences (exons and introns) of the given transcripts are extracted.

Output file: “transcr_exon_skip_CDS.fa”

```
~$ perl -ne '{if ($_=~/>/) {print "\n",$_}; if($_!~/^>/) {chomp; print $_}}' transcr_exon_skips_CDS_compr.fa | tail -n+2 > tran_exon_skips_CDS_compr.fa
```

Script 1.8 | “make_mrna_from_skipped_exons.py”:

This script removes skipped exons from “tran_exon_skips_CDS_compr.fa”, generating “transcripts_without_skipped_exons.fa”.

Script 1.9 | “find_stop_or_ptc_exon_skipped_transcripts.py”:

It detects introduction of PTC and distinguish them according to their potential effects (NMD or termination of translation) For each exon, position of every stop codon is calculated. “transcripts_NMD” , “transcripts_PTC”

Using Biomart to find out gene name and transcript type of each transcript.

7.2.3. Strategy for Alternative 3’ or 5’ splice site events

Script 1.10 | “find_alt_prime_pval_01_from_gff3.py”:

It matches statistically significant events with their sequence features, according to event ids. Creation of “alt_3prime_pval_01.gff3” and “alt_5prime_pval_01.gff3”.

```
~$awk '{if ($7=="+") print $0}' alt_3prime_pval_01.gff3 > alt_3prime_pval_01-PLUS.gff3
~$awk '{if ($7=="-") print $0}' alt_3prime_pval_01.gff3 > alt_3prime_pval_01-MINUS.gff3
~$awk '{if ($7=="+") print $0}' alt_5prime_pval_01.gff3 > alt_3prime_pval_01-PLUS.gff3
```

```
~$awk '{if ($7=="-") print $0}' alt_5prime_pval_01.gff3 >
alt_3prime_pval_01-MINUS.gff3
```

Script 1.11 | "find_only_alt_prime_pval_01_from_gff3.py":
It finds both isoforms of the region that is spliced for each event.

Creation of "only_alt_3prime_pval_01-PLUS.gff3", "only_alt_3prime_pval_01-MINUS.gff3", "only_alt_5prime_pval_01-PLUS.gff3", "only_alt_5prime_pval_01-MINUS.gff3"

```
~$ awk '{print $4":"$5,$0}' only_alt_3prime_pval_01-MINUS.gff3 |
sort > alt3_minus_mod
~$ join -o 1.10 alt3_minus_mod gencode_mod | sort -u >
annotated_isoforms_alt3_minus
~$ awk '{print $10}' alt3_minus_mod | sort -u >
all_isoforms_alt3_minus
~$ diff all_isoforms_alt3_minus annotated_isoforms_alt3_minus >
not_annotated_isoforms_alt3_minus
```

7.2.4. Functional analysis

Script 1.13 | "final_function_anal.r"

Functional profiling of alternative spliced genes using gProfiler and ClusterProfiler tool.

7.3. Results

Identification of alternative spliced transcripts - Premature termination codon detection

Table S1 | Events with adjusted p value less than 0.05.

Event Type	All	Unique genes
Alternative 3prime	1	1
Alternative 5prime	9	8
Exon skip	9	8
Intron retention	28	18
Multiple exon skip	0	0
Mutually exclusive exons	0	0
		35

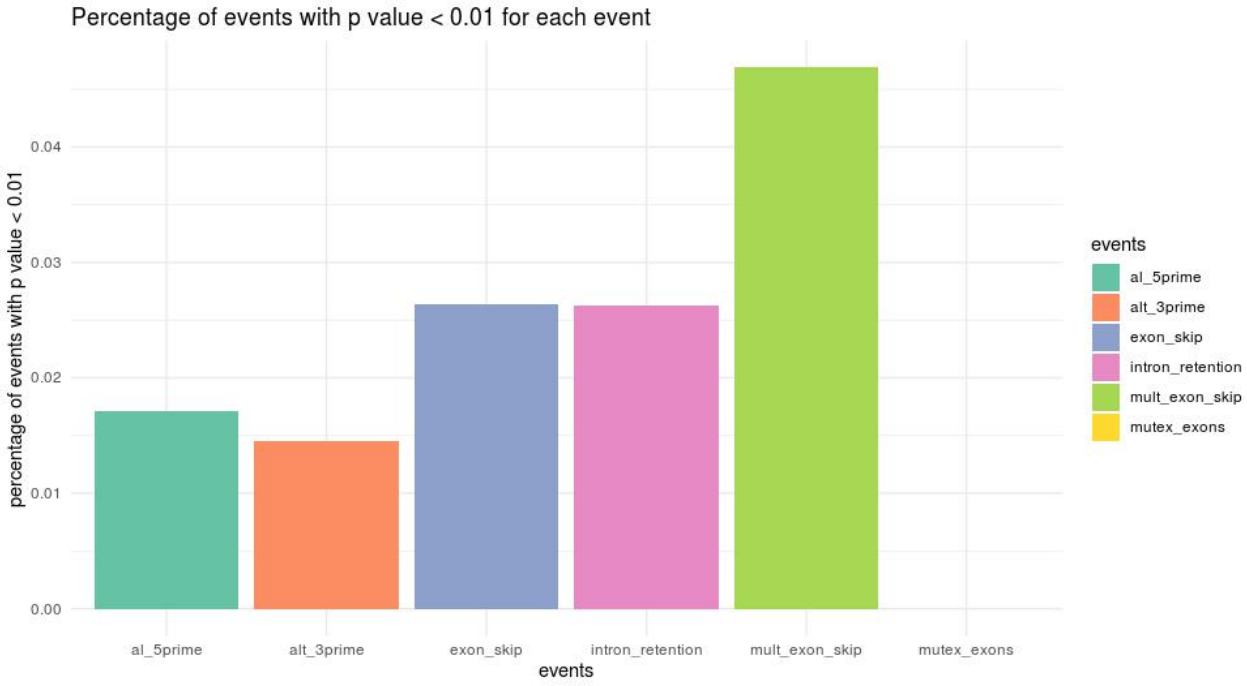


Figure S1 | Percentage of statistical significant events.

Table S2 | Alternative Splicing Events with p value less than 0.01.

event_type	gene	p_val	p_val_adj	Mean event count Healthy	Mean event count SLE	log2FC event count	symbol
alt5prime	ENSG00000122783.12	2.643E-11	2.232E-08	4.723	3.509	0.429	CYREN
alt5prime	ENSG00000122783.12	2.643E-11	2.232E-08	4.723	3.509	0.429	CYREN
intron_retention	ENSG00000187912.7	2.134E-11	5.119E-08	2.636	1.017	1.373	CLEC17A
intron_retention	ENSG00000250264.1	3.033E-11	5.119E-08	6.365	4.057	0.650	TAP2
intron_retention	ENSG00000122783.12	3.813E-11	5.119E-08	4.642	3.465	0.422	CYREN
exon_skip	ENSG00000124243.13	2.000E-09	4.719E-06	6.424	4.280	0.586	BCAS4
intron_retention	ENSG00000117748.5	5.569E-08	5.608E-05	31.253	24.445	0.354	RPA2
intron_retention	ENSG00000117748.5	1.102E-07	8.878E-05	12.638	9.510	0.410	RPA2
alt5prime	ENSG00000149925.12	2.804E-07	1.578E-04	65.620	49.311	0.412	ALDOA
intron_retention	ENSG00000187912.7	3.253E-07	2.184E-04	2.622	1.388	0.918	CLEC17A
intron_retention	ENSG00000187912.7	1.200E-06	6.155E-04	3.037	1.588	0.935	CLEC17A
intron_retention	ENSG00000123136.10	1.222E-06	6.155E-04	15.146	11.776	0.363	DDX39A
intron_retention	ENSG00000187912.7	2.347E-06	1.050E-03	3.487	1.843	0.920	CLEC17A
alt5prime	ENSG00000021355.8	4.265E-06	1.801E-03	5.707	4.876	0.227	SERPINB1

intron_retention	ENSG00000117748.5	5.725E-06	2.306E-03	11.875	9.520	0.319	RPA2
alt5prime	ENSG00000134882.11	7.358E-06	2.486E-03	15.900	12.458	0.352	UBAC2
alt3prime	ENSG00000187912.7	1.324E-06	2.825E-03	3.595	1.855	0.955	CLEC17A
intron_retention	ENSG00000123136.10	1.053E-05	3.856E-03	3.729	2.657	0.489	DDX39A
intron_retention	ENSG00000100226.11	1.347E-05	4.521E-03	2.213	3.874	-0.808	GTPBP1
alt5prime	ENSG00000100226.11	1.752E-05	4.932E-03	2.266	3.955	-0.804	GTPBP1
intron_retention	ENSG00000261971.2	1.605E-05	4.974E-03	3.601	2.637	0.450	RP11-473M20.7
intron_retention	ENSG00000123329.13	1.866E-05	5.369E-03	5.523	3.953	0.482	ARHGAP9
exon_skip	ENSG00000107099.11	5.343E-06	6.169E-03	7.656	5.135	0.576	DOCK8
alt5prime	ENSG00000161010.10	2.562E-05	6.181E-03	2.891	1.740	0.732	MRNIP
intron_retention	ENSG00000100813.10	2.745E-05	7.147E-03	11.145	8.714	0.355	ACIN1
intron_retention	ENSG00000104783.7	2.839E-05	7.147E-03	9.297	4.671	0.993	KCNN4
intron_retention	ENSG00000265148.1	3.488E-05	8.264E-03	3.029	1.693	0.839	TSPOAP1-AS1
intron_retention	ENSG00000081237.14	3.880E-05	8.683E-03	11.895	8.039	0.565	PTPRC
intron_retention	ENSG00000123136.10	6.460E-05	1.319E-02	19.958	16.925	0.238	DDX39A
intron_retention	ENSG00000204536.9	6.548E-05	1.319E-02	2.250	1.204	0.902	CCHCR1
alt5prime	ENSG00000100242.11	6.796E-05	1.435E-02	77.618	58.253	0.414	SUN2
intron_retention	ENSG00000100242.11	7.849E-05	1.506E-02	75.730	57.108	0.407	SUN2
intron_retention	ENSG00000081237.14	8.335E-05	1.526E-02	5.380	3.516	0.614	PTPRC
intron_retention	ENSG00000169180.7	8.909E-05	1.560E-02	6.275	4.870	0.366	XPO6
intron_retention	ENSG00000068831.14	9.437E-05	1.584E-02	8.246	5.597	0.559	RASGRP2
exon_skip	ENSG00000161010.10	2.135E-05	1.643E-02	2.964	1.771	0.743	MRNIP
alt5prime	ENSG00000136280.11	9.601E-05	1.802E-02	11.647	9.049	0.364	CCM2
exon_skip	ENSG00000149925.12	4.481E-05	2.113E-02	48.154	39.381	0.290	ALDOA
exon_skip	ENSG00000182473.16	4.576E-05	2.113E-02	1.598	0.900	0.828	EXOC7
exon_skip	ENSG00000160633.8	6.076E-05	2.317E-02	9.171	6.504	0.496	SAFB
exon_skip	ENSG00000124243.13	7.057E-05	2.317E-02	2.936	2.077	0.499	BCAS4
exon_skip	ENSG00000197142.6	8.026E-05	2.317E-02	8.622	5.433	0.666	ACSL5
intron_retention	ENSG00000104894.7	1.531E-04	2.467E-02	5.518	3.959	0.479	CD37
intron_retention	ENSG00000197530.8	2.675E-04	4.145E-02	3.967	5.510	-0.474	MIB2
intron_retention	ENSG00000100242.11	3.230E-04	4.702E-02	86.167	66.244	0.379	SUN2
intron_retention	ENSG00000100242.11	3.268E-04	4.702E-02	82.478	63.770	0.371	SUN2
exon_skip	ENSG00000061936.5	1.855E-04	4.759E-02	6.092	4.094	0.574	SFSWAP
alt3prime	ENSG00000139644.8	4.976E-05	5.307E-02	4.805	6.582	-0.454	TMBIM6
intron_retention	ENSG00000141086.13	4.192E-04	5.675E-02	3.716	5.610	-0.594	CTRL

intron_retention	ENSG00000197111.11	4.226E-04	5.675E-02	98.483	79.326	0.312	PCBP2
intron_retention	ENSG00000117748.5	4.386E-04	5.699E-02	8.149	6.973	0.225	RPA2
alt3prime	ENSG00000100226.11	8.763E-05	6.231E-02	1.690	3.029	-0.842	GTPBP1
alt5prime	ENSG00000100445.12	4.080E-04	6.265E-02	13.068	9.589	0.447	SDR39U1
alt5prime	ENSG00000100445.12	4.080E-04	6.265E-02	13.068	9.589	0.447	SDR39U1
mult_exon_skip	ENSG00000005844.13	6.331E-04	6.790E-02	16.876	12.137	0.476	ITGAL
mult_exon_skip	ENSG00000136280.11	7.073E-04	6.790E-02	21.572	18.432	0.227	CCM2
exon_skip	ENSG00000163565.14	2.969E-04	6.854E-02	7.453	8.185	-0.135	IFI16
intron_retention	ENSG00000131748.11	5.510E-04	6.935E-02	25.275	22.235	0.185	STARD3
intron_retention	ENSG00000077044.5	5.918E-04	7.041E-02	11.331	8.355	0.439	DGKD
intron_retention	ENSG00000148399.7	5.943E-04	7.041E-02	2.063	1.110	0.895	DPH7
intron_retention	ENSG00000197111.11	6.634E-04	7.634E-02	65.995	54.471	0.277	PCBP2
alt5prime	ENSG00000189430.8	5.828E-04	7.704E-02	5.014	2.421	1.050	NCR1
alt5prime	ENSG00000157873.13	5.930E-04	7.704E-02	11.081	9.429	0.233	TNFRSF14
mult_exon_skip	ENSG00000198089.10	1.497E-03	9.584E-02	1.972	1.097	0.846	SFI1
intron_retention	ENSG00000187741.10	8.631E-04	9.657E-02	3.750	5.454	-0.540	FANCA
exon_skip	ENSG00000136280.11	5.038E-04	9.885E-02	4.114	2.962	0.474	CCM2
exon_skip	ENSG00000148700.9	5.137E-04	9.885E-02	2.438	1.362	0.840	ADD3
intron_retention	ENSG00000108788.7	9.751E-04	1.062E-01	13.624	11.899	0.195	MLX
intron_retention	ENSG00000103510.15	1.002E-03	1.062E-01	9.845	7.360	0.420	KAT8
intron_retention	ENSG00000121716.14	1.071E-03	1.107E-01	18.192	13.018	0.483	PILRB
mult_exon_skip	ENSG00000143398.15	2.824E-03	1.132E-01	1.910	1.216	0.651	PIP5K1A
mult_exon_skip	ENSG00000136279.14	2.947E-03	1.132E-01	38.567	36.502	0.079	DBNL
intron_retention	ENSG00000182473.16	1.155E-03	1.163E-01	1.765	1.124	0.650	EXOC7
exon_skip	ENSG00000105968.14	7.704E-04	1.245E-01	8.486	9.864	-0.217	H2AZ2
exon_skip	ENSG00000076201.10	8.156E-04	1.245E-01	3.010	1.788	0.751	PTPN23
exon_skip	ENSG00000075151.15	8.785E-04	1.245E-01	2.282	1.664	0.456	EIF4G3
exon_skip	ENSG00000004534.10	8.806E-04	1.245E-01	8.035	5.239	0.617	RBM6
exon_skip	ENSG00000130254.7	9.166E-04	1.245E-01	4.143	2.761	0.585	SAFB2
intron_retention	ENSG00000185404.12	1.331E-03	1.307E-01	1.908	2.672	-0.485	SP140L
intron_retention	ENSG00000164828.13	1.411E-03	1.347E-01	2.556	1.629	0.649	SUN1
intron_retention	ENSG00000196187.7	1.438E-03	1.347E-01	7.977	5.571	0.518	TMEM63A
intron_retention	ENSG00000074370.13	1.537E-03	1.407E-01	2.472	1.502	0.718	ATP2A3
exon_skip	ENSG00000139613.7	1.121E-03	1.438E-01	11.215	7.549	0.571	SMARCC2
intron_retention	ENSG00000169230.5	1.638E-03	1.466E-01	40.626	39.904	0.026	PRELID1

intron_retention	ENSG00000197451.6	1.726E-03	1.494E-01	5.163	6.355	-0.300	HNRNPAB
intron_retention	ENSG00000108669.12	1.744E-03	1.494E-01	16.674	13.234	0.333	CYTH1
exon_skip	ENSG00000168397.12	1.261E-03	1.532E-01	8.799	6.418	0.455	ATG4B
intron_retention	ENSG00000172725.9	1.942E-03	1.630E-01	10.661	12.968	-0.283	CORO1B
intron_retention	ENSG00000187741.10	2.097E-03	1.724E-01	6.411	8.798	-0.457	FANCA
intron_retention	ENSG00000101158.8	2.175E-03	1.752E-01	7.790	5.667	0.459	NELFCD
alt3prime	ENSG00000197530.8	3.403E-04	1.815E-01	4.094	5.643	-0.463	MIB2
intron_retention	ENSG00000204482.6	2.356E-03	1.836E-01	65.924	55.713	0.243	LST1
intron_retention	ENSG00000100242.11	2.370E-03	1.836E-01	109.635	87.419	0.327	SUN2
intron_retention	ENSG00000149743.9	2.430E-03	1.846E-01	6.272	4.772	0.394	TRPT1
alt3prime	ENSG00000225484.2	4.926E-04	1.870E-01	4.375	3.546	0.303	NUTM2B-AS1
alt3prime	ENSG00000148399.7	5.261E-04	1.870E-01	2.097	1.117	0.909	DPH7
alt5prime	ENSG00000113119.8	1.895E-03	2.089E-01	2.499	3.234	-0.372	TMCO6
alt5prime	ENSG00000133313.10	1.899E-03	2.089E-01	11.994	8.969	0.419	CNDP2
alt5prime	ENSG00000075415.8	1.979E-03	2.089E-01	90.528	73.462	0.301	SLC25A3
alt3prime	ENSG00000121716.14	8.084E-04	2.098E-01	18.589	13.193	0.495	PILRB
alt3prime	ENSG00000169660.11	8.249E-04	2.098E-01	8.476	6.649	0.350	HEXD
alt3prime	ENSG00000033627.10	1.024E-03	2.098E-01	2.294	1.683	0.447	ATP6V0A1
alt3prime	ENSG00000182473.16	1.029E-03	2.098E-01	1.797	1.144	0.651	EXOC7
alt3prime	ENSG00000169230.5	1.195E-03	2.098E-01	41.405	40.401	0.035	PRELID1
alt3prime	ENSG00000204084.8	1.278E-03	2.098E-01	3.504	2.144	0.709	INPP5B
alt3prime	ENSG00000265148.1	1.370E-03	2.098E-01	3.510	2.345	0.582	TSPOAP1-AS1
alt3prime	ENSG00000108669.12	1.377E-03	2.098E-01	16.961	13.400	0.340	CYTH1
mult_exon_skip	ENSG00000188603.12	7.361E-03	2.098E-01	10.232	9.663	0.083	CLN3
mult_exon_skip	ENSG00000134882.11	7.946E-03	2.098E-01	6.707	5.472	0.294	UBAC2
mult_exon_skip	ENSG00000188603.12	9.446E-03	2.098E-01	14.782	14.147	0.063	CLN3
mult_exon_skip	ENSG00000064012.17	9.835E-03	2.098E-01	24.150	20.918	0.207	CASP8
exon_skip	ENSG00000121716.14	1.916E-03	2.161E-01	2.868	1.888	0.604	PILRB
exon_skip	ENSG00000132670.16	2.001E-03	2.161E-01	3.343	2.252	0.570	PTPRA
exon_skip	ENSG00000161010.10	2.119E-03	2.161E-01	2.694	1.817	0.568	MRNIP
exon_skip	ENSG00000182973.14	2.153E-03	2.161E-01	4.439	3.012	0.560	CNOT10
intron_retention	ENSG00000172932.10	2.992E-03	2.219E-01	10.417	9.364	0.154	ANKRD13D
intron_retention	ENSG00000204463.8	3.070E-03	2.219E-01	18.265	29.105	-0.672	BAG6
intron_retention	ENSG00000121716.14	3.138E-03	2.219E-01	2.698	1.853	0.542	PILRB

intron_retention	ENSG00000113108.13	3.140E-03	2.219E-01	2.366	1.719	0.461	APBB3
intron_retention	ENSG00000161547.10	3.212E-03	2.222E-01	1.803	2.495	-0.469	SRSF2
intron_retention	ENSG00000129968.11	3.255E-03	2.222E-01	3.886	2.627	0.565	ABHD17A
intron_retention	ENSG00000169230.5	3.313E-03	2.224E-01	28.578	28.548	0.001	PRELID1
intron_retention	ENSG00000166986.8	3.449E-03	2.244E-01	12.128	10.317	0.233	MARS1
intron_retention	ENSG00000072071.12	3.490E-03	2.244E-01	1.796	1.164	0.626	ADGRL1
intron_retention	ENSG00000141480.13	3.510E-03	2.244E-01	5.423	5.020	0.112	ARRB2
exon_skip	ENSG00000088833.13	2.472E-03	2.250E-01	19.296	16.901	0.191	NSFL1C
exon_skip	ENSG00000113648.12	2.492E-03	2.250E-01	9.362	9.026	0.053	MACROH2A1
exon_skip	ENSG00000105699.12	2.534E-03	2.250E-01	2.592	1.456	0.832	LSR
intron_retention	ENSG00000123136.10	3.618E-03	2.259E-01	17.559	15.555	0.175	DDX39A
intron_retention	ENSG00000158805.7	3.701E-03	2.259E-01	10.558	9.088	0.216	ZNF276
intron_retention	ENSG00000204482.6	3.726E-03	2.259E-01	11.255	9.583	0.232	LST1
intron_retention	ENSG00000135486.13	3.768E-03	2.259E-01	2.635	1.827	0.528	HNRNPA1
intron_retention	ENSG00000172725.9	3.813E-03	2.259E-01	17.972	20.825	-0.213	CORO1B
exon_skip	ENSG00000136280.11	2.662E-03	2.276E-01	21.347	18.646	0.195	CCM2
exon_skip	ENSG00000110713.11	2.763E-03	2.278E-01	3.774	2.455	0.620	NUP98
exon_skip	ENSG00000101084.12	2.876E-03	2.290E-01	4.766	4.672	0.029	RAB5IF
intron_retention	ENSG00000248333.3	4.009E-03	2.312E-01	14.804	12.336	0.263	CDK11B
intron_retention	ENSG00000182473.16	4.017E-03	2.312E-01	3.583	2.779	0.366	EXOC7
intron_retention	ENSG00000003756.12	4.158E-03	2.328E-01	8.046	6.591	0.288	RBM5
intron_retention	ENSG00000130489.8	4.162E-03	2.328E-01	3.533	11.317	-1.679	SCO2
exon_skip	ENSG00000109118.9	3.025E-03	2.329E-01	18.405	15.835	0.217	PHF12
exon_skip	ENSG00000231721.2	3.237E-03	2.343E-01	3.417	2.404	0.507	LINC-PINT
exon_skip	ENSG00000204482.6	3.258E-03	2.343E-01	11.785	9.989	0.238	LST1
exon_skip	ENSG00000172725.9	3.348E-03	2.343E-01	11.339	13.763	-0.279	CORO1B
exon_skip	ENSG00000019582.10	3.613E-03	2.454E-01	2.388	3.181	-0.414	CD74
exon_skip	ENSG00000142230.7	3.786E-03	2.498E-01	6.732	4.884	0.463	SAE1
intron_retention	ENSG00000100288.15	4.620E-03	2.539E-01	11.765	9.762	0.269	CHKB
intron_retention	ENSG00000204351.7	4.664E-03	2.539E-01	4.381	3.289	0.414	SKIV2L
intron_retention	ENSG00000172725.9	4.745E-03	2.541E-01	13.580	15.554	-0.196	CORO1B
intron_retention	ENSG00000178761.10	4.794E-03	2.541E-01	2.325	1.288	0.852	FAM219B
exon_skip	ENSG00000182944.13	3.978E-03	2.551E-01	25.270	21.529	0.231	EWSR1
alt5prime	ENSG00000172725.9	2.575E-03	2.558E-01	10.991	13.371	-0.283	CORO1B
intron_retention	ENSG00000213983.7	5.150E-03	2.666E-01	15.144	11.838	0.355	AP1G2

intron_retention	ENSG00000042445.9	5.163E-03	2.666E-01	3.305	2.323	0.508	RETSAT
exon_skip	ENSG00000109919.5	4.519E-03	2.695E-01	5.929	4.534	0.387	MTCH2
exon_skip	ENSG00000198089.10	4.635E-03	2.695E-01	4.909	3.336	0.558	SFI1
exon_skip	ENSG00000132155.7	4.640E-03	2.695E-01	19.945	18.311	0.123	RAF1
exon_skip	ENSG00000105699.12	4.762E-03	2.695E-01	4.260	2.588	0.719	LSR
exon_skip	ENSG00000142188.12	4.785E-03	2.695E-01	2.788	2.101	0.408	TMEM50B
exon_skip	ENSG00000213995.7	5.179E-03	2.807E-01	1.960	1.227	0.676	NAXD
exon_skip	ENSG00000140367.7	5.227E-03	2.807E-01	3.392	1.910	0.829	UBE2Q2
exon_skip	ENSG00000130723.13	5.387E-03	2.827E-01	3.006	1.807	0.735	PRRC2B
intron_retention	ENSG00000172732.7	5.625E-03	2.868E-01	9.002	7.046	0.353	MUS81
intron_retention	ENSG00000181090.13	5.716E-03	2.878E-01	7.442	5.441	0.452	EHMT1
exon_skip	ENSG00000183624.9	5.752E-03	2.951E-01	4.210	2.771	0.603	HMCES
intron_retention	ENSG00000174885.8	6.021E-03	2.969E-01	6.152	6.099	0.012	NLRP6
intron_retention	ENSG00000185033.10	6.044E-03	2.969E-01	11.765	10.748	0.130	SEMA4B
alt5prime	ENSG00000140995.12	3.233E-03	2.972E-01	5.148	4.218	0.287	DEF8
alt5prime	ENSG00000182473.16	3.343E-03	2.972E-01	3.659	2.831	0.370	EXOC7
intron_retention	ENSG00000180871.3	6.132E-03	2.976E-01	11.367	9.586	0.246	CXCR2
intron_retention	ENSG00000139192.7	6.372E-03	3.056E-01	2.888	2.314	0.320	TAPBPL
intron_retention	ENSG00000124831.14	6.504E-03	3.082E-01	5.301	4.136	0.358	LRRFIP1
intron_retention	ENSG00000100242.11	6.594E-03	3.089E-01	32.509	26.134	0.315	SUN2
exon_skip	ENSG00000103064.9	6.191E-03	3.107E-01	1.964	2.198	-0.162	SLC7A6
intron_retention	ENSG00000187741.10	6.762E-03	3.117E-01	2.582	3.479	-0.430	FANCA
intron_retention	ENSG00000141644.13	6.818E-03	3.117E-01	2.539	1.855	0.452	MBD1
intron_retention	ENSG00000117448.9	6.937E-03	3.117E-01	5.534	4.462	0.311	AKR1A1
intron_retention	ENSG00000165792.13	7.088E-03	3.117E-01	6.139	3.887	0.659	METTL17
intron_retention	ENSG00000160796.12	7.091E-03	3.117E-01	5.417	5.090	0.090	NBEAL2
intron_retention	ENSG00000064012.17	7.120E-03	3.117E-01	25.927	21.908	0.243	CASP8
exon_skip	ENSG00000166912.12	6.418E-03	3.135E-01	2.786	1.988	0.487	MTMR10
exon_skip	ENSG00000076513.12	6.517E-03	3.135E-01	16.080	14.142	0.185	ANKRD13A
intron_retention	ENSG00000010810.13	7.440E-03	3.205E-01	5.142	3.310	0.635	FYN
intron_retention	ENSG00000123136.10	7.481E-03	3.205E-01	27.238	24.987	0.124	DDX39A
intron_retention	ENSG00000140368.8	7.559E-03	3.205E-01	41.531	39.392	0.076	PSTPIP1
intron_retention	ENSG00000120889.8	7.722E-03	3.213E-01	2.731	1.975	0.468	TNFRSF10B
intron_retention	ENSG00000157873.13	7.737E-03	3.213E-01	1.749	1.397	0.324	TNFRSF14
alt3prime	ENSG00000113648.12	2.436E-03	3.352E-01	9.191	8.816	0.060	MACRO

							H2A1
alt3prime	ENSG00000172725.9	2.514E-03	3.352E-01	10.921	13.251	-0.279	CORO1B
exon_skip	ENSG00000182149.16	7.185E-03	3.357E-01	12.928	10.793	0.260	IST1
exon_skip	ENSG00000140750.12	7.270E-03	3.357E-01	3.753	2.578	0.542	ARHGAP17
intron_retention	ENSG00000038358.10	8.313E-03	3.382E-01	6.006	4.246	0.500	EDC4
intron_retention	ENSG00000187741.10	8.336E-03	3.382E-01	2.286	3.359	-0.555	FANCA
intron_retention	ENSG00000105122.8	8.450E-03	3.382E-01	28.763	21.722	0.405	RASAL3
intron_retention	ENSG00000135924.11	8.554E-03	3.382E-01	5.926	4.653	0.349	DNAJB2
intron_retention	ENSG00000128159.7	8.565E-03	3.382E-01	2.730	1.704	0.680	TUBGCP6
exon_skip	ENSG00000115548.12	7.795E-03	3.385E-01	4.718	3.133	0.591	KDM3A
exon_skip	ENSG00000123562.12	7.816E-03	3.385E-01	5.541	3.992	0.473	MORF4L2
exon_skip	ENSG00000188603.12	7.889E-03	3.385E-01	17.155	16.464	0.059	CLN3
exon_skip	ENSG00000003756.12	7.915E-03	3.385E-01	6.140	5.046	0.283	RBM5
alt5prime	ENSG00000272752.1	4.018E-03	3.393E-01	19.196	14.252	0.430	STAG3L5P-PVRIG2P-PILRB
intron_retention	ENSG00000197111.11	8.751E-03	3.422E-01	14.684	12.206	0.267	PCBP2
intron_retention	ENSG00000143258.11	8.964E-03	3.472E-01	5.700	4.020	0.504	USP21
alt5prime	ENSG00000177479.15	4.564E-03	3.514E-01	2.089	1.267	0.721	ARIH2
alt5prime	ENSG00000124831.14	4.716E-03	3.514E-01	5.403	4.125	0.389	LRRFIP1
alt5prime	ENSG00000130475.10	4.785E-03	3.514E-01	8.092	6.073	0.414	FCHO1
exon_skip	ENSG00000112305.10	8.660E-03	3.523E-01	3.446	2.350	0.552	SMAP1
exon_skip	ENSG00000112983.13	8.697E-03	3.523E-01	2.892	1.957	0.564	BRD8
exon_skip	ENSG00000112983.13	8.697E-03	3.523E-01	2.892	1.957	0.564	BRD8
intron_retention	ENSG00000168918.9	9.312E-03	3.572E-01	15.025	13.078	0.200	INPP5D
exon_skip	ENSG00000100225.13	9.142E-03	3.587E-01	6.198	5.558	0.157	FBXO7
exon_skip	ENSG00000125970.7	9.363E-03	3.587E-01	50.862	45.869	0.149	RALY
exon_skip	ENSG00000136279.14	9.566E-03	3.587E-01	38.660	37.329	0.051	DBNL
exon_skip	ENSG00000136933.12	9.767E-03	3.587E-01	2.124	1.520	0.483	RABEPK
intron_retention	ENSG00000168894.5	9.660E-03	3.671E-01	2.649	2.227	0.250	RNF181
alt5prime	ENSG00000237765.2	5.916E-03	4.163E-01	4.157	4.222	-0.022	FAM200B
alt5prime	ENSG00000177302.10	6.205E-03	4.192E-01	2.860	2.105	0.442	TOP3A
alt3prime	ENSG00000003756.12	3.535E-03	4.269E-01	8.178	6.664	0.295	RBM5
alt3prime	ENSG00000168883.15	3.602E-03	4.269E-01	9.741	7.833	0.314	USP39
alt5prime	ENSG00000007237.14	7.800E-03	4.979E-01	1.817	3.231	-0.830	GAS7

alt5prime	ENSG00000187741.10	7.960E-03	4.979E-01	2.639	3.566	-0.434	FANCA
alt5prime	ENSG00000136279.14	8.818E-03	5.319E-01	39.210	37.774	0.054	DBNL
alt5prime	ENSG00000114738.6	9.477E-03	5.519E-01	16.858	16.080	0.068	MAPKAPK3
alt3prime	ENSG00000126214.16	5.160E-03	5.793E-01	14.112	11.732	0.266	KLC1
alt3prime	ENSG00000140553.12	5.747E-03	6.129E-01	5.278	3.868	0.448	UNC45A
alt3prime	ENSG00000184009.5	6.305E-03	6.405E-01	368.018	318.773	0.207	ACTG1
alt3prime	ENSG00000197111.11	7.135E-03	6.625E-01	14.958	12.363	0.275	PCBP2
alt3prime	ENSG00000011275.14	7.144E-03	6.625E-01	4.803	3.358	0.516	RNF216
alt3prime	ENSG00000196323.7	7.768E-03	6.627E-01	3.475	4.282	-0.301	ZBTB44
alt3prime	ENSG00000196323.7	7.768E-03	6.627E-01	3.475	4.282	-0.301	ZBTB44
alt3prime	ENSG00000130475.10	8.392E-03	6.726E-01	12.045	9.221	0.385	FCHO1
alt3prime	ENSG00000102901.8	9.150E-03	6.726E-01	4.988	3.361	0.570	CENPT
alt3prime	ENSG00000102901.8	9.150E-03	6.726E-01	4.988	3.361	0.570	CENPT
alt3prime	ENSG00000123562.12	9.360E-03	6.726E-01	5.444	3.955	0.461	MORF4L2
alt3prime	ENSG00000170776.15	9.923E-03	6.726E-01	5.932	4.642	0.354	AKAP13
alt3prime	ENSG00000106009.11	9.995E-03	6.726E-01	16.996	13.655	0.316	BRAT1