Transcriptional and post-transcriptional regulation across biological units and time

Reza Sodaei

TESI DOCTORAL UPF / 2019

Thesis supervisors

Roderic Guigo

Manuel Irimia

Department of bioinformatics and genomics

Center for regulatory genomics(CRG), Barcelona

Abstract
How do genomes generate different tissues, behaviours, individuals or species? We try to identify parts of the human transcriptome that we found only in a tissue. Then, we focus on inter-individual extreme transcriptomic differences, and how this principle is extendable to species variation. In the second chapter, we measured the transcriptomic shifts during day-night and seasonal cycles in human postmortem tissues.

Resum
Com generen diferents genomes diferents teixits, conductes, individus o espècies? Tractem d'identificar parts del transcriptoma humà que hem trovat només en un teixit. Seguidament, ens centrem en diferències transcriptòmices extremes entre individus, i en com aquest principi es pot extendre a les variacions de espècie. Al segon capítol, mesurem els canvis transcriptòmics durant cicles de dia i nit així com estacionals en teixits humans postmortem.

Preface

For the first time in high school when I saw someone used a mathematical expression to predict the behaviour of an enzyme, thinking about biology suddenly felt much better. It opened a window that I have not opened. I have never been interested in curing diseases, but I always like to understand life. This knowledge, regardless of its application in engineering or medicine, matters because we never know the consequences of knowing something. The whole process of understanding for me is delicious.

I tried to follow what I found interesting as much as I could. Postmortem samples are great resources to study human biology. Although it is not clear how death affects the transcriptional regulation, analyzing transcriptomic postmortem data could prove fruitful.

One of the nagging questions in biology is how different tissues function distinctly at the molecular level. Analyzing the key features of tissue transcriptome could get us closer to this answer. Furthermore, I explore what the transcriptomic differences in protein-coding genes across individuals and species are. Individual differences are a smaller version of species differences, in some ways. Studying both could give us a way to understand how we evolve at the transcriptomic level.

I recently became more interested in the daily and seasonal changes across tissues. First, because I was curious whether it is possible to observe the circadian transcriptional behaviour from dead people without reading any literature about it. Second, seasonal changes for me was mysterious. Why so many genes were affected by season in human, human do not hibernate or change coat colour. So, I just followed the lead and discovered more.

Acknowledgement

The event horizon of the deadline is catching up, so I have to make it brief. While I do not know who had an influence on this thesis, I will do my best. Now it is 5 AM, and I am very tired.

Thanks to Roderic who gave me this opportunity, I will be grateful forever. I learned a lot from him in these years. Thanks to Manu for keeping my mind focused on the efficient path. I also learned a lot from him, and I really appreciate it. Thanks a lot to Romina and Gloria for their help all the time.

Thanks to football people Luciano and Moritz who organized this weekly dose of total disconnection.

Thanks to my family who love me.

Thanks to my very good friends for their existence and pleasant conversations; Bambino, Cebolla, Tigi, B, Savannah, Bathole, Amador, Vava, Ili, Linux, Edu, Dani, Geri, Chappa, Martini, Serendipity, Di Carlo, Tulido, Manu mexic, Digi, Bea, moca, cesi, Dimi, and dibartolos.

1 Chapter 1 - Introduction to biological units

Gemmules, pangenesis, and Lamarck

All humans have a brain, intestine, and many other organs and tissues. Independently of where humans grow up on Earth, they develop the same tissues. Although different in size and architecture, other mammals also have the same (homologous) organs that function fundamentally in the same manner. How is it possible that they always develop the same organs? What is the source of variation that creates differences between individuals of the same species? Or what is the source that makes humans different from other mammals? Ultimately, all these differences and similarities at the level of tissue, population, and species can be traced down to genomes, and are therefore heritable. For this reason, scientists have searched for the root of this heritability in past centuries.

One of the first recorded thoughts on inheritance belongs to Aristotle. He suggested that semen brings the active element, life, to the progeny and female provides the nutrition. Before the 19th century, scientists viewed the similarity between offspring and parents more as a blending of characters and the environment was part of it. For instance, it was common sense that environmental effects on the parents could reappear in the descendants. This theory was called the inheritance of acquired characters, or Lamarckism. It argued that organs and functions would improve or eliminated upon use or disuse (Lamarck. 1809). Charles Darwin needed some ground for an inheritance to explain the origin of species; he proposed the gemmules, hypothetical particles that all the cells of the body produced and which the environment could affect. Then, modified gemmules would assemble in the reproductive organs giving rise to a slightly modified new generation. He called his theory pangenesis (Darwin. 1859). Francis Galton, Darwin's half-cousin, set out to prove the pangenesis theory. He transfused the blood between dissimilar rabbits and screened their offsprings. Much to his surprise, he observed that none of the examined characters of the blood donors transmitting to the offspring (Galton, 1871).

Darwin responded in a letter clarifying his thesis (Darwin. 1871. response to Galton's experiment):

"Now, in the chapter on Pangenesis in my "Variation of Animals and Plants under Domestication," I have not said one word about the blood, or about any fluid proper to any circulating system. It is, indeed, obvious that the presence of gemmules in the blood can form no necessary part of my hypothesis; for I refer in illustration of it to the lowest animals, such as the Protozoa, which do not possess blood or any vessels; and I refer to plants in which the fluid, when present in the vessels, cannot be considered as true blood. The fundamental laws of growth, reproduction, inheritance, etc, are so closely similar throughout the whole organic kingdom, that the means by which the gemmules (assuming for the moment their existence) are diffused through the body, would probably be the same in all beings; therefore the means can hardly be diffusion through the blood."

After Galton's experiment, the idea of pangenesis received lots of criticism, nonetheless.

Weismann barrier

Decades after the pangenesis theory, August Weismann postulated the germ-plasm theory, by which only changes in the reproductive cells could go to the next generation, and that parents contribute to the traits by their distinct share of inheritance via the narrow street of the germ-plasm (Weismann, 1892). In his theory, the body (soma) could not influence the germ-plasm, in line with Galton's observation that blood of other rabbits could not influence the characters of the rabbit's progeny. In contrast, the germ-plasm could build the body, hence a one-way relationship. Weismann even tried to rule out the inheritance of acquired traits. He cut the tail of white mice for five generations, and he observed no changes in the tail length (Weismann, 1888). Public perceived it as a blow to the inheritance of acquired characters. However, Lamarck's main point was use and disuse of organs that could gradually be inherited, not violent mutilation of an organ[1].

After observing parental pairing of the chromosomes before meiosis —a cell division to produce gametes by reducing the number of chromosomes in half— and discovering Mendel's text after decades, scientists connected the chromosomes to inheritance of characters (Sutton. 1903). Soon after, chromosomes were under scrutiny. Morgan and his student Sturtevant by crossing flies and recording their phenotype observed that some phenotypes were linked (e.g. body colour and eye colour). Thus, they proposed that the linked traits must be on the same chromosomes. With the knowledge that crossing-over occurs probabilistically between the paired chromosomes, Sturtevant used the observed frequency of flies whose linked phenotypes were segregated to infer the distance between the location of the studied traits on a chromosome (Morgan. 1915). However, how the chromosome could influence the phenotype remained a mystery.

The molecule of inheritance

Although the chromosome theory does not sound outlandish now, it raised controversy back then (Grifiths et al. 2000). Scientist knew that chromosomes protein and thymo-nucleic acid (Miescher, 1874) are the components of chromosomes, and they initially suspected that proteins were responsible for the inheritance (Gulick 1941, Darlington. 1942, Cooper. 2000).

Following Mr Griffith's observation in 1920s' that non-pathogenic bacteria could take the form and phenotype of pathogenic bacteria in their vicinity from their dead cell lysates and, more importantly, that this acquired phenotype remained in the daughter cells[2], several scientists tested which component of cell lysate conferred the bacteria with this acquired hereditary phenotype. Hershey and Chase exposed bacteria to purified molecules of different classes, such as proteins, lipids, RNA, DNA, and carbohydrates from lysed bacteria, finding that only DNA could transform the bacteria from non-pathogenic to pathogenic[3]. Still, how DNA could be inherited or how it could affect the phenotype remained unclear, until Mr Crick and his colleagues, thanks to

data from Ms Franklin, modelled the structure of DNA, as a complementary double-strand helix. The sequence of nucleic acid in each strand of DNA would thus serve as a template for the synthesis of another strand[4]. Moreover, portions of it could be transcribed and, then, in the ribosomes —huge ribonucleoprotein machinery— translated into proteins. These proteins and, to some extent, RNA carry out most of the biochemical functions inside a cell. He called this specification of information from DNA to protein a one-way path called the central dogma of molecular biology[5]. However, understanding the general role of each molecular player does not explain how an organism, particularly complex multicellular ones, is assembled to give rise to life.

Multicellular organisms

Unicellular organisms divide and make copies of their system that acts and reacts more or less in the same manner as the original cell. Multicellular organisms, on the other hand, have a different strategy to survive Time. They produce a cell similar to an egg, the zygote, via sexual reproduction, that gives rise to a similar organism, and this process repeated billions of times throughout history. Each zygote, via cell division, develops different organs and cell-types that in turn produce the gametes, and these the zygote again.

Bacteria that diverged millions of years ago in the same environment could have major differences in behaviour, metabolism, morphology, and so forth. However, genetic differences that accumulate in their genomes throughout evolution, generating different protein sets could cause these differences. Does this process explains the differences we observe between tissues? A surge of cloning experiments in frogs during the mid 20th century exposed some of the mysteries of embryo development. King and Briggs devised a method to transplant the nuclei from different cells inside an egg[6]. Mr Gurdon took the nucleus from a differentiated cell and placed it into the nucleus-less unfertilized egg. The new patched-egg developed into a healthy frog, indicating that information necessary for the frog development resides in the nucleus of a differentiated cell[7,8]. However, trying to put the same nucleus in another differentiated cell would not lead to the development of anything, let alone a frog. This investigation suggest that while the information to form a frog was still intact in the nuclear DNA, it needed some factors that reside in ovum to trigger the developmental program. Decades of developmental biology have now shown that the genome is a blueprint of the plan that embryo development has to follow. Multiple genes orchestrate this plan. They are generally known as developmental regulators and are behind the generation of the unique proteomes that account for each cell and tissue type characteristics.

1.1 Transcription across biological units

Early research on the protein level differences between cell types, done by two-dimensional polyacrylamide gels, confirmed that different cells produce at least different concentrations of protein sets or distinct types of proteins, consistent with the idea that proteins ultimately specify the function and morphology of the cells (Albert et al. 2008). However, multiple studies have shown that these differences can be anticipated, by looking at the transcriptomes, the whole set of mRNAs of a cell (or group of cells). For instance, several studies have shown that the tissue of origin of cancer cells could be easily distinguishable by measuring gene expression (Albert et al. 2008).

DNA could regulate its expression at many levels such as transcription, RNA processing and editing, RNA localisation and degradation. Cells differ at many of those levels, and it is still not clear how each step contributes to the functional diversity of the cells. However, it is more economical for a cell to regulate itself at this first phenotypic layer, so that it does not synthesise superfluous molecules.

However, it is fair to ask: is the transcriptome a direct reflection of the proteome? Proteins have a higher dynamic range than the respective transcripts, and the low correlation between RNA level and protein concentration implied the involvement of other mechanisms in regulating the translation of RNAs[9–11]. Nonetheless, Many studies measured the transcription level differences to estimate the protein abundance[12];[13]; Thul et al. 2017). While it might seem more appropriate to study the proteins directly to infer the metabolic and functional differences between biological conditions, protein quantification has its caveats and difficulties[14]. For this reason, studies so far scratched the surface of the protein differences by quantifying mRNAs.

Comparing gene expression also has its caveats. Firstly, it is not clear which level of gene expression corresponds to a change in the stoichiometry of the proteins. A study comparing mouse and human showed that protein levels are more conserved, suggesting that protein expression is well-buffered. Accounting for buffering between RNA expression and proteins had a better predictive power in predicting protein concentration based on RNA measurements[15]. Secondly, an interesting study revealed that bacteria controlled its response to stress at the level of translation[16]. This observation indicates that considering only the regulation of transcription as the most cost-effective response would not be sufficient to predict cell behaviour.

Many studies compared gene expression levels between tissues[17]. Some cell types are unique to a given tissue, such as neurons for the brain, hepatocytes for liver, and

spermatids for testis; thus, we expect the corresponding tissues to have very different expression profiles. However, by analysing massive amounts of proteomic data across tissues and cell lines, the Uhlen team showed that the number of proteins that are uniquely expressed in one or few distinct cell types is below 2%, supporting the idea that the functions and phenotypes of each cell type is controlled more by changing the stoichiometry than by switching genes entirely on or off[18].

Besides, Kuster and his team surveyed the landscape of gene expression across tissues at both RNA level and protein level[19,20]. Interestingly, they found that, on average, only less than one per cent of the transcripts and proteins exhibited a tissue-enriched profile (0.73% of transcripts compared to 0.65% of proteins). Nevertheless, they found brain and testis to contain the highest number of tissue-enriched transcripts, consistent with previous studies[13]. It is still a mystery as to why testis and brain had the highest number of tissue-enriched transcripts. Perhaps, it indicates their high level of specialisation.

Despite these general patterns, however, it is well known that certain cell types need highly specific proteins to perform their action. Although they are the exception, they have strong functional relevance, and they are often conserved.

1.2 The substrate of alternative splicing

Importantly, transcriptional control is not the only tool the cell has to modify its transcriptome and proteome. Other mechanisms, referred to as post-transcriptional regulatory mechanisms, can also have a major impact. Furthermore, among those, alternative splicing —the differential processing of introns and exons in our genes to generate multiple mRNAs from each gene— is arguably the most crucial contribution to transcriptome diversification in mammals. However, what is the substrate in our genes for alternative splicing to occur? The first lead came from Berget's experiment. She and colleagues took the mRNA of adenovirus and hybridised it to the virus' DNA[21]. They observed that not all the DNA sequence of the gene hybridised with the mRNA, although all the mRNA hybridised with the DNA sequence. This observation meant mRNA lost part of its transcribed sequences from the original copy (i.e. introns) before the translation. Only a year after that discovery, Berk and Sharp[22] found that a given viral DNA could produce two different mRNA by choosing two different sites in the pre-mRNA to remove the intronic sequences. Those two mRNAs produced two different proteins, in a process called alternative splicing. Was it an exceptional or typical process? How many of the human genes could undergo alternative splicing to create different proteins?

Modrek et al. after analysing expression sequence tags (EST) —a method to collect the mRNA from a cell, clone it and perform Sanger shotgun sequencing — collected from human samples, estimated that 42% of human genes were alternatively

spliced[23]. This estimation soared up to nearly 100% of multiexonic genes using RNA-seq data[24]. While these results could mean that organisms had a larger protein repertoire than expected by their number of genes, studies needed to scrutinise their role in creating diverse biological functions. In one of the first studies of the functionality of alternative splicing, Early et al. found that one gene produced two different mRNAs, one membrane-bound and another secreted IgM[25]. Since then, a myriad of studies conducted on the role of both gene transcription and splicing in biological functions[26,27].
While alternative splicing could be a very complicated process such as selecting from multiple mutually exclusive exons (e.g. Dscam gene in fruit fly[28]), we can summarise it into major 4 types: Alternative 3' and 5' splice-site usage, exon skipping (also known as cassette exons), and intron retention(fig). Some researchers use additional subtypes, such as mutually exclusive exons, but in turn, these are a combination of two anti-correlated exon inclusion events. We also do not include alternative promoters and alternative poly-A sites as alternative splicing, since they are not in a first instance choices of the spliceosome (they are related to alternative transcription or choice of alternative termination sites of transcription, respectively).

In terms of their molecular impact on the proteomes, alternative splicing, besides producing different protein isoforms, could insert a disruptive sequence of RNA containing a stop codon in the middle of a transcript or alter the canonical open reading frame. Those transcripts do not result in functionally stable proteins in most cases and are typically degraded (at least in part) by the mRNA police, a piece of machinery called non-sense mediated decay (NMD), thus reducing the affected transcript concentration. However, ORF-disrupting splicing events can be used for downregulation of the final protein levels, thus contributing to regulate gene expression. For instance, NMD-causing splicing events auto-regulate the expression of many splicing factors [29].

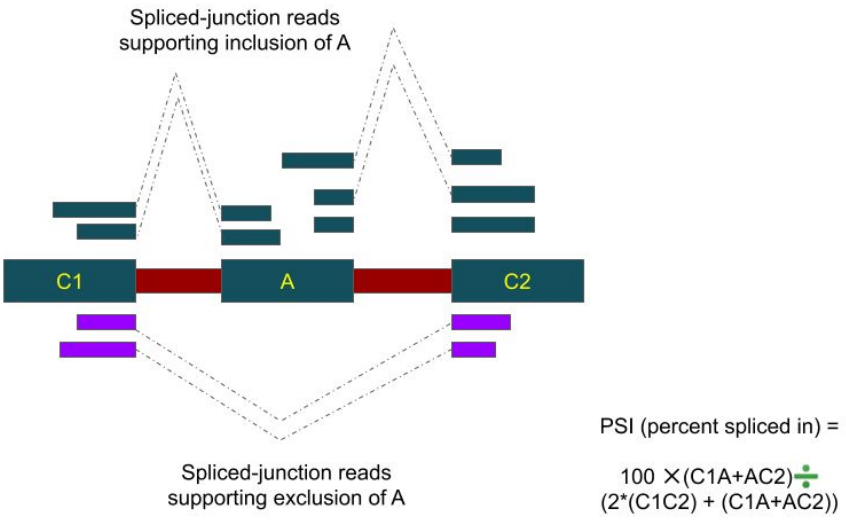


Figure I1. PSI calculation for exon skipping events

1.3 Regulation of alternative splicing across tissues

Are pre-mRNAs spliced differently across tissues? Lee and his group used EST data to first investigate this question in a comprehensive manner [30]. They found that 10-30% of human spliced genes exhibited tissue-specificity. By dividing the number of tissue-specific alternative splicing events of a given tissue by the total number of ESTs in the same tissue, they measured the enrichment ratio of tissue-specific alternative splicing. They found skin, retina and muscle to have the highest tissue-specific alternative splicing enrichment. However, in total, brain and testis expressed genes underwent the highest number of distinct alternative splicing events [30]. In a subsequent study, controlling for differences in the amount of EST for a given tissue, Burge and his group found brain and testis to have the highest number of genes with exon skipping events. The genes in the liver exhibited a high rate of alternative 3' and 5' splice-sites. They attributed liver's unusual pattern of splicing to the expression of serine-arginine and heterogeneous ribonucleoprotein genes based on microarray expression data [31]. More recently, using RNA-seq data, we surveyed the pattern of alternative splicing across tissues of human, mouse, and chicken. They found neural tissues, testis, and muscle to have the highest number of tissue-specific up-regulated exons in all the compared animals. Thus, such conservation could suggest a pivotal role for regulation of exon splicing in brain and testis[32], see Appendix).

Round spermatids' gene expression could be responsible for the unique gene expression and alternative splicing that was observed in testis because most of the RNA-seq reads of testis originates from round spermatids [33]. Therefore, it is likely

that the regulation of alternative splicing plays a role in sperm production or function. Similarly, alternative splicing might play a role during neurogenesis. SRRM4, a brain-specific splicing factor regulates alternative splicing of very short exons (microexons) that modulate protein-protein interactions that are necessary for neurogenesis. Moreover, disruption of microexon regulation was frequently observed in the brain of autistic individuals [34]. Furthermore, Rbfox and Ptbp1 antagonistically regulate splicing of microexons and other neural exons and consequently their roles are essential for differentiation of neural progenitor cells to neurons [35–37]. Thus, It seems that neurons need a different set of exons for neurogenesis.

Although a high percentage of genes have specific alternative exons in testis and brain, researchers found yet another tissue which topped the number of alternative splicing events. T-cells learn to recognise the body's antigens from non-self in the thymus. Epithelial thymic cells, with an extraordinary capacity to produce a large fraction of body antigen, present the body's proteins to the T-cell to build immunobiological self-tolerance. As expected, many studies found a high level of gene expression promiscuity in the thymic cells [38–40]. Similarly, Abramson's group found a high ratio of alternative splicing events in those cells compared to all the other screened tissues, and hence regulating the alternative splicing for a very different reason than cell differentiation. Interestingly, testis and brain shared the lowest number of alternative splicing events with the thymic cells ([41].

1.4 Individual and population transcriptomics

Individuals, similar to tissues, are biological spaces. However, unlike tissues of a body, individuals do not serve a bigger purpose, for example, at the species level [42,43]. Comparing cells from individuals across European and African ancestry highlighted a higher number of genes that vary across individuals within a population compared to those that vary between the two populations. Storey et al. found only 50 expressed genes that varied significantly between African and European populations (FDR < 20%), whereas 1,210 genes varied within populations (out of 5,194 gene probes expressed in lymphoid cell lines) [44]. In another study, genes with variable expression within-population found to be associated with genes that affect susceptibility to disease, particularly HIV, whereas the genes with low within-population variability were involved in fundamental biological processes such as translation and ribosome production [45]. The differences of gene expression across individuals and population, however, were small in magnitude. Besides, the heritability —a measure of genetic contribution to the variability of a phenotype, e.g. gene expression— of most genes that varied within a population was low, suggesting that environmental factors could be involved in transcriptional variation. On top of that, heritability estimates could not distinguish the familial genetic factors versus familial shared environment [46].

However, more recently, with availability of genotypic data corresponding to the RNAs samples, biologists could start associating varying RNA expression to polymorphic sites in the genomes (expression trait loci (eQTLs)). Using the genomic data, scientists could more confidently separate the heritable part of the between-individual transcriptional variation. Comparing eQTLs with genome-wide association studies (GWAS) showed that eQTLs were highly enriched in the risk loci of complex diseases[47,48], suggesting that inheritable variation in gene expression, although minor in magnitude, could affect predisposition to diseases. Studies on the mechanisms through which eQTL could influence gene expression showed that most of the eQTLs are caused by single nucleotide polymorphisms (SNPs) in the transcription factor binding sites of enhancers and promoters [49].

Interestingly, it seems that gene expression varies less among individuals compared to alternative splicing [50], indicating that alternative splicing could potentially contribute more to individual diversity and disease than gene expression. Pritchard and his team's research highlighted an equal role of sQTLs —polymorphic genomic sites that affect splicing patterns rather than gene expression— in contributing to complex traits as eQTLs. Instead of measuring the expression of an individual exon or transcript isoform, they measured the reads spanning the splice-site junctions to estimate the amount of splicing. The majority of the sQTLs variants were found to be near the splice-site regions. Remarkably, comparing splicing and expression QTLs with GWAS studies showed a more prominent role of sQTLs on complex diseases such as schizophrenia and multiple-sclerosis compared to eQTLs [51]; [51–54].

Recently, a group at Illumina applied a different method to discover the variants that could influence alternative splicing regulation [55]. Instead of associating SNPs to changes in alternative splicing profiles, they used all pre-mRNA genomic sequences as input and compared them to the GENCODE spliced transcripts as a training set to develop a deep learning model. Building that model, they could then predict cryptic splicing mutations in some individuals, which resulted in aberrant splicing with 75% rate of validation. They estimated 10% of cryptic splice-sites to be pathogenic variants in neurodevelopmental disorders, particularly autism. Besides, they found that losing a splice-site was more common than gaining it across individuals. More importantly, they catalogued many new regions in the genome that could influence splicing regulation. They could explain some of the regulated splicing events whose splicing motifs were degenerate by the exon to intron length or nucleosome positioning. One of the highlights of this study was that weak cryptic variants across the population could influence the profile of alternative splicing in a tissue-specific manner, indicating an unexpected role of cryptic variants in generating tissue-specific alternative splicing. Those cryptic variants could be favoured or get fixed in the population by drift, leading to the evolution of tissue-specific alternative splicing [55].

1.5 Evolution of gene expression and alternative splicing

In 1968 Mr Kimura proposed the neutral theory of evolution as the null model of DNA changes between organisms. This theory postulated that the majority of genomic sequence changes between species do not affect the phenotype of an organism [56]. He categorised nucleotides substitutions into two groups: those under great negative selection that seldom change, and neutral changes that get fixed by neutral drift. Besides, the theory predicts that substitutions accumulated with time and not phenotypic divergence. Does neutral model predict the pattern of gene expression and alternative splicing evolution as well?

Negative selection, an integral part of the neutral model, has been shown to have a strong influence on the within and between species gene expression evolution in the fruit flies [57–59]. Also, the gene expression distance between the flies increased with the number of genetic changes [57]. Similarly, in yeast, the majority of gene expression divergence was proportional to the genetic distance, and only a few genes in this study showed a correlation with phenotypic adaptations, suggesting a small contribution of positive selection on gene expression evolution [60]. Studies on natural populations of fish showed that phylogenetic distance could explain 18% of the gene expression changes, whereas positive selection explained 4% of gene expression variability across populations [61]. Those observations meant that the neutral model of evolution could predict gene expression variations between species more accurately than positive selection.

The neutral theory equips researchers with a tool to test a hypothesis based on a deviation observed from its prediction. Between-species variation over within-species divergence of gene expression could be an estimate to detect directional and stabilising selection or relaxation of selection at the level of transcription [62]. For example, low variations within species and between species indicate a scenario of stabilising selection; lineage-specific shifts in expression of the genes that have high within-species variation suggest a relaxation of negative selection; lineage-specific shifts of gene expression of lowly variable genes within the population, suggest positive selection [63,64]. However, this view is a bit simplistic. Environmental factors such as temperature, circadian clocks, season, and diet affect many transcriptomic events. Thus, highly controlled experiments are necessary to decode the mode of gene expression and alternative splicing evolution[65]. Comparing within and between-species variability of gene expression across primates' tissues showed that testis had the most substantial divergence of gene expression between species compared to brain and liver. This observation suggested that the testis' gene expression profile has evolved faster compared to other tissues [66].

Gene expression patterns across tissues have shown to be highly conserved compared to alternative splicing. Some studies compared gene expression and alternative splicing profile of orthologous genes and exons across tissues of vertebrate species[67,68]. They showed that alternative splicing could contribute to species divergence more than gene expression because alternative splicing varied more across species. However, those differences could also correspond to a more relaxed negative selection (i.e. more noise) in the case of alternative splicing pattern. Therefore, whether those alternative splicing changes across species denoted a neutral evolution or not is debatable.

Khaitovich and his team tried to profile the mode of evolution for exon splicing and found that the amount of splicing divergence correlates with the amount of time that species diverged from each other. Thus, the neutral theory might be an appropriate null model for the evolution of alternative splicing across species as well. Moreover, they found that most of the human-specific exons changed from constitutive in mammals to alternative in human [69]. In a null model of evolution in which most of the mutations are either neutral or weeded out by selective pressure, it thus seems that turning an alternative exon into constitutive will be selected against or is less likely to occur in the first place, due to mutational biases. Studying within-population differences could help us disentangle the two scenarios, mutation or selection. For instance, if in a population of individuals the change from constitutive to alternative exons were more common than increasing inclusion of an alternative exon, then it is most likely that it is the probability of mutation is more shifted to favour constitutive to alternative rather than a differential selection over those changes. This is particularly interesting, since changing from constitutive to alternative implies the creation of two different transcripts, thus expanding transcriptomic complexity, while the opposite trend converges two transcripts into one, simplifying the transcriptome. We will explore these questions in the Results section.

However, a neutral model does not mean that none of the species-specific alternative splicing events are non-adaptive. For example, alternative splicing conferred the species of vampire bats with infrared sensation in a tissue-specific manner [26], and several other examples have been identified, which likely represent only the tip of the iceberg of the contribution of alternative splicing to the evolutionary origin of novel traits.

1.6 Objectives

Many studies primarily focused on variability and quantitative differences of transcriptomes. Not to reinvent the wheel, in my PhD I set out to look for qualitative

differences: features that make the transcriptomes of a given tissue, individual or species unique.

Sergent and Dawid did one of the first studies on the qualitative differences at the transcriptome level by mixing the mRNA samples from two different developmental stages —blastula and gastrula— followed by the extraction of the mRNA from gastrula that did not hybridise with the copy of RNA from the blastula tissue. Therefore, the sequences were qualitatively unique to gastrula compared to blastula [70]. We took conceptually similar approaches to identify tissue, individual, and species-unique alternative splicing and gene expression events. Conceptually, if we could homogenise an entire human, take his/her mRNA and hybridised it to a pool of homogenised people in the population, what would be the mRNAs that would be unique to the tested individual? What if we do this with a homogenized brain and hybridize it against the rest of the human body? This experiment is the one we simulate with our RNA-seq data analyses in the first part of this PhD (Chapter 2). We quantified RNA output —alternative splicing and gene expression— of protein-coding genes in 8,378 postmortem samples derived from 48 tissues and 505 individuals to identify the mRNA species that are exclusively present or absent in the following three biological spaces: i) a tissue compared to the other tissue types, ii) one individual or a group of individuals —up to twenty percent of the studied individuals— compared to the rest of individuals, and iii) human as a species compared to other mammals.

2 Chapter 2 - Introduction to the transcriptome of time and its entourage

"Life is a cyclical chemical process that is regulated in four dimensions. We distinguish parts of the cycle: development describes the changes from a single cell to adult, and aging the changes from adult to death. Birth to death, a cycle, and there are cycles within cycles—circannual rhythms, menstrual cycles, semilunar cycles, and daily 24 hr or circadian cycles."[71]

Any space coordinate needs time in the real world. Similarly, a multicellular organism can regulate the timing of genomic output across its biological space. Internal timing is a crucial component in the cell cycle, development [72], ageing and maturation [73]. Time, also, could affect the organisms externally because of the circular movement of the earth around the sun and itself. Thus, the circadian and seasonal cycle evolved correspondingly. In the second part of the thesis, I focus on the circadian and circannual rhythms as the two main factors producing cyclical timing imposed on the organisms living on earth.

The circadian clock of the living system resonates every 24 hours for the organism to adapt and respond optimally to daily environmental cycles. Whether circadian regulation evolved as an adaption to more efficiently capture the energy of the environment and/or to avoid the direct consequences of solar radiation and/or something else is still unclear [74–78]. Although diurnal and nocturnal animals invert the periods of activity and rest, the presence of sleep and wake cycle is conserved among most animals. Moreover, circadian cycle influences hormones concentration in the blood, fasting and diet, core body temperature, cell regeneration, and photoperiodism in mammals [79,80].

Evidence for involvement of genetic factors in circadian rhythm came from fruit fly. Konopa and Benzer mutated many fruit flies, then screened them in search for deviation from wild type phenotypes. They detected flies with disrupted sleep and mutated PER gene. Following more genetic screens in fruit flies, researchers found more genes to be involved in the circadian cycle, and an autoregulatory negative and positive feedback loop at the level of transcription and translation emerged [81,82]: transcription factors ARNTL and CLOCK initiate the transcription of inhibitors (CRY and PER) of ARNTL itself and hence form a negative feedback loop. Takahashi's lab observed that the same transcriptional regulations existed in the mouse [83].

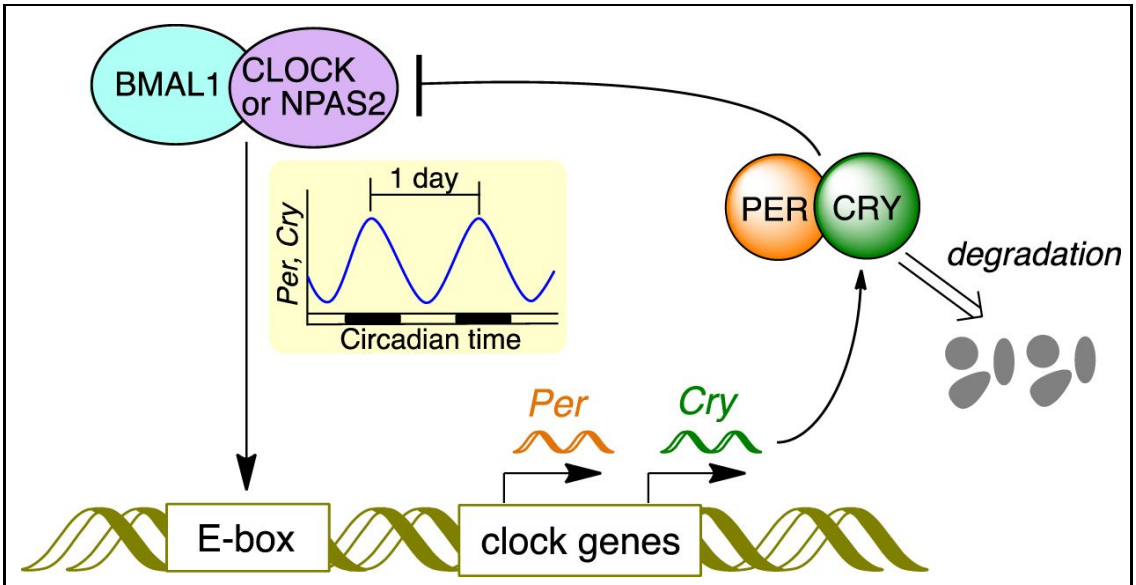


Figure I2. Transcriptional feedback loop of core clock genes [84].

An impressive result came out of studying humans. Red blood cells. Red blood cells do not possess a nucleus; thus, their daily rhythm must be independent of transcriptional regulation. Peroxiredoxins, highly conserved antioxidant proteins, oxidised in 24 hours cycles. This oscillation remained constant after days in the absence of external clues and was tunable by environmental stimuli and temperature. Observing this Peroxiredoxins cycle in prokaryotes, eukaryotes and archaea suggested an ancient origin and a pivotal role in the survival of life as a whole on earth [85,86]. Moreover, the oscillation in human red blood cells highlighted the interplay of transcriptional and non-transcriptional components in creating the circadian waves across cells of the body.

The circadian master synchronizer in mammals resides in the suprachiasmatic nucleus (SCN) of the hypothalamus. Removing it showed that animals had an irregular sleep-wake cycle. However, most tissues were found to exhibit circadian gene expression regulation when isolated from SCN [87]. This region receives input directly from ganglion cells of the retina, and send the output to the pineal gland. The pineal gland produces melatonin whose concentration peaks at night. It is still not clear how SCN synchronises different tissues' circadian rhythm. Later in 2006, Debruyne et al did not find any phenotype or substantial alteration of core circadian genes oscillation in the SCN of CLOCK-deficient mice. Thus, rendering CLOCK gene dispensable in regulating mammalian circadian behaviour[88]. Redundancy, also, could explain the observed phenotype; for example, ARNTL could dimerise with another gene (see results).

2.1 Circadian rhythm gene expression across tissues

In one of the first transcriptomic studies in mammalian tissues, Zhang et al. discovered that 43% of genes changes with circadian rhythm at least in one tissue out of 12 tested tissues [89]. The circadian regulation at the transcription level had a tissue-specific profile in mouse. Expression of protein-coding genes in the liver, kidney, and lung was the most affected by the time of the day, whereas that of skeletal muscle was the least modulated by the circadian rythm. Overall, 5-20% of the genes expressed in any given tissue are affected by daily oscillations. Protein-coding genes were affected by daily rhythms more than non-coding RNA, and a higher proportion of conserved non-coding RNA compared to non-conserved RNA were affected by daily rhythms.

Nonetheless, studying mouse to understand human biology has its downsides: mouse is a nocturnal animal, so the phase of its genes should be reversed to humans [90]. Besides, the lab mice have a disrupted sleep, usually fed ad libitum, and they do not produce melatonin during the sleep cycle. For these reasons, other researchers have used baboon, *Papio anubis*, as a diurnal primate model to study circadian patterns of transcription. They detected a 12 hr difference between the peak-phase of the core clock gene expression of baboon compared to a mouse, reflecting the nocturnality of mice. More importantly, the researchers found that 81% of the protein-coding genes and about 65% of the expressed genes were affected by circadian rhythms in at least one tissue. Limiting the gene set to highly expressed loci showed that only 382 out of 10,602 highly ubiquitous expressed genes had no rhythmicity across all the tested tissues. This 81 percent number was two times more than what Mr Zhang found in mouse mainly because of the higher number of screened tissues. However, both studies indicated that a high proportion of protein-coding genes were rhythmic.

Hogenesch's group studied human postmortem tissues from GTEx project to detect oscillating genes. Similarly, from the 13 tissues with a robust circadian signal of core clock genes, 44% of the 16,906 tested genes showed rhythmicity in at least one tissue [91]. However, in this study, they filtered out the tissues in which the core clock genes did not show the expected correlation or anti-correlation amidst each other. For example, we expect that ARNTL to be antiphase with NR1D1, but not in all the tissues the researcher observed the anti-correlation. Nevertheless, even in the baboon and mouse experiment that was well-designed to study circadian rhythms, not all the core clock genes oscillated in all tissues, suggesting that tissues could trigger the gene transcription oscillation by using part of the circadian core clock transcriptional feedback. Moreover, the circadian rhythmicity in all the studies above could come from other environmental factors beside dark-light cycle such as food uptake, temperature changes or sleeping.

2.2 Circadian regulation of alternative splicing in mammals

Post-transcriptional regulation in the context of the circadian clock has remained less explored compared to transcriptional regulation. Heyd and his colleagues revealed how body temperature rhythms could change the phosphorylation status of a splicing factor named U2AF26, that led to a change in the inclusion profile of mammalian exons [92]. Ule's group also showed that the inclusion of 55 exons in the liver cycles every 24 hours. The feeding cycle induced the inclusion of some these exons [93]. Another study found circadian-regulated alternative splicing to be involved in tumour metastasis [94].

2.3 Biology of seasonal (circannual) changes

The duration of the pineal production of melatonin is inversely related to day length (photoperiodism), and its secretion drives long-term changes in many physiological systems, including the hypothalamus-pituitary-adrenal, hypothalamus-pituitary-gonadal, and brain-gut axes, the autonomic nervous system, and the immune system[95].

Observing that some circadian genes also oscillate with the seasonal period could imply a role of circadian genes in determining the photoperiod and hence the seasonality. For example, Todd and co-workers showed that ARNTL expression peaked in white blood cells during summer; this could relate the circadian regulation to seasonal regulation[96].

Besides reproduction, animals exhibit a range of plasticity of behaviour and physiology during different seasons; for instance, altering the coat colour in polar animals, hibernation, and changing of diet. In human, many complex diseases exhibit a seasonal pattern such as cardiovascular disease [97–99], and various autoimmune disorders [100,101]. Human blood exhibited a seasonal variation of gene expression related to the immune system [96]. Infectious disease's seasonality is a well-supported case and causes a wave of reactions in the immune response genes [102,103]. Furthermore, gut microbiota populations exhibited seasonal variation. Given the increasing report on impact of microbiota, it is likely that seasonal variation influences the physiology via various biological pathways[104,105]. Psychiatric diseases such as Alzheimer's, schizophrenia, and bipolar disorders exhibited seasonal rhythmicity[106–109]. Furthermore, seasonal regulation of frontal cortex gene expression were arhythmic in the brain from Alzheimer samples [110]. Seasonal

affective disorder (SAD) is a famous example of the psychological influence of circannual rhythm.
A study in humans demonstrated that individuals with higher predisposition to SAD up-regulated the expression of the serotonin transporter in the brain, whereas non-affected individuals decreased its expression in winter [111]. Furthermore, Carlsson et al. recorded that 5-HT(serotonin) is at its maximum level in October and November and lowest in spring only in the hypothalamus[112], data from 56 human postmortem brain samples of people without neurological diseases and approximately balanced according to age and gender).

Even though responses to seasonal cycle play a pivotal role in organism adaptation and survival, only a few studies investigated the transcriptional and post-transcriptional variation across human tissues. With the ominous prediction of climate changes, studying the seasonal influence on human biology may seem even more relevant.

2.4 Objectives

Studies on circadian and seasonal rhythm in human tissues are scarce due to the difficulty of finding samples. Furthermore, recording seasonal transcriptome requires a considerable amount of investment in time. Thus, I found this opportunity golden to explore this angle of human biology.

First, I would explore whether it is possible from the time of death recorded from many different individuals to extract the circadian gene expression.
Second, do tissues autonomously show circadian oscillation? Are these circadian changes tissue-specific?
Third, I will compare the circadian transcriptome of human against a nocturnal animal and a diurnal animal.
Fourth, we add a potential list of genes with robust oscillaton across a high number of tissues.
Fifth, how circadian regulation could influence seasonal regulation, the effect of photoperiod.
Sixth, explore how alternative splicing changes diurnally and seasonally.
Seventh, as the role of circadian and seasonal variation in diseases, has been reported, I will explore whether disruption of the seasonal and circadian events could play a role in diseases predisposition.

3 Chapter 3 - Landscapes of unique human transcriptomes

3.1 R1 - Landscapes of unique human transcriptomes

Sodaei R, Permanyer J, Guigo R, Irimia M

3.2 Abstract

Myriad of studies tried to detect transcriptional and post-transcriptional regulation across different tissues, individuals and species. Most of those studies focused on quantitative differences. Here, using GTEx data, we set out to identify qualitative differences across biological units: transcriptomic elements (genes, exons, introns) that make tissues, individuals or the human species molecularly unique. Overall, we found that alternative splicing (AS) has a larger contribution to the generation of strictly unique transcriptomes than the gene regulation of steady-state levels (GE). With respect to tissues, whereas GE barely generates testis-unique transcripts, AS contributes to the brain, skeletal muscle and testis-unique transcriptomes. However, while neural unique splicing is enriched for microexons and highly conserved, testis events usually disrupt the reading frames and are not conserved. In the case of unique individual and species transcriptomes, AS had a much larger contribution than GE. Around 81% of individuals had at least one exon individual-unique exon, which in many cases could be associated with specific genomic variants. In both individual and species unique exons, we observed that most unique variants correspond to changes from constitutive to alternatively spliced patterns. Overall, our results provide a global assessment of the contribution of gene expression and AS to generate unique transcriptomes in time and space.

3.3 Introduction

Why neuronal tissues function differently from muscle tissues despite having a virtually identical genetic makeup [70], why individuals are different from each other (Francis Galton. 1889), or what makes humans and other mammals different from each other (Darwin C. 1871), were the questions many studies have tackled since long time ago.

Differences between tissues, individuals, and species are primarily emerging from different transcriptomes —either by regulation of the same genome or different genome or by differences in the genome, but all translate first into transcriptomic differences. One way to create different transcripts, as has been known, is to regulate the gene expression. This line of thought is embedded in the design of the GTEx project, an endeavour to assess the gene expression regulatory differences across 48 postmortem tissues in a large group of healthy individuals. However, besides gene expression(GE) regulation, other mechanisms could create different transcripts among which alternative splicing (AS) is the most prevalent in mammals.

Much work has been done to investigate the contribution of GE and AS to tissue, individuals, or species variation. At the tissue level, More recently studies showed that a significant fraction of human alternative splicing events and genes were tissue regulated, especially in the brain, muscles, and the testis [24,32,34].

At the population level, the alternative splicing was shown to be more variable between individuals than gene expression implying a more prominent role of alternative splicing in contributing more transcript variability at the population level [50]. Many of the inter-individual splicing differences were shown to be a consequence of the changes in the genomic splicing regulatory regions and their potential role in different diseases was discussed [51,52]

The alternative splicing and gene expression changes across population would provide a source of variability that eventually, upon fixation in the population, would add to the species differences. At the species level, studies on transcriptomic data of different mammalian species demonstrated that clustering of alternative splicing events based on inclusion levels (“per cent spliced in; PSI) is largely species-dominated, unlike gene expression, which has a tissue-dominated signature [67,68]. Those observations implied that exon splicing was more often affected by lineage-specific changes than gene expression and that alternative splicing potentially confers more sources of variability between species than gene expression. Interestingly, this variability may generate phenotypic novelties. For instance, an alternatively spliced exon in TRPV1 gene is involved in infrared detection in the vampire bat [26].

However, those studies primarily focused on variability and quantitative differences among transcriptomes. Here, we set out to look for qualitative differences: features that make the transcriptomes of a given tissue, individual or species “unique”. One of the first studies on the qualitative differences at the transcriptome level was conducted by mixing the mRNA samples from two different developmental stages —blastula and gastrula— followed by the extraction of the mRNA from gastrula that did not hybridize

with the copy of RNA from the blastula tissue, therefore, the sequences were unique to gastrula compared to blastula [70].

We took similar approaches to identify tissue-unique, individual-unique, and species-unique alternative splicing and gene expression events from RNA-seq data. We quantified RNA output —alternative splicing and gene expression— of protein-coding genes in 8378 postmortem samples derived from 48 tissues to identify the mRNA species that exclusively present or absent in the following three conditions; i) a tissue compared to other tissue types, ii) one individual or a group of individuals —up to twenty percent of the studied individuals— compared to the rest of 505 other individuals, and iii) human as a species compared to other mammals.

3.4 Results

3.5 Tissue-unique splicing events impact brain, muscle and testis

We defined alternatively spliced events or genes with tissue-unique regulation as follows (Figure R1.1A; see Methods for details):

I) Uniquely included splicing events: (i) the lower quartile of the PSI distribution of the target tissue (e.g. the brain) has to be > 10, whereas, for each of the other tissues, (ii) the lower quartiles of the PSI distributions have to be 0 and (iii) the upper quartiles < 10.

II) Uniquely excluded splicing events: (ii) the upper quartile of the PSI distribution of the target tissues has to be < 90, whereas, for each of the other tissues, (ii) the upper quartiles must be 0, and (iii) the lower quartiles have to be > 95.

III) Uniquely expressed gene: (i) the lower quartile of the RPKM distribution of the target tissue has to be > 5, whereas, for each of the other tissues, (ii) the lower quartiles of the cRPKM distributions have to be 0 and (iii) the upper quartile < 1.

IV) Uniquely repressed gene: (i) the upper quartile of the RPKM distribution of the target tissue has to be < 5, and (ii) the lower quartile must be 0, whereas, for each of the other tissues, (iii) the lower quartiles of the cRPKM distributions have to be >5.

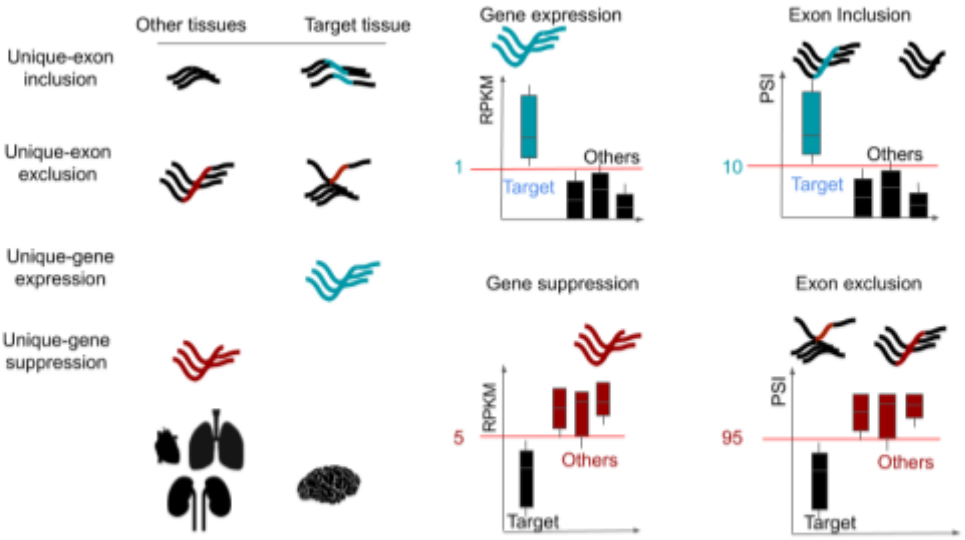


Figure R1.1A. Schematic summary of the approach used to define tissue-unique transcriptomic elements.

Genes with tissue-unique steady state levels were predominantly found in testis (131/147, 89%), and eight additional genes were uniquely expressed in the skin (involved in the keratinization) and a few in other tissues (Figure R1.1B). Different cut-offs for steady state levels (GE) resulted in only small differences in the total number of tissue-unique genes (data not shown). On the other hand, tissue-unique alternative splicing (AS) provided a richer and more diverse set of unique sequences in testis (54% of all tissue-unique AS events), brain (23%) and muscle tissues (12%). Except for uniquely excluded exons in testis, we found only a few cases of unique exclusion/repression in any tissue at the gene or splicing level (Figure R1.1B).

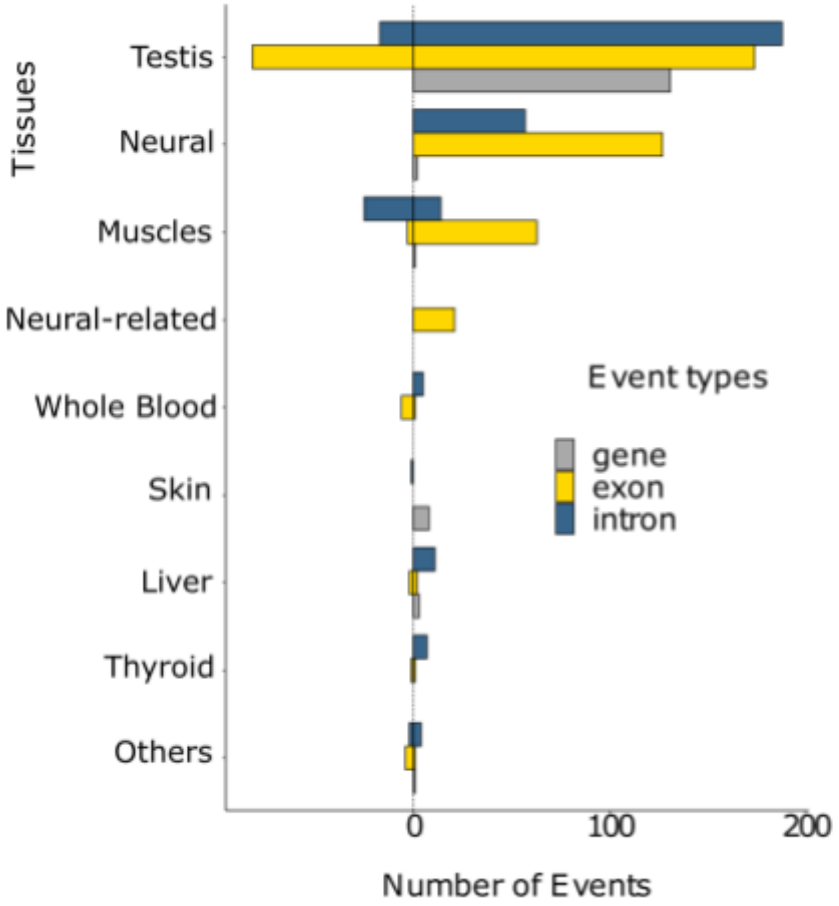


Figure R1.1B. Number of tissue-unique genes or AS events per category.

Given the way each set was defined (see Methods), we did not find overlap between the uniquely spliced gene and the unique genes expression. We validated 14/15 (93.3%) tissue-unique exons from testis, brain and muscle by RT-PCR from independent human samples (Figure R1.S1).

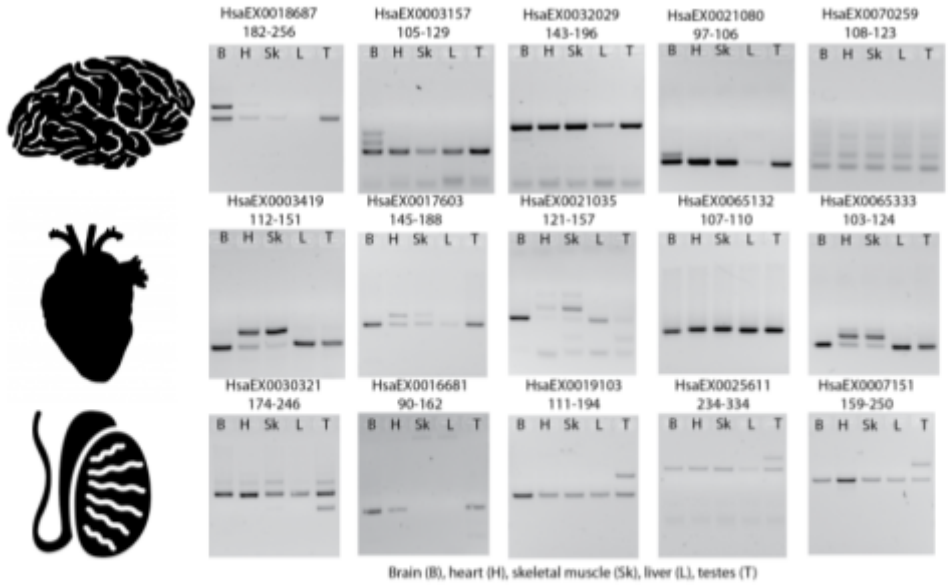


Figure R1.S1. RT-PCR assays validating the splicing patterns of tissue-unique alternative exons (by Jon Permanyer).

3.6 Neurons and spermatids drive neural- and testis-unique splicing profiles

Next, we explored which cell types could be driving the unique AS profiles of testis and brain (we omitted muscle tissues from this study because muscle tissues have low cellular heterogeneity). For brain, we used gene expression markers for neurons, various glial cells, and endothelial cells. We found a higher correlation of brain-unique exons with the neuronal markers (Figure R1.1C).

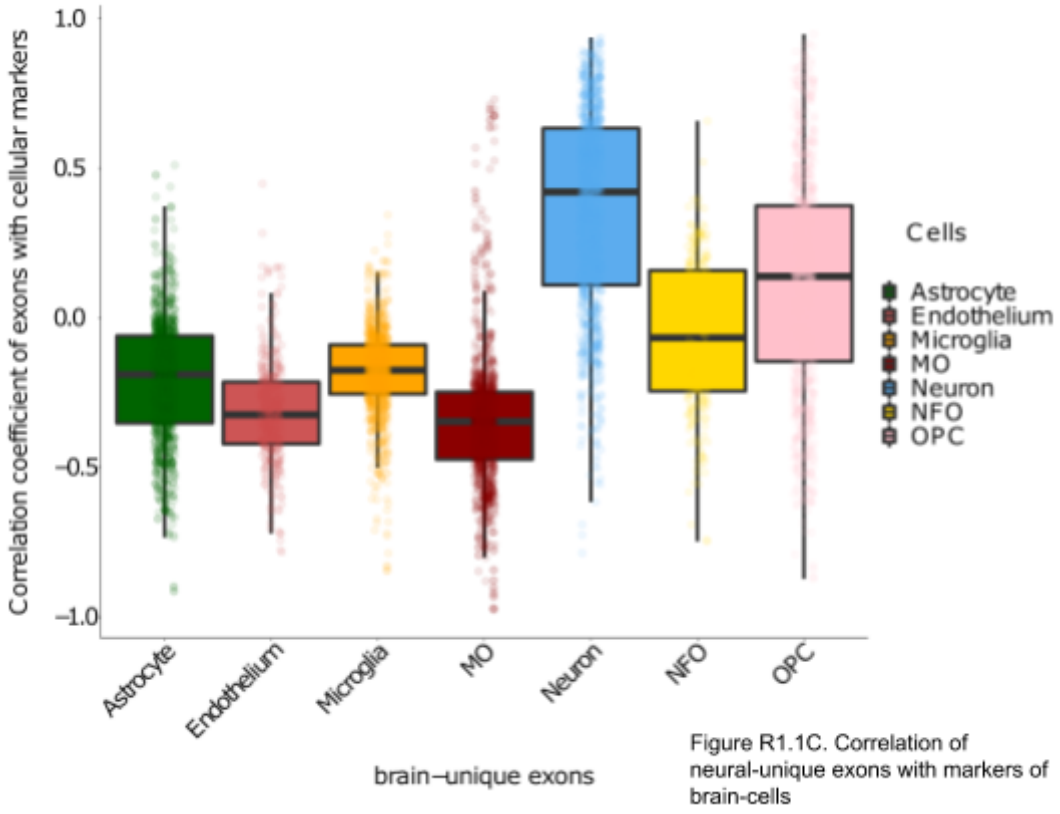


Figure R1.1C. Correlation of neural-unique exons with markers of brain-cells

Similarly, we used a set of markers for different testis cell types [113] and observed that both included and excluded testis-unique exons were only highly correlated with markers for spermatids, spermatocytes and, generally, meiotic cells (Figure R1.1D), which comprise more than 75% of the testis cellular repertoire [114]. Furthermore, using an RNA-seq dataset for different testis cell types [115], we found that testis-unique exons were only spliced uniquely in spermatids and spermatocytes (Figure S3). Therefore, the unique profile of the brain and testis-unique exons could be the result of unique cell-types in those tissues since neurons and spermatids were unique to neural tissues and testis, respectively.

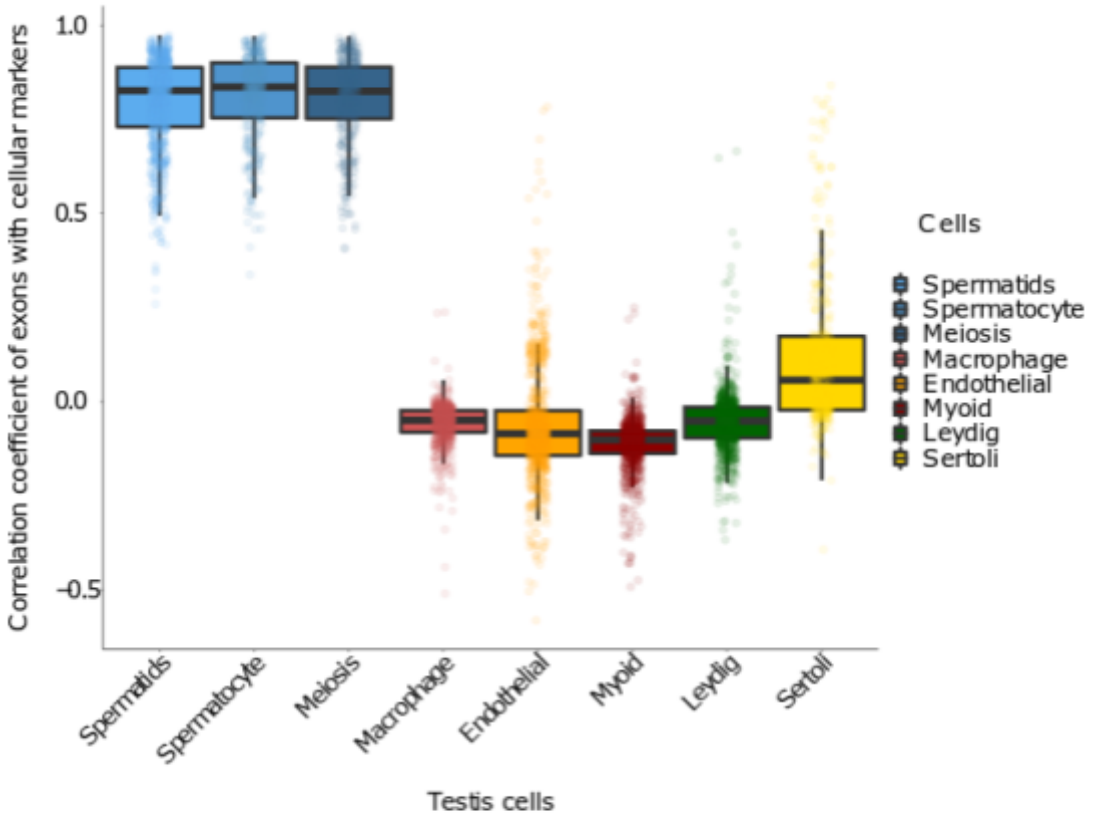


Figure R1.1D. Correlation of testis-unique exons with cell markers of testis.

3.7 Microexons are the main contributor to neural-unique AS

Next, we examined the length of tissue-unique exons. Strikingly, we observed that the majority (82/127, 65%) of brain-unique exons consisted of exons of shorter than 51 base-pairs, and (32/127, 25%) were microexons of length 3-15nt. This pattern is in line with the known microexon neural-specificity [34], and further suggests a dominant role of microexons in providing neural-unique sequences (figure R1.1E).

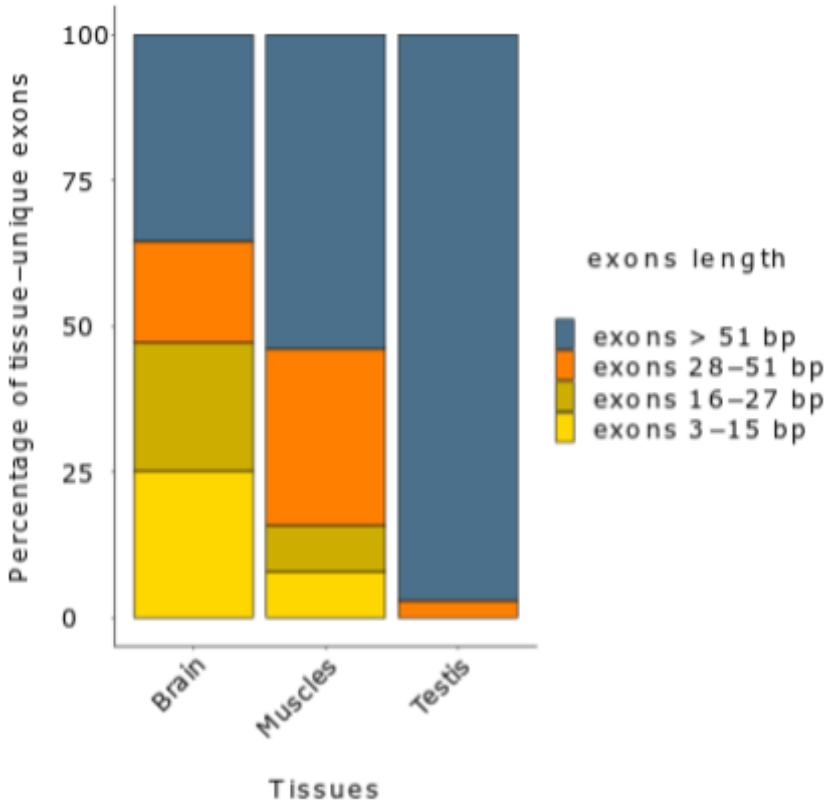


Figure R1.1E. Microexons provided unique transcriptome chiefly to brain.

3.8 Neural-uniquely included exons were highly conserved compared to testis-unique exons

We investigated the evolutionary dynamics of the inclusion pattern of the tissue-unique exons in other mammalian species. For this purpose, we first identified orthologous exons in chimpanzee, macaque and mouse using a liftover-based approach [32]. Next, to assess the evolutionary conservation of the regulatory patterns, we collected RNA-seq data from brain, muscle, testis, lung, liver and spleen of chimpanzee, macaque, and mouse and profiled AS genome-wide using vast-tools (see Methods). We observed the neural-unique exons were the most conserved and testis-unique exons the least conserved across mammalian species (figure R1.1F). This observation implied that neural-unique exons were under stronger negative selection in the mammalian clade. Moreover, the unique exons of the testis might evolve faster than in the other tissues.

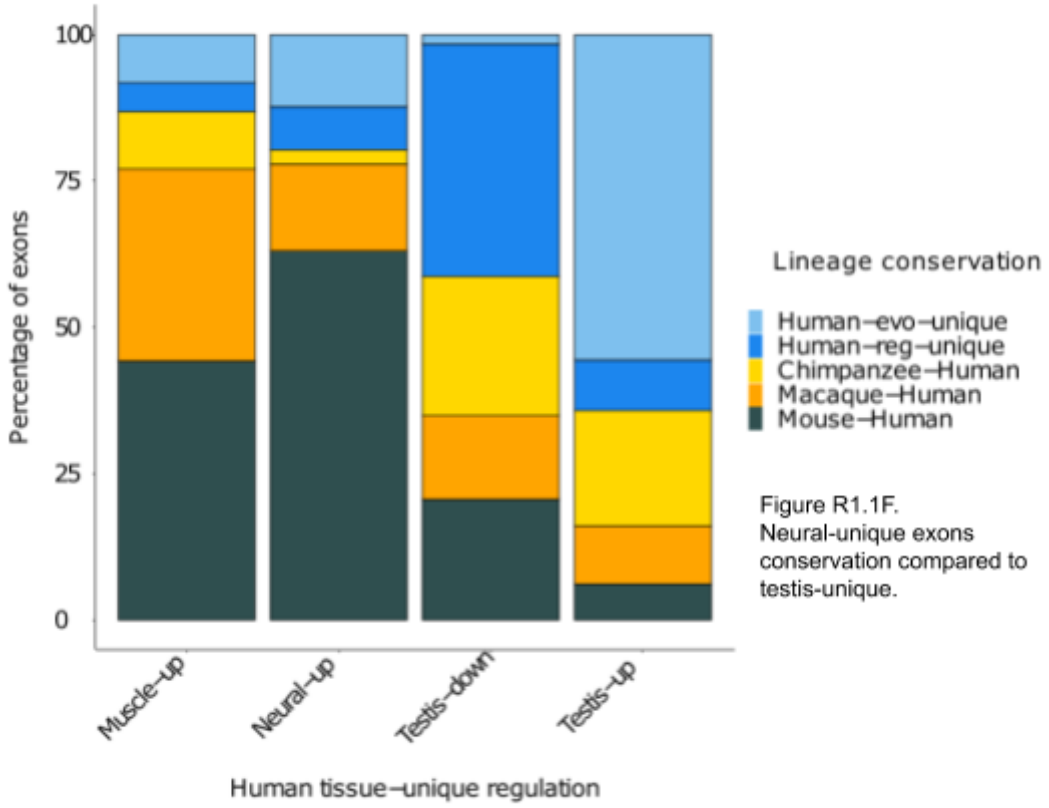


Figure R1.1F.
Neural-unique exons conservation compared to testis-unique.

Next, we examined if the disruptive unique testis exons were less conserved compared to the protein changing exons. We observed that only two of the testis-unique disruptive exons were conserved in the mouse, while the majority of the testis-uniquely included or excluded in the human testis. Besides, the isoform changing exons were highly conserved between human and mouse.

3.9 The majority of the testis-unique exons disrupted the reading frame

We investigated the effect that tissue-unique exons of protein-coding genes had on the ORFs to determine the possible effect they have on proteins. We found that, while the majority of neural- and muscle-unique exons preserve the ORF but changed the coding sequence of the proteins, testis-unique exons disrupted the reading frame of the mRNA and are predicted to trigger non-sense mediated decay (NMD) upon inclusion or skipping (figure R1.2A). Therefore, testis-unique events are predicted to affect gene's function negatively in the testis.

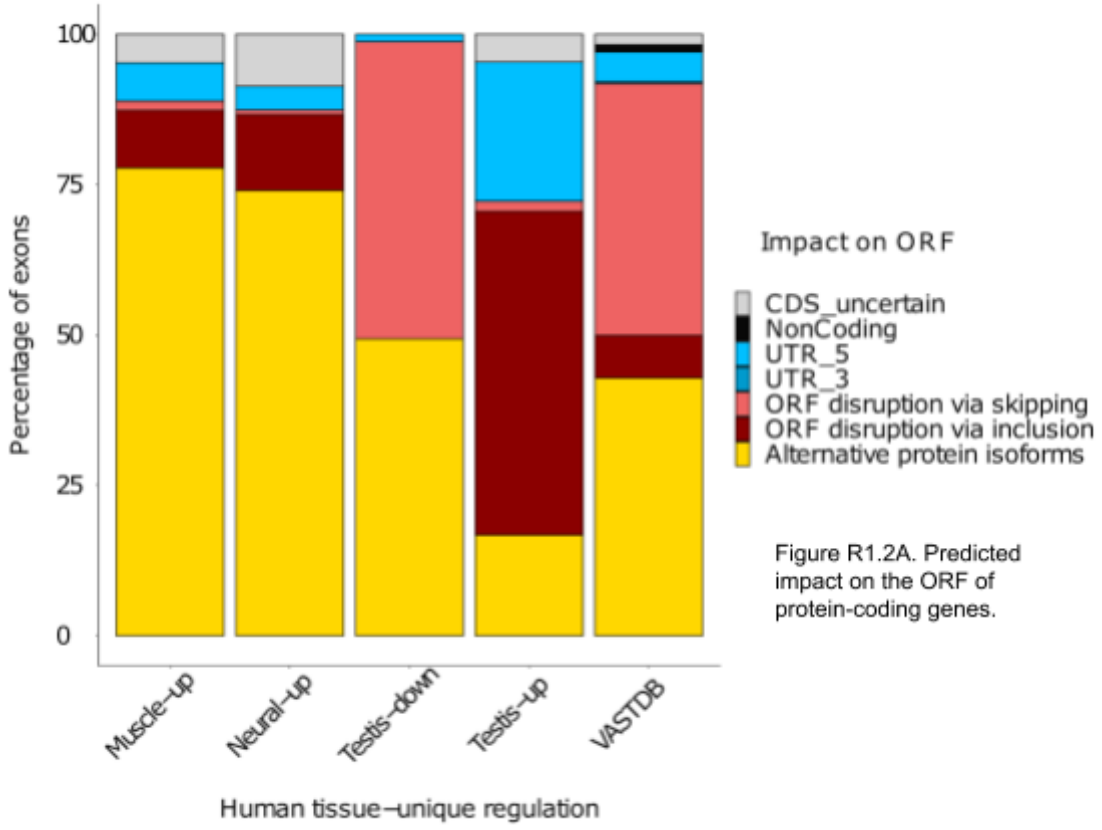


Figure R1.2A. Predicted impact on the ORF of protein-coding genes.

Observing a high percentage of disruptive events could mean that the NMD activity is low, in line with increased ratio of UPF3A over UPF3B in testis, and particularly in spermatocytes [116]. The NMD factors exhibited an elevated expression in testis, but so did they in other tissues. For instance, UPF3A is both elevated in testis and cerebellum, and the ratio of UPF3A to UPF3B is very high both in testis and spinal cord. If the NMD activity is low in testis, we should observe an increase in the mRNA species of the testis that disrupt the reading frame that under normal conditions would be eliminated, akin to mRNA collected from UPF1 knockdown samples compared to

control. By knocking down the UPF1 factor, the disruptive mRNA could survive longer in the cell; thus we expect an increase in PSI of exons that disrupt the reading frame by inclusion and a drop of PSI for exons that disrupt the reading frame by skipping compared to control. Therefore, we compared the differences in PSI value of all the exons of testis over all the other tissues to investigate a shift of NMD-triggering mRNA in testis. Clustering the NMD factors' expression showed that testis and cerebellar tissues were more similar, whereas liver clustered further. So, we chose liver, cerebellar tissues, testis, UPF1 knockdown cell line samples, and spinal cord (as it had the highest UPF3A to UPF3B ratio next to the testis) to scrutinize whether other tissues with similar NMD expression profile exhibited similar enrichment as testis did. We saw a distinct shift for UPF1 KD compared to control except for alternative protein isoforms as expected. Testis, unlike other organs, showed similarity to UPF1 knockdown shift in disruptive events except for the exon skipping (figure R1.2B).

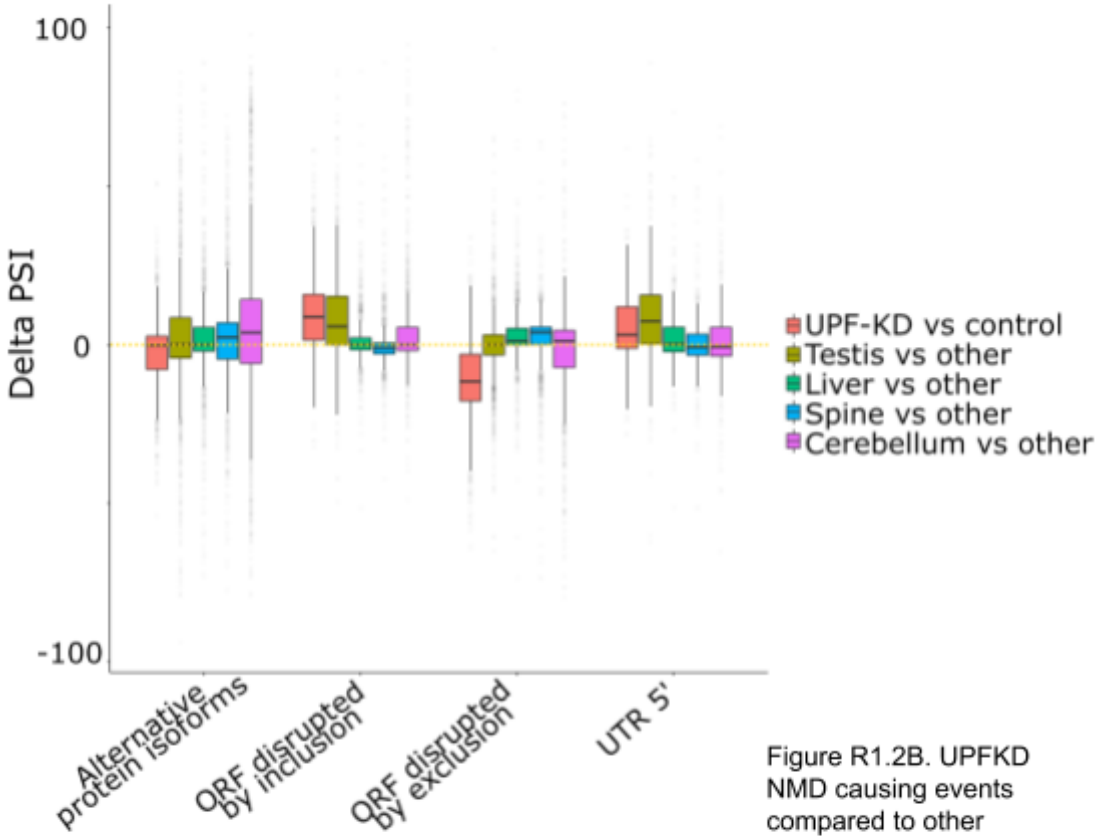


Figure R1.2B. UPFKD NMD causing events compared to other tissues.

Not observing a shift for disruptive skipping events might indicate that NMD is not globally suppressed in testis, instead, the disruptive exons of testis could be regulated or protected. Thus, we set out to investigate whether disrupted genes had a lower expression compared to genes without any NMD-triggering events because activated NMD digests the mRNA that are disrupted. Interestingly, we observed in all the compared tissues that the disrupted events had a lower expression as expected

(Wilcoxon test, p-value<0.0005), except for the UPF1 knockdown we found significantly higher expression for disrupted genes. Therefore, suppression of NMD is unlikely to occur in testis, sperm, or spermatids (figure R1.2C).

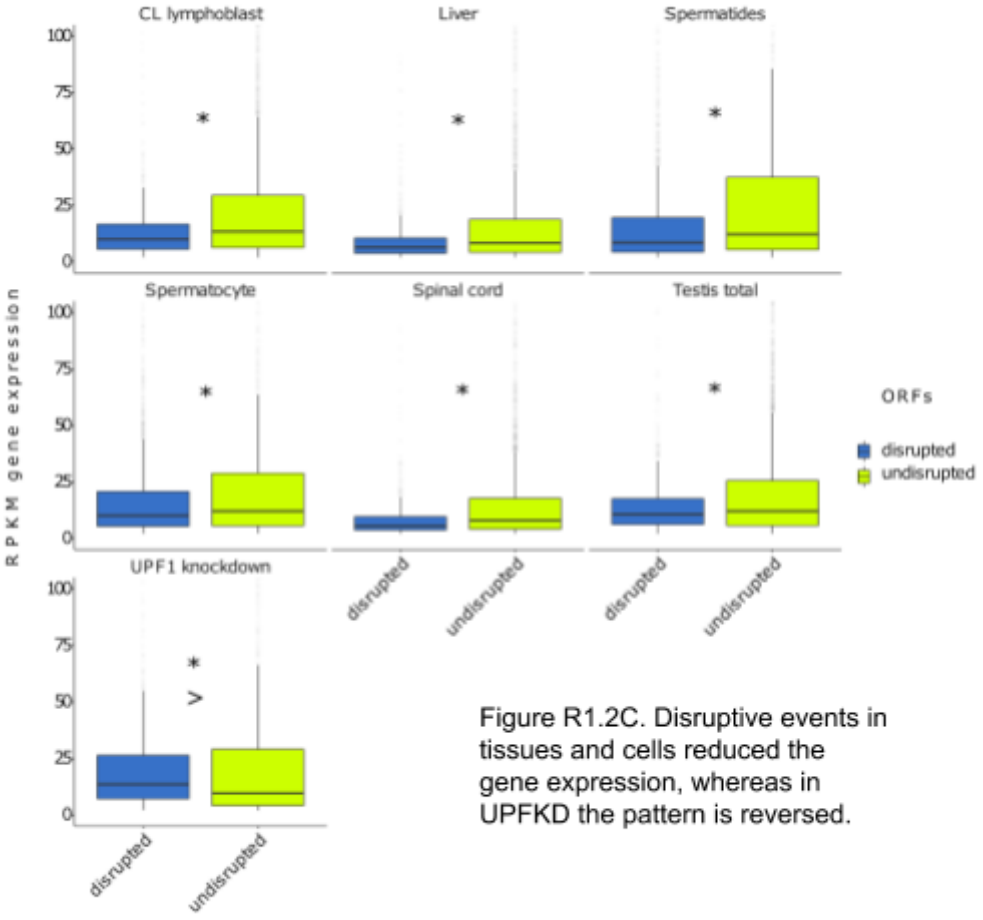


Figure R1.2C. Disruptive events in tissues and cells reduced the gene expression, whereas in UPFKD the pattern is reversed.

If the NMD is not suppressed, then how the NMD-triggering events could survive? We next compared the disrupted gene expression of testis-unique splicing events to all the disruptive events expressed in round spermatids (as round spermatids drove most of the testis-unique disrupted events). Surprisingly, we found that the genes of testis-unique disruptive events were significantly more expressed than the rest of the disrupted genes, suggesting a possibility of disrupted mRNAs being protected (figure R1.2D, adjusted p-value <0.001, Wilcoxon test).

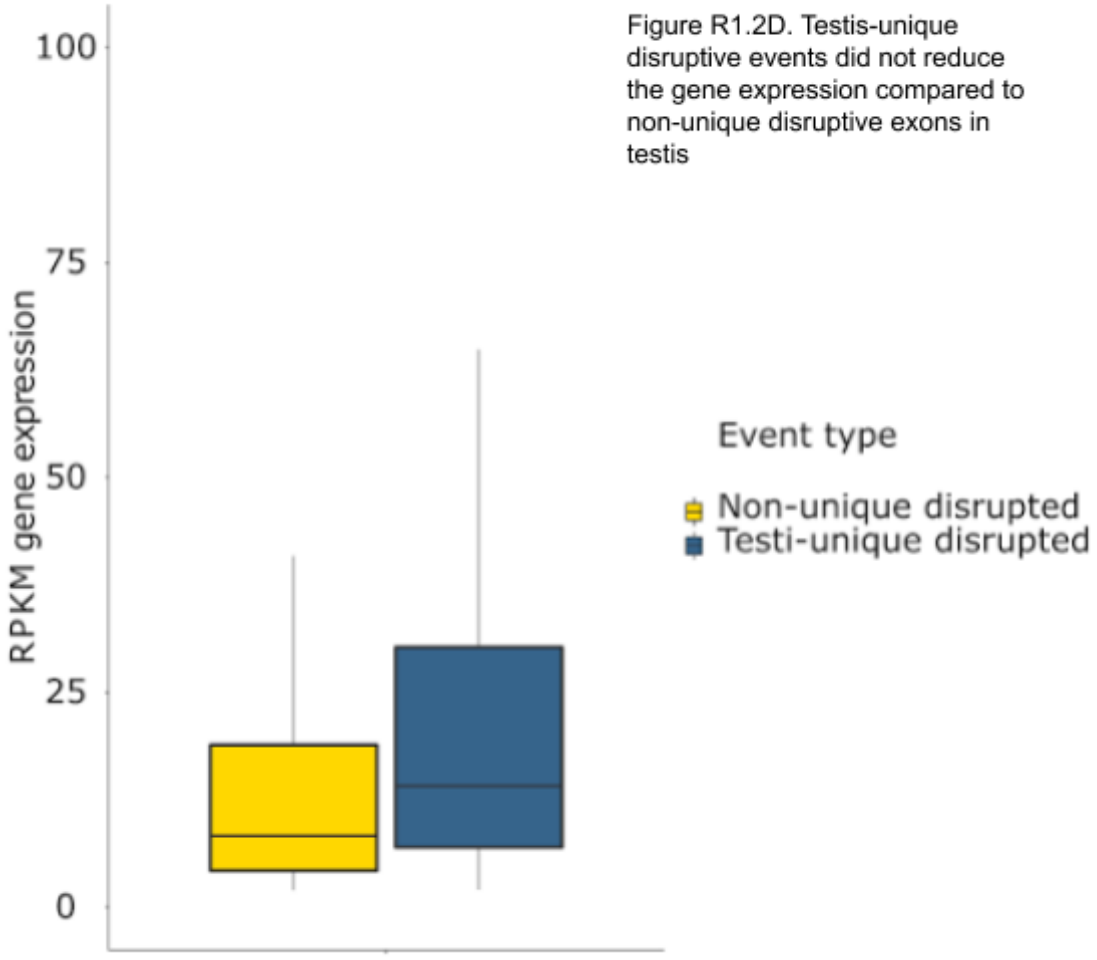


Figure R1.2D. Testis-unique disruptive events did not reduce the gene expression compared to non-unique disruptive exons in testis

We then investigated whether the high ratio of NMD-triggering events was conserved in mouse although the disruptive events of testis themselves were not (figure R1.2E).

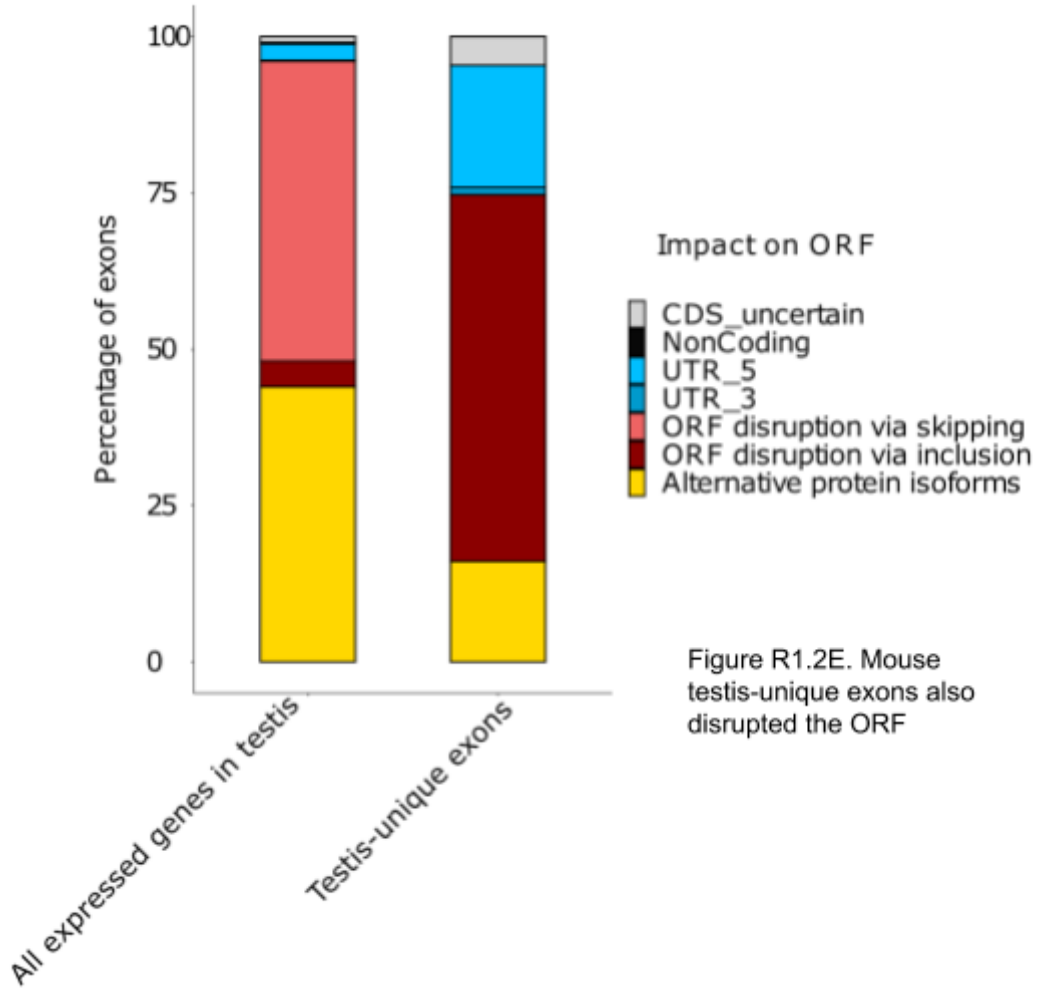
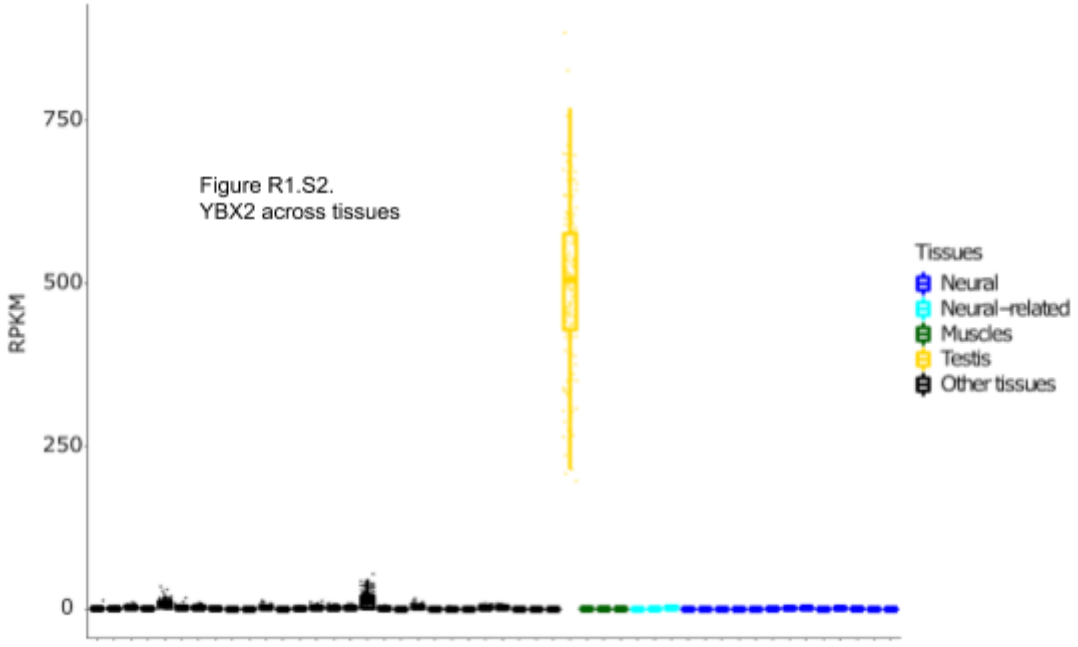


Figure R1.2E. Mouse testis-unique exons also disrupted the ORF

We found that mouse's testis-unique inclusion events to be enriched in disruptive events too. For this reason, we could investigate the effect of UPF3A knock-out samples with the mouse testis-unique exons. The exons that change the inclusion value upon loss of UPF3A (dPSI >= 10) overlapped only 1 out of 87 of mouse testis-unique exons (85 of the events had enough read coverage. So, the genes were expressed both in the cell line and spermatids). Therefore, it is unlikely that UPF3A alone regulated the suppression of NMD and give rise to testis-unique exons. Besides, observing no shift for spinal cord's exons indicated that the ratio of UPF3A to UPF3B is probably not the cause of testis disruptive inclusion or skipping in contrast to the results of a recent study [116].

Another scenario that could explain the presence of disruptive exons was that these set of exons were protected from NMD or translation machinery. For example, YBX2 exhibited the property of sequestering a set mRNA from translation, and hence NMD machinery, only to be translated later during spermatogenesis [117], and mutation in YBX2 in humans was associated with male infertility. We found YBX2 expressed strongly in testis and its expression to soar in round spermatids and spermatocytes. YBX2 had a testis and spermatid-enriched expression both in human and mouse (figure R1.S2-3). Protective role of YBX2 could be reconciled well with not observing a global increase of NMD-triggering mRNAs.



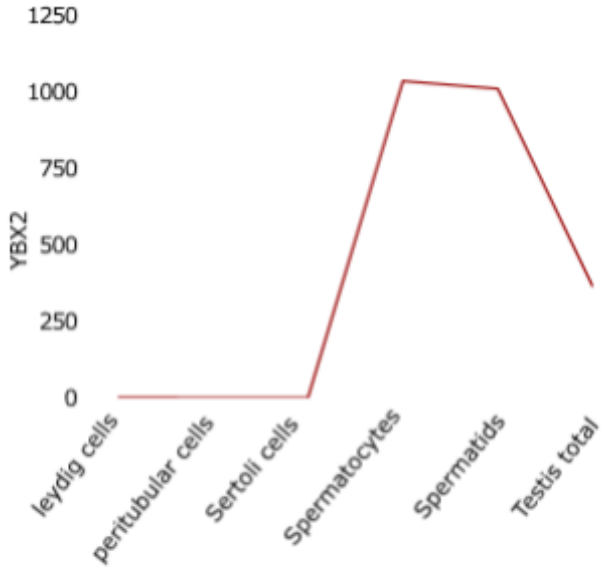


Figure R1.S3. YBX2 expression across cells of testis.

3.10 Disruption of the exon splicing was the main contributor to individual-unique transcriptomic differences

At the individual level, we profiled the events that had sufficient read-coverage in at least five samples of at least five tissues of the compared individuals. In addition, we filtered the exons that varied across tissues compared to individuals to avoid the confounding effect of unbalanced sampling of organs across individuals (see Methods). We define individual-unique genes or events as those with differential patterns in a minority of individuals (from one individual to fewer than 20% of the total population of 505 individuals). Consistent with previous studies showing a low level of expression variability of genes across individuals [50], we found only one protein-coding gene (SSX5) that was uniquely expressed in an individual. We discovered that the majority of the individual-unique cases were exons that were skipped in a given set of individuals (figure R1.3B). Our observation was consistent with a recent study demonstrating that the probability of an individual genetic variant causing exon skipping is much higher than those causing cryptic exon inclusion [55]. Remarkably, we found that the vast majority of individuals (81%) had at least one individual-unique event (either shared with other individuals or occurring only in a given individual). We found, on average each individual has 1.8 unique exons and individuals with a higher number of unique events were less likely to be found. We detected 7 unique events for a given individual to be the maximum (figure R1.3C). In addition, even though we allowed individual-unique events to be shared among 20% of the individuals, the majority of the events were found in only one individual, suggesting that rare genetic variants cause them in the population and that the individual-unique events likely have a low fixation rate in the population (figure R1.3A).

3.11 Mutation on the splice-site and splicing motifs explained the half of the individual-unique exon exclusion

The scenario for disruption of splicing factor's function in trans was possible but unlikely because we did not observe any individual with more than seven individual-unique events. Therefore, it is likely that mutations at cis could explain most of the individual-unique events. To test this, we separated splice-site of the individual-unique exon exclusion, and the motifs flanking the 5' splice-site (GTAAGT), and a set of splicing motifs previously shown to influence exon inclusion values in the human genome (Kishore Jaganathan et al. 2019). Then, we examined whether the genomic differences among individuals occurred on the splicing regulatory sites of the individual-unique exons plus the flanking 250 base-pairs. From 155 individual-unique exons of the individuals with genotype data, we found evidence for SNPs that fell on the splicing regulatory regions predicted to disrupt 70 exons (45% of the individual-unique exon exclusion with genotype data). Among these, 19 were caused by cis changes in the canonical splice-site dinucleotides, 8 in the extended 5' splice-site, and 43 in the deep-learning-inferred splicing motifs (figure R1.3D).

We next examined whether inter-individual genetic differences from an independent source (1000 genome project, phase III) were more enriched in individual-unique exons compared to human-unique and tissue-unique exons. We observed a higher number SNPs intersected the individual-unique exons compared to other unique sets, supporting the idea that the genetic differences were the main contributor to individual-unique transcriptomic profile (figure R1.3E).

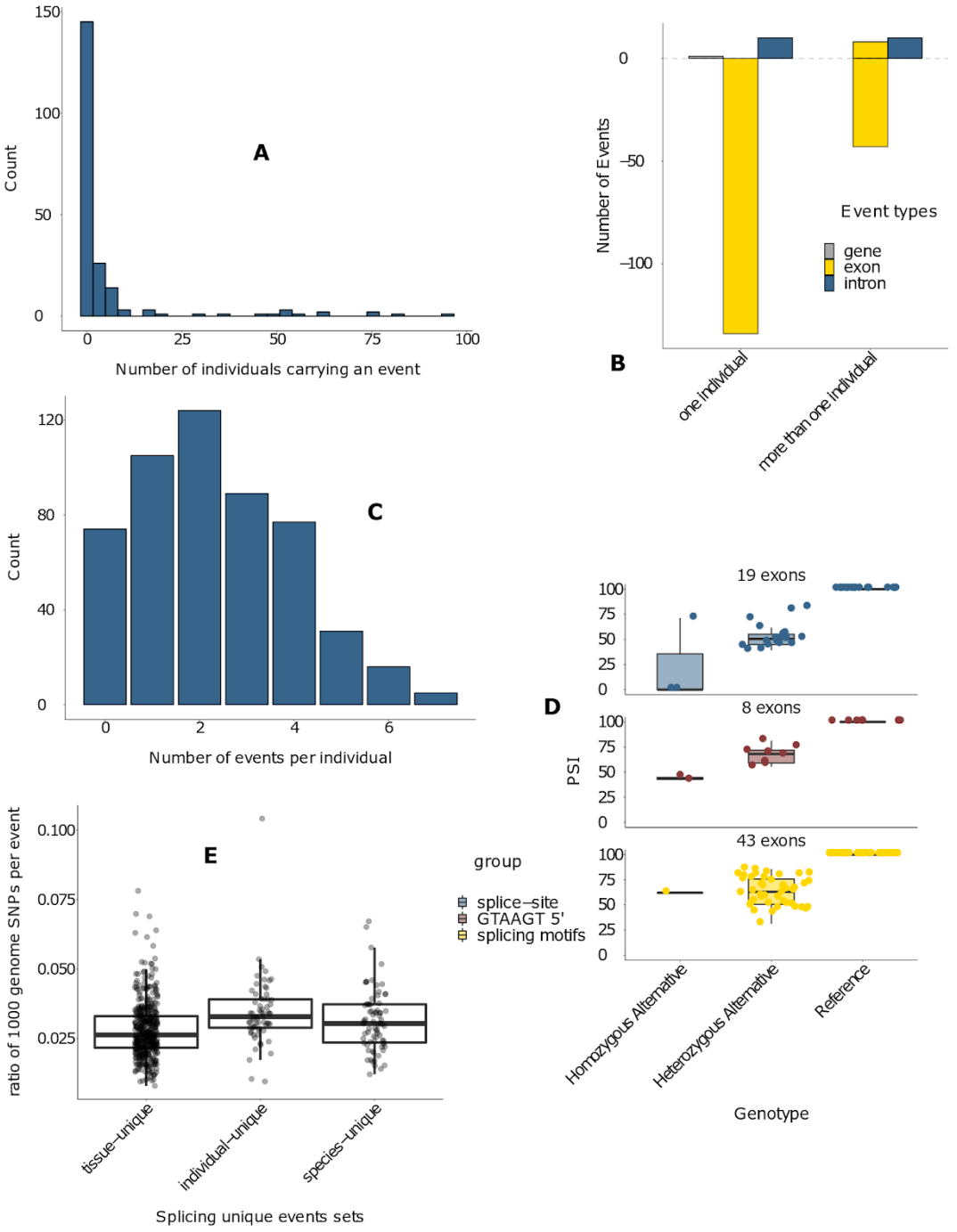


Figure R1.3. A. Higher distribution around zero indicates a low probability of fixation of an alternative new exon. B. Number of individual-unique events by type. C.Distribution of number of individuals carrying the alternative forms. D. Metaplot of individual-unique exons that we found a SNP on the splicing motifs. E. Distribution of 1000 genome SNPs on the unique sets.

3.12 Exon inclusion changes from constitutive to alternative was the prime factor in creating output differences in the species-unique transcriptome

We investigated what unique transcriptomic events have been fixed in the human lineage compared to other mammals (figure R1.4A). At the protein-coding gene level, we found only two genes with orthologs in other mammals that were uniquely excluded in human, and twenty-three human genes with no orthologs that were expressed significantly (median cRPKM > 10) in at least one tissue. In contrast, exons were the prime contributors to species-unique transcriptomic profiles. Using stringent homology filters, we found 15 exons uniquely included in at least one human tissue (median PSI > 10) via gain of novel splice-sites, and 53 exons that were alternative in most human tissues but constitutive in the other three studied mammalian species (mouse, macaque and chimp). These observations indicate that exon exclusion has been fixed in the human population more frequently than exon inclusion (figure R1.4B), and are in line with previous studies reporting that exon alternativization — change from constitutive to alternative — was the main contributor to species-specific AS (Jieyi Xiong et al. 2018).

We hypothesized that if constitutive to alternative form had occurred depending on mutational rate, then we would expect to observe more constitutive to alternative changes the higher the divergence of a species from the rest of the mammals. To assess this hypothesis we profiled unique exon exclusion (constitutive to alternative form) in chimpanzee, rhesus macaque, and mouse. We detected that the number of species-unique constitutive to alternative form increased linearly with divergence time of the species from the rest of the species, implying a neutral evolution of the species-unique exon exclusion (figure R1.4C).

Based on previous results [67], as well as the high fraction of individual-unique events likely caused by cis changes (figure R1.3D), we expect most species-unique AS events to be due to cis-regulatory differences between human and other mammals. Therefore, we investigated whether human-unique exons had a higher number of human fixed substitutions. We used the human divergent sites from chimpanzee from the 1000 Genomes Project (phase III) and removed those sites that were polymorphic in humans. We observed that the ratio of fixed substitution in the human-unique splicing events was higher than tissue and individual-unique events, which supported that cis changes were the primary factor that changes the constitutive to the alternative exons (figure R1.4E).

3.13 Human-unique exon exclusion mainly change the protein isoform or disrupt the reading frame by exclusion

We observed that the majority of the human-unique exon inclusion were included in the 5’ UTR of the transcripts (consistent with Fiszbein A et al. 2019) whereas human-excluded exons either change the protein isoform or disrupt the reading frame via the alternative transcript (figure R1.4D). This pattern is similar to that of individual-unique excluded exons, which either changed the protein isoform (64 per cent compared to 54 per cent in human-unique exon exclusion events) or disrupt the ORF by exclusion (figure R1.4D). However, the higher proportion of disruptive events in the human-unique exon exclusion compared to individual-unique exon exclusion may indicate a higher fixation rate for the disruptive events compared to CDS isoform changing events.

Next, we explored the potential function of the human-unique transcriptome. Even though we did not find any overlap between the human-unique genes and human-uniquely spliced genes, their protein interactors —only the validated and published interaction— were both enriched in epithelium development, tissue and organ development, and skin development. Human skin is one of the organs that had evolved very differently from other primates. This observation may indicate a cooperative evolution of alternative splicing and gene expression to function on the skin, mainly, and organ and tissue development generally. Besides, we observed an enrichment of human-unique genes in the hair growth phenotype(p.adj 0.002). On the other hand, we noticed an enrichment of human-unique exons’ interactors in the immune system process (p.adj 0.001).

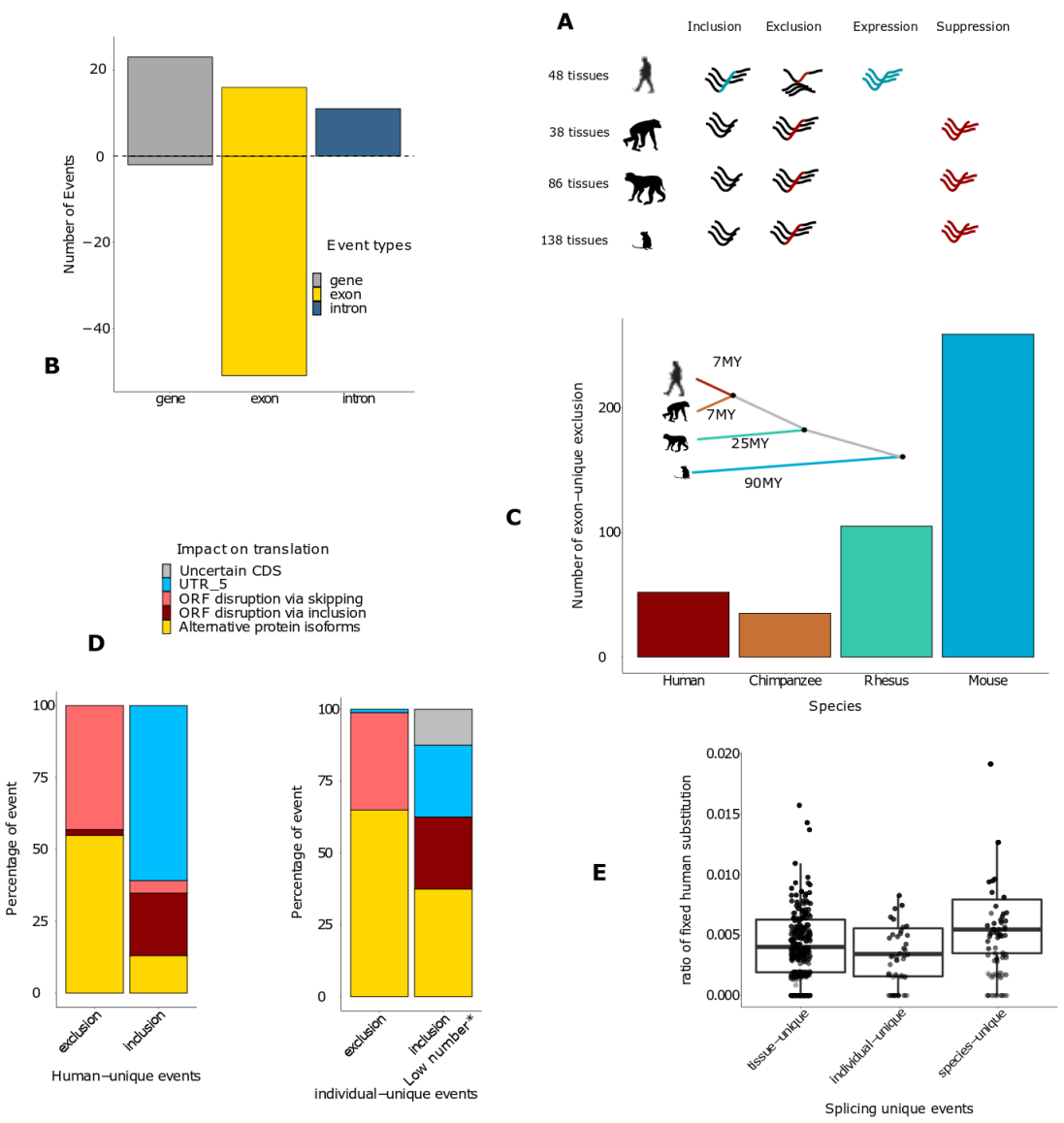


Figure R1.4. A. Human-unique events categorization. B. Number of human-unique transcriptomic events per category. C. Number of constitutive to alternative exons across lineages. D. Impact on ORF, and similarity to individual-unique. E. Human substitution across exon-unique events.

3.14 Discussion

We observed that alternative splicing and gene expression provided unique sequences to different tissues. However, we did not find unique exons or gene expression for most of the tissues. The heterogeneity of the cells inside a given organ could explain;

the organs contained many different cell types, and each cell type could have a different unique gene expression and alternative splicing events. However, when all the cell types of an organ were pooled together, we could only observe the aggregated transcriptomic profile. This aggregation could attenuate the signal for the unique transcripts of an organ, particularly, if the cell-types in that organs were similar with other organs. For instance, in a neural tissue mostly populated with neurons and glial cells that were rare in other organs, therefore, observing a large number of unique transcripts in the neural tissues might have been an indication of uniqueness of neural cell types in the brain that does not exist in other tissues.

The neural tissues-unique transcriptomic profile was mainly driven by the inclusion of the conserved microexons, while in the testis the majority of the unique exons disrupted the reading frame and were lowly conserved. Microexons are neural-specific, and we showed not only were they neural-specific but also they provided the majority of the unique transcripts to the neural tissues.

At the population level, we found in the healthy samples at least 81 per cent of the individuals had at least 1 unique-events, either shared with other individuals or occurring only in a single individual. Previously it has been shown that alternative splicing varied in a higher degree between the individuals than the gene expression. We also showed that the majority of the unique transcript repertoire of the individuals were driven by alternative splicing.

Strikingly, most of the individual-unique events changed a constitutive exon to an alternative exon. We did not detect many individual-unique exon inclusion. One explanation was that the VASTDB junction database was created from the junction that occurs in the screened sample, therefore, if an exon inclusion that was rare in the population would not get its way in the junction database. However, a recent study [55] showed by screening different individuals in the population that losing the acceptor and donor splicing site was more common than gaining it. Furthermore, we found fewer novel exon inclusion that was included in all the tissues of a given individual compared to exclusion events (fig S). It appeared to us that exon exclusion events occurred more frequently at the population level. However, the rate of the fixation of these events should be meagre because we observed that it is less likely to detect individual-unique exon exclusion to be excluded in a high number of individuals. Nonetheless, these events provide the raw material for the species divergence.

More interestingly, we saw this pattern of change from constitutive to alternative form was also more prominent at the species level, indicating the fixation of individual-unique events. These fixations may have arisen via a neutral drift. However, we observed a change in the proportion of the potential effect that exon exclusion had on the reading frame of the transcripts. Whereas the individual-unique events changed the CDS more frequently, at the species level this proportion dropped by 10 per cent and more events disrupted the reading frame by exclusion. This could be explained by either the differential selection of the disrupting events compared to CDS changing events or the CDS changing event became disruptive during evolution.

Finding that human-unique gene expression and splicing event genes interacting via proteins of the same biological process may indicate a cooperative evolution of gene

regulation and post-transcriptional modification. Their functions converged on the skin development, organ development, may help us explain the unique characteristics of the human skin compared to other mammals; Homo sapiens lost its body hair during evolution and developed different sweat glands from the other primates. Similar to the low DNA sequence differences between human and chimpanzee we did not observe a high number of unique transcripts that only belong to the human at least in the adult form, whether these human-unique differences could explain some of the phenotypic differences between human and chimpanzee is an open question to be addressed.

3.15 Material and methods

3.15.1 Quantification

We used RNA-seq samples from postmortem tissues to quantify gene expression and RNA splicing values [12]. We profiled the gene expression and splicing values for a total of 8378 samples from 48 different tissues and 505 individuals. We removed the tissues with less than 20 samples and all the cell-lines from our analysis. We used a good quality set of RNA-seq samples available in GEO for the chimpanzee(38 tissues and cell-lines), rhesus macaque(68 tissues and cell-lines), and mouse(138 tissues and cell-lines) for species comparison. We quantified the alternative splicing events —alternative exons(correlated and anti-correlated exons) and intron retention (IR) as previously described [32,34,118].

3.15.2 Tissue-unique events

We used a rigorous method to define tissue-unique splicing and gene expression events. Events were considered unique if they were included/excluded in just one tissue;
I) For unique inclusion, we required the events to be excluded in other tissues with a first quartile of PSI equal to zero and the upper quartile < 10 PSI, and the first quartile of the target tissue >= 10 PSI.
II) For exclusion, we required the events to be included in all other tissues to have third quartile equal to 100 PSI and the first quartile > 95 PSI, while the third quartile of the target tissue was < 90 PSI.
III) For the unique expression of genes, we required the genes to have their first quartile equal to zero and the third quartile < 1 RPKM, and for the target tissue, the first quartile > 5 RPKM.
IV) For the unique gene suppression, we required all the tissue to have a first quartile > 5 RPKM and the target tissue's third quartile < 1 RPKM.

Besides, for the splicing events, we required a sufficient read coverage — 10 reads per junction at least. The effect of exon inclusion on the reading frame of the transcript was obtained from a previous study(Irimia et al. 2014).

3.15.3 Conservation of the tissue-unique events

In order to examine the conservation of the tissue-unique events we used neural and brain tissues, muscular tissues, testis, and three organs such as the lung, liver and spleen RNA-seq samples from the chimpanzee, rhesus macaque, and mouse. We considered a given tissue-unique event, say a neural-unique included exon, conserved whose median PSI value was > 10 PSI than the other tested tissues and had the highest PSI difference compared to other pairwise comparison. We assigned the event to mammalian lineage if the event also had a high inclusion/exclusion in the same tissue in at least the mouse. We assigned the conservation to primates if it was included/excluded in the same tissue in the rhesus macaque but not in the mouse, and to apes, if it was only conserved in chimpanzee but neither in the rhesus macaque nor the mouse tissues. The event assigned tissue-unique exons to human-reg-unique if they had synteny in the other mammals while we did not discover any specificity for the same tissues in the rest of the tested mammals. We called the event as human-evolutionary-unique if we did not find any synteny in the other mammals.

3.15.4 Individuals-unique events

We used the same threshold of PSI and gene expression for profiling individual-unique and human-unique events. We Considered an event individual-unique if the event in all the samples from a given individual was excluded or included compared to all the samples from all other individuals. In addition, we restricted our analysis to 505 individuals with five tissue types and at least five samples with enough read coverage for a given splicing event, we considered an event individual-unique if it was included or excluded in less twenty per cent of the individuals, and the events did not have tissue variability.

In order to remove the individual-unique events with variability across tissues, we built a linear model to decompose the variance associated with individual differences and tissue differences. We used the lmer function from R package Lme4, and used individuals and tissues as random factors and calculated the relative explained-variance of tissue factor and individual factor.

```
lmer(PSI~(1|individual)+(1|tissue))
```

We kept the events whose individual relative variance explained were > 10 per cent and for the variance explained by tissue < 10 per cent.

3.15.5 Individual-unique genotype analysis

The genotype of 449 of individuals was profiled using the blood samples from the tissue donors(GTEx consortium, 2017). We restricted the genotype analysis to the called SNPs that fell on the splice-sites of the exons, the 5' (GTAAGT), and the sites that were identified to alter the splicing values [55] that were in the flanking of the individual-unique exons. In most of the cases, there was just one SNPs that occurred in the same individual that also had the individual-unique events, and in a minority of events, there were two or more SNPs were assigned to an individual-unique exon, but only one of them was correspondent to the same individuals with unique events. The probability of having the SNPs assigned to the same individual whose exons were uniquely excluded by chance was < 0.001(choose (number of individuals with a given SNPs, 449 individuals)/choose(number of the individuals without the SNPs, 449 individuals)). However, as the site of SNPs were chosen based on previous information about their effect on splicing the probability of false-positive would be even lower.

SNPs enrichment in the individual-unique events was measured by intersecting individual-unique exons plus 250 base-pairs flanking the exons with the called SNPs from the 1000 genome project phase III (The 1000 Genomes Project Consortium, 2015).

3.15.6 Human-unique events

We considered three types of human-unique exons. First, those exons that the genomics sequence only exist in human and it does not exist in any other species such as the chimpanzee, rhesus macaque, and mouse. However, the exons upstream and downstream of the exons should be present in all the other species. Besides, it should be included in at least one tissue with median > 10 PSI. We did not find any exons that met this condition.

Second, those exons in humans that have orthologous genomics regions in nine other species (chimpanzee, gorilla, orangutan, rhesus macaque, mouse, pig, dog, cow) but the complete splice site exists only in humans. These events were the human-unique exon inclusion via gain of splice-sites. We found 192 exons with splice-site gain in the human lineage. Also, we required the events to have the first quartile of > 10 PSI in at least 1 tissue; this filtering reduced the number to 25 exons.

Third, orthologous exons that included/excluded in all human tissue samples pooled together but were not included/excluded in all the tissues of the other three species. We used the same threshold as we used for the tissue-unique and individual-unique exclusion events.

For the gene expression, we applied the same threshold as we did for tissue-unique and individual-unique sets. Furthermore, we profiled the expression of the protein-coding genes that did not have any orthologous genes in other species. We required these genes to have an expression of 10 RPKM in at least one tissue.

3.15.7 Species-unique exon exclusion and divergence time

We assigned a mouse-unique exon exclusion if a given exon was excluded in all the tissues of mouse but included constitutively in the human, chimpanzee, and rhesus macaque. We profile the unique exons of all four species.
We assigned 7 million years for the distance of the human-chimpanzee common ancestor [119], 25 million years from the rhesus to the common ancestor of human and rhesus [120]. We assigned 90 million years for the divergence of the mouse from the common ancestor of mouse and human [121].

Enrichment of human-specific substitution compared to chimpanzee was obtained from a recent study [122]. We used the human-specific substitution that was shared with Neanderthal and Denisovan to reduce the probability of having a polymorphism in human population, in the case that 1000 genome project failed to screen individuals with the polymorphism across individuals and population. Moreover, we removed the human-specific substitutions if they had any called SNPs in the 1000 genomes phase III.

3.15.8 Functional enrichment of the human-unique events

As we did not find any enrichment regarding the functionality of the human-unique events, we investigate if the protein interactors of the unique event's gene had any enrichment in biological processes. We used the validated and published protein-protein interaction from human reference protein interactome mapping project (HuRI) to profile the proteins that interact with the genes of the unique events. Next, we investigated whether there is functional enrichment of these interacting proteins in human phenotype database and biological process using g:Profiler [123]. We filter out those terms that did not pass the threshold of 0.05 of adjusted p.value.

4 Chapter 4 - Tissue-specific circadian and circannual transcriptomic variation in human

4.1 R2. Tissue-specific circadian and circannual transcriptomic variation in human

Sodaei R*, Wucher V*, Amador R, Irimia M, Guigo R

4.2 Abstract

Our understanding of human physiological variation at the level of transcripts during day-night and seasonal cycle is limited. Here we tried to explore human transcriptome variation diurnally and seasonally, using 16825 postmortem tissue samples. While most of the core clock genes differentially expressed between day and night, they did not fluctuate diurnally in all the tissues. Furthermore, we observed that core clock genes' daily cycle of human was very similar to baboon, which is a diurnal animal and behave entirely the opposite of the mouse, which is a nocturnal animal. The general pattern of alternative exons might be affected by temperature during a day or a year. In contrast, we did not observe such a trend for microexons. We found 67 microexons suppressed in the anterior cingulate cortex, a subregion known to be associated with depression, in fall. This, and observing a large number of genes varying in fall in the anterior cortex might suggest a channel that seasonal changes could induce depression, and perhaps other cognitive changes.

4.3 Introduction

The presence of sleep and wake cycle is conserved among animals. But it is not the only sleep cycle that is affected by the daily movement of earth around itself, hormones and metabolome concentration in the blood, eating habits, core body temperature, cell regeneration, oral microbiome, and photoperiodism are among the physiological changes associated with circadian and sleep cycle in mammals [79,80]. Besides, clock disruption in mice results in cardio-metabolic, immunological, and neurological dysfunction. Moreover, the incidence of human diseases such as asthma, depression, and cardio-vascular diseases exhibited a diurnal association [124,125]. Many of those physiological changes are assumed to be regulated at the transcription level.

Mammalian transcriptome regulation showed a tissue-specific daily cycle. Surprisingly, more than 50% of protein-coding genes found to oscillate daily in at least one tissue and this number increased the more tissue were screened (in both mouse and

baboon) [89] [89,90]. Studies on 13 human tissues for diurnal transcriptional profile identified ubiquitous oscillators that oscillated in many mouse tissues, along with other human ubiquitous oscillators that did not oscillate in mouse tissue transcriptome [91]. Given the large variation in circadian behaviour between mammals, it is expected that the circadian regulation of gene expression to be malleable in an evolutionary time scale.

On top of diurnal fluctuations, animals' physiology changes during the seasonal cycle. Duration of pineal melatonin production is inversely related to day length, and its secretion drives long-term changes in many physiological systems, including the hypothalamus-pituitary-adrenal, hypothalamus-pituitary-gonadal, and brain-gut axes, the autonomic nervous system, and the immune system [95]. Observing that some circadian genes also oscillate with the seasonal period could imply the role of circadian genes in determining the photoperiod and hence the seasonality [96]. However, studies on seasonal variation at the transcription and post-transcription across human tissues have been minimal.

Here we set out to investigate the effect of circadian and circannual variation of gene expression and alternative splicing across human postmortem tissues.

4.4 Results

We used 16825 RNA-seq samples from 46 human postmortem tissues whose time and season of death were annotated to investigate the impact of circadian and circannual cycles on human tissues (see [12]). We compared the oscillation of known core clock genes in thyroid —on of the organs with a robust circadian core clock signature ([91]— with the time of death, and time of death plus the postmortem intervals in both baboon and human because it was still not clear whether the transcriptional regulation of circadian rhythm stops after death or continue to oscillate [126]. We observed oscillations only with the time of death. Besides, the pattern of oscillation with the time of death coincided with the baboon's core clock oscillation (figure R2.S1-2). Thus, we used the time of death to assess the circadian oscillation across human tissues. However, we assumed that the time of death annotated by the physicians was not as accurate as the study in baboons and mice, therefore, we divided the tissue samples whose owner died at night and compared them with samples that died during the day. Next, we used voom [127] to identify differentially expressed genes between day and night. We used the same method for the seasonal variation to identify the genes whose expression changed in one season compared to all other seasons. For alternative splicing variation, we used 8374 samples from GTEx with exon splicing quantification. Then we applied a linear model to investigate

whether including season and the day-night cycle could explain the exon inclusion variability significantly.

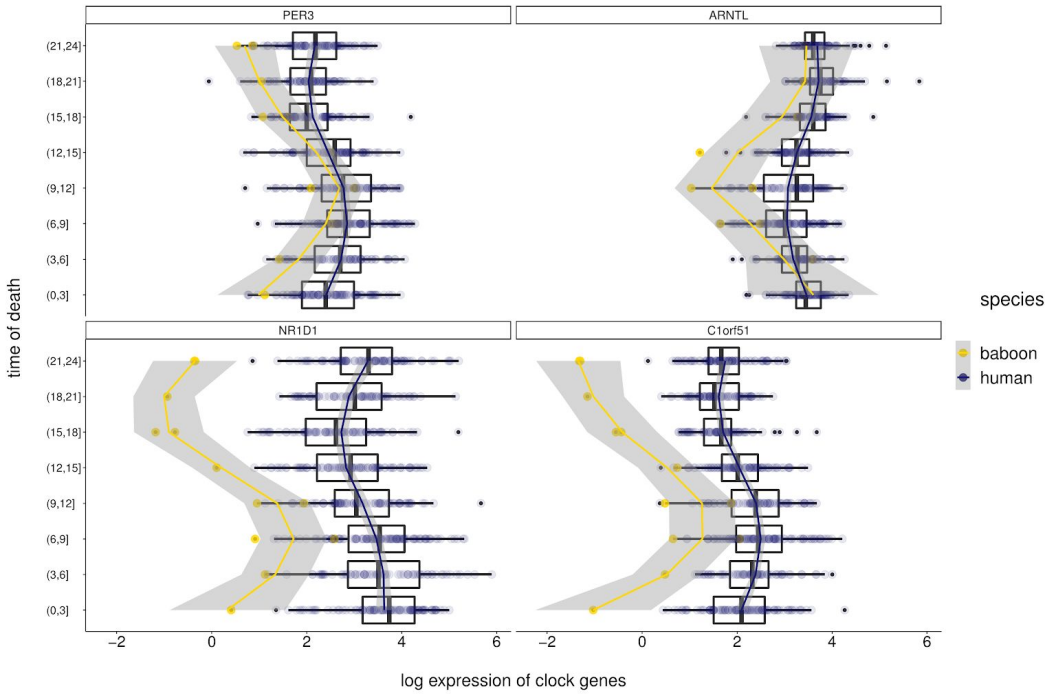


Figure R2.S1. Thyroid, time of death and the expression of core clock genes. The phase of the genes found to be similar to baboon's.

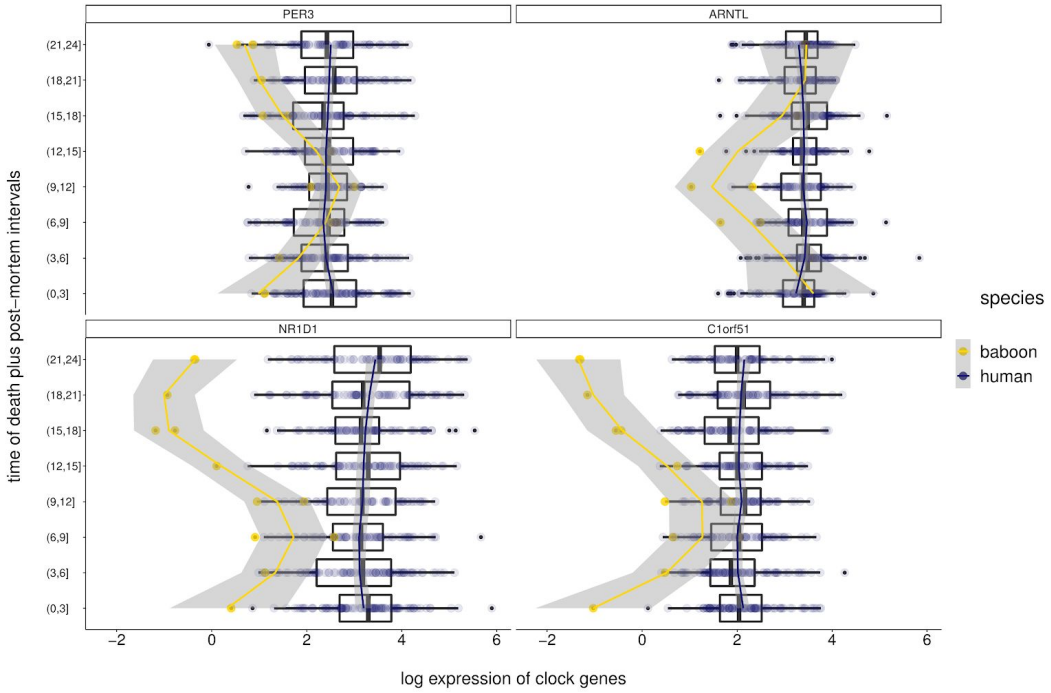


Figure R2.S2. The core clock genes do not appear to oscillate when we measured their expression with time of death plus the postmortem intervals, indicating the transcriptional feedback ceased after the individual deceased.

4.5 Diurnal variation of gene expression and alternative exon skipping

We observed 58% of the genes(protein-coding, non-coding RNA, and pseudogenes) to fluctuate with the diurnal cycle. This observation was consistent with other mammalian circadian gene expression studies ([89]; [89,90]). On average, the tissues had more up-regulated gene expression during the day than at night. This is in contrast to the mouse study, which reflects the effect of the nocturnal activity of a mouse. We observed all the tissues showing splicing and gene expression differences in the day compared to night. Among the screened tissue, we found lung (consistent with [89]) and oesophagus to have the highest number of transcriptional changes. Across brain subregions, we found many genes exhibiting diurnal-cycle regulation in the cerebellum and caudate of basal ganglia. In cerebellum higher number of genes were up-regulated in the day, whereas caudate's genes tended to be expressed more during the night. This observation is in line with the known role of the cerebellum and caudate in sleep-wake cycle regulation ([128,129], figure R2.1A). Interestingly, we found cerebellum to be enriched in the number of diurnal exon inclusion during the day, approximately two times more than any other tissues (figure R2.1B). Observing a high number of genes and splicing events fluctuate diurnally, might support the involvement of the cerebellum in the sleep-wake cycle [130].

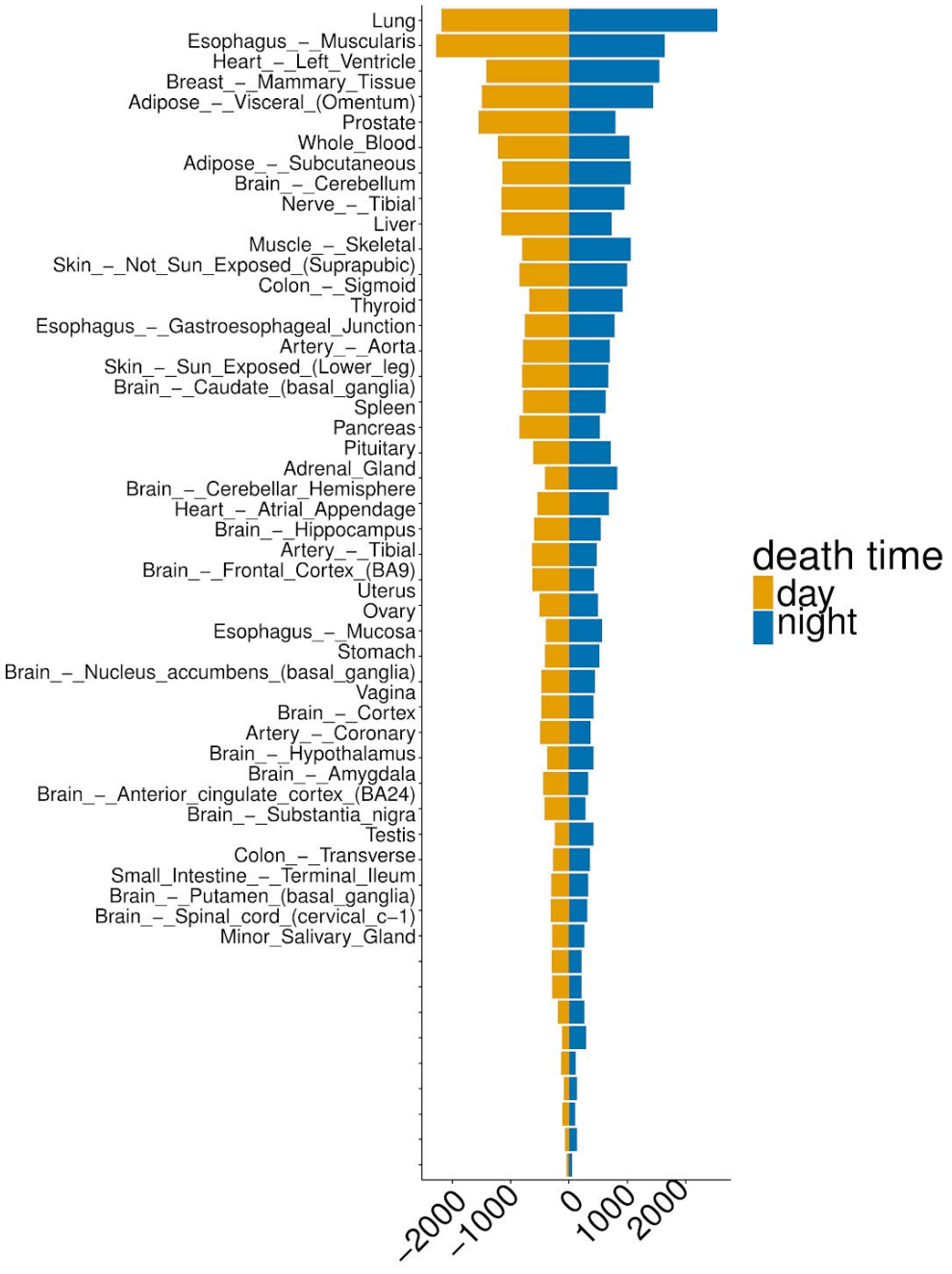


Figure R2.1A. Diurnally fluctuating genes across tissues.

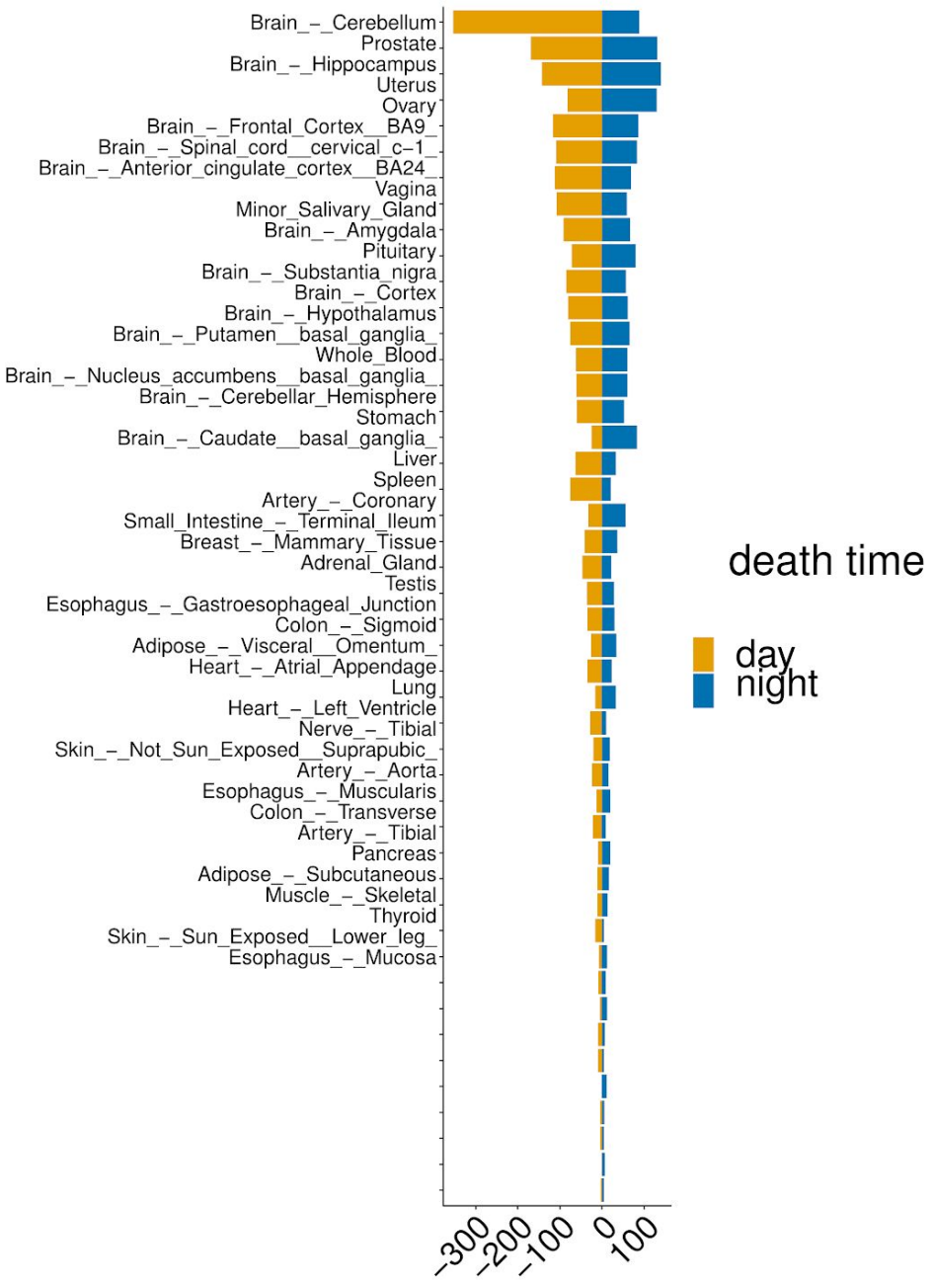


Figure R2.1B. Alternative splicing of exons during the daily cycle across tissues.

Furthermore, we found all the genes (8 genes) that fluctuated diurnally across more than 20 tissues to be known as core clock genes, indicating their general role in regulating human daily cycle physiology. We observed ARNTL to be most universal clock genes across tissues (figure R2.1C). We found that core clock genes had either

a night peak in the tissues or a day peak. However, adrenal gland exhibited a shifted pattern; NPAS2 and ARNTL were expressed during the day, whereas DBP and PER3, were expressed during the night. This pattern was in the contrary to the expression of the same genes in all the other tissues (figure R2.1D). Interestingly, we did not detect any core clock genes to express differently diurnally in testis, vagina, and ovary.

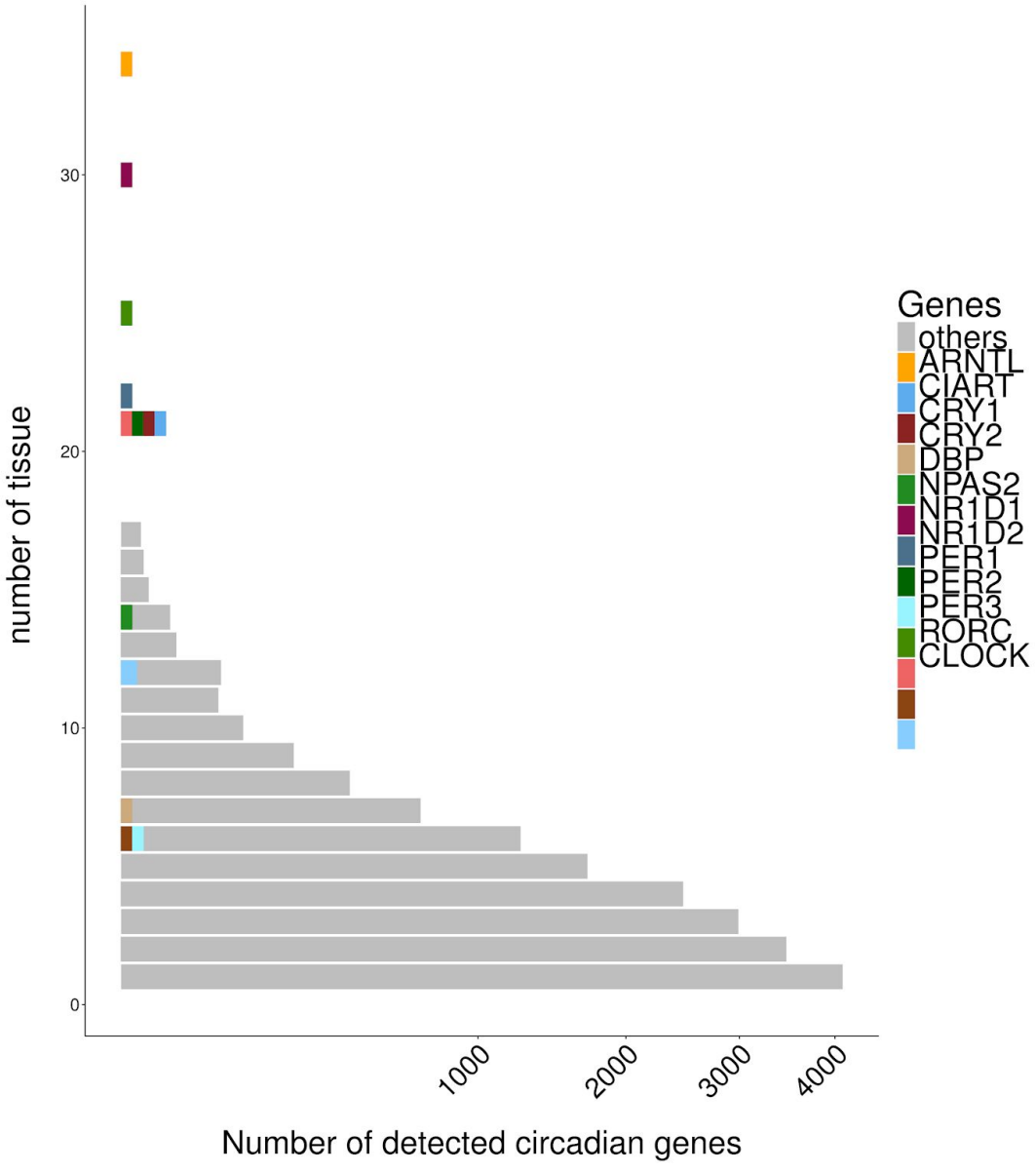


Figure R2.1C. Core clock genes found to fluctuate more frequently as expected.

Observing CLOCK genes exhibiting diurnal cycle expression in only 12 tissues, suggested a lower involvement of CLOCK in circadian regulation than expected. Moreover, NPAS2 shown to compensate the absence of CLOCK, and we observed NPAS2 oscillated diurnally in 30 tissues. Thus, it is likely that NPAS2 replaced the role of CLOCK in mammals because CLOCK was rhythmical only in a few tissues of baboon as well [131].

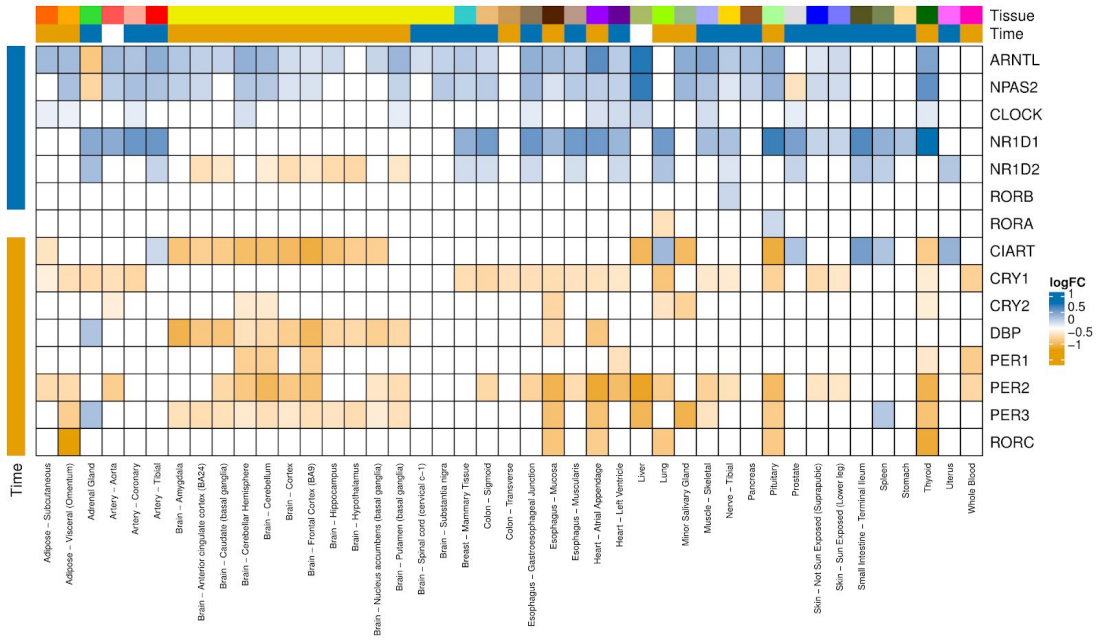


Figure R2.1D. Core clock genes log fold change across tissues. Colors at the top are as a result of summation over the columns. The same goes for the colors at the left, indicating the day or night of the core clock genes.

In addition, we profiled a set of diurnally regulated in more than 12 tissues (as we detected CLOCK's diurnal gene expression pattern only in 12 tissues), in order to investigate other potential factors involved in circadian regulation. For example, we found THRA (a nuclear receptor for thyroid hormones) increased during the night in 15 tissues without an increase in the brain. Interestingly, we observed fewer of these ubiquitous circadian genes to oscillate diurnally in the brain (figure R2.S3).

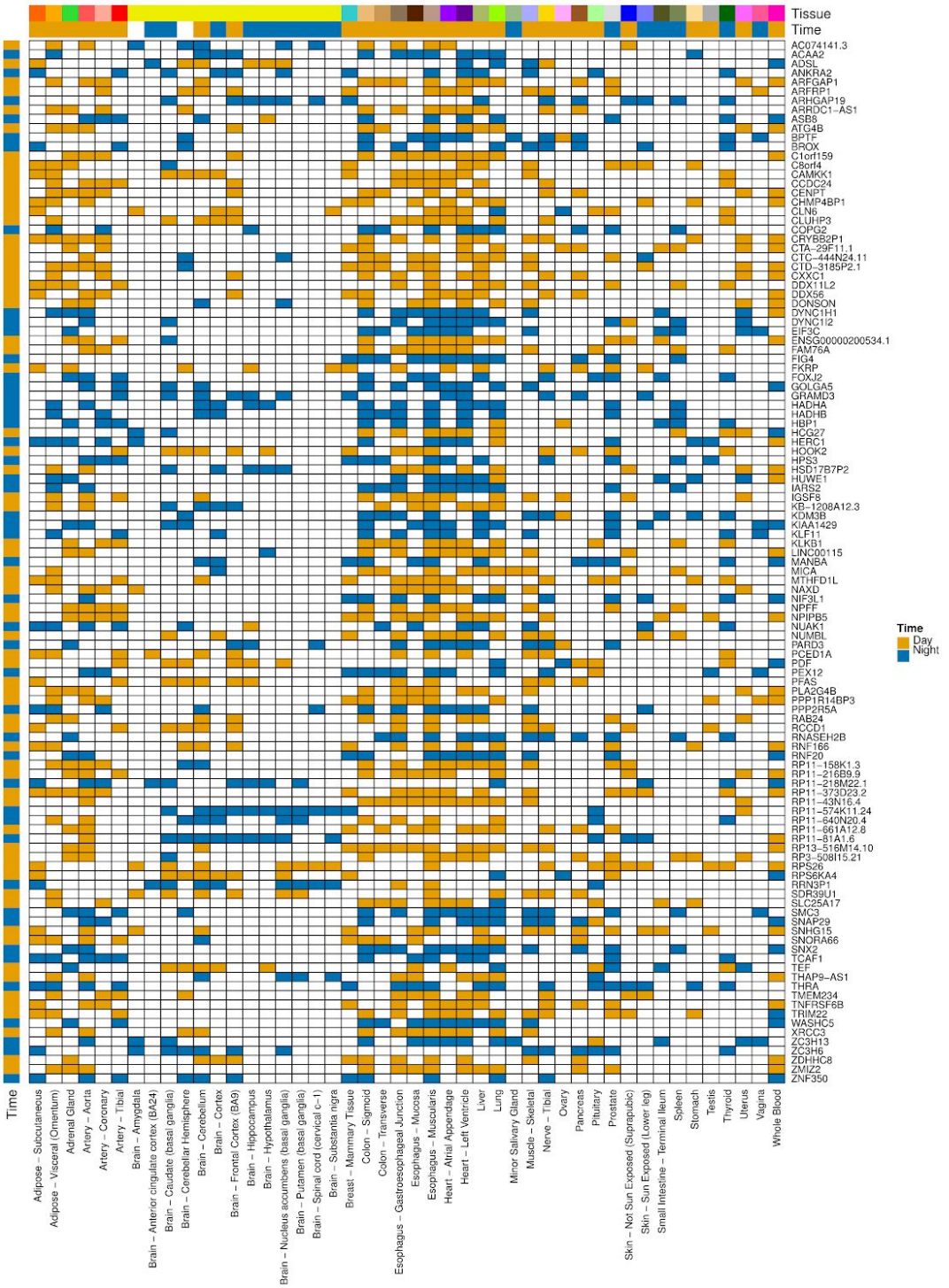


Figure R2.S3. Day-night changes across tissues of the genes which were not annotated as core clock but they fluctuate in more tissues than CLOCK gene.

4.6 Human and mammals core clock comparison

Next, we compared the core clock profile of human against mouse and baboon. Baboon is a diurnal animal, so we expected its core clock genes to be regulated similar to human (57 similar (79%) against 15 divergent). A mouse is a nocturnal animal, and we observed human core clock genes that were expressed during the night in human to express during the day in mouse (12 similar (26%) against 35 divergent, [132]). One exception was the CIART gene in the brain, whose expression increased during the day in human (figure R2.2A-B). However, CIART's peak in baboon was approximately in the early morning, which could result in its assignment to the night in the baboon analysis (figure R2.S4). Those observations not only meant that core clock genes adapt to the lifestyle of an organism.

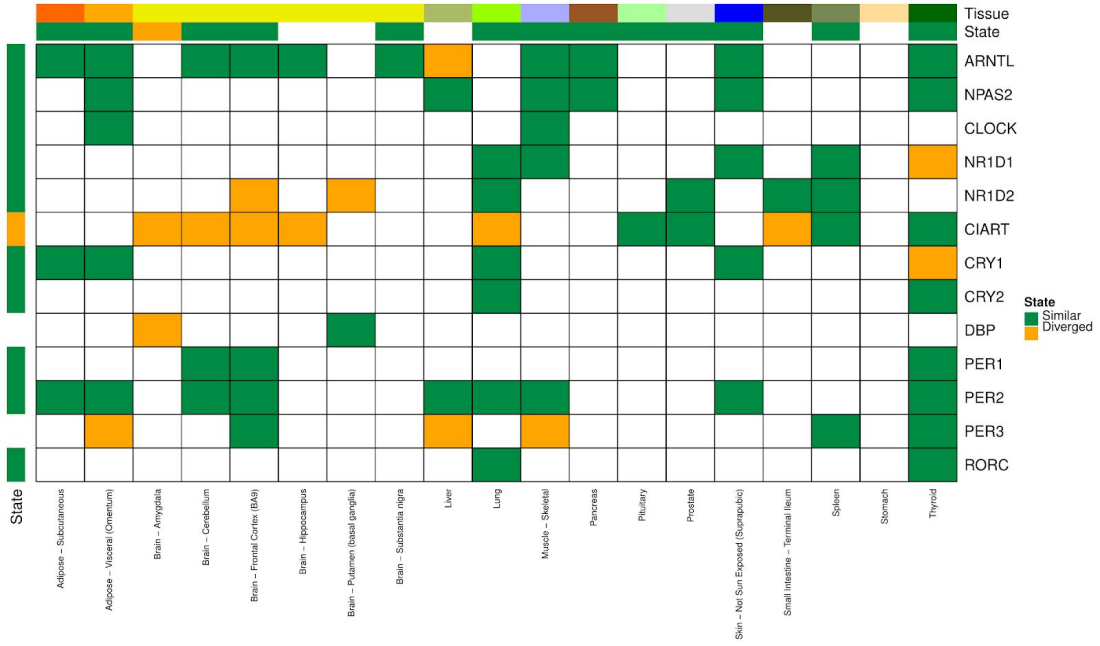


Figure R2.2A. Comparison of baboon and human core clock genes daily pattern.

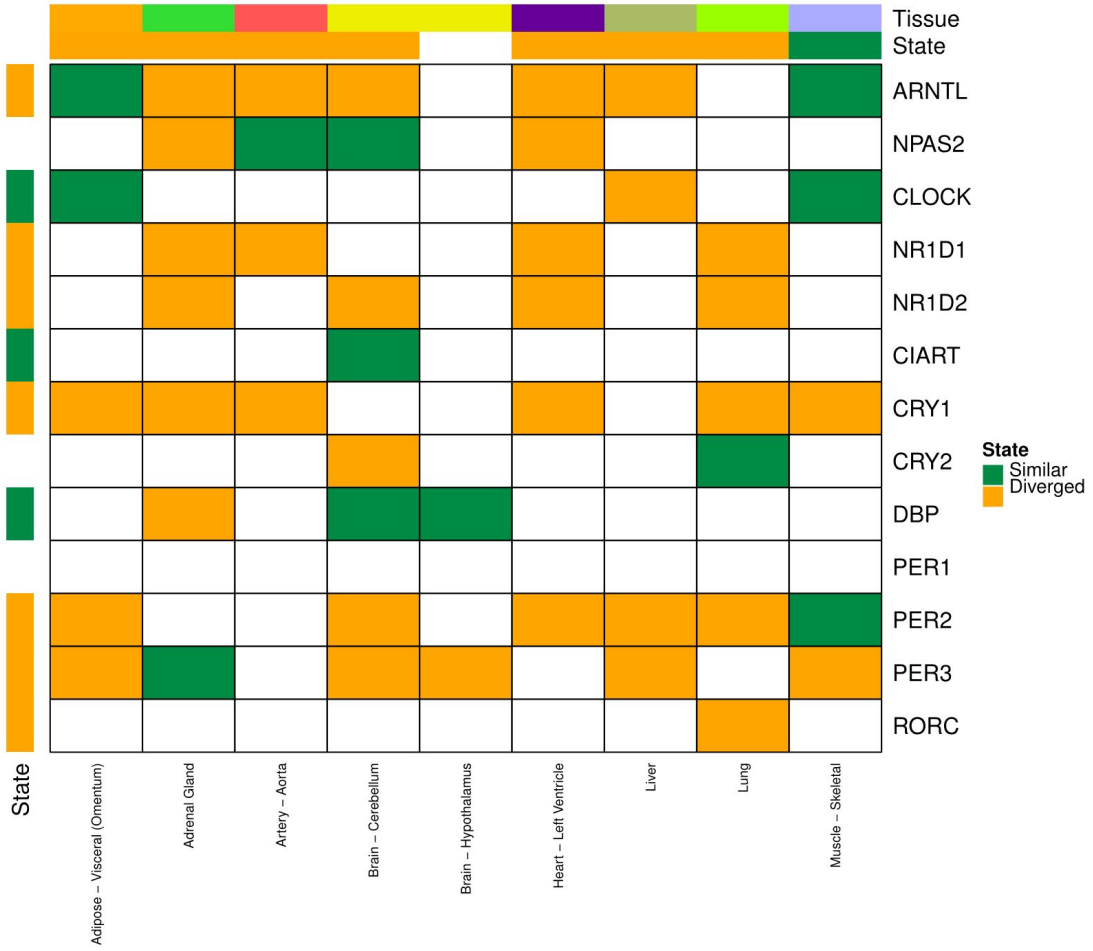


Figure R2.2B. Comparison with mouse core clock genes.

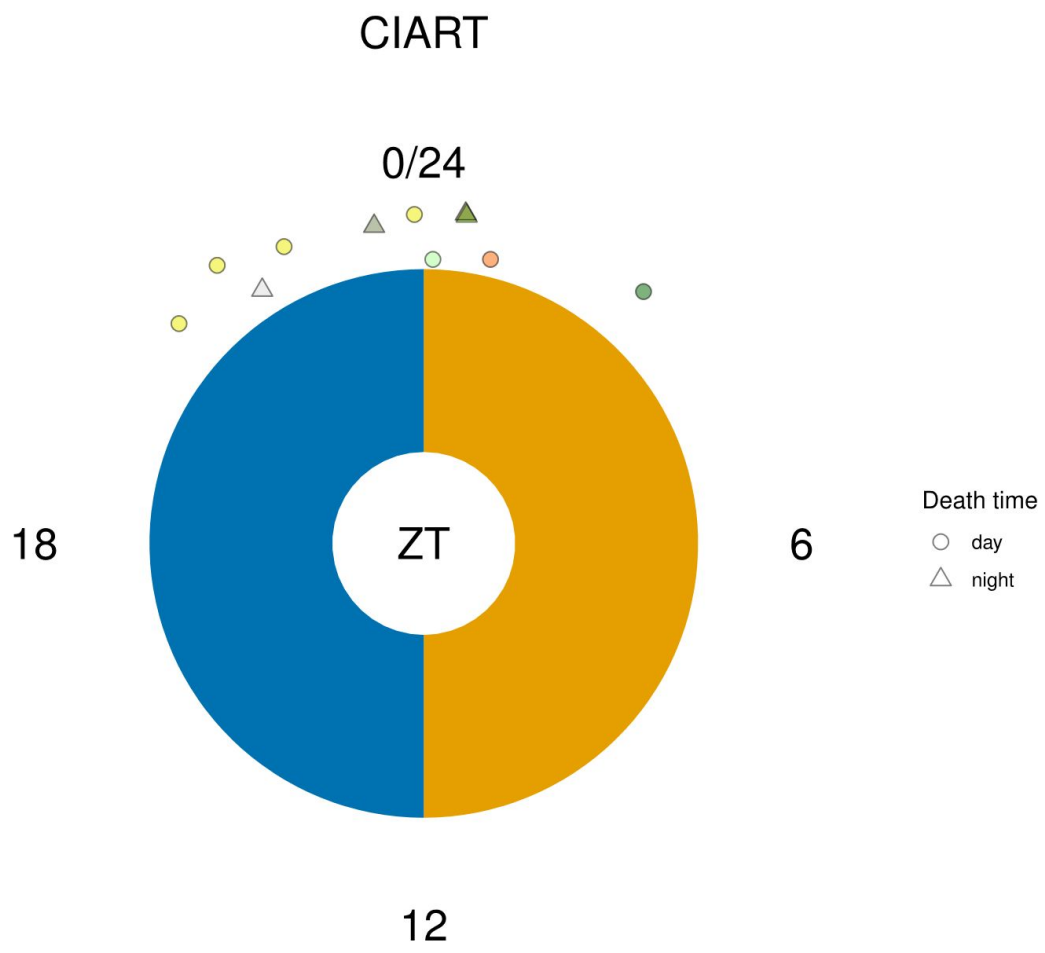


Figure R2.S4. Ciart phase of baboon is during the dusk, which is why we detected a shift for CIART phaste.

4.7 Landscape of seasonal variation across tissues

Seasonal changes and photoperiods are known to affect the brain, glandular systems, guts, and gonadal-brain axis. The role of circadian core clock genes in encoding the photoperiod has been suggested [133]. By comparing the transcriptome of tissues across seasons, we observed a high number of genes differentially up or down-regulated during fall and spring in the testis (7658 in the fall and 6101 in the spring). On the contrary, relatively fewer genes varied in winter and summer(figure R2.3A-D).

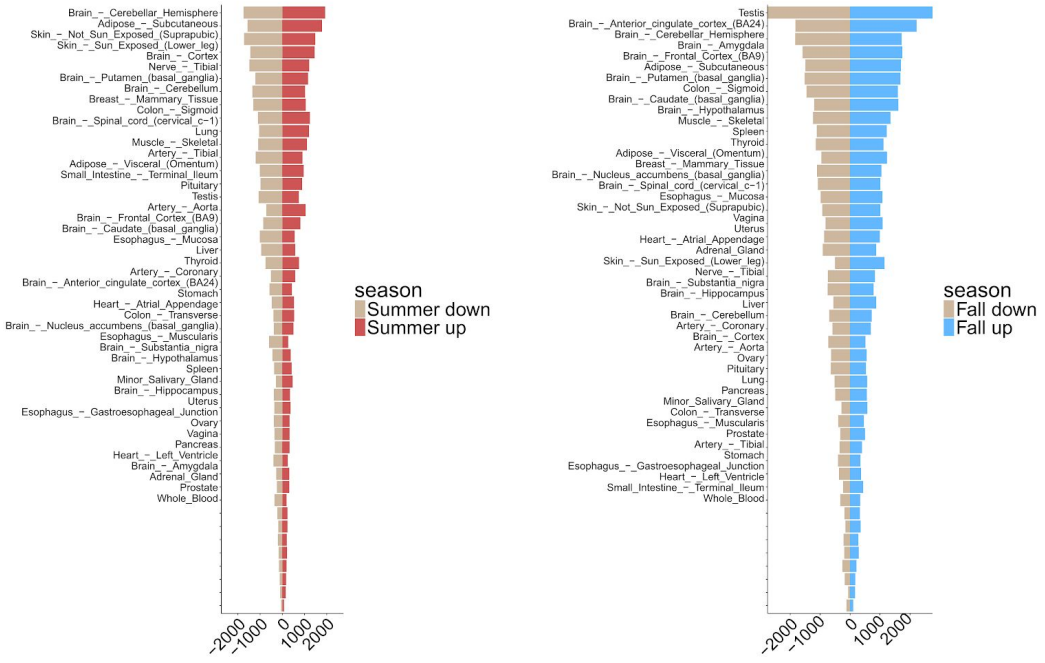


Figure R2.3A,B. Seasonal gene expression in summer and fall.

However, the movement of the earth around the sun could put pressure on a biological system, for example, via alteration of solar radiation. Solar radiation could affect the most exposed tissues. Thus, observing the two skin tissues exhibiting higher transcriptional changes mainly in summer and spring and spring compared to winter and fall, might suggest the influence of solar radiation (figure R2.3A-D). If solar radiation was the prime factor, we should observe a higher number of melanocyte in summer in the skin because the UV radiation could induce melanocyte differentiation. Therefore, we investigated the cell type composition of the tissues using Xcell [134] and, interestingly, we found melanocyte enrichment in summer only in exposed skin (FDR<0.1, figure R2.3E).

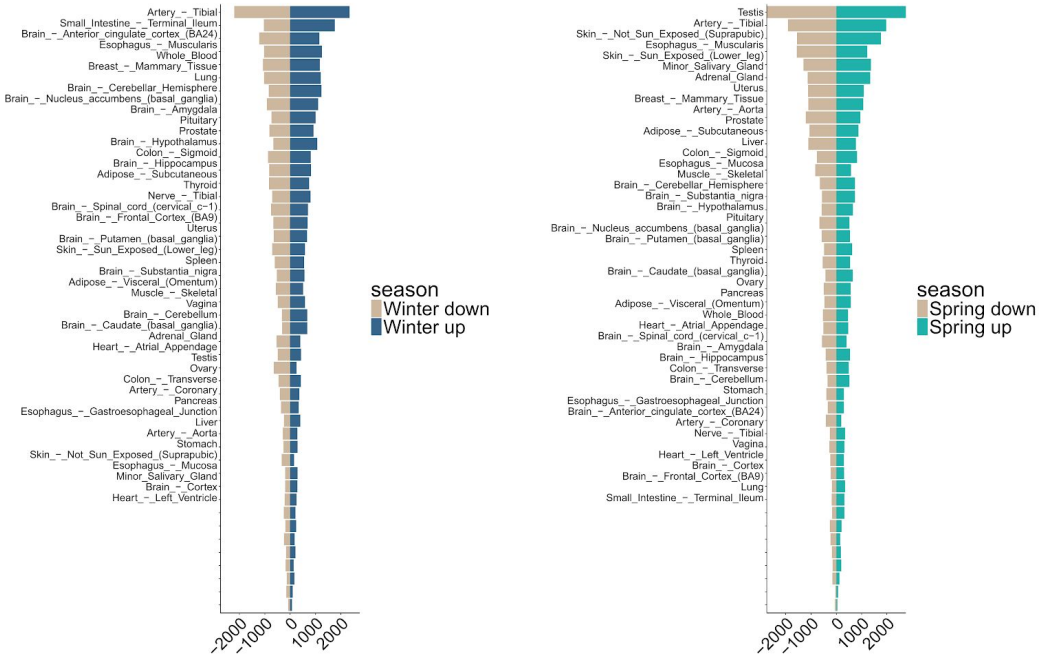


Figure R2.3C,D.Gene expression across tissues in winter and spring.

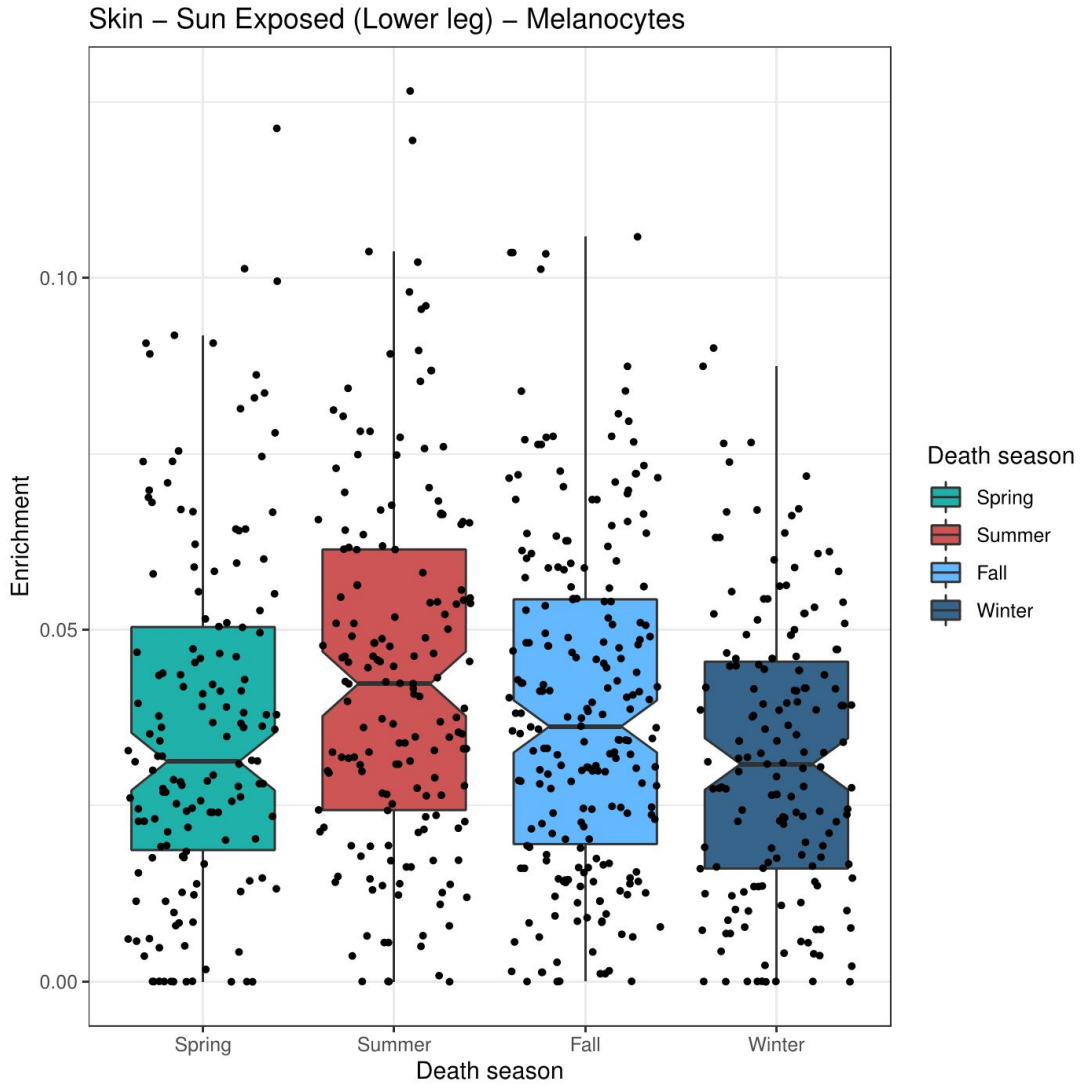


Figure R2.3E. An increase of melanocytes only in the skin with sun exposition.

Next, we investigated whether alternative splicing patterns also change across seasons. We noticed an overall trend in exon skipping in the colder season as opposed to exon inclusion in warmer seasons. Overall, we observed a higher number of exonic events in the brain. Particularly, cerebellar hemisphere exhibited a widespread increase in the inclusion of more than 800 exons (figure R2.4A-D).

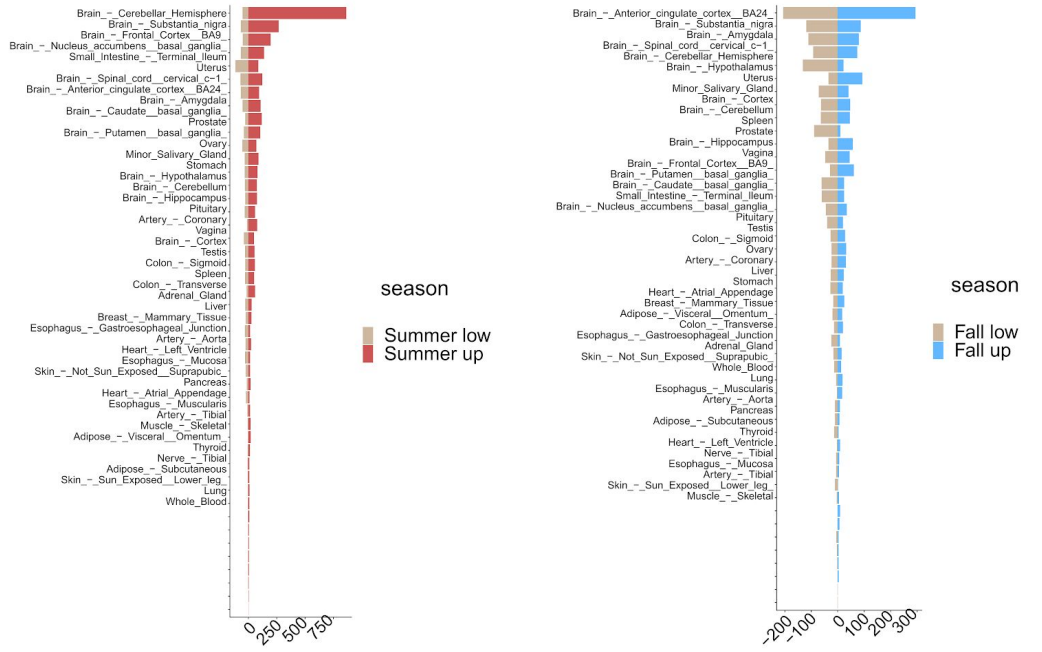


Figure R2.4A,B. Alternative splicing of exons in summer and fall.

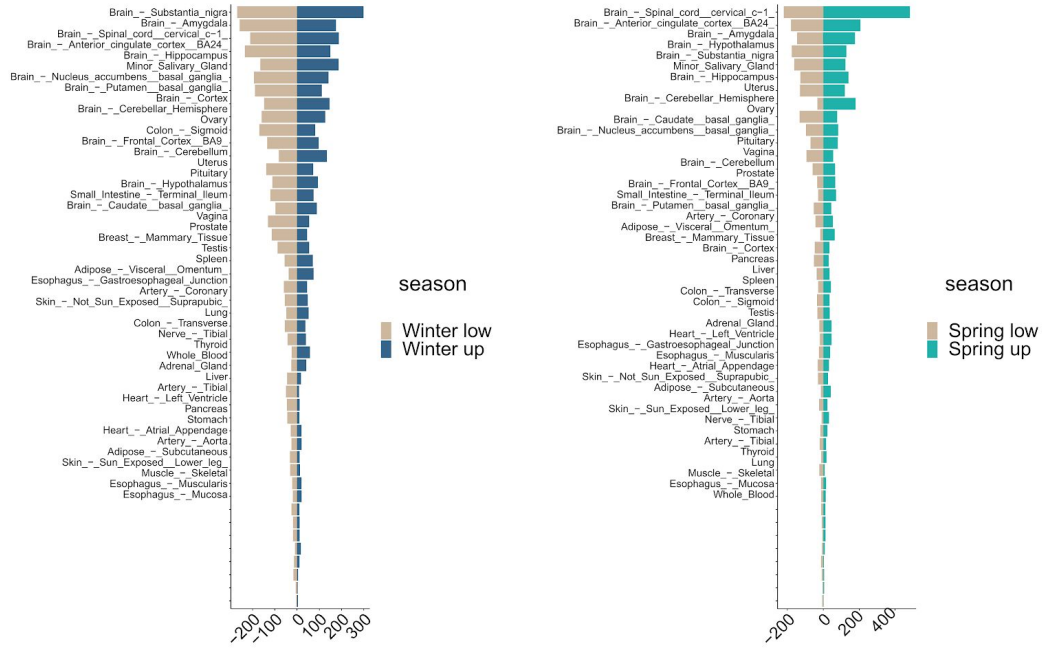


Figure R2.4C,D. Alternative splicing of exons in winter and spring.

4.8 PTBP1’s upregulation in fall caused a plunge in the microexon inclusion in anterior cingulate cortex.

After observing high dynamic of alternative exon splicing across seasons in the brain, we investigated the inclusion of microexons (3-27bp) because microexons shown to be very brain-specific [34]. We chose 1279 microexons, from which we found 304 to vary across seasons in at least one tissue. The majority of them only varied in the brain. We were surprised to observe anterior cingulate cortex exhibited a strong down-regulation of microexons in fall compared to all seasons across all tissues(Fig R2.5A).

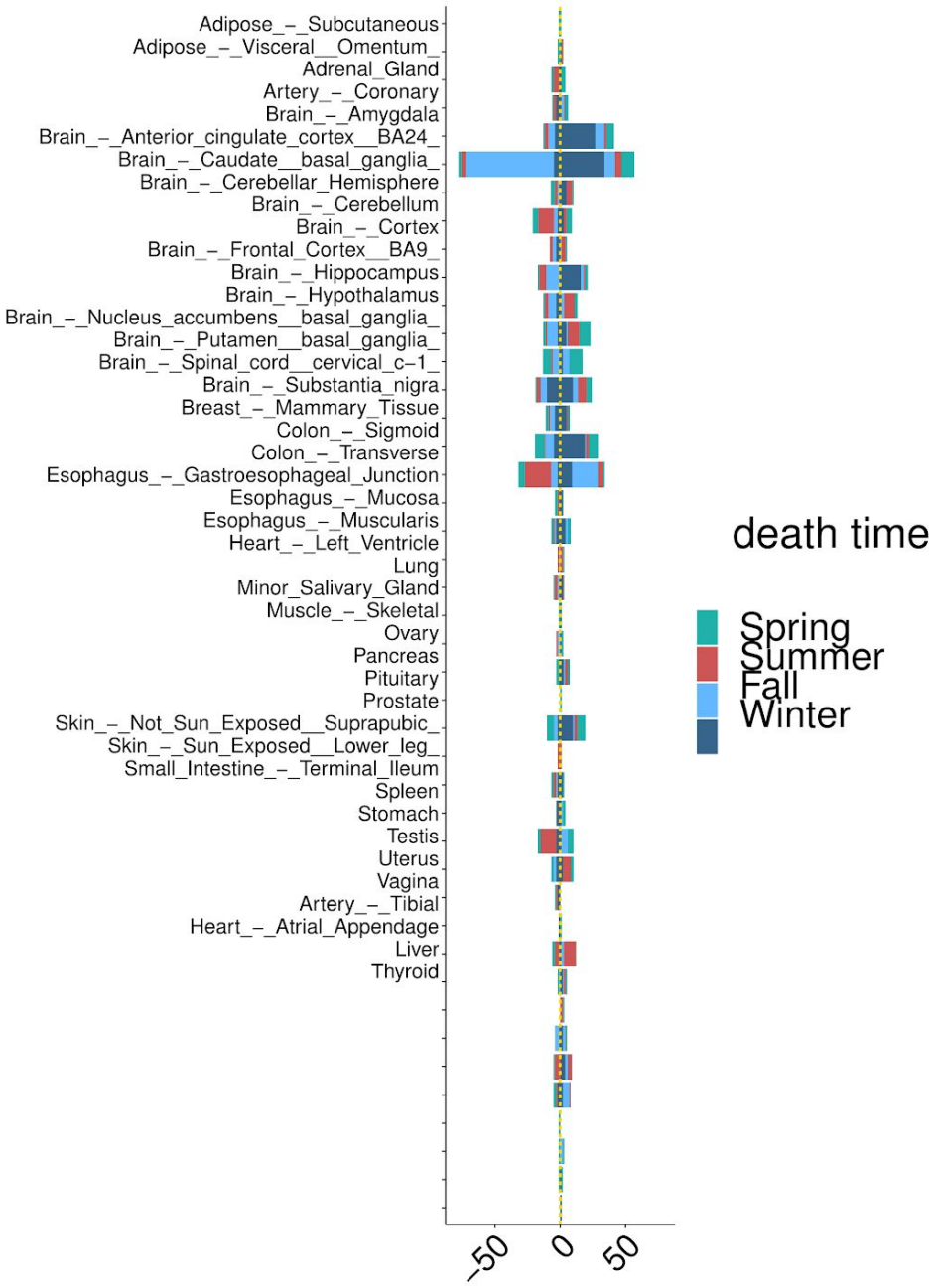


Figure R2.5A. The inclusion pattern of microexons, negative numbers are down regulation, and positives are up-regulation.

The pattern of microexons inclusion and exclusion did not appear to follow the temperature of the season. Instead, we hypothesise that the expression of splicing regulator of microexons could explain the plunge of microexons' inclusion (figure R2.S5).

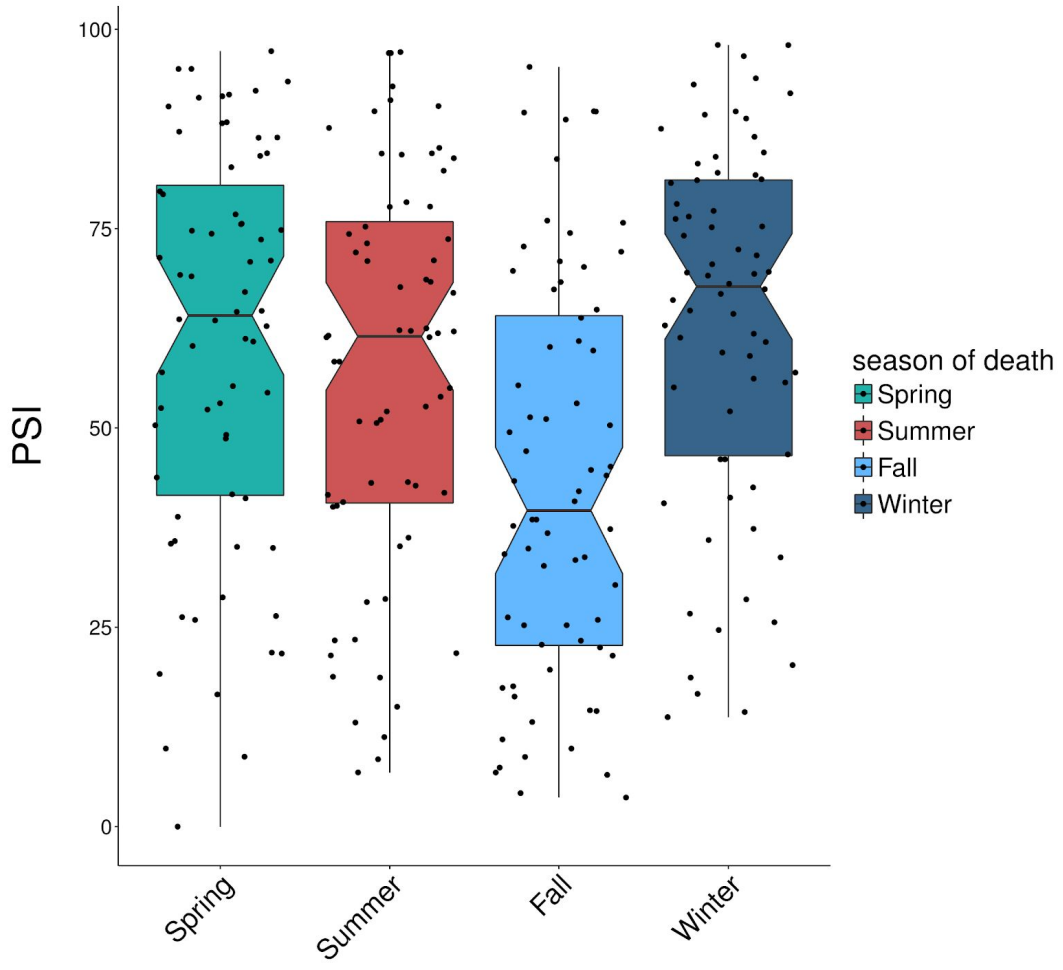


Figure R2.S5. Metaplot of significantly suppressed microexons in fall in anterior cingulate cortex.

We did not find evidence for down-regulation of SRRM4 (positive regulator of microexons) in the anterior cingulate cortex in fall. Thus, we investigate if PTBP1 (the inhibitor of microexons) could explain this pattern. Interestingly, we found PTBP1 up-regulated in the anterior cingulate cortex in fall (only in the anterior cingulate cortex in the brain, and only in fall), FDR<0.1, Fig R2.5B).

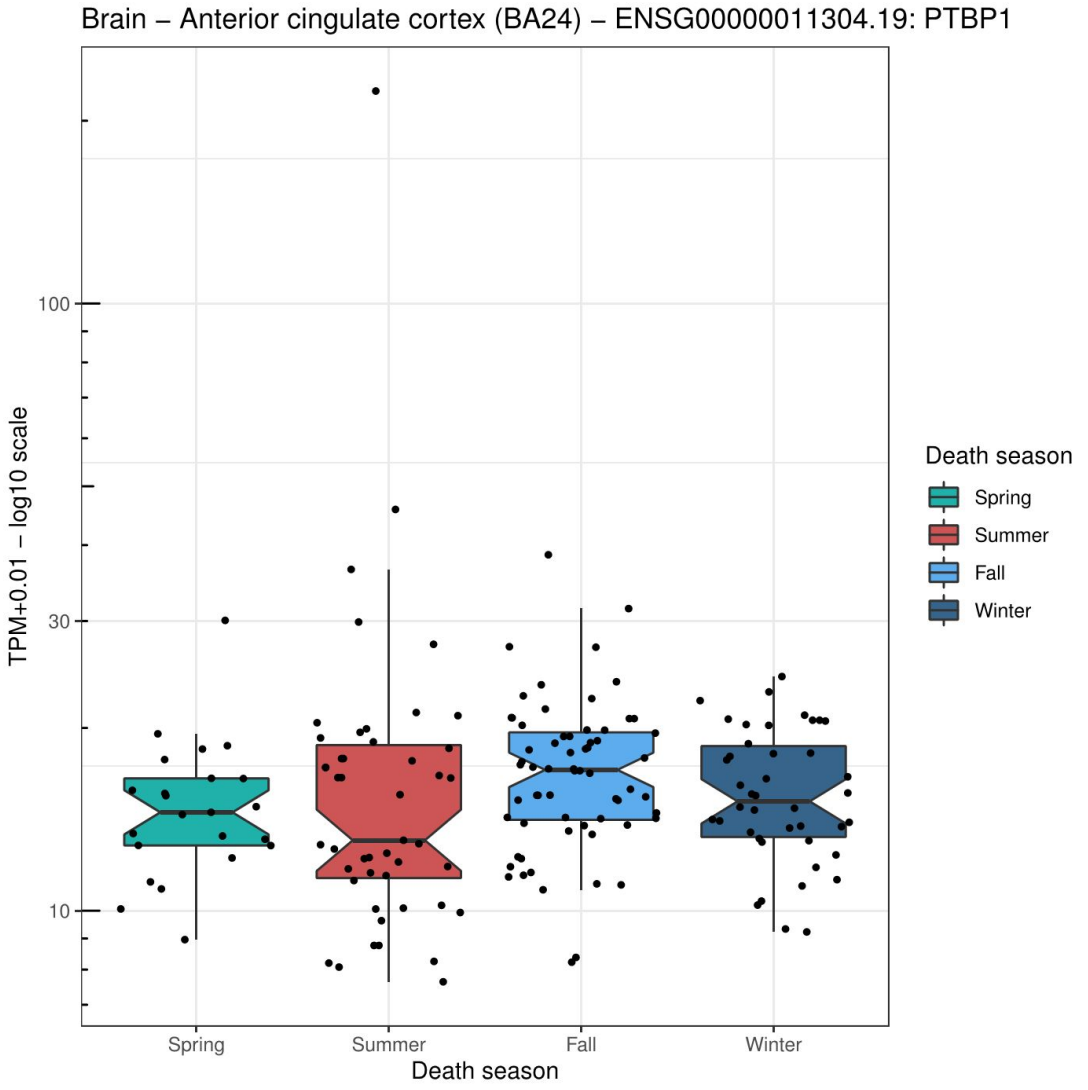


Figure R2.5B. PTBP1 up-regulation in fall.

Next, we used the microexons whose inclusion changed upon SRRM4 knockout in HEK293 cells [135], and found from 31 out of 67 down-regulated microexons in anterior cingulate cortex to be affected by SRRM4 KO. Moreover, we found those microexons whose inclusion drastically changed in fall, were also affected more drastically by the SRRM4 knockout (Fig R2.5C).

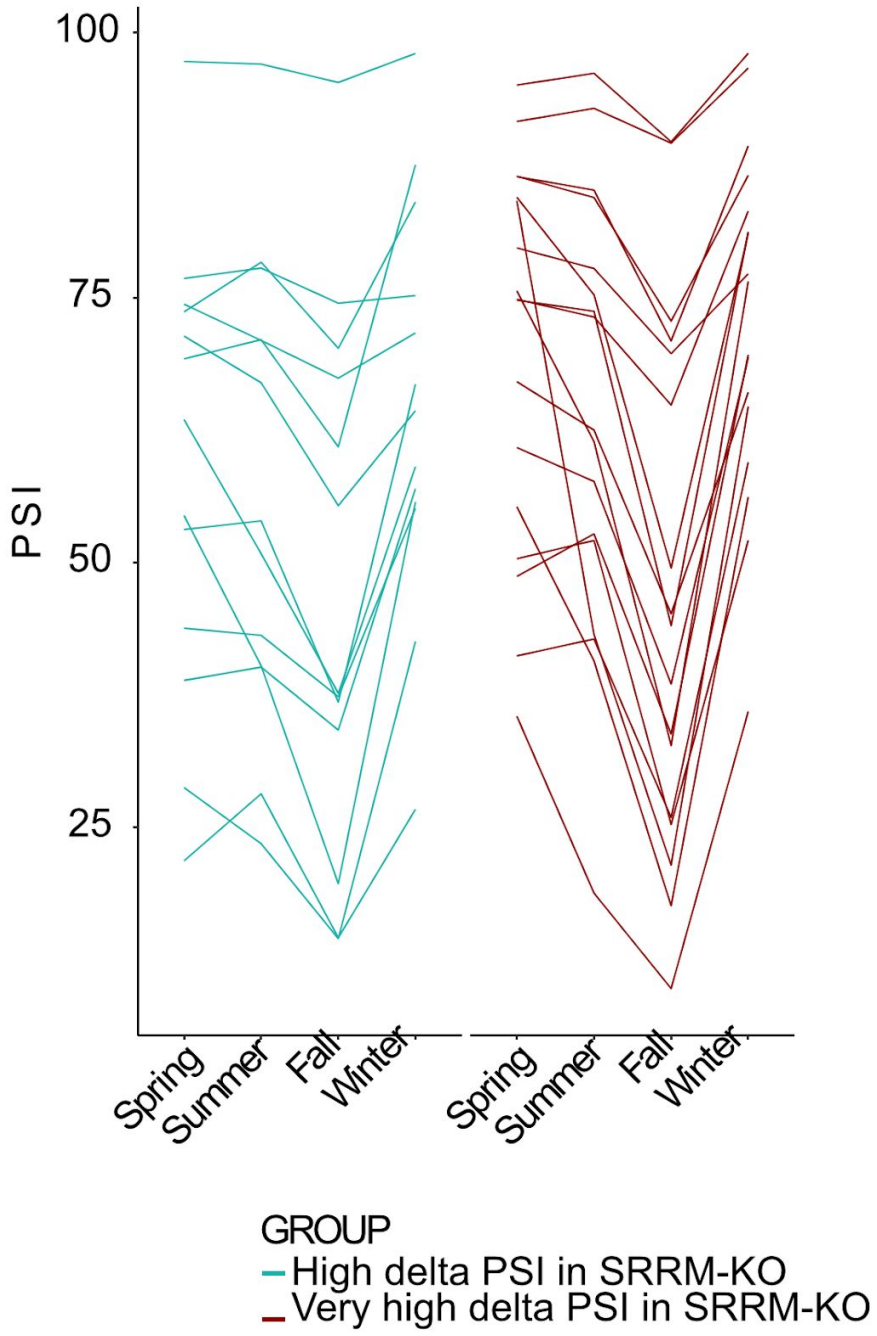


Figure R2.5C. Reduction of microexons inclusion in fall follows the same pattern in SRRM-KD.The difference between the exons whose value drastically changes in the SRRM-KD experiment, also changed drastically in fall (p-value, wilcoxon test 0.01).

Furthermore, we found MYC, the only annotated transcription factor regulator of PTBP1, increased during fall.Observing a high number of transcriptional and post-transcriptional and post-transcriptional events to change in fall, might suggest substantial physiological changes in anterior cingulate cortex. Anterior cingulate cortex

was shown to be involved in seasonal mood disorder and depression (figure R2.2B, figure R2.3B).

4.9 Material and methods

4.9.1 RNA-seq

The RNA-seq data was generated, mapped and quantified by the GTEx consortium (GTEx v8). Tissues with less than 100 individuals sequenced were discarded from the analyses (Kidney - Medulla, Cervix - Ectocervix , Fallopian Tube, Cervix - Endocervix, Bladder and Kidney - Cortex) as well as two cell lines (Cells - EBV-transformed lymphocytes and Cells - Cultured fibroblasts). For the Whole Blood tissue, all samples taken pre-mortem were discarded from the analysis for homogeneity with the others tissues. This lead to a list of 46 tissues, 932 individuals and 16,151 samples after filtering for samples with available covariates for the differential analyses.

4.9.2 Classification of the time of death as day or night

Using the time of death provided by the GTEx consortium, individuals have been classified between dead during the day or dead during the night if their time of death was falling into the following intervals [8:00-17:00[and [21:00-5:00[respectively. Other times of death have been discarded to avoid taking into account any RNA-seq samples coming from people where the day status was unsure, i.e. twilight. Using the above classification, 351, 315 and 266 individuals have been classified as day, night and twilight respectively, of which only 666 individuals (11,527 samples) were used for differential expression.

4.9.3 Season of death

The season of death was provided by the GTEx consortium with 282, 190, 221 and 239 individuals dying in Fall, Spring, Summer and Winter respectively.

4.9.4 Differential gene expression between day and night

Differential expression between day and night was performed separately on the 46 tissues using samples from the 666 individuals classified as day or night, going from 98 samples (Uterus) to 560 samples (Muscle - Skeletal). Genes were filtered per tissues, removing all genes with a median TPM value below 1. The analyses were run using R and the *voom-limma* pipeline [127,136] and TMM normalisation method from

edgeR [137,138]. The significance of the time of death was assessed correcting for the following covariates:

- Sex, if the tissue is not a sex-specific tissue;
- Age;
- BMI;
- Ischemic time.

All genes with an associate p-value below 0.05 were considered as differentially expressed between night and day.

4.9.5 Differential gene expression between seasons

Differential expression between season was performed separately on the 46 tissues using samples from the 932 individuals, going from 139 samples (Brain - Substantia nigra) to 789 samples (Muscle - Skeletal). Genes were filtered and the analyses was run as described for 'Differential expression between day and night'. The covariates used were the following:

- Time of death as classified as: day, night or twilight;
- Sex, if the tissue is not a sex-specific tissue;
- Age;
- BMI;
- Ischemic time.

The effect of each season was assessed by comparing one season against all the others, leading to four differential expression analyses, one for each season, for the whole dataset. All genes with an associated p-value below 0.05 were considered as circannual.

4.9.6 Exon PSI values calculation

The PSI values for the exons were obtained from Sodaei *et al*.

4.9.7 Differential exon inclusion between day and night

The differential exon inclusion has been computed as described above for the genes between day and night by replacing the *voom/limma* pipeline by a linear model using the 'lm()' function from R. The same covariates were used for the models.

4.9.8 Differential exon inclusion between seasons

The differential exon inclusion has been computed as described above for the genes between seasons by replacing the *voom/limma* pipeline by a linear model using the 'lm()' function from R. The same covariates were used for the models.

4.9.9 Baboon dataset

The baboon results were downloaded from the [90]. We extracted the significant genes using the same threshold as Mure *et al.* (pvalue <= 0.05). To compare GTEx results with baboon's results, we classified them as day if the phase of the oscillation was between 0ZT and 12ZT and night if the phase was between 12ZT and 24ZT. Common tissues between GTEx tissues list and the baboon tissues list have been manually curated.

4.9.10 Mouse dataset

The mouse circadian genes were extracted from the CirGRDB database (ref) which include genes already defined as circadian and filter for a p-value below 0.05. The genes have been classified into day and night for the baboon genes. Common tissues between GTEx tissues list and the mouse tissues list have been manually curated.

4.10 Discussion

Samples that allow researcher studying circadian rhythm on human tissues are rare and for the seasonal variation even more so. Here we tried to report what we observed in human tissues' transcriptome during the daily cycle and across seasons. We found that genes that ubiquitously expressed differently between day and night across tissues were the core clock genes. This observation both validated our analysis and opened new windows in understanding circadian regulation across mammalian tissues. Firstly, we confirmed that the diurnal pattern of gene expression of CLOCK in human tissues was less universal NPAS2, similar to baboon and mouse, suggesting that NPAS2 may have a dominant role in heterodimerization with ARNTL. Furthermore, we found 111 genes that found to fluctuate daily on par with CLOCK regarding the number of tissues. However, we do not know if they are upstream regulators of tissue-specific circadian regulation or most of them are near downstream of core clock genes. Observing that not all the core clock genes fluctuate between day and night, might suggest a tissue-specific combinatorial transcriptional regulation. However, one might explain this observation by a lack of data points across a large

window of time. The evidence from mouse and baboon data suggested that it is probably not the case.

Seasonal variation in gene expression could be adaptive, for example, for hibernating animals. However, many observed changes in the transcriptome could be a response to a different environment. For instance, changes that occur because of a change in temperature without any advantages. One accurate way, for an organism, to predict seasonal variation is via photoperiod. If, for example, photoperiod is regulated by the core clock genes, as some theoretical model suggested [139], then core clock genes could act as a robust system to also predict the seasonality. However, more experiments are needed to elucidate the possible scenarios. An animal living in an equatorial area would not experience a photoperiod shift during a year. However, it might experience a change in the availability of the food because of the rain, for instance [140]. Thus, even equatorial animals need to regulate their physiology, such as breeding and migration correspondingly.

Observing a large number of genes along with microexons down-regulation in the anterior cingulate cortex, a subregion which is suspected to be involved in mood changes might explain some symptoms of seasonal affective disorders. However, more experiments are needed to clarify, perhaps, other consequences of these robust changes. For instance, whether human cognition might be affected.

Observing skipping of events more in colder seasons compared to warmer seasons might suggest a temperature-dependent mechanism of splicing, as it has been shown in mammals previously. Interestingly, we also observed more exon inclusion during the day compared to the night.

5 Chapter 5 - General discussion

5.1 Mainly muscle, brain, and testis use a part of human transcriptome that no other tissue uses

Measuring unique sequences of a given tissue is an estimate of how specialized a tissue needs to be to function by controlling the mRNA at the level of transcription or post-transcription. It is different from comparing the tissues in pairwise. Pairwise comparison gives us an estimate of how two tissues differ from each other. For example, proteins that are expressed in the liver but not in the skin. We showed that each tissue uses part of the available transcriptome, but chiefly testis and brain (to some extent, muscles and heart) use a part of the transcriptome that other tissues do not use. Brain and muscles use the alternative splicing part of the transcriptomes predominantly to create their unique sequences, whereas testis uses a unique set of both transcriptional and post-transcriptional events.

We found testis-unique exons were much less conserved than brain and muscle unique exons, suggesting that testis-unique alternative exons had evolved recently. Analysis of Khaitovich et al. previously compared the gene expression of five tissues in human and chimpanzee supports more substantial divergence of human testis amidst tissues. Interestingly, testis-unique exons not only were less conserved, but the majority also disrupted the reading frame of their gene. We observed the testis-unique exons of mouse exhibiting the same pattern, although the exons themselves were not conserved. We were surprised to observe a tissue with unique disruptive events. Was NMD suppressed?

A recent study tried to relate the increase of UPF3A (a suppressor of NMD) in spermatocytes and necessity of NMD-regulation in spermatogenesis. They hypothesized that in spermatocytes UPF3A in the absence of UPF3B (the antagonist) could suppress the NMD. Even though we observed that the ratio of UPF3A to UPF3B was the highest in testis (very similar to some brain subregions), we did not observe any correlation between mouse testis-unique disruptive exons and UPF3A knockout experiment. Moreover, in a review (ref) pointed out that UPF3A only showed a minor capacity in stabilizing mRNAs in P19 cells. We observed that testis-unique exon splicing pattern does not match with the profile of UPF1 knockout experiments. Furthermore, despite expectations, we observed an increase in the gene expression of disrupted testis-unique exons compared to genes with disrupted non-unique exons. This meant NMD did not affect the disrupted gene of the testis-unique exons, and the increase of the expression might be as a result of the protection from NMD, which could result in less digestion and hence higher expression compared to unprotected transcripts. NMD occurs during the translation [141], and Kleene lab frequently showed that YBX2 could be a potential suppressor of translation for some spermatids mRNAs. Thus, we speculated YBX2 or similar factors could be involved in the suppression of translation and NMD. The expression pattern of YBX2 is conserved across mammals,

and it peaks in spermatocytes and round spermatids, exactly where we observed testis-unique disrupted exon inclusion. However, we also consider a scenario that involves many factors to explain testis-unique disruptive exon inclusion.

We found that in brain alternative splicing of microexons dominantly —compared to splicing of longer exons or transcription of genes— drove the unique specialization of neural tissues. Those microexons are conserved across vertebrates, thus indicating their crucial role in the survival of an organism.

5.2 Individual-unique differences exposed the genetic differences at the splice-sites and splicing motifs.

By choosing the individual-unique set, which was not variable across tissues, we ultimately reduce the events that could be affected by the environment of a given tissue. Therefore, most of the individual-unique exons are more robust in the face of medium changes in their surroundings, and hence more likely to be controlled by genetics. Interestingly, we found that individuals-unique exons occur mostly by disruption of splice-sites and motifs; thus, it is more likely that a constitutive exon turns into an alternative rather than the other way around. We observed the same pattern in the human-unique exons; changes from constitutive to alternative constituted the majority of changes. Likely, some of the past individual-unique variants, which occurred after human-chimpanzee divergence got fixed in the population. Furthermore, as we observed that the number of constitutive to alternative across species were dependent on the genetic divergence, and genetic divergence is dependent on the time of the divergence. Therefore, it is likely that these changes of exons from constitutive to alternative were overall neutral. However, even neutral changes could be influenced differently upon introduction of a novel change either in the genome or in the environment. Thus, neutral changes could provide the raw material for selection to act. Furthermore, some of those changes are not neutral, as has been shown in other species.

Overall, the universal differences between other mammals and humans were rare. Then it raises the question of whether those transcriptomic differences could create differences we observe at the level of species? The picture we have is very blurry as we do not know the component of genetics, environment, and their interaction that give rise into inter-species functional, and behavioural differences. Studies on translation and transcription between human and chimpanzee indicated that not all the transcriptional changes would lead to translational differences. This means the number of protein differences is even lower than that of the mRNAs.

However, other factors reduced the numbers in our study. We chose a very stringent set by specifying a robust expression in at least five tissues. Moreover, we required

that orthologous exons to exist in mouse, rhesus monkey, and chimpanzee, which will decrease the number of comparable exons. Besides the methodological differences, some protein changes could occur without a change in the concentration of mRNA, for instance, by replacing an amino acid inside an exon. Sometimes subtle changes at the genomic or proteomic level could give rise to significant effects, for example, during the development of an organism. More importantly, a recent study, done by Illumina, showed how weak cryptic splice site mutation could confer an exon with tissue specificity. Therefore, human-chimpanzee differences could have arisen from weak changes that affect transcriptional or post-transcriptional output in tissue or developmental specific manner while the overall transcriptome changes may remain similar because of averaging over all the tissues.

5.3 Circadian and circannual transcriptional and post-transcriptional variation across tissues

We found that most of the human transcriptome not only regulated for each tissue but also it is regulated diurnally and seasonally. Environmental factors could cause imbalances in the physiology of an organism. However, some of the observed transcriptomic differences could be adaptive responses. Most of the time, it is hard to distinguish the two scenarios. However, even in the absence of rhythmic environment fruit flies had a circadian rhythm in their eclosion, which persisted even in an arrhythmic environment after 600 generations. This experiment suggested that periodicity of eclosion was necessary for the fruit flies survival. For example, evolutionary studies on flies showed that a neutral trait persists 100-200 generations, and even faster (10-20 generations) for a neutral trait that has a high cost. Furthermore, sleep-wake cycle remained conserved in almost all the animal kingdom. So far, the core clock genes are involved in controlling this rhythm in animals.

We detected known core clock genes in a high number of tissues; this validates our method in detecting transcriptional events that differ diurnally. Moreover, we added a series of genes that oscillated in more than 12 tissues to human circadian profile. Some of those genes could be regulated at the downstream of clock transcription; some might have their role as a regulator. Interestingly, even though we detected a circadian profile for all the tissues for the core clock genes, testis, ovary, and vagina were the exceptions; none of core clock genes was found changing between day and night. Therefore, it seems each tissue uses part of this rhythmic transcriptome. We also observed the same pattern in baboon and mouse study.

Our analysis showed how human and baboon’s core clock genes peaked at roughly the same time, indicating a conserved pattern of oscillation between the two diurnal primates. Mouse, on the other hand, showed an almost complete reversion, e.g.

ARNTL and NPAS2 were more transcribed during the day, suggesting a plasticity of core clock genes in the adaptation of an organism to the rhythms of the day.

Furthermore, we observed that more than 50% of the genes oscillated diurnally, and strikingly the amount overlap between the circadian genes of a pair of tissues was very low. However, it is still not clear how this tissue-specificity arose. We still do not know if it is likely that all the genes follow some oscillation but with different periods. For example, had we had a narrow window in a well-designed experiment, we could detect more genes oscillate with different ranges of periods? A more extended period compared to the daily cycle is a seasonal cycle, with a yearly period. One way that season is related to the daily rhythms of an organism is by photoperiodism. However, it is still not clear how photoperiod in mammals works. We observed some of the core clock genes up or down-regulated across seasons, but so far, we do not have evidence that core clock genes could be responsible for the observed numerous changes.

Interestingly, we observed a high number of genes to change in testis during fall and spring, and we observed clock genes to be up or down-regulated in the testis in those two seasons. More startling, we did not detect circadian oscillation in the core clock genes in testis. For example, one could say we observed core clock genes in tissues with a large number of genes because simply more genes were found to be affected in that season in a particular tissue. Furthermore, we observed a general pattern in the splicing of exons, in the warmer months, exon inclusion was dominant, and in the colder seasons the exon skipping. Moreover, we observed a large number of changes in the brain during the fall.

We investigated the profile of mircoexons across seasons in the brain because we already knew that the majority of brain-unique events were microexons. Strikingly, we observed that many brain-unique exons (the majority of them were microexons) decreased in fall in the anterior cingulate cortex. It is unknown whether this effect on microexons of the anterior cingulate cortex is an adaptive response or as a result of homeostatic imbalance. We still do not know what the functional consequences of this suppression are. However, it seems the overexpression of PTBP1 caused this widespread suppression of microexons in fall. Given that anterior cingulate cortex has been known to affect mood and motivation, and its potential role in seasonal affective disorders [142,143], those observations might connect the large transcriptional and microexons splicing pattern in the anterior cortex to mood disorder.

6 Chapter 6 - Conclusion

6.1 Conclusion

Tissue-unique transcriptional regulation mainly occurred in testis. Alternative splicing provided unique sequences to brain, testis, and muscles; thus, these tissues uses part of the human transcriptome that no other tissue uses.

The majority of brain-unique exons were microexons, so microeoxns are the part of the unique transcriptome that brain uses.

Round spermatids unique transcriptional features drove the majority of testis-unique exons, whereas neurons were responsible for brain-unique exons.

Testis-unique exons were not conserved compared to brain-unique exons, across mammalian lineages. Furthermore, most of the testis-unique exons found to disrupt the open reading frame.

Both human-unique and individual-unique splicing pattern of exons followed a general trend from constitutive to alternative. Moreover, the number of species-unique exon skipping events is correspondent to the amount of genetic distance, and hence time. Neutral model of evolution supports these observations.

Human-unique exon splicing coincided with individual-unique alternative splicing patterns, and this support a neutral mode of exon splicing evolution, similar to gene expression.

Except for testis, ovary, and vagina, all the tissues had a circadian signature for at least one of the core clock genes. Even the pattern of core clock oscillation seems to be tissue-specific, which leads the more divergent downstream events.

Seasonal variation of transcriptome showed unexpectedly notable changes in the brain in fall, in the testis in spring and fall. Furthermore, alternative splicing exhibited a general trend for skipping in the cold and higher inclusion in the warmer seasons.

We observed a high number of microexons suppressed during fall in anterior cingulate cortex. We found that PRBP1 could be the cause. Interestingly, anterior cingulate cortex is a subregion in the brain involved in mood changes.

6.2 Appendix

List of publication during the PhD

Published

Javier Tapial, Kevin CH Ha, Timothy Sterne-Weiler, André Gohr, Ulrich Braunschweig, Antonio Hermoso-Pulido, Mathieu Quesnel-Vallières, Jon Permanyer, **Reza Sodaei**, Yamile Marquez, Luca Cozzuto, Xinchen Wang, Melisa Gómez-Velázquez, Teresa Rayon, Miguel Manzanares, Julia Ponomarenko, Benjamin J Blencowe, Manuel Irimia.(2017). An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms.
Genome research 27 (10), 1759-1768

Pedro G Ferreira, Manuel Muñoz-Aguirre, Ferran Reverter, Caio P Sá Godinho, Abel Sousa, Alicia Amadoz, **Reza Sodaei**, Marta R Hidalgo, Dmitri Pervouchine, Jose Carbonell-Caballero, Ramil Nurtdinov, Alessandra Breschi, Raziel Amador, Patrícia Oliveira, Cankut Çubuk, Joao Curado, François Aguet, Carla Oliveira, Joaquin Dopazo, Michael Sammeth, Kristin G Ardlie, Roderic Guigó.The effects of death and post-mortem cold ischemia on human tissue transcriptomes.Nature communications 9 (1), 490.

Bioarchive

Robert Y Yang, Jie Quan, **Reza Sodaei**, Francois Aguet, Ayellet V Segre, John A Allen, Thomas A Lanz, Veronica Reinhart, Matthew Crawford, Samuel Hasson, Kristin G Ardlie, Roderic Guigo, Hualin S Xi, GTEx Consortium. A systematic survey of human tissue-specific gene expression and splicing reveals new opportunities for therapeutic target identification and evaluation.
bioRxiv, 311563

In preparation

Sodaei R, Permanyer J, Guigo R, Irimia M. Landscapes of unique human transcriptomes.

Sodaei R*, Wucher V*, Amador R, Martin D.G, Sadeghi I, Irimia M, Guigo R. Tissue-specific transcriptional and post-transcriptional circadian and seasonal variation in human.

7 Bibliography

1. Koonin EV, Wolf YI. Is evolution Darwinian or/and Lamarckian? Biol Direct. 2009;4: 42. doi:10.1186/1745-6150-4-42
2. Griffith F. The significance of pneumococcal types. J Hyg (Lond). 1928;27: 113–159. doi:10.1017/S0022172400031879
3. Hershey AD, Chase M. Independent functions of viral protein and nucleic acid

in growth of bacteriophage. J Gen Physiol. 1952;36: 39–56. doi:10.1085/jgp.36.1.39

4. Watson JD, Crick FH. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. 1953. Rev Invest Clin. 2003;55: 108–109.

5. Crick F. [Central dogma of mollecular biology]. Tsitologiia. 1971;13: 906–910.

6. Briggs R, King TJ. Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. Proc Natl Acad Sci USA. 1952;38: 455–463. doi:10.1073/pnas.38.5.455

7. Gurdon JB, Elsdale TR, Fischberg M. Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature. 1958;182: 64–65.

8. Gurdon JB. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. J Embryol Exp Morphol. 1962;10: 622–640.

9. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature. 2011;473: 337–342. doi:10.1038/nature10098

10. Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell. 2016;165: 535–550. doi:10.1016/j.cell.2016.03.014

11. Franks A, Airoldi E, Slavov N. Post-transcriptional regulation across human tissues. PLoS Comput Biol. 2017;13: e1005535. doi:10.1371/journal.pcbi.1005535

12. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45: 580–585. doi:10.1038/ng.2653

13. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347: 1260419. doi:10.1126/science.1260419

14. Chandramouli K, Qian P-Y. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. Hum Genomics Proteomics. 2009;2009. doi:10.4061/2009/239204

15. Perl K, Ushakov K, Pozniak Y, Yizhar-Barnea O, Bhonker Y, Shivatzki S, et al. Reduced changes in protein compared to mRNA levels across non-proliferating tissues. BMC Genomics. 2017;18: 305. doi:10.1186/s12864-017-3683-9

16. Taylor RC, Webb Robertson B-JM, Markillie LM, Serres MH, Linggi BE, Aldrich JT, et al. Changes in translational efficiency is a dominant regulatory mechanism in the environmental response of bacteria. Integr Biol (Camb). 2013;5: 1393–1406. doi:10.1039/c3ib40120k

17. Sonawane AR, Platig J, Fagny M, Chen C-Y, Paulson JN, Lopes-Ramos CM, et al. Understanding Tissue-Specific Gene Regulation. Cell Rep. 2017;21:

1077–1088. doi:10.1016/j.celrep.2017.10.001

18. Pontén F, Gry M, Fagerberg L, Lundberg E, Asplund A, Berglund L, et al. A global view of protein expression in human cells, tissues, and organs. Mol Syst Biol. 2009;5: 337. doi:10.1038/msb.2009.93

19. Wang D, Eraslan B, Wieland T, Hallström B, Hopf T, Zolg DP, et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. Mol Syst Biol. 2019;15: e8503. doi:10.15252/msb.20188503

20. Uhlén M, Hallström BM, Lindskog C, Mardinoglu A, Pontén F, Nielsen J. Transcriptomics resources of human tissues and organs. Mol Syst Biol. 2016;12: 862. doi:10.15252/msb.20155865

21. Berget SM, Moore C, Sharp PA. Spliced segments at the 5’ terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci USA. 1977;74: 3171–3175. doi:10.1073/pnas.74.8.3171

22. Berk AJ, Sharp PA. Structure of the adenovirus 2 early mRNAs. Cell. 1978;14: 695–711. doi:10.1016/0092-8674(78)90252-0

23. Modrek B, Resch A, Grasso C, Lee C. Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic Acids Res. 2001;29: 2850–2859. doi:10.1093/nar/29.13.2850

24. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456: 470–476. doi:10.1038/nature07509

25. Early P, Rogers J, Davis M, Calame K, Bond M, Wall R, et al. Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. Cell. 1980;20: 313–319. doi:10.1016/0092-8674(80)90617-0

26. Gracheva EO, Cordero-Morales JF, González-Carcacía JA, Ingolia NT, Manno C, Aranguren CI, et al. Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. Nature. 2011;476: 88–91. doi:10.1038/nature10245

27. Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, et al. Function of alternative splicing. Gene. 2013;514: 1–30. doi:10.1016/j.gene.2012.07.083

28. Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, et al. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell. 2000;101: 671–684. doi:10.1016/s0092-8674(00)80878-8

29. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature. 2007;446: 926–929. doi:10.1038/nature05676

30. Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative

splicing in the human transcriptome. Nucleic Acids Res. 2002;30: 3754–3766. doi:10.1093/nar/gkf492

31. Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. Genome Biol. 2004;5: R74. doi:10.1186/gb-2004-5-10-r74

32. Tapial J, Ha KCH, Sterne-Weiler T, Gohr A, Braunschweig U, Hermoso-Pulido A, et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. Genome Res. 2017;27: 1759–1768. doi:10.1101/gr.220962.117

33. Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep. 2013;3: 2179–2190. doi:10.1016/j.celrep.2013.05.031

34. Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M, et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell. 2014;159: 1511–1523. doi:10.1016/j.cell.2014.11.035

35. Zhang X, Chen MH, Wu X, Kodani A, Fan J, Doan R, et al. Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. Cell. 2016;166: 1147-1162.e15. doi:10.1016/j.cell.2016.07.025

36. Jacko M, Weyn-Vanhentenryck SM, Smerdon JW, Yan R, Feng H, Williams DJ, et al. Rbfox splicing factors promote neuronal maturation and axon initial segment assembly. Neuron. 2018;97: 853-868.e6. doi:10.1016/j.neuron.2018.01.020

37. Li YI, Sanchez-Pulido L, Haerty W, Ponting CP. RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. Genome Res. 2015;25: 1–13. doi:10.1101/gr.181990.114

38. Anderson MS, Venanzi ES, Klein L, Chen Z, Berzins SP, Turley SJ, et al. Projection of an immunological self shadow within the thymus by the aire protein. Science. 2002;298: 1395–1401. doi:10.1126/science.1075958

39. Kyewski B, Klein L. A central role for central tolerance. Annu Rev Immunol. 2006;24: 571–606. doi:10.1146/annurev.immunol.23.021704.115601

40. Magalhães DAR, Silveira ELV, Junta CM, Sandrin-Garcia P, Fachin AL, Donadi EA, et al. Promiscuous gene expression in the thymus: the root of central tolerance. Clin Dev Immunol. 2006;13: 81–99. doi:10.1080/17402520600877091

41. Danan-Gotthold M, Guyon C, Giraud M, Levanon EY, Abramson J. Extensive RNA editing and splicing increase immune self-representation diversity in medullary thymic epithelial cells. Genome Biol. 2016;17: 219. doi:10.1186/s13059-016-1079-9

42. Charlesworth B, Charlesworth D. Population genetics from 1966 to 2016. Heredity. 2017;118: 2–9. doi:10.1038/hdy.2016.55

43. Hamilton WD. The genetical evolution of social behaviour. I. J Theor Biol. 1964;7: 1–16. doi:10.1016/0022-5193(64)90038-4

44. Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, Akey JM. Gene-expression variation within and among human populations. Am J Hum Genet. 2007;80: 502–509. doi:10.1086/512017

45. Li J, Liu Y, Kim T, Min R, Zhang Z. Gene expression variability within and between human populations and implications toward disease susceptibility. PLoS Comput Biol. 2010;6. doi:10.1371/journal.pcbi.1000910

46. Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet. 2008;24: 408–415. doi:10.1016/j.tig.2008.06.001

47. Peng S, Deyssenroth MA, Di Narzo AF, Lambertini L, Marsit CJ, Chen J, et al. Expression quantitative trait loci (eQTLs) in human placentas suggest developmental origins of complex diseases. Hum Mol Genet. 2017;26: 3432–3441. doi:10.1093/hmg/ddx265

48. O'Brien HE, Hannon E, Hill MJ, Toste CC, Robertson MJ, Morgan JE, et al. Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. Genome Biol. 2018;19: 194. doi:10.1186/s13059-018-1567-1

49. Brown CD, Mangravite LM, Engelhardt BE. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. PLoS Genet. 2013;9: e1003649. doi:10.1371/journal.pgen.1003649

50. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. The human transcriptome across tissues and individuals. Science. 2015;348: 660–665. doi:10.1126/science.aaa0355

51. Li YI, van de Geijn B, Raj A, Knowles DA, Petti AA, Golan D, et al. RNA splicing is a primary link between genetic variation and disease. Science. 2016;352: 600–604. doi:10.1126/science.aad9417

52. Takata A, Matsumoto N, Kato T. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. Nat Commun. 2017;8: 14519. doi:10.1038/ncomms14519

53. Fadista J, Vikman P, Laakso EO, Mollet IG, Esguerra JL, Taneera J, et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci USA. 2014;111: 13924–13929. doi:10.1073/pnas.1402665111

54. Zhao K, Lu Z, Park JW, Zhou Q, Xing Y. GLiMMPS: robust statistical model for regulatory variation of alternative splicing using RNA-seq data. Genome Biol.

2013;14: R74. doi:10.1186/gb-2013-14-7-r74

55. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell. 2019;176: 535-548.e24. doi:10.1016/j.cell.2018.12.015

56. Kimura M. Evolutionary rate at the molecular level. Nature. 1968;217: 624–626. doi:10.1038/217624a0

57. Rifkin SA, Kim J, White KP. Evolution of gene expression in the Drosophila melanogaster subgroup. Nat Genet. 2003;33: 138–144. doi:10.1038/ng1086

58. Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM. Common pattern of evolution of gene expression level and protein sequence in Drosophila. Mol Biol Evol. 2004;21: 1308–1317. doi:10.1093/molbev/msh128

59. Lemos B, Meiklejohn CD, Cáceres M, Hartl DL. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. Evolution. 2005;59: 126–137.

60. Fay JC, McCullough HL, Sniegowski PD, Eisen MB. Population genetic variation in gene expression is associated with phenotypic variation in Saccharomyces cerevisiae. Genome Biol. 2004;5: R26. doi:10.1186/gb-2004-5-4-r26

61. Whitehead A, Crawford DL. Neutral and adaptive variation in gene expression. Proc Natl Acad Sci USA. 2006;103: 5425–5430. doi:10.1073/pnas.0507648103

62. Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, et al. A neutral model of transcriptome evolution. PLoS Biol. 2004;2: E132. doi:10.1371/journal.pbio.0020132

63. Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP. Expression profiling in primates reveals a rapid evolution of human transcription factors. Nature. 2006;440: 242–245. doi:10.1038/nature04559

64. Blekhman R, Oshlack A, Chabot AE, Smyth GK, Gilad Y. Gene regulation in primates evolves under tissue-specific selection pressures. PLoS Genet. 2008;4: e1000271. doi:10.1371/journal.pgen.1000271

65. Khaitovich P, Enard W, Lachmann M, Pääbo S. Evolution of primate gene expression. Nat Rev Genet. 2006;7: 693–702. doi:10.1038/nrg1940

66. Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, et al. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science. 2005;309: 1850–1854. doi:10.1126/science.1108296

67. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, et al. The evolutionary landscape of alternative splicing in vertebrate species. Science. 2012;338: 1587–1593. doi:10.1126/science.1230612

68. Merkin J, Russell C, Chen P, Burge CB. Evolutionary dynamics of gene and

isoform regulation in Mammalian tissues. Science. 2012;338: 1593–1599. doi:10.1126/science.1228186

69. Xiong J, Jiang X, Ditsiou A, Gao Y, Sun J, Lowenstein ED, et al. Predominant patterns of splicing evolution on human, chimpanzee and macaque evolutionary lineages. Hum Mol Genet. 2018;27: 1474–1485. doi:10.1093/hmg/ddy058

70. Sargent TD, Dawid IB. Differential gene expression in the gastrula of Xenopus laevis. Science. 1983;222: 135–139. doi:10.1126/science.6688681

71. Dunlap JC. Molecular bases for circadian clocks. Cell. 1999;96: 271–290. doi:10.1016/s0092-8674(00)80566-8

72. Ebisuya M, Briscoe J. What does time mean in development? Development. 2018;145. doi:10.1242/dev.164368

73. Avendaño MS, Vazquez MJ, Tena-Sempere M. Disentangling puberty: novel neuroendocrine pathways and mechanisms for the control of mammalian puberty. Hum Reprod Update. 2017;23: 737–763. doi:10.1093/humupd/dmx025

74. Pittendrigh CS. Temporal organization: reflections of a Darwinian clock-watcher. Annu Rev Physiol. 1993;55: 16–54. doi:10.1146/annurev.ph.55.030193.000313

75. Bass J. Circadian topology of metabolism. Nature. 2012;491: 348–356. doi:10.1038/nature11704

76. Rosbash M. The implications of multiple circadian clock origins. PLoS Biol. 2009;7: e62. doi:10.1371/journal.pbio.1000062

77. Simons MJP. The evolution of the cyanobacterial posttranslational clock from a primitive “phoscillator”. J Biol Rhythms. 2009;24: 175–182. doi:10.1177/0748730409333953

78. Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M, et al. Coordinated transcription of key pathways in the mouse by the circadian clock. Cell. 2002;109: 307–320. doi:10.1016/s0092-8674(02)00722-5

79. Potter GDM, Skene DJ, Arendt J, Cade JE, Grant PJ, Hardie LJ. Circadian rhythm and sleep disruption: causes, metabolic consequences, and countermeasures. Endocr Rev. 2016;37: 584–608. doi:10.1210/er.2016-1083

80. Elkhenany H, AlOkda A, El-Badawy A, El-Badri N. Tissue regeneration: Impact of sleep on stem cell regenerative capacity. Life Sci. 2018;214: 51–61. doi:10.1016/j.lfs.2018.10.057

81. Konopka RJ, Benzer S. Clock mutants of Drosophila melanogaster. Proc Natl Acad Sci USA. 1971;68: 2112–2116. doi:10.1073/pnas.68.9.2112

82. Shearman LP, Sriram S, Weaver DR, Maywood ES, Chaves I, Zheng B, et al. Interacting molecular loops in the mammalian circadian clock. Science.

2000;288: 1013–1019. doi:10.1126/science.288.5468.1013

83. Vitaterna MH, King DP, Chang AM, Kornhauser JM, Lowrey PL, McDonald JD, et al. Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. Science. 1994;264: 719–725. doi:10.1126/science.8171325

84. Minegishi S, Sagami I, Negi S, Kano K, Kitagishi H. Circadian clock disruption by selective removal of endogenous carbon monoxide. Sci Rep. 2018;8: 11996. doi:10.1038/s41598-018-30425-6

85. O'Neill JS, Reddy AB. Circadian clocks in human red blood cells. Nature. 2011;469: 498–503. doi:10.1038/nature09702

86. Edgar RS, Green EW, Zhao Y, van Ooijen G, Olmedo M, Qin X, et al. Peroxiredoxins are conserved markers of circadian rhythms. Nature. 2012;485: 459–464. doi:10.1038/nature11088

87. Abe M, Herzog ED, Yamazaki S, Straume M, Tei H, Sakaki Y, et al. Circadian rhythms in isolated brain regions. J Neurosci. 2002;22: 350–356.

88. Debruyne JP, Noton E, Lambert CM, Maywood ES, Weaver DR, Reppert SM. A clock shock: mouse CLOCK is not required for circadian oscillator function. Neuron. 2006;50: 465–477. doi:10.1016/j.neuron.2006.03.041

89. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: implications for biology and medicine. Proc Natl Acad Sci USA. 2014;111: 16219–16224. doi:10.1073/pnas.1408886111

90. Mure LS, Le HD, Benegiamo G, Chang MW, Rios L, Jillani N, et al. Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. Science. 2018;359. doi:10.1126/science.aao0318

91. Ruben MD, Wu G, Smith DF, Schmidt RE, Francey LJ, Lee YY, et al. A database of tissue-specific rhythmically expressed human genes has potential applications in circadian medicine. Sci Transl Med. 2018;10. doi:10.1126/scitranslmed.aat8806

92. Preußner M, Goldammer G, Neumann A, Haltenhof T, Rautenstrauch P, Müller-McNicoll M, et al. Body temperature cycles control rhythmic alternative splicing in mammals. Mol Cell. 2017;67: 433-446.e4. doi:10.1016/j.molcel.2017.06.006

93. McGlincy NJ, Valomon A, Chesham JE, Maywood ES, Hastings MH, Ule J. Regulation of alternative splicing by the circadian clock and food related cues. Genome Biol. 2012;13: R54. doi:10.1186/gb-2012-13-6-r54

94. El-Athman R, Knezevic D, Fuhr L, Relógio A. A Computational Analysis of Alternative Splicing across Mammalian Tissues Reveals Circadian and Ultradian Rhythms in Splicing Events. Int J Mol Sci. 2019;20. doi:10.3390/ijms20163977

95. Walton JC, Weil ZM, Nelson RJ. Influence of photoperiod on hormones,

behavior, and immune function. Front Neuroendocrinol. 2011;32: 303–319. doi:10.1016/j.yfrne.2010.12.003

96. Dopico XC, Evangelou M, Ferreira RC, Guo H, Pekalski ML, Smyth DJ, et al. Widespread seasonal gene expression reveals annual differences in human immunity and physiology. Nat Commun. 2015;6: 7000. doi:10.1038/ncomms8000

97. Xu B, Liu H, Su N, Kong G, Bao X, Li J, et al. Association between winter season and risk of death from cardiovascular diseases: a study in more than half a million inpatients in Beijing, China. BMC Cardiovasc Disord. 2013;13: 93. doi:10.1186/1471-2261-13-93

98. Pell JP, Cobbe SM. Seasonal variations in coronary heart disease. QJM. 1999;92: 689–696. doi:10.1093/qjmed/92.12.689

99. Pell JP, Sirel J, Marsden AK, Cobbe SM. Seasonal variations in out of hospital cardiopulmonary arrest. Heart. 1999;82: 680–683. doi:10.1136/hrt.82.6.680

100. Iikuni N, Nakajima A, Inoue E, Tanaka E, Okamoto H, Hara M, et al. What's in season for rheumatoid arthritis patients? Seasonal fluctuations in disease activity. Rheumatology (Oxford). 2007;46: 846–848. doi:10.1093/rheumatology/kel414

101. Moltchanova EV, Schreier N, Lammi N, Karvonen M. Seasonal variation of diagnosis of Type 1 diabetes mellitus in children worldwide. Diabet Med. 2009;26: 673–678. doi:10.1111/j.1464-5491.2009.02743.x

102. Choe YJ, Smit MA, Mermel LA. Seasonality of respiratory viruses and bacterial pathogens. Antimicrob Resist Infect Control. 2019;8: 125. doi:10.1186/s13756-019-0574-7

103. Fisman DN. Seasonality of infectious diseases. Annu Rev Public Health. 2007;28: 127–143. doi:10.1146/annurev.publhealth.28.021406.144128

104. Sharon G, Cruz NJ, Kang D-W, Gandal MJ, Wang B, Kim Y-M, et al. Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice. Cell. 2019;177: 1600-1618.e17. doi:10.1016/j.cell.2019.05.004

105. Li Q, Han Y, Dy ABC, Hagerman RJ. The gut microbiota and autism spectrum disorders. Front Cell Neurosci. 2017;11: 120. doi:10.3389/fncel.2017.00120

106. Owens N, McGorry PD. Seasonality of symptom onset in first-episode schizophrenia. Psychol Med. 2003;33: 163–167. doi:10.1017/s0033291702006712

107. Watson CG, Kucala T, Tilleskjor C, Jacobs L. Schizophrenic birth seasonality in relation to the incidence of infectious diseases and temperature extremes. Arch Gen Psychiatry. 1984;41: 85–90. doi:10.1001/archpsyc.1984.01790120089011

108. Torrey EF, Miller J, Rawlings R, Yolken RH. Seasonal birth patterns of

neurological disorders. Neuroepidemiology. 2000;19: 177–185. doi:10.1159/000026253

109. Philpot M, Rottenstein M, Burns A, Der G. Season of birth in Alzheimer's disease. Br J Psychiatry. 1989;155: 662–666. doi:10.1192/s000712500001816x

110. Lim ASP, Klein H-U, Yu L, Chibnik LB, Ali S, Xu J, et al. Diurnal and seasonal molecular rhythms in human neocortex and their relation to Alzheimer's disease. Nat Commun. 2017;8: 14931. doi:10.1038/ncomms14931

111. Mc Mahon B, Andersen SB, Madsen MK, Hjordt LV, Hageman I, Dam H, et al. Seasonal difference in brain serotonin transporter binding predicts symptom severity in patients with seasonal affective disorder. Brain. 2016;139: 1605–1614. doi:10.1093/brain/aww043

112. Carlsson A, Svennerholm L, Winblad B. Seasonal and circadian monoamine variations in human brains examined post mortem. Acta Psychiatr Scand. 1980;61: 75–85. doi:10.1111/acps.1980.61.s280.75

113. Guo J, Grow EJ, Mlcochova H, Maher GJ, Lindskog C, Nie X, et al. The adult human testis transcriptional cell atlas. Cell Res. 2018;28: 1141–1157. doi:10.1038/s41422-018-0099-2

114. Soumillon M. [Transcriptome complexity in germ cells]. Med Sci (Paris). 2013;29: 1077–1079. doi:10.1051/medsci/20132912005

115. Jégou B, Sankararaman S, Rolland AD, Reich D, Chalmel F. Meiotic genes are enriched in regions of reduced archaic ancestry. Mol Biol Evol. 2017;34: 1974–1980. doi:10.1093/molbev/msx141

116. Shum EY, Jones SH, Shao A, Dumdie J, Krause MD, Chan W-K, et al. The Antagonistic Gene Paralogs Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay. Cell. 2016;165: 382–395. doi:10.1016/j.cell.2016.02.046

117. Cullinane DL, Chowdhury TA, Kleene KC. Mechanisms of translational repression of the Smcp mRNA in round spermatids. Reproduction. 2015;149: 43–54. doi:10.1530/REP-14-0394

118. Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis T, et al. Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res. 2014;24: 1774–1786. doi:10.1101/gr.177790.114

119. Chen FC, Li WH. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am J Hum Genet. 2001;68: 444–456. doi:10.1086/318206

120. Glazko GV, Nei M. Estimation of divergence times for major lineages of primate species. Mol Biol Evol. 2003;20: 424–434. doi:10.1093/molbev/msg050

121. Nei M, Glazko GV. The Wilhelmine E. Key 2001 Invitational Lecture. Estimation

of divergence times for a few mammalian and several primate species. J Hered. 2002;93: 157–164. doi:10.1093/jhered/93.3.157

122. Peyrégne S, Boyle MJ, Dannemann M, Prüfer K. Detecting ancient positive selection in humans using extended lineage sorting. Genome Res. 2017; doi:10.1101/gr.219493.116

123. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 2019;47: W191–W198. doi:10.1093/nar/gkz369

124. Yang G, Paschos G, Curtis AM, Musiek ES, McLoughlin SC, FitzGerald GA. Knitting up the raveled sleave of care. Sci Transl Med. 2013;5: 212rv3. doi:10.1126/scitranslmed.3007225

125. Fisk AS, Tam SKE, Brown LA, Vyazovskiy VV, Bannerman DM, Peirson SN. Light and cognition: roles for circadian rhythms, sleep, and arousal. Front Neurol. 2018;9: 56. doi:10.3389/fneur.2018.00056

126. Kimura A, Ishida Y, Hayashi T, Nosaka M, Kondo T. Estimating time of death based on the biological clock. Int J Legal Med. 2011;125: 385–391. doi:10.1007/s00414-010-0527-4

127. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15: R29. doi:10.1186/gb-2014-15-2-r29

128. Qiu M-H, Vetrivelan R, Fuller PM, Lu J. Basal ganglia control of sleep-wake behavior and cortical activation. Eur J Neurosci. 2010;31: 499–507. doi:10.1111/j.1460-9568.2009.07062.x

129. Lazarus M, Chen J-F, Urade Y, Huang Z-L. Role of the basal ganglia in the control of sleep and wakefulness. Curr Opin Neurobiol. 2013;23: 780–785. doi:10.1016/j.conb.2013.02.001

130. Canto CB, Onuki Y, Bruinsma B, van der Werf YD, De Zeeuw CI. The Sleeping Cerebellum. Trends Neurosci. 2017;40: 309–323. doi:10.1016/j.tins.2017.03.001

131. Landgraf D, Wang LL, Diemer T, Welsh DK. NPAS2 compensates for loss of CLOCK in peripheral circadian oscillators. PLoS Genet. 2016;12: e1005882. doi:10.1371/journal.pgen.1005882

132. Li X, Shi L, Zhang K, Wei W, Liu Q, Mao F, et al. CirGRDB: a database for the genome-wide deciphering circadian genes and regulators. Nucleic Acids Res. 2018;46: D64–D70. doi:10.1093/nar/gkx944

133. Yoshikawa T, Inagaki NF, Takagi S, Kuroda S, Yamasaki M, Watanabe M, et al. Localization of photoperiod responsive circadian oscillators in the mouse suprachiasmatic nucleus. Sci Rep. 2017;7: 8210.

doi:10.1038/s41598-017-08186-5

134. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017;18: 220. doi:10.1186/s13059-017-1349-1

135. Torres-Méndez A, Bonnal S, Marquez Y, Roth J, Iglesias M, Permanyer J, et al. A novel protein domain in an ancestral splicing factor drove the evolution of neural microexons. Nat Ecol Evol. 2019;3: 691–701. doi:10.1038/s41559-019-0813-6

136. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43: e47. doi:10.1093/nar/gkv007

137. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;40: 4288–4297. doi:10.1093/nar/gks042

138. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26: 139–140. doi:10.1093/bioinformatics/btp616

139. Myung J, Pauls SD. Encoding seasonal information in a two-oscillator model of the multi-oscillator circadian clock. Eur J Neurosci. 2018;48: 2718–2727. doi:10.1111/ejn.13697

140. Nelson RJ, Badura LL, Goldman BD. Mechanisms of seasonal cycles of behavior. Annu Rev Psychol. 1990;41: 81–108. doi:10.1146/annurev.ps.41.020190.000501

141. Hoek TA, Khuperkar D, Lindeboom RGH, Sonneveld S, Verhagen BMP, Boersma S, et al. Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay. Mol Cell. 2019;75: 324-339.e11. doi:10.1016/j.molcel.2019.05.008

142. Praschak-Rieder N, Willeit M, Wilson AA, Houle S, Meyer JH. Seasonal variation in human brain serotonin transporter binding. Arch Gen Psychiatry. 2008;65: 1072–1078. doi:10.1001/archpsyc.65.9.1072

143. Tyrer AE, Levitan RD, Houle S, Wilson AA, Nobrega JN, Meyer JH. Increased seasonal variation in serotonin transporter binding in seasonal affective disorder. Neuropsychopharmacology. 2016;41: 2447–2454. doi:10.1038/npp.2016.54