Review Article

Tracks through the genome to physiological events

Diane Lipscombe[1], Jen Q. Pan[2] and Stephanie Schorge[3]

[1]*Department of Neuroscience, Brown University, Providence, RI, USA*
[2]*Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA*
[3]*University College London Institute of Neurology, London, UK*

Diane Lipscombe is Professor of Neuroscience and interim Director of the Brown Institute for Brain Science. Member-at-Large and Fellow of the American Association for the Advancement of Science, Diane received the 2012 Joan Mott Prize from the Physiological Society and the Harriet W. Sheridan Award for Distinguished Contribution to Teaching and Learning at Brown University. Diane Lipscombe is nationally recognized for her research on voltage-gated calcium ion channels from genes to behaviour, including cell-specific pre-mRNA processing. Diane Lipscombe obtained her BSc and PhD degrees in Pharmacology from University College London and completed her postdoctoral training in Physiology at Yale University and Stanford University Schools of Medicine.

New Findings

- **What is the topic of this review?**
  We discuss tools available to access genome-wide data sets that harbour cell-specific, brain region-specific and tissue-specific information on exon usage for several species, including humans. In this Review, we demonstrate how to access this information in genome databases and its enormous value to physiology.
- **What advances does it highlight?**
  The sheer scale of protein diversity that is possible from complex genes, including those that encode voltage-gated ion channels, is vast. But this choice is critical for a complete understanding of protein function in the most physiologically relevant context.

Many proteins of great interest to physiologists and neuroscientists are structurally complex and located in specialized subcellular domains, such as neuronal synapses and transverse tubules of muscle. Genes that encode these critical signalling molecules (receptors, ion channels, transporters, enzymes, cell adhesion molecules, cell–cell interaction proteins and cytoskeletal proteins) are similarly complex. Typically, these genes are large; human Dystrophin (*DMD*) encodes a cytoskeletal protein of muscle and it is the largest naturally occurring gene at a staggering 2.3 Mb. Large genes contain many non-coding introns and coding exons; human Titin (*TTN*), which encodes a protein essential for the assembly and functioning of vertebrate striated muscles, has over 350 exons and consequently has an enormous capacity to generate different forms of Titin mRNAs that have unique exon combinations. Functional and pharmacological differences among protein isoforms originating from the same gene may be subtle but nonetheless of critical physiological significance. Standard functional, immunological and pharmacological approaches, so useful for characterizing proteins encoded by different genes, typically fail to discriminate among splice isoforms of individual genes. Tools are now available to access genome-wide data sets that harbour cell-specific, brain region-specific and

tissue-specific information on exon usage for several species, including humans. In this Review, we demonstrate how to access this information in genome databases and its enormous value to physiology.

Introduction

The combination of cell-specific use of alternative promoters, alternative polyadenylation sites and cell-specific alternative splicing of pre-mRNA in the cell nucleus generates a wide range of mRNAs that may have different stabilities or encode functionally different proteins. During transcription and pre-mRNA processing, the spliceosome orchestrates intron removal and exon splicing. Appropriate caps (5′-7-methylguanylate) and tails [3′ poly-(A)] are added to form mature mRNAs. Cell-specific factors recruit transcriptional machinery in the cell nucleus to control the choice of alternative promoters and first exons, promote recognition or skipping of alternatively spliced exons by the spliceosome and control the choice of alternative polyadenylation sites. The large number of RNA variants that can be derived from a single gene is a powerful mechanism to support functional diversity among proteins without increasing genetic load. Cell-specific use of alternative exons, alternative promoters, RNA editing and alternative sites of polyadenylation in 3′ untranslated regions give rise to multiple mRNA isoforms according to cell and tissue type, developmental stage, activity level and disease state. These forms of RNA processing are no longer considered exceptional, but rather a property that is characteristic of the vast majority of multi-exon eukaryotic genes. Certain genes are subject to extensive alternative splicing, including *Drosophila Dscam* and mammalian neurexin (*Nrxn*) genes that have the potential to generate thousands of functionally different protein isoforms (Schmucker *et al.* 2000; Aoto *et al.* 2013; Lah *et al.* 2014).

In our discussions below, we focus on examples of tissue-specific regulation of exon choice that expand the proteome. However, tissue-specific control of certain disruptive alternative exons is also a cellular mechanism to turn on or off the expression of proteins downstream of transcriptional activity (Lareau & Brenner, 2015). Insertion or skipping of disruptive alternative exons in mRNAs shifts the open reading frame, introduces premature stop codons and results in mRNA degradation by nonsense-mediated decay (O'Brien *et al.* 2012; Eom *et al.* 2013; Yan *et al.* 2015; Lareau & Brenner, 2015).

The functional importance of alternative splicing has already been shown for a wide range of physiological processes, including the following: acquisition of cell phenotype during early differentiation; correct axonal targeting and synaptogenesis during brain development; sex determination in insects; muscle fibre type specification; synaptic scaling; cell signalling; pluripotent stem cell renewal; and, possibly, speciation (Aoto *et al.* 2013; Han *et al.* 2013; Jepson *et al.* 2013; Lipscombe *et al.* 2013a; Venables *et al.* 2013; Lah *et al.* 2014; Raj *et al.* 2014; Lu *et al.* 2014). Cell-specific pre-mRNA processing greatly expands the coding capacity of genes, and the resulting pool of mRNA splice isoforms in specific tissues carries a richer source of information about the function and the state of the cell than the genotype alone. The potential to produce functional diversity by regulated alternative splicing is now well appreciated, and the molecular mechanisms that regulate cell-specific and developmental stage-specific alternative splicing are under active investigation. Our understanding of where, when, why and how mRNA splice isoforms are expressed is currently incomplete, but the landscape is changing rapidly. Whole-transcriptome analyses of data sets derived from massively parallel sequencing are added to publicly available databases with high frequency, and contain sequences for expressed RNAs for an expanding number of human and mouse tissues, cells and diseases (Voineagu *et al.* 2011; Li *et al.* 2014; Martin *et al.* 2014; Zhang *et al.* 2014). The sheer scale of protein diversity that is possible from complex genes, including those that encode voltage-gated ion channels, can be daunting when it comes to selecting the most physiologically relevant protein splice isoform to study. Nonetheless, this choice is likely to be critical for a complete understanding of protein function – including subcellular targeting, activation, modulation and pharmacological sensitivity – in the most physiologically relevant context.

Identifying splice isoforms in specific cells and tissues

Identifying and manipulating protein isoforms in specific tissues is technically challenging because of their structural, pharmacological and functional similarities, the range of tissues and cell types where they are expressed, the short temporal expression pattern of certain splice isoforms and the short length of some alternative exons, which may encode only one or two amino acids. Where traditional physiological and pharmacological tools

cannot meet the challenge, large-scale, deep-sequencing methods now give unprecedented details about the mRNAs expressed in a variety of human tissues. Critically, these methods now provide comprehensive sequence coverage for complex genes, such as ion channels and other membrane proteins whose mRNAs may be present at relatively low levels. Using available visualization tools, it is now easier to gain substantial information on the transcribed sequences of genes of interest, including the major mRNA isoforms expressed in various mammalian tissues, the location of major splice junctions and the location and type of RNA and DNA regulatory elements. Such information can provide a useful starting point to develop hypotheses about the role of different protein isoforms in specific cells and tissues for direct testing *in vitro* and *in vivo*. We demonstrate the utility of these genome-wide deep-sequencing data sets by illustrating how to visualize alternatively spliced exons of the *Cacna1* genes, which encode the functional core $\alpha 1$ subunits of voltage-gated calcium (Ca_V) channels in mammals.

Site of alternative splicing in the *Cacna1b* gene

Two decades ago, we set out to identify the molecular correlates of different $Ca_V 2.2$ channel activities (N-type currents) observed in patch-clamp recordings from neurons. The $Ca_V 2.2$ channels support many different cell functions, including excitation–secretion coupling at many different synapses in the nervous system, and are encoded by the *Cacna1b* gene. At that time, genomic sequence information was limited. We, and others, therefore relied heavily on time-intensive targeted RT-PCR amplification across exon junctions, based on the available *Cacna1b* sequence, and using mRNA isolated from microdissected brain regions, different peripheral ganglia and different tissues. We identified four sites of cell-specific alterative splicing in *Cacna1b* based on the presence of $Ca_V 2.2$ mRNA isoforms as well as alternative sites of polyadenylation in the 3′ untranslated region that generates $Ca_V 2.2$ mRNAs with different stabilities in neurons (Schorge *et al.* 1999). In subsequent studies, we showed that cell-specific control of alternative pre-mRNA splicing positions $Ca_V 2.2$ splice isoforms with unique biophysical properties and sensitivities to modulation by G-protein-coupled receptors in specific subpopulations of neurons and brain regions (Lipscombe *et al.* 2013*b*). Today these same $Ca_V 2.2$ mRNA isoforms can be identified in minutes by viewing $Ca_V 2.2$ RNA-seq data alignments and the conservation track of *Cacna1b* homologues across a range of vertebrates. This type of integrated visualization of data sets is nicely illustrated using the UCSC Genome Browser (Kent *et al.* 2002), for four cell type-specific alternatively spliced exons in *Cacna1b* (Fig. 1A–1D; see below).

Tools to visualize tissue-specific isoforms from genome-wide expression data sets

Massively parallel sequencing technology (next-generation sequencing) has reduced the cost of DNA sequencing dramatically (Lander, 2011), and many genome-wide expression data sets have been produced for a range of tissues and species and deposited to the public database. At the same time, interactive informatics tools have been developed to help researchers access, align and visualize any genome-wide information that is annotated by genome co-ordinate positions for species whose genomes have been sequenced and reference genome assembly and co-ordinates established. The UCSC Genome Browser and the Integrated Genomics Viewer (IGV) visualization tools offer different ways to explore large, heterogeneous, integrated genomic data sets from different species. While the UCSC Genome Browser is a Web application that offers the pre-integration of many data sets (Kent *et al.* 2002; http://genome.ucsc.edu/), IGV is optimized to provide high-performance data visualization as a stand-alone software application (IGV, http://www.broadinstitute.org/igv/; Thorvaldsdóttir *et al.* 2013, 2015). Both tools can upload custom data sets for visualization and both provide access to the Encyclopedia of DNA elements (ENCODE; Birney *et al.* 2007). Genotype-Tissue Expression project (GTEx; http://www.gtexportal.org/; GTEx, 2013) is a resource that can be used to visualize human gene expression and regulation for multiple tissues from a database collected and analysed within the GTEx project. The type of information highlighted here represents only a glimpse of what can be accessed from sequence databases using these tools; new sequence information is added regularly and new tools are being developed constantly.

UCSC Genome Browser. In the UCSC Genome Browser, the user can zoom and scroll over genomic locations on the chromosomes and display genome-wide annotation data sets ('tracks'), including mRNA alignments, gene-expression data and epigenetic regulatory markers, viewed beneath genome co-ordinate positions (Kent *et al.* 2002). The annotation tracks allow rapid vertical integration to visualize different types of information aligned to the same genome co-ordinates. The following description is best read while viewing the UCSC Genome Browser (see legend to Fig. 1). The UCSC Genome Browser contains > 10 major categories of genome annotations. For example, 'Expression and Regulation' is particularly useful for viewing cell- and tissue-specific events. Under each category, various tracks are available for viewing, and each track is derived from specially formatted files that can be read by the Genome Browser. These large genome-wide data sets have been preloaded at the UCSC server and

can be configured by the user to display individual tracks of interest at several different resolutions (dense, squish, pack and full). For example, the 'Burge-RNA seq' track located under the 'Expression and Regulation' tab contains RNA-seq data from five human cell lines and nine human tissues (Wang *et al.* 2008) that can be displayed by selecting a view option (dense, squish, pack and full). In Fig. 1, we show an output from the UCSC Genome Browser to demonstrate its value for visualizing tissue-specific alternative exons in *Cacna1b*. Previously
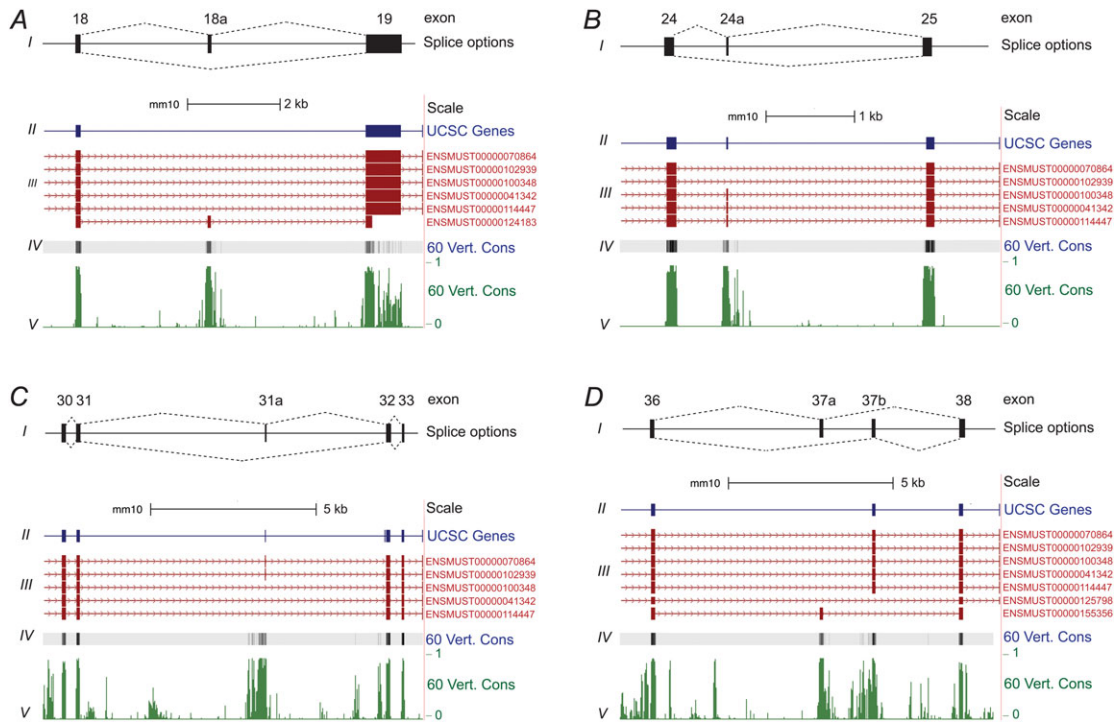

Figure 1. Visualizing various tracks using UCSC Genome Browser reveals the location of four alternatively spliced exons of mouse *Cacna1b* based on mouse genome version Dec. 2011 (GRCm38/mm10; http://genome.ucsc.edu/cgi-bin/hgGateway; Kent *et al.* 2002)
The genomic regions flanking alternatively spliced exons (18a, 24a, 31a and 37a/37b) of *Cacna1b* are shown in *A–D*. In each panel, scale bars indicate the size of the genomic region shown. Horizontal layers I–IV represent different display options, as follows: layer I, possible splice patterns in the region of *Cacna1b*; and layers II–V, tracks of UCSC Genes (blue), Ensemble Genes (maroon), and two different display options for Vertebrate Conservation (grey and green). Exons are shown as rectangles (layer I), and the direction of transcription is indicated by small arrows (layer III). The conservation track (layers IV and V) displays the PhastCon scores (from 0 to 1) calculated based on the genome sequence alignment of 60 vertebrates. In each panel, an alternatively spliced exon is captured in at least one Ensemble transcript. Not all alternatively spliced exons appear in mm10 UCSC Genes (or Refseq Genes track, not shown), but they align perfectly with peaks of highest conservation in the Vertebrate Conservation track. The following steps will recreate the display shown. (i) In the UCSC Genome Browser, choose 'Genomes' top left, group 'mammal', genome 'mouse', assembly 'GRCm38/mm10', search term '*Cacna1b*', location chr2:24603889-24763152, submit. (ii) Scroll down to 'Genes and Genes Predictions' header category, set 'Ensembl genes' tab to 'full' and 'UCSC Genes' tabs to 'dense' and set all other tabs to 'hide'. (iii) Scroll down to Comparative Genomics header, set 'Conservation' to display 'full' and click on the 'Conservation' link; here you can select subtracks by clade, select 'phastCons' scores in 'full', and 'hide' other features, further restrict the PhastCons scores to '60 Vert. Cons' in the subtrack lists, 'submit'. The complete *Cacna1b* gene is displayed. *A–D* are zoomed in to resolve regions that contain alternatively spliced exons. To recreate each panel type in locations: chr2:24678405-24686581 (18a region; *A*); chr2:24653647-24657925 (24a region; *B*); chr2:24638102-24649580 (31a region; *C*); and chr2:24621473-24632972 (37 region; *D*). Co-ordinates of exons are as follows: chr2:24682976-24683038 (18a); chr2:24656711-24656722 (24a); chr2:24642853-24642858 (e31a); chr2:24626823-24626919 (37a); and chr2:24625238-24625334 (37b). Alternatively spliced exons are flanked by consensus dinucleodide splice junction motifs AG and GT. Functions and tissue-specific distributions of these alternative exons have been described by our laboratory (Lipscombe *et al.* 2013*a*).

characterized alternatively spliced exons 18a, 24a, 31a, 37a and 37b in *Cacna1b* are captured by visualizing *UCSC Genes* and *Ensembl Genes* tracks aligned to the *60-vertebrate Conservation* track (Fig. 1). Notably, all five alternative exons are found in all vertebrate *Cacna1b* genes and captured in tracks such as '*Other RefSeq*' (non-human), '*Human mRNAs*' and '*Spliced ESTs*', but none is documented in the human reference RNA (human RefSeq) for the human genome (data not shown). Even exon 31a, which is only six nucleotides, located within a 10–12 kb intron and expressed at very low levels in mammalian brains, is captured in at least one *Ensembl* transcript (Fig. 1).

In Fig. 2, we show the 5′ end of *Cacna1c*, the gene that encodes dihydropyridine-sensitive $Ca_V1.2$ channels (high-voltage-activated L-type currents) to illustrate tissue-specific choice of promoters and first exons. $Ca_V1.2$ channels are expressed throughout the body, including the heart, smooth muscle, endocrine cells and neurons. They couple excitation to muscle contraction in heart and smooth muscle, to vesicle release in endocrine cells and to postsynaptic calcium-mediated synaptic plasticity and gene expression in neurons. Consistent with their very different cellular roles, unique $Ca_V1.2$ channel isoforms and exons dominate in different types of cells and tissues. The tissue-specific choice of alternative promoters of mouse *Cacna1c* can be inferred from visualizing tissue-specific histone modifications displayed in ChIP-seq track from the track '*ENCODE/Ludwig Institute for Cancer Research*' (LICR; ENCODE, 2012) in parallel with RNA-seq data from different mouse tissues (Fig. 2). Here, we show the *H3K4me3* subtrack that is predictive of active promoters, but this track can also be used to display an extensive number of *cis*-regulatory elements genome wide for a range of mouse tissue and cell lines (ENCODE, 2012).

Integrated Genomics Viewer. The IGV is a stand-alone application that is optimized for high-performance data visualization and offers functionality such as displaying splicing and exon junctions quantitatively (Robinson *et al.* 2011; Thorvaldsdóttir *et al.* 2013, 2015; Katz *et al.* 2015). The IGV comes with a set of highly used publicly available data sets, including ENCODE. Custom genome annotation data sets can be loaded from a server or local files. Many large-scale RNA-seq projects have produced mRNA expression data sets across multiple tissues and multiple cell lines. For example, human BodyMap 2.0 data from Illumina® can be downloaded, and it contains deep-sequencing alignments and coverage for RNA-seq derived from 16 human tissues (http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513/). Human BodyMap 2.0 data can be visualized within IGV. Sashimi plots and junction box displays provide quantitative visualizations of tissue-specific expression patterns of alternative splice junctions, as shown in Fig. 3.

The GTEx portal. The GTEx portal is an atlas of gene expression and regulation across multiple human tissues (GTEx, 2013). Since 2012, 3797 tissues from 150 donors have been collected and genotyped, and there is a target of gathering data from 900 donors and ∼20,000 tissue samples. GTEx is especially rich in mRNA sequences from different brain regions (13 subregions), lending greater resolution to region-specific expression patterns. Analysed data from GTEx are available to download using a simple user interface that generates an output showing both expression levels and frequency of exon usage across tissue and brain regions. In Fig. 4, we again use *Cacna1c* to illustrate how GTEx can be used to visualize tissue-specific use of alternative promoters and exons in human heart and coronary artery smooth muscle. Alternative exons 1a and 8a dominate in $Ca_V1.2$ mRNAs expressed in heart, whereas alternative exons 1b and 8b dominate in $Ca_V1.2$ mRNAs of coronary artery (Fig. 4). All other *Cacna1c* exons marked in GTEx appear to be expressed at about the same frequency in these two tissues. Brainspan is a complementary site that, although not specifically designed to identify alternative splicing, shows a heat map of gene expression levels in different regions of the human nervous system across development (http://www.brainspan.org/; BrainSpan, 2011). The Brainspan project contains mRNA sequencing data from 42 brain specimens spanning pre- and postnatal developmental stages for both sexes. For example, the 'Developmental transcriptome' tab can be used to visualize expression levels of *CACNA1* genes in different brain regions.

Splicing factors identify cohorts of synergistically regulated exons across different genes

The new genome-wide data sets also contain valuable information about correlated exon choice across multiple genes, a process essential for acquisition of cell phenotype and that is controlled by the action of cell-specific and tissue-specific RNA binding proteins (Wang *et al.* 2008). Nuclear cell-specific trans-acting splicing factors bind pre-mRNAs to promote or inhibit the recognition of certain alternatively spliced exons by the spliceosome during splicing. For example, the RBM20 RNA binding protein regulates alternative splicing of an exon in Titin pre-mRNA that is essential for adaptive changes in cardiac ventricular filling (Hidalgo & Granzier, 2013). In neurons, Nova RNA binding proteins regulate exon selection during pre-mRNA splicing of many genes essential for synaptic transmission (Zhang *et al.* 2010). RbFox RNA binding proteins are critical for exon selection of genes expressed in many tissues and cells, and they are essential for

normal cell differentiation, cell survival and cell signalling (Yeo *et al.* 2009). Analyses of splicing factor binding to RNA, by cross-linking immunoprecipitation combined with high-throughout sequencing (CLIP-HITS), have provided genome-wide maps of protein binding sites in brain regions (Licatalosi *et al.* 2008; Lambert *et al.* 2014). The location of splicing factor binding, in some cases, is predictive of splicing factor action. For example, serine–arginine-rich (SR) splicing factors Nova, RbFox and PTB act as repressors or enhancers of alternative exon inclusion in RNA in neurons depending on where their consensus binding motifs (*cis* elements) are relative to the target exon (upstream or downstream; Zhang *et al.* 2010). Genome-wide mapping of splicing factor binding represents the starting framework to define a splicing code that may eventually predict certain cell-specific splicing events across families of genes (Wang *et al.* 2008; Licatalosi *et al.* 2008; Yeo *et al.* 2009). Trans-acting RNA binding proteins are the best studied of the cell-specific splicing factors, but other molecules regulate cell-specific splicing,

including epigenetic modulators of DNA (Kornblihtt *et al.* 2013; Fukuda *et al.* 2013; Yan *et al.* 2015). For example, a ubiquitous DNA-binding protein, CCCTC binding factor (CTCF) regulates cell-specific splicing of alternative exons. In this case, CTCF binding promotes recognition of a weak splice site by the spliceosome, but its binding is inhibited by cell-specific methylation within the CTCF binding motif (Shukla *et al.* 2011). These data emphasize that transcription and pre-mRNA processing should not be viewed as sequential events but rather as processes that are strongly coupled temporally and mechanistically (Shukla & Oberdoerffer, 2012; Kornblihtt *et al.* 2013).

Developmental changes in ion channel splicing

In the examples discussed in the foregoing subsections, we highlight isoforms of voltage-gated calcium ion channel genes (*Cacna1*) that vary across different tissues and brain regions, but splicing of certain alternatively spliced
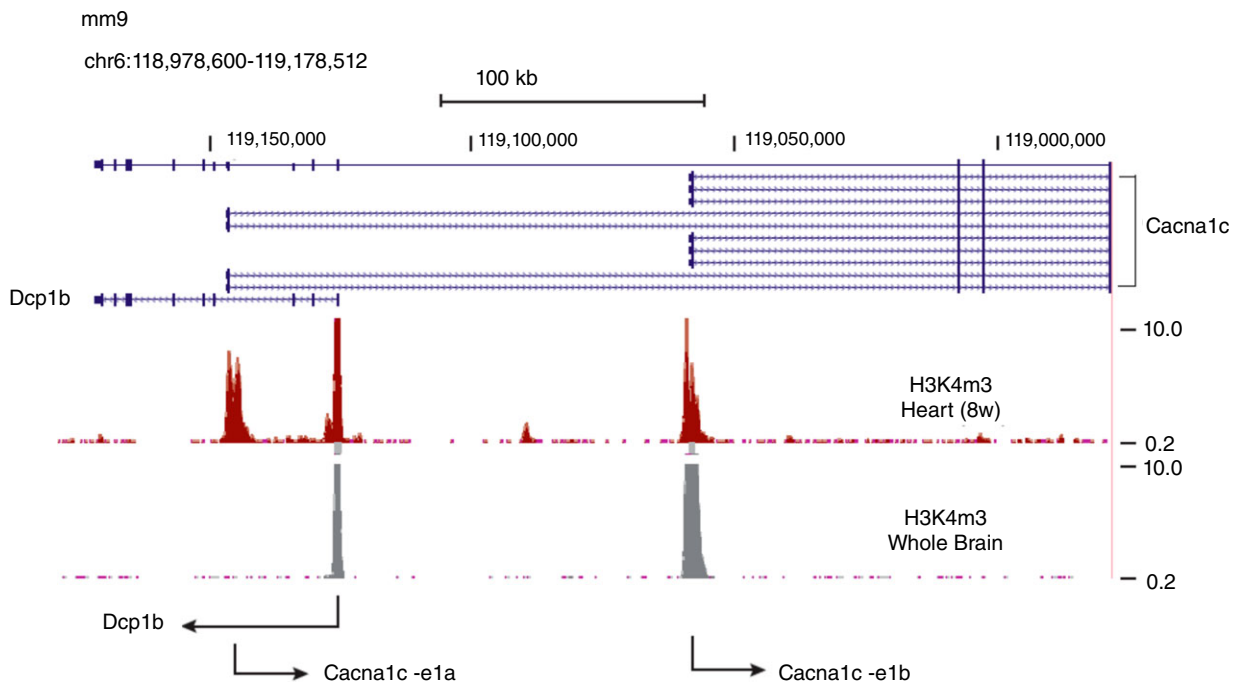

Figure 2. The use of different first exons in mouse *Cacna1c* in heart and brain is readily visualized using the UCSC Genome Browser by viewing H3K4Me3 regulatory markers that often correlate with the sites of transcription initiation
Mouse July 2007 (NCBI37/mm9) assembly is used for display, with genomic co-ordinate chr6:118534231-119182730. Genomic location and scale bar indicate the position and coverage of the genome. Each line represents a reference sequence aligned to the mouse genome (mm9) from UCSC Genes track (blue). Tissue-specific histone modifications are displayed in the ChIP-seq subtrack from the ENCODE/Ludwig Institute for Cancer Research (LICR; ENCODE, 2012) in parallel with RNA-seq data from mouse heart (8 weeks) and mouse embryonic whole brain (day 14.5). Exon 1a is used in heart and is located > 80 kb upstream of exon 1b that is used in both heart and brain. Three H3K4Me peaks are shown that correspond to two different transcription start sites for *Cacna1c* and a single transcription start for *Dcp1b*. The *Dcp1b* gene is transcribed on the reverse strand.

exons is also strongly dependent on development. In this section, we turn to studies of closely related voltage-gated sodium ion channel genes (*SCN*) to highlight examples of developmentally regulated alternatively spliced exons.

Splice isoforms of several neuronal sodium channels have been described, but perhaps most interesting is a splicing event that is highly conserved across several human *SCN* genes [*SCN1A* ($Na_V1.1$), *SCN2A* ($Na_V1.2$), *SCN3A* ($Na_V1.3$) and *SCN8A* ($Na_V1.6$)] that define 'neonatal' (exon N) and 'adult' (exon A) forms of human Na_V channels (Copley, 2004; Gazina *et al.* 2015). Mutually exclusive exons (usually exon 5) encode different amino acid sequences in the first (D1) of four canonical structurally homologous domains that typify all tetrameric voltage-gated ion channels.

Interestingly, this conserved neonatal/adult splice site is a modifier of disease severity. For example, a mutation in *SCN2A* ($Na_V1.2$ L1563V), which is located in a region of *SCN2A* that is downsteam of both the neonantal and adult exons, is causative of benign familial neonatal–infantile seizures. Although this mutation has no functional impact in adult $Na_V1.2$ channel isoforms, it increases the overall activity of the neonatal isoform of $Na_V1.2$. The benign familial neonatal–infantile seizures-causing mutation alters several biophysical properties of neonatal $Na_V1.2$ channels, including reduced kinetics of inactivation, faster recovery from inactivation and less voltage-sensitive inactivation. Notably, the mutation makes the neonatal $Na_V1.2$ channels behave like wild-type adult $Na_V1.2$ channels, perhaps explaining why this mutation is disease causing in only neonatal infants and not in adults (Xu *et al.* 2007; Gazina *et al.* 2015).

Human *SCNA8A* ($Na_V1.6$) contains a different pair of exons (18A and 18N) that are also differentially expressed during development. Exon 18N-containing mRNAs dominate in neurons at birth, but by 10 months of age in
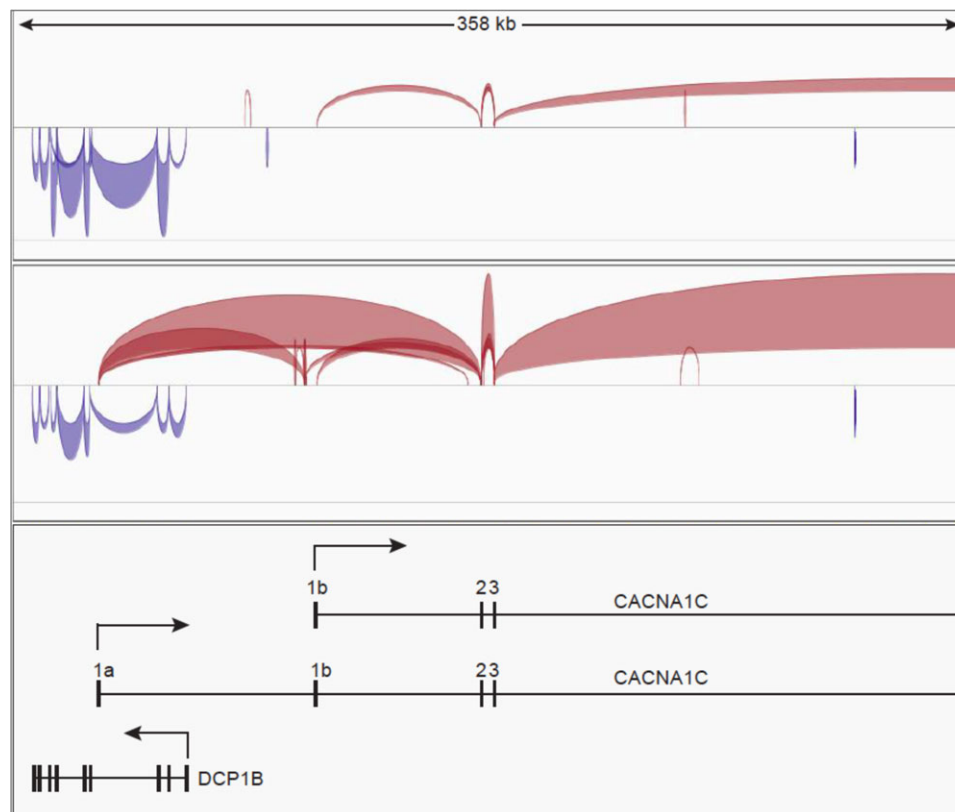

Figure 3. Visualization of tissue-specific use of alternative promoters in human *CACNA1C* using the integrative genomics viewer (IGV; Thorvaldsdóttir *et al.* 2013)
The 2009 human (GRCh37/hg19) assembly was used in this figure to visualize the two alternative promoters in *CACNA1C*. Junction Box option is turned on for quantitative visualization of exon–exon junction from mRNA sequencing reads aligned to gene annotations. These plots show tissue-specific usage of exon 1a (heart) and exon 1b (heart and brain). The thickness of the connections indicates the frequency of the reads that connect exon–exon junctions. The sense strand (red) and antisense strand (blue) are shown. Data were loaded from 'Bodymap 2.0' from IGV. The *DPC1B* gene is short and transcribed in the reverse direction. There are two promoters in *CACNA1C*, marked 1a and 1b.

mice there is a complete switch to 18A-containing mRNAs. Interestingly, exon 18N is disruptive; its inclusion shifts the open reading frame. $Na_V 1.6$ mRNAs containing 18N are non-functional and are degraded by nonsense-mediated decay (O'Brien *et al.* 2012). Thus, in this case alternative pre-mRNA splicing is used as a mechanism for cell-specific and development-dependent control of $Na_V 1.6$ channel expression post-transcription. Subtle changes in splicing regulation at the conserved neonatal/adult splice site of *SCN* genes have also been proposed to arise from common single nucleotide polymorphisms (SNPs), and these are linked to altered clinical outcomes. For example, common SNPs in *SCNA1* ($Na_V 1.1$) are hypothesized to influence the development of some types of epilepsy (Kasperaviciute *et al.* 2013) and the dosage of anti-epileptic drugs (Tate *et al.* 2005).

The function of voltage-gated calcium ion channels of skeletal muscle, $Ca_V 1.1$, is also modified during development as a result of alternative pre-mRNA splicing. Compared with adult channels, $Ca_V 1.1$ channels that dominate in embryonic skeletal muscle lack exon 29, an alternatively spliced exon that encodes 19 amino acids in the putative extracellular linker between transmembrane domains IVS3–IVS4. When present in $Ca_V 1.1$, the exon 29 encoding sequence reduces calcium currents and shifts channel activation thresholds to more positive voltages compared with embryonic $Ca_V 1.1$ channels that lack exon 29 (Tuluc *et al.* 2009; Benedetti *et al.* 2015). The amino acid sequence that links domains IVS3 and IVS4 of $Ca_V 1$ and $Ca_V 2$ channels is hypervariable and a hotspot for alterative splicing in *Cacna1* genes from *Drosophila* to humans (see Lipscombe *et al.* 2013*b*).

The neonatal and adult alternative splice isoforms of *SCN* genes and the developmentally regulated *CACNA1S* encoding $Ca_V 1.1$ channels can all be identified in the UCSC genome browser or IGV by aligning the mRNAs from all species and cross-referencing with conservation scores, as shown in Figs 1 and 2 for *Cacna1* genes, demonstrating the utility of the genome-wide transcription data in identifying developmentally regulated splicing events.

Current directions: alternative splicing disease and therapy

As discussed elsewhere, aberrant patterns of alternative pre-mRNA splicing can also contribute to disease pathology, as the primary cause of disease, as a consequence of disease pathology or as disease modifiers (Lipscombe *et al.* 2013*b*; Singh & Cooper, 2012), while natural genetic variations among individuals might account for differences in disease susceptibility. Consequently, there is interest in therapeutic strategies to regulate the pattern of alternative pre-mRNA splicing, for example, to correct mis-splicing that may result from mutations that disrupt pre-mRNA splicing or to shift the balance of splicing to modify disease severity (Yoshida *et al.* 2015).

Aberrant splicing patterns arising from rare mutations and common nucleotide variants across the human genome contribute to the aetiology of diseases and disorders including cancers, spinal muscular atrophy, autism spectrum disorders and other developmental disorders (Moulton *et al.* 2013; Tavassoli *et al.* 2014; Xiong *et al.* 2015; Zhang *et al.* 2015). In human heart, *SCN5A* pre-mRNAs ($Na_V 1.5$) are mis-spliced in patients with myotonic dystrophy type 1 (DM1) and Brugada syndrome ventricular tachyarrhythmias (Wahbi *et al.* 2013). There are abnormally high levels of the neonatal splice isoforms of $Na_V 1.5$ mRNA in ventricular myocardial tissue from patients with DM1 compared with control subjects, but this is only one example of a large number of abnormal splicing patterns found in
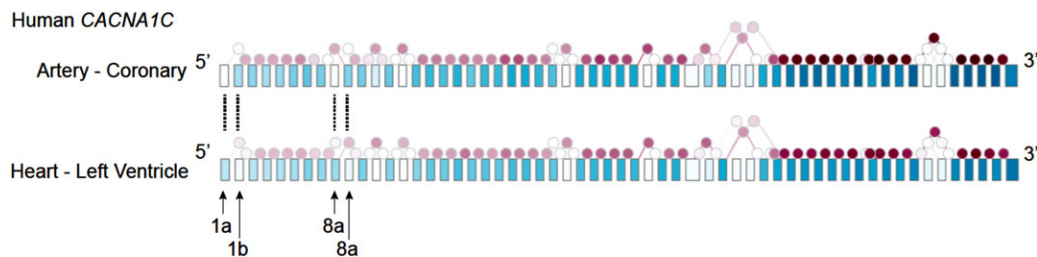

Figure 4. Tissue-specific use of alternative exons in human *CACNA1C* visualized using the GTEx portal (GTEx, 2013)
Using the GTEx portal, the major exons and patterns of alternative splicing are displayed. The blue rectangles represent exons and the red circles the splice options. The colour intensity represents the frequency of sequence reads for the specified tissue. Shown are analyses of RNA-seq data from human coronary artery and heart ventricles. The majority of $Ca_V 1.2$ expressed sequence in heart differs from that expressed in coronary artery at two mutually exclusive sites. In heart, exons 1a and 8a are expressed, whereas in coronary artery exons 1b and 8b are expressed.

patients with DM1. Myotonic dystrophy type 1 is caused by heterozygous trinucleotide repeat expansion [(CTG)*n*] in the 3′ untranslated region of the dystrophia myotonica protein kinase gene. A secondary consequence of this trinucleotide expansion is widespread dysregulation of pre-mRNA splicing in muscle, because of the sequestration of nuclear trans-acting splicing factors to the expanded trinucleotide repeat, notably Muscleblind-like member 1 (*MBNL1*) and CUG-binding protein Elav-like family member 1 (*CELF1*). Strategies designed to correct levels of these critical splicing factors are being explored as novel therapies for this devastating disease (Lee *et al.* 2012).

The causal gene for Rett syndrome, *MECP2*, regulates splicing, and aberrant alternative splicing patterns are observed in mouse models of Rett syndrome (Young *et al.* 2005). Several other neurological and motor neuron disorders are linked by common dysfunction in RNA processing, including fused in sarcoma (FUS)-amyotrophic lateral sclerosis (FUS-ALS), TAR DNA binding protein (TDP43)-ALS, frontotemporal dementia (FTD), spinal muscular atrophy (SMA) and autism spectrum disorders (Achsel *et al.* 2013; Ling *et al.* 2013; Corominas *et al.* 2014; Xiong *et al.* 2015). Approaches to normalize the RNA splicing events are therefore being explored as therapeutic strategies to compensate for diseases such as SMA (Hua *et al.* 2011) and DM1 (Oana *et al.* 2013; Wojciechowska *et al.* 2014).

Conclusion

The importance of individual and correlated RNA splicing events in normal and pathological cell function is indisputable. Here, we demonstrate how to access publicly available genome-wide sequence databases and to visualize and vertically integrate this information for a given gene or sets of genes using available Web-based and informatics tools. The value-content of genome-wide data sets is already high, but it will continue to grow as massively parallel sequencing of RNA and maps of regulatory elements in RNA and DNA are added from different cell types and tissues, in healthy and disease-affected individuals. Deep, genome-wide sequencing of expressed mRNAs at different developmental time points for a range of specific brain regions and tissues will be necessary for an accurate spatial and temporal map of splicing regulation. Knowledge of cell- and tissue-specific exon combination patterns, as well as the cell-specific regulatory elements that regulate them, will lead to a comprehensive set of data that should be highly predictive of the function and the state of the cell, and may inform strategies aimed at correcting abnormal splicing events associated with disease.

References

Achsel T, Barabino S, Cozzolino M & Carri MT (2013). The intriguing case of motor neuron disease: ALS and SMA come closer. *Biochem Soc Trans* **41**, 1593–1597.

Aoto J, Martinelli DC, Malenka RC, Tabuchi K & Sudhof TC (2013). Presynaptic neurexin-3 alternative splicing trans-synaptically controls postsynaptic AMPA receptor trafficking. *Cell* **154**, 75–88.

Benedetti B, Tuluc P, Mastrolia V, Dlaska C & Flucher BE (2015). Physiological and pharmacological modulation of the embryonic skeletal muscle calcium channel splice variant CaV1.1e. *Biophys J* **108**, 1072–1080.

Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermuller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaoz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Loytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman

N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN, Yu Y, Ruan Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, deBakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrimsdottir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B & deJong PJ (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816.

BrainSpan (2011). Atlas of the developing human brain. Funded by ARRA Awards 1RC2MH089921-01, 1RC2MH090047-01 and 1RC2MH089929-01. Available from: http://developinghumanbrain.org

Copley RR (2004). Evolutionary convergence of alternative splicing in ion channels. *Trends Genet* **20**, 171–176.

Corominas R, Yang X, Lin GN, Kang S, Shen Y, Ghamsari L, Broly M, Rodriguez M, Tam S, Trigg SA, Fan C, Yi S, Tasan M, Lemmens I, Kuang X, Zhao N, Malhotra D, Michaelson JJ, Vacic V, Calderwood MA, Roth FP, Tavernier J, Horvath S, Salehi-Ashtiani K, Korkin D, Sebat J, Hill DE, Hao T, Vidal M & Iakoucheva LM (2014). Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat Commun* **5**, 3650.

ENCODE (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.

Eom T, Zhang C, Wang H, Lay K, Fak J, Noebels JL & Darnell RB (2013). NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. *Elife* **2**, e00178.

Fukuda K, Ichiyanagi K, Yamada Y, Go Y, Udono T, Wada S, Maeda T, Soejima H, Saitou N, Ito T & Sasaki H (2013). Regional DNA methylation differences between humans and chimpanzees are associated with genetic changes, transcriptional divergence and disease genes. *J Hum Genet* **58**, 446–454.

Gazina EV, Leaw BT, Richards KL, Wimmer VC, Kim TH, Aumann TD, Featherby TJ, Churilov L, Hammond VE, Reid CA & Petrou S (2015). 'Neonatal' Nav1.2 reduces neuronal excitability and affects seizure susceptibility and behaviour. *Hum Mol Genet* **24**, 1457–1468.

GTEx (2013). The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580–585.

Guo W, Schafer S, Greaser ML, Radke MH, Liss M, Govindarajan T, Maatz H, Schulz H, Li S, Parrish AM, Dauksaite V, Vakeel P, Klaassen S, Gerull B, Thierfelder L, Regitz-Zagrosek V, Hacker TA, Saupe KW, Dec GW, Ellinor PT, MacRae CA, Spallek B, Fischer R, Perrot A, Özcelik C, Saar K, Hubner N & Gotthardt M (2012). *RBM20*, a gene for

hereditary cardiomyopathy, regulates titin splicing. *Nat Med* **18**, 766–773.

Han H, Irimia M, Ross PJ, Sung HK, Alipanahi B, David L, Golipour A, Gabut M, Michael IP, Nachman EN, Wang E, Trcka D, Thompson T, O'Hanlon D, Slobodeniuc V, Barbosa-Morais NL, Burge CB, Moffat J, Frey BJ, Nagy A, Ellis J, Wrana JL & Blencowe BJ (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* **498**, 241–245.

Hidalgo C & Granzier H (2013). Tuning the molecular giant titin through phosphorylation: role in health and disease. *Trends Cardiovasc Med* **23**, 165–171.

Hua Y, Sahashi K, Rigo F, Hung G, Horev G, Bennett CF & Krainer AR (2011). Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model. *Nature* **478**, 123–126.

Jepson J, Sheldon A, Shahidullah M, Fei H, Koh K & Levitan IB (2013). Cell-specific fine-tuning of neuronal excitability by differential expression of modulator protein isoforms. *J Neurosci* **33**, 16767–16777.

Kasperaviciute D, Catarino CB, Matarin M, Leu C, Novy J, Tostevin A, Leal B, Hessel EV, Hallmann K, Hildebrand MS, Dahl HH, Ryten M, Trabzuni D, Ramasamy A, Alhusaini S, Doherty CP, Dorn T, Hansen J, Krämer G, Steinhoff BJ, Zumsteg D, Duncan S, Kalviainen RK, Eriksson KJ, Kantanen AM, Pandolfo M, Gruber-Sedlmayr U, Schlachter K, Reinthaler EM, Stogmann E, Zimprich F, Théâtre E, Smith C, O'Brien TJ, Meng Tan K, Petrovski S, Robbiano A, Paravidino R, Zara F, Striano P, Sperling MR, Buono RJ, Hakonarson H, Chaves J, Costa PP, Silva BM, da Silva AM, deGraan PN, Koeleman BP, Becker A, Schoch S, vonLehe M, Reif PS, Rosenow F, Becker F, Weber Y, Lerche H, Rössler K, Buchfelder M, Hamer HM, Kobow K, Coras R, Blumcke I, Scheffer IE, Berkovic SF, Weale ME, Delanty N, Depondt C, Cavalleri GL, Kunz WS & Sisodiya SM (2013). Epilepsy, hippocampal sclerosis and febrile seizures linked by common genetic variation around *SCN1A*. *Brain* **136**, 3140–3150.

Katz Y, Wang ET, Silterra J, Schwartz S, Wong B, Thorvaldsdóttir H, Robinson JT, Mesirov JP, Airoldi EM & Burge CB (2015). Quantitative visualization of alternative exon expression from RNA-seq data. *Bioinformatics*. DOI: 10.1093/bioinformatics/btv034.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM & Haussler D (2002). The human genome browser at UCSC. *Genome Res* **12**, 996–1006.

Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E & Muñoz MJ (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* **14**, 153–165.

Lah GJ, Li JS & Millard SS (2014). Cell-specific alternative splicing of *Drosophila Dscam2* is crucial for proper neuronal wiring. *Neuron* **83**, 1376–1388.

Lambert N, Robertson A, Jangi M, McGeary S, Sharp PA & Burge CB (2014). RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell* **54**, 887–900.

Lander ES (2011). Initial impact of the sequencing of the human genome. *Nature* **470**, 187–197.

Lareau LF & Brenner SE (2015). Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible. *Mol Biol Evol* **32**, 1072–1079.

Lee JE, Bennett CF & Cooper TA (2012). RNase H-mediated degradation of toxic RNA in myotonic dystrophy type 1. *Proc Natl Acad Sci USA* **109**, 4221–4226.

Li HD, Menon R, Omenn GS & Guan Y (2014). The emerging era of genomic data integration for analyzing splice isoform function. *Trends Genet* **30**, 340–347.

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

Ling SC, Polymenidou M & Cleveland DW (2013). Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **79**, 416–438.

Lipscombe D, Allen SE & Toro CP (2013a). Control of neuronal voltage-gated calcium ion channels from RNA to protein. *Trends Neurosci* **36**, 598–609.

Lipscombe D, Andrade A & Allen SE (2013b). Alternative splicing: functional diversity among voltage-gated calcium channels and behavioral consequences. *Biochim Biophys Acta* **1828**, 1522–1529.

Lu Y, Loh YH, Li H, Cesana M, Ficarro SB, Parikh JR, Salomonis N, Toh CX, Andreadis ST, Luckey CJ, Collins JJ, Daley GQ & Marto JA (2014). Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells. *Cell Stem Cell* **15**, 92–101.

Martin AR, Costa HA, Lappalainen T, Henn BM, Kidd JM, Yee MC, Grubert F, Cann HM, Snyder M, Montgomery SB & Bustamante CD (2014). Transcriptome sequencing from diverse human populations reveals differentiated regulatory architecture. *PLoS Genet* **10**, e1004549.

Moulton VR, Grammatikos AP, Fitzgerald LM & Tsokos GC (2013). Splicing factor SF2/ASF rescues IL-2 production in T cells from systemic lupus erythematosus patients by activating IL-2 transcription. *Proc Natl Acad Sci USA* **110**, 1845–1850.

Oana K, Oma Y, Suo S, Takahashi MP, Nishino I, Takeda S & Ishiura S (2013). Manumycin A corrects aberrant splicing of *Clcn1* in myotonic dystrophy type 1 (DM1) mice. *Sci Rep* **3**, 2142.

O'Brien JE, Drews VL, Jones JM, Dugas JC, Barres BA & Meisler MH (2012). Rbfox proteins regulate alternative splicing of neuronal sodium channel SCN8A. *Mol Cell Neurosci* **49**, 120–126.

Raj B, Irimia M, Braunschweig U, Sterne-Weiler T, O'Hanlon D, Lin ZY, Chen GI, Easton LE, Ule J, Gingras AC, Eyras E & Blencowe BJ (2014). A global regulatory mechanism for activating an exon network required for neurogenesis. *Mol Cell* **56**, 90–103.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G & Mesirov JP (2011). Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26.

Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE & Zipursky SL (2000). *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671–684.

Schorge S, Gupta S, Lin Z, McEnery MW & Lipscombe D (1999). Calcium channel activation stabilizes a neuronal calcium channel mRNA. *Nat Neurosci* **2**, 785–790.

Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R & Oberdoerffer S (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79.

Shukla S & Oberdoerffer S (2012). Co-transcriptional regulation of alternative pre-mRNA splicing. *Biochim Biophys Acta* **1819**, 673–683.

Singh RK & Cooper TA (2012). Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med* **18**, 472–482.

Tate SK, Depondt C, Sisodiya SM, Cavalleri GL, Schorge S, Soranzo N, Thom M, Sen A, Shorvon SD, Sander JW, Wood NW & Goldstein DB (2005). Genetic predictors of the maximum doses patients receive during clinical use of the anti-epileptic drugs carbamazepine and phenytoin. *Proc Natl Acad Sci USA* **102**, 5507–5512.

Tavassoli T, Kolevzon A, Wang AT, Curchack-Lichtin J, Halpern D, Schwartz L, Soffes S, Bush L, Grodberg D, Cai G & Buxbaum JD (2014). De novo *SCN2A* splice site mutation in a boy with Autism spectrum disorder. *BMC Med Genet* **15**, 35.

Thorvaldsdóttir H, Robinson JT & Mesirov JP (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178–192.

Thorvaldsdóttir H, Robinson JT, Turner D & Mesirov JP (2015). A genomic data viewer for iPad. *Genome Biol* **16**, 46.

Tuluc P, Molenda N, Schlick B, Obermair GJ, Flucher BE & Jurkat-Rott K (2009). A Ca$_V$1.1 Ca^{2+} channel splice variant with high conductance and voltage-sensitivity alters EC coupling in developing skeletal muscle. *Biophys J* **96**, 35–44.

Venables JP, Brosseau JP, Gadea G, Klinck R, Prinos P, Beaulieu JF, Lapointe E, Durand M, Thibault P, Tremblay K, Rousset F, Tazi J, Abou Elela S & Chabot B (2013). RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. *Mol Cell Biol* **33**, 396–405.

Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ & Geschwind DH (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384.

Wahbi K, Algalarrondo V, Becane HM, Fressart V, Beldjord C, Azibi K, Lazarus A, Berber N, Radvanyi-Hoffman H, Stojkovic T, Béhin A, Laforêt P, Eymard B, Hatem S & Duboc D (2013). Brugada syndrome and abnormal splicing of SCN5A in myotonic dystrophy type 1. *Arch Cardiovasc Dis* **106**, 635–643.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP & Burge CB (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476.

Wojciechowska M, Taylor K, Sobczak K, Napierala M & Krzyzosiak WJ (2014). Small molecule kinase inhibitors alleviate different molecular features of myotonic dystrophy type 1. *RNA Biol* **11**, 742–754.

Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RK, Hua Y, Gueroussov S, Najafabadi HS, Hughes TR, Morris Q, Barash Y, Krainer AR, Jojic N, Scherer SW, Blencowe BJ & Frey BJ (2015). RNA splicing. The human

splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806.

Xu R, Thomas EA, Jenkins M, Gazina EV, Chiu C, Heron SE, Mulley JC, Scheffer IE, Berkovic SF & Petrou S (2007). A childhood epilepsy mutation reveals a role for developmentally regulated splicing of a sodium channel. *Mol Cell Neurosci* **35**, 292–301.

Yan Q, Weyn-Vanhentenryck SM, Wu J, Sloan SA, Zhang Y, Chen K, Wu JQ, Barres BA & Zhang C (2015). Systematic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators. *Proc Natl Acad Sci USA* **112**, 3445–3450.

Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD & Gage FH (2009). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* **16**, 130–137.

Yoshida M, Kataoka N, Miyauchi K, Ohe K, Iida K, Yoshida S, Nojima T, Okuno Y, Onogi H, Usui T, Takeuchi A, Hosoya T, Suzuki T & Hagiwara M (2015). Rectifier of aberrant mRNA splicing recovers tRNA modification in familial dysautonomia. *Proc Natl Acad Sci USA* **112**, 2764–2769.

Young JI, Hong EP, Castle JC, Crespo-Barreto J, Bowman AB, Rose MF, Kang D, Richman R, Johnson JM, Berget S & Zoghbi HY (2005). Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc Natl Acad Sci USA* **102**, 17551–17558.

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

Zhang X, Joehanes R, Chen BH, Huan T, Ying S, Munson PJ, Johnson AD, Levy D & O'Donnell CJ (2015). Identification of common genetic variants controlling transcript isoform variation in human whole blood. *Nat Genet* **47**, 345–352.

Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, Phatnani HP, Guarnieri P, Caneda C, Ruderisch N, Deng S, Liddelow SA, Zhang C, Daneman R, Maniatis T, Barres BA & Wu JQ (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci* **34**, 11929–11947.

Additional information

Competing interests

None declared.

Author contrbutions

All authors contributed equally to the compiling of the literature, its analysis and the writing of the paper.

Funding