Bhuiyan et al. BMC Genomics (2018) 19:637
https://doi.org/10.1186/s12864-018-5013-2

BMC Genomics

RESEARCH ARTICLE Open Access

Systematic evaluation of isoform function in literature reports of alternative splicing

Shamsuddin A. Bhuiyan[1,2,3], Sophia Ly[1], Minh Phan[1], Brandon Huntington[1], Ellie Hogan[1], Chao Chun Liu[1], James Liu[1] and Paul Pavlidis[1,2*]

Abstract

Background: Although most genes in mammalian genomes have multiple isoforms, an ongoing debate is whether these isoforms are all functional as well as the extent to which they increase the functional repertoire of the genome. To ground this debate in data, it would be helpful to have a corpus of experimentally-verified cases of genes which have functionally distinct splice isoforms (FDSIs).

Results: We established a curation framework for evaluating experimental evidence of FDSIs, and analyzed over 700 human and mouse genes, strongly biased towards genes that are prominent in the alternative splicing literature. Despite this bias, we found experimental evidence meeting the classical definition for functionally distinct isoforms for ~ 5% of the curated genes. If we relax our criteria for inclusion to include weaker forms of evidence, the fraction of genes with evidence of FDSIs remains low (~ 13%). We provide evidence that this picture will not change substantially with further curation and conclude there is a large gap between the presumed impact of splicing on gene function and the experimental evidence. Furthermore, many functionally distinct isoforms were not traceable to a specific isoform in Ensembl, a database that forms the basis for much computational research.

Conclusions: We conclude that the claim that alternative splicing vastly increases the functional repertoire of the genome is an extrapolation from a limited number of empirically supported cases. We also conclude that more work is needed to integrate experimental evidence and genome annotation databases. Our work should help shape research around the role of splicing on gene function from presuming large general effects to acknowledging the need for stronger experimental evidence.

Keywords: Alternative splicing, Functional diversity, Isoform function, Literature curation

Background

An ongoing debate is whether most mammalian genes produce more than one functional isoform [1–3]. The mere presence of multiple isoforms in public sequence databases is clearly insufficient to settle the question [4]. Arguments against widespread functional alternative isoforms include the fact that the splicing machinery's limited fidelity causes the stochastic generation of "junk" isoforms [5, 6]. Analyses using proteomics and molecular evolution approaches have also failed to support the expression and conservation of most splice isoforms [3, 7–10]. Nevertheless, the question lingers because the lack of evidence is not generally accepted as evidence and the function of most splice isoforms remain unknown [1, 4]. Beyond the question of whether most genes have more than one functional isoform is a critical issue: whether these isoforms increase the functional repertoire of genes, or are merely functionally redundant [11–14]. In this paper we take steps to address the gap between the commonplace assumption that most genes have more than one distinct functional product and the evidence-based reality.

Establishing whether a gene has functionally distinct isoforms requires experimental validation. While databases that contain information on transcript isoforms gather information on isoform features, none attempt to assess functionally distinct isoform reports from the experimental literature. For example, Ensembl, RefSeq, and UniProt catalog and annotate splice isoforms based on

* Correspondence: paul@msl.ubc.ca
[1]Michael Smith Laboratories, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[2]Department of Psychiatry, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
Full list of author information is available at the end of the article

© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

evidence that they exist as a transcript or protein [4, 15–17]. However, the existence of a splice isoform alone does not provide direct support for its functionality, much less functional distinctness.

To establish the extent to which splice isoforms increase the functional repertoire of the genome, we need data on which genes have *functionally distinct splice isoforms (FDSIs).* Identification of genes with FDSIs requires experimental support to demonstrate the necessity of each splice isoform. A classical method to determine the function of a given gene is to knock it out and observe the phenotypic consequence [18, 19]. This idea readily extends to isoforms; if a single isoform is made absent and that isoform is necessary for the normal function of the gene, then a consequence (change in phenotype) would be expected. A gene has FDSIs if two or more isoforms meet this criterion independently (Fig. 1a). In contrast, the depletion of an unnecessary or redundant splice isoform will not cause a phenotype. Another approach that is often used to probe the function of isoforms is overexpression. However, overexpression is well known to be fraught with interpretational challenges including artifacts so the gold standard is to generate loss-of-function alleles [20]. Note that a negative result from experiments is not evidence of a lack of functional distinctness, as it is possible the functional distinction between the isoforms may be eventually discovered. Curating the genes with FDSIs is of obvious importance to evaluate the state of the literature support for the commonplace claim that alternative splicing increases the functional repertoire of the genome.

Beyond identifying knowledge gaps, establishing a set of genes with FDSIs provides potential avenues for improving computational approaches to analyzing alternative splicing. For example, classifiers, such as PULSE, attempt to predict genes with multiple functional splice isoforms [21]. Hao et al. trained PULSE using a set of splice isoforms confirmed by Western blot experiments. PULSE predicted that one-third of human protein-coding genes have multiple functional isoforms (not necessarily functionally distinct). A difficulty cited by Hao et al. was in the identification of training data, an issue which is even worse if one is interested in functional distinctness. Having lists of experimentally validated genes with FDSIs could open the door to improved algorithmic approaches in characterizing isoform function.

Here we present a literature-based analysis of experimental evidence for functionally distinct splice isoforms (FDSIs) for over 700 human and mouse genes. Despite a gene selection strategy that was highly biased towards

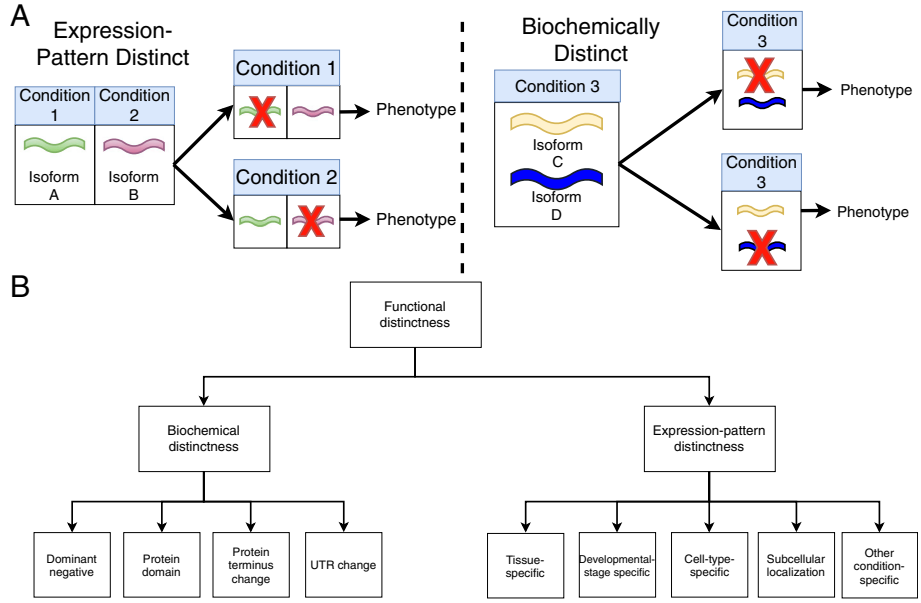

Fig. 1 Non-mutually exclusive types of functional distinctness for literature reported genes with FDSIs. **a** Generally, the distinctness of FDSIs of the same gene can be attributed to expression-pattern distinctness or biochemical distinctness. Expression-pattern distinctness is defined as a gene having specific splice isoforms necessary in distinct conditions. The depletion of the splice isoform in its distinct condition causes a phenotype. Biochemical distinctness is defined as a protein structure difference between splice isoforms of the same genes. While the FDSIs of the gene can be expressed in the same condition, the depletion of either splice isoform causes a phenotype. **b** For genes with FDSIs, we categorized the specific subtypes of functional distinctness which contributed to the distinctness between the splice isoforms of the gene (summarized in Table 4). Expression-pattern distinctness can be further categorized as "cell-type-specific", "tissue-specific", "developmental-stage-specific", "subcellular localization-specific" and "other condition-specific". Biochemical distinctness can be further categorized as "dominant-negative", "protein domain", "UTR change" and "protein terminus change"

genes suggested to have multiple functional isoforms, we found good experimental evidence for FDSIs for fewer than 10% of genes.

Results

Landscape of the alternative splicing literature

To generate a starting set of papers to curate, we queried PubMed on August 2017 using the term "alternative splicing". We found 19,049 human studies and 8197 mouse studies representing 12,891 human genes and 7585 mouse genes. While the median number of papers per gene was one, there was a large variance (Additional file 1: Figure S1 and Figure S2). Most human genes (7738) had only one such paper associated with them, while some have up to 100 (for example, SRSF1). We also observed that genes with many "alternative splicing"-mentioning papers tend to have many papers in PubMed overall (Spearman's rank correlation = 0.55, see Additional file 1: Figure S3). For example, we identified 86 studies linked to human *TP53* with the term "alternative splicing" (rank 2), but this is not particularly remarkable because overall, PubMed contains 8261 studies linked to TP53 – the most studies for any single gene. This suggests, unsurprisingly, that heavily studied genes tend to have more research done on their splicing.

Curation summary

We manually curated primary studies which provide evidence for the function of splice isoforms. As described in Methods, we selected genes and publications for curation in a manner that we expected should enrich for documentation of functional distinctness – for example, using review articles on splicing function. The curation process primarily focused on determining whether the elimination of expression of each splice isoform from a single gene caused an observable phenotype. Table 1 provides a summary of the knowledgebase as of July 20th, 2018, and Additional file 1 and Additional file 2 contain full details of all curated studies for human and mouse. In total, we curated 1127 human and mouse studies. This encompasses 903 human studies (555 genes) and 272 mouse studies (227 genes). We have curated a median of 1 study per human gene and 1 study per mouse gene (mean = 1.5 studies and 1.2 studies, respectively). Our curation evaluations (see Methods) revealed that the curators agree on the interpretation of a paper 98% of the time. Errors were generally false positives for functional distinctness, which we addressed in the final review (see Methods).

Identification of 23 human genes with direct evidence of functionally distinct splice isoforms

By definition, a gene with functionally distinct splice isoforms (FDSIs) has at least two splice isoforms necessary for the gene's normal function. We find that genes with such evidence are rare: about 4% of curated human genes (9% of mouse genes) have FDSIs, based on reports in a total of 64 studies out of 1127 studies. Note that 138 studies depleted only one splice isoform of a gene and no other study we curated had depleted any other isoforms of the same gene. We provided the full list of the 23 human genes and the 20 mouse genes with FDSIs in Table 2 with additional information in Additional file 3. RNAi knockdown experiments provided support for over 75% of these FDSIs, while the remaining FDSIs were characterized using gene knockouts combined with isoform-specific rescue.

We sought genes with negative evidence for FDSIs. For these cases, experiments individually depleted multiple splice isoforms for a single gene, however, only one splice isoform's depletion caused a phenotype and while the depletion of the other splice isoforms caused no phenotype. We found 16 genes with such evidence (shown in Table 3).

As mentioned, we biased our gene and paper selection in such a way that our estimate of ~ 4% (~ 9% for mouse) might be too high. To help clarify this issue, we randomly selected 100 human genes (from those that had at least one alternative splicing related paper) for gene-centric curation (listed in Additional file 1: Table S1). Of these

Table 1 Curation of alternative splicing literature has reveals 23 human genes and 20 mouse genes with functionally distinct splice isoforms (FDSIs)

Species	Curated genes	Genes with FDSIs	Studies curated	Study Type			
				Isoform Removal	Overexpression	Localization	Other study types
Human	555	23	903	149	294	80	380
Mouse	227	20	272	82	70	37	83
Total	782	43	1127	222	353	109	443

The 23 human genes with FDSIs accounted for almost 4% of human genes annotated in this knowledgebase, while the 20 mouse genes accounted for 9% of the all mouse genes annotated. The majority of curated studies could be classified into three different types: "isoform removal", "overexpression" and "localization". Isoform removal studies have experiments where expression of at least one splice isoform is eliminated and a phenotypic change is evaluated. Overexpression studies have experiments where at least one splice isoform is overexpressed. This "abundance" of the splice isoform can cause a phenotype (not necessarily distinct). Localization studies have experiments that characterize where in the cell or organism the splice isoform is expressed. A single study can report experiments with multiple study types. The total number of human and mouse studies curated do not sum to 1158 studies because some publications investigated both human and mouse forms of a single gene

Table 2 Genes with FDSIs identified

	Gene	Number of FDSIs	Number of Ensembl Transcripts	Number of Studies	Mappable to Ensembl?	PULSE	Gene	Number of FDSIs	Number of Ensembl Transcripts	Number of Studies	Mappable to Ensembl?	PULSE
Human	*AR*	3	9	31	Yes	NA	*BCAR1*	3	17	1	No	Missed
	BDNF	3	19	12	No	NA	*BIRC5*	3	11	34	Yes	NA
	BOK	2	2	1	No	Missed	*CD44*	2	39	58	Yes	Predicted
	CFLAR	2	25	6	Yes	Predicted	*CSPP1*	2	7	1	Yes	NA
	DPF3	2	16	2	Yes	NA	*EIF4G1*	2	38	6	No	NA
	EIF4G2	3	32	3	No	NA	*HBS1L*	2	14	1	Yes	Predicted
	KLF6	2	7	16	No	Missed	*MADD*	2	23	2	Yes	Predicted
	MST1R	2	15	11	Yes	Predicted	*PML*	2	22	12	Yes	Predicted
	PGAM5	2	4	1	Yes	Missed	*PRMT5*	2	20	4	Yes	Missed
	STIM2	2	12	2	No	NA	*SUN1*	2	35	2	No	NA
	TICAM1	2	2	1	No	NA	*TICAM2*	2	2	1	No	NA
	TP63	2	14	27	Yes	NA						
Mouse	*Cacna1b*	2	10	2	No	Missed	*Calca*	2	7	10	Yes	NA
	Cdc42	2	2	8	Yes	NA	*Enc1*	2	1	3	No	NA
	Homer1	2	12	7	Yes	NA	*Il1rap*	2	7	4	Yes	Training
	Lpin1	2	8	5	No	Missed	*Lrp8*	2	12	13	No	Predicted
	Mecp2	2	6	9	Yes	Missed	*Myh10*	2	10	7	No	Predicted
	Nf1	2	9	5	Yes	Missed	*Opn4*	2	3	2	Yes	NA
	Oprm1	3	31	20	Yes	Predicted	*Rbfox1*	2	2	22	No	Predicted
	Robo3	2	6	5	Yes	Missed	*Rock2*	2	12	4	No	Missed
	Ryr3	2	12	5	No	NA	*Sirt3*	2	10	3	No	Missed
	Snap25	2	3	12	Yes	NA	*Tp63*	2	8	9	Yes	NA

Studies have provided positive evidence of functional distinctness for these genes in experiments where individual splice isoforms were eliminated and a phenotypic change was observed. See Additional file 3 for study demonstrating functional distinctness. "Number of FDSIs" indicates the number of splice isoforms where depletion of splice isoforms causes a phenotype. "Number of Ensembl Transcripts" indicates number of transcripts found in Ensembl entry for gene. "Number of studies" indicates the number of studies associated with the gene retrieved with the term "alternative splicing" on PubMed. The highest number of FDSIs found in a single gene is three. "Mappable to Ensembl" indicates genes where we successfully linked all FDSIs back to Ensembl. "PULSE" indicates whether the gene was used at all by Hao and colleagues in their computational predictions. "Training" in this column means that the gene was used as part of PULSE's training set. "Predicted" means that PULSE predicted that the gene has multiple functional splice isoforms. "Missed" means that PULSE failed to predict that the gene has multiple functional splice isoforms. "NA" means that the gene was not an input for PULSE

100 genes, two genes (PML and DPF3, 2%, of the curated genes) had experimental evidence of FDSIs.

We also curated gain-of-function experiments where investigators overexpressed multiple splice isoforms of the same gene. From our 555 curated human genes and 227 curated mouse genes, we found 50 human genes (~ 9%) and 14 mouse genes (~ 4%) where investigators overexpressed individual splice isoforms and yielded multiple distinct phenotypes. Such studies did not meet our criteria for FDSIs, but we report them in case this relaxed criterion is of interest to others.

Genes tend to express functionally distinct splice isoforms in the same condition

To further explore functional distinctness in splicing, we identified non-mutually exclusive types of functional distinctness between FDSIs of the same gene, summarized in Table 4. We classified two main types of distinctness, expression-pattern distinctness and biochemical distinctness. Genes with expression-pattern distinct FDSIs have splice isoforms necessary for specific conditions while genes with biochemical distinctness have FDSIs with distinct biochemical properties that cannot compensate for each other even when co-expressed (for further description see Methods and Fig. 1).

The majority of genes (27/43) have biochemically distinct isoforms, rather than expression-pattern distinct. We identified "dominant-negative" as the most common subtype of biochemically distinct FDSIs (12/31 genes). For example, the mouse gene *Enc1* has two FDSIs, named "57 kDa" and "67 kDa" by the authors, interacting in the Wnt-signalling pathway [22]. Knockdown of 57 kDa promoted osteoblast mineralization while the knockdown of 67 kDa inhibited osteoblast mineralization.

Table 3 Genes with evidence failing to support FDSIs (negative results)

Gene	Experimental method	Tissue/Cell Type	Reference (PubMed ID)
Ank3	Isoform-specific rescue	Neuron	25,552,556
Ar	Knockdown	Prostate cancer cell line	20,823,238
Ccnd1	Isoform-specific rescue	Embryonic fibroblast	21,200,149
Dab1	Isoform-specific rescue	Neuron	28,968,791
Dntt	Isoform-specific rescue	Bone marrow	11,136,823
FANCE	Isoform-specific rescue	Breast cancer cell line	26,277,624
FNBP1L	Isoform-specific rescue	MDCK cell line	26,063,734
Pcdha1	Isoform-specific rescue	Brain	18,973,563
PDE4D	Knockdown	Kidney	16,030,021, 17,673,687
PEX19	Isoform-specific rescue	Fibroblast	11,883,941
Pparg	Knockdown and isoform-specific rescue	Adipose	11,782,442
RAP1GSDS1	Knockdown	Breast cancer cell line	24,197,117
RREB1	Knockdown	Bladder	21,703,425
SIRT1	Knockdown	Colon cancer cell line	22,124,156
Smad2	Isoform-specific rescue	Embryonic stem cells	15,630,024
STAT1	Knockdown	Embryonic cells	21,914,475

These genes had multiple isoforms tested however only one splice isoform caused a change in phenotype

In contrast to biochemical distinctness, we identified fewer cases of genes with expression-pattern distinct FDSIs. Only a total of 17 human and mouse genes had FDSIs in which the distinctness arises from distinct expression patterns. For example, the mouse gene *Myh10* has two FDSIs, named B1 and B2 by the authors [23]. Cells in the brainstem express B1 to promote normal migration of facial neurons, while cells in the cerebellum expressed B2 to promote normal cerebellar Purkinje cell development.

Challenges linking FDSIs to sequence databases

We attempted to link all identified FDSIs back to Ensembl transcript identifiers and were successful in 25/43 cases. Our process was as follows. First, in the studies for seven genes, investigators provided a GenBank or RefSeq ID. We were able to map three of these to Ensembl (which includes GenBank and RefSeq data), but not for the other four (for more details see Additional file 3), accounting for four of the 16 failures. Next, for 36 genes with missing accession information, we used sequence alignment or other information to identify likely matches (See Methods). This was successful in 25 cases. In a further 6 cases, we were able to determine a sequence by referring to other papers by the same authors. Despite extensive efforts, we were unable to find matching Ensembl transcripts or sequence data for the isoforms of 5 genes. This situation was not specific to

Table 4 Most genes with FDSIs have biochemically distinct splice isoforms

	Types of distinctness	Human genes	Mouse genes
Distinct expression patterns	Cell-type-specific	*AR, MADD*	
	Developmental-stage-specific	*CD44*	*Myh10, Robo3*
	Cellular localization	*BIRC5, CSPP1, PRMT5, PML*	*Myh10, Rbfox1, Robo3, Sirt3*
	Tissue-specific	*MST1R*	*Calca, Rock2*
	Other-condition-specific	*BOK*	
Biochemically distinct	Protein domain	*CFLAR, DPF3, EIF4G1, TICAM1, TP63*	*Lrp8*
	Dominant negative	*BIRC5, HBS1L, KLF6, Nf1, PRMT5, STIM2, SUN1, TICAM*	*Enc1, Nf1, Robo3, Ryr3, Tp63*
	Protein terminus change	*BCAR1, BDNF, EIF4G2, IL1RAP, PGAM5*	*Cacna1b, Mecp2, Oprm1, Pn4*
	UTR Change	*BDNF*	

Genes with FDSIs were categorized on functional type based on the literature that reported on the FDSIs using the scheme outlined in Fig. 1. Genes categorized as "distinct expression patterns" express FDSIs in specific conditions. Genes categorized as "biochemically distinct" have FDSIs whose functional distinctness is a consequence of biochemical differences in their final protein product. Genes can be categorized as both "distinct expression patterns" and "biochemically distinct" such as Myh10 and Robo3

Ensembl as we failed to link isoforms of 8 genes to UniProt; see Additional file 3.

Only a quarter of genes with FDSIs are predicted by a computational classifier

Hao et al. [21] developed a machine learning algorithm (PULSE) that predicted 1/3 of human genes have more than one functional isoform (but not necessarily functionally distinct). We hypothesized that our curated genes with FDSIs would be enriched among those predictions, because even though Hao and colleagues were not attempting to predict functional distinctness, genes with FDSIs by definition have more than one functional isoform. Though we included PULSE's training genes in our gene-centric curation, only two gene with FDSIs (including human orthologues of our curated mouse genes) were used by Hao et al., in their training data. In their validation gene set of 212 genes, we found none of our genes with FDSIs. Hao et al. predicted 2419 genes to each have multiple functional splice isoforms. Ten of our genes with FDSIs are included in this set. Based on input set used for PULSE predictions, the classifier failed to predict 12 of our genes with experimentally-validated FDSIs to have multiple functional splice isoforms. However, our interpretation of these is limited because of the small number of genes with FDSIs.

Discussion

This paper represents progress towards documenting and evaluating the breadth of evidence for functionally distinct splice isoforms (FDSIs) for human and mouse genes. The inspiration for our study was strong arguments against the likelihood of most genes having multiple functional isoforms, contrasted with the ubiquitous claim that splicing vastly increases the functional repertoire of the genome [3, 11–14, 24, 25]. This led us to ask where this latter claim comes from: while surely there are interesting cases of multi-isoform genes, has this been optimistically extrapolated to the entire genome? Our analysis suggests this is the case and supports the hypothesis that the majority of splice isoform functions remain unknown [4, 26, 27]. While we were not surprised that there is no evidence of FDSIs for most genes, we were surprised by the low fraction for which there is supporting data, a mere 4% in human genes and 9% in mouse genes. Regardless of whether this number holds true with more curated studies, by contributing a list of genes with documented functionally distinct isoforms, we start to identify the scope of the gaps, the parameters for future experimental work, and assist computational methods that require training examples.

The low fraction of genes surveyed for which we found evidence of FDSIs (~ 4–9%) agrees with the general sense that we still have limited concrete evidence of more than one functional splice isoform per gene [3, 9, 28]. Even if we loosen our criteria to include overexpression studies, this fraction rises only to ~ 12–13% Furthermore, we only considered genes for which some literature exists for their isoforms, so the range 4% to 9% is relative to genes that have at least one publication about them associated with splicing. Based on our PubMed queries, we estimate that one-third of human protein-coding genes do not have any type of specific experimental study of differences among their isoforms. For most genes the main available sources of information come from genome-wide studies of transcript expression patterns, which do not address function.

One might question whether the fraction 4% will rise substantially as we continue our curation efforts, but we hypothesize a lower true fraction of genes with documentation of FDSIs in the literature. First, we aimed the gene-centric aspect of our curation at genes mentioned in review articles or otherwise prominent genes, and thus is highly biased towards genes with experimentally-backed function, yielding an over-estimate. Second, the gene-centric survey of 100 randomly-selected human genes yielded only two genes with evidence of FDSIs. Third, we found a median of only one study per gene from PubMed. Since the genes with FDSIs tended to be genes with relatively more associated studies (Table 2), genes with few associated studies seem less likely to yield existing positive evidence for functional distinctness. Fourth, investigators face technical and/or resource challenges when testing the functional distinctness of isoforms, requiring either the ability to conduct isoform-specific depletion experiments, or isoform-specific rescue following a complete gene knockout. Reasonably, one might suppose that in many cases the experiments have not been done. The essential problem remains that most genes simply have not had their isoforms tested in such a way as to establish distinct functions.

We also sought *negative* evidence of genes having FDSIs from experiments where the depletion of only one splice isoform causes a phenotype while the depletion of the remaining splice isoforms of the same gene causes no phenotype. However, we only identified eight human genes and eight mouse genes from 16 studies with this type of evidence in our current curation of 1127 studies (Table 3). Since most studies consider only one type of functional assay, it remains possible that tests of different functions would yield positive results for these genes. Nevertheless, the "file-drawer effect" – a type of publication bias against negative results – potentially plays a role in the dearth of negative evidence [29].

A natural question is whether genes with FDSIs have distinguishing features compared to genes without FDSIs. However, we identified too few genes to perform an adequately powered analysis. Furthermore, both the

literature and our curation process have large biases in the identification of FDISs. They tend to involve highly studied genes, while at the same time the extent and types of investigations into isoform function are highly variable. If there are biological principles that explain the distribution of FDSIs in the genome, discovering them will require a larger and less biased source of data than is currently available.

Evaluating the evidence for FDSIs at the gene level

After the curation of over 1000 alternative splicing studies, we identified 23 human genes and 20 mouse genes with evidence for functionally distinct splice isoforms, mostly determined by RNAi knockdown experiments. RNAi knockdowns naturally align with our definition of a functional splice isoform and how researchers traditionally determine function in molecular biology. One question that arises in discussing RNAi is target specificity and efficacy. In most, but not all, of the papers we curated as having FDSIs, the authors demonstrate the target specificity of their siRNA to effectively deplete a single isoform. We raise this as a reminder that reports of evidence for functional distinctness may vary in quality.

Isoform-specific rescues demonstrating functional distinctness provide an alternative option to knockdown studies but the method has limitations when determining whether the splice isoforms rescue distinct phenotypes. In some studies, splice isoforms of the same gene clearly rescued distinct functions. For example, Candi and colleagues performed rescue experiments on Tp63-null mice [30]. The knockout of Tp63 impeded the development of skin. In the rescue experiments, the splice isoform ΔNp63 restored the skin's basal layer while the TAp63 restored the skin's upper layers. In contrast, other studies rescued the same phenotype with each splice isoform, which makes evidence of functional distinctness unclear. For example, in the investigation by Coldwell and colleagues, each splice isoform of *EIF4G1* (eIF4G1e and eIF4G1f) rescued the phenotype of translation by restoring the translation rate [31]. It is unclear whether this constitutes evidence of functional distinctness. Since both splice isoforms rescued the same phenotype, they appear functionally redundant. Nevertheless, in cases such as these, we accepted the claim of the authors that the gene has FDSIs.

We resisted accepting overexpression studies as demonstrating FDSIs for two reasons. First, overexpression experiments are known to be subject to a variety of artifacts [20]. Second, and more importantly, overexpression experiments fail to provide evidence for a splice isoform's necessity. In molecular biology, a molecule's necessity can only be supported by the effects of the molecule's absence [32, 33]. Thus, we have more confidence in isoform depletion experiments to provide support for genes with FDSIs compared to overexpression. We draw a parallel to the standards of evidence for characterizing gene function, in which evaluation of a loss of function is the gold standard [34]. We argue that the same criteria used to establish gene function must be applied to isoforms.

Types of functional distinctness in FDSIs

It has been speculated that many poorly-characterized isoforms may have function because genes express splice isoforms in specific conditions, perhaps yet to be studied [1, 24, 35]. It is therefore relevant that the minority (17) of genes had functional distinctness due to condition-specificity. This may simply be due to a lack of study of condition-specific studies, as it might be generally easier to study isoforms expressed in the same conditions. Our results thus point to a potential gap in the literature.

Disconnect between the literature and gene databases

In one-third of the genes with FDSIs, the isoforms studied in a paper could not be matched to transcripts in Ensembl (as mentioned, this is not an Ensembl-specific problem; ~20% of genes had functional isoforms that could not be matched to UniProt). Conversely, Ensembl contains many transcripts that the literature ignores. This observation has fairly serious implications for basing splice isoform research on the contents of Ensembl (or related databases). If one developed experiments to functionally test the splice isoforms of the genes we identified to have FDSIs based on Ensembl transcripts for that gene, their experiments would not contain the correct FDSIs in at least one-third of the genes. In bioinformatics research, computational methods that make predictions based on Ensembl transcripts might be valueless to experimental biologists as Ensembl does not reflect the literature. Large-scale databases specialized for alternative splicing, such as the Alternative Splicing encyclopedia (ASpedia) and the APPRIS database, tend to anchor to Ensembl [36, 37]. Of note, previous discussions used APPRIS to understand the functional impact of alternative splicing [3]. The disconnect between Ensembl the literature also impacts datasets not specific to splicing but where splice isoform information is important. For example, the GTEx consortium provides transcript-level quantification based on the Ensembl transcriptome [38]. The FDSIs that are not in Ensembl are therefore not included in GTEx. Given the few known cases of genes with FDSIs and PULSE's inability to predict all our genes with FDSIs, it remains crucial that computational resources contain FDSIs and experimentalists ensure that they submit their sequence data to these resources.

Implications for alternative splicing's impact on gene function

Recent studies have challenged whether *most* genes can produce multiple functional splice isoforms and our results can offer something to both sides of the debate. We acknowledge that other researchers may have different definitions of a functional splice isoform, but we view the debate within our operational definition – a functional splice isoform is one that is necessary for the gene's overall function.

One side of the debate claims that most genes have multiple functionally distinct isoforms [1]. Viewing our findings optimistically, we provide what is to our knowledge the only substantial list of human and mouse genes for which this is actually documented to be true. The low number of genes with such evidence can be interpreted as a vast opportunity for experimentalists to identify the functions of the isoforms for > 80% of genes. The other side of the debate approaches alternative splicing with a less Panglossian view, with the null hypothesis being that most isoforms do *not* have a specific distinct function [39]. Multiple studies taking a genomic or evolutionary perspective have concluded that it is unlikely that most genes have multiple functional splice isoforms [3–6, 8–10, 40–44]. Viewed pessimistically, our data is consistent with this body of work. If the literature lacks supporting evidence for widespread FDSIs, the null hypothesis should be maintained and claims that every observed isoform has a function to be discovered should be viewed skeptically.

Conclusions

To our knowledge, this report represents the first effort to curate the literature in order to determine the genes where splicing increases the genome's functional potential. Such individual reports have been generally ignored in the debate about the function of alternative splicing, which has instead focused on databases and high-throughput data sets. Our estimate that only 4% of human and 9% of mouse genes have evidence for functionally distinct isoforms serves both a sobering reminder of the limited evidence, and a motivation for increased experimental efforts to settle the debate. At the same time, we also recognize there are likely genes with FDSIs that we did not curate and should be included. We invite contributions from the community of corrections or suggestions for papers or genes to curate, which can be sent to the authors. Future updates to our curation can be found in our dataverse (http://hdl.handle.net/11272/10628).

Methods

Determining the type of functional distinctness

We developed a scheme to describe non-mutually exclusive types of functional distinctness found in genes with FDSIs. We recognize two general biological mechanisms by which functional distinctness could arise, schematized in Fig. 1a, and elaborated on further below: "expression-pattern distinctness" or "biochemical distinctness". Figure 1b outlines our full scheme for classifying functional distinctness. The subclasses we identified were designed to accommodate how functional distinctness is reported in the literature we curated, that is, we did not create this classification wholly ab initio. We determined the type of functional distinctness using the publication which provided the evidence for FDSI, but some cases required an inference based on other literature by the authors. We stress that a gene can have multiple types of functional distinctness. For example, biochemically distinct isoforms could also have expression pattern distinctness. We annotated as many types of functional distinctness as were provided by the literature reports.

Expression-pattern distinctness

Expression-pattern distinctness requires the condition-dependent expression of isoforms of a single gene. Generally, in this category, splice isoforms of the same gene have functional relevance in distinct conditions. We further specified expression-pattern distinctness as "subcellular-localization-specific", "cell-type-specific", "tissue-specific", "developmental stage-specific", and "other--condition-specific". Thus, genes with cell-type-specific FDSIs express their splice isoforms in distinct cell types, and the elimination of expression of either splice isoform causes a phenotype (Fig. 1a). These isoforms' final products could be identical (that is, they are not biochemically distinct). However, they are still functionally distinct because they have partially different expression patterns and one cannot fully compensate for the other.

Biochemical distinctness

Biochemical distinctness is defined as differences in biochemical properties or activities, and which cannot compensate for each other even if co-expressed in the same condition. We further specified biochemical distinctness as "protein domain change", "dominant negative", "subcellular localization", "UTR change" and "protein terminus change". Genes categorized as FDSIs with distinct protein-domains indicate that each splice isoform has a unique structural or functional unit in their final protein product. We manually extracted information about the specific protein domain from the authors providing the evidence of functional distinctness. In some cases, this could involve the presence or absence of one or more protein domains. Genes categorized as "protein terminus change" indicates that the FDSIs' final protein product differ from each other either in their C-terminus or their N-terminus. These changes to the C- or N-termini usually

do not affect the presence or absence of protein domains (or the paper did not make any note of changes to protein domains). Genes with dominant-negative FDSIs have splice isoforms with antagonistic phenotypes. Typically, these splice isoforms regulate each other's function. The loss of one splice isoform generally affects the function of the other splice isoform. Gene categorized as "UTR change" indicates that the FDSIs of the same gene differ in the UTRs of the mRNA (coding regions may change as well).

Literature selection

On July 17th 2017, we generated a "starting set" of publications associated with human and mouse genes to curate using PubMed e-utilities and the search term "alternative splicing". From here curation was both "gene--centric" and "paper-centric."

Gene-centric curation

The gene-centric approach attempted to curate all relevant studies associated with a specific gene. PubMed linked each study from our starting set to a specific gene which provided a list of genes with literature. The genes we selected to curate from this list were genes suggested to us by the community, PULSE's training genes or commonly discussed by the literature [21]. As suggestions form the community might be biased, 100 random genes were also selected for gene-centric curation.

Paper-centric curation

The paper-centric approach attempted to curate literature likely enriched for evidence of genes with FDSIs. Using this approach, we make no attempt to curate all relevant reports for any specific gene. As a targeted source of literature likely to be enriched for functional evidence, we used review articles on the function of alternative splicing that provided citations for 603 genes [3, 11, 28, 45–47]. We further extended paper-centric curation with specific search phrases in PubMed. Search terms were: "functionally distinct splice isoforms", "CRISPR alternative splicing", "alternative splicing knockdown" and "alternative splicing knockout." These queries identified an additional 260 papers for our starting set of papers. The genes found in the publications retrieved by these PubMed queries and provided in the aforementioned reviews further informed us of which genes to gene-centrically curate. For example, *BDNF* and *XBP1* were commonly reviewed in the literature and consequently, we gene-centrically curated them.

Curation process

For each paper, a trained curator first identified general features of that study by manually extracting the following information: the investigated gene, the reported number of the splice isoforms for the gene, the names used by the authors for the splice isoforms, the number of splice isoforms specifically investigated in the paper ("the investigated isoforms"), the experiments performed, the organism where the gene was identified, the organism or cell line used for the experiments, and any claims of functional distinctness.

Next, using a decision tree (Fig. 2), we annotated each paper as to whether the data provided positive evidence of functional distinctness for the investigated splice isoforms. We sought evidence where the loss of one isoform (via knockdown, knockout or other means of isoform-specific depletion) produced a phenotype in the test system. We also curated experiments which performed overexpression analyses, which were retained as a separate category from the isoform loss studies (as an example, see study by Scotton and colleagues [48]). We did not accept studies of aberrant isoforms caused by rare mutations (for example in cancer), as we deemed these as not relevant to the normal function of the gene as we have defined it (as an example, see Cogan et al. [49]). If a study provided evidence where investigators depleted multiple splice isoforms of a single gene but at most one splice isoform caused a phenotype, we classified the gene as having *negative* support for FDSIs. Finally, regardless of study type, the curators provided a concise explanation of the functions investigated.

For our definition of genes with FDSIs, we required evidence for the independent depletion of at least two splice isoforms of the same gene. If the curated study investigated the outcome of the absence of a single isoform for a given gene, then that study alone insufficiently provides evidence of FDSIs. While such studies demonstrate an existence of a single functional isoform, the support for FDSIs requires data on at least two isoforms from the same gene. However, we subsequently attempted to identify a second paper that provided functional evidence for a different splice isoform of the same gene. In situations where a second paper identified evidence of a different functional splice isoform, we recorded the gene as having FDSIs.

Curator validation

To ensure consistent curation, we evaluated the curators. These tests consisted of all curators curating the exact same randomly selected 50 papers. After the test, we addressed any discrepancy between curators and we updated the curation standards with any necessary clarifications (curation standards provided in Additional file 1). This evaluation process was conducted three times. We also further scrutinized papers annotated as providing positive evidence of a gene with FDSI to eliminate any false positives.

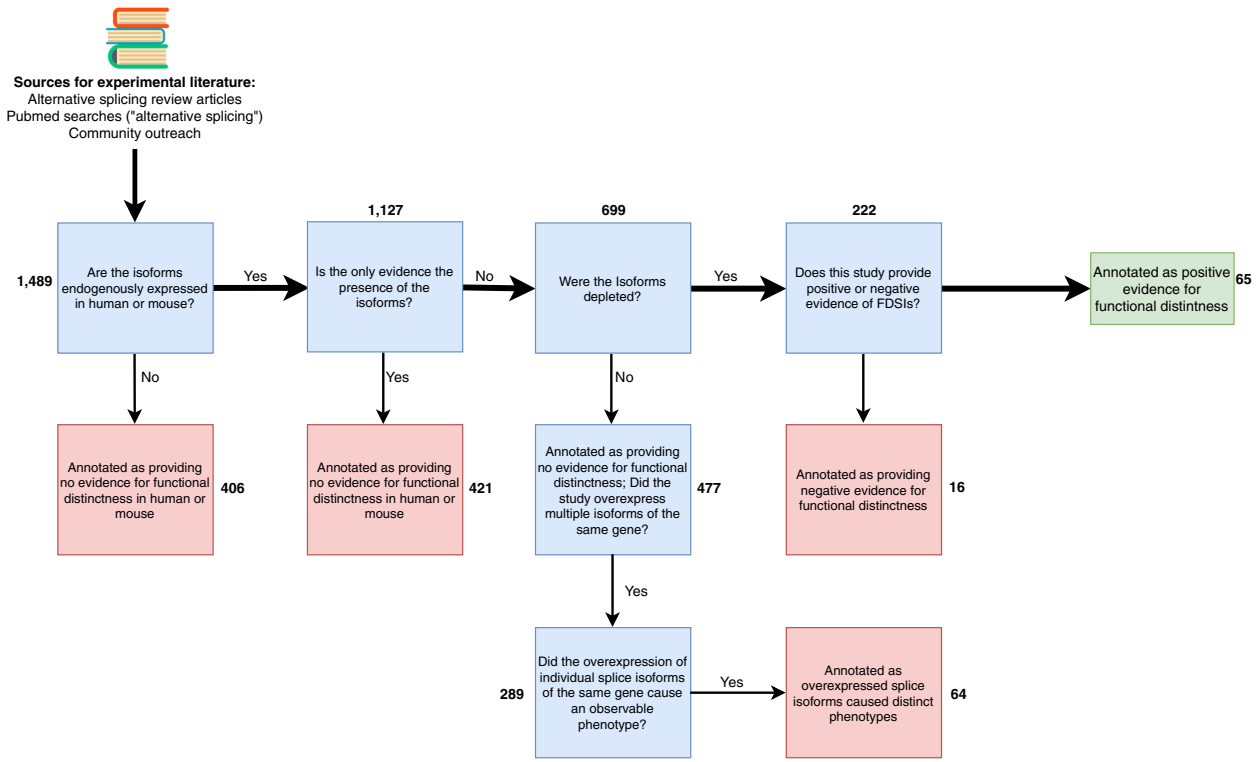

Fig. 2 Overview of literature curation scheme. We sought papers which study the functional distinctness of a single human or mouse gene's splice isoforms. Positive studies are those that provide evidence where multiple splice isoforms of a single gene are depleted and at least two isoforms show a phenotype. We annotated studies as providing negative evidence for functional distinctness when investigators deplete multiple splice isoforms of the same gene but only one produces an observable phenotype. The numbers in bold represent the number of studies in each category. Clip art designed from Flaticon (free license with attribution)

Linking FDSIs to Ensembl

If a paper provided positive evidence for FDSIs, we linked the splice isoforms with the appropriate Ensembl transcript ID. Generally, studies provided GenBank or RefSeq accession IDs and these accession IDs linked to Ensembl. In the absence of an accession ID, we referred to the literature for sequence information about the splice isoforms and aligned splice isoform sequences to Ensembl using ClustalOmega [50].

Computational predictions of genes with FDSIs

PULSE, a computational classifier developed by Hao et al., predicted 2419 of 15,639 UniProt genes to have multiple functional isoforms based on a training set of 145 genes [21]. We downloaded the supplementary data provided by Hao et al. to determine whether PULSE predicted our genes with FDSIs to have multiple functional splice isoforms. We also investigated whether any of our genes with FDSIs were part of PULSE training and validation set of genes. This was of interest because a training set enriched for genes with FDSIs may yield predictions for genes with FDSIs, even though PULSE was only designed to detect function, not distinct function. For our comparison to PULSE prediction, we used the human orthologue for any mouse gene with FDSI as determined by BioMart [51].

Additional files

Additional file 1: Contains explanations of each heading found in Additional file 2, standards curators used to evaluate studies for evidence of functionally distinct splice isoforms, three genes where literature reported splice isoforms were not found in Ensembl, supplemental tables and figures, and citations for all human and mouse literature curated for this study. (PDF 1624 kb)

Additional file 2: Contains a list of all human and mouse literature we curated for this study, and curators's annotations for each study. (XLSX 187 kb)

Additional file 3: Contains the list of human and mouse genes with literature evidence for functionally distinct splice isoforms found in our curation. (XLSX 23 kb)

Abbreviations

FDSI: Functionally distinct splice isoforms; UTR: Untranslated region

Acknowledgements

The authors would like to thank and acknowledge Dr. Sanja Rogic, Dr. Lilah Toker, Dr. Shreejoy Tripathy and Benjamin Callaghan for their useful feedback throughout the project. We also thank all contributors who provided genes or studies to our curation efforts.

Funding
This work was supported by a Natural Sciences and Engineering Research Council (http://www.nserc-crsng.gc.ca/) CREATE scholarship for the University of British Columbia Bioinformatics Training Program (awarded to SB), and Natural Sciences and Engineering Research Council Discovery grant (RGPIN-2016-05991) and National Institutes of Health (www.nih.gov) grant MH111099 to PP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials
Authors can confirm that all relevant data are included in the article and its Additional files.

Authors' contributions
SB and PP designed the overall project and curation framework. SB and PP supervised the curators. SL, MP, BH, EH, CL, JL curated the literature. All curators provided feedback to SB and PP for the curation framework. JL queried PubMed for literature to curate. SL referenced splice isoforms to genomic databases. SB determined types of functional distinctness. SB and PP prepared the manuscript. All authors have read and approved the manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Michael Smith Laboratories, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. [2]Department of Psychiatry, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. [3]Graduate Program in Bioinformatics, University of British Columbia, Vancouver, Canada.

Received: 6 June 2018 Accepted: 14 August 2018
Published online: 28 August 2018

References

1. Blencowe BJ. The Relationship between Alternative Splicing and Proteomic Complexity. Trends Biochem Sci [Internet]. 2017 [cited 2017 May 9];0. Available from: https://www.cell.com/trends/biochemical-sciences/fulltext/S0968-0004(17)30070-1.
2. Tress ML, Abascal F, Valencia A. Most Alternative Isoforms Are Not Functionally Important. Trends Biochem Sci [Internet]. 2017 [cited 2017 May 9]; Available from: http://www.sciencedirect.com/science/article/pii/S0968000417300713.
3. Tress ML, Abascal F, Valencia A. Alternative splicing may not be the key to proteome complexity. Trends Biochem Sci. 2017;42:98–110.
4. Light S, Elofsson A. The impact of splicing on protein domain architecture. Curr Opin Struct Biol. 2013;23:451–8.
5. Hsu S-N, Hertel KJ. Spliceosomes walk the line: splicing errors and their impact on cellular function. RNA Biol. 2009;6:526.
6. Melamud E, Moult J. Stochastic noise in splicing machinery. Nucleic Acids Res. 2009;37:4873–86.
7. Abascal F, Tress ML, Valencia A. Alternative splicing and co-option of transposable elements: the case of TMPO/LAP2α and ZNF451 in mammals. Bioinformatics. 2015;31:2257–61.
8. Pickrell JK, Pai AA, Gilad Y, Pritchard JK. Noisy splicing drives mRNA isoform diversity in human cells. PLoS Genet. 2010;6:e1001236.
9. Reyes A, Anders S, Weatheritt RJ, Gibson TJ, Steinmetz LM, Huber W. Drift and conservation of differential exon usage across tissues in primate species. Proc Natl Acad Sci. 2013;110:15377–82.
10. Saudemont B, Popa A, Parmley JL, Rocher V, Blugeon C, Necsulea A, et al. The fitness cost of mis-splicing is the main determinant of alternative splicing patterns. Genome Biol. 2017;18:208.
11. Lipscombe D, Andrade A, Allen SE. Alternative splicing: functional diversity among voltage-gated calcium channels and behavioral consequences. Biochim Biophys Acta. 1828;2013:1522–9.
12. Kriventseva EV, Koch I, Apweiler R, Vingron M, Bork P, Gelfand MS, et al. Increase of functional diversity by alternative splicing. Trends Genet. 2003;19:124–8.
13. Stetefeld J, Ruegg MA. Structural and functional diversity generated by alternative mRNA splicing. Trends Biochem Sci. 2005;30:515–21.
14. Lipscombe D, Pan JQ, Gray AC. Functional diversity in neuronal voltage-gated calcium channels by alternative splicing of ca(v)alpha1. Mol Neurobiol. 2002;26:21–44.
15. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, et al. The Ensembl gene annotation system. Database. 2016;2016:baw093.
16. Bely B, Martin MJ, Apweiler R. Source of annotations in the UniProt Knowledgebase. F1000Posters [Internet]. 2010 [cited 2014 May 29];1. Available from: http://f1000.com/posters/browse/summary/243
17. Zhao S, Zhang B. A comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. BMC Genomics. 2015;16:97.
18. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Studying Gene Expression and Function. 2002 [cited 2017 Apr 19]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK26818/
19. Shehu A, Barbará D, Molloy K. A Survey of Computational Methods for Protein Function Prediction. In: Wong K-C, editor. Big Data Anal Genomics [Internet]. Springer International Publishing; 2016 [cited 2016 Oct 31]. p. 225–98. Available from: http://link.springer.com/chapter/10.1007/978-3-319-41279-5_7
20. Gibson TJ, Seiler M, Veitia RA. The transience of transient overexpression [Internet]. Nat. Methods. 2013 [cited 2018 Apr 24]. Available from: https://www.nature.com/articles/nmeth.2534
21. Hao Y, Colak R, Teyra J, Corbi-Verge C, Ignatchenko A, Hahne H, et al. Semi-supervised learning predicts approximately one third of the alternative splicing isoforms as functional proteins. Cell Rep. 2015;12:183–9.
22. Worton LE, Shi Y-C, Smith EJ, Barry SC, Gonda TJ, Whitehead JP, et al. Ectodermal-neural cortex 1 isoforms have contrasting effects on MC3T3-E1 osteoblast mineralization and gene expression. J Cell Biochem. 2017;118:2141–50.
23. Ma X, Kawamoto S, Uribe J, Adelstein RS. Function of the neuron-specific alternatively spliced isoforms of nonmuscle myosin II-B during mouse brain development. Mol Biol Cell. 2006;17:2138–49.
24. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456:470–6.
25. Auboeuf D. Alternative mRNA processing sites decrease genetic variability while increasing functional diversity. Transcription. 2018;9:75–87.
26. Frankish A, Mudge JM, Thomas M, Harrow J. The importance of identifying alternative splicing in vertebrate genome annotation. Database J Biol Databases Curation. 2012;2012:bas014.
27. Mudge JM, Frankish A, Fernandez-Banet J, Alioto T, Derrien T, Howald C, et al. The origins, evolution, and functional potential of alternative splicing in vertebrates. Mol Biol Evol. 2011;28:2949–59.
28. Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, et al. Function of alternative splicing. Gene. 2013;514:1–30.
29. Kennedy D. The old file-drawer problem. Science. 2004;305:451.
30. Candi E, Rufini A, Terrinoni A, Dinsdale D, Ranalli M, Paradisi A, et al. Differential roles of p63 isoforms in epidermal development: selective genetic complementation in p63 null mice. Cell Death Differ. 2006;13:1037–47.
31. Coldwell MJ, Sack U, Cowan JL, Barrett RM, Vlasak M, Sivakumaran K, et al. Multiple isoforms of the translation initiation factor eIF4GII are generated via use of alternative promoters, splice sites and a non-canonical initiation codon. Biochem J. 2012;448:1–11.
32. Gannett L. What's in a cause?: the pragmatic dimensions of genetic explanations. Biol Philos. 1999;14:349–73.
33. Gifford F. Genetic traits. Biol Philos. 1990;5:327–47.
34. Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. Cell. 2018;172:393–407.
35. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008;40:1413–5.
36. Hyung D, Kim J, Cho SY, Park C. ASpedia: a comprehensive encyclopedia of human alternative splicing. Nucleic Acids Res. 2018;46:D58–63.

37. Rodriguez JM, Rodriguez-Rivas J, Di Domenico T, Vázquez J, Valencia A, Tress ML. APPRIS 2017: principal isoforms for multiple gene sets. Nucleic Acids Res [Internet]. 2018 [cited 2017 Oct 23]; Available from: https://academic.oup.com/nar/article/doi/10.1093/nar/gkx997/4561658/APPRIS-2017-principal-isoforms-for-multiple-gene
38. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. Nat Genet. 2013;45:580–5.
39. Gould SJ, Lewontin RC. The spandrels of san Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond B. 1979;205:581–98.
40. Hu J, Boritz E, Wylie W, Douek DC. Stochastic principles governing alternative splicing of RNA. PLoS Comput Biol. 2017;13:e1005761.
41. Abascal F, Ezkurdia I, Rodriguez-Rivas J, Rodriguez JM, del Pozo A, Vázquez J, et al. Alternatively spliced homologous exons have ancient origins and are highly expressed at the protein level. PLoS Comput Biol. 2015;11:e1004325.
42. Kurmangaliyev YZ, Gelfand MS. Computational analysis of splicing errors and mutations in human transcripts. BMC Genomics. 2008;9:13.
43. Wang M, Zhang P, Shu Y, Yuan F, Zhang Y, Zhou Y, et al. Alternative splicing at GYNNGY 5′ splice sites: more noise, less regulation. Nucleic Acids Res. 2014;42:13969–80.
44. Zhang Z, Xin D, Wang P, Zhou L, Hu L, Kong X, et al. Noisy splicing, more than expression regulation, explains why some exons are subject to nonsense-mediated mRNA decay. BMC Biol. 2009;7:23.
45. Kovacs E, Tompa P, Liliom K, Kalmar L. Dual coding in alternative reading frames correlates with intrinsic protein disorder. Proc Natl Acad Sci U S A. 2010;107:5429–34.
46. Ramanouskaya TV, Grinev VV. The determinants of alternative RNA splicing in human cells. Mol Gen Genomics. 2017:1–21.
47. Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, et al. Function of alternative splicing. Gene. 2005;344:1–20.
48. Scotton P, Bleckmann D, Stebler M, Sciandra F, Brancaccio A, Meier T, et al. Activation of muscle-specific receptor tyrosine kinase and binding to dystroglycan are regulated by alternative mRNA splicing of agrin. J Biol Chem. 2006;281:36835–45.
49. Cogan J, Austin E, Hedges L, Womack B, West J, Loyd J, et al. Role of BMPR2 alternative splicing in heritable pulmonary arterial hypertension penetrance. Circulation. 2012;126:1907–16.
50. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. Mol Syst Biol. 2011;7:539.
51. Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, et al. BioMart--biological queries made easy. BMC Genomics. 2009;10:22.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions