D58–D63 Nucleic Acids Research, 2018, Vol. 46, Database issue
doi: 10.1093/nar/gkx1014
Published online 2 November 2017

ASpedia: a comprehensive encyclopedia of human alternative splicing

Daejin Hyung[1,2,†], Jihyun Kim[1,†], Soo Young Cho[1,*] and Charny Park[1,*]

[1]Research Institute, National Cancer Center, 323 Ilsan-ro, Goyang-si, Kyeonggi-do 10408, Republic of Korea and [2]Department of Computer Engineering, Dong-A University, 37 Nakdong-daero 550 Beon-gil, Saha-gu, Busan 49315, Republic of Korea

Received August 11, 2017; Revised October 02, 2017; Editorial Decision October 12, 2017; Accepted October 12, 2017

ABSTRACT

Alternative splicing confers the human genome complexity by increasing the diversity of expressed mRNAs. Hundreds or thousands of splicing regions have been identified through differential alternative splicing analysis of high-throughput datasets. However, it is hard to explain the functional impact of each splicing event. Protein domain formation and nonsense-mediated decay are considered the main functional features of splicing. However, other functional features such as miRNA target sites, phosphorylation sites and single-nucleotide variations are directly affected by alternative splicing and affect downstream function. Hence, we established ASpedia: a comprehensive database for human alternative splicing annotation, which encompasses a range of functions, from genomic annotation to isoform-specific function (ASpedia, http://combio.snu.ac.kr/aspedia). The database provides three features: (i) genomic annotation extracted from DNA, RNA and proteins; (ii) transcription and regulation elements analyzed from next-generation sequencing datasets; and (iii) isoform-specific functions collected from known and published datasets. The ASpedia web application includes three components: an annotation database, a retrieval system and a browser specialized in the identification of human alternative splicing events. The retrieval system supports multiple AS event searches resulting from high-throughput analysis and the AS browser comprises genome tracks. Thus, ASpedia facilitates the systemic annotation of the functional impacts of multiple AS events.

INTRODUCTION

Alternative splicing (AS) events can be estimated from ∼95% of multi exon human genes, and these confer functional diversity to the gene (1). Various cases of AS events have been reported to play key roles in molecular and cellular functions (2). Currently, AS events are inferred mainly using exon tiling microarray or RNA-Seq. High-throughput platform analysis focuses on precise statistical model development to estimate differential AS (3–5). Although many statistical methods have been proposed, the functional impacts of many AS event sites remain unknown. To understand the functionality of AS, protein evidences such as protein domain loss and nonsense-mediated decay (NMD) have been used. For example, a stand-alone application, AltAnalyze, tests differential splicing and integrates the splicing events with protein domains, miRNA-binding sites, molecular interactions or pathways (6). SpliceR is an R package that can be used to predict protein coding potential and NMD (7). Exon ontology (EXONT) offers functional impact terms derived from various protein features, and PoSAS identifies functional protein structures from splicing sites and isoforms (8,9). Several AS databases such as ASAP II, ASD and H-DBAS have been established for non-protein evidence, including splicing event classification and cross-species conservation (10–12). Tissue and disease-specific splicing events have also been cataloged in TCGASpliceSeq and ASPA II (10,13). These applications predict protein activities induced by AS. However, the functional impact of AS is not determined only by the protein domain or protein-coding potential.

Recently, novel clues for AS functionality have accumulated in various genomic domains. Functional phosphorylation sites and subcellular localization on AS sites have been systemically investigated in epithelial-mesenchymal transition cells and breast cancer cell lines (8). Isoform-specific protein interaction network has been built at a genome-wide level (14). AU-rich repeat elements of 3′ UTR retained intron of beta-catenin have been shown to influ-

*To whom correspondence should be addressed. Tel: +82 31 920 2581; Email: charn78@ncc.re.kr
Correspondence may also be addressed to Soo Young Cho. Tel: +82 31 920 2556; Email: sooycho@ncc.re.kr
†These authors contributed equally to the paper as first authors.

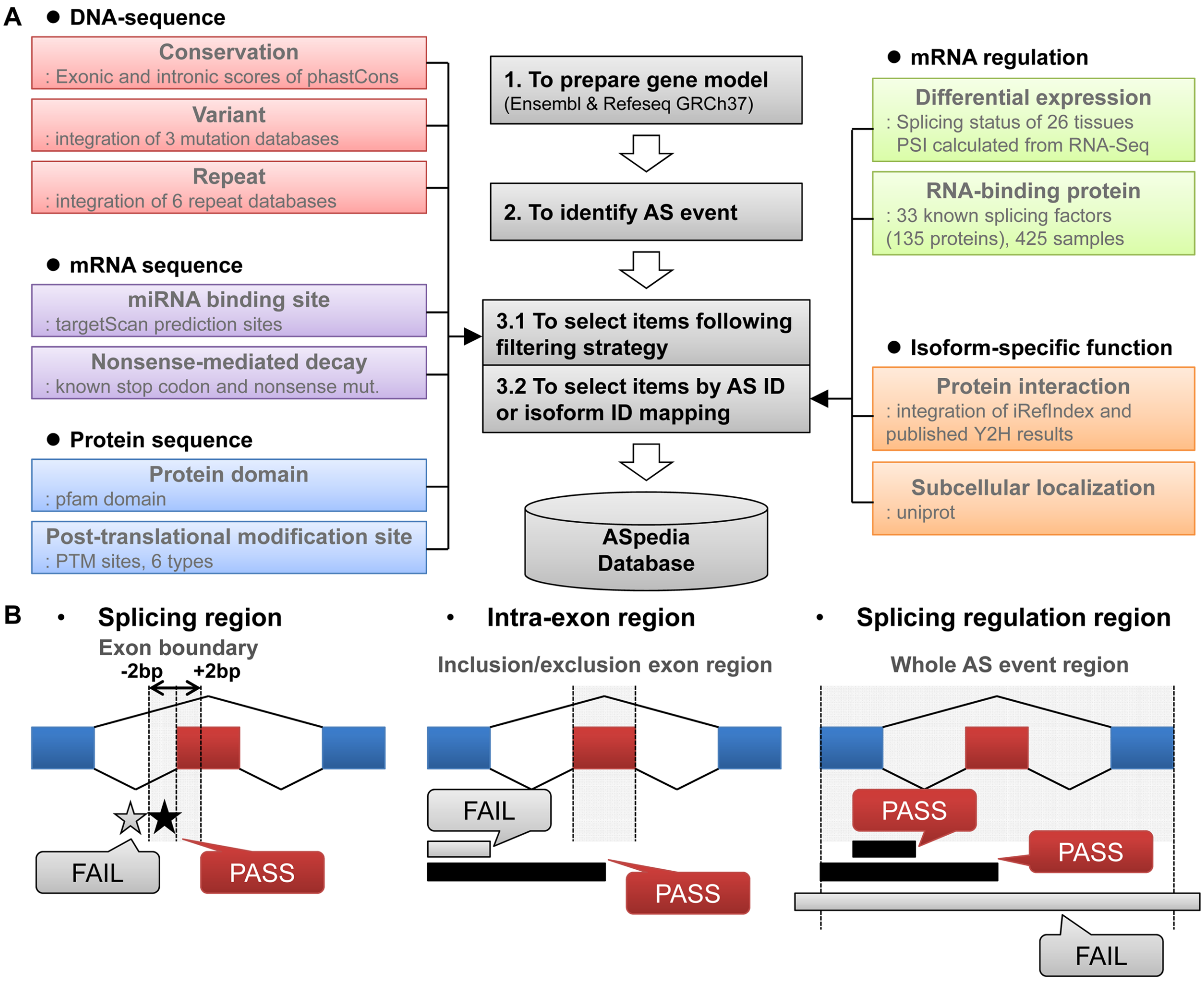
A
● DNA-sequence
Conservation
: Exonic and intronic scores of phastCons
Variant
: integration of 3 mutation databases
Repeat
: integration of 6 repeat databases
● mRNA sequence
miRNA binding site
: targetScan prediction sites
Nonsense-mediated decay
: known stop codon and nonsense mut.
● Protein sequence
Protein domain
: pfam domain
Post-translational modification site
: PTM sites, 6 types
1. To prepare gene model
(Ensembl & Refeseq GRCh37)
2. To identify AS event
3.1 To select items following filtering strategy
3.2 To select items by AS ID or isoform ID mapping
ASpedia Database
● mRNA regulation
Differential expression
: Splicing status of 26 tissues PSI calculated from RNA-Seq
RNA-binding protein
: 33 known splicing factors (135 proteins), 425 samples
● Isoform-specific function
Protein interaction
: integration of iRefIndex and published Y2H results
Subcellular localization
: uniprot
B
• Splicing region
Exon boundary
-2bp
+2bp
FAIL
PASS
• Intra-exon region
Inclusion/exclusion exon region
FAIL
PASS
• Splicing regulation region
Whole AS event region
PASS
PASS
FAIL

Figure 1. (**A**) Database construction overview of ASpedia. All annotation database components are showed in left and right side. Each annotation item associated with AS event were selected by workflow step 1∼3 in the center. (**B**) Sequence filtering strategy is classified into three cases and is applied in workflow step 4 of (A).

ence the stability of the human beta-catenin mRNA (15). miRNA let-7D binds to the splicing site of 3′ UTR of DMT1, and the miRNA controls its regulation (16). Variants or point mutations that affect splicing have been systematically predicted (17). To reveal these novel functional features induced by AS, we need a database expansion to annotate AS events and a retrieval system to query splicing sites.

Here, we introduce ASpedia, a comprehensive human AS database encompassing genomic features, as well as isoform-level functions. Genomic features associated with the functional impact of AS were investigated from DNA, RNA and protein sequences. Large-scale transcriptional regulation and its elements were also integrated. ASpedia supports a retrieval system that identifies unique AS IDs and a browser to visualize each AS event. The retrieval system can also be used for searching multiple AS sites from the results of differential AS analysis using RNA-Seq. We believe that ASpedia could be an extensive annotation application and a powerful search tool. More details, a user manual, and statistics are provided on the ASpedia web site (http://combio.snu.ac.kr/aspedia).

DATABASE OVERVIEW

ASpedia database was established from the integration of AS events inferring from gene models and associated annotation dataset. The integration steps are shown in Figure 1A. First, human genome GRCh37 Ensembl release 82 and RefSeq release 105 were prepared as reference gene model, we identifies AS events using AStalavista (18). Then, the events were classified into five types: alternative 3′ splice site (A3SS), alternative 5′ splice site (A5SS), skipping exon (SE), mutual exclusive exon (MXE) and retained intron (RI). Functional annotation information associated with AS events was collected from various datasets: DNA, mRNA, protein, mRNA regulation and isoform-specific function.

The characteristics and mining status of each annotation dataset are described in next section. Annotation dataset shown Figure 1A left items include genomic coordinate information. These items have to exist in particular AS region. For example, functional miRNA-binding site depending on AS should locate in exclusion or inclusion exon. Therefore, we designed filtering strategy that genomic co-

Table 1. Summary statistics of alternative splicing events and annotation items for each gene model

Status		Ensembl		RefSeq	
		Gene	AS event (or isoform)	Gene	AS event (or isoform)
Gene model	Total gene	57 773		27 778	
	AS gene/AS	22 183	39 804	15 058	26 918
Annotation database	NMD	2849	5120	1754	2722
	miRNA binding site	1118	1911	1034	1582
	Repeat	4823	9017	3019	5208
	Protein domain	7303	19 162	5556	11 060
	PTM	9888	32 516	8505	20 752
	Protein interaction	2445	43 172	2355	44 145
	Subcellular localization	1105	1711	1327	3384

All annotations were assigned distinct gene IDs and matched with AS events. Protein interactions and subcellular localizations were assigned unique transcript IDs.

ordinate of annotation item precisely belongs to particular region shown in Figure 1B. The rest annotation dataset, mRNA regulation and isoform-specific function were simply identified by mapping with isoform IDs or AS event Ids. Specific filtering strategy and ID mapping status for each AS annotation item are described in next section. ASpedia database was finally constructed after proceeding annotation data selection step. The overall database components and workflow are shown in Figure 1A and the status and counts of the databases are summarized in Table 1.

Genomic annotation of AS event regions

The genomic annotation database was extracted from DNA, RNA and protein sequences. All functional sequences around splicing sites were collected. We then eliminated noise sequences to avoid conferring AS specificity. The filtering strategies are categorized in three cases, as shown in Figure 1B. First, a splicing region filtering strategy selects genomic sequences around the exon boundary (±2). Variants belong to this type. Repeats, miRNA-binding sites, protein domains and post-translational modification (PTM) sites play a functional role in only inclusion or exclusion exon regions. These belong to the intra-exon region filtering case as shown in Figure 1B. The splicing regulatory region shown in Figure 1B indicates regions that could affect the splicing regulatory machinery. The details for establishing each annotation database are described below.

- *Evolutionary conservation*: we calculated the conservation scores in terms of average score from exons and introns in AS structure. The original data were obtained from phastCons100ways and phastConst45ways of the UCSC genome browser (19). The scores for datasets of primates, placental mammals and vertebrates were calculated separately.
- *Variants*: single nucleotide variants (SNVs) and somatic mutations around donor and acceptor sites that affect splicing were investigated from dbSNP v138 and COSMIC v77 (20,21). We supplied disease-associated SNVs to irregulate splicing from SPIDEX (17). The result was also integrated into this category.
- *miRNA-binding sites*: we referred to the results of targetScan prediction from the UCSC genome browser. The miRNA binding sites were considered to overlap with 3′ UTR regions.
- *Repeat sequences*: we collected five repeat databases from the UCSC genome browser: interrupted Rpts, Microsatellite, RepeatMasker, SelfChain and Simple Repeats (19). RepeatMasker provides 10 different classes of repeats. All repeat sequences were filtered using the splicing regulation region filtering strategy and repeat sequence size. The cut-off of repeat sequence size was decided by confirming the distribution described in Supplementary Section 1 and Figure S1.
- *NMD sites*: NMD sites were inferred from known and novel stop-codons located >50 nt upstream of the 3′-end of the splice-generated exon-exon junction. Known stop-codons were inferred using the gene model data, and novel stop codons were inferred from nonsense mutations in dbSNP and COSMIC.
- *Protein domains*: Pfam domains were investigated for all transcript sequences (22). To match with AS events, each domain was represented as its genomic coordinate. Next, we extracted domains that overlapped with AS regions using the intra-exon filtering strategy.
- *PTM sites*: nine types of PTM sites were collected from phosphoSitePlus (23). Only PTMs in the intra-exon region were suggested to confer functional impact.

Transcriptional regulation and its elements

To establish a database for transcriptional regulation, we investigated tissue-specific differential AS events and splicing regulatory elements known as RNA-binding protein from various NGS platforms. Tissue-specific AS events were collected from the EBI ArrayExpress RNA-Seq (Accession ID E-MTAB-1733) (24). We built up a RNA-Seq alignment pipeline by referring to hg19. This dataset of 26 tissues from 241 samples was aligned using STAR after read quality trimming using Trimmomatic (25,26). A representative AS event measurement, percent spliced-in index (PSI), signifies the exon inclusion percentage corresponding to the AS event. The PSI value for each tissue was estimated by rMATS (27).

The ASpedia database supports transcription regulatory element information including splicing factors to elucidate splicing machinery. RBP NGS datasets (RIP-Seq and CLIP-Seq) were collected from the ENCODE project. The

Table 2. RNA-binding protein dataset status analyzed from ENCODE CLIP-Seq and RIP-Seq

Platform	RBP	Total sample	Splicing factor (SF)	SF sample
eCLIP-Seq	112	322	33	102
iCLIP-Seq	5	10	3	6
RIP-Seq	23	93	5	25
Total	135	425	38	133

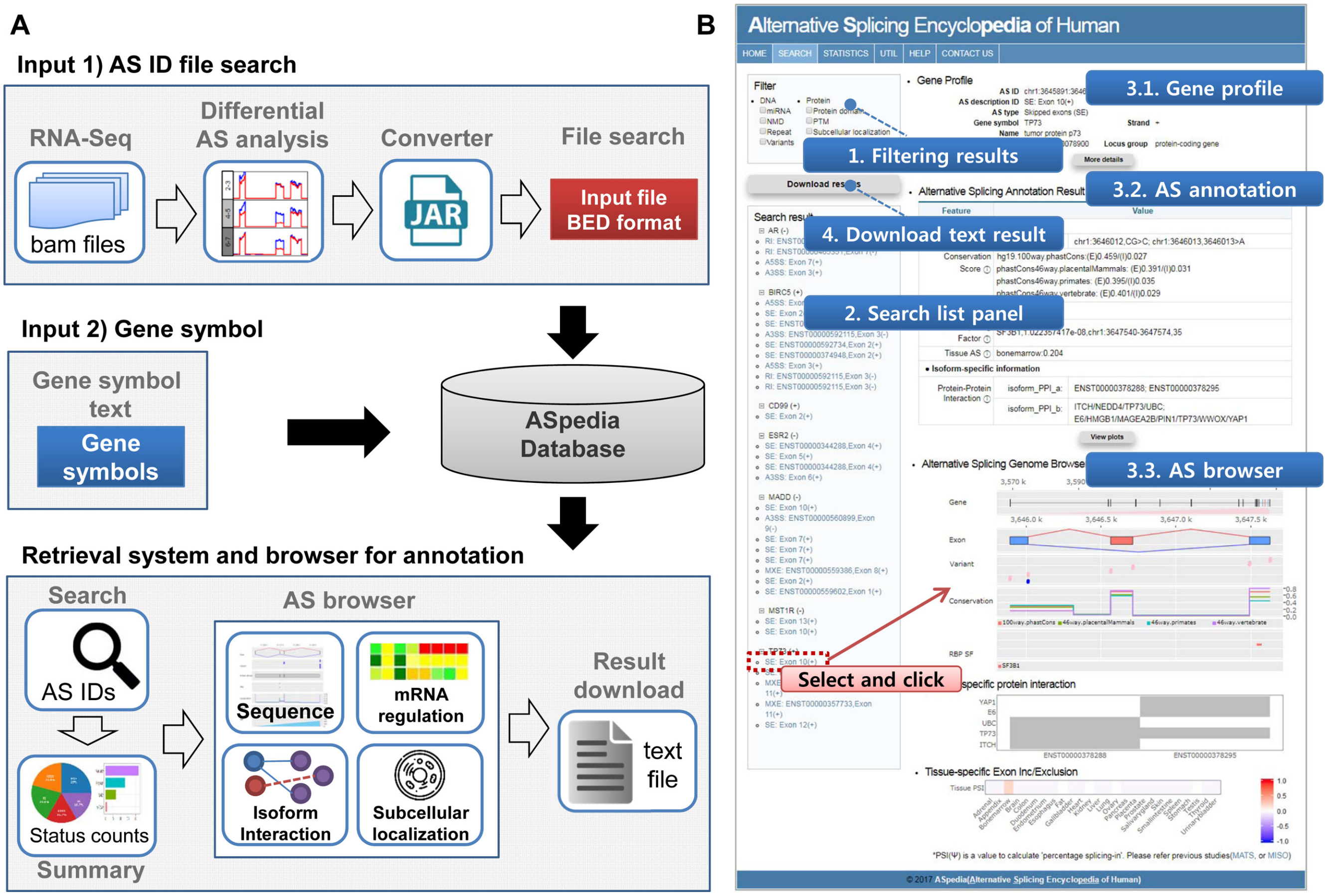

Figure 2. (**A**) System overview of ASpedia. Our system allows two input types both gene names and BED file. Retrieval system and browser are specialized to confirm each AS event. Final results can be downloaded in a text file format. (**B**) User interface of ASpedia search results. Left panel shows filtering options and the matched AS event list. Right panel presents AS event genomic profiles, text result and AS browser.

status of the RBP dataset is summarized in Table 2. The specific RBP analysis workflow is described in Supplementary Section 2 and Figure S2. Three peak calling methods were used for each platform (28–30). The results were filtered out based on the *P*-value, peak length, RPKM and splicing regulation region filtering strategy. Peak length and RPKM cut-offs were decided based on the length distribution for each of the platforms shown in Supplementary Figure S3.

Isoform-specific functions

The protein interaction and subcellular localization are characterized at the isoform level. An isoform-specific protein interaction database was integrated from genome-wide yeast two-hybrids and the protein interaction database iRefIndex (14,31,32). The subcellular localization data were collected from UniProt. We converted promiscuous IDs of raw datasets to Ensembl and RefSeq transcript IDs. Finally, we matched the transcript IDs of an isoform-specific function database to the corresponding transcript IDs of AS events. Over 2300 and 1300 RefSeq genes were found for protein interaction and subcellular localization, respectively.

SYSTEM OVERVIEW

The ASpedia system contains three main components: an annotation database, a retrieval system and a browser. The system can be queried using two input formats, shown in Figure 2A. In the annotation database, we defined key AS IDs originating from chromosomal position without any gene model dependency and the keys have one-to-one relationships with AS events. These IDs can be used to search AS events and annotate relevant information. The file-based search allows the user to explore multiple AS event sites. The system requires a BED format file, including matched AS IDs. The file-based AS event query is suitable for annotating differential AS analysis results of RNA-Seq. Furthermore, our system supports the file converter for RNA-Seq differential AS analysis results. The converter makes several program results to the ASpedia BED format. The converter jar execution file is downloadable from the ASpedia web site, and it is developed under Java version 1.6. Specific usage to query differential alternative splicing analysis results is described in the Supplementary Section 3, Figure S4–5 and web site manual. Thus, users can easily prepare ASpedia input files. The gene symbol-based search mode requires only the gene name list as the input. ASpedia searches all saved AS events for the input genes.

Search results are categorized and visualized in the ASpedia browser. The mapping annotation status and the search result count are summarized in the first result web page. The search results can be confirmed in the left side panel of ASpedia, and the list can be selected using the different filtering options shown in Figure 2B. Each AS event can be visualized in the browser panel with basic genomic information. Gene-tracks are extensible depending on genomic annotation and RBP peak datasets. PSI values of tissue-specific AS events and isoform-specific function are represented in the heatmap figure. The user can also download annotation results in a tab-delimited text file format and get all AS browser figures via e-mail. The detailed user manual is available on ASpedia web site.

DISCUSSION

ASpedia offers an alternative splicing annotation system integrating heterogeneous genomic information. The database includes unique novel contents like isoform-specific protein interaction as well as fundamental information like protein domains. To support user convenience, each annotated AS event is presented as genome track figure in the AS browser. An additional advantage is the ability to query multiple AS events estimated from differential AS analysis using RNA-Seq. Finally, our AS annotation result supports crucial evidence to decide the functional impact of AS.

ASpedia was established to support gene models for Ensembl and RefSeq. These gene models confer data source dependency to construct the database. ASpedia has a limitation to identify custom gene models or novel exons. GENCODE or UCSC Known gene shares considerable gene models with Ensembl and RefSeq, but a few unique AS events existing in custom gene models remain missing. To reduce user data loss by gene model dependency, we will extend the database to include well-known gene models, and consider adding human genome version hg18 and GRCh39.

Known isoform-specific protein interactions were relatively unrevealed; therefore, published dataset are insufficient. Hence, isoform-specific functions are rarely collected in the ASpedia database. In the next update, we plan to quantitatively and qualitatively complement isoform-specific function. Several prediction algorithms or novel analysis approaches can be considered for updating this database, such as subcellular localization prediction (8,33).

Despite the functional importance and diversity of human AS events, approaches for the functional investigation of AS events at the genome-wide level are insufficient. ASpedia resolves this issue. We believe that ASpedia can be an excellent tool for understanding and studying the functional impact of AS at the genome scale. In further studies, we plan to maintain and expand this database with well-characterized protein features and disease-specific transcript regulation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Cancer Center Grant (NCC-1611200) through National Cancer Center Korea; Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A1B03930929); the Collaborative Genome Program for Fostering New Post-Genome industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C9A3064815).

Conflict of interest statement. None declared.

REFERENCES

1. Pan,Q., Shai,O., Lee,L.J., Frey,B.J. and Blencowe,B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, **40**, 1413–1415.
2. Kelemen,O., Convertini,P., Zhang,Z., Wen,Y., Shen,M., Falaleeva,M. and Stamm,S. (2013) Function of alternative splicing. *Gene*, **514**, 1–30.
3. Clark,T.A., Schweitzer,A.C., Chen,T.X., Staples,M.K., Lu,G., Wang,H., Williams,A. and Blume,J.E. (2007) Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol.*, **8**, R64.
4. Katz,Y., Wang,E.T., Airoldi,E.M. and Burge,C.B. (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods*, **7**, 1009–1015.
5. Shen,S., Park,J.W., Huang,J., Dittmar,K.A., Lu,Z.X., Zhou,Q., Carstens,R.P. and Xing,Y. (2012) MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. *Nucleic Acids Res.*, **40**, e61.
6. Emig,D., Salomonis,N., Baumbach,J., Lengauer,T., Conklin,B.R. and Albrecht,M. (2010) AltAnalyze and DomainGraph: analyzing and visualizing exon expression data. *Nucleic Acids Res.*, **38**, W755–W762.
7. Vitting-Seerup,K., Porse,B.T., Sandelin,A. and Waage,J. (2014) spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics*, **15**, 81.
8. Tranchevent,L.C., Aube,F., Dulaurier,L., Benoit-Pilven,C., Rey,A., Poret,A., Chautard,E., Mortada,H., Desmet,F.O., Chakrama,F.Z. *et al.* (2017) Identification of protein features encoded by alternative exons using Exon Ontology. *Genome Res.*, **27**, 1087–1097.
9. Birzele,F., Kuffner,R., Meier,F., Oefinger,F., Potthast,C. and Zimmer,R. (2008) ProSAS: a database for analyzing alternative splicing in the context of protein structures. *Nucleic Acids Res.*, **36**, D63–D68.
10. Kim,N., Alekseyenko,A.V., Roy,M. and Lee,C. (2007) The ASAP II database: analysis and comparative genomics of alternative splicing in 15 animal species. *Nucleic Acids Res.*, **35**, D93–D98.
11. Thanaraj,T.A., Stamm,S., Clark,F., Riethoven,J.J., Le Texier,V. and Muilu,J. (2004) ASD: the Alternative Splicing Database. *Nucleic Acids Res.*, **32**, D64–D69.

12. Takeda,J., Suzuki,Y., Sakate,R., Sato,Y., Gojobori,T., Imanishi,T. and Sugano,S. (2010) H-DBAS: human-transcriptome database for alternative splicing: update 2010. *Nucleic Acids Res.*, **38**, D86–D90.
13. Ryan,M., Wong,W.C., Brown,R., Akbani,R., Su,X., Broom,B., Melott,J. and Weinstein,J. (2016) TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res.*, **44**, D1018–D1022.
14. Yang,X., Coulombe-Huntington,J., Kang,S., Sheynkman,G.M., Hao,T., Richardson,A., Sun,S., Yang,F., Shen,Y.A., Murray,R.R. *et al.* (2016) Widespread expansion of protein interaction capabilities by alternative splicing. *Cell*, **164**, 805–817.
15. Thiele,A., Nagamine,Y., Hauschildt,S. and Clevers,H. (2006) AU-rich elements and alternative splicing in the beta-catenin 3'UTR can influence the human beta-catenin mRNA stability. *Exp. Cell Res.*, **312**, 2367–2378.
16. Andolfo,I., De Falco,L., Asci,R., Russo,R., Colucci,S., Gorrese,M., Zollo,M. and Iolascon,A. (2010) Regulation of divalent metal transporter 1 (DMT1) non-IRE isoform by the microRNA Let-7d in erythroid cells. *Haematologica*, **95**, 1244–1252.
17. Xiong,H.Y., Alipanahi,B., Lee,L.J., Bretschneider,H., Merico,D., Yuen,R.K., Hua,Y., Gueroussov,S., Najafabadi,H.S., Hughes,T.R. *et al.* (2015) The human splicing code reveals new insights into the genetic determinants of disease. *Science*, **347**, 1254806.
18. Foissac,S. and Sammeth,M. (2007) ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Res.*, **35**, W297–W299.
19. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
20. Forbes,S.A., Beare,D., Gunasekaran,P., Leung,K., Bindal,N., Boutselakis,H., Ding,M., Bamford,S., Cole,C., Ward,S. *et al.* (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Aids Res.*, **43**, D805–D811.
21. Coordinators,N.R. (2017) Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **45**, D12–D17.
22. Finn,R.D., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Mistry,J., Mitchell,A.L., Potter,S.C., Punta,M., Qureshi,M., Sangrador-Vegas,A. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
23. Hornbeck,P.V., Zhang,B., Murray,B., Kornhauser,J.M., Latham,V. and Skrzypek,E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.*, **43**, D512–D520.
24. Fagerberg,L., Hallstrom,B.M., Oksvold,P., Kampf,C., Djureinovic,D., Odeberg,J., Habuka,M., Tahmasebpoor,S., Danielsson,A., Edlund,K. *et al.* (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics*, **13**, 397–406.
25. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
26. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
27. Shen,S., Park,J.W., Lu,Z.X., Lin,L., Henry,M.D., Wu,Y.N., Zhou,Q. and Xing,Y. (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E5593–E5601.
28. Kucukural,A., Ozadam,H., Singh,G., Moore,M.J. and Cenik,C. (2013) ASPeak: an abundance sensitive peak detection algorithm for RIP-Seq. *Bioinformatics*, **29**, 2485–2486.
29. Li,Y., Zhao,D.Y., Greenblatt,J.F. and Zhang,Z. (2013) RIPSeeker: a statistical package for identifying protein-associated transcripts from RIP-seq experiments. *Nucleic Acids Res.*, **41**, e94.
30. Lovci,M.T., Ghanem,D., Marr,H., Arnold,J., Gee,S., Parra,M., Liang,T.Y., Stark,T.J., Gehman,L.T., Hoon,S. *et al.* (2013) Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.*, **20**, 1434–1442.
31. Corominas,R., Yang,X., Lin,G.N., Kang,S., Shen,Y., Ghamsari,L., Broly,M., Rodriguez,M., Tam,S., Trigg,S.A. *et al.* (2014) Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat. Commun.*, **5**, 3650.
32. Razick,S., Magklaras,G. and Donaldson,I.M. (2008) iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics*, **9**, 405.
33. Petersen,T.N., Brunak,S., von Heijne,G. and Nielsen,H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, **8**, 785–786.