ASD: the Alternative Splicing Database

T.A. Thanaraj[1,*], Stefan Stamm[2], Francis Clark[3], Jean-Jack M. Riethoven[1], Vincent Le-Texier[1] and Juha Muilu[1,4]

[1]EMBL Outstation - European Bioinformatics Institute (EBI), Hinxton, Cambridge CB10 1SD, UK ; [2]University of Erlangen Nurenburg, Department of Biochemistry, Fahrstrasse 17, 91054 Erlangen, Germany ; [3]Advanced Computational Modelling Centre, University of Queensland, St. Lucia, 4072, Australia ; [4]Current address: Finnish Genome Centre, P.O. Box 63, 00014 Helsinki, Finland.
[*]To whom correspondence should be addressed; Fax +44 (0)1223 494468; e-mail thanaraj@ebi.ac.uk

INTRODUCTION. Alternative splicing is widespread in mammalian gene expression, and variant splice patterns are often specific to different stages of development, particular tissues, or a disease state. There is a need to systematically collect data on alternatively spliced exons, introns and splice isoforms, and to annotate this data. The Alternative Splicing Database (ASD) consortium has been addressing this need, and is committed to maintaining and developing a value-added database of alternative splice events, and of experimentally verified regulatory mechanisms that mediate splice variants. In this poster we present two of the products from this project; namely a database of computationally delineated alternative splice events as seen in alignments of EST/cDNA sequences with genome sequences, and a database of alternatively spliced exons collected from literature. The reported splice events are from nine different organisms and are annotated for various biological features including expression states and cross-species conservation.

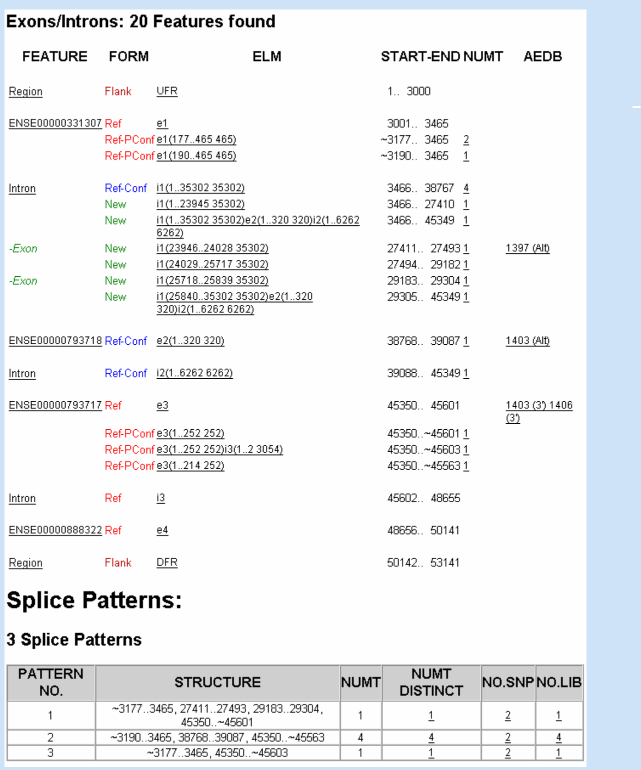
ALTSPLICE. AltSplice is a computer generated high quality data set of human transcript-confirmed splice patterns, alternative splice events, and the associated annotations.
"Known" genes from Ensembl and transcripts from EMBL are used as the start-up data for the computational pipeline that results in a large dataset of splice patterns, alternative splice events, and an extensive annotation on various levels.
The AltSplice database stores per gene with alternative events:

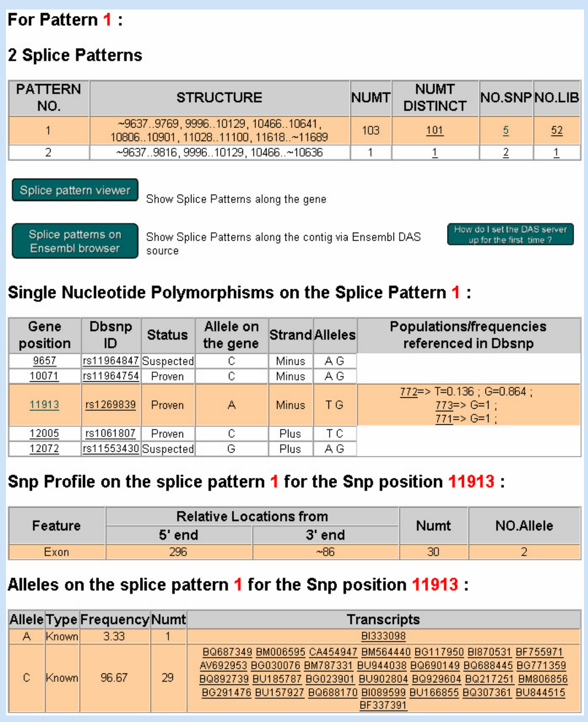
- generic gene information (Ensembl ID, location, cross-references, description, GO)
- gene structure (reference/confirmed/new features)
- splice patterns (transcript id, SNP mapping, expression states using standards vocabularies, preliminary protein changes)
- splicing events (various types, conservation human<->mouse)
- features (introns, exons and their characterisation)

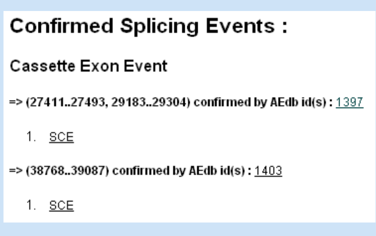
AltSplice detects the following splicing events: exon or intron isoform, cassette exon, intron retention, and mutual exclusive exon. For each indications are in place where exactly the change occurs (5′ or 3′, or both), and for the last three types simple or complex qualifiers are added as well. AltSplice[a)] brings earlier work with the AltExtron[b)] database to production standards and works on larger datasets.

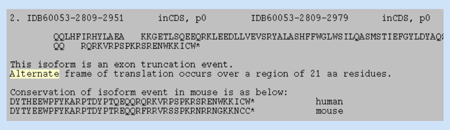
Contact for the AltSplice database is T.A. Thanaraj (EBI), e-mail thanaraj@ebi.ac.uk

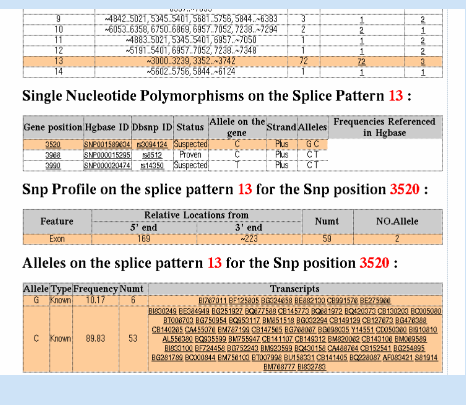
a) Thanaraj T.A., Stamm S., Clark F., Riethoven J.J.M, Le Texier V., and Muilu J. ASD: the Alternative Splicing Database. Nucl. Acids. Res. 32: D64-D69 (2004).
b) Clark, F. and T. A. Thanaraj. Categorisation and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. Human Molecular Genetics 11:451-464 (2002)











Screenshots from the AltSplice interface: features and splice patterns (above left), SNP mapping and information (right), splice events with integration (left) and preliminary protein changes (left-bottom).

INTEGRATION. The presented databases can be queried separately, but it is also possible to view the data in an integrated way. Two levels of integration exist: 1) a wrapper sends queries to both databases at the same time and displays the results together and 2) within the individual databases we have integration of information with the other database.
For example, if certain splice events are reported in literature this will be shown in AltSplice (see above), and vice versa if events are confirmed by the AltSplice pipeline this is reported in AEdb.

ACKNOWLEDGEMENTS. AltSplice and AEdb are part of the ASD project (www.ebi.ac.uk/asd/asd-ec/) which has the following partners: Erlangen University (Stamm), EBI (Thanaraj), ExonHit (Bracco), EMBL (Bork), CRG (Valcarcel, Guigo), and Hebrew University Israël (Soreq).
This work is supported by a grant (QLRT-CT-2001-02062) from the European Commission.
Andreas Kahari and Ewan Birney (EBI) are gratefully acknowledged for setting up a DAS server for the ASD data.

STATISTICS. AltSplice and AltExtron* database statistics

	Confirmed		Genes with at least one confirmed		
	Introns (a)	Exons (b)	Intron/exon (c)	Alternative event (d)	d/c
AltSplice Human	162552	124105	15298	8400	55%
AltSplice Mouse	122946	91802	13414	6221	46%
Drosophila*	33241	20595	7881	1302	16%
C.elegans*	41214	23043	9552	598	6%
Arabidopsis*	67455	51733	13302	758	6%
Chicken*	1115	811	228	42	
Cow*	594	367	177	20	
Rat*	1263	846	339	34	
Zebrafish*	1227	829	257	21	

Distribution of confirmed alternative events

	Intron retention	Exon isoforms	Intron isoforms	Cassette exons	Mutual exclusive exons
AltSplice Human	3677	5527	10090	12764	936
AltSplice Mouse	2210	3474	6862	6285	364
Drosophila*	111	494	1294	797	
C.elegans*	50	229	424	267	
Arabidopsis*	156	372	633	70	
Chicken*	4	26	37	27	
Cow*	0	12	20	25	
Rat*	1	15	26	22	
Zebrafish*	0	11	16	9	

AEdb database statistics: 1555 records in AEdb-Sequence (Cassette exon = 830; Alternating exon = 107; Alternate 3′ splice site = 136; Alternate 5′ splice site = 116; Intron retention = 101; Alternate poly-A = 48; unknown = 217) ; 171 records in AEdb-Function.

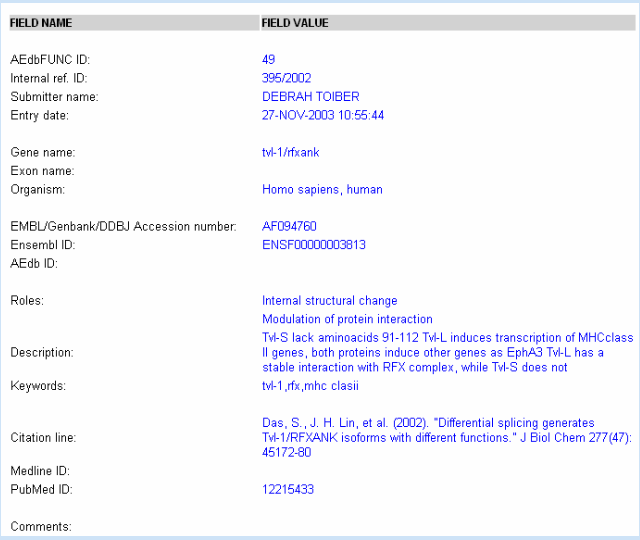
* AltExtron uses gene entries from the EMBL Nucleotide Database.

AEDB. AEdb[c)] is a manual generated database for human alternative exons and their properties - the data is gathered from literature where these exons have been experimentally verified. AEdb consists of two databases: one which focuses on the reported nucleotide sequence of the alternative event (AEdb-Sequence), and one which consists of the specific roles or functions that can be assigned to alternative events (AEdb-Function).
For AEdb-Sequence the following information is stored:

- administrative data (submitter, date, etc)
- bibliographical data (including PubMed)
- gene information (species, gene name, xrefs)
- splice event data (type of event, tissue, association, expression)
- splice event sequence data (sequence alternative exon, 5′ and 3′ constitutive exons)

Contact for the AEdb database is Stefan Stamm (Univ. Erlangen), e-mail stefan@stamms-lab.net

c) Stamm, S., J. Zhu, K. Nakai, P. Stoilov, O. Stoss, and M.Q. Zhang. An alternative exon database (AEDB) and its statistical analysis. DNA and Cell Biol., 19: 739-756 (2000).



FIELD NAME	FIELD VALUE
AEdbFUNC ID:	49
Internal ref. ID:	395/2002
Submitter name:	DEBRAH TOIBER
Entry date:	27-NOV-2003 10:55:44
Gene name:	tvl-1/rfxank
Exon name:	
Organism:	Homo sapiens, human
EMBL/Genbank/DDBJ Accession number:	AF094760
Ensembl ID:	ENSF00000003813
AEdb ID:	
Roles:	Internal structural change Modulation of protein interaction
Description:	Tvl-S lack aminoacids 91-112 Tvl-L induces transcription of MHCclass II genes, both proteins induce other genes as EphA3 Tvl-L has a stable interaction with RFX complex, while Tvl-S does not
Keywords:	tvl-1,rfx,mhc clasii
Citation line:	Das, S., J. H. Lin, et al. (2002). "Differential splicing generates Tvl-1/RFXANK isoforms with different functions." J Biol Chem 277(47): 45172-80
Medline ID:	
PubMed ID:	12215433
Comments:	

AEdb-sequence example data for an alternative exon

BIBLIOGRAPHY DATA
Publication: Das, S., J. H. Lin, et al. (2002). "Differential splicing generates Tvl-1/RFXANK isoforms with different functions." J Biol Chem 277(47): 45172-80.
Comments: Stimulus Interferon ? Effect: shift in splicing pattern toward theTvl-1 form (inclusion of exon 5) (these leads to MNC class II genes transcription induction.) Tvl-S lack aminoacids 91-112 Tvl-L induces transcription of MHCclass II genes, both proteins induce other genes as EphA3 Tvl-L has a stable interaction with RFX complex, while Tvl-S does not
PubMed ID: 12215433

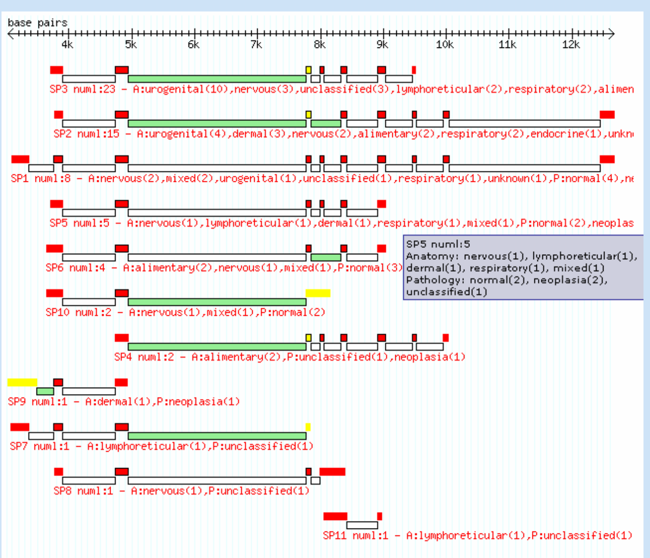
GENE INFORMATION
Species: HS, Homo sapiens, human
Gene name/exon number: Tvl-1/RFXANK, exon5
EMBL accession: AF094760
Ensembl accession number: ENSF00000003813
AEdb id is associated with AltSplice: ENSG00000064490
Type of Sequence: Genomic DNA

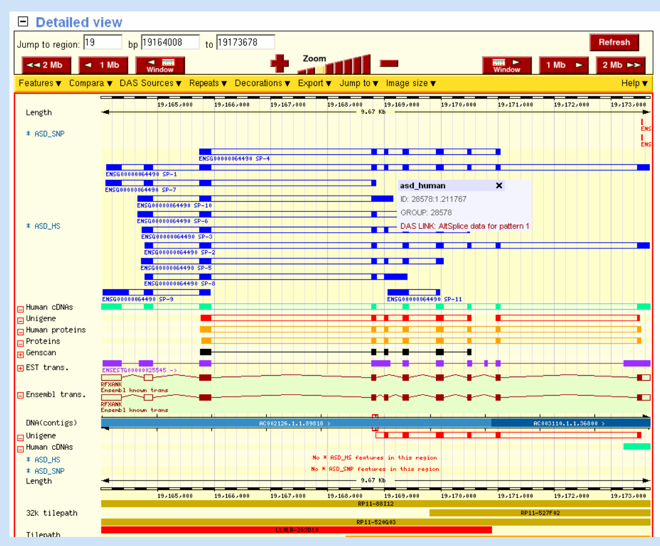
SPLICE EVENT DATA
Splicing Mechanism: CA, cassette
Technique used: PCR+cloning
Tissue in which expressed: Unknown
Regulated during development: Unknown
Association with disease: Unknown
Regulatory features: Unknown
Keywords: Interferon γ; Tvl-1,RFX, MHC clasII

SPLICE EVENT SEQUENCE DATA
For alternatively spliced exon:
12 nucleotides of the downstream intron: GTCCTGCCCCAG
Sequence of alternative exon: CCCTGTCCATCCACCAGCTCGCAGCACAGGGGGAGCTGGACCAGCTGAAGGAGCATTTGCGGAAAG
Reading frame: 3
6 nucleotides of the upstream intron: GTGCGT

For constitutive exons flanking alternative exon
Frame of 5' exon: 3
Sequence of 5' exon: GCAGCTCCCTGAAGCACTCCACCACTCTCACCAACCGGCAGCGAGGGAACGAGGTGTCAGCTCTGCCGGCCACCCTAGACT
Frame of 3' exon: 1
Sequence of 3' exon: GGTGCCGACCCCCACATCCTGGCAAAAGAGCGAGAGAGCGCCCTGTCGCTGGCCAGCACAGGCGGCTACACAGACATTGTGGGGCTGCTGCTGGAGCGTGACGTGGACATCAACATCTATGATTGG

An example of an AEdb-Function entry (above) and specific data that is collected for the AEdb-Sequence database (right).





Splice patterns from AltSplice can be viewed graphically on either the AltSplice Pattern Viewer (left) or as a DAS track on the Ensembl Genome Browser (right).

EBI Home Page www.ebi.ac.uk
ASD Pages www.ebi.ac.uk/asd
Email asd-ebi@ebi.ac.uk
FTP ftp.ebi.ac.uk
Telephone +44(0) 1223 494
Fax +44(0) 1223 494468