Databases for Gene Expression analysis

Perform a THOROUGH in silico analysis before doing benchwork

What is the difference between databases and papers?

For your pet gene:
determine all names using genecards
how many known exonic SNPs are there
find two commercial antisera for the protein
get the complete sequence for your pet gene using the UCSC database and compare the exon intron-organisation with the one predicted by exonscan

There are multiple databases for expression analysis

http://www.science.co.il/biomedical/RNA-Databases.asp

Abbreviation	Database
	5S ribosomal RNA database
ASTRA	Alternative Splicing and TRanscription Archives
AAARS	Aminoacyl-tRNA synthetases database
	Codon Usage Database
	Comparative RNA Web Site
Mamit-tRNA	Compilation of mammalian mitochondrial tRNA genes
	Compilation of tRNA sequences and sequences of tRNA genes
PolyA_DB	Database for mammalian mRNA polyadenylation
snoRNABase	Database of Human H/ACA and C/D Box snoRNAs
BPS	Database of RNA Base pair Structures
Recode	Database of translational recoding events
	DNA and RNA Molecular Weights and Conversions
	European Ribosomal RNA Database
FSDB	Frameshift Database
	Free Energy and Enthalpy Tables for RNA Folding
fRNAdb	Functional RNA Database
sRNAMap	Genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes
GtRDB	Genomic tRNA Database
	Greengenes: 16S rDNA gene database and tools
H-InvDB	H-Invitational database of human genes and transcripts
RNAdb	Mammalian noncoding RNA database
	Methylation Guide snoRNA Database
miRBase	microRNA Sequence Database
	Modomics: a database of RNA modification pathways
ncRNA	Noncoding RNA database
RNaseP	Ribonuclease P Database
RDP	Ribosomal Database Project of 16S rRNA sequences
	Ribosomal RNA Mutation Database
RAG	RNA As Graphs Web resource
Rfam	RNA families database
	RNA Modification Database
STRAND	RNA secondary STRucture and statistical ANalysis Database
	Silva rRNA database
	Small Subunit rRNA Modification Database
SCOR	Structural Classification of RNA
tmRDB	tmRNA Database
	tmRNA Website
	Viral RNA Structure Database
	Yeast Small Nucleolar RNA (snoRNA) Database

Five useful sites

Genecards, http://www.genecards.org/
Fast overview over a gene, link to reagents

UCSC genome browser, http://genome.ucsc.edu/
Most comprehensive (and best to use) database

Fast DB, http://www.fast-db.com/fastdb2/frame.html
Specialized in splicing, very easy to use,
Log in: Stefan@stamms-lab.net , PW: OUSMqllerX

Gil Ast lab page, http://www.tau.ac.il/~gilast/
Links to splicing analysis tools

Christopher Burge lab, http://genes.mit.edu/burgelab/software.html
Links to splicing analysis tools

Genecards

Provided free to academic non-profit institutions. ALL other users require a commercial license from Xennex, Inc.

TRA2B Gene
protein-coding GIFtS: 50
GC03M185632

transformer 2 beta homolog (Drosophila)
(Previous names: **splicing factor, arginine/serine-rich 10 (transformer 2 homolog, Drosophila)**)
Symbol approved by the HUGO Gene Nomenclature Committee (HGNC) database
(Previous symbol: **SFRS10**)

Aliases & Descriptions for TRA2B gene

(According to [1]HGNC, [2]Entrez Gene, [3]UniProtKB/Swiss-Prot, [4]UniProtKB/TrEMBL, [5]OMIM, [6]GeneLoc , and/or [7]Ensembl, [8]miRBase)
About This Section

Aliases & Descriptions

transformer 2 beta homolog (Drosophila)[1] [2]	DKFZp686F18120[2]
Htra2-beta[1] [2]	TRA2-BETA[2]
splicing factor, arginine/serine-rich 10 (transformer 2 homolog, Drosophila)[1] [2]	transformer-2 protein homolog beta[2]
Transformer-2 protein homolog B[2] [3]	transformer-2-beta[2]
TRA-2 beta[2] [3]	transformer 2 homolog[2]
SFRS10[2] [3]	Splicing factor, arginine/serine-rich 10[3]
SRFS10[2] [5]	TRA2-beta[3]
TRAN2B[2]	hTRA2-beta[3]

External Ids: HGNC: 10781[1] Entrez Gene: 6434[2] Ensembl: ENSG00000136527[7] UniProtKB: P62995[3]

Search outside databases for aliases for TRA2B gene

Previous GC identifer: GC03M187110

Summaries for TRA2B gene

(According to Entrez Gene, Tocris Bioscience, Wikipedia's Gene Wiki, UniProtKB/Swiss-Prot, and/or UniProtKB/TrEMBL)
About This Section

UniProtKB/Swiss-Prot: TRA2B_HUMAN, P62995
Function: Sequence-specific RNA-binding protein which participates in the control of pre-mRNA splicing

Gene Wiki entry for TRA2B (SFRS10)

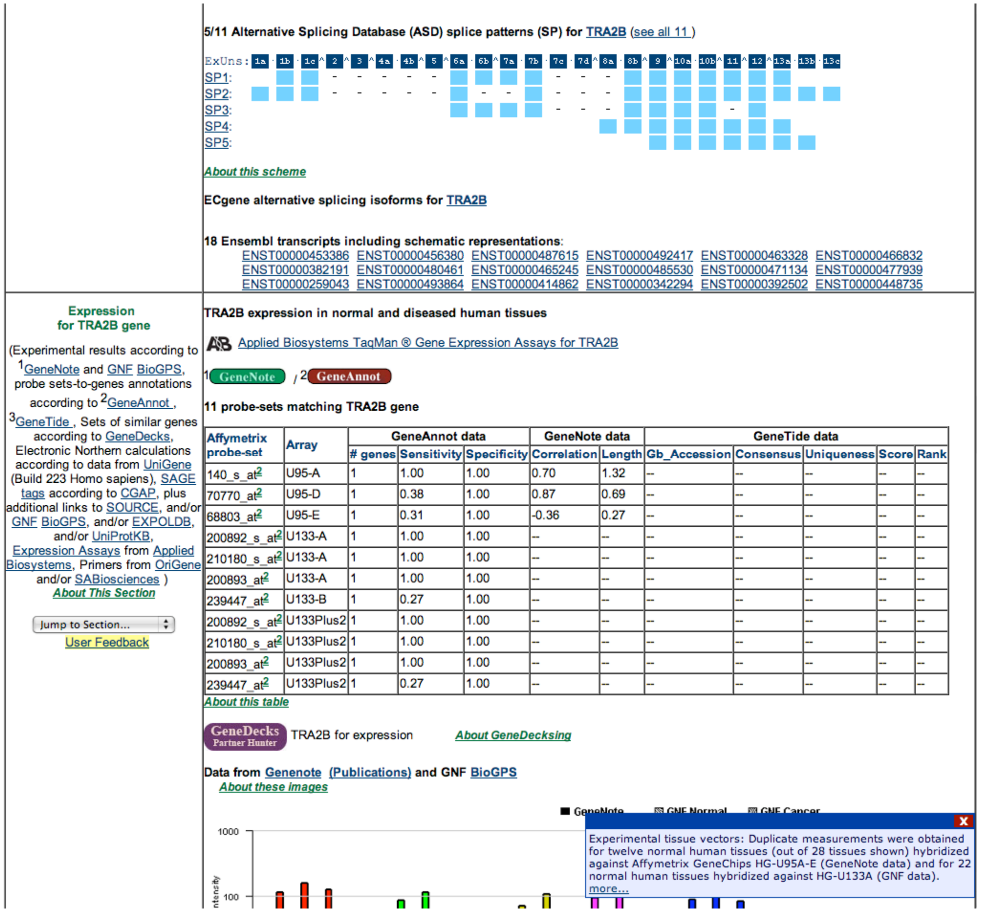
Genomic Views for TRA2B gene

(According to GeneLoc and/or HGNC, and/or Entrez Gene (NCBI build 37), and/or miRBase, Genomic Views according to UCSC and Ensembl (release 56), Regulatory elements and Epigenetics data according to

Regulatory elements:
Search SABiosciences Regulatory transcription factor binding sites for TRA2B

Epigenetics:
SABiosciences Methyl-Profiler DNA Methylation qPCR Primer Assays for TRA2B: MePH05827-2A

Genomic Location:
Genomic View: UCSC Golden Path with GeneCards custom track



5/11 Alternative Splicing Database (ASD) splice patterns (SP) for TRA2B (see all 11)

About this scheme

ECgene alternative splicing isoforms for TRA2B

18 Ensembl transcripts including schematic representations:
ENST00000453386 ENST00000456380 ENST00000487615 ENST00000492417 ENST00000463328 ENST00000466832
ENST00000382191 ENST00000480461 ENST00000465245 ENST00000485530 ENST00000471134 ENST00000477939
ENST00000259043 ENST00000493864 ENST00000414862 ENST00000342294 ENST00000392502 ENST00000448735

Expression for TRA2B gene

(Experimental results according to [1]GeneNote and GNF BioGPS, probe sets-to-genes annotations according to [2]GeneAnnot , [3]GeneTide , Sets of similar genes according to GeneDecks, Electronic Northern calculations according to data from UniGene (Build 223 Homo sapiens), SAGE tags according to CGAP, plus additional links to SOURCE, and/or GNF BioGPS, and/or EXPOLDB, and/or UniProtKB, Expression Assays from Applied Biosystems, Primers from OriGene and/or SABiosciences)
About This Section

TRA2B expression in normal and diseased human tissues

Applied Biosystems TaqMan ® Gene Expression Assays for TRA2B

[1] GeneNote / [2] GeneAnnot

11 probe-sets matching TRA2B gene

Affymetrix probe-set	Array	GeneAnnot data			GeneNote data		GeneTide data				
		# genes	Sensitivity	Specificity	Correlation	Length	Gb_Accession	Consensus	Uniqueness	Score	Rank
140_s_at[2]	U95-A	1	1.00	1.00	0.70	1.32	--	--	--	--	--
70770_at[2]	U95-D	1	0.38	1.00	0.87	0.69	--	--	--	--	--
68803_at[2]	U95-E	1	0.31	1.00	-0.36	0.27	--	--	--	--	--
200892_s_at[2]	U133-A	1	1.00	1.00	--	--	--	--	--	--	--
210180_s_at[2]	U133-A	1	1.00	1.00	--	--	--	--	--	--	--
200893_at[2]	U133-A	1	1.00	1.00	--	--	--	--	--	--	--
239447_at[2]	U133-B	1	0.27	1.00	--	--	--	--	--	--	--
200892_s_at[2]	U133Plus2	1	1.00	1.00	--	--	--	--	--	--	--
210180_s_at[2]	U133Plus2	1	1.00	1.00	--	--	--	--	--	--	--
200893_at[2]	U133Plus2	1	1.00	1.00	--	--	--	--	--	--	--
239447_at[2]	U133Plus2	1	0.27	1.00	--	--	--	--	--	--	--

About this table

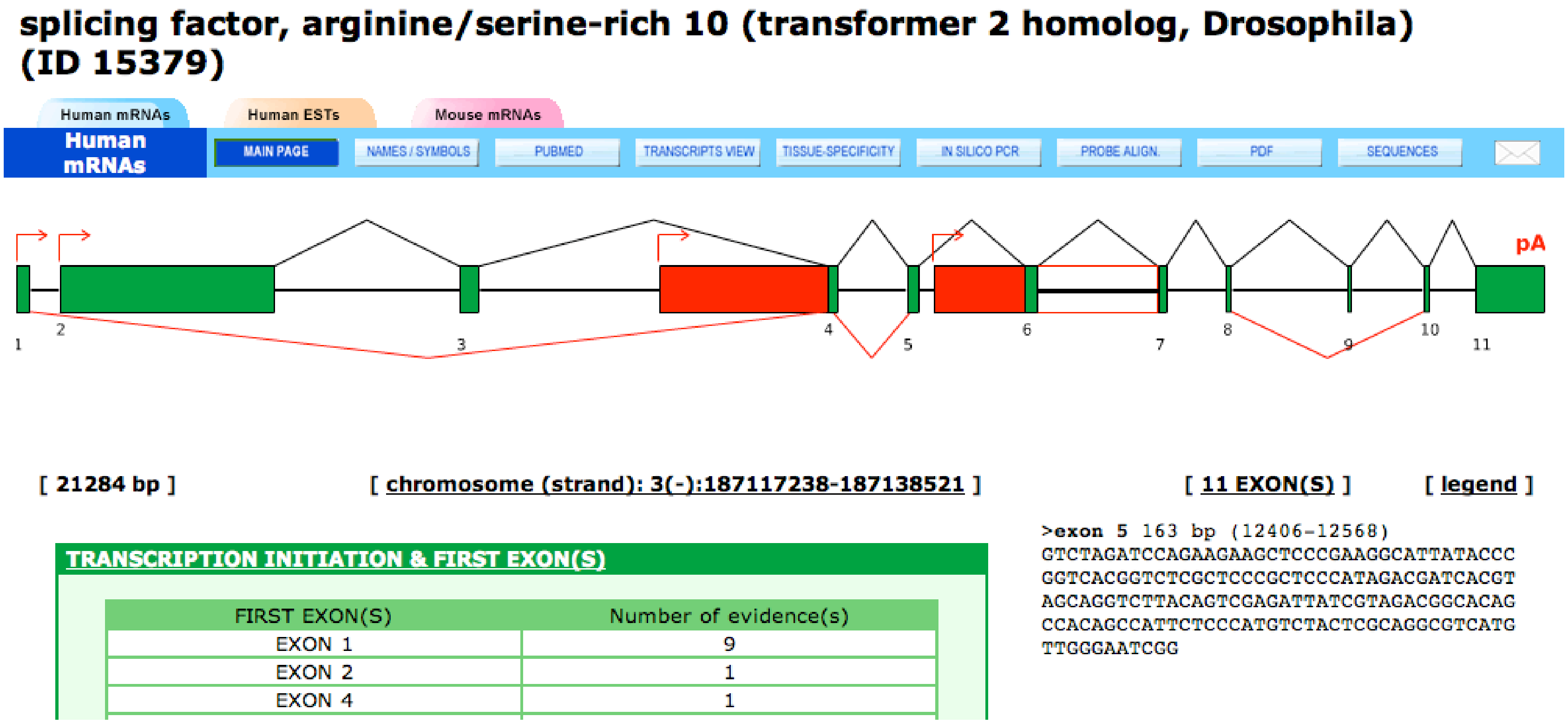
TRA2B for expression ***About GeneDecksing***

Data from Genenote (Publications) and GNF BioGPS
About these images

Experimental tissue vectors: Duplicate measurements were obtained for twelve normal human tissues (out of 28 tissues shown) hybridized against Affymetrix GeneChips HG-U95A-E (GeneNote data) and for 22 normal human tissues hybridized against HG-U133A (GNF data). more...

Fast, comprehensive overview for a given gene (expression, names, publications…)

FAST-DB



splicing factor, arginine/serine-rich 10 (transformer 2 homolog, Drosophila) (ID 15379)

[21284 bp] **[chromosome (strand): 3(-):187117238-187138521]** **[11 EXON(S)]** **[legend]**

TRANSCRIPTION INITIATION & FIRST EXON(S)

FIRST EXON(S)	Number of evidence(s)
EXON 1	9
EXON 2	1
EXON 4	1

```
>exon 5 163 bp (12406-12568)
GTCTAGATCCAGAAGAAGCTCCCGAAGGCATTATACCC
GGTCACGGTCTCGCTCCCGCTCCCATAGACGATCACGT
AGCAGGTCTTACAGTCGAGATTATCGTAGACGGCACAG
CCACAGCCATTCTCCCATGTCTACTCGCAGGCGTCATG
TTGGGAATCGG
```

http://www.fast-db.com

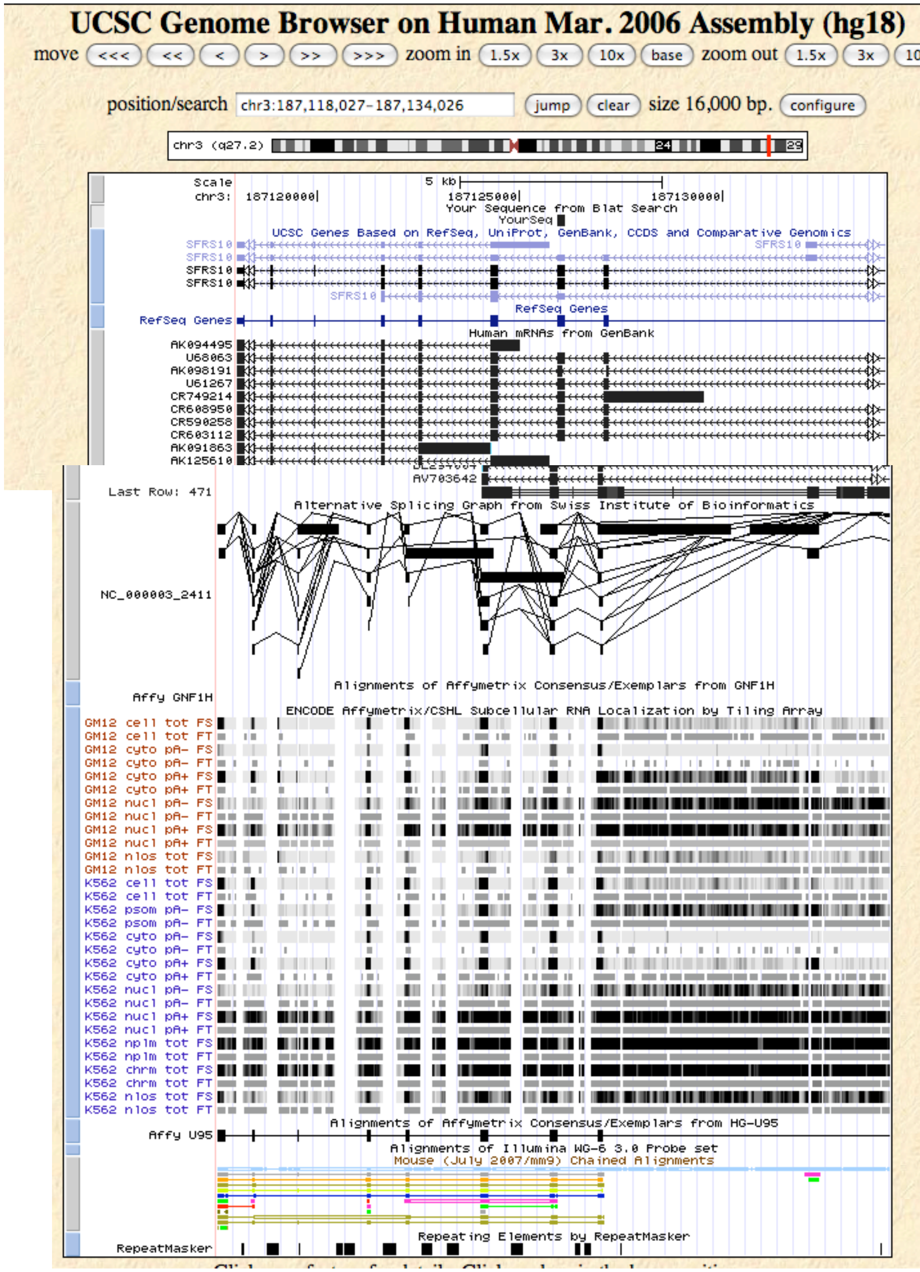
User-friendly

Stefan@stamms-lab.net

PW: OUSMqIlerX

UCSC genome browser

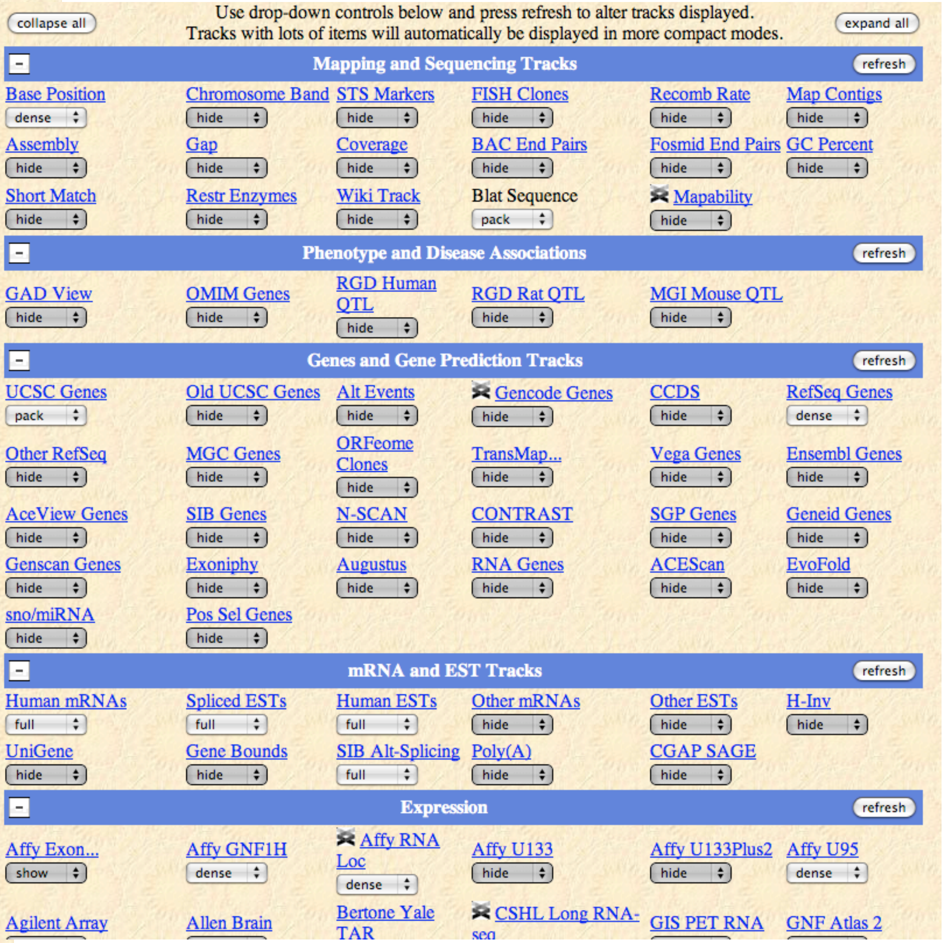
UCSC blat



UCSC Genome Browser on Human Mar. 2006 Assembly (hg18)

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chr3:187,118,027-187,134,026 jump clear size 16,000 bp. configure



Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.

Mapping and Sequencing Tracks refresh

Base Position dense	Chromosome Band hide	STS Markers hide	FISH Clones hide	Recomb Rate hide	Map Contigs hide
Assembly hide	Gap hide	Coverage hide	BAC End Pairs hide	Fosmid End Pairs hide	GC Percent hide
Short Match hide	Restr Enzymes hide	Wiki Track hide	Blat Sequence pack	Mapability hide	

Phenotype and Disease Associations refresh

GAD View hide	OMIM Genes hide	RGD Human QTL hide	RGD Rat QTL hide	MGI Mouse QTL hide

Genes and Gene Prediction Tracks refresh

UCSC Genes pack	Old UCSC Genes hide	Alt Events hide	Gencode Genes hide	CCDS hide	RefSeq Genes dense
Other RefSeq hide	MGC Genes hide	ORFeome Clones hide	TransMap... hide	Vega Genes hide	Ensembl Genes hide
AceView Genes hide	SIB Genes hide	N-SCAN hide	CONTRAST hide	SGP Genes hide	Geneid Genes hide
Genscan Genes hide	Exoniphy hide	Augustus hide	RNA Genes hide	ACEScan hide	EvoFold hide
sno/miRNA hide	Pos Sel Genes hide				

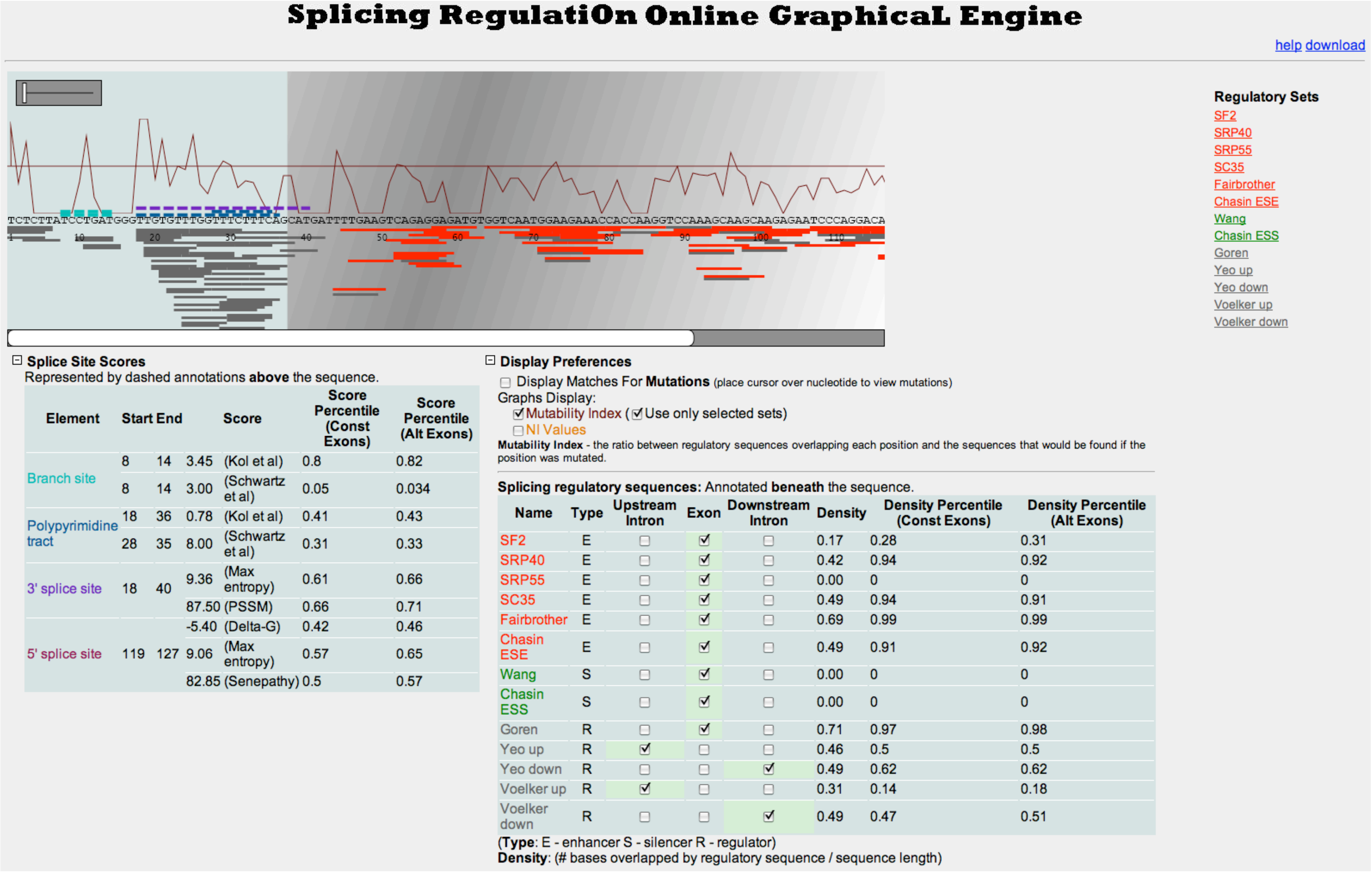
mRNA and EST Tracks refresh

Human mRNAs full	Spliced ESTs full	Human ESTs full	Other mRNAs hide	Other ESTs hide	H-Inv hide
UniGene hide	Gene Bounds hide	SIB Alt-Splicing full	Poly(A) hide	CGAP SAGE hide	

Expression refresh

Affy Exon... show	Affy GNF1H dense	Affy RNA Loc dense	Affy U133 hide	Affy U133Plus2 hide	Affy U95 dense
Agilent Array	Allen Brain	Bertone Yale TAR	CSHL Long RNA-seq	GIS PET RNA	GNF Atlas 2

Gil Ast homepage



⊟ Splice Site Scores

Represented by dashed annotations **above** the sequence.

Element	Start	End	Score	Score Percentile (Const Exons)	Score Percentile (Alt Exons)
Branch site	8	14	3.45 (Kol et al)	0.8	0.82
	8	14	3.00 (Schwartz et al)	0.05	0.034
Polypyrimidine tract	18	36	0.78 (Kol et al)	0.41	0.43
	28	35	8.00 (Schwartz et al)	0.31	0.33
3' splice site	18	40	9.36 (Max entropy)	0.61	0.66
			87.50 (PSSM)	0.66	0.71
			-5.40 (Delta-G)	0.42	0.46
5' splice site	119	127	9.06 (Max entropy)	0.57	0.65
			82.85 (Senepathy)	0.5	0.57

⊟ Display Preferences

☐ Display Matches For **Mutations** (place cursor over nucleotide to view mutations)

Graphs Display:

☑ Mutability Index (☑ Use only selected sets)

☐ NI Values

Mutability Index - the ratio between regulatory sequences overlapping each position and the sequences that would be found if the position was mutated.

Splicing regulatory sequences: Annotated **beneath** the sequence.

Name	Type	Upstream Intron	Exon	Downstream Intron	Density	Density Percentile (Const Exons)	Density Percentile (Alt Exons)
SF2	E	☐	☑	☐	0.17	0.28	0.31
SRP40	E	☐	☑	☐	0.42	0.94	0.92
SRP55	E	☐	☑	☐	0.00	0	0
SC35	E	☐	☑	☐	0.49	0.94	0.91
Fairbrother	E	☐	☑	☐	0.69	0.99	0.99
Chasin ESE	E	☐	☑	☐	0.49	0.91	0.92
Wang	S	☐	☑	☐	0.00	0	0
Chasin ESS	S	☐	☑	☐	0.00	0	0
Goren	R	☐	☑	☐	0.71	0.97	0.98
Yeo up	R	☑	☐	☐	0.46	0.5	0.5
Yeo down	R	☐	☐	☑	0.49	0.62	0.62
Voelker up	R	☑	☐	☐	0.31	0.14	0.18
Voelker down	R	☐	☐	☑	0.49	0.47	0.51

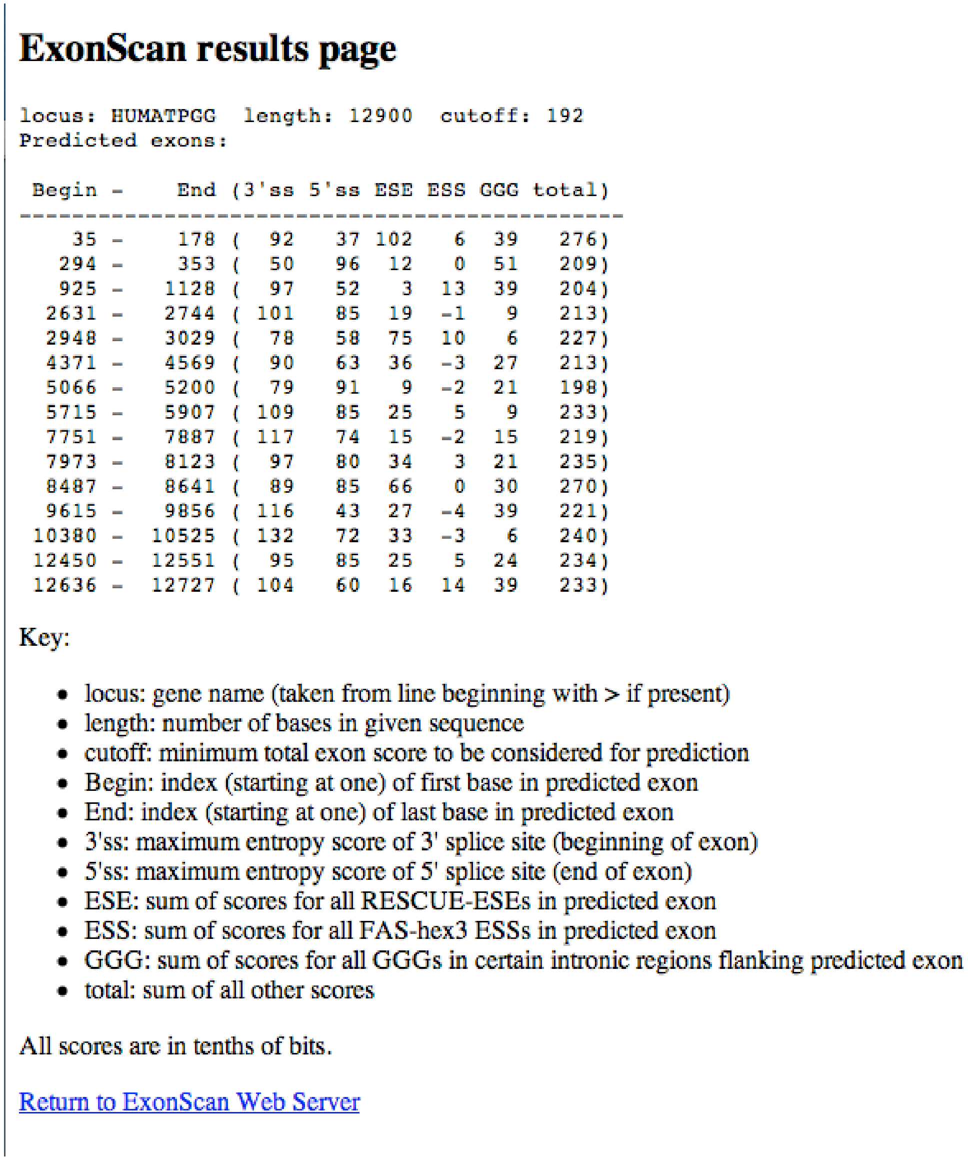
(**Type**: E - enhancer S - silencer R - regulator)

Density: (# bases overlapped by regulatory sequence / sequence length)

SCROOGLE

Christopher Burge lab

EXONSCAN



ExonScan results page

```
locus: HUMATPGG  length: 12900  cutoff: 192
Predicted exons:

 Begin -    End (3'ss 5'ss ESE ESS GGG total)
----------------------------------------------
    35 -    178 (  92   37 102   6  39   276)
   294 -    353 (  50   96  12   0  51   209)
   925 -   1128 (  97   52   3  13  39   204)
  2631 -   2744 ( 101   85  19  -1   9   213)
  2948 -   3029 (  78   58  75  10   6   227)
  4371 -   4569 (  90   63  36  -3  27   213)
  5066 -   5200 (  79   91   9  -2  21   198)
  5715 -   5907 ( 109   85  25   5   9   233)
  7751 -   7887 ( 117   74  15  -2  15   219)
  7973 -   8123 (  97   80  34   3  21   235)
  8487 -   8641 (  89   85  66   0  30   270)
  9615 -   9856 ( 116   43  27  -4  39   221)
 10380 -  10525 ( 132   72  33  -3   6   240)
 12450 -  12551 (  95   85  25   5  24   234)
 12636 -  12727 ( 104   60  16  14  39   233)
```

Key:

- locus: gene name (taken from line beginning with > if present)
- length: number of bases in given sequence
- cutoff: minimum total exon score to be considered for prediction
- Begin: index (starting at one) of first base in predicted exon
- End: index (starting at one) of last base in predicted exon
- 3'ss: maximum entropy score of 3' splice site (beginning of exon)
- 5'ss: maximum entropy score of 5' splice site (end of exon)
- ESE: sum of scores for all RESCUE-ESEs in predicted exon
- ESS: sum of scores for all FAS-hex3 ESSs in predicted exon
- GGG: sum of scores for all GGGs in certain intronic regions flanking predicted exon
- total: sum of all other scores

All scores are in tenths of bits.

Return to ExonScan Web Server

SUMMARY AND OUTLOOK

Lots of information available about gene expression that can be visualized using browsers

Problem with databases is their upkeep, UCSC and ENSEMBLE up to date for nucleic acids

Multiple analysis tools are available on the web

For your pet gene:
determine all names using genecards
how many known exonic SNPs are there
find two commercial antisera for the protein
get the complete sequence for your pet gene using the UCSC database and
compare the exon intron-organisation with the one predicted by exonscan