

Assessing Noncoding Functional Elements by Experimental and Computational Means

Current Topics in Genome Analysis

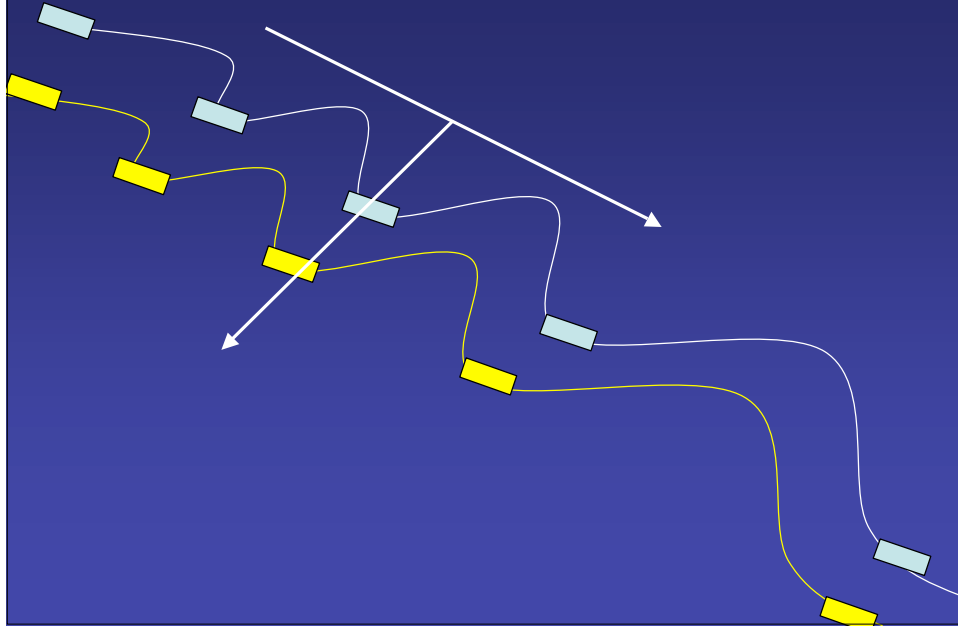
Laura Elnitski, Ph.D.
NHGRI/NIH

Comparative Sequence Analysis

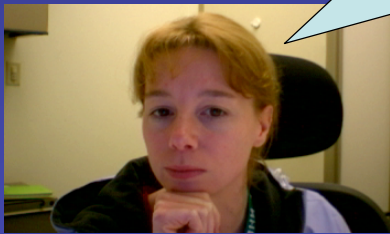
Dr. Elliott Margulies
Current Topics Lecture 3-1-05

- Tools for aligning genomic sequences
- Electing to use multiple species
- Conservation as an indicator of functional elements
- Sequence similarities reflect evolutionary relationships

Intra- and Interspecies Sequence Comparisons



How do I use genomic databases to find information about functional elements that might regulate expression of my gene of interest?



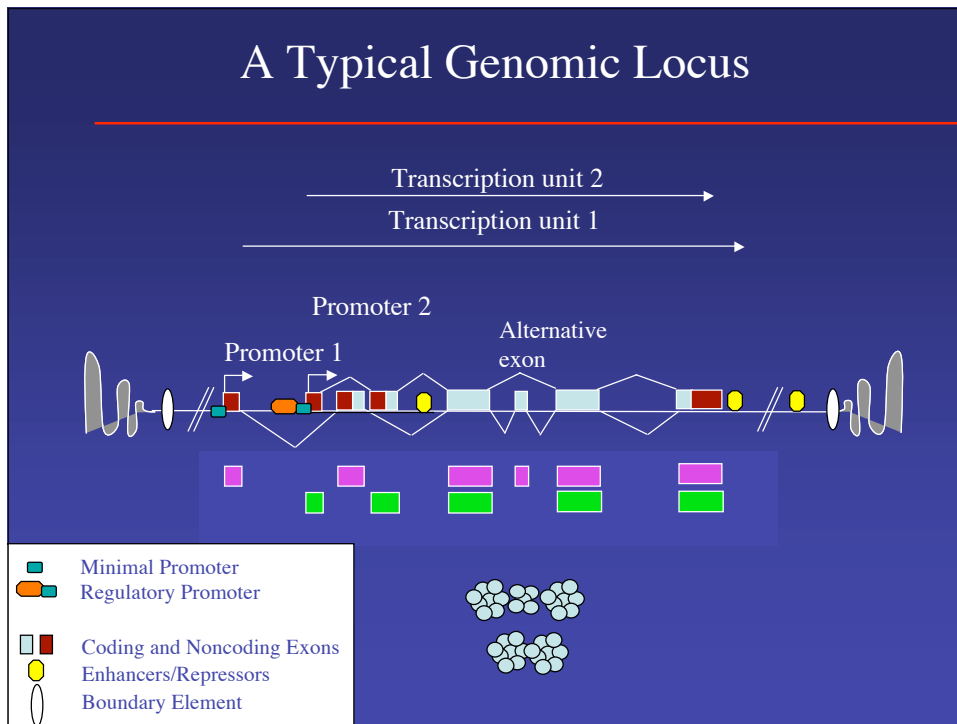
Outline

- Classes of Functional Elements
- Computational Analyses
- Experimental Analyses/Validation
- Preponderance of Evidence

Functional Elements

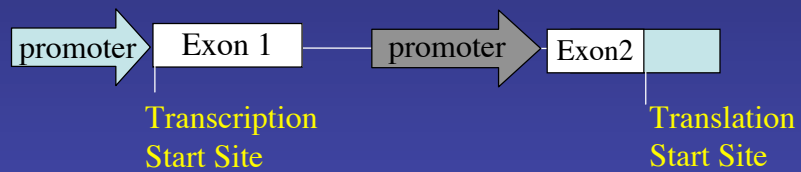
- **Intragenic Functional Elements**
 - Coding Exons
 - Untranslated Regions (UTRs)
 - Introns
- **Intergenic Functional Elements**
 - Boundary Elements
- **Regulators of Gene Expression**
 - Promoters
 - Enhancers/Repressors

A Typical Genomic Locus



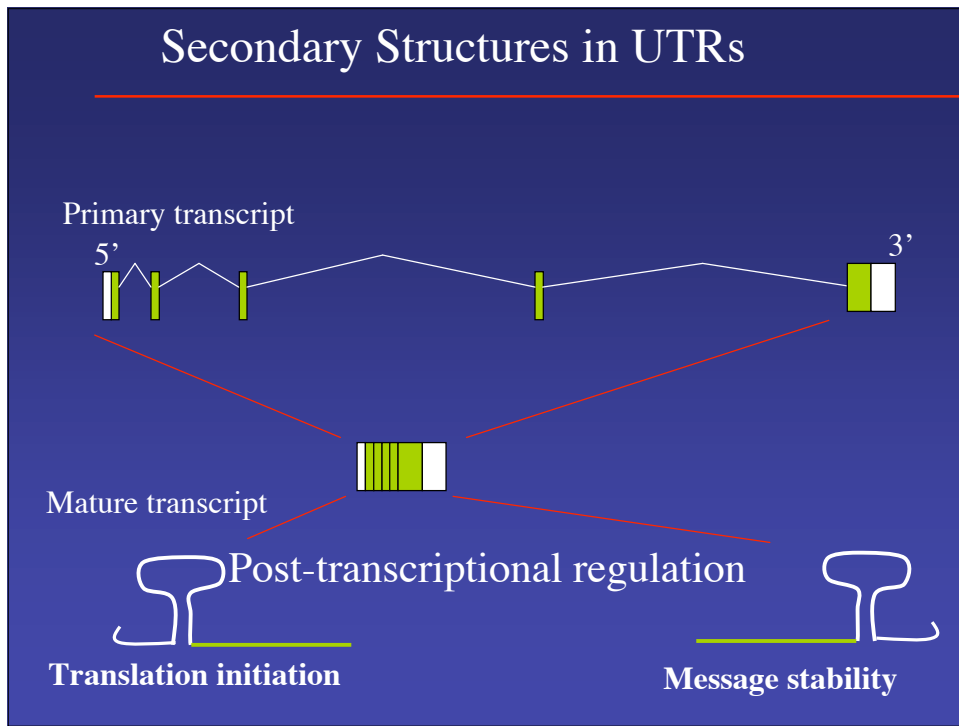
UTRs

- Often omitted from gene annotations
- 5' UTRs indicate transcription start site

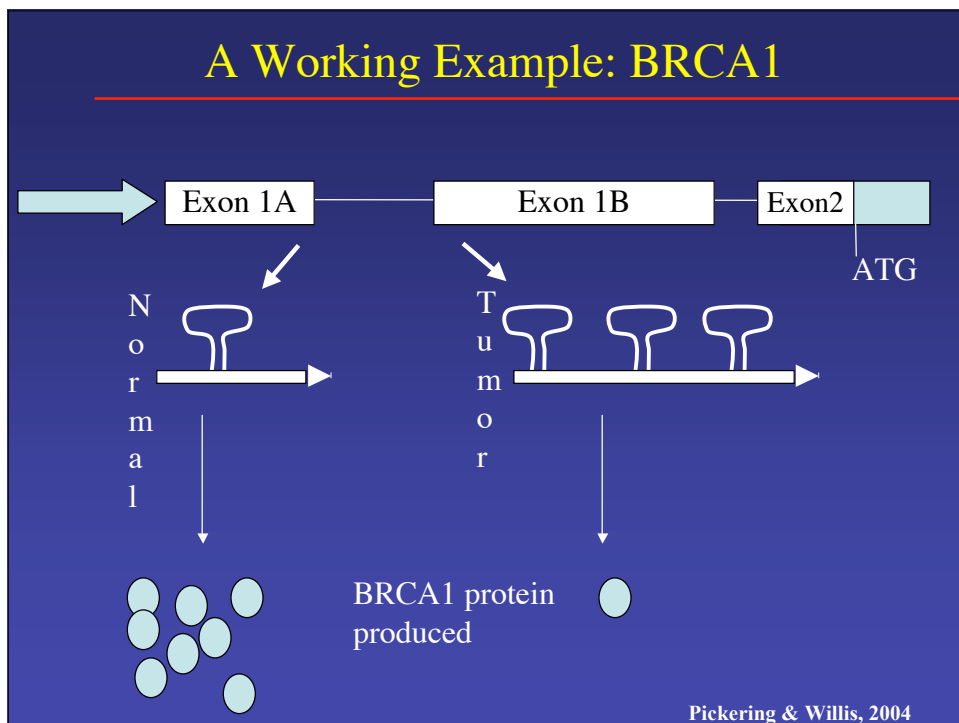


- Can be regulatory in nature

Secondary Structures in UTRs



A Working Example: BRCA1



Pickering & Willis, 2004

Genome Browser - Description Page

Home - Genomes - Genome Browser - Gene Sorter - Blat -

Human Gene BRCA1 Description and Page Index

Description: BRCA1 protein.
Alternate Gene Symbols: NM_007294
Representative mRNA: [BC072418](#) **Protein:** [Q6IN79](#)
RefSeq Summary: BRCA1, which functions as a tumor suppressor in hum phosphoprotein which associates with RNA polymerase II holoenzyme. Mu responsible for approximately 45% of inherited breast cancer and more than cancer. BRCA1 may function as a transcriptional regulator, due to an amino nuclear localization signals, and an acidic carboxy terminal domain. BRCA1 as a secreted growth inhibitory protein. BRCA1 may normally serve as a neg growth. This function is compromised in breast cancer either by direct mutat BRCA1 participates in transcription-coupled repair of oxidative DNA dama human chromosome 17, and consists of 24 exons, 22 of which are coding ex an unusually high density of Alu repetitive DNA (41.5%), but a relatively lo sequences. BRCA1 intron lengths ranged in size from 403 bp to 9.2 kb and located in introns 12, 19, and 20. Other genes have been localized close to B genes on the chromosome is: centomere-IFP35-VAT1-RHO7-BRCA1-M1 play a significant role in modulating the subcellular localization and physiolo

Page Index	Quick Links	Sequence	Microarray	RNA St
Other Species	GO Annotations	mRNA Descriptions	Pathways	Methods

Data Retrieval via Table Browser

UCSC Genome Browser Home

http://genome-test.cse.ucsc.edu/

UCSC Genome Bioinformatics

Genomes - Gene Sorter - Blat - PCR - **Tables** - Proteome - FAQ - Help

Gene Browser
Gene Sorter

About the UCSC Genome Bioinformatics Site

This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also shows the CFTR (cystic fibrosis) region in 13 species and provides a portal to the ENCODE project.

table: Describe Table Schema

Home Genomes Blast Gene Portal PTR FAQ Help

Retrieving UTRs by Location

Output knownGene as BED

Table Browser

Use this program to get the data associated with a track in text format, to calculate intersections between sequence covered by a track. See [Using the Table Browser](#) for a description of the controls in this form. [Page](#) is still available for a limited period.

clade: Vertebrate genome: Human assembly: July 2003

group: Genes and Gene Prediction Tracks track: Known Genes

table: knownGene describe table schema

region: 76-41662412 lookup

identifiers (names): all fields from selected table
chosen fields from selected and related tables
sequence

filter: edit

intersection: GTF - gene transfer format
BED - browser extensible data
query results to GALA

output format: custom track

output file: hyperlinks to Genome Browser (leave blank to keep output in browser)

file type returned: plain text gzip compressed

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream order to avoid extending past the edge of the chromosome.

get BED cancel

Regulatory Splice Junctions in Introns

AG

ELN:10

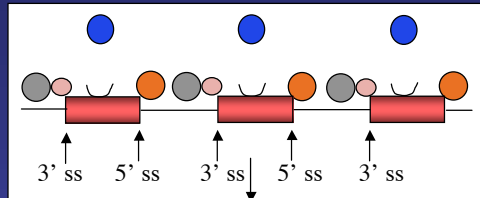
49247:	-----TCACTGCTTTGTCCCCGGCAG	AG	SAGCTCGGTTCCCCGGTGTGGGGG	human
1156719:	AGCTC.CTCC.....C...A--T	G.....	cat
1870778:	---A..TGT.....C.A.T---	T.....	mouse
1305970:	---C..TGT.....C.A.T---	T.....A.....	rat
187645:	..C..CT..A.CC.....TATC..C.....C..A..C		chicken

ELN:10

49293:	TGCTCCCTGGAGTTCCTCACTGGAGCAGGAGTTAAGCCCAAGGCTCCAGGT	GT	human
1156767:C..G.....C...A.....C...G.....A.C.....		cat
1870821:CA.....C..AG.....		mouse
1306013:CA.....C..G.....T..G.....		rat
187685:	...G..C..C.....C..CA.C..GA.C...G.G..A.G.....		chicken

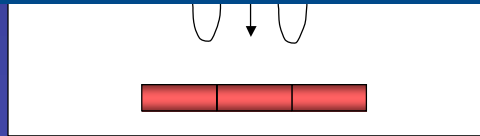
GT

Splicing Enhancers (Exonic)



RESCUE-ESE Web Server

An online tool for identifying candidate ESEs in vertebrate exons



<http://genes.mit.edu/burgelab/rescue-ese/>

RESCUE-ESE Server

Choice of Vertebrate ESEs

Check if **YES**, default is **NO**

- Human (Homo sapiens)
- Mouse (Mus musculus)
- Zebrafish (Danio rerio)
- Pufferfish (Fugu rubripes)

Find ESEs Reset

Enter Your Sequence (RNA or DNA) as **plain text** or **multi-FASTA** format
(maximum 4k bases of sequence data)

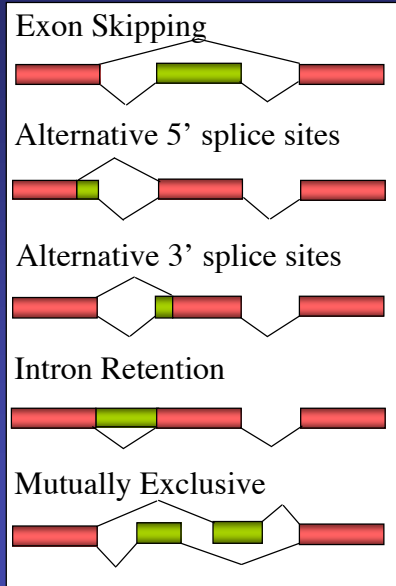
```
ACCAAGGTCAGAACATTACCGAAAGACAACAGCATCCACACGAAAAGTG
TCACTGGCCCTCAGGCAAACCTTGACTGAAGTGGATATATTTCAAGAAG
GTTATCTCAAGAACTGGCTTGAAATAAGTGAAGAAATTAACGAAGAAG
ACTTAAAGGTAGGTATACATCGCTTG
```

G
AG
ATTAG
..|..
280 290 300 310 320 330 340 350 360

GAAT
GAA
GA
G

GAATT
.....

Alternative splicing events



Conserved across species:

38%

18%

8%

3%

N/A

Nature Reviews Genetics
5; 773-782 (2004)

UCSC Alt-splicing Data

Base Position 116700000 | 116750000 | 116800000 | 116850000

table browser query on knowGene
Known Genes Based on SWISS-PROT, TrEMBL, mRNA, and RefSeq

ABCC7

Alternate Gene Symbols: ABCC7, CF, MRP7, NM_000492
 Representative mRNA: [M28668](#) Protein: [P13569](#) (aka CFTR_HUMAN)

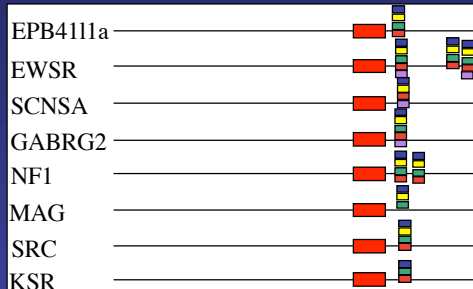
RefSeq Summary: The protein encoded by this gene is a member of the superfamily of ATP-binding cassette (ABC) membranes. ABC genes are divided into seven distinct subfamilies (ABC1, MDR/TAP, MRP, ALD, OABP, GCN2 resistance). This protein functions as a chloride channel and controls the regulation of other transport pathways. Mutations in this gene cause cystic fibrosis (CF) and congenital bilateral aplasia of the vas deferens (CBAVD). Alternative splice variants have been described.

Page Index	Quick Links	UniProt Comments	Sequence	Microarray	Alt-splicing
	Protein Structure	Other Species	GO Annotations	mRNA Descriptions	Pathways
Methods					

Alternative Splicing

<http://genome-test.cse.ucsc.edu>

Splicing Enhancers (Intronic)



UGCAUG

- Associated with alternatively spliced exons that have brain specific expression pattern.
- Show conservation across mammalian species

Minovitsky et al. NAR 2005

Cis-Acting Regulatory Elements

Defined with respect to location and orientation from the transcription start site

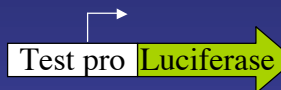
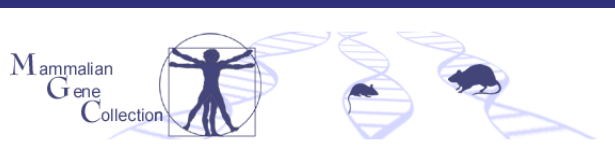
- **Promoters**
 - Proximal to transcription start site
 - Determine timing of gene expression during development
 - Recruit basal transcription machinery
- **Enhancers/(Repressors)**
 - Act in a distance and orientation independent manner
 - Augment the level of expression or choice of tissue specificity

• Core Promoters

Elements located proximal to transcription start sites that determine the timing of gene expression during development and act to recruit the basal transcription machinery.

High Throughput Promoter Analysis

Identification and functional analysis of human transcriptional promoters
Trinklein et al. (2003) Genome Research 13(2):308-312.



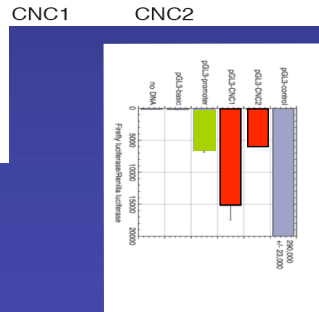
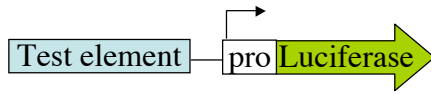
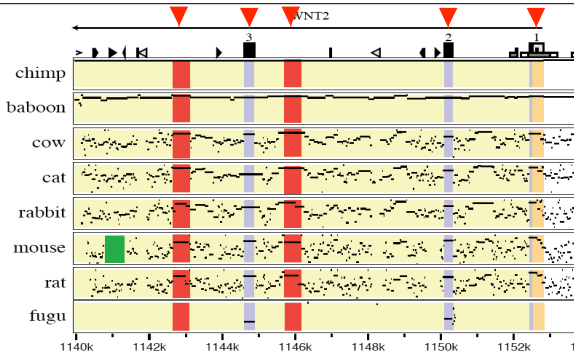
- Upstream regions from the Mammalian Gene Collection.
- 90% of tested predictions are positive for function
- Over 40,000 mapped candidate promoters in the human genome.

Available at genome-test.cse.ucsc.edu

April 2003

• Enhancers

- Identify candidate enhancers through multiple sequence alignment



Transfection Assay

Annotations of Putative Regulatory Regions

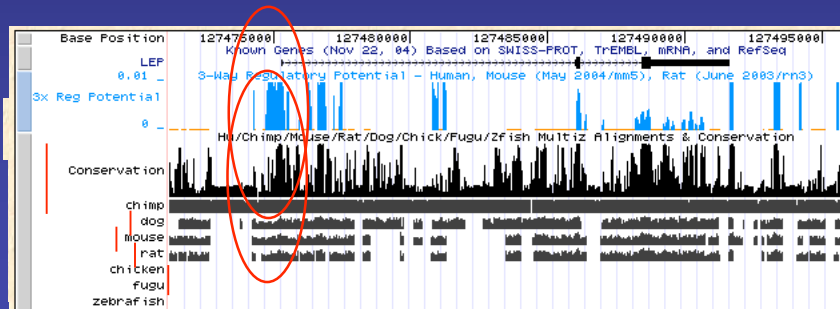
Expression and Regulation				
GNF Atlas 2 <input type="button" value="hide"/>	UCLA Tissues <input type="button" value="hide"/>	GNF Ratio <input type="button" value="hide"/>	Affy U133 <input type="button" value="hide"/>	Affy GNF1H <input type="button" value="hide"/>
Affy U133Plus2 <input type="button" value="hide"/>	Affy U95 <input type="button" value="hide"/>	CpG Islands <input type="button" value="hide"/>	FirstEF <input type="button" value="hide"/>	3x Reg Potential <input type="button" value="full"/>
2x Reg Potential <input type="button" value="hide"/>	TFBS Conserved <input type="button" value="hide"/>	Transfrags <input type="button" value="hide"/>		
Comparative Genomics				
10-Way Conservation <input type="button" value="full"/>	10-Way Most Conserved <input type="button" value="hide"/>	Conservation (Std.) <input type="button" value="full"/>	Most Cons. (Std) <input type="button" value="hide"/>	phastCons HCE <input type="button" value="hide"/>

Regulatory Predictions

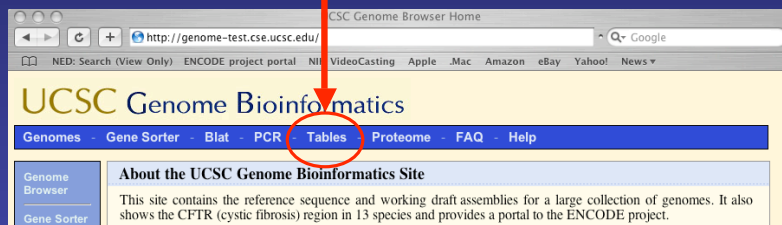
Conservation

PhyloHMM (PhastCons)

Regulatory Potential Scores



Data Retrieval via Table Browser



<http://genome-test.cse.ucsc.edu/>

Home - Genomes - Gene Sorter - Blat - PCR - Tables - FAQ - Help

Table Browser

Use this program to get the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. See [Using the Table Browser](#) for a description of the command line form. The [old Table Browser Page](#) is still available for a limited period.

clade: Vertebrate genome: Human assembly: May 2004

group: Expression and Regulation track: 3x Reg Potential

table: regPotential3X Describe Table Schema

region: genome ENCODE position chr4:56214201-56291736

filter: Create

intersection: Create

output format: data points

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

Note: output is limited to 100,000 lines returned. Use the filter setting to change this limit.

Get Output Summary/Statistics

Apple .Mac Amazon eBay Yahoo! News

Home - Genomes - Blat - Gene Sorter - PCR - FAQ - Help

Table Browser

Use this program to get the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. See [Using the Table Browser](#) for a description of the command line form. The [old Table Browser Page](#) is still available for a limited period.

clade: Vertebrate genome: Human assembly: May 2004

group: Comparative Genomics track: 10-Way Conservation

table: phastCons10way describe table schema

region: genome ENCODE position chr4:56214201-56291736

filter: create

intersection: create

output format: data points

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

Binding Site Prediction



Zlab Home Page Boston University BU Bioinformatics

Gene Regulation Tools Protein Engineering Tools

People Publications Lab Schedule Resources Links

Zlab

Gene Regulation Tools

<http://zlab.bu.edu/zlab/gene.shtml>

- Regulatory Networks
- Clusters of binding sites
- High-throughput approaches
- Precomputed, Genome-Wide Analyses (rVista/zPicture, Genome Browser, Table-Browser/GALA)

Visualizing Binding Sites

on Alignments
as Genome Annotations



<http://www.dcode.org> - Comparative Genomics Center at Lawrence Livermore National Laboratory

rVista 2.0

<http://rvista.dcode.org>

Expression and Regulation				
GNF Atlas 2 <input type="button" value="hide"/>	UCLA Tissues <input type="button" value="hide"/>	GNF Ratio <input type="button" value="hide"/>	Affy U133 <input type="button" value="hide"/>	Affy GNF1H <input type="button" value="hide"/>
Affy U133Plus2 <input type="button" value="hide"/>	Affy U95 <input type="button" value="hide"/>	CpG Islands <input type="button" value="hide"/>	FirstEF <input type="button" value="hide"/>	3x Reg Potential <input type="button" value="hide"/>
2x Reg Potential <input type="button" value="hide"/>	TFBS Conserved <input type="button" value="full"/>	Transfrags <input type="button" value="hide"/>		

Genome Browser Conserved Binding Sites

Expression and Regulation

GNF Atlas 2 hide	UCLA Tissues hide	GNF Ratio hide	Affy U133 hide	Affy GNF1H hide
Affy U133Plus2 hide	Affy U95 hide	CpG Islands hide	FirstEF hide	3x Reg Potential full
2x Reg Potential hide	TFBS Conserved hide	Transfrags hide		

LEP HMR Conserved Transcription Factor Binding Sites

Filtering the dataset

filter:

Filter on Fields from hg16.tfbsCons

bin	is	ignored		
chrom	does	match	*	AND
chromStart	is	ignored		AND
chromEnd	is	ignored		AND
name	does	match	*	AND
score	is	ignored		AND
strand	does	match	*	AND
species	does	match	human	AND
factor	does	match	MEF-2	AND
id	does	match	*	AND

Free-form query:

Filtering on Binding Sites

Home - Genomes - Gene Sorter - Blat - PCR - Tables - FAQ - Help

Table Browser

Use this program to get the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. See [Using the Table Browser](#) for a description of the controls in this form. The [old Table Browser Page](#) is still available for a limited period.

clad: Vertebrate genome: Human assembly: July 2003
 group: Expression and Regulation track: TFBS Conserved
 table: tfbsCons [Describe Table Schema](#)
 region: genome ENCODE position chr7:115365024-117242449 [Lookup](#)
 identifiers (names/accessions): [Paste List](#) [Upload List](#)
 filter: [Edit](#) [Clear](#)
 intersection: [Create](#)
 output format: all fields from selected table
 output file: (leave blank to keep output in browser)
 file type returned: plain text gzip compressed
[Get Output](#) [Summary/Statistics](#)

Table Browser - Table Schema

Sample Rows

bin	chrom	chromStart	chromEnd	name	score	strand	species	factor	id
585	chr1	1679	1686	VSSRY_01	860	-	human	SRY	Q05066
585	chr1	1679	1686	VSSRY_01	860	-	mouse	SRY	Q05738
585	chr1	2423	2433	V\$MYOD_Q6	850	-	N	N	N
585	chr1	3779	3787	V\$MZF1_01	890	-	human	MZF-1	P28698
585	chr1	3949	3956	V\$SNKX25_01	1000	-	mouse	Nkx-2.5	P42582
585	chr1	3949	3956	V\$SNKX25_01	1000	-	mouse	Nkx-2.5	P97335
585	chr1	4020	4042	V\$MEF2_02	850	-	human	MEF-2	N
585	chr1	4020	4042	V\$MEF2_02	850	-	mouse	MEF-2	N
585	chr1	4020	4042	V\$MEF2_03	870	-	human	MEF-2	N
585	chr1	4020	4042	V\$MEF2_03	870	-	mouse	MEF-2	N

- Descriptions of all information relating to the track
- Sample data

GALA: Genome Alignment and Annotation database on Human July 2003 Freeze

Powered by DB2

[Menu](#)

[Query form](#)

[History page](#)

Expression and Regulation

Transcription factor binding sites

The transcription factor binding sites were produced with [tfind](#), [tfloc](#), and [TRANSEAC](#) (free registration required)

Query all TF binding sites (interval is required)

only binding sites conserved in hg16Gg2 cutoff used was 0.85

only binding sites conserved in hg16Mm3Rn3 cutoff used was 0.85

only binding sites conserved in hg16Pt1Mm3Rn3Gg2 cutoff used was 0.75

To select/add factor names [Click here](#) To select/add factor IDs [Click here](#) Strand Score (range between 0.75 and 1) to

www.bx.psu.edu

Experimental Approaches

- Inference

Altered chromatin structure

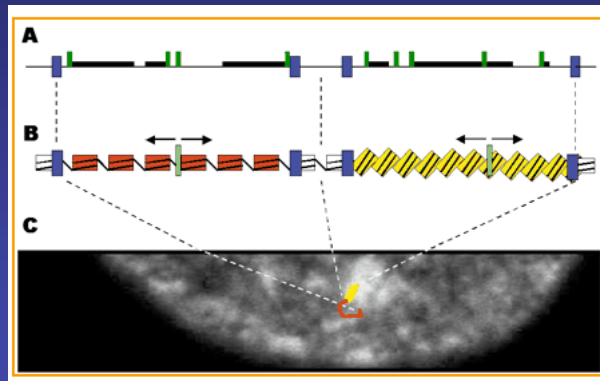
- Verification

Protein binding sites

- Prediction

High throughput assays

Levels of Gene Regulation



Linear Sequence

Chromatin Conformation

Nuclear Location

Roel van Driel Journal of Cell Science: 4067-4075 (2003)

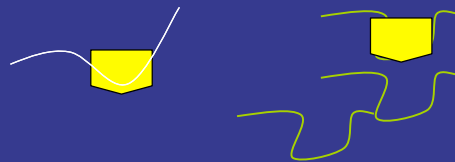
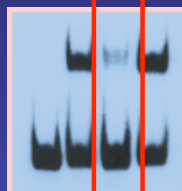
Detecting Protein Binding

Gel Shift / EMSA (Electrophoretic Mobility Shift Assay)

Purified Proteins

Nuclear Extracts

Free probe

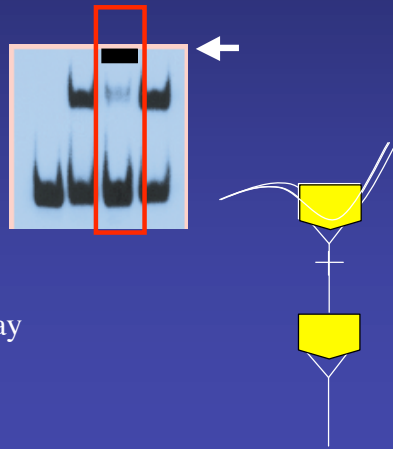


Nuclear Extracts

- Competition Assay
- Supershift assay

Detecting Protein Binding

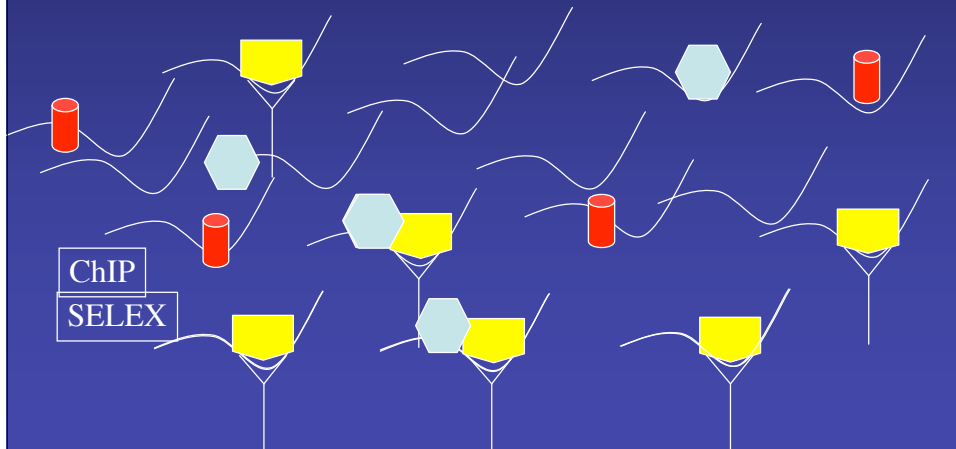
Gel Shift / EMSA (Electrophoretic Mobility Shift Assay)



- Supershift assay

Identifying Preferred Motifs

- Chromatin immunoprecipitation “ChIP”
- SELEX (Casting)



PWM

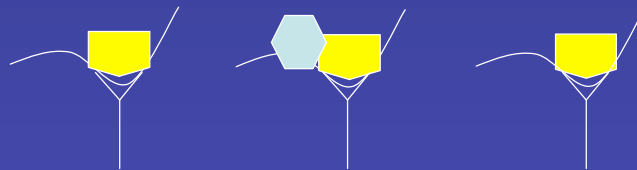


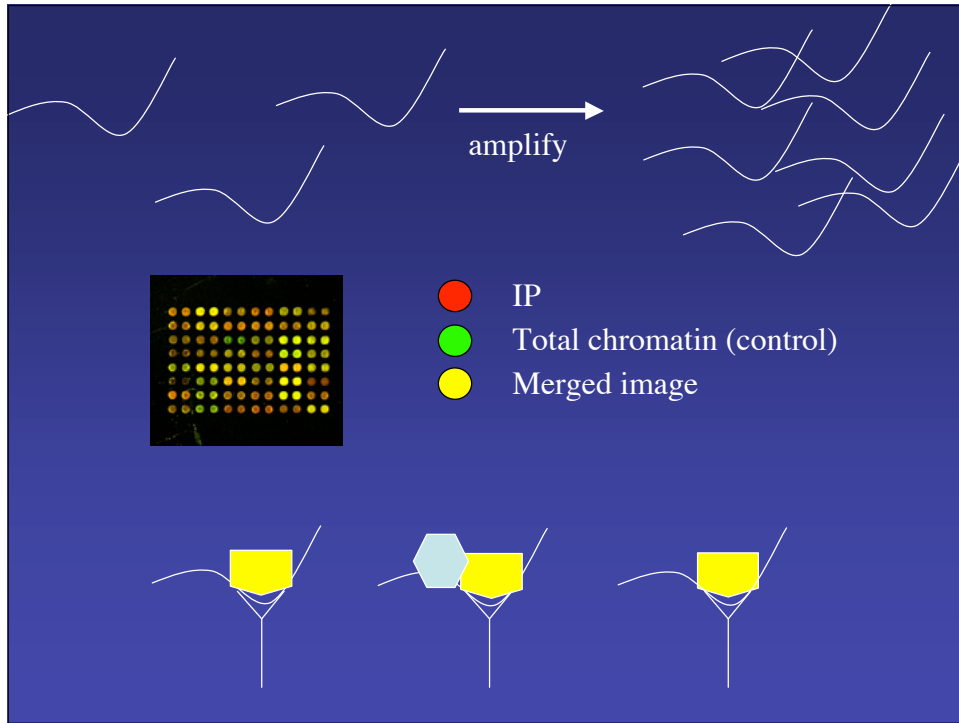
Gene Regulation <http://www.gene-regulation.com/>

Position	A	C	G	T	Base
01	6	0	16	3	G
02	2	0	24	0	G
03	1	0	0	25	T
04	2	0	0	24	T
05	25	0	0	1	A
06	20	2	1	3	A
07	2	0	0	24	T
08	8	3	8	7	N
09	16	0	1	9	W
10	3	1	1	21	T
11	1	4	0	21	T
12	16	1	6	3	A
13	9	11	3	3	M
14	7	14	0	5	M
15	11	7	4	3	N
XX					
BA	26 binding sites from 20 genes				

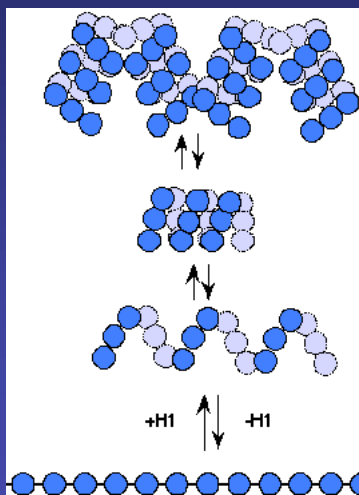
High Throughput Binding Assays

- **ChIP-chip**
chromatin immunoprecipitation using DNA microarray chips
 - Detects Protein-DNA Interactions (Direct)
 - Protein-Protein Interactions (Indirect)





Chromatin Compaction



solenoid

- Transcriptionally inactive
- Nuclease insensitive

~ 30 nm fiber

- Decreasing levels of compaction

Beads on a string

10 nm fiber, with exposed linker region

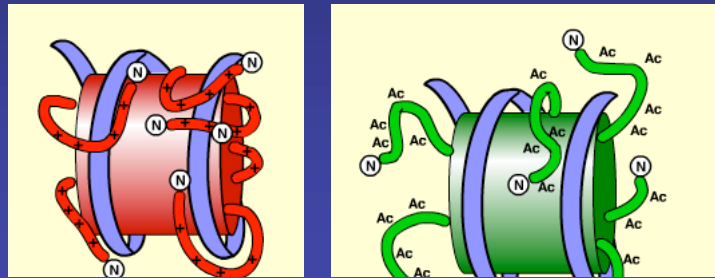
Detectable Histone Modifications

Gross rearrangements : movement or removal

Detected by DNase I sensitivity

Subtle alterations: methylation, acetylation, phosphorylation

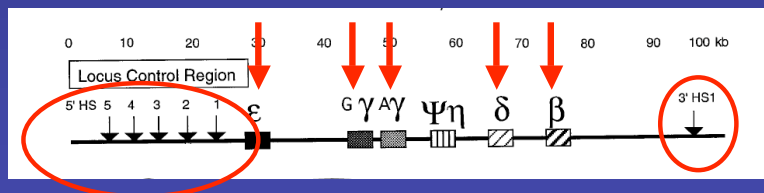
Detected by ChIP assays



<http://www-schreiber.chem.harvard.edu/home/animation.html>

Chromatin and Gene Expression

- Changes in chromatin conformation can be measured by accessibility of DNA to cleavage by DNase I
- DNase I *hypersensitivity* is an indicator of relaxed chromatin
- Often seen prior to detectable gene expression



DNase I Assays

- Targeted Loci

In vitro and *in vivo*

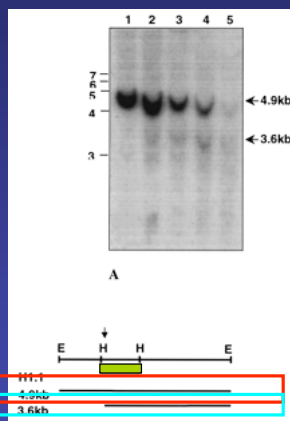
Indirect end-labeling/ Southern Blot

Quantitative PCR

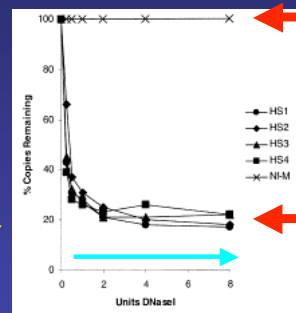
- High-Throughput Analyses

Cloning endpoints of enzymatic cleavage

DNase I Assays

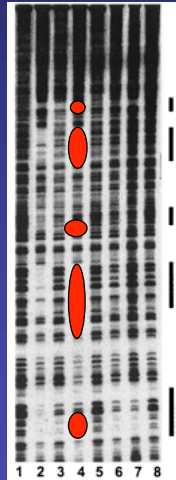


Indirect end-labeling/
Southern Blot



Quantitative PCR

DNase I Footprinting

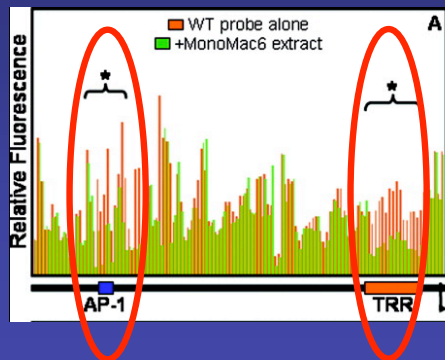


Identifies regions inaccessible to the enzyme that cleaves the DNA

“DNase Protection”

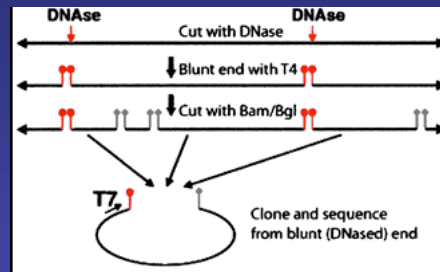
Advancing Technologies

Automated sequencing machines for hypersensitivity assays



Base pair resolution

High Throughput DNase I Hypersensitivity



Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites.

Crawford et al. (2004) Proc Natl Acad Sci U S A. 101:992-7.

Characteristics Implicating Function

- Location in genome
- Conservation in a multiple sequence alignment
- Predictive tracks
- Collections of Protein binding sites




ENCODE Project at UCSC

Regions - Data Submission - Downloads - Tools - Terms - Help

The ENCODE project- Amassing high-throughput data indicative of functional elements

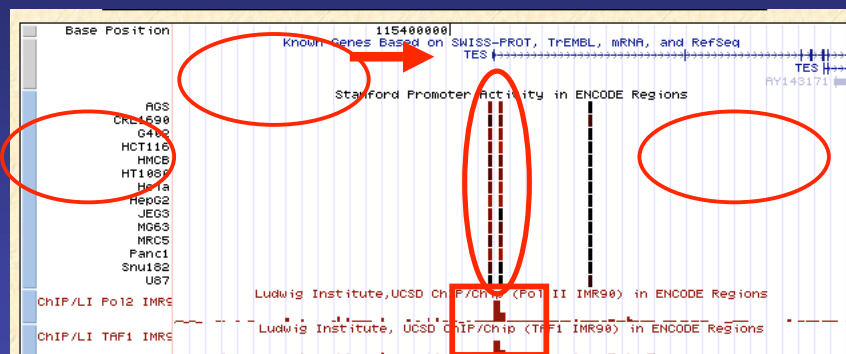
The ENCODE Project: ENCyclopedia Of DNA Elements

- ① [Overview](#)
- ① [Consortium Membership](#)
- ① [Data Release Policy](#)
- ① [Accessing ENCODE Data](#)
- ① [Common Consortium Resources](#)
- ① [Target Selection Process and Target Regions](#)
- ① [Comparative Sequence Analysis](#)
- ① [Meeting Reports](#)
- ① [Request for Application \(RFA\)](#)
- ① [Press Releases and Publications](#)
- ① [Program Staff](#)



<http://www.genome.gov/10005107>

Features that identify promoters



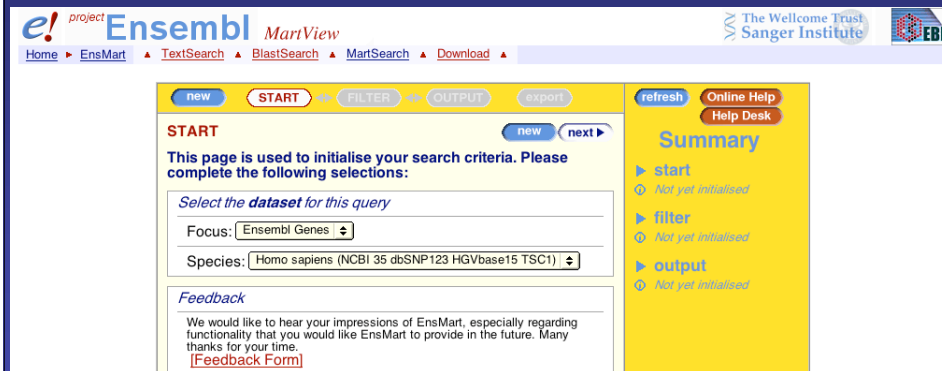
DNase I hypersensitivity ?

Integrative Data Management

Servers providing tools to:

- Compare and contrast annotation tracks
- Include or exclude features on command
- Progressively refine the search criteria

EnsMart Genome Data Mining Tool



The screenshot displays the EnsMart MartView web interface. At the top, there is a navigation bar with the 'e! project Ensembl MartView' logo on the left and logos for 'The Wellcome Trust Sanger Institute' and 'EBI' on the right. Below the navigation bar, a breadcrumb trail reads: Home > EnsMart > TextSearch > BlastSearch > MartSearch > Download. The main content area is divided into several sections. On the left, there is a 'START' section with a 'new' button and a 'next' button. Below this, a message states: 'This page is used to initialise your search criteria. Please complete the following selections:'. Underneath, there are two dropdown menus: 'Focus: Ensembl Genes' and 'Species: Homo sapiens (NCBI 35 dbSNP123 HGVbase15 TSC1)'. Below the species dropdown is a 'Feedback' section with a message: 'We would like to hear your impressions of EnsMart, especially regarding functionality that you would like EnsMart to provide in the future. Many thanks for your time.' and a link to a 'Feedback Form'. On the right side of the interface, there is a yellow sidebar titled 'Summary' with three expandable sections: 'start', 'filter', and 'output'. Each section has a right-pointing arrow and a status indicator that says 'not yet initialised'. At the top of the sidebar, there are buttons for 'refresh', 'Online Help', and 'Help Desk'. At the bottom of the screenshot, the URL 'http://www.ensembl.org/Multi/martview' is displayed.

<http://www.ensembl.org/Multi/martview>

Compound Queries with Table Browser

Home - Genomes - Blat - Gene Sorter - PCR - FAQ - Help

Table Browser

Use this program to get the data associated with a track in text format, to calculate intersected tracks, and to retrieve DNA sequence covered by a track. See [Using the Table Browser](#) for the controls in this form. The [old Table Browser Page](#) is still available for a limited period.

clade: Vertebrate ▾ **genome:** Human ▾ **assembly:** May 2004 ▾

group: Genes and Gene Prediction Tracks ▾ **track:** Known Genes ▾

table: knownGene ▾ describe table schema

region: genome ENCODE position chr7:127471196-127495720 lookup

identifiers (names/accessions): paste list upload list

filter: create

intersection: create

Intersect with Known Genes

Select a group and track to intersect with:

group: Expression and Regulation ▾ **track:** CpG Islands ▾

These combinations will maintain the gene/alignment structure (if any) of Known Genes.

- All Known Genes records that have any overlap with CpG Islands
- All Known Genes records that have no overlap with CpG Islands
- All Known Genes records that have at least 80 % overlap with CpG Islands
- All Known Genes records that have at most 80 % overlap with CpG Islands

These combinations will discard the gene/alignment structure (if any) of Known Genes position ranges.

- Base-pair-wise intersection (AND) of Known Genes and CpG Islands
- Base-pair-wise union (OR) of Known Genes and CpG Islands

Check the following boxes to complement one or both tables. To complement a table in the intersection if it is *not* included in the table.

- Complement Known Genes before intersection/union
- Complement CpG Islands before intersection/union

Submit Cancel

Additional Query/Analysis Tools

GALA Database

clade: genome: assembly:

group: track:

table: [Describe Table Schema](#)

region: genome ENCODE position [Lookup](#)

identifiers (names/accessions): [Paste List](#) [Upload List](#)

filter: [Create](#)

intersection: [Create](#)

output format: all fields from selected table

output file: [chosen fields from selected and related tables](#) (browser)

file type returned: sequence GTF - gene transfer format BED - browser extensible data query results to GALA custom track hyperlinks to Genome Browser

[Get Output](#) [Summary](#)

To reset all user queries, click [here](#).

GALA: Genome Alignment and Annotation database on Freeze

GALA History Page

Powered by DB2

[Menu](#)

[Query form](#)

GALA Common queries

03 Freeze

Powered These can be used for viewing or compound queries. For the display button these queries u

- A: all genes (default set, UCSC Known Genes) Found 18238 range(s)
- B: all CpG islands Found 257361 range(s)
- C: all SNPs Found 4880901 range(s)
- D: alignments human vs. mouse, min 100bps, 70%identity Found 585026 range(s)
- E: Union of exons from all gene models Found 452256 range(s)
- F: nonrepetitive DNA aligned with both mouse and rat Found 2134502 range(s)

Previous user queries

These queries will stay in the history for 14 days from last use.

- 1: table browser query on ChIP/LI Pol2 HeLa status: ready Found 24339 range(s)
- 2: table browser query on Promoter/Stanford status: ready Found 642 range(s)

[DELETE selected user queries](#)

[EDIT a query description](#)

[EDIT a previous user query](#) including changing the output format

GALA Compound Queries

Operations that can be performed on selected queries

On 1 of the above queries

- NOT
- Restrict region size: greater than or equal to less than or equal to
- Get aligning coordinates from pairwise alignments for chicken mouse rat
- Get orthologous genes and view in corresponding GALA
- Get orthologous regions using net alignments and liftOver

On 2 or more of the above queries

for help with INTERSECTION [click here](#)

- UNION
- INTERSECTION
- INTERSECTION and trim regions

On 2 of the above queries

- INTERSECTION returning ranges from the earlier query that overlap anything in the later query
- INTERSECTION returning ranges from the later query that overlap anything in the earlier query

- SUBTRACTION earlier minus later query
- SUBTRACTION later minus earlier query

Remove:

Proximity

Return regions from query number that

- lie within bp
- lie more than bp

(for genes only upstream downstream both) from a region in query number

Clusters

Return regions from query that have at least regions from query within +/- bps of a region in first query

Table Browser Query

Request:

Identify promoters that are regulated by muscle-specific factors

- Of all functional ENCODE promoters (Stanford Promoters Track)
- *How many correspond to conserved regions in mammals?*
- *How many have a conserved MEF-2 site?*

Table Browser - GALA Query

Request:

Identify promoters that are regulated by muscle-specific factors

- Of all functional ENCODE promoters (Stanford Promoters Track)
- *How many correspond to conserved regions in mammals?*
- *How many have a conserved MEF-2 site?*

- *Can we identify clusters of binding sites in active promoters?*

GALA Query

Rationale: Identify clusters of MEF-2 and MYOD binding sites in promoters analyzed by the Stanford Group

Expression and Regulation

Transcription factor binding sites

The transcription factor binding sites were produced with [tffind](#), [tfloc](#), and [TRANSFAC](#) (free registration required)

Query all TF binding sites (interval is required)

only binding sites conserved in hg16Gg2 cutoff used was 0.85

only binding sites conserved in hg16Mm3Rn3 cutoff used was 0.85

only binding sites conserved in hg16Pt1Mm3Rn3Gg2 cutoff used was 0.75

Previous user queries

These queries will stay in the history for 14 days from last use.

1: table browser query on [ChIP/ELISA/HeLa](#) status: ready Found 24339 range(s)

2: table browser query on Promoter/Stanford status: ready Found 642 range(s)

3: (only binding sites conserved in hg16Mm3Rn3) AND (factor ID = VSMEF2_01, VSMEF2_02, VSMEF2_03) status: ready Found 2284 range(s)

4: (only binding sites conserved in hg16Mm3Rn3) AND (factor ID = VSMYOD_01, VSMYOD_Q6) status: ready Found 158646 range(s)

VSMEF2_01
 VSMEF2_02
 VSMEF2_03
 VSMEF2_04

VSMYOD_01
 VSMYOD_Q6

