The UCSC Genome Browser

Shamith Samarajiwa
Integrative Systems Biomedicine Group
MRC Cancer Unit
University of Cambridge
ss861@mrc-cu.cam.ac.uk

17th May 2016,
CRUK Cambridge Institute

UNIVERSITY OF CAMBRIDGE

MRC Cancer Unit

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Genomic Coordinate Systems

- There are two major coordinate systems in genomics.

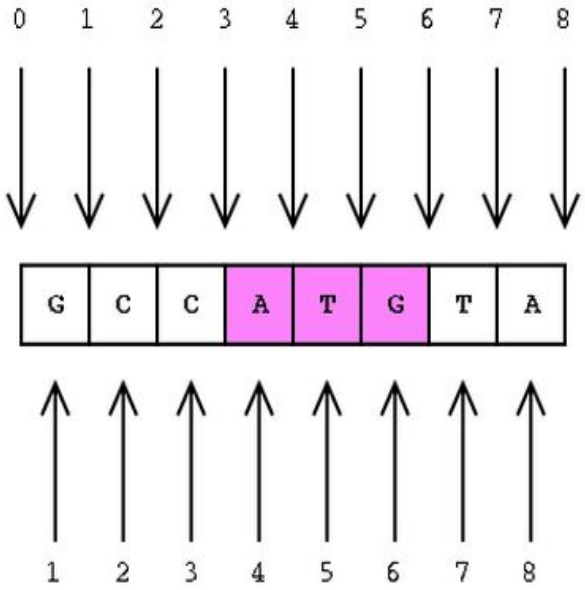
- **Base coordinate system** anchors genomic feature to nucleotide positions while the **Interbase coordinate system** anchor genomic feature between nucleotide positions.

- Most genome annotation portals (e.g. **NCBI or Ensembl**), bioinformatics software (e.g. BLAST) and annotation file formats (e.g. **SAM, VCF, GFF and Wiggle**) use the base coordinate system, which represents a feature starting at the first nucleotide as **position 1**.

- Other systems (e.g. **UCSC, Chado, DAS2**) and formats **(BAM, BCFv2, BED, and PSL)** use the interbase coordinate system, whereby a feature starting at the first nucleotide is represented as **position 0**.

Genomic Coordinate Systems

- The UCSC genome browser uses both systems and refer to the base coordinate system as "**one-based, fully-closed**" (used in the UCSC genome browser display) and interbase coordinate system as "**zero-based, half-open**" (used in their tools and file formats).
- The interbase coordinate system is also referred to as "space-based" by some authors.



There are several advantage for using the interbase coordinate system including:

1. the ability to represent features that occur between nucleotides (like a splice site),
2. simpler arithmetic for computing the length of features (length=end-start) and overlaps (max(start1,start2), min(end1, end2))
3. more rational conversion of coordinates from the positive to the negative strand

Ref: http://bergmanlab.ls.manchester.ac.uk/?p=36

UCSC genome browser: Introduction

main sections:

1. UCSC Genome Browser
2. BLAT
3. Custom tracks, Sessions and Track Hubs
4. Table Browser
5. Other UCSC tools

- what does it do?
- How do I use it?
- What problems does it help me solve?

UCSC Genome Bioinformatics

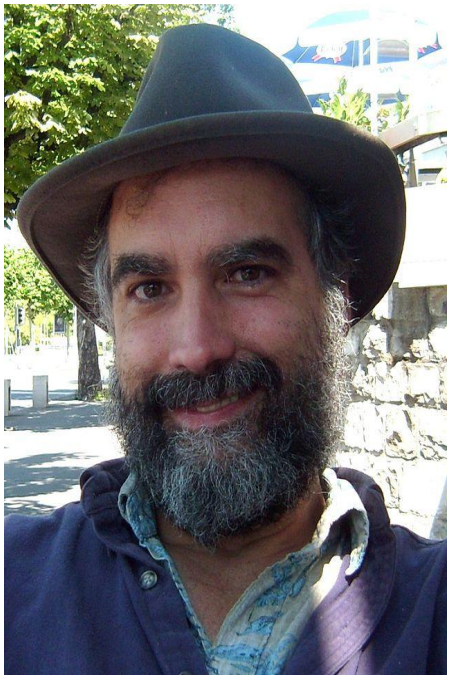
GENOME RESEARCH
CSHL Press | Journal Home | Subscriptions | eTOC Alerts | BioSupplyNet

Genome Res. 2001 Sep; 11(9): 1541–1548. PMCID: PMC311095
doi: 10.1101/gr.183201

Assembly of the Working Draft of the Human Genome with GigAssembler

W. James Kent[1,3] and David Haussler[2]

David Haussler



Jim Kent

About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to ENCODE data at UCSC (2003 to 2012) and to the Neandertal project. Download or purchase the Genome Browser source code, or the Genome Browser in a Box (GBiB) at our online store.

We encourage you to explore these sequences with our tools. The Genome Browser zooms and scrolls over chromosomes, showing the work of annotators worldwide. The Gene Sorter shows expression, homology and other information on groups of genes that can be related in many ways. Blat quickly maps your sequence to the genome. The Table Browser provides convenient access to the underlying database. VisiGene lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. Genome Graphs allows you to upload and display genome-wide data sets.

The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the UC Santa Cruz Genomics Institute at the University of California Santa Cruz (UCSC). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our public mailing list.

The Genome Browser project team relies on public funding to support our work. Donations are welcome -- we have many more ideas than our funding supports! If *you* have ideas, drop a comment in our suggestion box.

DONATE NOW

News

News Archives ▶

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the genome-announce mailing list. Please see our blog for posts about Genome Browser tools, features, projects and more.

---

29 June 2015 - GENCODE Genes Now the Default Gene Set on the Human (GRCh38/hg38) Assembly

In a move towards standardizing on a common gene set within the bioinformatics community, UCSC has made the decision to adopt the GENCODE set of gene models as our default gene set on the human genome assembly. Today we have released the GENCODE v22 comprehensive gene set as our default gene set on human genome assembly GRCh38 (hg38), replacing the previous default UCSC Genes set generated by UCSC. To facilitate this transition, the new gene set employs the same familiar UCSC Genes schema, using nearly all the same table names and fields that have appeared in earlier versions of the UCSC set.

By default, the browser displays only the transcripts tagged as "basic" by the GENCODE Consortium. These may be found in the track labeled "GENCODE Basic" in the Genes and Gene Predictions track group. However, all the transcripts in the GENCODE comprehensive set are present in the tables, and may be viewed by adjusting the track configuration settings for the All GENCODE super-track. The most recent version of the UCSC-generated genes can still be accessed in the track "Old UCSC Genes".

The new release has 195,178 total transcripts, compared with 104,178 in the previous version. The total number of canonical genes has increased from 48,424 to 49,534. Comparing the new gene set with the previous version:

- 9,459 transcripts did not change.
- 22,088 transcripts were not carried forward to the new version.
- 43,681 transcripts are "compatible" with those in the previous set, meaning that the two transcripts show consistent splicing. In most cases, the old and new transcripts differ in the lengths of their UTRs.
- 28,950 transcripts overlap with those in the previous set, but do not show consistent splicing (i.e., they contain overlapping introns with differing splice sites)

More details about the new GENCODE Basic track can be found on the GENCODE Basic track description page.

---

26 June 2015 - New Bonobo (panPan1) Assembly Now Available in the Genome Browser

We are pleased to announce the release of a Genome Browser for the May 2012 assembly of bonobo, *Pan paniscus* (Max-Planck Institute panpan1, UCSC version panPan1). The assembly was provided by the Max-Planck Institute for Evolutionary Anthropology. There are 10,867 scaffolds with a total size of 2,869,190,071 bases.

Bulk downloads of the sequence and annotation data are available via the Genome Browser FTP server or the Downloads page. These data have specific conditions for use. The bonobo (panPan1) browser annotation tracks were generated by UCSC and collaborators worldwide. See the Credits page for a detailed list of the organizations and individuals who contributed to this release.

---

12 June 2015 - Data Integrator: Have you ever wished that the Table Browser could associate your custom track items with some other track, while retaining the item names from both? We have released a new tool that can do just that, and more: the Data Integrator. Read more.

28 May 2015 - New UCSC Genes Track Released for GRCm38/mm10: We're happy to announce the release of an updated UCSC Genes track for the GRCm38/mm10 mouse Genome Browser. Read more.

Genome Browser Gateway

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | Help | About Us

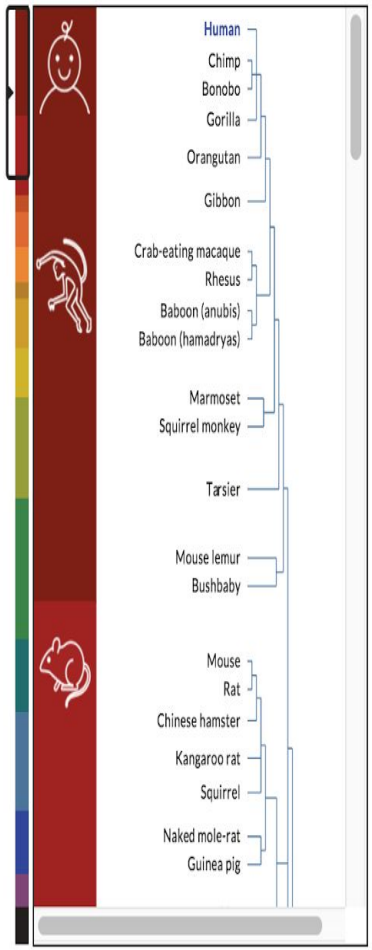
Browse/Select Species

POPULAR SPECIES

Human | Mouse | Rat | Fruitfly | Worm | Yeast

Enter species or common name

REPRESENTED SPECIES



Find Position

Human Assembly: Dec. 2013 (GRCh38/hg38)

GO

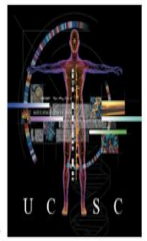
Position/Search Term: Enter position, gene symbol or search terms

Current position: chr9:133,252,000-133,280,861

Human Genome Browser - hg38 assembly

view sequences

UCSC Genome Browser assembly ID: hg38
Sequencing/Assembly provider ID: GRCh38 Genome Reference Consortium Human Reference 38 (GCA_000001405.15)
Assembly date: Dec. 2013
GenBank accession ID: GCA_000001305.2
NCBI Genome information: NCBI genome/51 (Homo sapiens)
NCBI Assembly information: NCBI assembly/883148 (GRCh38/GCA_000001405.15)
BioProject information: NCBI Bioproject: 31257



Homo sapiens (Graphic courtesy of CBSE)

Search the assembly:

- **By position or search term:** Use the "position or search term" box to find areas of the genome associated with many different attributes, such as a specific chromosomal coordinate range; mRNA, EST, or STS marker names; or keywords from the GenBank description of an mRNA. More information, including sample queries.
- **By gene name:** Type a gene name into the "search term" box, choose your gene from the drop-down list, then press "submit" to go directly to the assembly location associated with that gene. More information.
- **By track type:** Click the "track search" button to find Genome Browser tracks that match specific selection criteria. More information.

Download sequence and annotation data:

- Using rsync (recommended)
- Using FTP
- Using HTTP
- Data use conditions and restrictions
- Acknowledgments

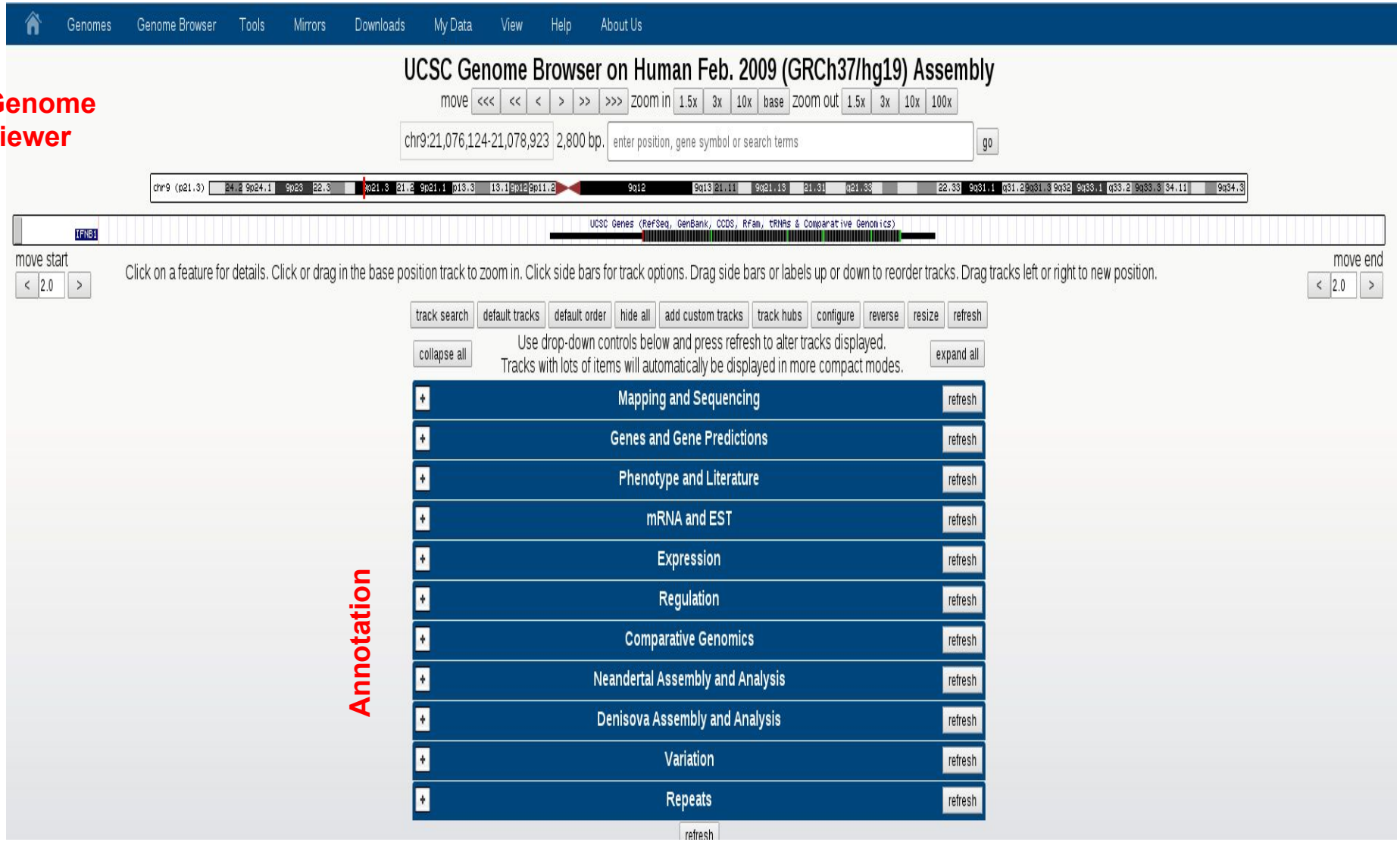
Assembly Details

The GRCh38 assembly is the first major revision of the human genome released in more than four years. As with the previous GRCh37 assembly, the Genome Reference Consortium (GRC) is now the primary source for human genome assembly data submitted to GenBank. Beginning with this release, the UCSC Genome Browser version numbers for the human assemblies now match those of the GRC to minimize version confusion. Hence, the GRCh38 assembly is referred to as "hg38" in the Genome Browser datasets and documentation. For a glossary of assembly-related terms, see the GRC Assembly Terminology page.

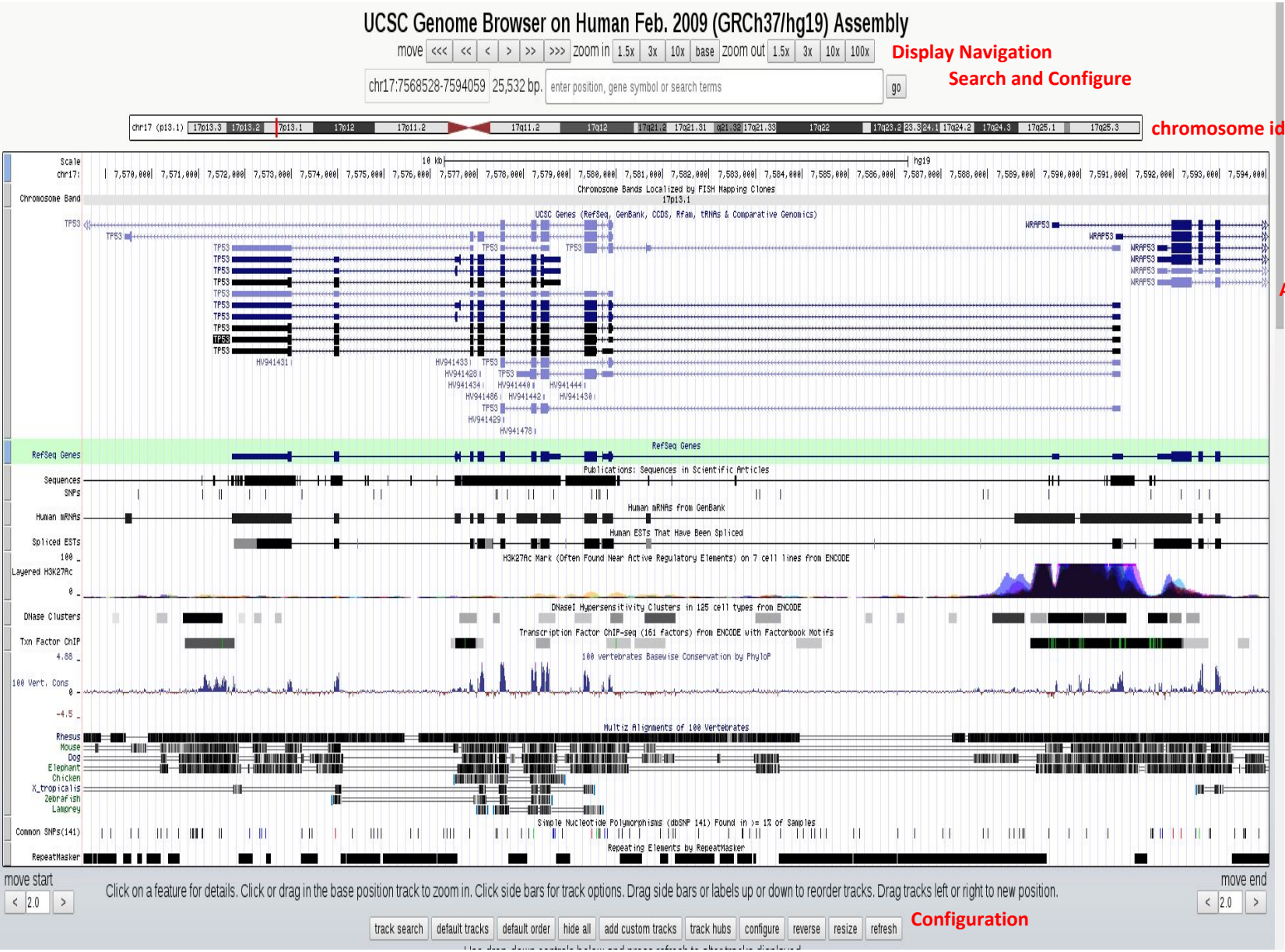
1. UCSC Browser

- Understanding the browser interface
- Basic searches
- Viewing tracks
- Configuring the display
- Navigating
- Printing images
- Retrieving DNA sequences and annotation

Graphical view of genes, gene structure and annotation



Browser Interface

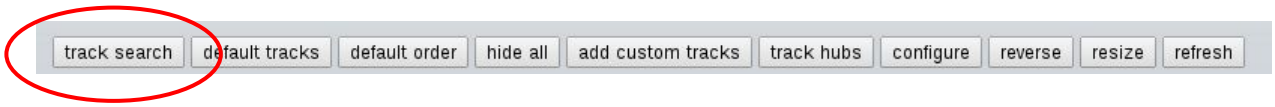


Track Configuration

- Track configuration depends on track type and enables you to;
  - Set data thresholds
  - Include or exclude data from a specific source
  - Choose data labels
  - Choose graph type, height, range and scale

- Track and element descriptions contain additional information

Configuring the genome browser display



Search for data types

Search | Advanced

e2f1

search | clear | cancel

return to browser (0 of 10 selected)

+ -	Visibility		Track Name	Sort: by Relevance / Alphabetically / by Hierarchy
☐	hide		HeLa E2F1 Std	HeLa-S3 E2F1 Standard ChIP-seq Signal from ENCODE/SYDH
☐	hide		HeLa E2F1 Std	HeLa-S3 E2F1 Standard ChIP-seq Peaks from ENCODE/SYDH
☐	hide		MCF-7 E2F1	MCF-7 TFBS Uniform Peaks of HA-E2F1 from ENCODE/USC/Analysis
☐	hide		HeLa-S3 E2F1 c2	HeLa-S3 TFBS Uniform Peaks of HA-E2F1 from ENCODE/USC/Analysis
☐	hide		HeLa-S3 E2F1 c1	HeLa-S3 TFBS Uniform Peaks of E2F1 from ENCODE/USC/Analysis
☐	hide		MCF7 HAE2 UCD	MCF-7 HA-E2F1 UC Davis ChIP-seq Signal from ENCODE/SYDH
☐	hide		MCF7 HAE2 UCD	MCF-7 HA-E2F1 UC Davis ChIP-seq Peaks from ENCODE/SYDH
☐	hide		HeLa HAE2 Std	HeLa-S3 HA-E2F1 Standard ChIP-seq Signal from ENCODE/SYDH
☐	hide		HeLa HAE2 Std	HeLa-S3 HA-E2F1 Standard ChIP-seq Peaks from ENCODE/SYDH
☐	hide	🔧	SYDH TFBS	Transcription Factor Binding Sites by ChIP-seq from ENCODE/Stanford/Yale/USC/Harvard

Return to Browser (0 of 10 selected)

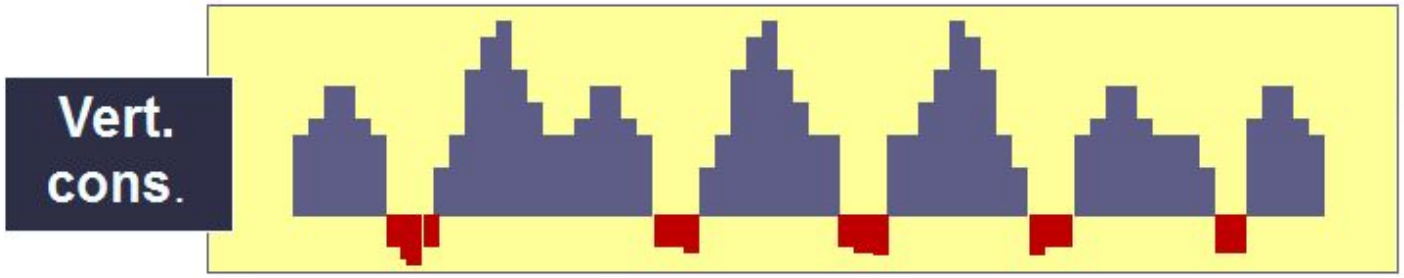
🔧 Tracks so marked are containers which group related data tracks. Containers may need additional configuration (by clicking on the 🔧 icon) before they can be viewed in the browser.

Visual cues

Intron and direction of transcription <<< or >>>

Track colors *may* have meaning—for example, UCSC Gene track:

- If there is a corresponding PDB entry = black
- If there is a corresponding reviewed/validated seq = dark blue
- If there is a non-RefSeq seq = lightest blue

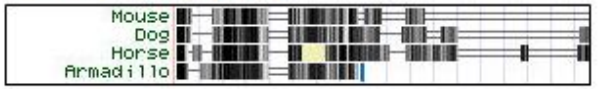


Wiggle

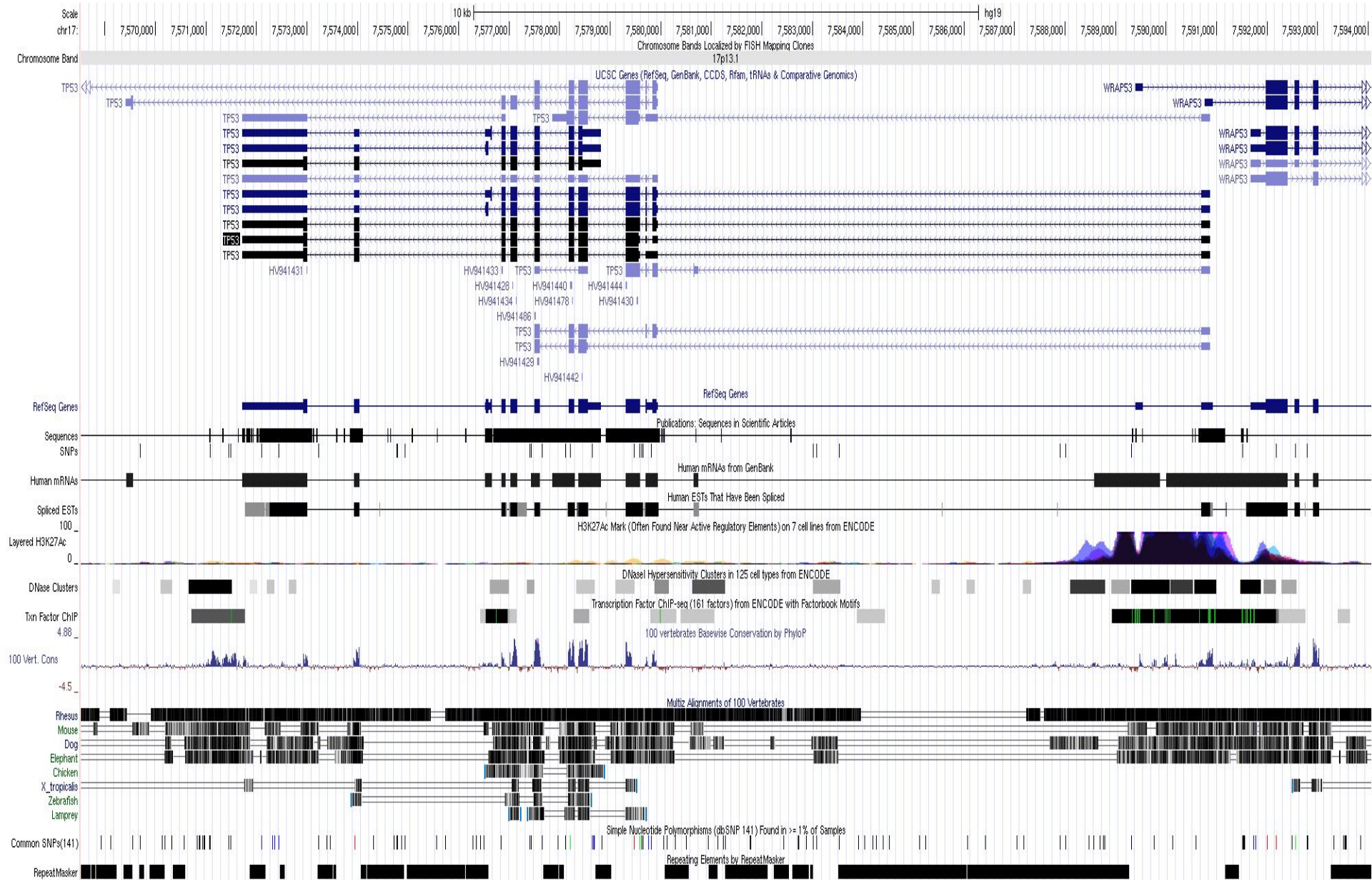
height of a blue bar is increased likelihood of conservation,
red indicates a likelihood of faster-evolving regions

Alignment indications (Conservation pairs: "chain" or "net" style)

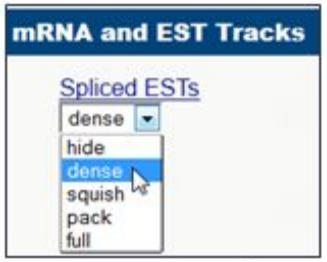
- Alignments = boxes, Gaps = lines



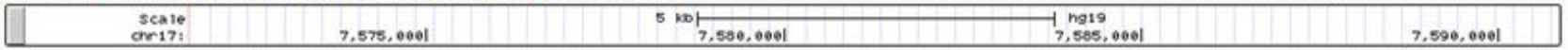
Example search for human TP53



Annotation Track menu options



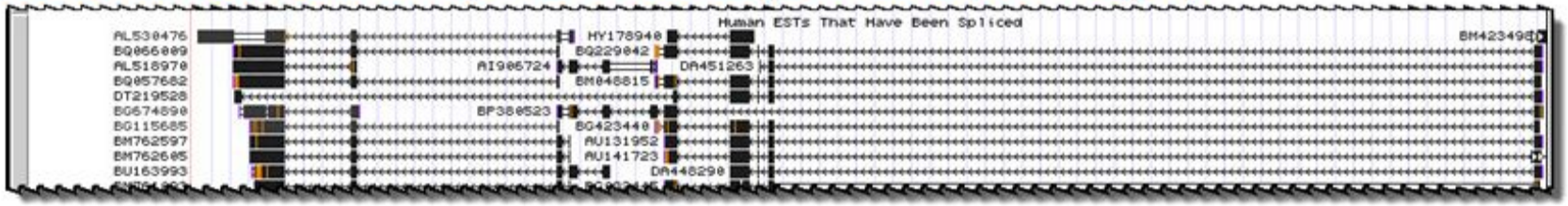
- Hide: removes a track from view



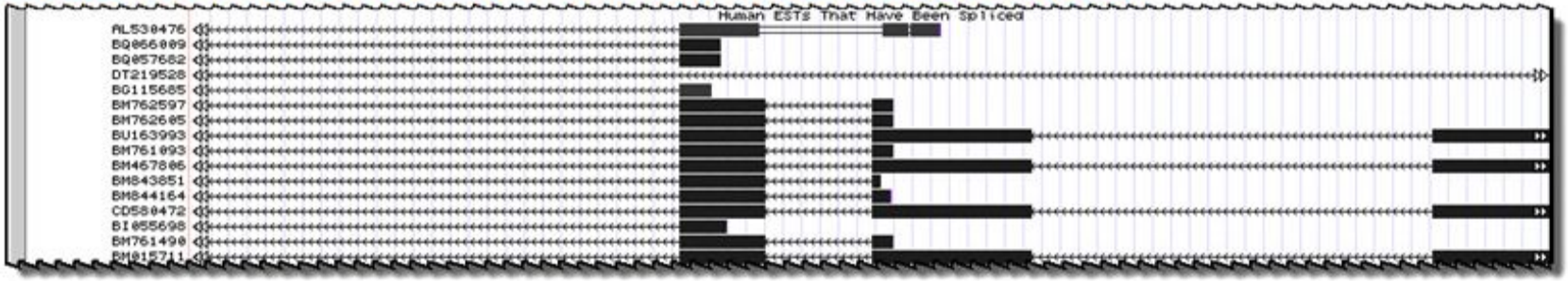
- Dense: all items collapsed into a single line

- Squish: each item = separate line, but 50% height + packed

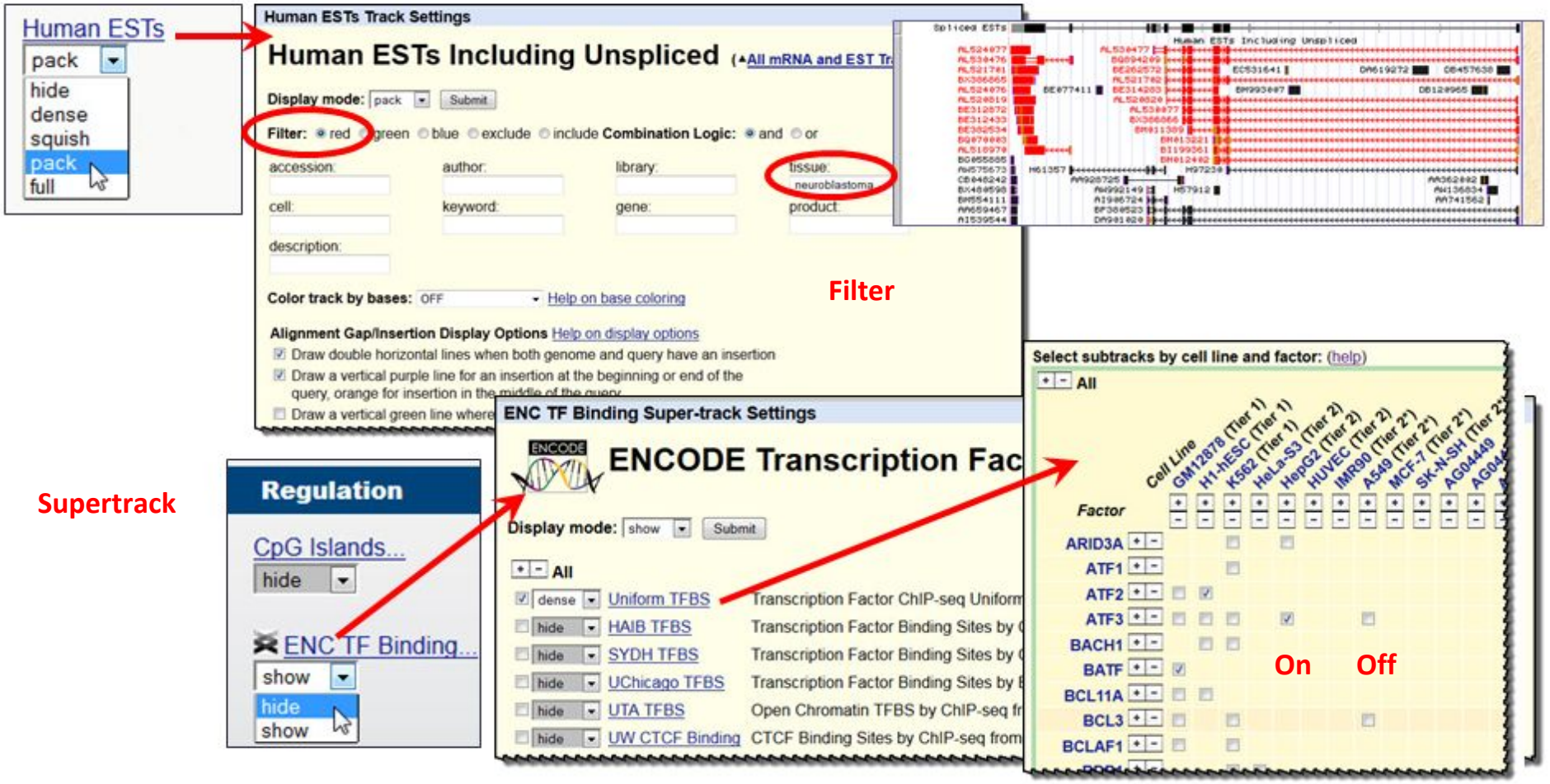
- Pack: each item separate, but efficiently stacked (full height)



- Full: each item on separate line (may need to zoom to fit)

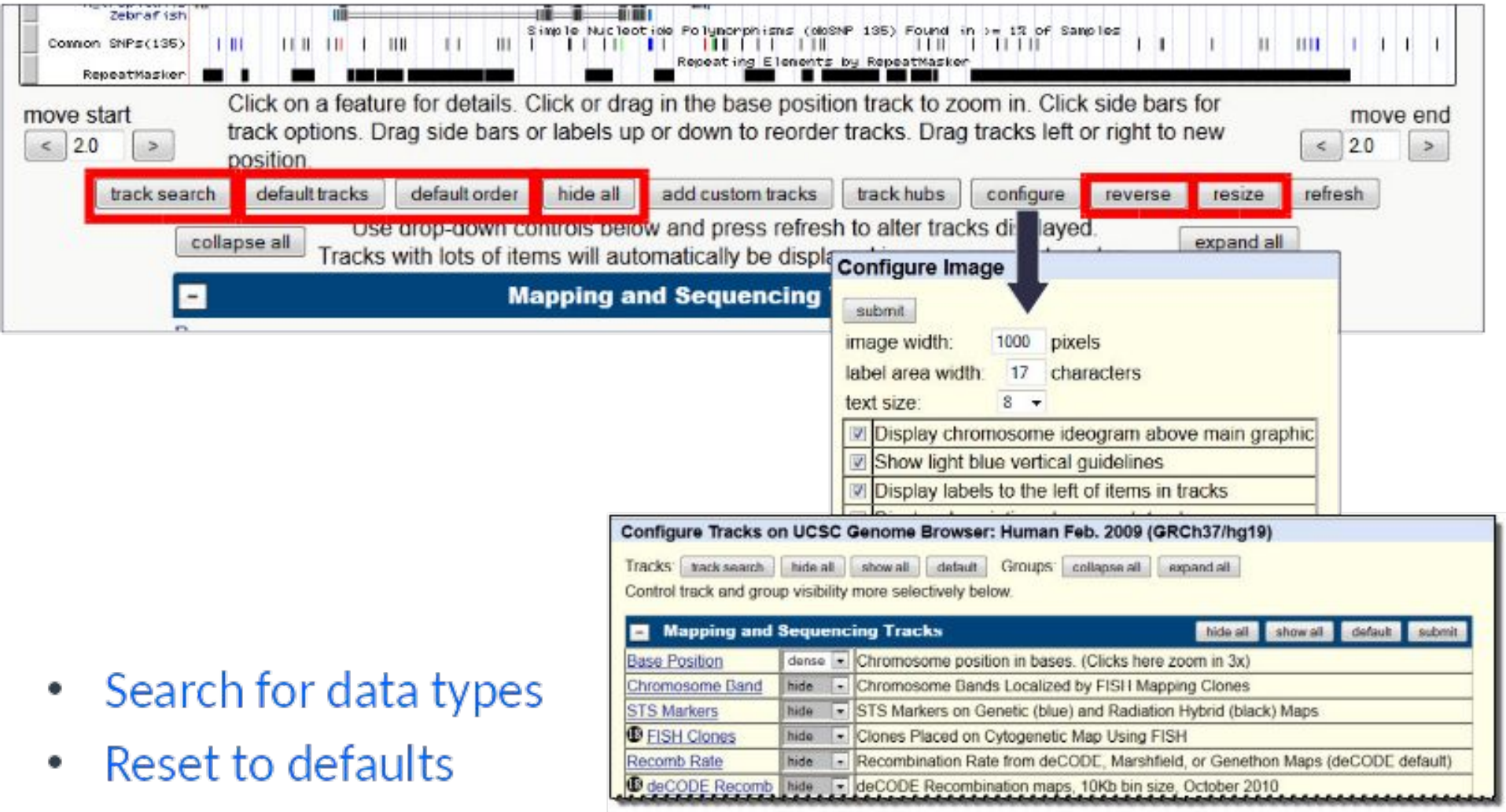


Additional Options: Filters, Supertracks …



- Some tracks have filters *(ESTs shown; SNPs other good example)*
- Super-tracks may have multiple components, various settings
- Some tracks may have un-displayed data

Mid page options to change settings

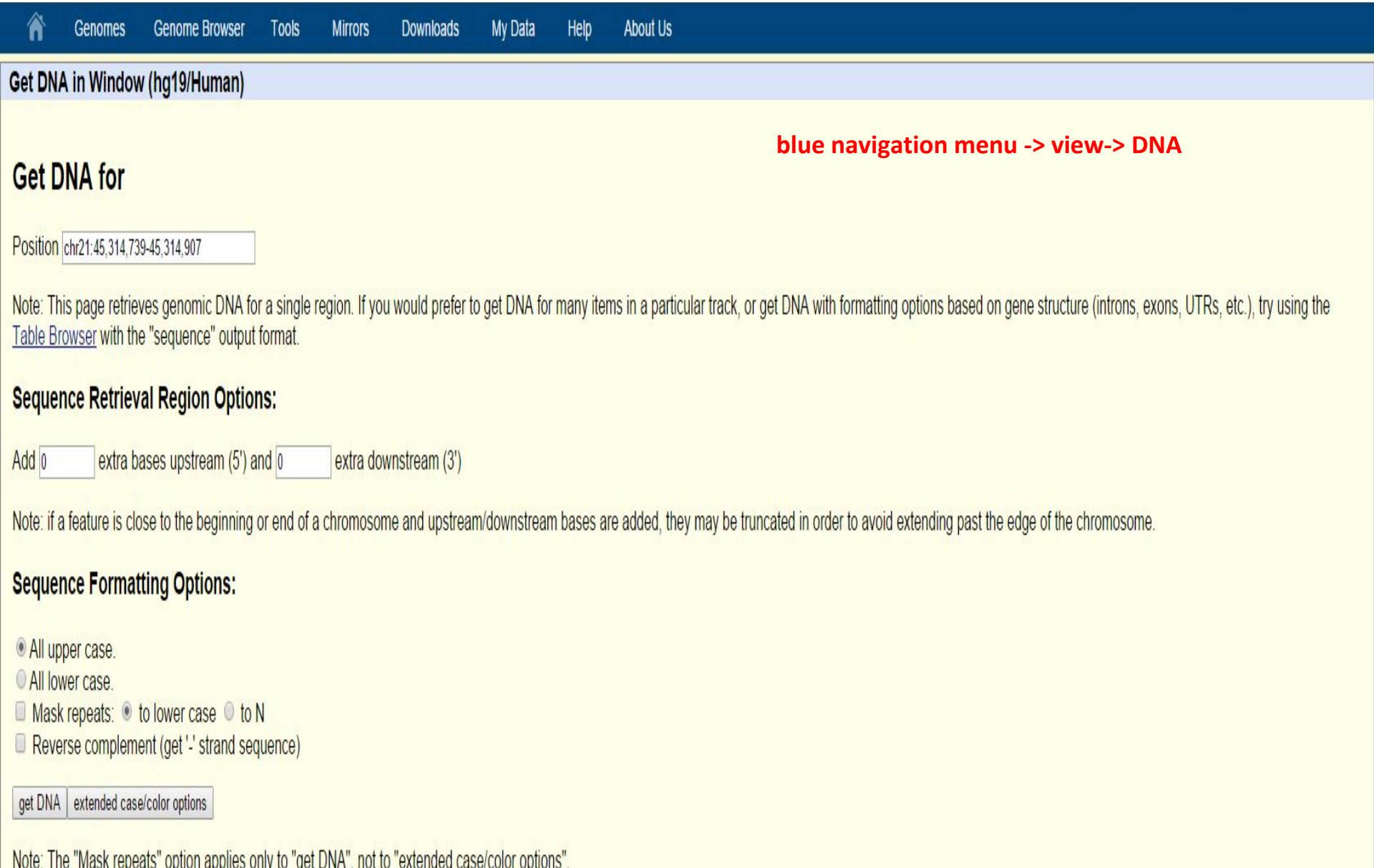


- Search for data types
- Reset to defaults
- Configure options page
- You control the views with numerous features

Printing track figures

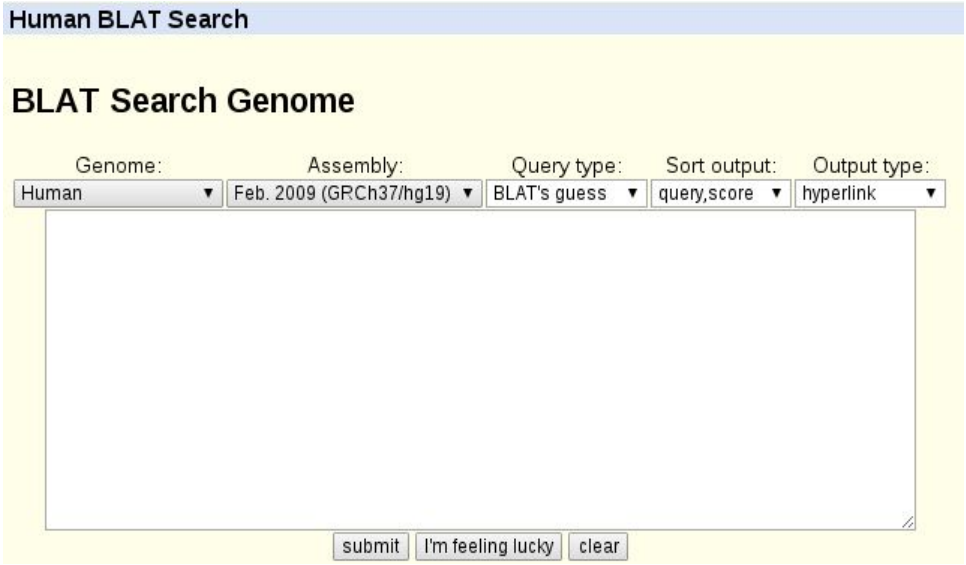
- Customize track
- Add title
- consider showing only one transcript per gene by turning off splice variants
- Increase the font size and remove the light blue vertical guide lines in the image configuration menu
- Change image size
- Click on blue navigation menu-> view ->**PDF/PS** link

Retrieve DNA sequence

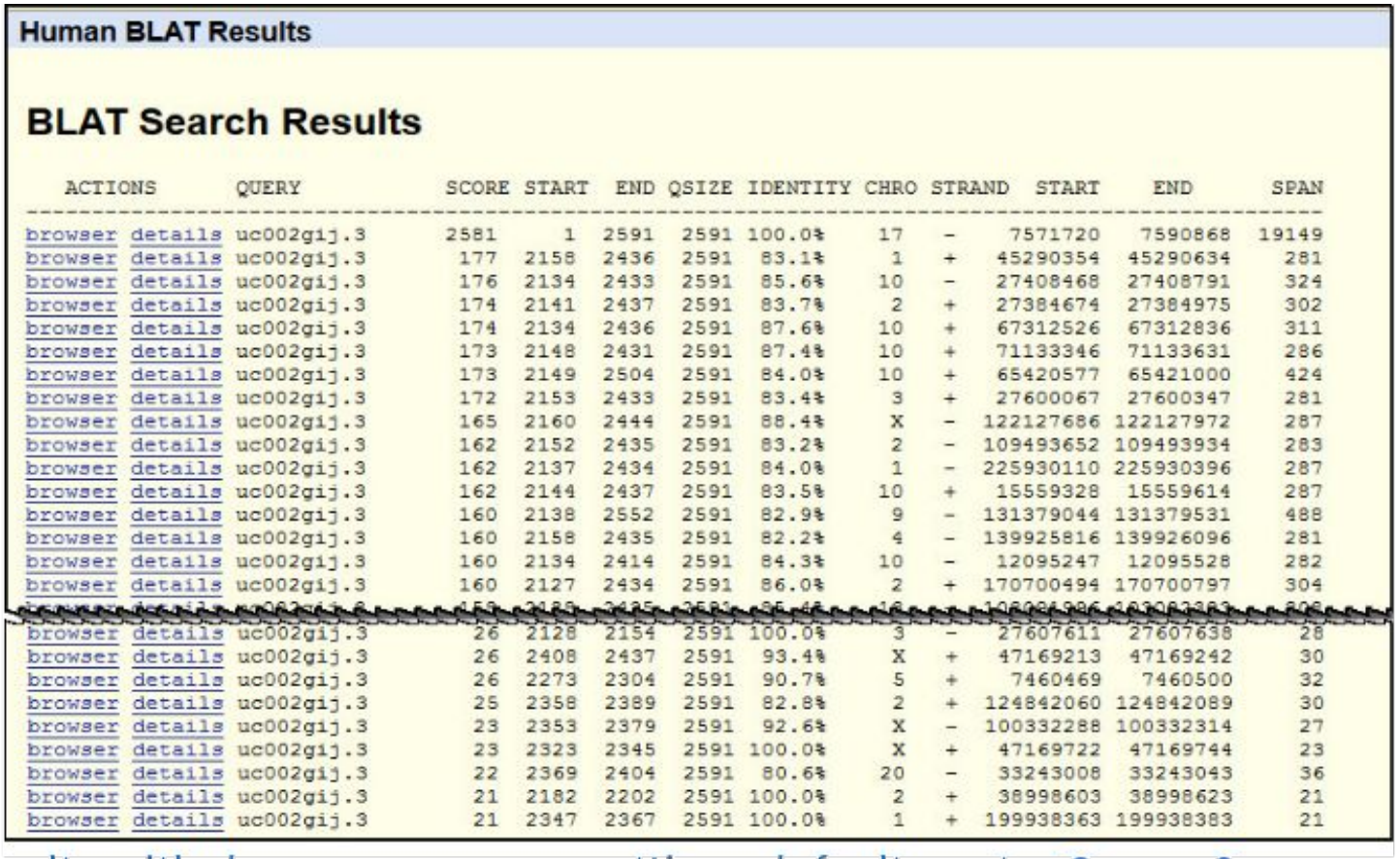


2. BLAT (Blast Like Alignment Tool)

- Rapid sequence search by indexing entire genome
- Useful for finding high similarity matches
- 95% and greater similarity of length 25 bases or more OR sequences of 80% and greater similarity of length 20 amino acids or more
- Limits: DNA (25000 bp), Protein (10000 aa) or 25 sequences
- Can be installed and run locally



BLAT results



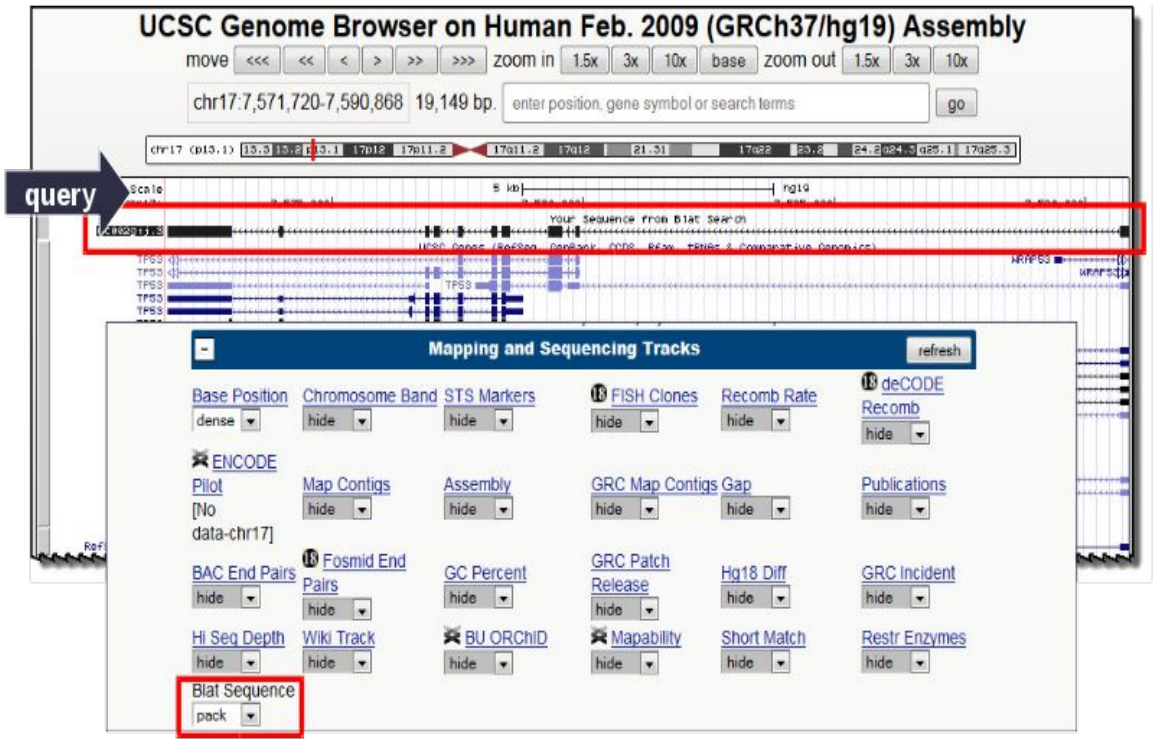
Human BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	uc002gij.3	2581	1	2591	2591	100.0%	17	-	7571720	7590868	19149
browser details	uc002gij.3	177	2158	2436	2591	83.1%	1	+	45290354	45290634	281
browser details	uc002gij.3	176	2134	2433	2591	85.6%	10	-	27408468	27408791	324
browser details	uc002gij.3	174	2141	2437	2591	83.7%	2	+	27384674	27384975	302
browser details	uc002gij.3	174	2134	2436	2591	87.6%	10	+	67312526	67312836	311
browser details	uc002gij.3	173	2148	2431	2591	87.4%	10	+	71133346	71133631	286
browser details	uc002gij.3	173	2149	2504	2591	84.0%	10	+	65420577	65421000	424
browser details	uc002gij.3	172	2153	2433	2591	83.4%	3	+	27600067	27600347	281
browser details	uc002gij.3	165	2160	2444	2591	88.4%	X	-	122127686	122127972	287
browser details	uc002gij.3	162	2152	2435	2591	83.2%	2	-	109493652	109493934	283
browser details	uc002gij.3	162	2137	2434	2591	84.0%	1	-	225930110	225930396	287
browser details	uc002gij.3	162	2144	2437	2591	83.5%	10	+	15559328	15559614	287
browser details	uc002gij.3	160	2138	2552	2591	82.9%	9	-	131379044	131379531	488
browser details	uc002gij.3	160	2158	2435	2591	82.2%	4	-	139925816	139926096	281
browser details	uc002gij.3	160	2134	2414	2591	84.3%	10	-	12095247	12095528	282
browser details	uc002gij.3	160	2127	2434	2591	86.0%	2	+	170700494	170700797	304
[illegible]											
browser details	uc002gij.3	26	2128	2154	2591	100.0%	3	-	27607611	27607638	28
browser details	uc002gij.3	26	2408	2437	2591	93.4%	X	+	47169213	47169242	30
browser details	uc002gij.3	26	2273	2304	2591	90.7%	5	+	7460469	7460500	32
browser details	uc002gij.3	25	2358	2389	2591	82.8%	2	+	124842060	124842089	30
browser details	uc002gij.3	23	2353	2379	2591	92.6%	X	-	100332288	100332314	27
browser details	uc002gij.3	23	2323	2345	2591	100.0%	X	+	47169722	47169744	23
browser details	uc002gij.3	22	2369	2404	2591	80.6%	20	-	33243008	33243043	36
browser details	uc002gij.3	21	2182	2202	2591	100.0%	2	+	38998603	38998623	21
browser details	uc002gij.3	21	2347	2367	2591	100.0%	1	+	199938363	199938383	21

- Results with demo sequences, settings default; sort = Query, Score
  - **Score is a count of matches—higher number, better match**
- Click browser to go to Genome Browser image location (next slide)
- Click details to see the alignment to genomic sequence (2nd slide)

Browser link



- From browser click in BLAT results
- A new track line with *Your Sequence from BLAT Search* appears

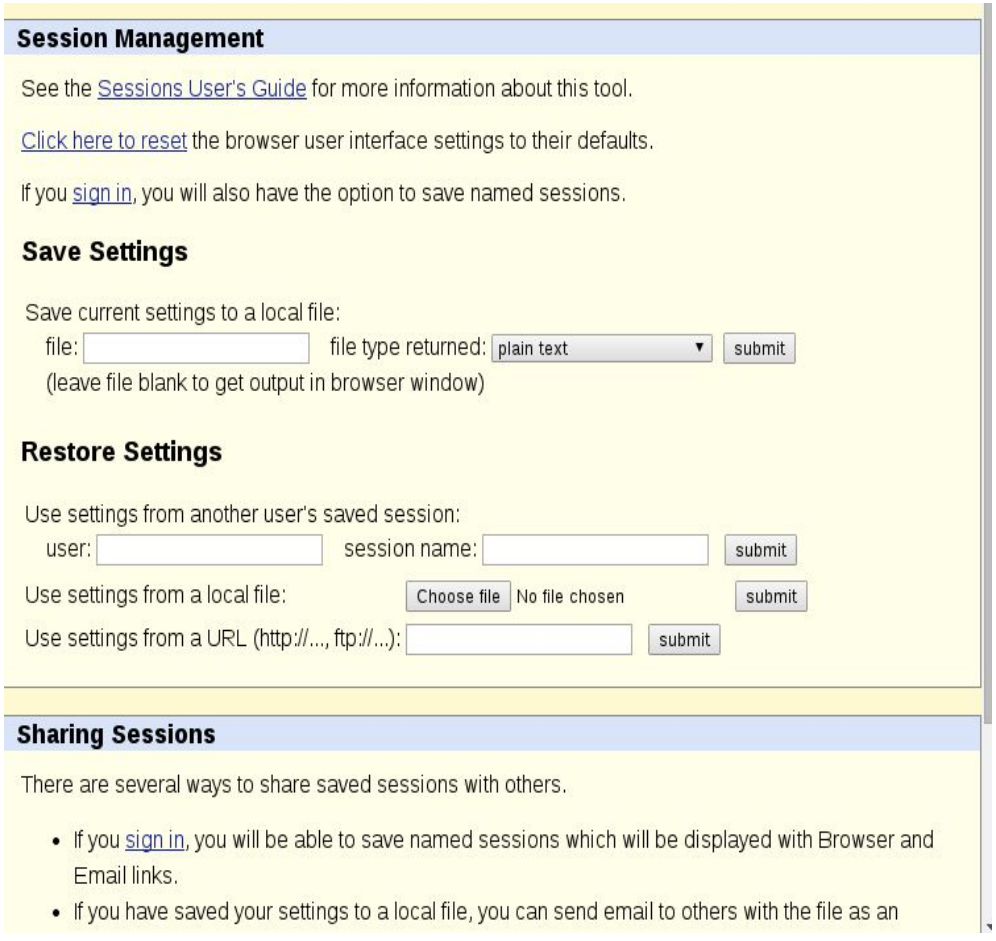
Details link

3. Custom tracks, session and track Hubs

Sessions

- **Signing in** enables you to save current settings into a named session, and then restore settings from the session later.

- lifespan: 4 months
- If you wish, you can share named sessions with other users.
- Individual sessions may be designated as either *shared* or *non-shared* to protect the privacy of confidential data.

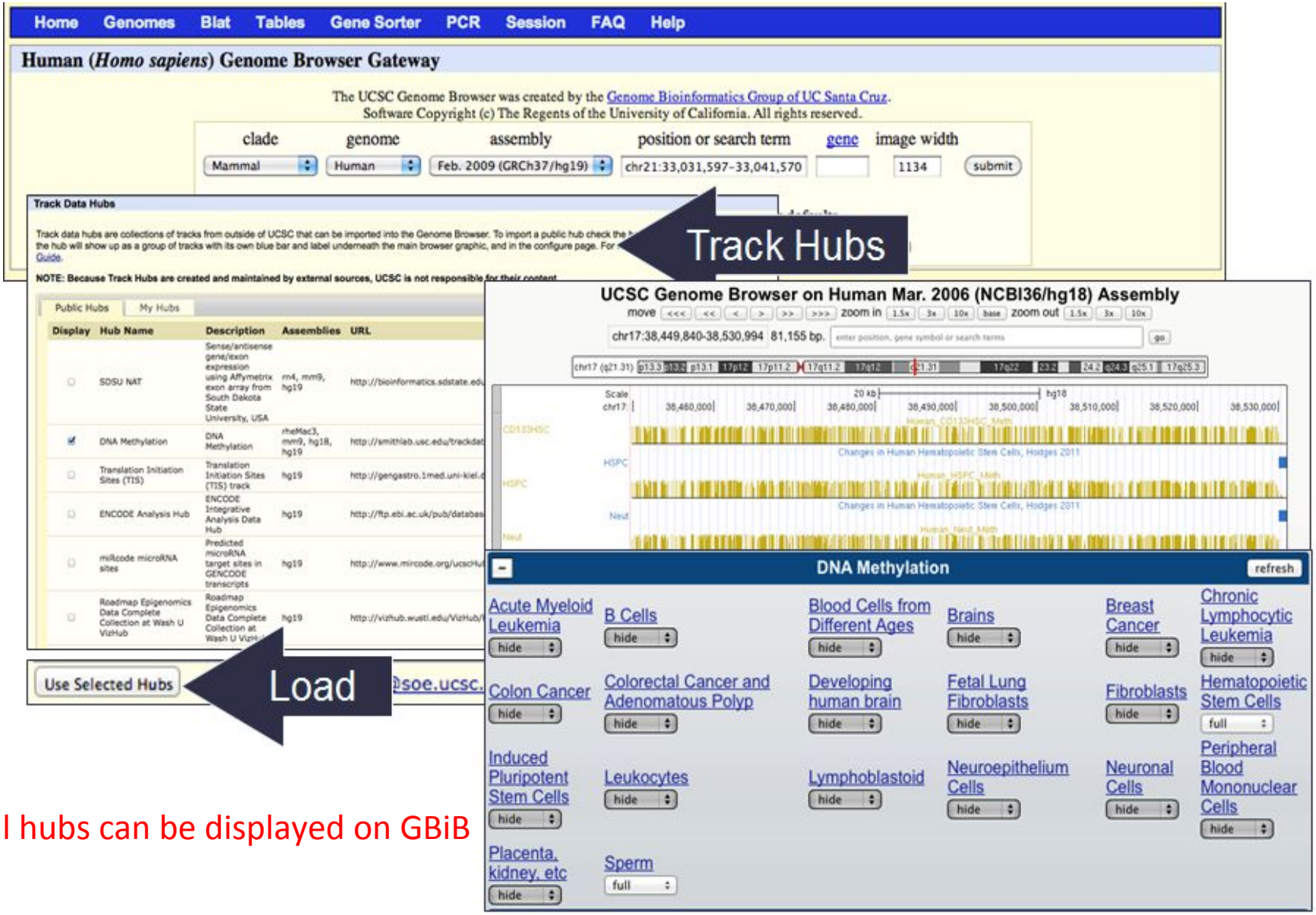


Custom tracks

- It is possible for users to upload their own annotation data for temporary display in the browser.
- These custom annotation tracks are viewable only on the machine from which they were uploaded and are automatically deleted 48 hours after the last time they are accessed, unless they are saved in a Session.
- Optionally, users can make custom annotations viewable by others as well.

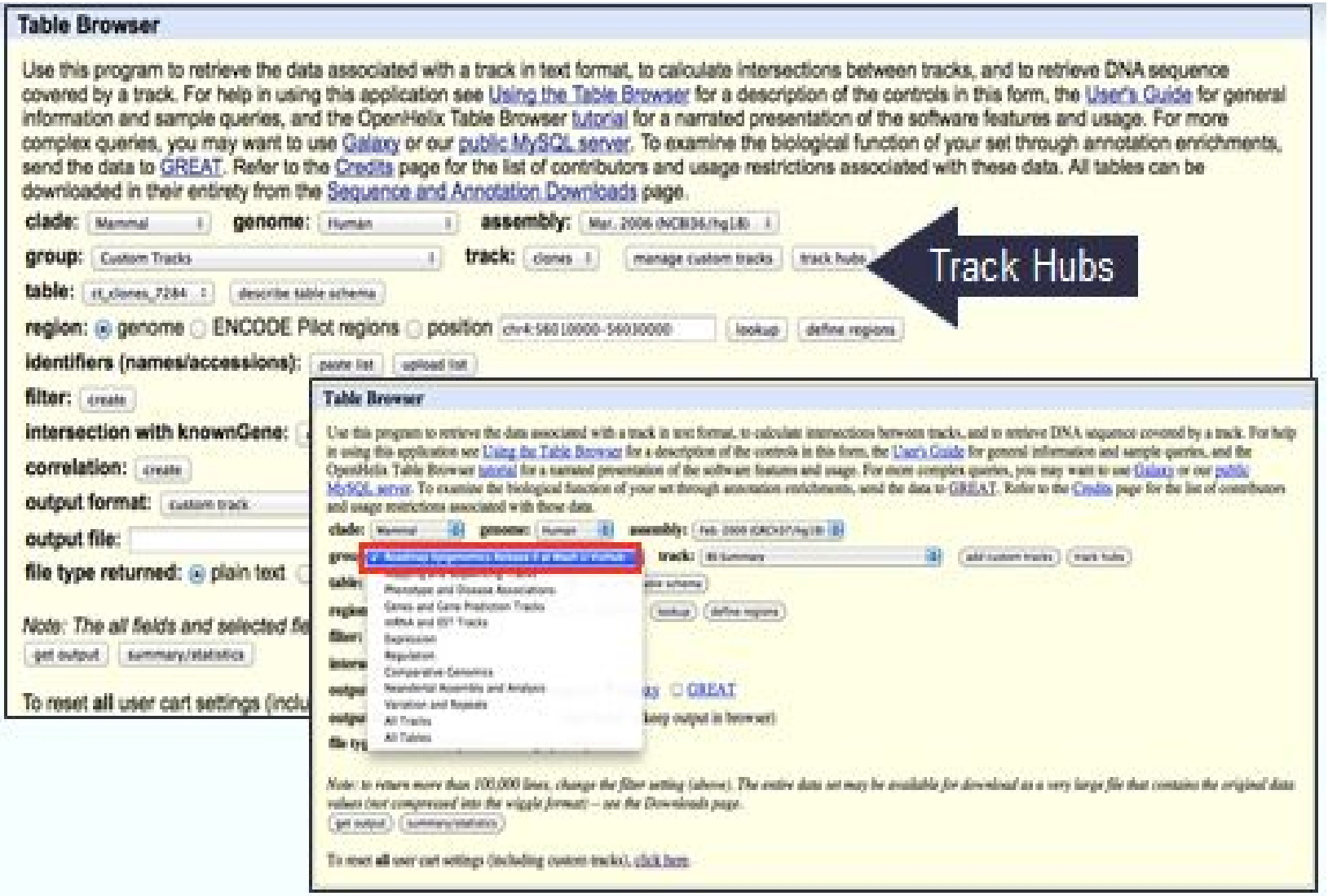
- Format your data
- Define browser characteristics
- Define track characteristics
- Upload and view your track
- Add URL for annotation details (option)

Track Hubs



Local hubs can be displayed on GBiB

Track Hubs



4. UCSC Table Browser

- Search for genes and annotation
- Setup and filters
- Join tables
- Retrieve sequences
- Intersecting tracks
- Export to external resources

Table browser interface

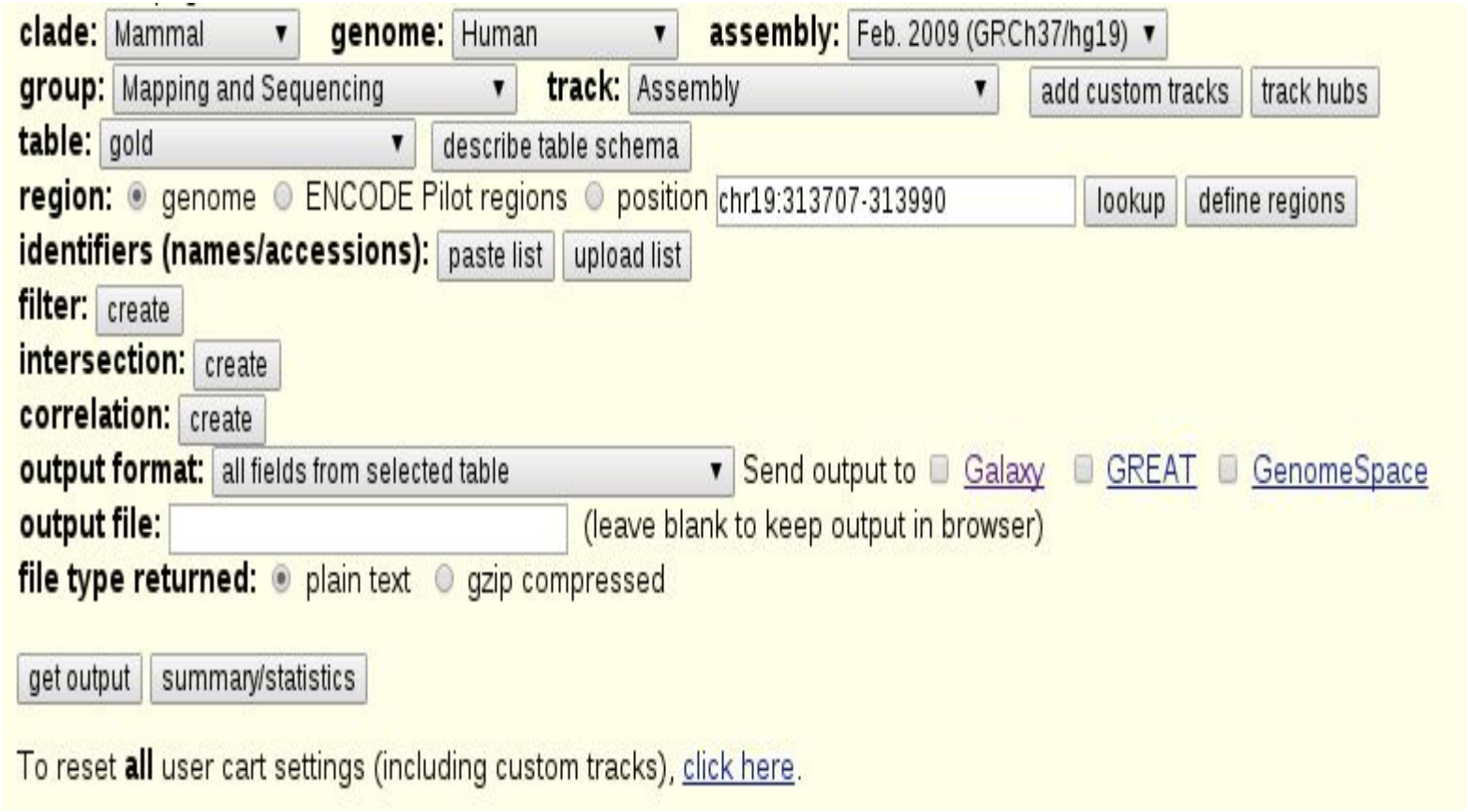


Table browser usage

- Retrieve the DNA sequence data or annotation data underlying Genome Browser tracks for the entire genome, a specified coordinate range, or a set of accessions
- Apply a filter to set constraints on field values included in the output
- Generate a custom track and automatically add it to your session so that it can be graphically displayed in the Genome Browser
- Conduct both structured and free-from SQL queries on the data
- Combine queries on multiple tables or custom tracks through an intersection or union and generate a single set of output data
- Display basic statistics calculated over a selected data set
- Display the schema for table and list all other tables in the database connected to the table
- Organize the output data into several different formats for use in other applications, spreadsheets, or databases

Table Browser driven discovery

Task: Search entire genome for "CAG" trinucleotide repeats from USCS tables.

- Choose genome [hg19]
- Choose table [Repeats->Simple Repeats]
- Describe table -find correct data fields
- Choose region [genome]
- Upload locations
- Data summary - approx. 1 million simple repeats

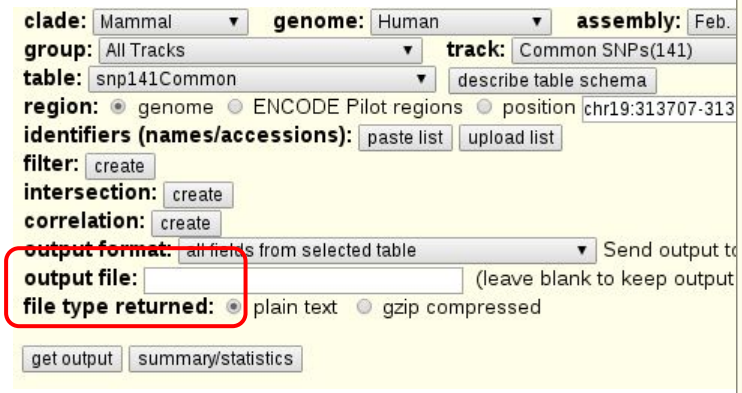
Features of trinucleotide expansion in humans

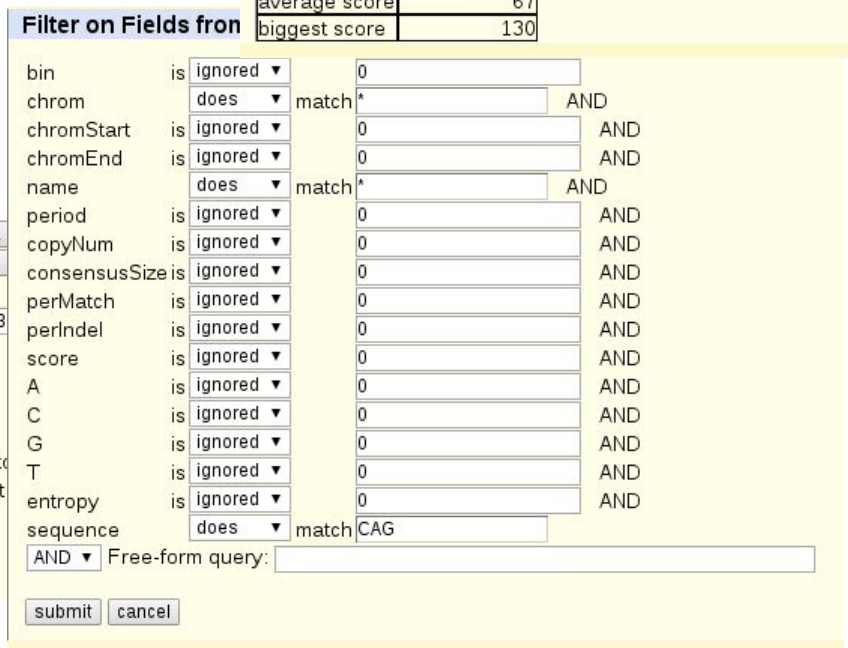
Disease	Sequence	Location	Parent of origin of expansion	Repeat number (normal)	Repeat number (pre-mutation)	Repeat number (disease)	Somatic instability
Diseases with coding TNRs							
DRPLA	CAG	*ATN1* (exon 5)	P	6–35	35–48	49–88	Yes
HD	CAG	*HTT* (exon 1)	P	6–29	29–37	38–180	Yes
OPMD	GCN	*PABPN1* (exon 1)	P and M	10	12–17	>11	None found in tissue tested (hypothalamus)
SCA1	CAG	*ATXN1* (exon 8)	P	6–39	40	41–83	Yes
SCA2	CAG	*ATXN2* (exon 1)	P	<31	31–32	32–200	Unknown
SCA3 (Machado–Joseph disease)	CAG	*ATXN3* (exon 8)	P	12–40	41–85	52–86	Unknown
SCA6	CAG	*CACNA1A* (exon 47)	P	<18	19	20–33	None found
SCA7	CAG	*ATXN7* (exon 3)	P	4–17	28–33	>36 to >460	Yes
SCA17	CAG	*TBP* (exon 3)	P > M	25–42	43–48	45–66	Yes
SMBA	CAG	*AR* (exon 1)	P	13–31	32–39	40	None found

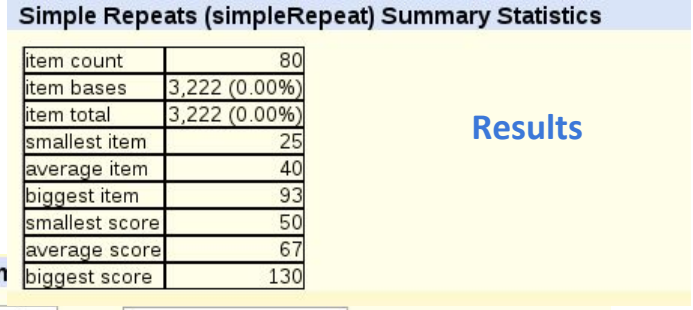
McMurray CT. Mechanisms of trinucleotide repeat instability during human development. Nat Rev Genet. 2010 Nov;11(11):786-99.

modified from openhelix UCSC tutorial

Table Browser:Filtering







Results

search for simple repeats in the entire genome with “CAG” sequence and extract data table.

Table Browser: Intersections

- Combines the output of two queries into a single set of data based on specific join criteria.
- For example, this can be used to find all SNPs that intersect with RefSeq coding regions. The intersection can be configured to retain the existing alignment structure of the table with a specified amount of overlap, or discard the structure in favor of a simple list of position ranges using a base-pair intersection or union of the two data sets.
- The button functionalities are similar to those of the *filter* option.

Accessing genome browsers via R

- rtracklayer bioconductor package
- Download tracks from UCSC table browser

get repeat-masked regions in and around the transcription start site (TSS) of the human E2F3 gene, in hg19:

```
library (rtracklayer)
mySession = browserSession("UCSC")
genome(mySession) <- "hg19"
e2f3.tss.grange <- GRanges("chr6", IRanges(20400587, 20403336))
tbl.rmsk <- getTable( ucscTableQuery(mySession, track="rmsk", range=e2f3.tss.grange, table="rmsk"))
```

Other tools

- Gene sorter
- *In silico* PCR
- VisiGene browser
- Cancer Browser and Encode portal
- Genome graphs
- Other tools:
  - liftOver
  - Dusters
  - Tree maker

Search for related genes

UCSC Human Gene Sorter

genome Human assembly Mar. 2006 (NCBI36/hg18) search tp53 Go!
sort by Expression (GNF Atlas2) configure filter (now off) display 50 output sequence text

About the Gene Sorter

This program displays a sorted table of genes that are related to one another. The relationship can be one of several types, including protein-level homology, similarity of gene expression profiles, or genomic proximity.

To display a gene and its relatives:

1. Select a genome and assembly from the corresponding pull-down menus.
2. Type a word or phrase into the *search* text box to specify which gene should be displayed in the Gene Sorter. Examples of search terms include FOXA2, HOXA9, and MAP kinase.
3. Choose the gene relationship with which you would like to sort the list by selecting an option from the *sort by* pull-down menu.
4. Press the Go! button to display your results.

Following a successful search, the Gene Sorter displays a table containing the specified gene -- highlighted in light green -- and its relatives, each on a separate line. To adjust the number of rows shown, select an option from the *display* pull-down menu.

The default set of table columns -- which can be expanded, reduced, and rearranged via the *configure* button -- shows additional information about the genes. Some of the column data, such as those in the *BLAST E-value* and *%ID* columns, are calculated relative to the highlighted gene. To select a different gene in the list, click on its name. Clicking on a gene's *Genome Position* will open the UCSC Genome Browser to the location of that gene. Similarly, clicking on a gene's *Description* will open a page showing detailed information about the gene.

One of the most powerful features of the Gene Sorter is its filtering capabilities, accessed via the *filter* button. Use the filter to fine-tune the list of displayed genes to a subset based on a selection of detailed and flexible criteria. For example, the filter may be used to select all human genes over-expressed in the cerebellum that have GO-annotated G-protein coupled receptor activity.

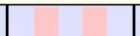
The Gene Sorter offers two options for displaying and downloading sequence associated with the genes in the table. Clicking on the *sequence* button will fetch associated protein, mRNA, promoter, or genomic sequence. To dump the table into a simple tab-delimited format suitable for import into a spreadsheet or relational database, click the *text* button.

The UCSC Gene Sorter was designed and implemented by Jim Kent, Fan Hsu, Donna Karolchik, David Haussler, and the UCSC Genome Bioinformatics Group. This work is supported by a grant from the National Human Genome Research Institute and by the Howard Hughes Medical Institute.

Gene Sorter

UCSC Human Gene Sorter

genome Human assembly Mar. 2006 (NCBI36/hg18) search uc002gij.2 Go!
sort by Expression (GNF Atlas2) configure filter (now off) display 25 output sequence text

#	Name	VisiGene	fetal brain	whole brain	amygdala	thymus	bone marrow	PB-CD4+ Tcells	skin	adipocyte	pancreatic islets	heart	lung	kidney	liver	ovary	testis	BLASTP E-Value	Genome Position	Description
1	TP53	n/a																0	chr17 7,522,016	tumor protein p53 isoform a
2	RPS20	n/a																n/a	chr8 57,148,895	ribosomal protein S20
3	H2AFV	n/a																n/a	chr7 44,846,994	H2A histone family, member V isoform 1
4	RPL7A	187765																n/a	chr9 135,206,495	ribosomal protein L7a
5	RPS13	n/a																n/a	chr11 17,054,155	ribosomal protein S13
6	SNRPG	181122																n/a	chr2 70,368,191	small nuclear ribonucleoprotein polypeptide G
7	EIF4A1	176036																n/a	chr17 7,419,687	eukaryotic translation initiation factor 4A
8	ADSL	77625																n/a	chr22 39,082,485	adenylosuccinate lyase isoform a
9	CR601950	n/a																n/a	chr17 72,069,204	Homo sapiens primary hepatoblastoma cDNA, clone:HKMT0728, full insert sequence.
10	UBE2A	182203																n/a	chrX 118,597,467	ubiquitin-conjugating enzyme E2A isoform 1
11	GMPS	176663																n/a	chr3 157,104,616	guanine monophosphate synthetase
12	G3BP1	176455																n/a	chr5 151,148,388	Ras-GTPase-activating protein SH3-domain-binding
13	NUP37	187198																n/a	chr12 101,014,297	nucleoporin 37kDa
14	QARS	180161																n/a	chr3 49,112,772	glutaminyl-tRNA synthetase
15	ZNF207	26352																n/a	chr17 27,711,425	zinc finger protein 207 isoform c
16	XRCC5	n/a																n/a	chr2 216,730,812	ATP-dependent DNA helicase II
17	LOC647099	n/a																n/a	chr17 24,073,314	similar to ribosomal protein L23A
18	PABPC4	36799																n/a	chr1 39,807,039	poly A binding protein, cytoplasmic 4 isoform 2
19	RPS18	180521																n/a	chr6 33,350,044	ribosomal protein S18
20	RPS18	n/a																n/a	chr6_cox_hap1 4,622,203	ribosomal protein S18
21	RPS18	n/a																n/a	chr6_qbl_hap2 4,428,251	ribosomal protein S18
22	PSMA5	180067																n/a	chr1 109,758,277	proteasome alpha 5 subunit
23	LOC441743	n/a																n/a	chr16 376,999	Uncharacterized protein ENSP00000332117.
24	PHF10	27218																n/a	chr6 169,855,917	PHD finger protein 10 isoform a
25	RPS27	59894																n/a	chr1 152,230,551	ribosomal protein S27

Configure

Configure Gene Sorter

submit | Columns: hide all | show all | default | Settings: save | load

Expression ratio colors: red high/green low | Show all splicing variants: | custom columns

Name	On	Position	Description	Configuration
#	☑	▼	Item Number in Displayed List/Select Gene	n/a
Name	☑	▲▼	Gene Name/Select Gene	n/a
UniProtKB	☐	▲▼	UniProtKB Protein Display ID	n/a
UniProtKB Acc	☐	▲▼	UniProtKB Protein Accession	n/a
RefSeq	☐	▲▼	NCBI RefSeq Gene Accession	n/a
Entrez Gene	☐	▲▼	NCBI Entrez Gene/LocusLink ID	n/a
UCSC ID	☐	▲▼	UCSC Transcript ID	n/a
GenBank	☐	▲▼	GenBank mRNA Accession	n/a
Ensembl	☐	▲▼	Ensembl Transcript ID	n/a
KEGG	☐	▲▼	KEGG Pathway ID	n/a
GNF Atlas 2 ID	☐	▲▼	ID of Associated GNF Atlas 2 Expression Data	n/a
Gene Category	☐	▲▼	High Level Gene Category - Coding, Antisense, etc.	n/a
CDS Score	☐	▲▼	Coding potential score from txCdsPredict	n/a
VisiGene	☑	▲▼	UCSC VisiGene In Situ Image Browser	n/a
Allen Brain	☐	▲▼	Allen Brain Atlas In Situ Images of Adult Mouse Brains	n/a
U133 ID	☐	▲▼	ID of Associated Affymetrix U133 Expression Data	n/a
U133Plus2 ID	☐	▲▼	ID of Associated Affymetrix U133 Plus 2.0 Expression Data	n/a
U95 ID	☐	▲▼	ID of Associated Affymetrix U95 Expression Data	n/a
GNF Atlas 2	☑	▲▼	GNF Expression Atlas 2 Data from U133A and GNF1H Chips	brightness: 1.0 tissues: selected values: ratio
H-Inv	☐	▲▼	H-Invitational Gene Database	n/a
Max GNF Atlas 2	☐	▲▼	Maximum Expression Value of GNF Expression Atlas 2	n/a
GNF Atlas 2 Delta	☐	▲▼	Normalized Difference in GNF Expression Atlas 2 from Selected Gene	n/a
GNF U95	☐	▲▼	GNF Expression Atlas 1 Human Data on Affy U95 Chips	brightness: 1.0 tissues: selected values: ratio
Max GNF U95	☐	▲▼	Maximum Expression Value of GNF Expression Atlas 1	n/a
GNF Atlas1 Delta	☐	▲▼	Normalized Difference in GNF Atlas 1 Expression from Selected Gene	n/a
Affy Exons	☐	▲▼	Affymetrix All Exon Microarrays	brightness: 1.0
Affy Exon Dst	☐	▲▼	Affymetrix All Exon Microarrays Distance	n/a
BLASTP Bits	☐	▲▼	NCBI BLASTP Bit Score	n/a
BLASTP E-Value	☑	▲▼	NCBI BLASTP E-Value	n/a

Filter

Gene Sorter Filter

On this page you can restrict which genes appear in the main table based on the values in any column. Click the *submit* button to return to the main Gene Sorter page with the current filter settings applied.

submit | clear filter | save filter | load filter

Quickly obtain a list of gene names that pass the filter: list names

Filter Controls for Displayed Columns:

Name - Gene Name/Select Gene

Name search (including * and ? wildcards):

Include if any ▾ words in search term match.

Limit to items (no wildcards) in list: paste list | upload list

VisiGene - UCSC VisiGene In Situ Image Browser

VisiGene search (including * and ? wildcards):

Include if any ▾ words in search term match.

Limit to items (no wildcards) in list: paste list | upload list

GNF Atlas 2 - GNF Expression Atlas 2 Data from U133A and GNF1H Chips

Note: the values here range from about -5.0 to 5.0.
These are calculated as logBase2(tissue/reference).

Tissue	Minimum	Maximum
fetal brain		
whole brain		
amygdala		
thymus		
bone marrow		
PB-CD4+ Tcells		
skin		
adipocyte		
pancreatic islets		
heart		
lung		
kidney		
liver		
ovary		
testis		

In silico PCR

UCSC In-Silico PCR

Genome: Human | Assembly: Mar. 2006 (NCBI36/hg18) | Target: genome assembly | Forward Primer: | Reverse Primer: | submit

Max Product Size: 4000 | Min Perfect Match: 15 | Min Good Match: 15 | Flip Reverse Primer:

About In-Silico PCR

In-Silico PCR searches a sequence database with a pair of PCR primers, using an indexing strategy for fast performance.

Configuration Options

Genome and Assembly - The sequence database to search.
Target - If available, choose to query transcribed sequences.
Forward Primer - Must be at least 15 bases in length.
Reverse Primer - On the opposite strand from the forward primer. Minimum length of 15 bases.
Max Product Size - Maximum size of amplified region.
Min Perfect Match - Number of bases that match exactly on 3' end of primers. Minimum match size is 15.
Min Good Match - Number of bases on 3' end of primers where at least 2 out of 3 bases match.
Flip Reverse Primer - Invert the sequence order of the reverse primer and complement it.

Output

When successful, the search returns a sequence output file in fasta format containing all sequence in the database that lie between and include the primer pair. The fasta header describes the region in the database and the primers. The fasta body is capitalized in areas where the primer sequence matches the database sequence and in lower-case elsewhere. Here is an example from human:

```
>chr22:31000551+31001000  TAACAGATTGATGATGCATGAAATGGG CCCATGAGTGGCTCCTAAAGCAGCTGC
TtACAGATTGATGATGCATGAAATGGGgggtggccaggggtgggggtga
gactgcagagaaaggcagggctggttcataacaagctttgtgcgtcccaa
tatgacagctgaagttttccaggggctgatggtgagccagtgagggtaag
tacacagaacatcctagagaaaccctcattccttaaagattaaaaataaa
gacttgctgtctgtaagggattggattatcctatttgagaaattctgtta
tccagaatggcttaccccacaatgctgaaaagtgtgtaccgtaatctcaa
agcaagctcctcctcagacagagaaacaccagccgtcacaggaagcaaag
aaattggcttcacttttaaggtgaatccagaacccagatgtcagagctcc
aagcactttgctctcagctccacGCAGCTGCTTTAGGAGCCACTCATGaG
```

The + between the coordinates in the fasta header indicates this is on the positive strand.

Author

In-Silico PCR was written by Jim Kent. Interactive use on this web server is free to all. Sources and executables to run batch jobs on your own server are available free for academic, personal, and non-profit purposes. Non-exclusive commercial licenses are also available. Contact Jim for details.

In silico PCR usage

- Select genome
- Genomic or transcript?
- Enter primers
- Set configuration options

About In-Silico PCR

In-Silico PCR searches a sequence database with a pair of PCR primers, using an indexing strategy for fast performance.

Configuration Options

Genome and Assembly - The sequence database to search.
Target - If available, choose to query transcribed sequences.
Forward Primer - Must be at least 15 bases in length.
Reverse Primer - On the opposite strand from the forward primer. Minimum length of 15 bases.
Max Product Size - Maximum size of amplified region.
Min Perfect Match - Number of bases that match exactly on 3' end of primers. Minimum match size is 15.
Min Good Match - Number of bases on 3' end of primers where at least 2 out of 3 bases match.
Flip Reverse Primer - Invert the sequence order of the reverse primer and complement it.

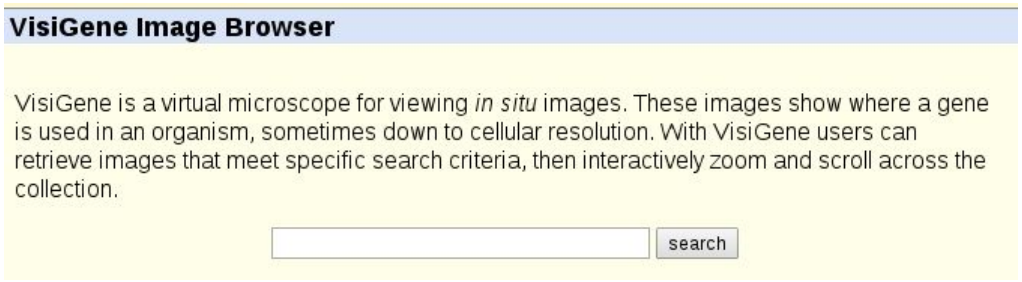
Output

When successful, the search returns a sequence output file in fasta format containing all sequence in the database that lie between and include the primer pair. The fasta header describes the region in the database and the primers. The fasta body is capitalized in areas where the primer sequence matches the database sequence and in lower-case elsewhere. Here is an example from human:

```
>chr22:31000551+31001000  TAACAGATTGATGATGCATGAAATGGG CCCATGAGTGGCTCCTAAAGCAGCTGC
TtACAGATTGATGATGCATGAAATGGGgggtggccaggggtgggggggtga
gactgcagagaaaggcagggctggttcataacaagctttgtgcgtcccaa
tatgacagctgaagttttccaggggctgatggtgagccagtgagggtaag
tacacagaacatcctagagaaaccctcattccttaaagattaaaaataaa
gacttgctgtctgtaagggattggattatcctatttgagaaattctgtta
tccagaatggcttaccccacaatgctgaaaagtgtgtaccgtaatctcaa
agcaagctcctcctcagacagagaaacaccagccgtcacaggaagcaaag
aaattggcttcacttttaaggtgaatccagaacccagatgtcagagctcc
aagcactttgctctcagctccacGCAGCTGCTTTAGGAGCCACTCATGaG
```

The + between the coordinates in the fasta header indicates this is on the positive strand.

Visigene



Images Available

The following image collections are currently available for browsing:

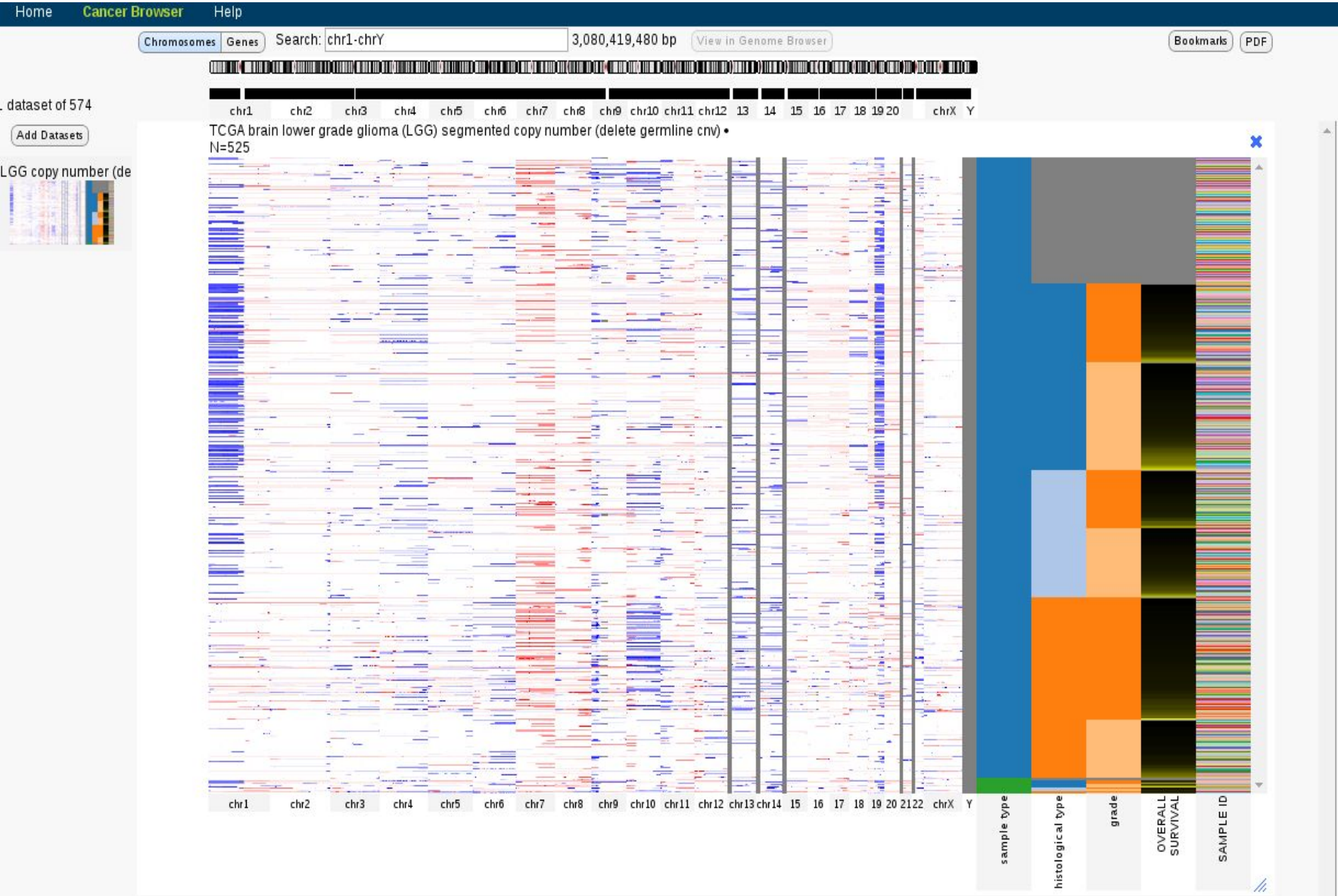
- High-quality high-resolution images of eight-week-old male mouse sagittal brain slices with reverse-complemented mRNA hybridization probes from the Allen Brain Atlas, courtesy of the Allen Institute for Brain Science
- Mouse *in situ* images from the Jackson Lab Gene Expression Database (GXD) at MGI
- Transcription factors in mouse embryos from the Mahoney Center for Neuro-Oncology
- Mouse head and brain *in situ* images from NCBI's Gene Expression Nervous System Atlas (GENSAT) database
- *Xenopus laevis in situ* images from the National Institute for Basic Biology (NIBB) XDB project

ALLEN INSTITUTE for BRAIN SCIENCE
Fueling Discovery

MGI
CELEBRATING 25 YEARS

GENSAT
GENE EXPRESSION NERVOUS SYSTEM ATLAS

Cancer Browser



Encode

Encyclopedia of DNA Elements at UCSC 2003 - 2012

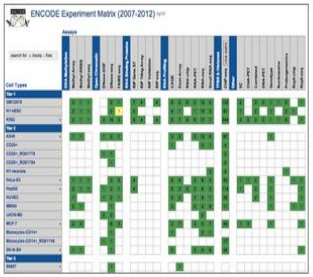
About

The Encyclopedia of DNA Elements (ENCODE) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

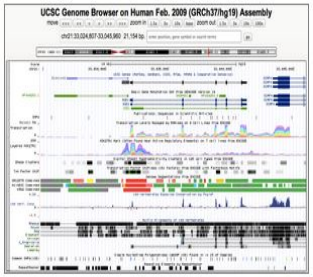
UCSC coordinated data for the ENCODE Consortium from its inception in 2003 (Pilot phase) to the end of the first 5 year phase of whole-genome data production in 2012. All data produced by ENCODE investigators and the results of ENCODE analysis projects from this period are hosted in the UCSC Genome browser and database. Explore ENCODE data using the image links below or via the left menu bar. ***All ENCODE data at UCSC are freely available for download and analysis***.

ENCODE results from 2013 and later are available from the ENCODE Project Portal, encodeproject.org. The ENCODE Project Portal also hosts ENCODE data from the first production phase, additional ENCODE access tools, and ENCODE project pages including up-to-date information about data releases, publications, and upcoming tutorials.

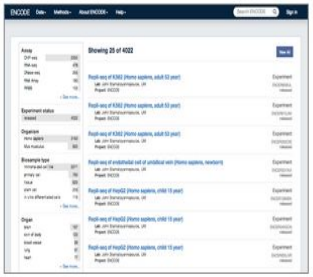
Explore ENCODE data at UCSC



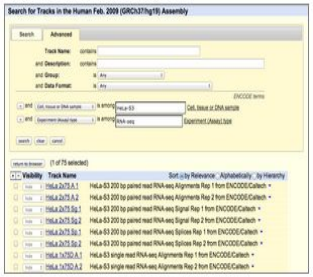
View ENCODE data in the UCSC Genome Browser



Search for data at the ENCODE Portal



Search for ENCODE tracks in the UCSC Browser



Other utilities

UCSC Genome Bioinformatics

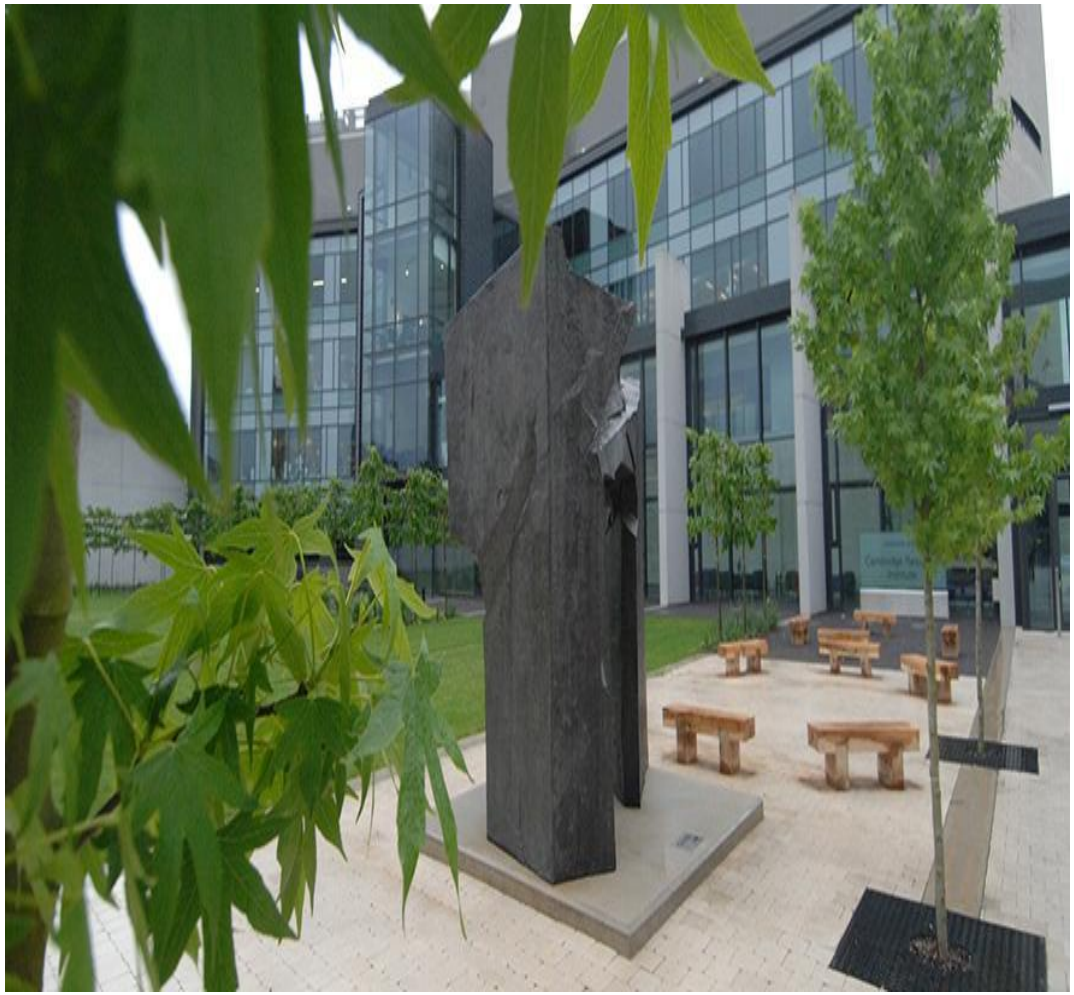
Home · Genomes · Blat · Tables · Gene Sorter · PCR · Session · FAQ · Help

UCSC Genome Browser Utilities

This page contains links to tools and utilities created by the UCSC Genome Bioinformatics Group.

- Batch Coordinate Conversion (liftOver) - converts genome coordinates and genome annotation files between assemblies. The current version supports both forward and reverse conversions, as well as conversions between selected species.
- DNA Duster - removes formatting characters and other non-sequence-related characters from an input sequence. Offers several configuration options for the output format, including translated protein.
- Protein Duster - removes formatting characters and other non-sequence-related characters from an input sequence. Offers several configuration options for the output format.
- Phylogenetic Tree Gif Maker - creates a gif image from the phylogenetic tree specification given. Offers several configuration options for branch lengths, normalized lengths, branch labels, legend etc.
- Executable and Source Code Downloads - executable and source code downloads of the Genome Browser, Blat and liftOver.

Acknowledgements



Some slides were modified from UCSC and OpenHelix course material.