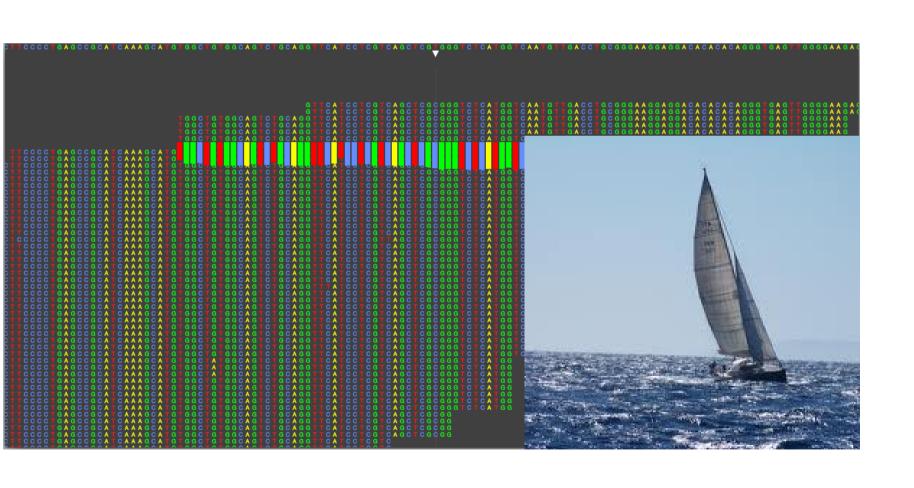
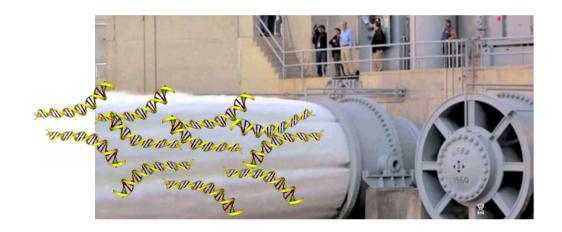
HTS data visualization in genome browsers: the UCSC Genome Browser



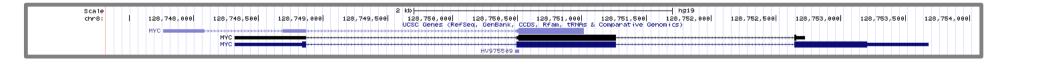
What is a Genome Browser

 The Human Genome Project (2001) marks the beginning of a massive production of sequencing and genomic annotation data.

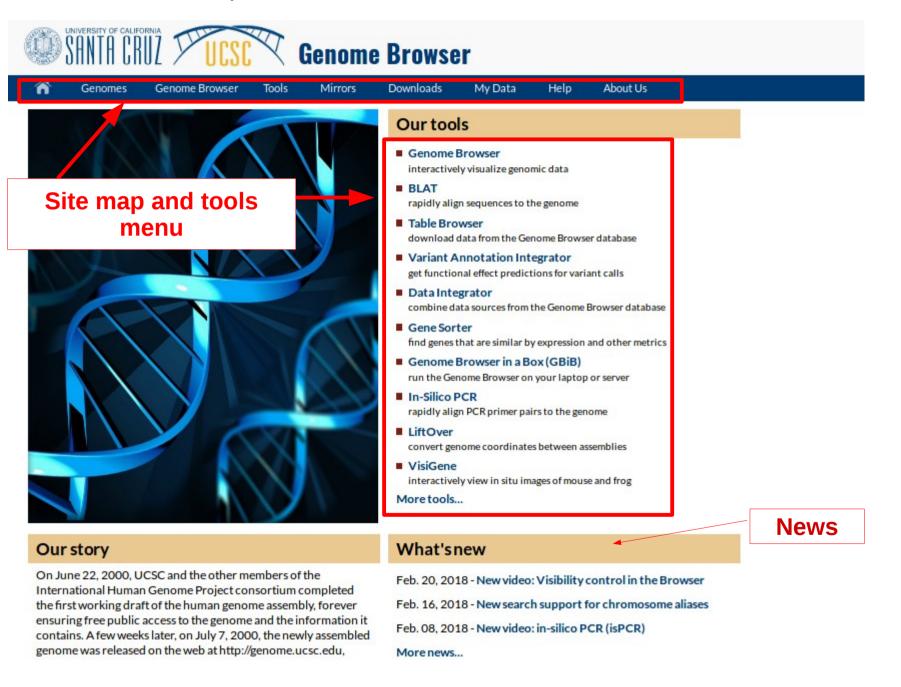


For the first time there is a need to develop *ad hoc* tools to organize and manage this huge data amount → the first genome browsers appear:

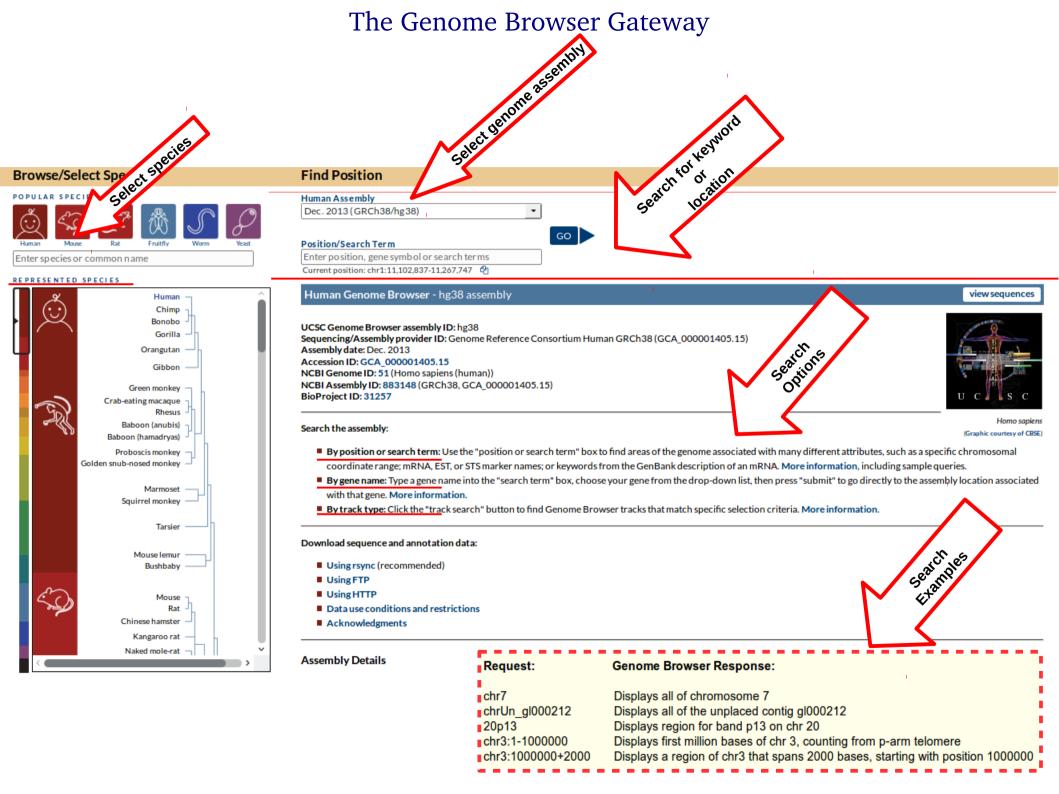
MapViewer (http://www.ncbi.nlm.nih.gov/mapview/), l'**Ensembl Genome Browser** (http://www.ensembl.org/) and the **UCSC Genome Browser** (http://genome.ucsc.edu/)



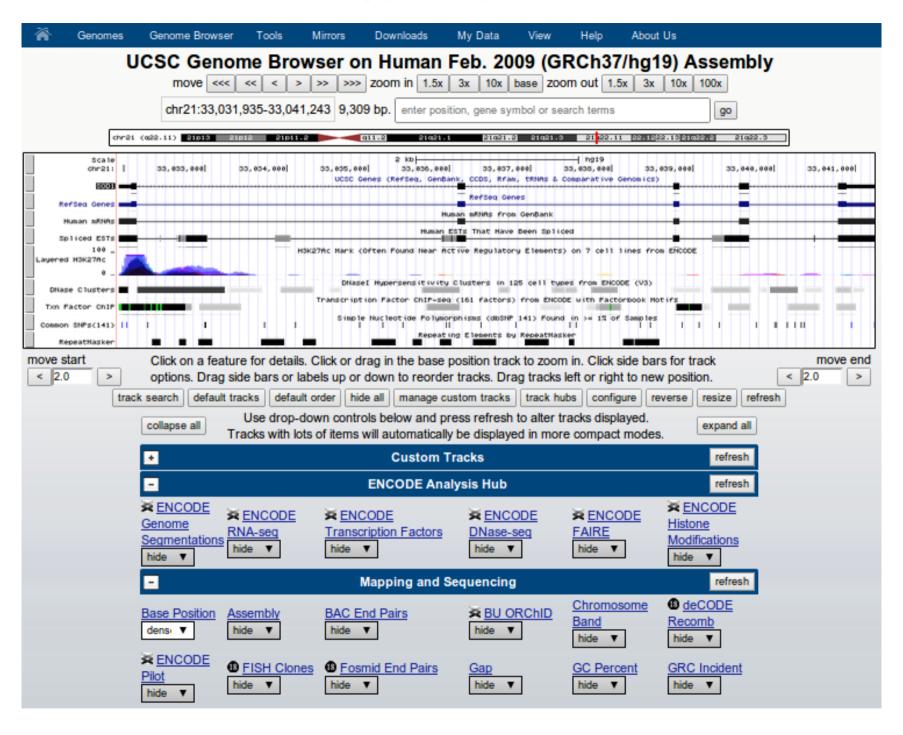
University of California Santa Cruz (UCSC) Genome Browser



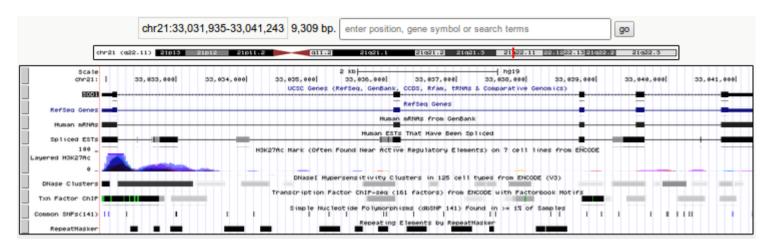
http://genome.ucsc.edu/



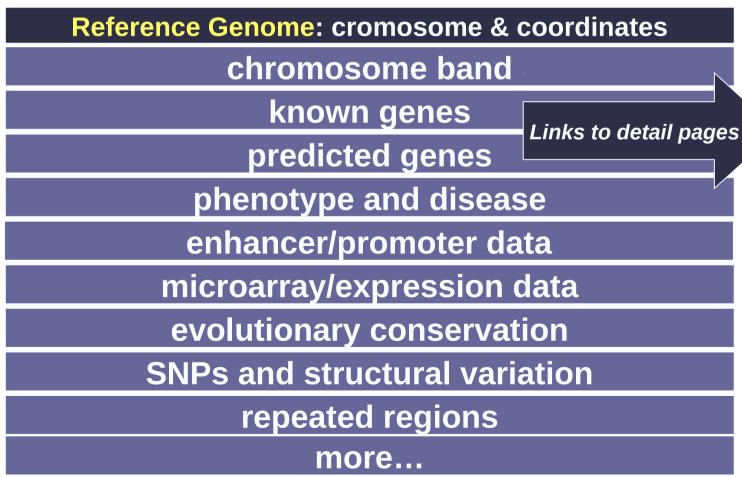
Search result

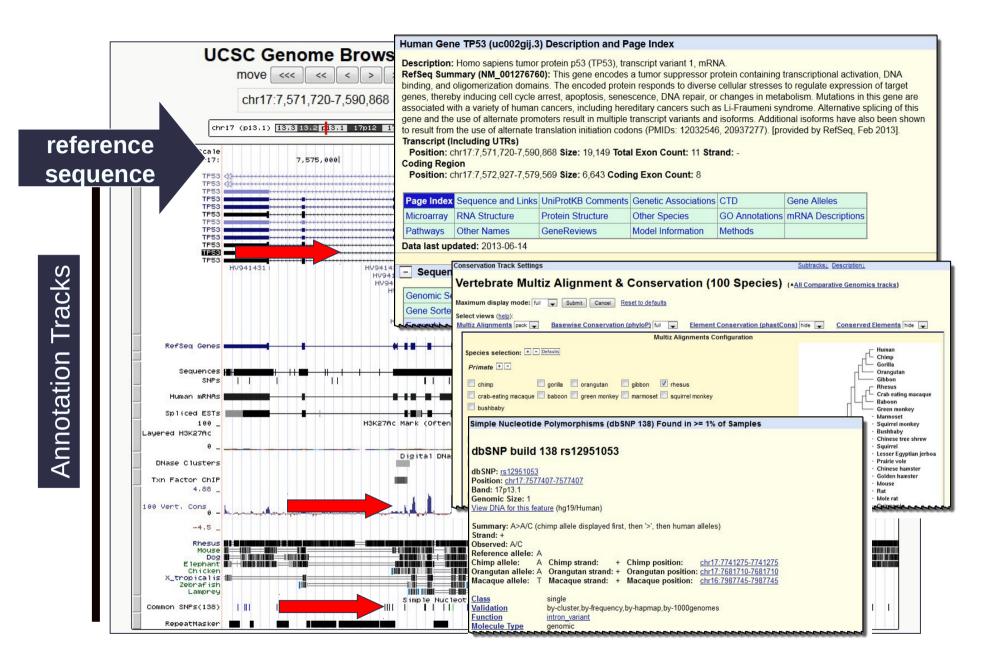


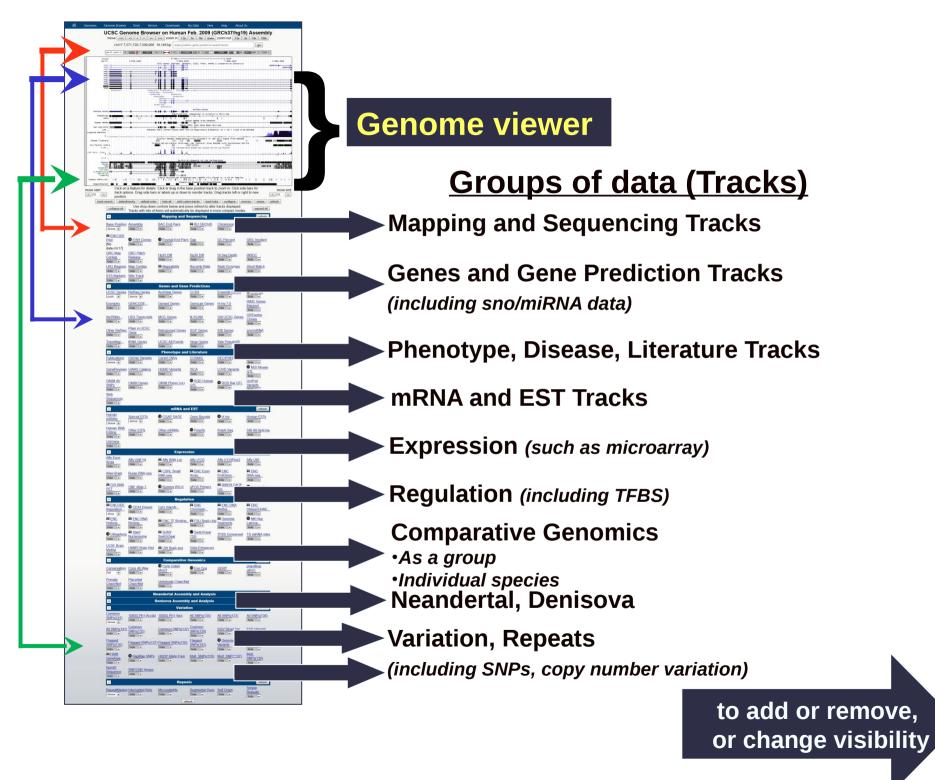
Search results: page structure



Annotation Tracks

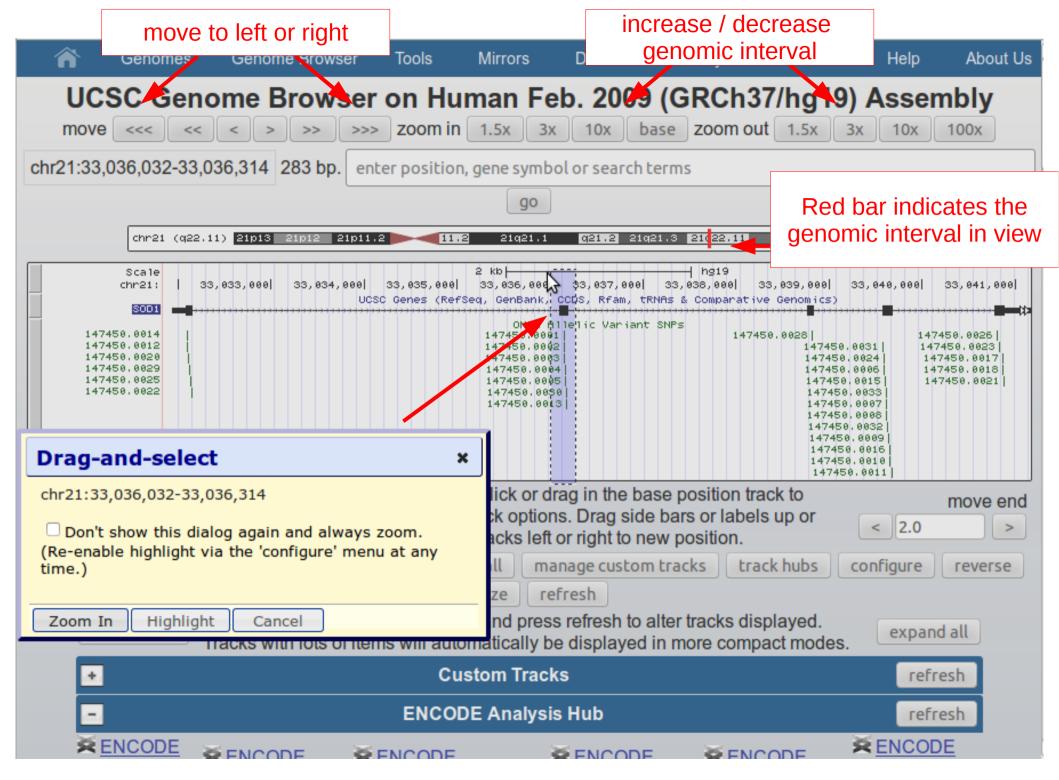




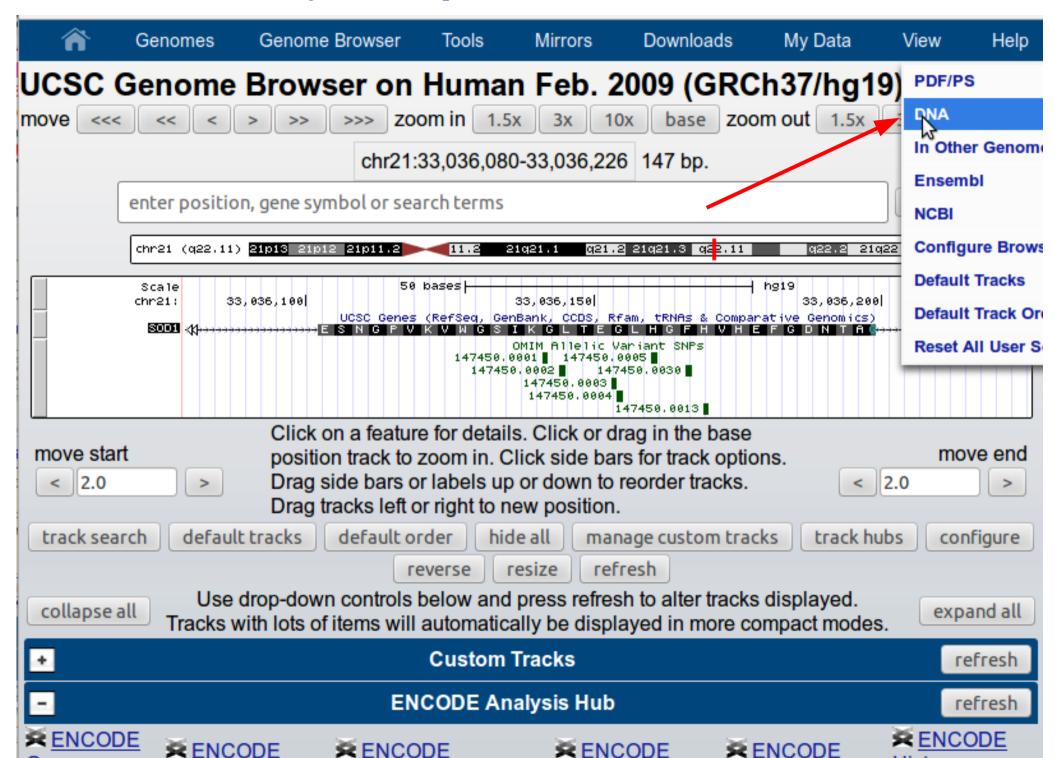


hide dense squish pack full

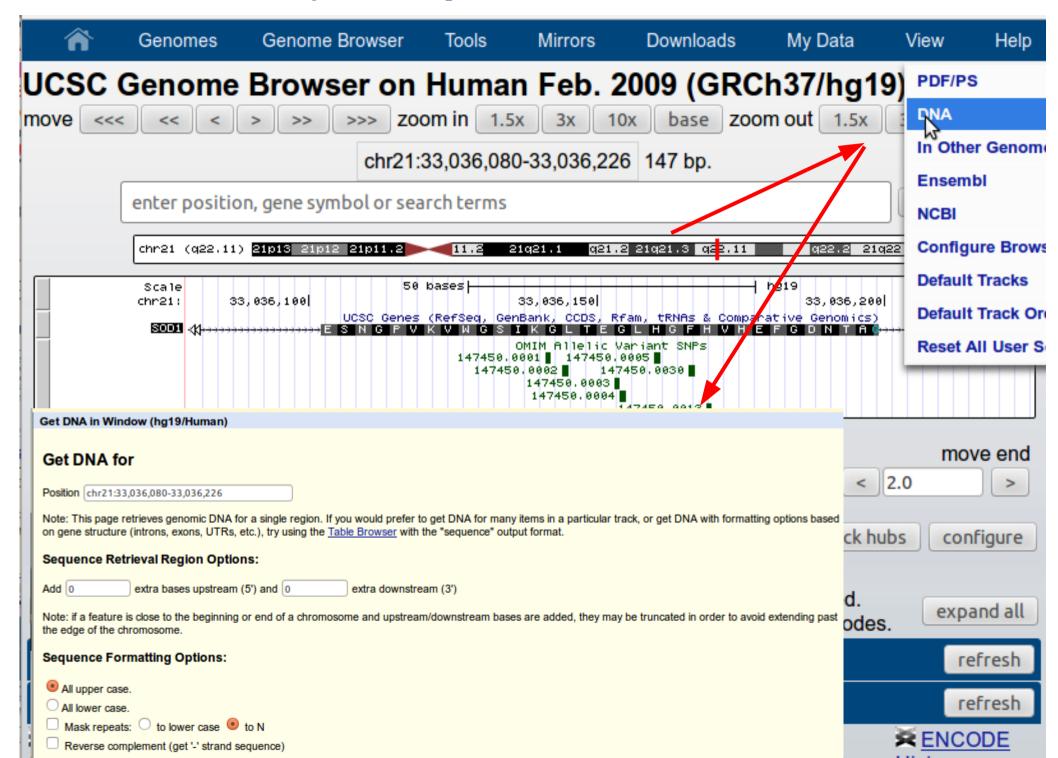
dense V



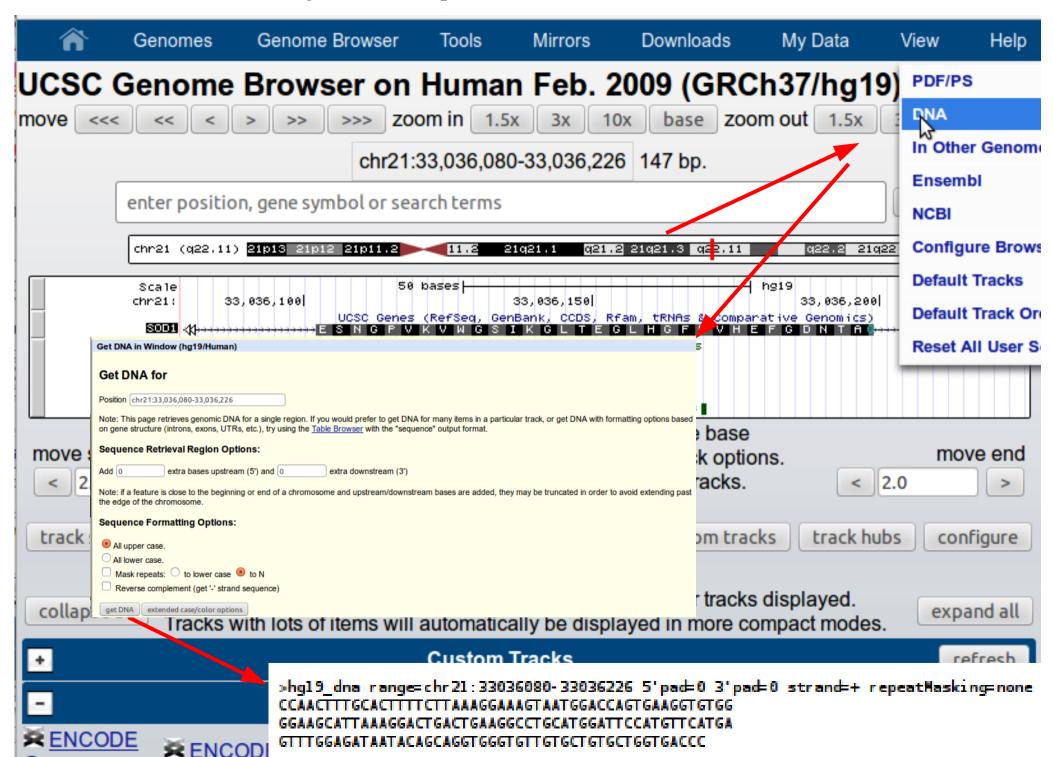
Extract a genomic sequence of interest ('View' → 'DNA')



Extract a genomic sequence of interest ('View' → 'DNA')



Extract a genomic sequence of interest ('View' → 'DNA')



Mapping of a sequence (Tool → 'Blat') Genome Browser Genomes Mirrors **Downloads** My Data About Us Tools: View. Help 2009 (GRCh37/hg19) Assembly UCSC Genome Brows Blat Tal. Browser base zoom out 1.5x 3x 10x move <<< << Variant Annotation Integrator chr8:128.747.765-128.751.01 he symbol or search terms go Gene Sorter chr6 (q24.21) 23.1 600 Genome Graphs 19019 古とみりか In-Silico PCR ches: 128,748,500 749.594 128,750,000 128,750,500 128,751,000 el dati in formato BED LiftOver. Rfam. TRNAS & Comparative Genomics) VisiGene HUSCHER & .. Other Utilities ic Variant stess 190000,00001 190000,0001 190050.0004 194404, 44421 Alignments of Affymetrix Consensus/Exemplans from MG-U133 Flus 2.8 238381_x_at / 238386_X_At di 会会会431_m_at 1 244959_AT DESCRIPTION Repeating Elements by RepeatMasker 23565 LINE LT的 Distant \$ 6500 50 Low Complexity Matellitte. Click on a feature for details. Click or drag in the base position track to zoom in. Click move start move end side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag < 2.0 2.0 tracks left or right to new position. default tracks default order hide all manage custom tracks track hubs configure track search refresh reverse resize Use drop-down controls below and press refresh to alter tracks displayed.

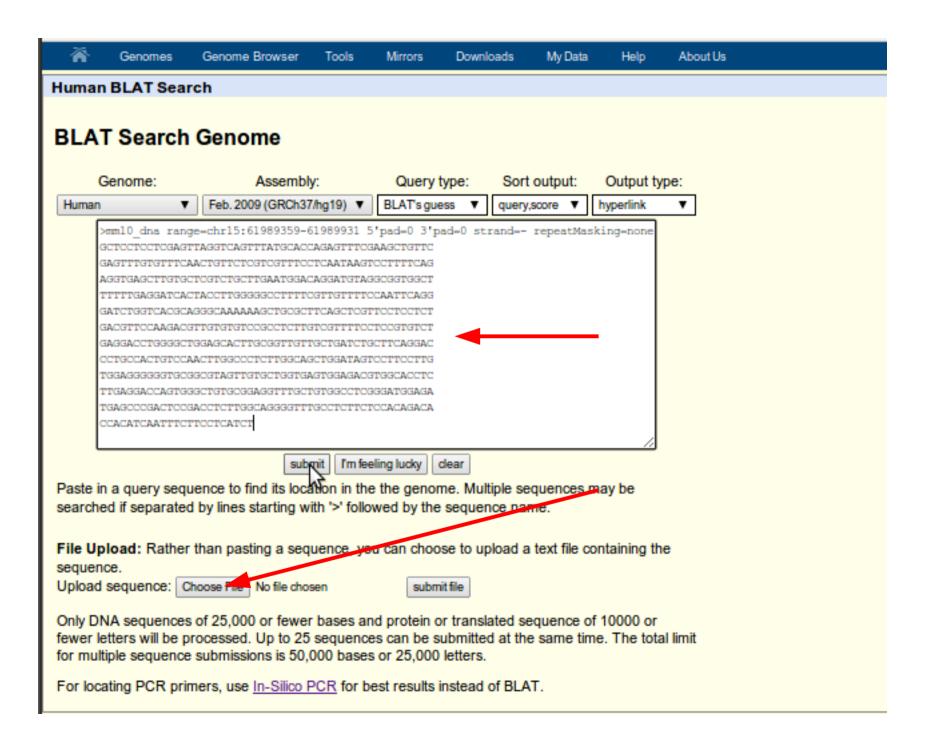
Tracks with lots of items will automatically be displayed in more compact modes.

BLAT = BLAST-like Alignment Tool

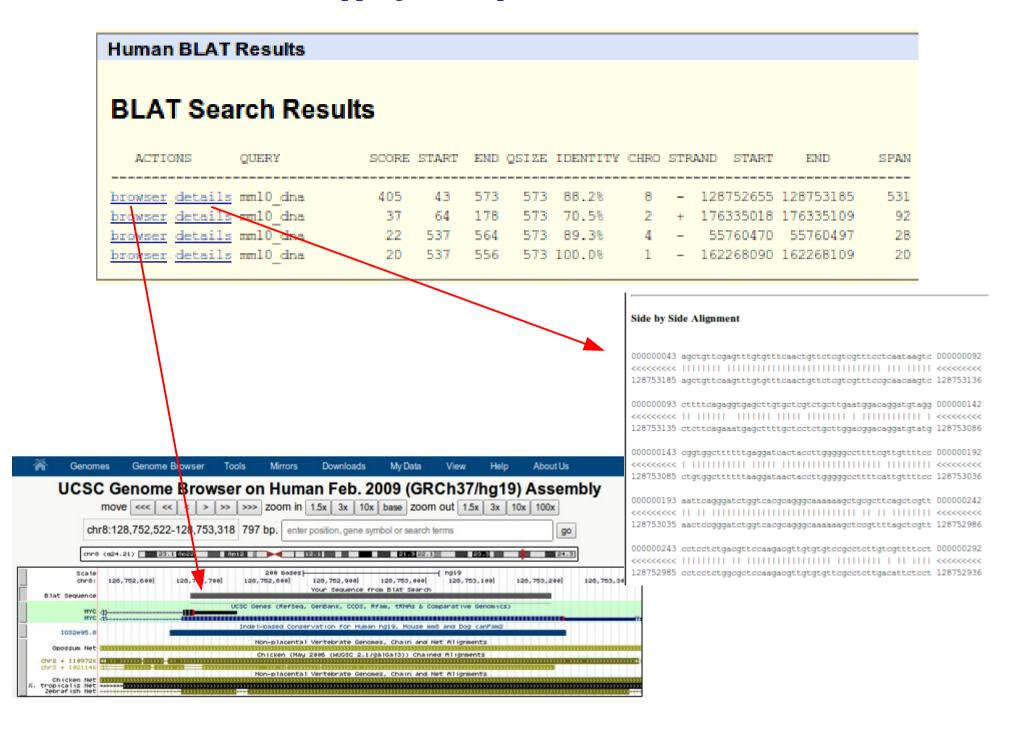
collapse all

expand all

Mapping of a sequence (Tool → 'Blat')



Mapping of a sequence (Tool → 'Blat')



Viewing our data in the genome browser: The BED format

The BED (Browser Extensible Data) format is a tab-delimited text format where each line specifies a genomic interval.

The first three BED fields (mandatory) are:

chrom - The name of the chromosome (e.g. chr3, chrY, chr2 random).

chromStart - The starting position of the feature in the chromosome (0-based).

chromEnd - The ending position of the feature in the chromosome. The chromEnd base is not included in the display of the feature. For example, the first 100 bases of a chromosome are defined as chromStart=0, chromEnd=100, and span the bases numbered 0-99.

The above can be followed by up to 9 additional (optional) BED fields, of which the most commonly used are:

name - Defines the name of the BED line.

SCORE - A score between 0 and 1000.

strand - Defines the strand - either '+' or '-'.

Finally, a BED file can also include an header line describing its visual settings in the UCSC Genome Browser

Example of BED format (including header line):

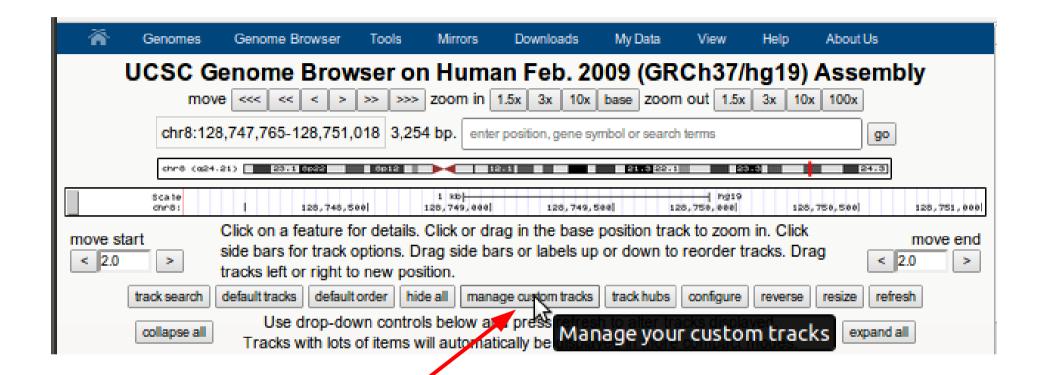
track name='my track' description='my track description' color=255,0,0 visibility=2

chr8 128867449 128867455 geneA 99 +

chr8 128902915 128902921 geneB 1000 +

chr8 129001512 129001518 geneC 0 +

Viewing our data in the Genome Browser ('add custom track')



Querying the UCSC Genome Browser

#chrom

chr17

chromStart

38451241

chromEnd

38451242

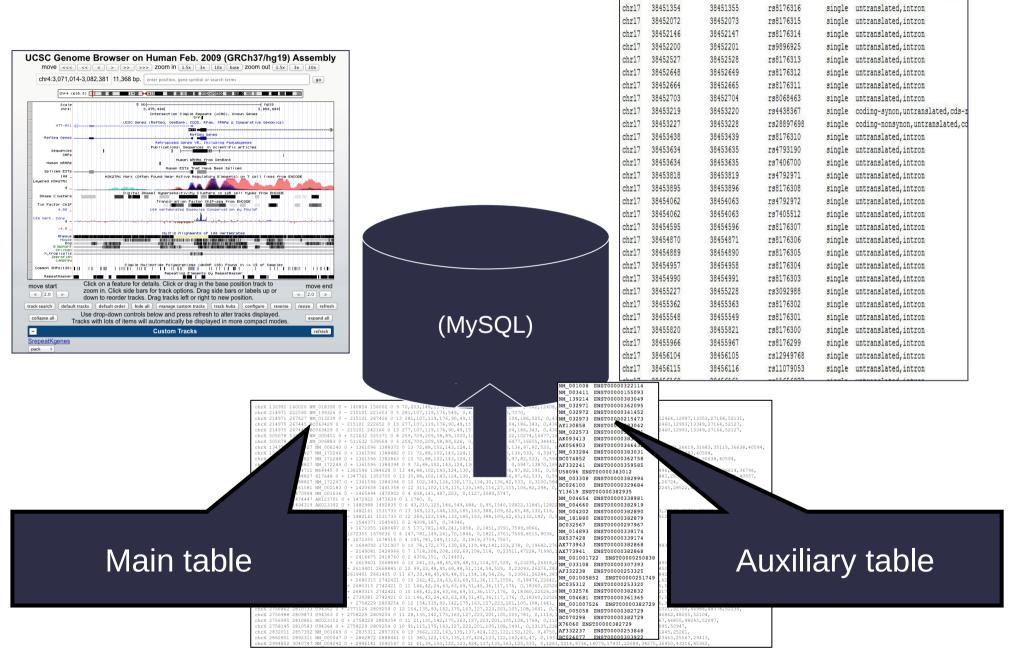
class

func

single coding-nonsynon, untranslated, co

name

rs28897699



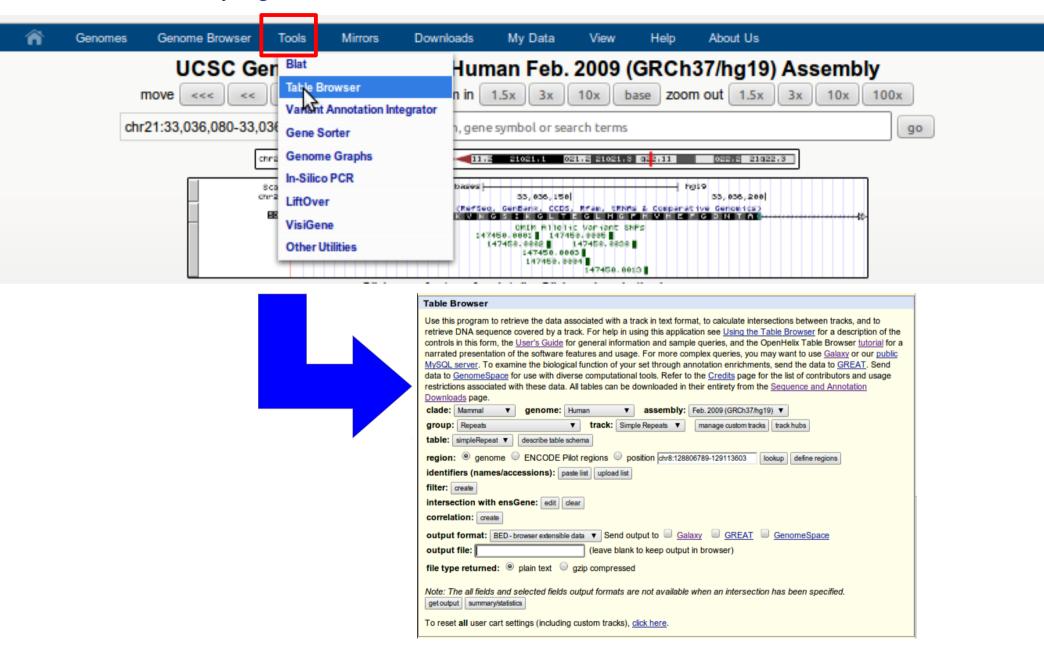
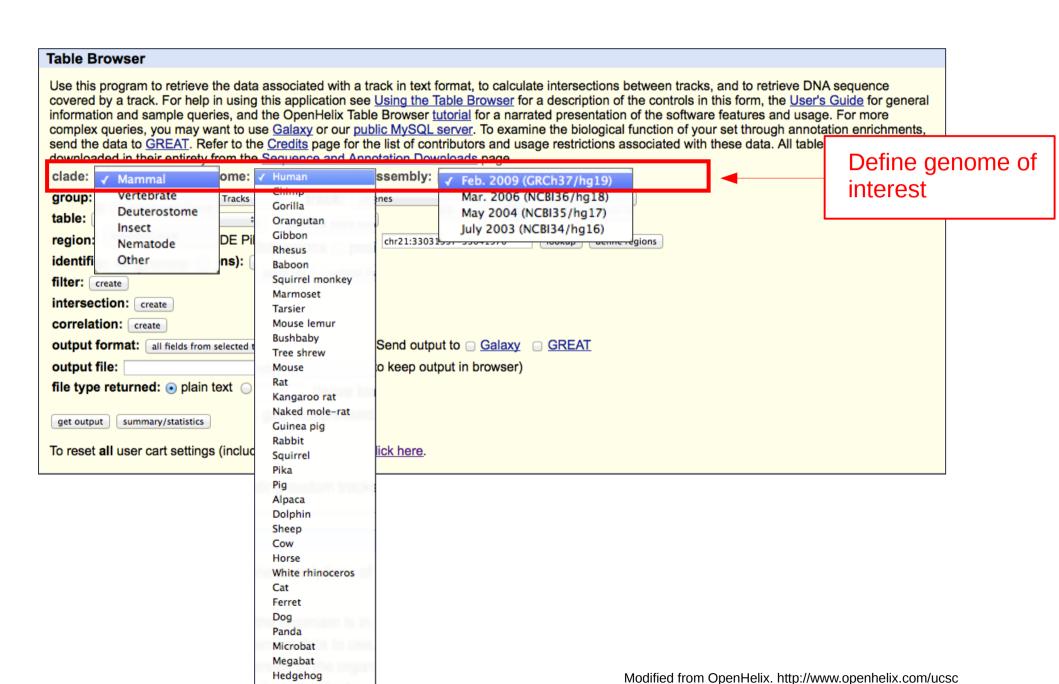
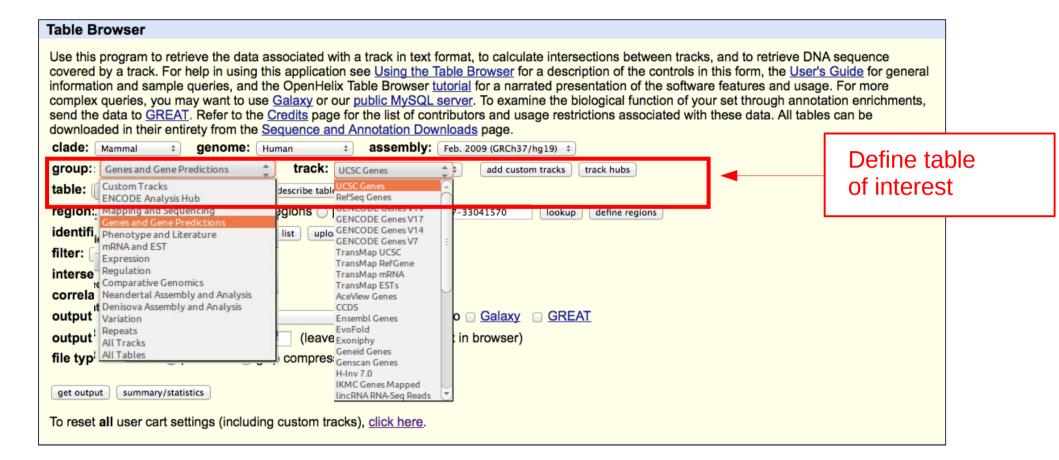
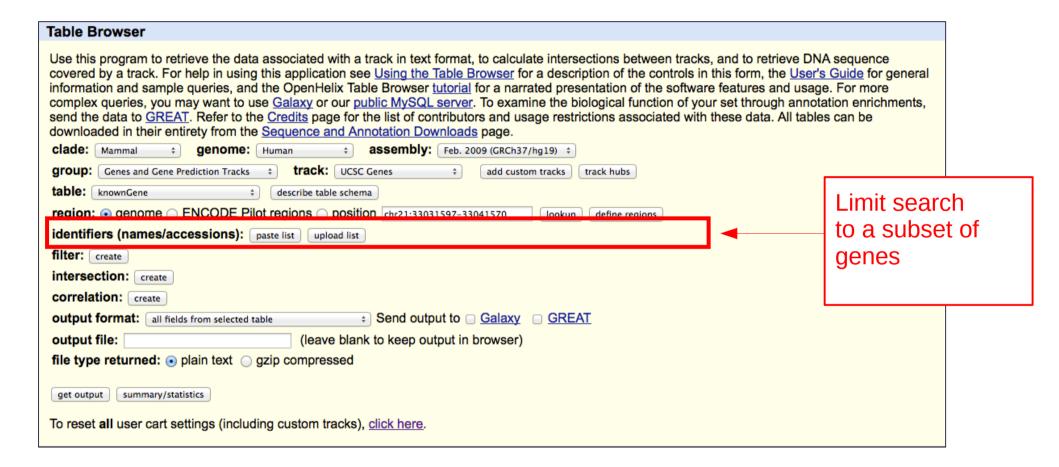


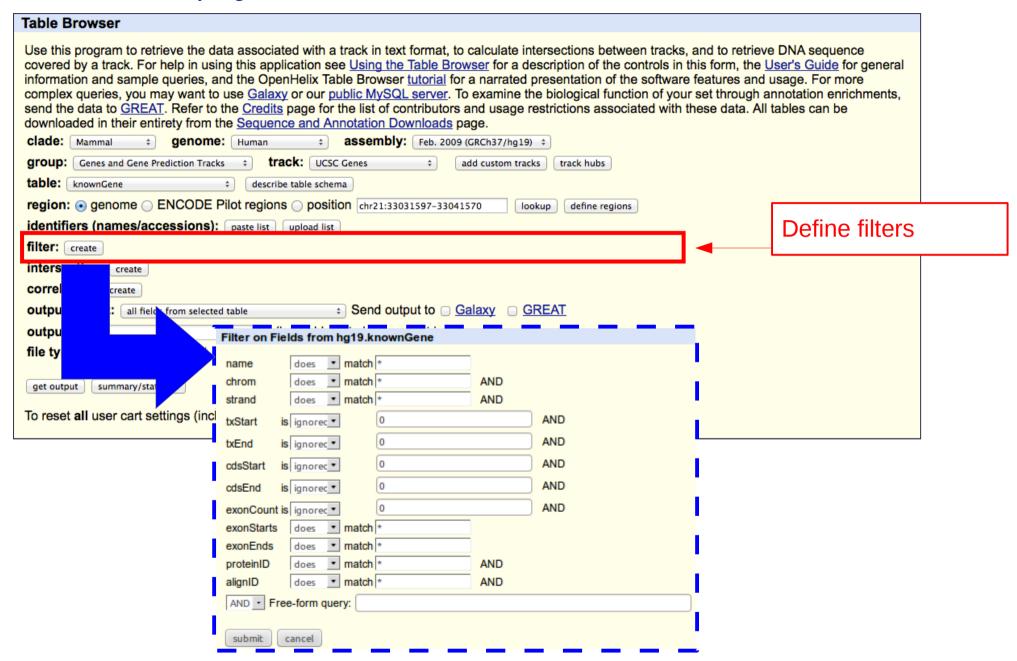
Table Browser Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see Using the Table Browser for a description of the controls in this form, the User's Guide for general information and sample gueries, and the OpenHelix Table Browser tutorial for a narrated presentation of the software features and usage. For more complex gueries, you may want to use Galaxy or our public MySQL server. To examine the biological function of your set through annotation enrichments, send the data to GREAT. Send data to GenomeSpace for use with diverse computational tools. Refer to the Credits page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the Sequence and Annotation Downloads page. clade: Mammal genome: Human assembly: Feb. 2009 (GRCh37/hg19) ▼ track: Simple Repeats ▼ group: Repeats manage custom tracks | track hubs table: simpleRepeat ▼ describe table schema region: genome ENCODE Pilot regions position chr8:128806789-129113603 define regions identifiers (names/accessions): paste list upload list filter: create intersection with ensGene: edit | dear correlation: create output format: BED-browser extensible data ▼ Send output to □ Galaxy □ GREAT □ GenomeSpace output file: (leave blank to keep output in browser) file type returned: plain text gzip compressed Note: The all fields and selected fields output formats are not available when an intersection has been specified. get output | summary/statistics To reset all user cart settings (including custom tracks), click here.















Same item, different names in different databases

Ensembl Gene ID	Associated Gene Name	HGNC symbol
ENSG00000108958	AC016292.3	
ENSG00000123009	NME2P1	NME2P1
ENSG00000124399	NDUFB4P12	NDUFB4P12
ENSG00000126005	MMP24-AS1	MMP24-AS1
ENSG00000128692	EIF2S2P4	EIF2S2P4
ENSG00000144158	RP11-395L14.17	
ENSG00000146677	AC004453.8	
ENSG00000153363	LINC00467	LINC00467
ENSG00000173727	CMB9-22P13.1	
ENSG00000174977	AC026271.5	
ENSG00000175701	LINC00116	LINC00116
ENSG00000178660	ARMC10P1	ARMC10P1
ENSG00000179082	<u>C9orf106</u>	C9orf106
ENSG00000179799	OR7E22P	OR7E22P
ENSG00000179967	PPP1R14BP3	PPP1R14BP3
ENSG00000180574	EIF2S3L	
ENSG00000181126	HLA-V	HLA-V
ENSG00000181524	RPL24P4	RPL24P4
ENSG00000185275	CD24P4	CD24P4

\rightarrow	A	В	С	D	
1	Acc	TGM 1h Notx	TGM 1h Kdo2	Log2 Fold	
2	NM_010510	5.675	13.865	8.19	
3	NM_008361	7.865	16.015	8.15	
4	NM_008176	8.47	15.975	7.505	
5	NM_009140	6.785	14.055	7.27	
6	NM_013693	9.775	16.77	6.995	
7	AK152177	9.09	16.06	6.97	
8	L38281	9.1	15.94	6.84	
9	NM_009404	6.085	12.765	6.68	
10	NM_030612	7.95	14.485	6.535	
11	NM_010554	7.715	13.98	6.265	
12	NM_013652	11.255	17.215	5.96	
13	NM_011337	10.305	15.65	5.345	
14	NM_138648	7.26	12.26	5	
15	NM_021274	8.99	13.93	4.94	
16	NM_008331	7.775	12.415	4.64	
17	NM_010907	12.6	17.17	4.57	
18	NM_010276	6.29	10.725	4.435	
19	NM_007707	11.895	16.23	4.335	
- Collection					

