

UCSC Genome Browser Table Browser

The screenshot shows the UCSC Genome Browser homepage with a blue navigation bar at the top. The menu items include Home, Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. A dropdown menu for 'Tools' is open, showing options like Blat, Table Browser, Variant Annotation Integrator, Data Integrator, Gene Sorter, Genome Graphs, In-Silico PCR, LiftOver, VisiGene, and Other Utilities. The 'Table Browser' option is highlighted. The main content area features a large blue DNA helix graphic on the left and a yellow sidebar on the right titled 'tools' with descriptions for various tools.

Blat

Table Browser

Variant Annotation Integrator

Data Integrator

Gene Sorter

Genome Graphs

In-Silico PCR

LiftOver

VisiGene

Other Utilities

tools

ome Browser

actively visualize genomic data

T

lly align sequences to the genome

le Browser

ownload data from the Genome Browser database

ariant Annotation Integrator

unctional effect predictions for variant calls

a Integrator

bine data sources from the Genome Browser database

le Sorter

find genes that are similar by expression and other metrics

■ **Genome Browser in a Box (GBiB)**

run the Genome Browser on your laptop or server

■ **In-Silico PCR**

rapidly align PCR primer pairs to the genome

■ **LiftOver**

convert genome coordinates between assemblies

■ **VisiGene**

interactively view *in situ* images of mouse and frog

[More tools...](#)

Our story

On June 22, 2000, UCSC and the other members of the International Human Genome Project consortium completed the first working draft of the human genome assembly, forever changing the way we think about our genetic information it

<https://genome.ucsc.edu/cgi-bin/hgTables>

What's new

Nov. 07, 2016 - **New CRISPR track for many assemblies**

Nov. 03, 2016 - **New chromosome aliases search support**

Oct. 17, 2016 - **UCSC Genome Browser 2017 release**

Select your genome

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Help](#)[About Us](#)

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 queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, 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

group: Repeats

table: chainSelf [describe table](#)

region: genome position

identifiers (names/accessions)

filter: [create](#)

intersection: [create](#)

output format: all fields from selected table

output file:

file type returned: plain text

[get output](#) [summary/statistics](#)

To reset **all** user cart settings (including this one)

Using the Table Browser

This section provides brief line-by-line descriptions of the controls in the Table Browser.

- **clade:** Specifies which clade to search. The default is Mammal.
- **genome:** Specifies which organism's genome sequence to use. The default is Human.
- **assembly:** Specifies which genome assembly to use. The default is Dec. 2013 (GRCh38/hg38).
- **group:** Selects the type of tracks to display. The default is Repeats. Other options include All Tracks, All Tables, and All Annotations.
- **database:** (with "All Tables") Selects which database should be used for options in table menu.
- **track:** Selects the annotation tracks to display. This list displays all tracks belonging to the group specified in the group list. Some tracks are not available when

assembly: Dec. 2013 (GRCh38/hg38)

genome: Human

Mouse

Alpaca

Armadillo

Baboon

Bonobo

Bushbaby

Cat

Chimp

Chinese hamster

Cow

Crab-eating macaque

Dog

Dolphin

Elephant

Ferret

Gibbon

Gorilla

Green monkey

Guinea pig

Hedgehog

Horse

Kangaroo rat

Malayan flying lemur

Manatee

Marmoset

Megabat

Microbat

Minke whale

Mouse lemur

Naked mole-rat

Opossum

Orangutan

Panda

Pig

Pika

add custom tracks

track hubs

lookup

define regions

Send output to Galaxy GREAT GenomeSpace
(keep output in browser)

[click here.](#)

Table Browser controls. For more information on using this program, see the [Table Browser User's Guide](#).

m's genome sequence to use.

n the track list. The options correspond to the track groupings shown in the Genome Browser. Select 'All Tracks' for groups. Select 'All Tables' to see all tables including those not associated with a track.

ines which database should be used for options in table menu.

ith. This list displays all tracks belonging to the group specified in the group list. Some tracks are not available when

Select your assembly

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Help](#)[About Us](#)

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 queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, 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: ✓ Dec. 2013 (GRCh38/hg38)
group: Repeats track: Self Chain add
table: chainSelf describe table schema
region: genome position chr6:31995680-31996859 lookup define
identifiers (names/acceessions): paste list upload list
filter: create
intersection: create
output format: all fields from selected table Send output to Galaxy GREAT GenomeSpace
output file: (leave blank to keep output in browser)
file type returned: plain text gzip compressed

[get output](#) [summary/statistics](#)

To reset **all** user cart settings (including custom tracks), [click here](#).

Using the Table Browser

This section provides brief line-by-line descriptions of the Table Browser controls. For more information on using this program, see the [Table Browser User's Guide](#).

- **clade:** Specifies which clade the organism is in.
- **genome:** Specifies which organism data to use.
- **assembly:** Specifies which version of the organism's genome sequence to use.
- **group:** Selects the type of tracks to be displayed in the *track* list. The options correspond to the track groupings shown in the Genome Browser. Select 'All Tracks' for an alphabetical list of all available tracks in all groups. Select 'All Tables' to see all tables including those not associated with a track.
- **database:** (with "All Tables" group option) Determines which database should be used for options in table menu.
- **track:** Selects the annotation track data to work with. This list displays all tracks belonging to the group specified in the *group* list. Some tracks are not available when

Let's get RefSeq Genes

Use this help in OpenH MySQL computer the [Se clade: group table:](#)

Table [Mapping and Sequencing](#) [Genes and Gene Predictions](#)

Phenotype and Literature
mRNA and EST
Expression
Regulation
Comparative Genomics
Neandertal Assembly and Analysis
Denisova Assembly and Analysis
Variation
Repeats

✓ All Tracks
All Tables

region: genome ENCODE Pilot regions position chr21:33031597-33041570 [lookup](#) [define regions](#)

identifiers (names/acccessions): [paste list](#) [upload list](#)

filter: [create](#)

intersection: [create](#)

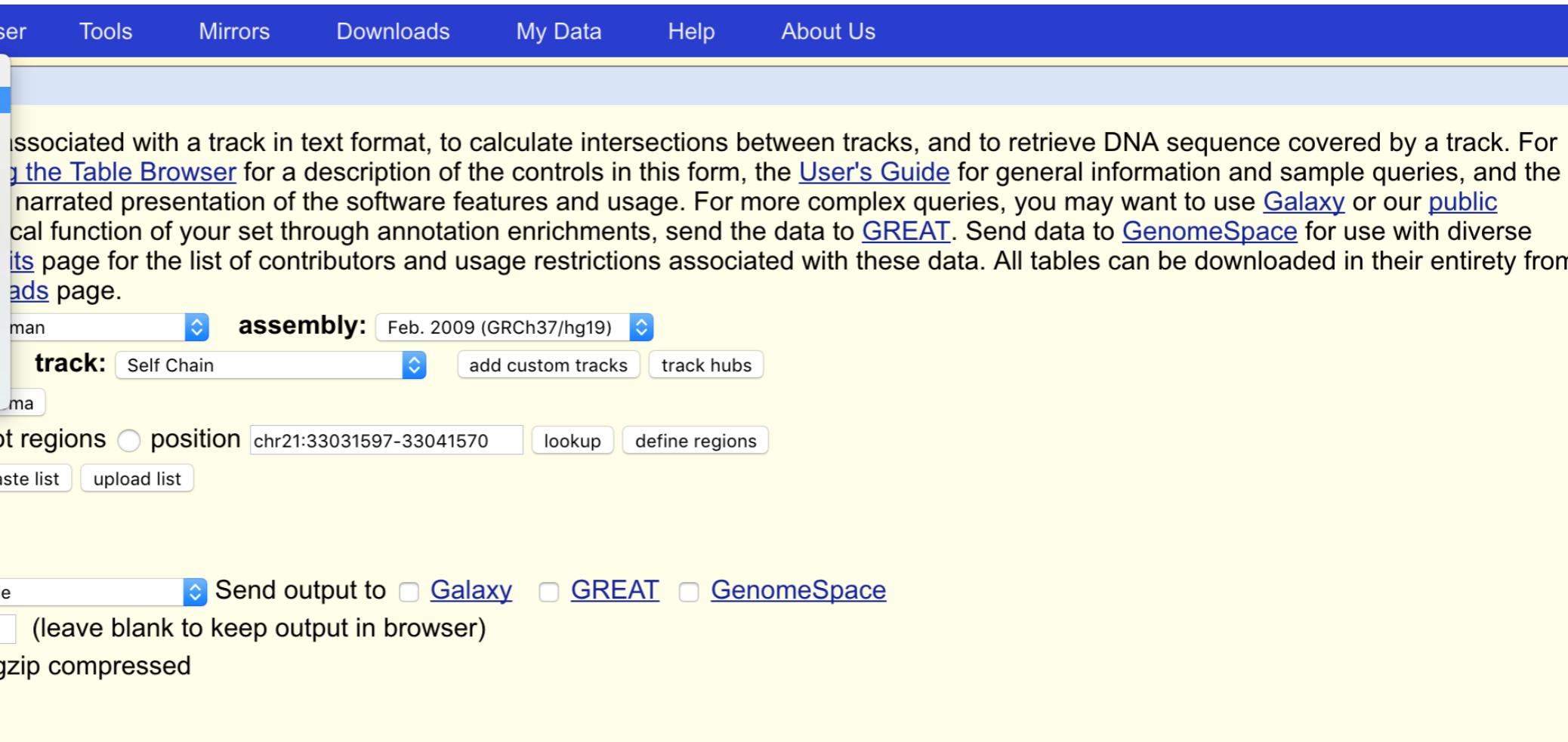
output format: [all fields from selected table](#) Send output to Galaxy GREAT GenomeSpace

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

file type returned: plain text gzip compressed

[get output](#) [summary/statistics](#)

To reset **all** user cart settings (including custom tracks), [click here](#).



Using the Table Browser

This section provides brief line-by-line descriptions of the Table Browser controls. For more information on using this program, see the [Table Browser User's Guide](#).

- **clade:** Specifies which clade the organism is in.
- **genome:** Specifies which organism data to use.
- **assembly:** Specifies which version of the organism's genome sequence to use.
- **group:** Selects the type of tracks to be displayed in the *track* list. The options correspond to the track groupings shown in the Genome Browser. Select 'All Tracks' for an alphabetical list of all available tracks in all groups. Select 'All Tables' to see all tables including those not associated with a track.
- **database:** (with "All Tables" group option) Determines which database should be used for options in table menu.
- **track:** Selects the annotation track data to work with. This list displays all tracks belonging to the group specified in the *group* list. Some tracks are not available when

Now you can select RefSeq genes

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Help](#)[About Us](#)

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 queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, 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)

group: Genes and Gene Predictions track: UCSC Genes

table: knownGene description: RefSeq Genes

region: genome ENCODE Pilot regions

identifiers (names/acccessions):

filter:

intersection:

correlation:

output format: all fields from selected table

output file: (leave empty to get output in browser)

file type returned: plain text gzip compressed

To reset **all** user cart settings (including custom tracks)

✓ UCSC Genes
RefSeq Genes

GENCODE Gene V24lift37
GENCODE Genes V19
GENCODE Genes V17
GENCODE Genes V14
GENCODE Genes V7
TransMap UCSC
TransMap RefGene
TransMap mRNA
TransMap ESTs
AceView Genes
Augustus
CCDS
CRISPR Regions
CRISPR Targets
Ensembl Genes
EvoFold
Exoniphy
Geneid Genes
Genscan Genes
H-Inv 7.0
IKMC Genes Mapped
lincRNA RNA-Seq Reads
lincRNA Transcripts
LRG Transcripts
MGC Genes
N-SCAN
Old UCSC Genes
ORFeome Clones
Other RefSeq
Pfam in UCSC Gene
Retroposed Genes
SGP Genes

-33041570

Galaxy GREAT GenomeSpace

Table Browser)

Using the Table Browser

This section provides brief line-by-line descriptions of the controls.

- **clade:** Specifies which clade the organism belongs to.
- **genome:** Specifies which organism data is used.
- **assembly:** Specifies which version of the genome assembly is used.
- **group:** Selects the type of tracks to be displayed. This dropdown menu lists an alphabetical list of all available tracks.
- **database:** (with "All Tables" group option) Selects the database to use.

These controls. For more information on using this program, see the [Table Browser User's Guide](#).

ence to use.

The options correspond to the track groupings shown in the Genome Browser. Select 'All Tracks' for 'Tables' to see all tables including those not associated with a track. The base should be used for options in table menu.

We want BED format...

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Help](#)[About Us](#)

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 queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, 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)

group: Genes and Gene Predictions track: RefSeq Genes add custom tracks track hubs

table: refGene describe table schema

region: genome ENCODE Pilot regions position chr21:33031597-33041570 lookup define regions

identifiers (names/acceessions): [paste list](#) [upload list](#)

filter: [create](#)

intersection: [create](#)

correlation: [create](#)

output format: all fields from selected table
 selected fields from primary and related tables
 sequence
 GTF - gene transfer format
 CDS FASTA alignment from multiple alignment
 BED - browser extensible data
 custom track
 hyperlinks to Genome Browser

output file:

file type return:

[get output](#) [summary](#)

To reset all user cart settings (including custom tracks), [click here](#).

Let's get a file, “refseq.bed”

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

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 queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, 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)

group: Genes and Gene Predictions **track:** RefSeq Genes [add custom tracks](#) [track hubs](#)

table: refGene [describe table schema](#)

region: genome ENCODE Pilot regions position chr21:33031597-33041570 [lookup](#) [define regions](#)

identifiers (names/acceessions): [paste list](#) [upload list](#)

filter: [create](#)

intersection: [create](#)

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

[get output](#) [summary/statistics](#)

To reset **all** user cart settings (including custom tracks), [click here](#).

Get output

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Help](#)[About Us](#)

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 queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, 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)

group: Genes and Gene Predictions **track:** RefSeq Genes [add custom tracks](#) [track hubs](#)

table: refGene [describe table schema](#)

region: genome ENCODE Pilot regions position chr21:33031597-33041570 [lookup](#) [define regions](#)

identifiers (names/acceessions): [paste list](#) [upload list](#)

filter: [create](#)

intersection: [create](#)

correlation: [create](#)

output format: BED - browser extensible data **Send output to** [Galaxy](#) [GREAT](#) [GenomeSpace](#)

output file: refseq.bed (leave blank to keep output in browser)

file type returned: plain text gzip compressed

[get output](#) [summary/statistics](#)

To reset **all** user cart settings (including custom tracks), [click here](#).

Get BED, we want the “Whole Gene” in this case

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

Output refGene as BED

Include custom track header:

name=
description=
visibility=
url=

Create one BED record per:

Whole Gene
 Upstream by bases
 Exons plus bases at each end
 Introns plus bases at each end
 5' UTR Exons
 Coding Exons
 3' UTR Exons
 Downstream by bases

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

After saving the file wherever you want, run less

```
chr1 66999251 67216822 NM_001308203 0 + 67000041 67208778 0 22 104,123,64,25,57,55,176,25,52,86,93,75,128,127,66,112,156,133,203,65,165,8067,  
0,677,92278,99501,106208,109241,109975,137426,138375,139712,143435,146109,155579,156621,160870,185725,195695,200179,205766,207089,207703,209504,  
chr1 66999638 67216822 NM_032291 0 + 67000041 67208778 0 25 413,64,25,72,57,55,176,12,12,25,52,86,93,75,501,128,127,60,112,156,133,203,65,16  
5,8067, 0,91891,99114,101988,105821,108854,109588,126557,133574,137039,137988,139325,143048,145722,147913,155192,156234,161478,185338,195308,199792,205379,206702,207316,209117,  
chr1 16767166 16786584 NM_001145277 0 + 16767256 16785491 0 7 182,101,105,82,109,178,1248, 0,2960,7198,7388,8421,11166,18170,  
chr1 16767166 16786584 NM_001145278 0 + 16767256 16785385 0 8 104,101,105,82,109,178,76,1248, 0,2960,7198,7388,8421,11166,15146,18170,  
chr1 16767166 16786584 NM_018090 0 + 16767256 16785385 0 8 182,101,105,82,109,178,76,1248, 0,2960,7198,7388,8421,11166,15146,18170,  
chr1 33547778 33567493 NR_126031 0 + 33567493 33567493 0 8 177,174,173,172,166,163,113,60, 0,1776,9872,11067,12370,14529,15889,19655,  
chr1 48998526 50489626 NM_001323575 0 - 48999844 50489468 0 13 1439,97,163,153,112,115,90,40,217,95,125,123,192, 0,6787,54149,57978,10163  
8,120482,130297,334336,512729,712915,1164458,1318541,1490908,  
chr1 48998526 50489626 NM_001323574 0 - 48999844 50489468 0 14 1439,27,97,163,153,112,115,90,40,217,95,161,123,192, 0,2035,6787,54149,57978,  
101638,120482,130297,334336,512729,712915,1164422,1318541,1490908,  
chr1 48998526 50489626 NR_136623 0 - 50489626 50489626 0 13 1439,97,163,153,112,115,90,124,40,217,95,123,192, 0,6787,54149,57978,10163  
8,120482,130297,156925,334336,512729,712915,1318541,1490908,  
chr1 25071759 25170815 NM_013943 0 + 25072044 25167428 0 6 357,110,126,107,182,3552, 0,52473,68825,81741,94591,95504,  
chr1 48998526 50489626 NM_032785 0 - 48999844 50489468 0 14 1439,27,97,163,153,112,115,90,40,217,95,125,123,192, 0,2035,6787,54149,57978,  
101638,120482,130297,334336,512729,712915,1164458,1318541,1490908,  
chr1 33546713 33586132 NM_001293562 0 + 33547850 33585783 0 11 182,118,177,174,173,135,166,163,113,215,488, 0,278,1065,2841,10937,12169,1343  
5,15594,16954,36789,38931,  
chr1 48998526 50489626 NM_001323573 0 - 48999844 50489468 0 13 1439,97,163,153,112,115,90,40,217,95,161,123,192, 0,6787,54149,57978,10163  
8,120482,130297,334336,512729,712915,1164422,1318541,1490908,  
chr1 33546713 33586132 NM_052998 0 + 33547850 33585783 0 12 182,121,212,177,174,173,135,166,163,113,215,488, 0,275,488,1065,2841,1093  
7,12169,13435,15594,16954,36789,38931,  
chr1 8378144 8404227 NM_001080397 0 + 8378168 8404073 0 9 102,421,93,225,728,154,177,206,421, 0,6221,7213,7733,12124,17352,19731,21408,25662,  
chr1 33547778 33567493 NM_001301826 0 + 33547850 33567493 0 8 177,174,173,135,166,163,113,60, 0,1776,9872,11104,12370,14529,15889,19655,  
chr1 33547778 33586132 NM_001301825 0 + 33547850 33585783 0 9 177,174,173,135,166,163,173,215,488, 0,1776,9872,11104,12370,14529,15829,3572  
4,37866,  
chr1 33546729 33586132 NM_001301824 0 + 33557656 33585783 0 8 380,173,135,166,163,113,215,488, 0,10921,12153,13419,15578,16938,36773,38  
915,  
:
```

Now let's get CpG islands!

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Help](#)[About Us](#)

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 queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, 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: Human assembly: Feb. 2009 (GRCh37/hg19)

group: ✓ Genes and Gene Predictions
table: Phenotype and Literature
mRNA and EST
region: Expression
identification: Regulation
Comparative Genomics
Neandertal Assembly and Analysis
Denisova Assembly and Analysis
variation: Variation
repeats: Repeats
output: All Tracks
output: All Tables

Send output to Galaxy GREAT GenomeSpace
(leave blank to keep output in browser)

file type returned: plain text gzip compressed

[get output](#) [summary/statistics](#)

To reset **all** user cart settings (including custom tracks), [click here](#).

The default selection should be CpG islands

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

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 queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, 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)

group: Regulation **track:** CpG Islands [add custom tracks](#) [track hubs](#)

table: cpgIslandExt [describe table schema](#)

region: genome ENCODE Pilot regions position chr21:33031597-33041570 [lookup](#) [define regions](#)

identifiers (names/acceessions): [paste list](#) [upload list](#)

filter: [create](#)

intersection: [create](#)

correlation: [create](#)

output format: BED - browser extensible data **Send output to:** [Galaxy](#) [GREAT](#) [GenomeSpace](#)

output file: refseq.bed (leave blank to keep output in browser)

file type returned: plain text gzip compressed

[get output](#) [summary/statistics](#)

To reset **all** user cart settings (including custom tracks), [click here](#).

Let's call it cpg.bed

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Help](#)[About Us](#)

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 queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, 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)

group: Regulation track: CpG Islands add custom tracks track hubs

table: cpgIslandExt describe table schema

region: genome ENCODE Pilot regions position chr21:33031597-33041570 lookup define regions

identifiers (names/accessions): paste list upload list

filter: create

intersection: create

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

[get output](#) [summary/statistics](#)

To reset all user cart settings (including custom tracks), [click here](#).

And after clicking get output...click get BED

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

Output cpGIslandExt as BED

Include [custom track](#) header:

name=
description=
visibility=
url=

Create one BED record per:

Whole Gene
 Upstream by bases
 Downstream by bases

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

We want the “Whole Gene”

Output looks like this in less:

```
chr1 28735 29810 CpG:_116
chr1 135124 135563 CpG:_30
chr1 327790 328229 CpG:_29
chr1 437151 438164 CpG:_84
chr1 449273 450544 CpG:_99
chr1 533219 534114 CpG:_94
chr1 544738 546649 CpG:_171
chr1 713984 714547 CpG:_60
chr1 762416 763445 CpG:_115
chr1 788863 789211 CpG:_28
chr1 801975 802338 CpG:_24
chr1 805198 805628 CpG:_50
chr1 839694 840619 CpG:_83
chr1 844299 845883 CpG:_153
chr1 854765 854973 CpG:_16
chr1 858970 861632 CpG:_257
chr1 869332 871872 CpG:_178
chr1 875730 878363 CpG:_246
chr1 886356 886602 CpG:_18
chr1 894313 902654 CpG:_615
chr1 906296 906538 CpG:_23
chr1 912869 913153 CpG:_28
chr1 919726 919927 CpG:_15
chr1 933387 937410 CpG:_413
chr1 948670 948894 CpG:_19
chr1 949329 949851 CpG:_35
chr1 954768 956343 CpG:_148
chr1 963795 964507 CpG:_54
chr1 967966 970238 CpG:_185
:|
```



Ensembl Homepage

www.ensembl.org/index.html?redirect=no

Login/Register

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Search all species...

Search: All species for
e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [rs699](#) or [coronary heart disease](#)

Browse a Genome
Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotates genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Favourite genomes

 Human GRCh38.p7	 Human ↗ GRCh37
 Mouse GRCm38.p4	 Zebrafish GRCz10

[Edit favourites](#)

All genomes

-- Select a species -- ↗

[View full list of all Ensembl species](#)

Other species are available in [Ensembl Pre!](#) ↗ and [EnsemblGenomes](#) ↗

Still using Human GRCh37?

Variant Effect Predictor

Gene expression in different tissues

Find SNPs and other variants for my gene

Retrieve gene sequence
GCCTGACTTCCGGGTGC
GGGCTTGCGCGCGAGC
GCGCCTCGCTCGCGCCT
AGGGGACAGATTTGTGA
CACCTCTGGACCGCGTT
CCCAGTCAGCGGGCG

Compare genes across species

Use my own data in Ensembl

ENCODE data in Ensembl

What's New in Ensembl Release 86 (October 2016)

- [Mouse Strains](#)
- [Chicken new assembly and gene set](#)
- [Macaque new assembly and genebuild](#)
- [Mouse lemur new assembly and genebuild](#)
- [Zebrafish: update to Ensembl-Havana merged gene set](#)

[Full details](#) | [All web updates, by release](#) | [More news on our blog](#) ↗

- 28 Oct 2016: [Ensembl genomes 33 is out!](#) ↗
- 17 Oct 2016: [Ensembl helpdesk maintenance](#) ↗
- 05 Oct 2016: [Ensembl 86 has been released!](#) ↗

[Go to Ensembl blog](#) ↗

Select a species

Login/Register



BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Search all species...



Search: for
e.g. **BRCA2** or **rat 5:62797383-63627669** or **rs699** or **coronary heart disease**

Browse a Genome

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Favourite genomes



Human
GRCh38.p7



Human
GRCh37

✓ -- Select a species --

Favourites

Human GRCh38

Human GRCh37

Mouse

Zebrafish

Primates

Bushbaby

Chimpanzee

Gibbon

Gorilla

Human

Macaque

Marmoset

Mouse Lemur

Olive baboon

Orangutan

Tarsier

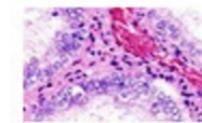
Still using Human GRCh37?



Variant Effect Predictor



Gene expression in different tissues



Find SNPs and other variants for my gene



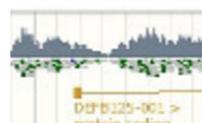
Retrieve gene sequence

```
GCCTGACTTCCGGGTGC  
GGGCTTGTCGGCGGAGC  
GCGCCTCTGCTGCCCT  
AGGGACAGATTTGTGA  
CACCTCTGGAGCGGGTT  
CCCAGTCCAGCGTGGCG
```

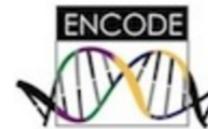
Compare genes across species



Use my own data in Ensembl



ENCODE data in Ensembl



What's New in Ensembl Release 86 (October 2016)

- Mouse Strains
- Chicken new assembly and gene set
- Macaque new assembly and genebuild
- Mouse lemur new assembly and genebuild
- Zebrafish: update to Ensembl-Havana merged gene set

[Full details](#) | [All web updates, by release](#) | [More news on our blog](#)

- 28 Oct 2016: [Ensembl genomes 33 is out!](#)
- 17 Oct 2016: [Ensembl helpdesk maintenance](#)
- 05 Oct 2016: [Ensembl 86 has been released!](#)

[Go to Ensembl blog](#)

Ensembl supports data from external projects through



If necessary, select assembly

e!Ensembl Login/Register

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human (GRCh38.p7) ▾ Search Human...

Human
Homo sapiens

Search all categories ▾ Search Human...

e.g. [BRCA2](#) or [17:63973115-64437414](#) or [rs1333049](#) or [osteoarthritis](#)

Genome assembly: GRCh38.p7 (GCA_000001405.22)

- More information and statistics
- Download DNA sequence (FASTA)
- Convert your data to GRCh38 coordinates
- Display your data in Ensembl

Other assemblies
GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart

What's New in Human release 86

- Human: updated cDNA alignments
- Human: updated RefSeq gene import
- External database references update

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

- More about this genebuild, including RNASeq gene expression models
- Download genes, cDNAs, ncRNA, proteins (FASTA)
- Update your old Ensembl IDs

Additional manual annotation can be found in Vega

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

- More about comparative analysis
- Download alignments (EMF)

Variation

What can I find? Short sequence variants and longer structural variants; disease and other phenotypes

- More about variation in Ensembl
- Download all variants (GVF)
- Variant Effect Predictor

Regulation

What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as enhancers and repressors, and

Pax6 INS FOXP2 DMD ssh
Example gene

Example transcript

ATCGAGCT ATCAGCT ATCGAGAT
Example variant

Example phenotype

Now that you chose your assembly, go to BioMart

[Login/Register](#)

e!GRCh37

BLAST/BLAT | [BioMart](#) | Tools | Downloads | Help & Documentation | Blog

 Search Human... 

Human (GRCh37.p13) ▾

 **Human**
Homo sapiens

Search all categories ▾ Search Human... 

e.g. [BRCA2](#) or [17:63973115-64437414](#) or [rs1333049](#) or [osteoarthritis](#)

Genome assembly: GRCh37.p13 (GCA_000001405.14)

-  [More information and statistics](#)
-  [Download DNA sequence \(FASTA\)](#)
-  [Convert your data to GRCh37.p13 coordinates](#)
-  [Display your data in Ensembl](#)

Other assemblies

GRCh38 (Ensembl release 86)  

Comparative genomics

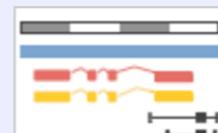
What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

-  [More about comparative analysis](#)
-  [Download alignments \(EMF\)](#)

Regulation

What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as enhancers and repressors, and grch37.ensembl.org/biomart/martview

View karyotype 

Example region 

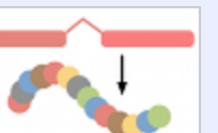
Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

-  [More about this genebuild, including RNASeq gene expression models](#)
-  [Download genes, cDNAs, ncRNA, proteins \(FASTA\)](#)
-  [Update your old Ensembl IDs](#)

 Additional manual annotation can be found in [Vega](#)

Pax6 INS FOXP2 BRCA2 DMD ssh 

Example gene 

Example transcript 

Variation

What can I find? Short sequence variants and longer structural variants; disease and other phenotypes

-  [More about variation in Ensembl](#)
-  [Download all variants \(GVF\)](#)
-  [Variant Effect Predictor](#)



ATCGAGCT ATTCAGCT ATCGAGAT 

Example variant 

Example phenotype 

Let's get some data. Choose Ensembl Gene

e!GRCh37 Login/Register

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Search all species... 

New Count Results URL XML Perl Help

Dataset [None selected]

✓ - CHOOSE DATABASE -

- Ensembl Gene
- Ensembl Variation
- Ensembl Regulation
- Vega

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Pick assembly again (you have access to GRCh37)

e!GRCh37 Login/Register

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Search all species... 

New | Count | Results

Dataset
Homo sapiens genes (GRCh37.p13)

Filters
[None selected]

Attributes
Ensembl Gene ID
Ensembl Transcript ID

Dataset
[None Selected]

- CHOOSE DATASET -
Danio rerio genes (Zv9)
Gallus gallus genes (Galgal4)
✓ Homo sapiens genes (GRCh37.p13)
Mus musculus genes (GRCm38.p2)
Rattus norvegicus genes (Rnor_5.0)

Ailuropoda melanoleuca genes (ailMel1)
Anas platyrhynchos genes (BGI_duck_1.0)
Anolis carolinensis genes (AnoCar2.0)
Astyanax mexicanus genes (AstMex102)
Bos taurus genes (UMD3.1)
Caenorhabditis elegans genes (WBcel235)
Callithrix jacchus genes (C_jacchus3.2.1)
Canis familiaris genes (CanFam3.1)
Cavia porcellus genes (cavPor3)
Chloepus hoffmanni genes (choHof1)
Ciona intestinalis genes (KH)
Ciona savignyi genes (CSAV2.0)
Dasypus novemcinctus genes (Dasnov3.0)
Dipodomys ordii genes (dipOrd1)
Drosophila melanogaster genes (BDGP5)
Echinops telfairi genes (TENREC)
Equus caballus genes (EquCab2)
Erinaceus europaeus genes (eriEur1)
Felis catus genes (Felis_catus_6.2)
Ficedula albicollis genes (FicAlb_1.4)
Gadus morhua genes (gadMor1)
Gasterosteus aculeatus genes (BROADS1)
Gorilla gorilla genes (gorGor3.1)
Ictidomys tridecemlineatus genes (spetri2)
Latimeria chalumnae genes (LatCha1)
Lepisosteus oculatus genes (LepOcu1)

URL | XML | Perl | Help

Outputs) -> Attributes (desired output) -> Results

Select Attributes



BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Login/Register

Search all species...



New Count Results

URL XML Perl Help

Dataset

Homo sapiens genes
(GRCh37.p13)

Filters

[None selected]

Attributes

Ensembl Gene ID

Ensembl Transcript ID

Dataset

[None Selected]

Ensembl Gene

Homo sapiens genes (GRCh37.p13)



Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Select Structures



BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Login/Register

Search all species...



New Count Results

URL XML Perl Help

Dataset

Homo sapiens genes
(GRCh37.p13)

Filters

[None selected]

Attributes

Ensembl Gene ID

Ensembl Transcript ID

Dataset

[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

Missing non coding genes in your mart query output, please check the following [FAQ](#)

- Features Sequences
- Structures Variant (Germline)
- Homologues Variant (Somatic)

GENE:

EXON:

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Let's get starts and ends for transcripts, and chrom



New Count Results

URL XML Perl Help

Dataset

Homo sapiens genes
(GRCh37.p13)

Filters

[None selected]

Attributes

Ensembl Gene ID
Ensembl Transcript ID
Chromosome Name
Transcript Start (bp)
Transcript End (bp)

Dataset

[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

Missing non coding genes in your mart query output, please check the following [FAQ](#)

- Features Sequences
 Structures Variant (Germline)
 Homologues Variant (Somatic)

GENE:

Ensembl

- Ensembl Gene ID
 Ensembl Transcript ID
 Ensembl Protein ID
 Chromosome Name
 Gene Start (bp)
 Gene End (bp)
 Transcript Start (bp)
 Transcript End (bp)
 Transcription Start Site (TSS)
 Transcript length (including UTRs and CDS)
 Strand

- Associated Gene Name
 Associated Gene DB
 5' UTR Start
 5' UTR End
 3' UTR Start
 3' UTR End
 CDS Length
 Transcript count
 Description
 Gene type

EXON:

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

And starts and ends for exons

e!GRCh37 Login/Register

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Search all species... 

New Count Results URL XML Perl Help

Dataset
Homo sapiens genes (GRCh37.p13)

Filters
[None selected]

Attributes

Ensembl Gene ID
Ensembl Transcript ID
Chromosome Name
Transcript Start (bp)
Transcript End (bp)
Exon Chr Start (bp)
Exon Chr End (bp)

Dataset
[None Selected]

GENE:

Ensembl

Ensembl Gene ID
 Ensembl Transcript ID
 Ensembl Protein ID
 Chromosome Name
 Gene Start (bp)
 Gene End (bp)
 Transcript Start (bp)
 Transcript End (bp)
 Transcription Start Site (TSS)
 Transcript length (including UTRs and CDS)
 Strand

Associated Gene Name
 Associated Gene DB
 5' UTR Start
 5' UTR End
 3' UTR Start
 3' UTR End
 CDS Length
 Transcript count
 Description
 Gene type

EXON:

Exon Information

Exon Chr Start (bp)
 Exon Chr End (bp)
 Constitutive Exon
 Exon Rank in Transcript
 start phase
 cDNA coding start

cDNA coding end
 Genomic coding start
 Genomic coding end
 Ensembl Exon ID
 CDS Start
 CDS End

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

And lastly let's display the results

e!GRCh37 Login/Register

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

 Search all species... 

New Count Results

URL XML Perl Help

Dataset
Homo sapiens genes (GRCh37.p13)

Filters
[None selected]

Attributes

Ensembl Gene ID
Ensembl Transcript ID
Chromosome Name
Transcript Start (bp)
Transcript End (bp)
Exon Chr Start (bp)
Exon Chr End (bp)

Dataset
[None Selected]

GENE:

Ensembl

Ensembl Gene ID
 Ensembl Transcript ID
 Ensembl Protein ID
 Chromosome Name
 Gene Start (bp)
 Gene End (bp)
 Transcript Start (bp)
 Transcript End (bp)
 Transcription Start Site (TSS)
 Transcript length (including UTRs and CDS)
 Strand

Associated Gene Name
 Associated Gene DB
 5' UTR Start
 5' UTR End
 3' UTR Start
 3' UTR End
 CDS Length
 Transcript count
 Description
 Gene type

EXON:

Exon Information

Exon Chr Start (bp)
 Exon Chr End (bp)
 Constitutive Exon
 Exon Rank in Transcript
 start phase
 cDNA coding start

cDNA coding end
 Genomic coding start
 Genomic coding end
 Ensembl Exon ID
 CDS Start
 CDS End

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Output looks like this

e!GRCh37 Login/Register

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Search all species... 

New Count Results   

URL XML Perl Help     

Dataset: Homo sapiens genes (GRCh37.p13)

Filters: [None selected]

Attributes:

- Ensembl Gene ID
- Ensembl Transcript ID
- Chromosome Name
- Transcript Start (bp)
- Transcript End (bp)
- Exon Chr Start (bp)
- Exon Chr End (bp)

Export all results to: File TSV Unique results only 

Email notification to:

View: 100 rows as: HTML Unique results only

Ensembl Gene ID	Ensembl Transcript ID	Chromosome Name	Transcript Start (bp)	Transcript End (bp)	Exon Chr Start (bp)	Exon Chr End (bp)
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66119285	66119659
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66298434	66298819
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66314236	66314392
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66320895	66321004
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66339743	66339847
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66341024	66341071
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66424056	66424100
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66440552	66440621
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66447170	66447234
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66448221	66448294
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66455382	66456619
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66320895	66321004
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66339743	66339847
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66341024	66341071
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66424056	66424100
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66440552	66440621
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66447170	66447234

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Let's just download the first 200 transcripts

e!GRCh37 Login/Register

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Search all species... 

New | Count | Results | URL | XML | Perl | Help

Dataset
Homo sapiens genes (GRCh37.p13)

Filters
[None selected]

Attributes
Ensembl Gene ID
Ensembl Transcript ID
Chromosome Name
Transcript Start (bp)
Transcript End (bp)
Exon Chr Start (bp)
Exon Chr End (bp)

Dataset
[None Selected]

Export all results to
Email notification to

View
10
20
50
100
150
200
All

Rows as TSV Unique results only Go

Ensembl Gene ID	Ensembl Transcript ID	Chromosome Name	Transcript Start (bp)	Transcript End (bp)	Exon Chr Start (bp)	Exon Chr End (bp)
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66119285	6611965
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66298434	6629881
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66314236	6631439
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66320895	6632100
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66339743	6633984
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66341024	6634101
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66424056	6642410
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66440552	6644062
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66447170	6644721
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66448221	6644829
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66455382	6645661
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66320895	6632100
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66339743	6633984
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66341024	6634101
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66424056	6642410
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66440552	6644062
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66447170	6644721
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66448221	6644829
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66448894	6644910

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Click “Go”

e!GRCh37 Login/Register

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Search all species... 

New **Count** **Results** **URL** **XML** **Perl** **Help**

Dataset
Homo sapiens genes (GRCh37.p13)

Filters
[None selected]

Attributes
Ensembl Gene ID
Ensembl Transcript ID
Chromosome Name
Transcript Start (bp)
Transcript End (bp)
Exon Chr Start (bp)
Exon Chr End (bp)

Dataset
[None Selected]

Export all results to **File** **TSV** Unique results only **Go**

Email notification to

View **200** rows as **TSV** Unique results only

Ensembl Gene ID	Ensembl Transcript ID	Chromosome Name	Transcript Start (bp)	Transcript End (bp)
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Save as “bedtest” to whatever folder

Ensembl Gene ID	Ensembl Transcript ID	Chromosome Name	Transcript Start (bp)	Transcript End (bp)	Exon Chr Start (bp)	Exon Chr End (bp)
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66119285	66119659
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66298434	66298819
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66314236	66314392
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66320895	66321004
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66339743	66339847
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66341024	66341071
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66424056	66424100
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66440552	66440621
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66447170	66447234
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66448221	66448294
ENSG00000261657	ENST00000566782	HG991_PATCH	66119285	66456619	66455382	66456619
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66320895	66321004
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66339743	66339847
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66341024	66341071
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66424056	66424100
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66440552	66440621
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66447170	66447234
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66448221	66448294
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66448894	66449105
ENSG00000261657	ENST00000562780	HG991_PATCH	66320895	66455748	66455382	66455748
ENSG00000261657	ENST00000569579	HG991_PATCH	66320895	66456619	66320895	66321004
ENSG00000261657	ENST00000569579	HG991_PATCH	66320895	66456619	66339743	66339847
ENSG00000261657	ENST00000569579	HG991_PATCH	66320895	66456619	66341024	66341071
ENSG00000261657	ENST00000569579	HG991_PATCH	66320895	66456619	66382145	66382229
ENSG00000261657	ENST00000569579	HG991_PATCH	66320895	66456619	66424056	66424100
ENSG00000261657	ENST00000569579	HG991_PATCH	66320895	66456619	66440552	66440621
ENSG00000261657	ENST00000569579	HG991_PATCH	66320895	66456619	66447170	66447234
ENSG00000261657	ENST00000569579	HG991_PATCH	66320895	66456619	66448221	66448294
ENSG00000261657	ENST00000569579	HG991_PATCH	66320895	66456619	66455382	66456619
ENSG00000261657	ENST00000568242	HG991_PATCH	66320895	66465398	66320895	66321004
ENSG00000261657	ENST00000568242	HG991_PATCH	66320895	66465398	66339743	66339847
ENSG00000261657	ENST00000568242	HG991_PATCH	66320895	66465398	66341024	66341071
ENSG00000261657	ENST00000568242	HG991_PATCH	66320895	66465398	66424056	66424100
ENSG00000261657	ENST00000568242	HG991_PATCH	66320895	66465398	66433723	66433776
ENSG00000261657	ENST00000568242	HG991_PATCH	66320895	66465398	66440552	66440621
ENSG00000261657	ENST00000568242	HG991_PATCH	66320895	66465398	66447170	66447234
ENSG00000261657	ENST00000568242	HG991_PATCH	66320895	66465398	66448221	66448294
ENSG00000261657	ENST00000568242	HG991_PATCH	66320895	66465398	66455382	66455762
ENSG00000261657	ENST00000568242	HG991_PATCH	66320895	66465398	66456102	66456194
ENSG00000261657	ENST00000568242	HG991_PATCH	66320895	66465398	66457809	66458645
ENSG00000261657	ENST00000568242	HG991_PATCH	66320895	66465398	66463876	66465398
ENSG00000261657	ENST00000565530	HG991_PATCH	66339287	66448276	66339287	66339342
ENSG00000261657	ENST00000565530	HG991_PATCH	66339287	66448276	66339743	66339847
ENSG00000261657	ENST00000565530	HG991_PATCH	66339287	66448276	66341024	66341071
ENSG00000261657	ENST00000565530	HG991_PATCH	66339287	66448276	66424056	66424100
ENSG00000261657	ENST00000565530	HG991_PATCH	66339287	66448276	66447170	66447234
ENSG00000261657	ENST00000565530	HG991_PATCH	66339287	66448276	66448221	66448276
ENSG00000223116	ENST00000411184	13	23551994	23552136	23551994	23552136
ENSG00000233440	ENST00000418454	13	23708313	23708703	23708313	23708703

Data looks like this

Let's turn this into BED format shall we?

- After saving “bedtest.txt”, run:
 - `awk '{print $3,$6,$7,$4,$5,$1,$2}' FS='\t' OFS='\t' bedtest.txt | sed '1d' > transcripts.bed`
 - And now it's in BED format! Check with less `transcripts.bed`

Tabix

- Tabix is packaged with samtools, and bgzip
- Can create indices for any standard genome format:
 - BAM, GFF, SAM, BED, VCF
 - Compress files with bgzip first to save space
- Allows for rapid querying of a genome position file
- Makes for quick searching in IGV, less data needs to be loaded into RAM

Tabix

- Example with the refseq.bed file we used earlier.
 - First, make sure the file is sorted, then compress it:
 - sort -k1,1 -k2,2n refseq.bed | bgzip -c > refseq.bed.gz
 - Then index it with tabix, this makes a .tbi file:
 - tabix refseq.bed.gz
 - Then try to query it (note that I used “chr”, because we used it):
 - tabix refseq.bed.gz chr1:1-100000

```
[semja:Documents] $ sort -k1,1 -k2,2n refseq.bed | bgzip -c > refseq.bed.gz
[semja:Documents] $ tabix refseq.bed.gz
[semja:Documents] $ tabix refseq.bed.gz chr1:1-100000
chr1 11873 14409 NR_046018 0 + 14409 14409 0 3 354,109,1189, 0,739,1347,
chr1 14361 29370 NR_024540 0 - 29370 29370 0 11 468,69,152,159,198,136,137,147,99,154,50, 0,608,1434,2245,2496,2871,3244,3553,3906,10376,14959,
chr1 17368 17436 NR_106918 0 - 17436 17436 0 1 68, 0,
chr1 17368 17436 NR_107062 0 - 17436 17436 0 1 68, 0,
chr1 17368 17436 NR_107063 0 - 17436 17436 0 1 68, 0,
chr1 17368 17436 NR_128720 0 - 17436 17436 0 1 68, 0,
chr1 30365 30503 NR_036051 0 + 30503 30503 0 1 138, 0,
chr1 30365 30503 NR_036266 0 + 30503 30503 0 1 138, 0,
chr1 30365 30503 NR_036267 0 + 30503 30503 0 1 138, 0,
chr1 30365 30503 NR_036268 0 + 30503 30503 0 1 138, 0,
chr1 34610 36081 NR_026818 0 - 36081 36081 0 3 564,205,361, 0,666,1110,
chr1 34610 36081 NR_026820 0 - 36081 36081 0 3 564,205,361, 0,666,1110,
chr1 69090 70008 NM_001005484 0 + 69090 70008 0 1 918, 0,
```

Use of tabix in IGV

- First of all, IGV will ask to index a file that you have not indexed yet (it will use igvtools to do so, but it is not as efficient as tabix)
- If you passed a tabixed file into IGV, it will be able to access parts of the file much faster as IGV can read a .tbi index

Acknowledgements

This exercise is provided by Jim Havrilla, PhD

Modified by Kai Wang and Qian Liu to suit this class