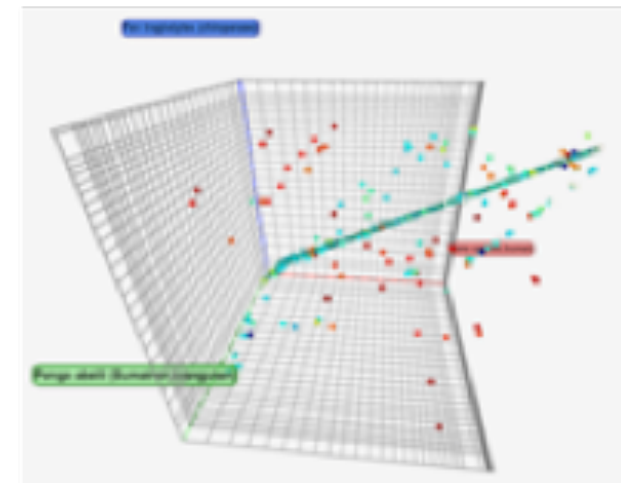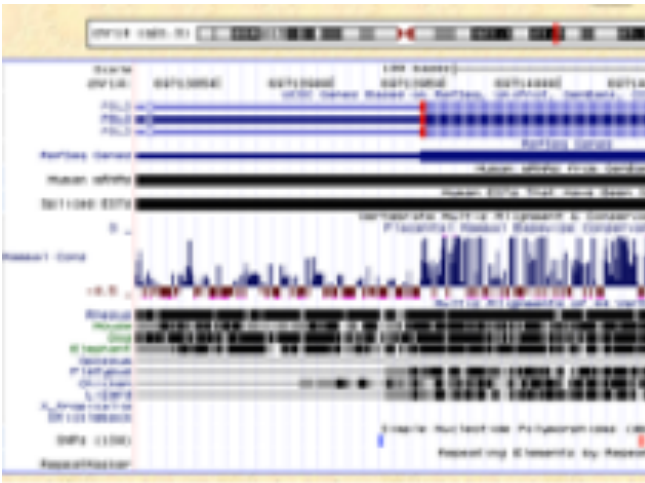# Computational Genomics

Data Retrieval I

Data Retrieval
Retrieving Data From UCSC

Table Browser

Table Browser

Use this tool to retrieve and export data from the Genome Browser annotation track database. You can limit retrieval based on data attributes and intersect or merge with data from another track, or retrieve DNA sequence covered by a track. More...

Select dataset

clade: Mammal ⌄ **genome:** Human ⌄ **assembly:** Dec. 2013 (GRCh38/hg38) ⌄
group: Genes and Gene Predictions ⌄ **track:** GENCODE V38 ⌄
table: knownGene ⌄ [describe table schema]

Define region of interest

region: ⦿ genome ◯ position [chrX:15,560,138-15,602,945] [lookup] [define regions]
identifiers (names/accessions): [paste list] [upload list]

Optional: Subset, combine, compare with another track

filter: [create]
intersection: [create]

Retrieve and display data

output format: [all fields from selected table ⌄] Send output to ☐ Galaxy ☐ GREAT
output filename: [   ] (leave blank to keep output in browser)
file type returned: ⦿ plain text ◯ gzip compressed

[get output] [summary/statistics]

Data Retrieval
Retrieving Data From UCSC

Table Browser

- Retrieving all coding exons corresponding to: ***Homo sapiens* ChrX**

Select dataset

clade: Mammal ⌄ **genome:** Human ⌄ **assembly:** Dec. 2013 (GRCh38/hg38) ⌄
group: Genes and Gene Predictions ⌄ **track:** GENCODE V38 ⌄
table: knownGene ⌄ [describe table schema]

Define region of interest

region: ○ genome ● position chrX:15,560,138-15,602,945 [lookup] [define regions]
identifiers (names/accessions): [paste list] [upload list]

Optional: Subset, combine, compare with another track

filter: [create]
intersection: [create]

Retrieve and display data

output format: all fields from selected table ⌄ Send output to ☐ Galaxy ☐ GREAT
output filename: [   ] (leave blank to keep output in browser)
file type returned: ● plain text ○ gzip compressed

[get output] [summary/statistics]

Data Retrieval
Retrieving Data From UCSC

Table Browser

- Retrieving all coding exons corresponding to: ***Homo sapiens* ChrX**

```
#chrom  chromStart  chromEnd  name  score  strand  thickStart  thickEnd  reserved  blockCount  blockSizes  chromStarts  name2  cdsStartStat  cdsEndStat  exonFrames  type  geneName  geneName2  geneType  transcriptClass
source  transcriptType  tag  level  tier
chrX  15494565  15600960  ENST00000679162.1  0  -  15500533  15600911  789624  19  284,88,195,117,101,59,173,123,99,145,227,170,98,106,113,144,94,159,235,
0,5932,69458,71687,73160,75729,77058,77635,78801,81100,83523,86655,90909,93187,94778,97147,97663,100279,106160,  uc288ugk.1  none  none  -1,2,2,2,0,1,2,2,2,1,2,0,1,0,1,1,0,0,0,  none  ACE2  none  none  coding  havana_homo_sapiens  nonsense_mediated_decay
RNA_Seq_supported_only,inferred_transcript_model,overlapping_locus  2  all
chrX  15494565  15600960  ENST00000679212.1  0  -  15494839  15600911  789624  18  284,195,117,101,59,173,123,99,145,227,170,98,106,113,144,94,159,235,
0,69458,71687,73160,75729,77058,77635,78801,81100,83523,86655,90909,93187,94778,97147,97663,100279,106160,  uc288uii.1  none  none  2,2,2,0,1,2,2,2,1,2,0,1,0,1,1,0,0,0,  none  ACE2  none  none  coding  havana_homo_sapiens  protein_coding
RNA_Seq_supported_only,appris_alternative_2,basic,inferred_transcript_model,overlapping_locus  2  basic,all
chrX  15518188  15600960  ENST00000679278.1  0  -  15518927  15600911  789624  18  791,195,117,101,59,173,123,99,145,227,170,98,106,113,144,94,159,235,
0,45835,48064,49537,52106,53435,54012,55178,57477,59900,63032,67286,69564,71155,73524,74040,76656,82537,  uc288ukw.1  none  none  2,2,2,0,1,2,2,2,1,2,0,1,0,1,1,0,0,0,  none  ACE2  none  none  coding  havana_homo_sapiens  protein_coding
RNA_Seq_supported_only,appris_alternative_2,basic,overlapping_locus  2  basic,all
chrX  15550714  15600960  ENST00000678046.1  0  -  15551072  15600911  789624  18  365,195,117,101,59,173,123,99,145,227,170,98,106,113,144,94,159,235,
0,13309,15538,17011,19580,20909,21486,22652,24951,27374,30506,34760,37038,38629,40998,41514,44130,50011,  uc288spv.1  none  none  2,2,2,0,1,2,2,2,1,2,0,1,0,1,1,0,0,0,  none  ACE2  none  none  coding  havana_homo_sapiens  protein_coding
RNA_Seq_supported_only,appris_alternative_2,basic,inferred_transcript_model,overlapping_locus  2  basic,all
chrX  15557038  15580438  ENST00000677282.1  0  -  15561904  15580312  789624  11  4975,195,117,101,59,173,123,99,145,227,158,  0,6985,9214,10687,13256,14585,15162,16328,18627,21050,23242,  uc288rms.1  none  none
2,2,2,0,1,2,2,2,1,2,0,  none  ACE2  A0A7D6JAD5  none  coding  havana_homo_sapiens  protein_coding  CAGE_supported_TSS,RNA_Seq_supported_only,basic,inferred_transcript_model,overlapping_locus  2  basic,all
chrX  15557038  15601208  ENST00000680121.1  0  -  15561904  15600911  789624  19  4975,195,117,101,59,173,123,99,145,227,170,98,106,113,144,94,159,289,26,
0,6985,9214,10687,13256,14585,15162,16328,18627,21050,24182,28436,30714,32305,34674,35190,37806,43687,44144,  uc289ajk.1  none  none  2,2,2,0,1,2,2,2,1,2,0,1,0,1,1,0,0,0,-1,  none  ACE2  Q9BYF1  none  coding  havana_homo_sapiens  protein_coding
CAGE_supported_TSS,CCDS,RNA_Seq_supported_only,appris_principal_2,basic,overlapping_locus  2  basic,all
chrX  15557038  15602155  ENST00000427411.2  0  -  15561904  15600911  789624  19  4975,195,117,101,59,173,123,99,145,227,170,98,106,113,144,94,159,289,200,
0,6985,9214,10687,13256,14585,15162,16328,18627,21050,24182,28436,30714,32305,34674,35190,37806,43687,44917,  uc004cxb.3  none  none  2,2,2,0,1,2,2,2,1,2,0,1,0,1,1,0,0,0,-1,  none  ACE2  Q9BYF1  none  coding  ensembl_havana_transcript_homo_sapiens  protein_coding
CAGE_supported_TSS,CCDS,RNA_Seq_supported_only,alternative_5_UTR,appris_principal_2,basic,inferred_transcript_model,overlapping_locus  2  basic,all
chrX  15557038  15607236  ENST00000678073.1  0  -  15561904  15600911  789624  19  4975,195,117,101,59,173,123,99,145,227,170,98,106,113,144,94,159,289,100,
0,6985,9214,10687,13256,14585,15162,16328,18627,21050,24182,28436,30714,32305,34674,35190,37806,43687,50098,  uc288sqw.1  none  none  2,2,2,0,1,2,2,2,1,2,0,1,0,1,1,0,0,0,-1,  none  ACE2  Q9BYF1  none  coding  havana_homo_sapiens  protein_coding
CCDS,RNA_Seq_supported_only,alternative_5_UTR,appris_principal_2,basic,inferred_transcript_model,overlapping_locus  2  canonical,basic,all
chrX  15561032  15600960  ENST00000252519.8  0  -  15561904  15600911  789624  18  981,195,117,101,59,173,123,99,145,227,170,98,106,113,144,94,159,235,
0,2991,5220,6693,9262,10591,11168,12334,14633,17056,20188,24442,26720,28311,30680,31196,33812,39693,  uc004cxa.3  none  none  2,2,2,0,1,2,2,2,1,2,0,1,0,1,1,0,0,0,  none  ACE2  Q9BYF1  none  coding  havana_homo_sapiens  protein_coding
CAGE_supported_TSS,CCDS,Ensembl_canonical,MANE_Select,appris_principal_2,basic,overlapping_locus  2  basic,all
chrX  15561035  15566287  ENST00000471548.5  0  -  15561904  15566287  789624  4  270,498,195,35,  0,480,2988,5217,  uc064yct.1  none  none  -1,2,2,0,  none  ACE2  none  none  coding  havana_homo_sapiens
nonsense_mediated_decay  cds_start_NF,overlapping_locus  2  all
chrX  15561046  15675608  ENST00000649243.1  0  -  15600989  15659062  789624  20  967,195,117,123,99,145,227,170,98,106,113,144,94,159,289,195,114,86,59,166,
0,2977,5206,11154,12320,14619,17042,20174,24428,26706,28297,30666,31182,33798,39679,78515,83869,97969,103290,114396,  uc286crf.1  none  none  -1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,2,2,2,0,-1,-1,  none  RP11-478H11.3-001  A0A3B3IT09  none  coding
havana_homo_sapiens  nonsense_mediated_decay  Ensembl_canonical,appris_principal_1,basic,overlapping_locus,readthrough_transcript  2  canonical,basic,all
chrX  15561879  15566709  ENST00000473851.1  0  -  15561879  15561879  16646144  3  134,195,457,  0,2144,4373,  uc064ycu.1  none  none  -1,-1,-1,  none  ACE2  none  none  nonCoding  havana_homo_sapiens
retained_intron  overlapping_locus  2  all
chrX  15602880  15621484  ENST00000421585.1  0  +  15602880  15602880  25600  2  101,473,  0,18131,  uc064ycw.1  none  none  -1,-1,  none  GS1-594A7.3  none  none  nonCoding  havana_homo_sapiens  lncRNA
Ensembl_canonical,basic  2  canonical,basic,all
```

This is a BED File

Example:

Here's an example of an annotation track, introduced by a **header line**, that is followed by a complete BED definition:

```
track name=pairedReads description="Clone Paired Reads" useScore=1
chr22 1000 5000 cloneA 960 + 1000 5000 0 2 567,488, 0,3512
chr22 2000 6000 cloneB 900 - 2000 6000 0 2 433,399, 0,3601
```

Data Retrieval
Retrieving Data From UCSC

Table Browser

Example:

Here's an example of an annotation track, introduced by a **header line**, that is followed by a complete BED definition:

```
track name=pairedReads description="Clone Paired Reads" useScore=1
chr22 1000 5000 cloneA 960 + 1000 5000 0 2 567,488, 0,3512
chr22 2000 6000 cloneB 900 - 2000 6000 0 2 433,399, 0,3601
```

- The first three required BED fields are:

  chrom - The name of the chromosome (e.g. chr3, chrY, chr2_random) or scaffold (e.g. scaffold10671).

  chromStart - The starting position of the feature in the chromosome or scaffold. The first base in a chromosome is numbered 0.

  chromEnd - The ending position of the feature in the chromosome or scaffold. The chromEnd base is not included in the display of the feature, however, the number in position format will be represented. For example, the first 100 bases of chromosome 1 are defined as chrom=1, chromStart=0, chromEnd=100, and span the bases numbered 0-99 in our software (not 0-100), but will represent the position notation chr1:1-100. Read more here.

  chromStart and chromEnd can be identical, creating a feature of length 0, commonly used for insertions. For example, use chromStart=0, chromEnd=0 to represent an insertion before the first nucleotide of a chromosome.

Data Retrieval
Retrieving Data From UCSC

Table Browser

Example:

Here's an example of an annotation track, introduced by a **header line**, that is followed by a complete BED definition:

```
track name=pairedReads description="Clone Paired Reads" useScore=1
chr22 1000 5000 cloneA 960 + 1000 5000 0 2 567,488, 0,3512
chr22 2000 6000 cloneB 900 - 2000 6000 0 2 433,399, 0,3601
```

- The 9 additional optional BED fields are:

  name - Defines the name of the BED line. This label is displayed to the left of the BED line in the Genome Browser window when the track is open to full display mode or directly to the left of the item in pack mode.

  score - A score between 0 and 1000. If the track line useScore attribute is set to 1 for this annotation data set, the score value will determine the level of gray in which this feature is displayed (higher numbers = darker gray). This table shows the Genome Browser's translation of BED score values into shades of gray:

shade									
score in range	≤ 166	167-277	278-388	389-499	500-611	612-722	723-833	834-944	≥ 945

Data Retrieval
Retrieving Data From UCSC

Table Browser

Example:
Here's an example of an annotation track, introduced by a **header line**, that is followed by a complete BED definition:

```
track name=pairedReads description="Clone Paired Reads" useScore=1
chr22 1000 5000 cloneA 960 + 1000 5000 0 2 567,488, 0,3512
chr22 2000 6000 cloneB 900 - 2000 6000 0 2 433,399, 0,3601
```

- **strand** - Defines the strand. Either "." (=no strand) or "+" or "-".
- **thickStart** - The starting position at which the feature is drawn thickly (for example, the start codon in gene displays).

  When there is no thick part, thickStart and thickEnd are usually set to the chromStart position.
- **thickEnd** - The ending position at which the feature is drawn thickly (for example the stop codon in gene displays).
- **itemRgb** - An RGB value of the form R,G,B (e.g. 255,0,0). If the track line itemRgb attribute is set to "On", this RBG value

  will determine the display color of the data contained in this BED line. NOTE: It is recommended that a simple color

  scheme (eight colors or less) be used with this attribute to avoid overwhelming the color resources of the Genome

  Browser and your Internet browser.
- **blockCount** - The number of blocks (exons) in the BED line.
- **blockSizes** - A comma-separated list of the block sizes. The number of items in this list should correspond to blockCount.
- **blockStarts** - A comma-separated list of block starts. All of the blockStart positions should be calculated relative

  to chromStart. The number of items in this list should correspond to blockCount.

In BED files with block definitions, the first blockStart value must be 0, so that the first block begins at chromStart.

Similarly, the final blockStart position plus the final blockSize value must equal chromEnd. Blocks may not overlap.