

Getting the genes from regions of homozygosity using the UCSC browser

For use with SimulConsult's Regions of Homozygosity feature

The regions of homozygosity are typically reported as a BED file (Browser Extensible Data), with many lines in one of the following 2 types of format:

chr2 27300000 27330000 or chr2:27,300,000-27,330,000

You can use either of the 2 formats above to get a list of genes using the UCSC Table Browser using the following steps:

1. Go to <https://genome.ucsc.edu/cgi-bin/hgTables>. If you've used this page before, your browser will save your settings in cookies and these will persist even if you refresh the page.
2. In "assembly" choose the desired assembly, typically Dec. 2013 (GRCh38/hg38) or Feb. 2009 (GRCh37/hg19). The example shown uses GRCh38/hg38.

clade: Mammal genome: Human assembly: Dec. 2013 (GRCh38/hg38) Feb. 2009 (GRCh37/hg19) Mar. 2006 (NCBI36/hg18) May 2004 (NCBI35/hg17) July 2003 (NCBI34/hg16)

group: Genes and Gene Predictions track: NCBI RefSeq

table: RefSeq All (ncbiRefSeq) describe table schema

region: genome position chrX:15,560,138-15,602,945 lookup define regions

identifiers (names/accessions): paste list upload list

filter: create

subtrack merge: create

intersection: create

correlation: create

output format: selected fields from primary and related tables Send output to Galaxy GREAT

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](#).

3. In "track" choose "NCBI RefSeq"

clade: Mammal genome: Human assembly: Feb. 2009 (GRCh37/hg19)

group: Genes and Gene Predictions track: NCBI RefSeq

table: RefSeq All (ncbiRefSeq) describe table schema

region: genome ENCODE Pilot regions

identifiers (names/accessions): paste list

filter: create

subtrack merge: create

intersection: create

correlation: create

output format: selected fields from primary and related tables

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

file type returned: plain text gzip compressed

get output summary/statistics

4. In **"region"**:

- a) On the first run click the **"define regions"** button

region: ☒ genome ☐ ENCODE Pilot regions ☐ position chrX:15,578,261-15,621,068

On the resulting screen, paste many lines of regions such as " chr2:27,300,000-27,330,000" into the text area.

Enter region definition

Paste regions: Or upload file: no file selected

chr2:27,300,000-27,330,000

- b) Note: On subsequent runs, **"defined regions"** will be remembered by the browser and already selected; choose **"change"** or **"clear"** if you want to change the regions.

region: ☐ genome ☐ ENCODE Pilot regions ☐ position chrX:15,578,261-15,621,068 ☒ defined regions

5. Click the **"submit"** button

6. In **"output format"**, choose **"selected fields from primary and related tables"**

correlation:

output format: ☐ Galaxy ☐ GREAT

output file:

file type return:

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

7. Click the **"get output"** button

- Check the "**name2**" checkbox

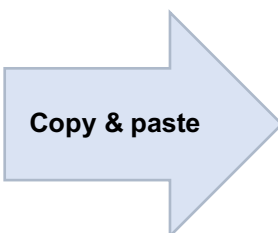
Select Fields from hg19.ncbiRefSeq

<input type="checkbox"/>	bin	
<input type="checkbox"/>	name	Name of gene (usually transcript_id from GTF)
<input type="checkbox"/>	chrom	Reference sequence chromosome or scaffold
<input type="checkbox"/>	strand	+ or - for strand
<input type="checkbox"/>	txStart	Transcription start position (or end position for minus strand item)
<input type="checkbox"/>	txEnd	Transcription end position (or start position for minus strand item)
<input type="checkbox"/>	cdsStart	Coding region start (or end position for minus strand item)
<input type="checkbox"/>	cdsEnd	Coding region end (or start position for minus strand item)
<input type="checkbox"/>	exonCount	Number of exons
<input type="checkbox"/>	exonStarts	Exon start positions (or end positions for minus strand item)
<input type="checkbox"/>	exonEnds	Exon end positions (or start positions for minus strand item)
<input type="checkbox"/>	score	score
<input checked="" type="checkbox"/>	name2	Alternate name (e.g. gene_id from GTF)
<input type="checkbox"/>	cdsStartStat	Status of CDS start annotation (none, unknown, incomplete, or complete)
<input type="checkbox"/>	cdsEndStat	Status of CDS end annotation (none, unknown, incomplete, or complete)
<input type="checkbox"/>	exonFrames	Exon frame {0,1,2}, or -1 if no frame for exon

get output cancel check all clear all

- Click the "**get output**" button
- Paste the full output from the UCSC tool into the text area at the top of the SimulConsult "Regions of Homozygosity analysis" screen accessible under the helix symbol on the top menu. Don't bother to remove the header or the repeated HGNC symbols; the Regions of Homozygosity analysis removes these automatically.

```
#name2
TRIM54
TRIM54
TRIM54
TRIM54
TRIM54
UCN
MPV17
MPV17
MPV17
MPV17
MPV17
MPV17
MPV17
MPV17
MPV17
GTF3C2
GTF3C2
GTF3C2
GTF3C2
```



Copy & paste

Regions of Homozygosity analysis

```
#name2
TRIM54
TRIM54
TRIM54
TRIM54
TRIM54
TRIM54
UCN
MPV17
MPV17
MPV17
```

Load genes

When the parents are first cousins, thousands of lines of HGNC gene symbols are obtained, with many repeated symbols but typically ~2,000 unique genes, of which only hundreds will have known human biallelic phenotypes. From the tiny gene list above, the last time we checked, only the MPV17 gene has a known human phenotype.

When you click the Next button on the “Regions of Homozygosity analysis” screen in SimulConsult, you will see the "Genotype" tab of the Diagnose screen. If you’ve already entered a robust set of findings for the patient, the genes with highest pertinence for the differential diagnosis for the patient will be high in the list of pertinent genes. You can enter additional pertinent positive or pertinent negative findings to help distinguish among the most pertinent genes and choose which to sequence. The useful findings suggested on the "Add findings" and "Add tests" tabs of the Diagnose screens can help choose findings on which to comment or other types of testing that can be helpful.