

Retrieving data from gnomAD

Marie Saitou

Purpose

- We have a list of genes
- We would like to access an online database of genetic variation, and obtain the variation data of particular genes (web scraping?)
- They offer a downloadable files but it is 500Gb. So, working on the web seems to be faster

a database - gnomAD

- <https://gnomad.broadinstitute.org/>

The image shows a screenshot of the gnomAD browser interface. A yellow box on the left, labeled "gene list", contains a table of gene symbols. A yellow arrow points from the "OR6S1" entry in this list to a dropdown menu in the browser header, which also displays "OR6S1". Another yellow arrow points from the "OR6S1" entry to a search box labeled "Search each gene from the gene list". The browser header includes the text "gnomAD browser" and navigation links: "About", "Downloads", "Terms", and "Contact". A blue banner below the header states "This is a new version of gnomAD available at <http://gnomad-old.broadinstitute.org>". The main content area features the "gnomAD" logo, a blue mountain range graphic, and the text "genome aggregation database". At the bottom, there is a search bar with the placeholder text "Search by gene, region, or variant" and examples: "Examples - Gene: [PCSK9](#), Variant: [1-55516888-G-GA](#)".

SYMBOL
OR52M1
OR51B5
OR52N4
OR11H1
OR4K15
OR6S1
OR5AP2
OR5H1
OR13C2
OR2K2
OR8G5
OR1L3
OR2T27

gnomAD browser

OR6S1

About Downloads Terms Contact

This is a new version of gnomAD available at <http://gnomad-old.broadinstitute.org>

gene list

Search each gene from the gene list

gnomAD

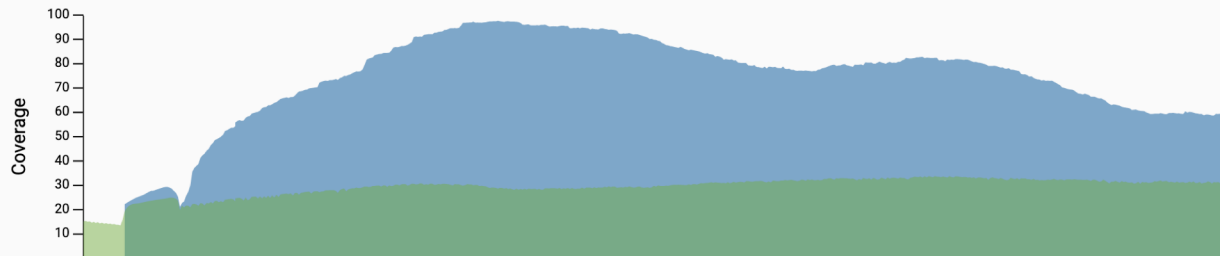
genome aggregation database

Search by gene, region, or variant

Examples - Gene: [PCSK9](#), Variant: [1-55516888-G-GA](#)

OR6S1 olfactory receptor, family 6, subfamily S, member 1

Ensembl gene ID [ENSG00000181803](#)
Ensembl transcript ID [ENST00000320704 \(canonical\)](#)
UCSC Browser [14:21108856-21109851](#)
GeneCards [OR6S1](#)
OMIM —



get a CSV for each gene

Show tissues
Mean
pext

ClinVar
pathogenic
and likely
pathogenic
variants (0)

gnomAD v2.1.1
(360)

Viewing in table

21,108,784 21,108,899 21,109,013 21,109,127 21,109,241 21

☒ LoF only ☒ Missense only ☒ Synonymous only ☒ Other only

Export variants to CSV

The output csv (genetic variation) head

Home

Insert

Page Layout

Formulas

Data

Review

View

</

I'd like to add a column
of gene name

Merge the output files to a single csv

header
CSV of GENE1
CSV of GeNE2
CSV of GENE3

⋮