Just-DNA-Seq

Release 0.1

VovixSW

Oct 07, 2022

CONTENTS

Just-DNA-Seq is a project to facilitate working with human genomes at all levels - from clinical cases to personal curiosity, education and longevity. We envision the future when genomics is becoming more *available*, *understandable* and *useful* for everybody, especially those interested in life extension and improving human condition. For this purpose we use OpenCravat framework to integrate *annotators* which are tools or databases accumulating what we actually know about the genome - genes, their influence on health or drug response, polygenic risk scores and so on.

Check out the *Getting Started* section for further information, including the section *Genome assemblies supported by OpenCravat* to understand whether you have the appropriate genome assembly.

Note: This project is under active development.

CONTENTS

1.1 Getting Started

1.1.1 OakVar vs OpenCravat

1.1.2 Installing OakVar/OpenCravat

1.1.3 Uploading Your Genome

In OakVar:

1. Open OakVar in your browser. You will see the index page:

![Index page](https://github.com/dna-seq/just-dna-seq/tree/master/docs/source/index.png)

2. In the **Variants** section you should choose the right assembly version of the **Genome:** hg38/GRCh38, hg19/GRCh37, or hg18/GRCh36.

3. Click **Add input files**.

1.1.4 Register with OpenCravat

1.1.5 Genome assemblies supported by OpenCravat

1.1.6 Uploading files

1.2 Annotators

Annotators can be divided into 2 groups:

1) Tools that predict pathogenicity (bold)

2) Tools that provide information like databases

- **cadd_exome** (1.6.1) CADD is a tool for scoring the deleteriousness of single nucleotide variants as well as insertion/deletions variants in the human genome.

- gnomad_gene (2.2.1) Gene level population statistics from gnomAD

- pubmed (1.1.5) articles related to a particular gene

- **clingen** (1.0.1) - NIH-funded resource that defines the clinical relevance of genes and variants

- **clinpred** (1.0.0) - prediction tool to identify disease-relevant nonsynonymous single nucleotide variants

- clinvar (2021.10.01) - ClinVar is an archive of reports of the relationships among human variations and phenotypes, archive of interpretations of clinically relevant variants (Uncertain significance, Likely pathogenic, Pathogenic etc.)

- mitomap (1.1.0) A human mitochondrial genome database

- ncbigene (2019.08.02) - gene descriptions from NCBI (National Center for Biotechnology Information)Gene database.

- omim (1.0.0) Catalog of human genes and genetic disorders and traits.

- **prec** (3.6.0)provides a database identifying rare and likely deleterious loss-of-function (LoF) alleles

- **provean** (1.0.0), tool which predicts whether an amino acid substitution or indel has an impact on the biological function of a protein

- **revel** (2020.12.02), ensemble method for predicting the pathogenicity of missense variants based on a combination of scores from 13 individual tools

- **sift** (1.2.0) predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids

- GnomADD aggregating and harmonizing both exome and genome sequencing data from a wide variety of large-scale sequencing projects

- PharmGKB

- dbSNP

The most important annotators for us are **ClinVar** and **dbSNP**.

1.3 Filters

Filters in OpenCravat/OakVar allow to select those variants which are relevant. As the number of variants in a genome usually is very large, you need to filter them first. OpenCravat cannot load more than 100,000 variants at once.

1.3.1 Filter tab

Select the **Filter** tab in the Result Viewer. There are the following sections:

•Samples

This is for oncological samples and is not used in Just-DNA-Seq.

•Genes

Here you can type in any particular gene names, one per line. Also you can load them from a file by clicking **Browse...**

•Variant properties

This section has 2 subtabs in its turn:

••Smart Filters

Here are various useful filters:

Population AF «==*

Sequence Ontology

Chromosome

Coding

ClinGen

ClinVar

dbSNP Common ID

Revel Rank Score >=

SIFT Prediction

••Query Builder

Here you can create a set of filter rules.

By default, an opening (left) parenthesis appears with buttons **+** and **(** in the lower left corner, and a greyed out switch **NOT** appears if you hover the mouse in the upper left corner, which allows to make the following rule negative by clicking on it. Clicking **NOT** once again deactivates it.

Click **+** to add a rule. A line of boxes will appear:

The first drop-down box is the source to which the rule will apply. For example: Variant Annotation, ClinVar, PharmGKB etc. The second drop-down box allow to select an item in the source to apply the rule. E.g. UID, Chrom, Position, Gene etc. The following switch "not", greyed out (inactive) by default, allows to select if the following condition should apply or should not apply. For the latter, click the word "not", and it will become black (active). To remove "not" from the condition, just click it again, and it will be greyed out. The next drop-down allows to select the condition from one of the following: **has data** - if the item being searched contains any data **equals** - opens one more box where you can enter what the item should be equal to **is empty** - if the item being searched is empty **in range** - opens two boxes where you enter the boundaries of the range where the item should be **<=** - if the item is less or equal to the value in the following box **>=** - if the item is greater or equal to the value in the following box At the end of the line, a small "x" allows to delete the whole rule by clicking on it.

If you click **+** once again, another rule is added, and between them the operator **and** is displayed by default, meaning that to satisfy the filter, both rules should apply. You can change it to **or** by clicking on it, so that to satisfy the filter, one of rules being true may be enough. Clicking **or** once again turns it to **and** again.

You can add as many rules as you wish, and the operators **and / or** between them will follow the general priority logic of boolean operations, i.e. **and** has the priority over **or**, as in any program code.

To change the priority and build more complex logical rules, you can click **(** which makes a separate set of rules (in parentheses), which have higher priority, as in mathematical operations. Note than the **and / or** operator which appears before the parentheses depends of the previous operator selected, i.e. if it was **or**, the next one will also be **or**, and vice versa. You can always change the operators by clicking on them.

Within the parentheses, you can create any number of rules, and there are separate **+** and **(** buttons to add new rules and nested parentheses inside the parentheses. Also in the upper left corner a separate **NOT** switch appears if you hover the mouse over it.

You can also move any rule to another rule. To do this, drag an anchor **||** which appears from the left side of the rule if you hover the mouse there, and drop it on any rounded **+** anchors which appear between rules and/or parentheses (not on the **+** button that adds rules).

BUG: if dropping a rule on certain location, it redirects browser to an error page

1.3.2 Clearing Filters

Under any section you can click the **Clear** button to remove any filter settings from that section.

1.3.3 Applying the Filter

When you finish building the filter, click **Apply Filter** in the lower right corner of the page. The number of variants in the lower left corner (the first number before the slash, while the second one is the total number of variants and is not changed) should now become smaller. These numbers also appear in the **FILTER** tab header, making it e.g. **FILTER 22/4,727,413**. If the number is small enough, you can switch to the **VARIANT** tab, or make the filter more strict to reduce the number.

1.3.4 Saving and Importing Filters

You can save the filter (the whole set of rules) in OpenCravat/OakVar for further loading, as well as exporting to a file, or import it from a file.

To save the filter, click the floppy disk icon in the lower right corner of the page, and enter the name.

NOTE: Filters are saved internally in OpenCravat/OakVar, i.e. on the server if using a remote installation. To have a filter saved into a local file, export it after saving.

The saved filter appears in the left part of the page in the **Saved Filters** list. To load a saved filter, just click its name. To export a saved filter into a file, click the icon with a down arrow next to its name. To delete a saved filter, click the X icon in its line.

To import a filter from a file, click the "up error" (rightmost) button in the lower right corner of the page, and browse for a file to import (e.g. pathogenic.json). Clicking **Open** in the browse window loads the filter. **NOTE:** the filter is not saved automatically, you need to save it with "Save filter" (floppy) icon if you want to save in on the server for further working.

1.4 Working With Annotation Information

After applying the necessary filters, click the **Variant** tab.

You will see a view with a table in the upper part and widgets below. You can move the border between them or switch to another view using the button in the upper right corner of the window:

View table - switch to table-only view View detail pane - switch to widgets-only view View table - switch to the default view (both)

For our purposes we'll need the table first of all.

The table contains columns and column sets with general info about the filtered variants, as well as connected to certain annotators. Some logically grouped column sets (by a particular annotator) can be extended or collapsed by clicking the +/- sign in the upper right corner of the column set (the topmost row). If you filtered by particular annotators, especially using "has data" condition, for other annotators it may show nothing for that particular variants, and they can be collapsed for convenience.

Each row of the table represents a variant that you can research.

1.4.1 Variant Annotation

UID - the variant number in this (filtered) sequence

Chrom - chromosome where the variant is located. Chromosome names are 'chr1' to 'chr22', 'chrX', 'chrY' and 'chrM'.

Position - Chromosomal position of the variant. THe first position in each chromosome is position 1.

Ref Base - Reference allele at this chromosomal position (one of A,C,G,T, and N).

Alt Base - Alternative allele; called based on reads mapping to this chromosomal position.

Note -

Coding -

Gene -

Transcript -

RefSeq -

Sequence Ontology -

cDNA change -

Protein Change -

All Mappings -

Sample Count -

Samples -

Tags -

1.4.2 CADD Exome

Score -

Phred -

1.4.3 CardioBoost

1.4.4 Cardiovascular Disease Knowledge Portal

1.4.5 ClinPred

1.4.6 ClinVar

Clinical Significance -

Disease Ref Nums -

Disease Names -

Review Status -

ClinVar ID -

Significance Detail -

1.4.7 dbSNP

rsID - Database identifier („rs" number) of this variant in dbSNP.

This column is empty if the observed variant is not described in dbSNP. Such variants can be extremely rare variants or technical artifacts.

1.4.8 Extra VCF INFO Annotations

1.4.9 Gencode Gene Mapper

1.4.10 gnomAD Gene

1.4.11 GWAS Catalog

1.4.12 NCBI Gene

1.4.13 OMIM

1.4.14 Original Input

1.4.15 PharmGKB

1.4.16 PubMed

1.4.17 REVEL

1.4.18 SIFT

1.4.19 VCF Info

Phred -

VCF Filter -

Zygosity - Most likely zygosity of the variant this chromosomal position, computed from the observed variant frequency (column 8) and can be "FP/HET" (<15%), "HET" (15- 75%), "HET/HOM" (75-85%), or "HOM" (>85%).

Alternate reads - Number of reads showing the alternative allele.

Total reads - Total number of reads.

Variant AF -

Haplotype block ID -

Haplotype strand ID -

1.5 Use Cases

Blablabla Again blablabla Second blablabla annotators wtf?

1.6 For Advanced Users

Blablabla Again blablabla Second blablabla annotators wtf?