HGVA Home

Welcome to HGVA project!

Human Genomic Variation Archive (HGVA) is an open-access genomic variation resource that integrates variants from the main reference human projects. HGVA also adds valuable information such as variant annotation: consequence types, population frequencies, protein effect predictions, variant-associated phenotype, etc.

Overview

HGVA aims to provide a high-performance and scalable resource to store, query and visualise variants from main open-access human datasets. We have put special emphasis on making HGVA very responsive even with complex queries and to make the data available to researchers and bioinformaticians in three different ways: a rich web interface based on **OpenCB IVA**, client libs (Python, Java and JavaScript) and through a command-line. Also, different datasets are normalised and annotated using **OpenCB CellBase**.

HGVA does not intend to replace or provide full archiving services like **NCBI dbSNP** or **EMBL-EBI EVA** p rojects, these provide excellent submission and accessioning services and play a crucial role allowing scientists to submit variation data for many different species. Instead, HGVA is focused on human and only the most relevant datasets are selected and indexed. HGVA provides different high-performance user interfaces to allow researchers to query and visualise human datasets or to use this data in genomic pipelines.

We would like to thank very much the authors of the different projects such as **1000 Genomes Project** or **ExAC** (see below for a complete list) for making all these invaluable data open and freely accessible for the biomedical community, we hope HGVA will help to make all these data more accessible to them.

HGVA was born during 2016 as a response to the necessity of having most relevant human datasets centralised, normalised and annotated for different analysis pipelines. It is currently developed and maintained by researchers at University of Cambridge and Genomics England and it is freely available at **http://hgva.opencb.org**

Main Features

- Most important high-quality variant studies normalised and integrated in one single server database
- High-performance complex queries to variants. Faceted seach also implemented
- Datasets are organised in four main projects *Reference Studies GRCh37*, *Reference Studies GRCh38*, *Cancer GRCh37* and *Platinum* (see below)
- Rich variant annotation performed using **OpenCB CellBase**, including HPO terms, consequence types, substitution effect prediction scores, Gene Ontology terms, etc.
- Population frequencies calculated, including populations and super-populations
- Data is indexed in the server using **OpenCB OpenCGA**. This provides a high-performance and scalable variant storage solution for *big data* analysis and visualisation
- Rich interactive web-based data mining tool based on **OpenCB IVA**
- Clients in Python, Java, JavaScript for fast programmatic access
- Command-line interface developed

Latest news:

HGVA 1.1.0 released!
Javier Lopez posted on Jun 09, 2017
Improved web interface appearance Web interface includes new beta features: Beta Genome Browser available to visualise variant context Summary (beta) tab to get a quick, visual description of filtering result Facets to get graphical descriptions of the data Upgraded backend. HGVA backend is now powered by OpenCGA 1.1.1 Try now! http://hgva.opencb.org http://hgva.opencb.org

HGVA paper published!
Javier Lopez posted on May 30, 2017
HGVA paper has been recently published. Please, cite: Lopez, J., Coll, J., Haimel, M., Kandasamy, S., Tarraga, J., Furio-Tari, P., Rendon, A, Dopazo, J & Medina, I. (2017). HGVA: the Human Genome Variation Archive. Nucleic Acids Research. https://academic.oup.com/nar/article/3848351

gnomAD population frequencies
Javier Lopez posted on May 30, 2017
gnomAD genomes and gnomAD exomes population frequencies are now provided as part of the advanced annotation tab! http://gnomad.broadinstitute.org/

Current Projects and Studies

Reference Studies GRCh37

- 1000 Genomes Phase 3
- Exome Aggregation Consortium (ExAC)
- Exome Sequencing Project (ESP6500)
- Genome of the Netherlands (GoNL)
- UK10K project
- Spanish Medical Genome Project (MGP)

Refernce Studies GRCh38

- 1000 Genomes Phase 3

Cancer GRCh37

Statistics

More than **250M** variants reported and about **120M** of unique variants

Developers

Source Code

Web based on IVA project at https://github.com/opencb/iva/tree/app/hgva

Server based on OpenCGA at https://github.com/opencb/opencga

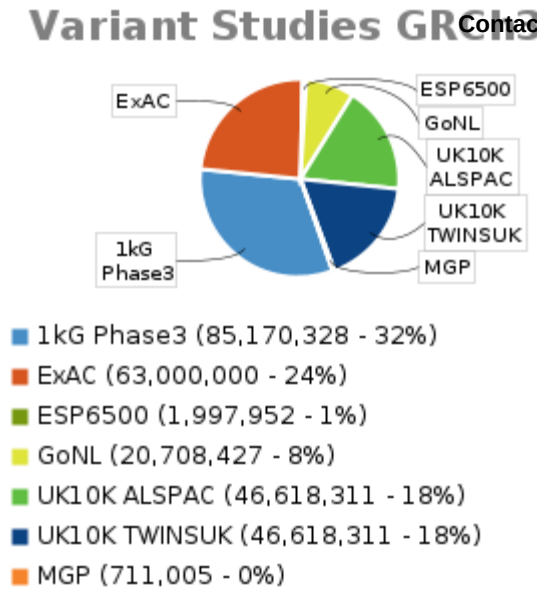
Contributing

HGVA is a collaborative project that aims to integrate as many reference human studies as possible, you can contact us for feature request. If you want to contribute to the code you are more than welcome to contribute to IVA and OpenCGA

- QIMR Berghofer Melanoma
- Chronic Myeloid Leukemia - Russian Academy of Medical Sciences

Platinum

- Illumina Platinum



Contact

Javier Lopez (*javier.lopez@genomicsengland.co.uk*)
Ignacio Medina (*im411@cam.ac.uk*)

Recent space activity

Nacho Medina
User Interfaces updated Apr 16, 2019 view change

Javier Lopez
Using the Python REST client updated Jul 04, 2018 view change
Using the Command Line updated Jul 04, 2018 view change

Nacho Medina
Datasets and Studies updated Jan 20, 2018 view change
RESTful Web Services and Clients updated Oct 24, 2017 view change