Overview

OpenCGA is an open-source platform that aims to provide a full stack solution for big data analysis and visualisation of genomic data. OpenCGA has been designed to provide a secure, high-performance and scalable solution for genomics analysis and visualisation.

OpenCGA implements a complete solution that covers all aspects of genomic analysis: metadata database, authentication and security, variant normalisation and aggregation, variant storage and annotation, highly scalable variant NoSQL storage engine, alignment and coverage, big data variant analysis, RESTful web services, visualisation

OpenCGA is developed and maintained in the University of Cambridge and it is currently used by several big data projects such as GEL (Genomics England).

Main Features

OpenCGA provides a complete solution for genomics data analysis:

- **Authenticated** and **secure** platform to query and visualise data, advanced **permission** system
- A **metadata database** to keep track of registered users, projects, studies, files, samples, families, jobs, ...
- You can store the **clinical data** for sample, patients or families
- **Alignment storage** allows to index BAM/CRAM, calculate index and query data and coverage
- The most advanced, high-performance and scalable **Variant storage** solution, you can normalise, load, index and aggregate thousands of whole genomes per day
- **Genomic Analysis** implemented on top of variant and alignment storage layer using advanced technologies such as Spark
- Full **clinical analysis platform** implemented, you can create the cases and run different clinical interpretations algorithms from your scripts or from a web application
- Comprehensive **RESTful web service API** with more than 150 endpoints to fully query and manage all metadata and clinical data
- Four different **client libraries** implemented in Java, Python, R and Javascript
- Interactive **web-based application** for the analysis and visualisation of variants and reads

Projects

OpenCGA is used by several projects being the most importan Genomics England (NHS).