**GA4GHclient: accessing federated genomic databases through GA4GH**

W. Souza1, B. S. Carvalho2, C. Rocha1, I. Lopes-Cendes1

1Department of Medical Genetics, School of Medical Sciences, 2Department of Statistics, Institute of Mathematics, Statistics and Computer Science; University of Campinas, UNICAMP, Campinas, SP, BRAZIL and the Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas, SP, BRAZIL.

**Introduction:** Federated databases consist in using local resources for hosting data instead of submitting data to centralized data servers [1]. This concept facilitates sharing genomic data with restricted access to sensitive information such as individuals’ information. It also avoids legal issues associated to data protection. Hosting data locally leads to difficulties in data integration from different databases because each provider may implement its own non-standard interface for publishing data. The Global Alliance for Genomics and Health (GA4GH) was formed to help accelerate the potential of genomic medicine to advance human health. They developed data model schemas and application program interfaces (APIs) for genomic data. These APIs are specifically designed to allow for standardized genomics data exchange. There are server-side implementations of these APIs for hosting genomic data via the Internet. Due to the lack of software for retrieving information from GA4GH-based data servers, we developed the GA4GHclient for accessing these federated data servers and retrieving genomic data through the GA4GH APIs. GA4GHclient provides a graphical web interface for easy data interaction and programming libraries for integrative data analysis development.

**Materials and Methods:** The GA4GH API schemas are developed continuously to address many issues regarding genomic and clinical data sharing. We used the latest version of these schemas for retrieving variant genomic and sequence alignment data from GA4GH-based data servers. We implemented our software in R programming language using Bioconductor packages for manipulating API data [2]. The web-based user interface was designed to show general data while omitting sensitive data such as patient information. We tested our software package by retrieving genomic data from 1000 Genomes Project, Ensembl, BRCA Exchange and the Brazilian Initiative on Precision Medicine (BIPMed).

**Results:** We developed GA4GHclient, a Bioconductor package that provides easy access to GA4GH-based public data servers. Our package provides programming tools for integrating genomic data from different databases. The programming framework contain tools for converting retrieved data to common file formats such as VCF. GA4GHclient also provides a graphical web application for interacting with genomics data with search engine by gene name and genomic location.

**Discussion:** Federated genomic databases that use the GA4GH API implementations combined with the GA4GHclient package open new opportunities for integrating genomic data from thousands of individuals. GA4GHclient’s graphical interface allows interacting with GA4GH-based databases. The graphical interface is also useful for making available genomic databases with restricted access to anonymous users.

**Conclusion:** We developed GA4GHclient, a Bioconductor package that provides programmatic access to GA4GH-based databases and a user-friendly web interface. The GA4GHclient package is freely available at https://github.com/labbcb/GA4GHclient.

**References:** [1] The Global Alliance for Genomics and Health, Science 352(6291): 1278-1280, 2016; [2] Gentleman RC et al., Genome Biology 5(10): R80, 2004.

Supported by FAPESP