

GenomeSpace: Open source interoperability platform with crowd-sourced analysis recipes

Ted Liefeld¹, Sara Garamszegi², Felix Wu², Marco Ocana¹, Barbara Hill², Helga Thorvaldsdottir², Michael Reich¹, Jill Mesirov¹

¹ University of California San Diego, La Jolla, CA, USA. Email: liefeld@ucsd.edu

² The Broad Institute of MIT and Harvard, Cambridge, MA, USA.

Project Website: <http://www.genomespace.org>

Source Code: <https://bitbucket.org/GenomeSpace/combined/>

License: GNU Lesser General Public License (LGPL).

Main Text of Abstract

GenomeSpace is a cloud-based environment that provides interoperability between best-of-breed computational tools, enabling scientists to easily combine tool capabilities without needing to program. It offers a common space to utilize, contribute, and share an ever-growing range of genomic analysis tools. GenomeSpace provides support for cloud-based data storage and analysis, multi-tool analysis scenarios, automatic conversion of data between tools, and ease of connecting new tools to the environment. GenomeSpace is free for all users and is open source under the GNU Lesser General Public License (LGPL).

The GenomeSpace architecture consists of (1) a server running in the Amazon cloud that manages communications and data transfer between tools, and data storage; (2) the collection of computational tools, updated to communicate with the GenomeSpace server while retaining their original user interfaces; and (3) connections to Amazon, Dropbox, and Google Drive cloud-based data services. The connection between server, tools, and data is provided by a RESTful API that can be easily adopted by new tools. A web-based user interface provides an integrated tool bar and data view from which users can launch tools, perform analyses, and transfer data. GenomeSpace also provides single-sign-on across compatible tools via OpenID.

We will describe the open source GenomeSpace platform and how it can be used to support genomic analyses that utilize multiple independent tools including GenePattern, Galaxy, Cytoscape, and IGV. We will also show how developers can add their own tools to the GenomeSpace ecosystem. We will also describe the newly released GenomeSpace Recipe Resource, a repository of short, standalone guides for performing integrative bioinformatic analyses. Each recipe walks users step-by-step through the process of obtaining and analyzing data across multiple GenomeSpace-enabled tools. The Recipe Resource provides the GenomeSpace community with the ability to create, share, and collaborate on analytic recipes of common value.