



# GenAp

Bringing computing to genetics researchers and clinicians

The genetic blueprint of an individual has never been more accessible due to the recent availability of advanced and inexpensive DNA sequencing technology. The study of genomics is of immense value in studying human ancestry, genetic diseases, and biological systems. It also provides a powerful new tool in clinical work, allowing medical diagnosis, treatment, and risk assessment based on an individual's genetic make-up. Unfortunately its impact on everyday clinical medicine has been modest to date because existing genomic tools are exclusively focused on the research community and difficult to use, and because working with large-scale genomic data is challenging.

Enter the Genetics and Genomics Analysis Platform (GenAP), a Platform that is not only furthering the advanced analysis of genomic data but also making this data much more accessible to biologists and physicians.

## Big, individualized data

A full set of one person's DNA data requires a stack of 50 DVDs while a large study with 1,000 patients can be hundreds of terabytes of data. This makes it impractical to transfer genomic data using traditional methods, challenging to store it, and virtually impossible to use it without the proper tools.

GenAP transparently pulls together hundreds of genomic software components and high performance computing systems into a single research tool with the requisite storage capacity and processing power. This distributed system keeps patient medical information behind a firewall using stringent privacy and access considerations while allowing widespread access to public genome datasets. The GenAP platform allows both researchers and clinicians alike to search and share genomic

data more easily, helping speed breakthroughs in research and treatment.

Discoveries in genomics will one day enable family doctors to routinely find rare genetic mutations and prescribe drugs specifically designed to correct them. Families afflicted by these rare disorders will be the first beneficiaries of these tools but everyone else will ultimately benefit, too. As it turns out, conditions that affect many of us – such as asthma, certain cancers, heart disease, and diabetes – may also be the effect of existing or new genetic variants in one's genome.

## Working together

GenAP is the result of collaborative work between the genomics and high-performance computing (HPC) groups at McGill University and the Université de Sherbrooke.

The GenAP Software Platform is available through the CANARIE Software Registry. This includes Services for managing large distributed files for genome processing and analysis, as well as the base set of 100+ genomic tools.

---

*GenAP transparently pulls together hundreds of genomic software components and high performance computing systems into a single research tool and allows both researchers and clinicians alike to search and share genomic data more easily, helping speed breakthroughs in research and treatment.*

---

# Technical Details

## Platform: Genetics and Genomics Analysis Platform (GenAP)

<b>Description</b>	A computing Platform for life sciences researchers that leverages both the CANARIE high-speed network and Compute Canada's High Performance Computing (HPC) resources by providing researchers access to modern and specialized Web services and analysis tools closely integrated to HPC resources.
<b>Creator(s)</b>	McGill University, Université de Sherbrooke
<b>Collaborator(s)</b>	Génome Québec, Compute Canada
<b>Research Subject</b>	Bioinformatics
<b>Managed Version<sup>i</sup></b>	Yes - available to Compute Canada users
<b>Self-hosted Version<sup>ii</sup></b>	No
<b>Cloud Support</b>	Integrated with Compute Canada's HPC infrastructure
<b>Host OS</b>	Linux
<b>Licence</b>	Various open source licences
<b>Details</b>	<a href="http://canarie.ca/software/genap-en">canarie.ca/software/genap-en</a>

## Contributed Services:

	<b>CVMFS Stratum 0 and Stratum 1 Services</b>	<b>Squid Cache VM Service</b>	<b>VM with Galaxy (Life Sciences Platform) Image</b>
<b>Description</b>	Used to distribute software over the web using CVMFS. No additional remote software administration is required.	Provides Squid resource caching for CVMFS service for rapid, robust software distribution.	Provides web access to the Galaxy Life Sciences Platform.
<b>Category</b>	Resource/Cloud Management	Resource/Cloud Management	Resource/Cloud Management
<b>Research Subject</b>	Genomics	Genomics	Genomics
<b>Managed Version<sup>i</sup></b>	Yes - used via the GenAP platform to configure software on remote servers.	Yes - used as a service for the GenAP platform to cache CVMFS content.	Yes - for use via the GenAP platform.
<b>Self-deployed Version<sup>ii</sup></b>	Yes - software available via Compute Canada.	Yes - software available via Compute Canada.	Yes - available as VM image via Compute Canada.
<b>Cloud Support</b>	Integrated with Compute Canada's HPC infrastructure	Integrated with Compute Canada's HPC infrastructure	Integrated with Compute Canada's HPC infrastructure
<b>Host OS</b>	Linux	Linux	Linux
<b>Licence</b>	GNU General Public Licence	GNU General Public Licence	GNU General Public Licence
<b>Details</b>	<a href="http://canarie.ca/software/cvfmfs">canarie.ca/software/cvfmfs</a>	<a href="http://canarie.ca/software/squid">canarie.ca/software/squid</a>	<a href="http://canarie.ca/software/galaxy">canarie.ca/software/galaxy</a>

<sup>i</sup> Managed version: Creators host a live instance of the software on their infrastructure, available for use by others

<sup>ii</sup> Self-deployed version: Users host a private instance of the software on their own infrastructure