

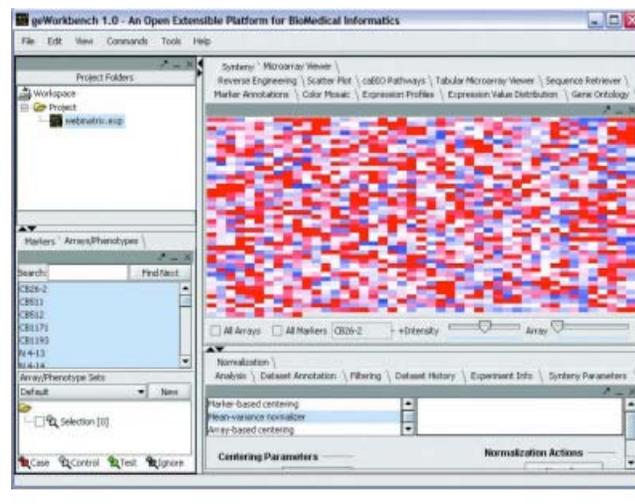
geWorkbench

Microarray gene expression and sequence data management

geWorkbench provides an innovative, open-source software platform for genomic data integration, bringing together analysis and visualization tools for gene expression, sequences, pathways, and other biomedical data. It gives scientists transparent access to a number of external data sources and algorithmic services, combining these with many built-in tools for analysis and visualization (at present more than 40 distinct analysis and visualization modules are part of the platform).

Features:

- Analysis and visualization tools for microarray-based gene expression profiling data from a variety of systems, including Affymetrix MAS5/GCOS; Matrix format (geWorkbench); tab-delimited simple spreadsheet (GEO series matrix, RMAExpress); and GenePix
- Analysis and visualization tools for gene and protein sequence data (FASTA); plug-in components including filter and normalize; promoter analysis; regulatory networks; differential expression; enrichment analysis; annotation; sequence analysis; and pattern discovery
- Support for pathways (BioCarta), gene ontologies, networks, and patterns based on regular expressions
- Integrated access to many external data sources and computational services (e.g., GoldenPath at Santa Cruz, NCBI BLAST, BioCarta diagrams through caBIO)
- Front-end to caArray and access-provider to caGrid-enabled computational resources
- Additional viewers including molecular structure and raw expression



geWorkbench interface

Categories of Use:

- | | | | |
|--|--|--|---|
| <input type="checkbox"/> Biospecimens | <input type="checkbox"/> Data Sharing | <input type="checkbox"/> Imaging | <input type="checkbox"/> Proteomics |
| <input type="checkbox"/> Clinical Trials Management | <input type="checkbox"/> Genome Annotation | <input checked="" type="checkbox"/> Microarrays | <input type="checkbox"/> Translational Research |
| <input checked="" type="checkbox"/> Data Analysis & Statistical Tools | <input type="checkbox"/> Infrastructure | <input checked="" type="checkbox"/> Pathways | <input type="checkbox"/> Vocabularies |



Architecture Overview

- **Application type:** geWorkbench is a desktop application that also provides access to remote data and analysis services.
- **System requirements:** geWorkbench is downloaded as a self-installing package with support for Windows, Linux, and MAC OS-X; it requires installation of Sun Java J2SE 5.0 or higher (Java 1.5 JRE).

Installation and Administration:

- **Skill sets needed:** Users should have basic computer skills, plus familiarity with basic biology and the goals of the data analysis one wishes to perform.
- **Infrastructure needed:** Installation of Java 1.5 JRE is required before installing geWorkbench; a Windows version of geWorkbench can be downloaded with its own local Java 1.5 JRE included.
- **Long-term administration needs:** None is required; however, some modules maintain local copies of genomic or sequence information that one may wish to update periodically (or wait for the next geWorkbench release).

Key Contributors:

- Center for Computational Biology and Bioinformatics at Columbia University
- Jackson Labs
- Northwestern University
- Washington University

Other Life Sciences Distribution Components:

- caArray
- Cancer Genome-Wide Association Studies (caGWAS)
- caTissue Core
- Clinical Trials Object Data System (CTODS)
- National Cancer Imaging Archive (NCIA)

Resources

Tool Overview Page	https://cabig.nci.nih.gov/tools/geWorkbench
Primary Workspace	Integrative Cancer Research (ICR) https://cabig.nci.nih.gov/workspaces/ICR
Molecular Analysis Tools Knowledge Center	https://cabig-kc.nci.nih.gov/Molecular/KC/index.php/Main_Page
geWorkbench Email List	http://gforge.nci.nih.gov/mailman/listinfo/geworkbench-announce
caBIG[®] Tool Inventory	https://cabig.nci.nih.gov/inventory
NCI Center for Bioinformatics Applications Support	ncicb@pop.nci.nih.gov



NIH Publication No. 08-6528

Printed June 2008

**Online Version Updated
October 2008**

