

Cancer Genome-Wide Association Studies (caGWAS)



Genome-wide data management system

Cancer Genome-Wide Association Studies (caGWAS) allows researchers to integrate, query, report, and analyze significant associations between genetic variations and disease, drug response, or other clinical outcomes. caGWAS accelerates the process of analyzing results from various biomedical studies and allows researchers and bioinformaticians to access and analyze clinical and experimental data across multiple clinical studies.

Features:

- Standardized model to represent SNP genotype data, SNP association findings, population frequency data, and clinical phenotype
- Support for search and retrieval of genome-wide association findings in the context of genes or chromosomal regions of interest
- Allows users to load GWAS studies and provides powerful search capabilities for small datasets or bulk downloads of large genotype files
- Allows for genome-wide SNP association studies
- Provides touch points to clinical and specimen annotations through the caBIG™-compatible interface

The screenshot shows the 'caGWAS/data portal' interface. At the top, it features the National Cancer Institute logo and navigation links like 'Home', 'Login', and 'Contact Us'. The main content area is titled 'Browse Data:' and includes a 'Select a Study:' dropdown menu with options 'GWAS STUDY I', 'GWAS STUDY II', and 'GWAS STUDY III'. Below this is a 'Select a Version:' dropdown menu. A 'Study Description:' box contains the text: 'There are multiple studies available. Please review the context-sensitive description when you use the list above to select a specific study.' At the bottom, there is a 'Select a Dataset:' section with radio buttons for 'Association Finding' (selected) and 'Population Frequency'.

caGWAS 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 checked="" type="checkbox"/> Genome Annotation | <input type="checkbox"/> Microarrays | <input checked="" type="checkbox"/> Translational Research |
| <input type="checkbox"/> Data Analysis & Statistical Tools | <input type="checkbox"/> Infrastructure | <input type="checkbox"/> Pathways | <input type="checkbox"/> Vocabularies |



Cancer Genome-Wide Association Studies (caGWAS)

Genome-wide data management system



Architecture Overview

- **Application type:** Open-source, Web-based application based on J2EE middle-ware technologies such as n-tier service oriented architecture and software design patterns
- **System requirements:** Java Software Development Kit (JDK) version 1.5.0_10; JBoss Container (JBoss version 4.0.4 recommended); Jakarta Ant version 1.7; Oracle 9i; MySQL 5.0.27; FTP server

Installation and Administration:

- **Skill sets needed:** Experience with Unix/Linux environment (configuring environment variables); Installing Ant; JDK; Apache and JBOSS server; Ant build scripts; J2EE Web application development using the Struts framework, Servlet/JSPs, JavaScript, AJAX, and XML/XSLT; J2EE middle-ware technologies such as n-tier service oriented architecture and software design patterns; database administration to properly configure the database (Oracle 9i or MySQL)
- **Infrastructure needed:** caGWAS is a read-only application that allows you to query and retrieve your analyzed Genome-Wide Association Studies Data; all data for caGWAS must be loaded via DB scripts
- **Long-term administration needs:** Creation of users in CSM (Common Security Module) tables and continuous data loads using provided data load scripts

Key Contributors:

- Lockheed Martin
- NARTec
- NCI Center for Bioinformatics
- Science Applications International Corporation
- SenPro
- TerpSys

Other Life Sciences Distribution Components:

- caArray
- caTissue Core
- Clinical Trials Object Data System (CTODS)
- geWorkbench
- National Cancer Imaging Archive (NCIA)

Resources

Tool Overview Page	https://cabig.nci.nih.gov/tools/caGWAS
Primary Workspace	Integrative Cancer Research (ICR)
ICR Workspace Coordinator	Elaine Freund: efreund@3rdmill.com
caBIG™ Tool Inventory	https://cabig.nci.nih.gov/inventory
NCI Center for Bioinformatics Applications Support	ncicb@pop.nci.nih.gov



March 2008