LIFE SCIENCES DISTRIBUTION

caIntegrator

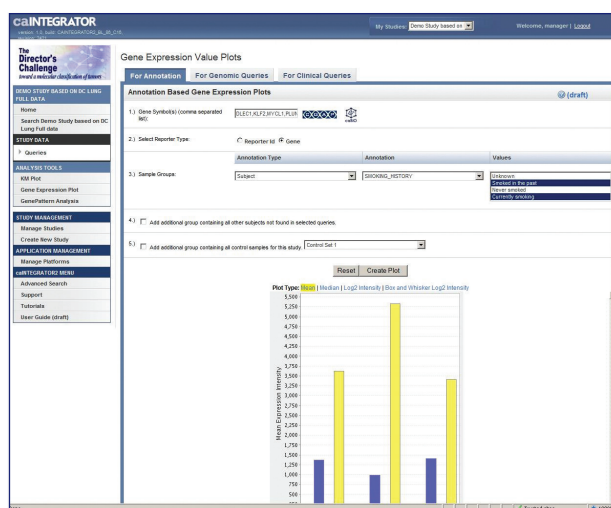## *A translational research tool for 21st Century Biomedicine*

caIntegrator is a Web-based software package that allows researchers to set up custom caBIG®-compatible Web portals to conduct integrative research without requiring programming experience. These portals bring together heterogeneous clinical, microarray and medical imaging data to enrich multidisciplinary research.

Using caIntegrator, researchers can execute, save and share queries to identify and collect many types of data, combining clinical information with genetic and genomic data to enable multidimensional analysis. caIntegrator uses caGrid analytical services such as GenePattern and BioConductor to perform analysis on the integrated study data, including copy number variation detection.

caIntegrator leverages the Cancer Data Standards Registry and Repository (caDSR) to map experimental data to well-defined datatypes and utilizes caGrid and Java client APIs to access data from caBIG® applications such as caArray and the National Biomedical Imaging Archive (NBIA). caIntegrator is also integrated with cancer Biomedical Infrastructure Objects (caBIO) to perform queries on genes and pathways.

Capabilities

- Web-based deployment and study management
- Remotely access multiple caGrid-accessible datasets including integration with caArray and NBIA
- Map study annotations to caDSR datatypes
- Create lists of genes, subjectIDs, and other annotations that can be queried and used for complex analyses
- Run complex queries across clinical, microarray and imaging datasets
- Create and save pre-computed views of the data

caIntegrator interface

Categories of Use

☐ Biospecimens	☐ Data Sharing	■ **Imaging**	☐ Proteomics
☐ Clinical Trials Management	■ **Genome Annotation**	■ **Microarrays**	■ **Translational Research**
■ **Data Analysis & Statistical Tools**	☐ Infrastructure	■ **Pathways**	☐ Vocabularies

caIntegrator
A translational research tool for 21st Century Biomedicine

(capabilities continued)

- Create Kaplan-Meier survival curves based on clinical or gene expression data

- Search caBIO with complex annotation search capabilities

- Analyze gene expression data from Agilent platform

- Restrict studies to internal research teams or open data for public display and analysis

- Leverage existing grid capabilities for data analysis through integration with GenePattern and BioConductor

Architecture Overview

- **Application type:** Open-source, Web-based J2EE Java application employing Spring, Struts 2 and Hibernate technologies.

- **System requirements**: Full installation requires installation of JBoss; MySQL; ANT (Apache); JDK.

Installation and Administration

- **Skill sets needed:** Linux/Windows server administrator, MySQL database administrator, software engineer with experience in Java desirable.

- **Infrastructure needed**: Windows/Red Hat Enterprise Linux-based server (32- or 64-bit).

- **Long-term administration needs**: caIntegrator requires ongoing system and database administration such as disk space, database management and database and file system backup. Application maintenance is required for user account management.

Key Contributors

- NCI Center for Biomedical Informatics and Information Technology (NCI CBIIT)

- 5AM Solutions

- Science Applications International Corporation (SAIC)

- ScenPro

- ESAC

Resources

Tool Overview Page	https://cabig.nci.nih.gov/tools/caIntegrator
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
caBIG® Tool Inventory	https://cabig.nci.nih.gov/inventory
NCI Center for Bioinformatics Application Support	ncicb@pop.nci.nih.gov