

ICR Project Summary

SIG	Project Name	Project Summary	Name	Role	Center
Data Analysis and Statistical Tools	caWorkbench 2.0	caWorkbench 2.0 (also known as BioWorks) is a bioinformatics platform that makes available tools for data management, analysis and visualization to the community in a convenient fashion. Some of the most fully developed capabilities of the platform include microarray data analysis, pathway analysis and reverse engineering, sequence analysis, transcription factor binding site analysis and pattern discovery. caWorkbench allows structured communication between disparate software modules which may be written with little or no knowledge of each other. For caBIG, caWorkbench 2.0 will be fully documented and integrated with caArray.	Andrea Califano	Developer	Columbia
	Distance-Weighted Discrimination	DWD is a tool that performs statistical corrections to reduce systematic biases that are due to different sources of RNA, different batches of microarrays and especially different microarray platforms. This tool, currently implemented in MatLab, will be ported to an open source, caBIG-compliant platform this year.	Steve Marron	Developer	Lineberger
			Everett Zhou	Developer	Lineberger
			Michael Nebozhyn	Adopter	Wistar
			Michael Showe	Adopter	Wistar
	GenePattern	GenePattern is a flexible analysis platform developed to support multidisciplinary genomic research programs. Its architecture and environment are expressly designed to allow rapid prototyping and integration of new technologies. For caBIG, an adapter to GenePattern will be created to allow the application to read data directly from caBIO/caArray.	Ted Liefeld	Developer	MIT/Broad
			Michael Reich	Developer	MIT/Broad
			Jomol Mathew	Adopter	New York
			Judith Goldberg	Adopter	New York
	Magellan	Magellan is a web-based system that allows biologists to perform complex analyses on heterogeneous data in an environment that does not require a background in computer programming or statistics. Stored data and annotations are treated as abstract entities such that arbitrary, user-defined types of information can be stored. In the context of caBIG, Magellan will consume of data from caBIO/caArray and adapt its architecture to Silver-level compliance.	David Fenstermacher	Adopter	Penn
			Craig Street	Adopter	Penn
			Chris Kingsley	Developer	UC San Francisco
			Ajay Jain	Developer	UC San Francisco
	VISDA	Visual and Statistical Data Analyzer (VISDA) has been developed with the goal of revealing all of the interesting patterns in a dataset. The main application of VISDA is for multivariate cluster modeling, discovery, and visualization, particularly for data sets living in high dimensional space. Currently implemented in MatLab, VISDA will be ported to a caBIG-compliant, open-source architecture.	Joseph Wang	Developer	Georgetown
			Malik Yousef	Adopter	Wistar
Michael Nebozhyn			Adopter	Wistar	
Michael Showe			Adopter	Wistar	
Cancer Molecular Pages	This system will be based on technology originally developed by the Joint Center for Structural Genomics, to aid cancer researchers in with keeping up with all of the information being generated on a gene or gene set of interest. It is a fully functional database and automated annotation system combining automated computer-based annotations and automated data collection from experimental stations and set of web based visualization tools. This system will be ported to a caBIG-compliant architecture this year and will consume of data from caArray.	Kutbuddin Doctor	Developer	Burnham	
		Edith Zang	Adopter	IFCP	
		Rakesh Nagarajan	Developer	Wash U	
		John Rux	Adopter	Wistar	
FunctionExpress	FunctionExpress is an environment for the integrated analysis and visualization of complementary data sets. The system provides a mechanism for regular updates of integrated annotation data that can be readily associated with probes on microarrays. The application allows for plotting of raw and transformed microarray data, for viewing orthologous probesets from other experiments, and for creating literature-based gene networks. This system, currently implemented in a 2-tier, proprietary architecture will be ported to an open source, caBIG-compliant architecture this year.	Harold Riethman	Adopter	Wistar	
		David Kane	Developer	NCI - CCR	
		John Rux	Adopter	Wistar	
GOMiner	GOMiner leverages the Gene Ontology (GO) to identify the biological processes, functions and components represented in gene lists. Instead of analyzing microarray results with a gene-by-gene	David Kane	Developer	NCI - CCR	
		John Rux	Adopter	Wistar	

ICR Project Summary

Genome Annotation	GOMiner	approach, GoMiner classifies the genes into biologically coherent categories and assesses these categories. For caBIG, GoMiner will be adapted to a caBIG-compliant architecture and made available as a web service.	Harold Riethman	Adopter	Wistar
	HapMap, PromoterDB	The HapMap database is a repository of human SNPs, their genotypes, and the linkage disequilibrium relationships among them. The Vertebrate Promoter Database (VPD) is a curated resource for vertebrate transcription factor binding sites and their corresponding regulatory regions. Both of these datasets will be made available through caBIO in this caBIG project.	Lincoln Stein	Developer	Cold Spring
			Alex Lash (PromoterDB)	Adopter	Sloan
	PIR	The Protein Information Resource (PIR) is an integrated public bioinformatics resource that supports genomic and proteomic research. PIR maintains the Protein Sequence Database (PSD), an annotated protein database containing over 283,000 sequences covering the entire taxonomic range. PIR is also a member of the UniProt consortium, the central international resource of protein sequence and function that unifies the PIR, Swiss-Prot, and TrEMBL databases. For caBIG, the PIR database will be grid-enabled to demonstrate how such a datasource can be discovered and consumed in a grid environment.	Harold Riethman (HapMap)	Adopter	Wistar
			Cathy Wu	Developer	Georgetown
Craig Street			Adopter	Penn	
SEED	SEED is a framework that supports peer-to-peer annotation of genomes. Investigators can work independently on their own instances of the SEED database and synchronize their work when desired or update code versions quickly via the network. Several hundred microbial organisms are in SEED now and pipelines for high-throughput processing, e.g., BLASTing, exist for rapidly including new organisms. For caBIG, access to curated SEED data will be made available through a caBIG-compliant interface. In addition, eukaryotic data will be added to SEED.	David Fenstermacher	Adopter	Penn	
		Cathy Wu	Adopter	Georgetown	
Microarray Repositories	caArray	NCICB is developing their next generation microarray data repository, termed caArray. Phase I of the system, to be released in September, will be caBIG-compliant and will contain the following key features: MIAME 1.1 compliant; support for MAGE-ML import and export; utilities for the submission and retrieval of Affymetrix and GenePix native file formats; use controlled vocabularies; accessibility through a MAGE-OM API. Many Centers in the ICR Workspace are interested in adopting caArray.	Ed Frank	Developer	Holden
			Arnie Miles	Adopter	Georgetown
			Michael Showe	Adopter	Wistar
			Walter Mankowitz	Adopter	Wistar
			Wen Hwai Horng	Adopter	Wistar
			Jomol Mathew	Adopter	New York
	Judith Goldberg	Adopter	New York		
	NCI-60 Data Sharing	The NCI-60 are 60 diverse human cancer cell lines used by the NCI Developmental Therapeutics Program to screen >100,000 chemical compounds since 1990 for anticancer activity. The Weinstein and other laboratories have assessed potential molecular targets and modulators of activity in those cells one or a few at a time and have developed a number of "omics" databases for this data. For this project, these datasets will be made available through standards and interfaces developed/approved by caBIG.	David Kane	Developer	CCR
TBD			Adopter	TBD	
Zebrafish microarray data sharing	The Thomas Jefferson University Zebrafish Microarray Service provides microarray data generation services for the community of scientists using zebrafish as a model organism for the study of cancer. They have developed a custom microarray using the commercial Zebrafish oligo library (Compugen/Sigma-Genosys). Through this caBIG project, Thomas Jefferson will make datasets from this repository available to the research community by identifying sharable datasets, creating a web application to select these datasets, and making these datasets available via a caBIG-defined exchange format.	Alex Lash	Adopter	Sloan	
		Jack London	Developer	Thomas Jefferson	