

# Molecular and Cellular Characterization of Screen-Detected Lesions -Coordinating Center and Data Management Group



Christopher Amos, Ph.D.

# The CDMG Team



Kristen Anton, M.S.



Amar Das, M.D., Ph.D.  
(Psychiatry)



Eugene Demidenko, Ph.D.



Tor Tosteson, Ph.D.



Judi Forman, M.S.



Dakai Zhu, M.S.



Lamar D. Moss, B.A.

# Goals

- **To provide network coordination.** This aim includes support for meeting development, conference call support and documentation of all meetings and protocols as these develop. An interactive web interface will be maintained to facilitate communications among investigators and with NCI and the coordinating team.
- **To provide statistical and computational support.** For this aim we will provide assistance to investigative teams for development and application of state of the art statistical methods for biomarker studies. Additionally, we will facilitate research across the network by providing access to the extensive computational resources available at Dartmouth.
- **To support the development of cross network studies by developing new protocols and databases that will support careful collection of new data.**



# Provide Network Coordination

## Meeting planning

Experience as Secretary-Treasurer of International Genetic Epidemiology Society for 5 years – planning meetings, working with hotel contracts

Experience as coleader with Brian Henderson of the Genetic Associations and Mechanisms in Oncology – for 5 years

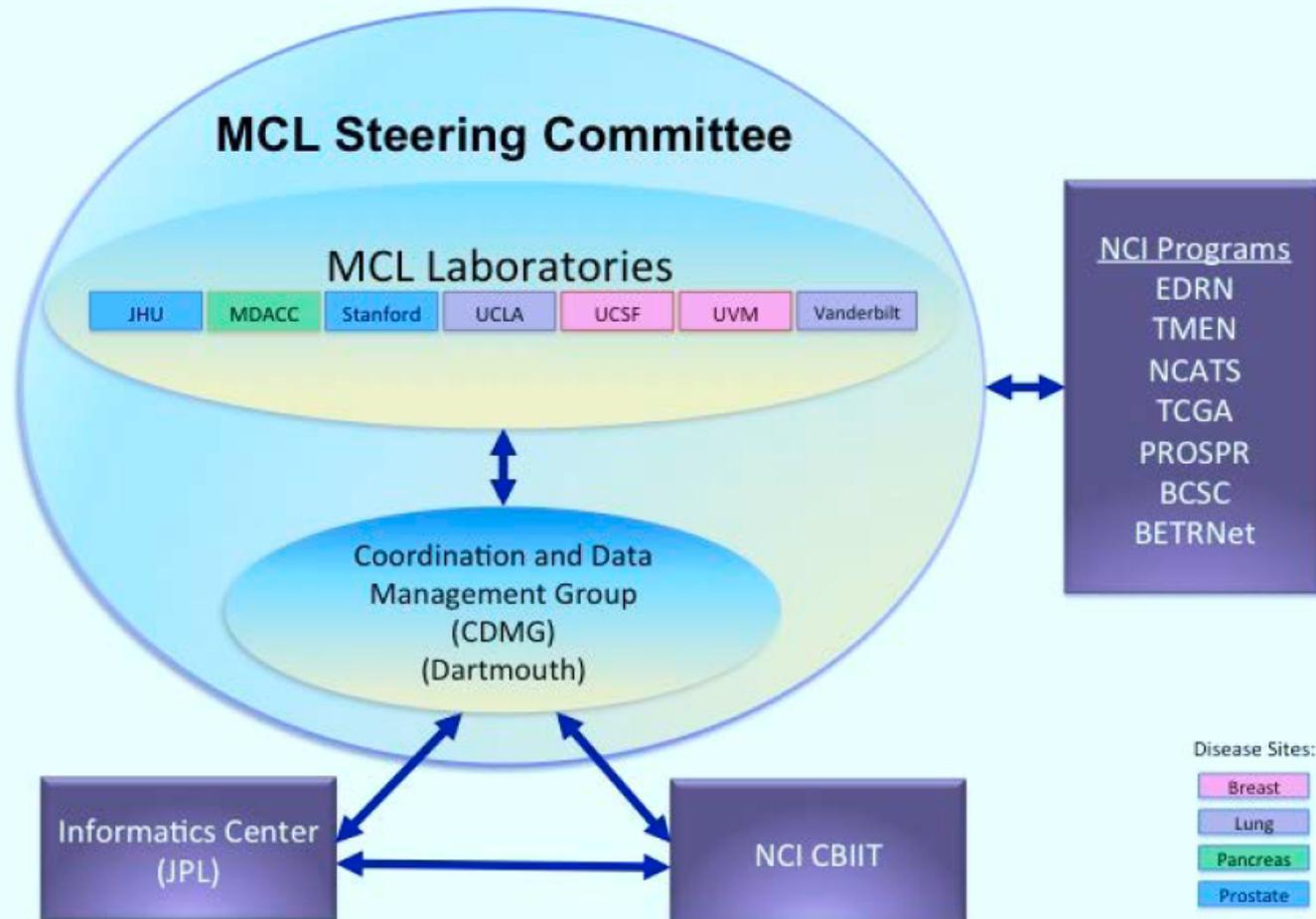
Supplemental funding derived for donor funds from Dartmouth which will facilitate travel (contracts and food for meeting) and helps support statistical elements

Set up annual meeting October

<https://www.eventbrite.com/e/mcl-consortium-orientation-meeting-tickets-18251183782>



# Organizational Structure of the MCL Consortium



# Provide statistical and computational support.

**Sample Size Calculations for Logistic Regression with Exposure Measurement Error**

This program provides sample size calculations for logistic regression with a continuous exposure variable and an additional continuous covariate or confounding variable. A classical measurement error model is assumed for this implementation. Please refer to the following reference for details on the methods used:

Power and sample size calculations for generalized regression models with covariate measurement error.  
 Tor D. Tosteson, Jeffrey S. Buzas, Eugene Demidenko, Margaret Karagas. *Statistics in Medicine* Volume 22, Issue 7, 2003. Published Online: 10 Mar 2003

Odds ratio for a one standard deviation increase in the true exposure:

Prevalence of the binary outcome at the mean for the exposure and the covariate:

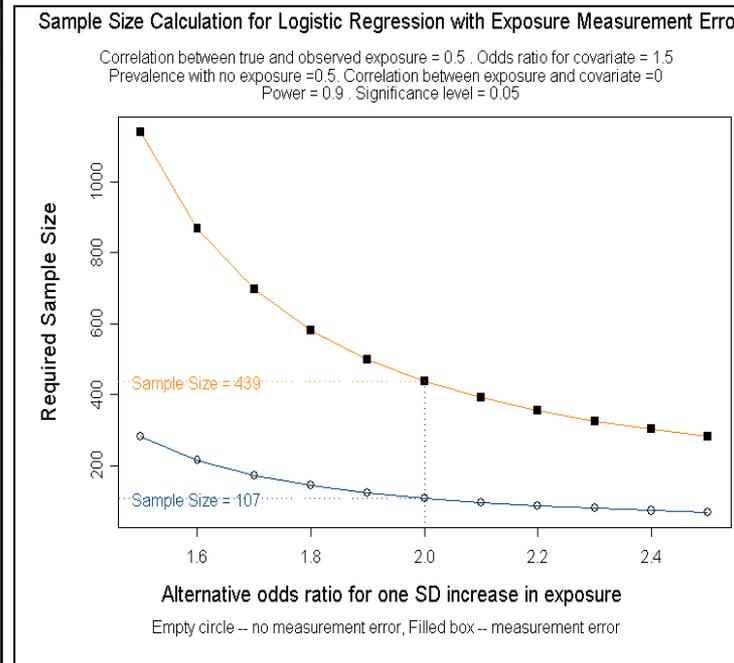
Power of score test:

Two-sided significance level:

Correlation between the true and the measured exposure:

Odds ratio for a one standard deviation increase in the covariate:

Correlation between the covariate and the exposure:



# Provide Network Coordination

Facilitate Development of a Manual of Operations:

Develop Publication Policies: Adapted from existing policies derived from GAME-ON initiative or EDRN where applicable

Assist with IRB-related issues

Develop Data Sharing Plans

Assist with Developing Biobanking Protocols

Establish Cross-Collaborations by developing working groups



Figure 1. Statistical Methods Server Example: Sample Size Calculations with Exposure Measurement Error

Software | Center for x

← → ↻ <https://morgan.dartmouth.edu/node/3>

Apps Microarray Sam... northern stage ... richard barth h... Scientific Journ... Journal / Autho... New Tab



[Home](#)

## Software

posted by dzhu on Mon, 01/21/2013 - 18:08

### Software Available in Open Systems

Program	Information/Documentation
<a href="#">454</a>	454 Sequencing tools
<a href="#">act</a>	Analysis of Complex Traits, a group of programs we wrote.
<a href="#">alleqro</a>	Version 2.0f is available as of April 2011. <a href="#">Documentation</a> is available too. Version 1.2c is also available, in case anyone needs it. Its <a href="#">documentation</a> is also posted.
<a href="#">ancestrymap</a>	Version 2-rel is available as of April 2011. User should go to the author's Web site for documentation.
<a href="#">APT</a>	Affymetrix Power Tools version 1.15.0. This group of programs is installed on Linux servers
<a href="#">beaglecall</a>	Version 0.9.4.
<a href="#">bimbam</a>	Version 1.0 is available as of February 2, 2013. It was built to support parallel computation with

### Navigation

- ▶ [Chaos Tools AJAX Demo](#)
- [Form builder example](#)
- ▶ [Forums](#)



# To provide statistical and computational support.



HOME **SOFTWARE** FREE TRIAL SALES PUBLICATIONS SUPPORT CONTACT US



## Statistical Analysis Software You Can Count On

Partek Genomics Suite is the software scientists worldwide have come to trust for their genomic data analysis and visualization needs. In 2013 alone, it was featured over 500 peer-reviewed publications. It's fast, agile, and memory efficient, allowing you to analyze large data sets right from your desktop. With its user-friendly interface, comprehensive workflows, and ability to support all next generation sequencing technology, and NGS platforms, Partek Genomics Suite gives biologists, bioinformaticians, and statisticians a single, integrated solution for NGS quality results. Get your free demo today and see what Partek Genomics Suite can do for you.

## Guided Analysis Workflows

Partek Genomics Suite software offers an intuitive user interface and built-in workflows for a variety of genomic applications that guide researchers through every step of the analysis process.

### Workflows include:

#### Microarray

- Gene expression
- ChIP-chip analysis
- Biomarker discovery
- Copy number
- Whole genome copy number

#### Next Generation Sequencing

- SNP-Seq
- ChIP-Seq
- ChIP-Seq
- RNA-Seq
- ChIP-Seq



## SNP & VARIATION SUITE

YOUR DATA. YOUR ANALYSIS. YOUR DISCOVERY.



SNP & Variation Suite is a powerful analysis tool created specifically to empower biologists and other researchers to easily perform complex analysis and visualizations on genomic and phenotypic data. With SVS you can focus on your research instead of learning to be a programmer or waiting in line for bioinformaticians.

Released MyTool! Released MyTool!

### Genome-Wide Association Studies (GWAS)

SVS includes a broad range of analysis tools built to improve your data quality and easily perform quality assurance and statistical tests for genetic association studies.

### Small Sample DNA Sequencing Workflows

SVS gives users access to the latest association sources for filtering and analyzing raw variants from microarray analysis pipelines to obtain a short list of potentially pathogenic variants.

http://www.clcbio.com/products/clc-genomics-workbench/

File Edit View Favorites Tools Help

Our website uses cookies. By using our website and agreeing to this policy, you consent to our use of cookies. [OK](#) [MORE INFO](#)

CLC bio A QIAGEN Company

Buy online CLC websites Subscribe to newsletter Search...

GET STARTED DESKTOP SOFTWARE ENTERPRISE PLATFORM PRODUCTS CONSULTING SUPPORT CORPORATE

# CLC GENOMICS WORKBENCH

A comprehensive and user-friendly analysis package for analyzing, comparing, and visualizing next generation sequencing data.

Introduction Features Screenshots Latest Improvements Download To Top

## Introduction

Analyze, compare and visualize NGS data

Dominating the high-throughput Feature overview



# Support Statistical Coordination

Reviewing submitted grants for places of synergy

Will reach out to groups to identify statistical/computational needs

Establish connections to Dartmouth where relevant

Can facilitate data transfer to dbGAP or among members



# Support Statistical Analyses

Reviewing submitted grants for places of synergy – potential development of collaborations

Will reach out to groups to identify statistical/computational needs

Establish connections to Dartmouth for computational support where relevant  
(available 2400 node computational cluster with various architectures)

Can facilitate data transfer to dbGAP or among members



# Network Support and Development

**To support the development of cross network studies by developing new protocols and databases that will support careful collection of new data.**

Develop procedures for development of new studies, management of the studies, process flow of samples and monitoring. Worksheets and information management systems for collection of data and specimen tracking in individual and multi-center Consortium studies, and verify all generated data deposited at the CDMG.

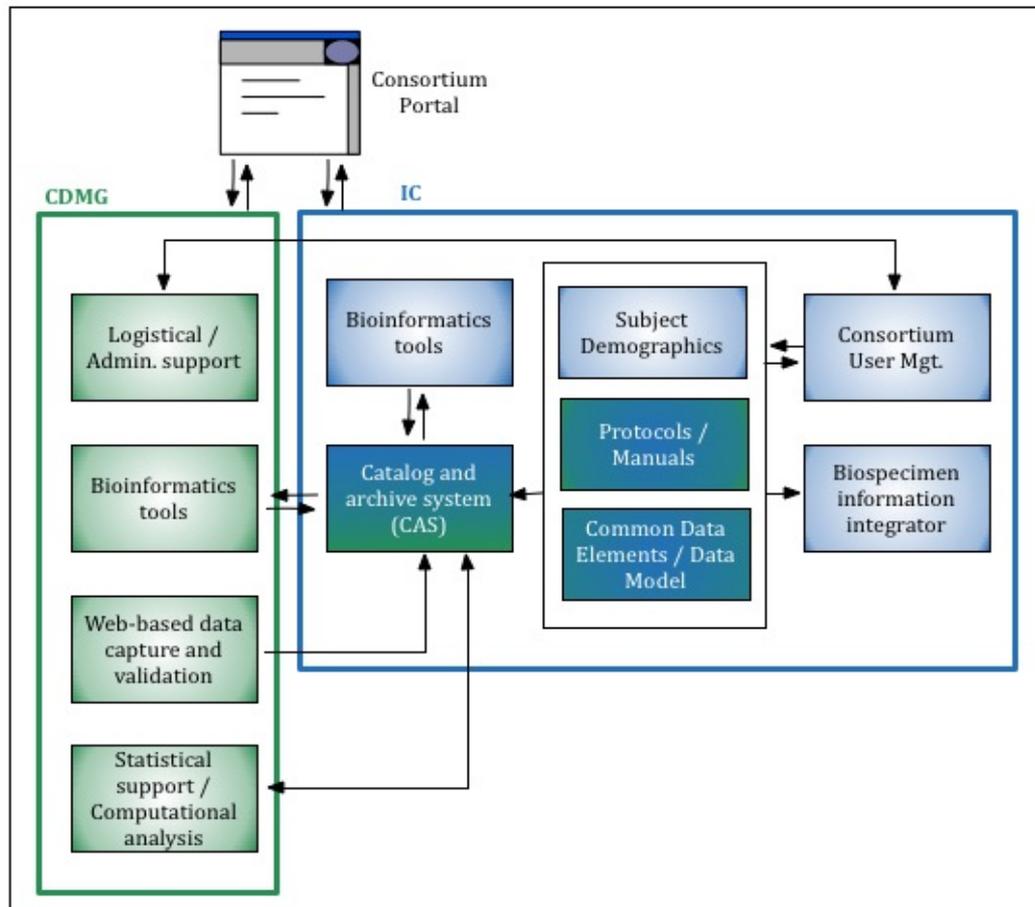
# Network Support and Development

## Five key informatics goals:

- Defining an information model for describing the Consortium data landscape;
- Collaborating with the EDRN-IC to adapt the EDRN Knowledge Environment for the needs of this Consortium, including searchable archives for protocols, research data, analytic data sets, a data capture and validation system to support proactive collection of data, system to support monitoring activities, and access to bioinformatics tools and pipelines;
- Providing a secure transfer and distribution infrastructure to meet HIPAA, collaborating institution, and United States federal regulations for data sharing; and
- Providing an integrated portal environment, based on the structure of the EDRN Portal, for access to Consortium operational information and to capture and distribute the Consortium data;
- Develop value return for researchers to motivate interaction with and participation in the Consortium Knowledge System.



# Network Support and Development



High level view of  
Relationship between  
CDMG and IC



# Process for Database Development and Support

- Data Model and Information Standards. The successful implementation of a Consortium Knowledge System depends upon the successful definition and adoption of an information model for describing the Consortium data.
- Definition and submission of Common Data Elements (CDEs)
- Portal. The centerpiece of the Consortium Knowledge Environment will be the Consortium Portal. The portal will provide access to all Consortium information and functional interfaces, including the interfaces through which all archived and distributed data resources are queried for retrieval of a data set. The portal will serve as the Consortium communication “hub” for all authorized Consortium users, providing access to information like Consortium operational information, publicizing opportunities within the Consortium, and posting minutes.
- Adapt Molecular Characterization Knowledge System Informatics Infrastructure.
- **Adapt an existing “data grid” technology** called the Object Oriented Data Technology (Apache OODT), developed by JPL and released as open source software to the Apache Software Foundation, to enable data and tool sharing and computational services.



# Additional Responsibilities for CDMG

- **Monitor Consortium protocol adherence, monitor data collection and submission, and report violations to the Steering Committee.**
- **Ensure that data are collected to determine the benefits and risks that follow from positive or negative test results.** Element of current review of proposals
- **Risk model performance evaluation:** graded by calibration, discrimination and accuracy. Calibration is most important and often overlooked element.
- **Technology Transfer** – Beyond scope of CDMG but also rests upon technical feasibility of biomarker assays. Data about technical requirements also needs to be curated.

