



Center for  
Causal  
Discovery

# Scientific Catalysts Program Kickoff Meeting The Center for Causal Discovery (CCD)

April 21, 2015

University of Pittsburgh  
Carnegie Mellon University  
Pittsburgh Supercomputing Center  
Yale University

PIs: Ivet Bahar, Jeremy Berg, Gregory Cooper



# Agenda

- Introductions to CAC team and Scientific catalysts
- Information on CCD – Presentation and Plan
- Goals for Scientific Catalyst program
  - Date for initial workshop in summer
  - How to get help for researchers/consortium to use CCD tools
- Next steps and timeline

# Introductions to CAC team and Scientific catalysts

## Scientific Catalysts

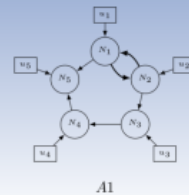
The CCD Scientific Catalysts will directly and personally promote the use of CCD software and algorithms and awareness of the BD2K Consortium among colleagues. They will also solicit feedback from their colleagues regarding the challenges faced with regard to Big Data in their research and with implementing CCD or other BD2K Consortium tools. Investigators who would like to serve as CCD Scientific Catalysts are encouraged to contact [Dr. Mike Becich](#).

- [Derek Angus, MD MPH](#) – Clinical Investigation and Systems Modeling of Acute Illness
- [Yuan Chang, MD](#) – Cancer Virology and Viral Oncogenesis
- [Mary Edgerton, MD, PhD](#) – Mathematical Modeling of Cancer Development and Progression
- [Kevin Gibson, MD](#) – Interstitial Lung Disease and GRADS Lung Study
- [Mark Gladwin, MD](#) – Pulmonary Disease, Allergy and Critical Care Medicine
- [Adrian Lee, PhD](#) – Breast Cancer Genetics and Tumor Heterogeneity
- [Cecelia Lo, PhD](#) – Developmental Biology and Cardiac Developmental Abnormalities
- [Michael Lotze, MD](#) – Programmed Cell Death, Autophagy and Immunology in Cancer
- [Bradley Malin, PhD](#) – Data Mining Privacy Protection and Computable Phenotype
- [Patrick Moore, MD MPH](#) – Cancer Virology and Cellular and Molecular Cancer Cell Biology
- [Alison Morris, MD MS](#) – Genomics and Microbiome in Interstitial Lung Disease
- [Michael Ochs, PhD](#) – Bayesian Modeling of Cell Signaling
- [Steffi Oesterreich, PhD](#) – Breast Cancer Genetics and Nuclear Receptors
- [C. David Page, PhD](#) – Machine Learning and Data Mining
- [Gunaretnam Rajagopal, PhD](#) – Drug Discovery, Systems Biology, Cyber Infrastructure
- [Mark Roberts, MD MPP](#) – Decision Analysis and Simulation of Problems in Health and Medicine
- [Frank Sciruba, MD](#) – Lung SCCOR, COPD and Interstitial Lung Disease Center
- [Bennett Van Houten, PhD](#) – Molecular Oncology, Cancer Pharmacology, Aging & DNA Repair

# Website: [www.ccd.pitt.edu](http://www.ccd.pitt.edu)



Center for  
Causal  
Discovery



[About](#)

[Research](#) ▼

[Training](#) ▼

[People](#)

[Calendar](#)

[News](#)

[Contact](#)



## News



Distinguished Lecture  
in Causal Discovery -  
Dr. Kathleen Gates



Distinguished Lecture  
in Causal Discovery -  
Dr. C. David Page



Summer Short Course:  
Causal Discovery with  
Graphical Models



Distinguished Lecture  
in Causal Discovery -  
Dr. Clark Glymour



Center for Causal  
Discovery featured in  
Pitt SOM Annual  
Report

## Home

### Data Science

Development of  
algorithms, software, &  
system architecture to  
discover causality in big  
data

### Biomedical Science

Discovery of causal  
knowledge in big data  
for cancer driver  
mutations, lung fibrosis, &  
brain connectome

### Training

Interactive &  
downloadable materials  
for data scientists,  
biomedical investigators,  
& software users at all  
levels

### Consortium

Dissemination of  
CCMD resources through  
the Web, Technical  
Catalyst, Scientific  
Catalysts, & collaborations

## Upcoming Events

APR  
23  
Thu

11:00 am CCD Colloquia  
Series @ 407A/B BAUM,  
5607 Baum Blvd

MAY  
21  
Thu

11:00 am CCD Colloquia  
Series @ Steinberg  
Auditorium, Baker Hall  
Room A53, CMU

JUN  
8  
Mon

all-day CCD Summer  
Workshop @ Carnegie  
Mellon University



Add ▼

[View Calendar](#) →

# A Strong Team of Investigators

- University of Pittsburgh investigators are in Biomedical Informatics, Computational and Systems Biology, Computer Science, and Human Genetics
- University of Pittsburgh is in partnership with CMU, Yale, and PSC
- Collaborative projects with Stanford, Harvard, and other BD2K Centers

# Scientific Aims of CCD

- **Aim 1.** Develop and implement state-of-the-art methods for causal modeling and discovery (CMD) of knowledge from biomedical big data
  - Make the best existing CMD methods available (“one-stop shopping”)
  - Develop new CMD methods
- **Aim 2.** Investigate three biomedical projects
  - Evaluate the usefulness of CMD methods on these problems
  - Drive further the development of the CMD methods
- **Aim 3.** Disseminate the CMD methods widely to biomedical researchers and data scientists
  - Available as Application Programming Interfaces (APIs)
  - Available through an easy-to-use and powerful desktop interface

# Workshop:

## Causal Discovery with Graphical Models

- **June 8-11, 2015** (Carnegie Mellon University, Pittsburgh, PA)
- Introduction to causal modeling and discovery in biomedical research
- Appropriate for graduate students, postdocs, new & established investigators seeking training in causal discovery – both biomedical and data scientists
- Hands-on, individualized training on the use of Tetrad with practice data & attendee's own data
- Poster session with dinner
- Attendees bring laptops
- No registration fee – discounted housing available (attendee expense) at CMU dormitories & nearby hotels
- Registration Closes Friday, May 15, 2015
- <http://www.ccd.pitt.edu/summer-short-course-causal-discovery-with-graphical-models/>

# Workshop Registration:

<https://adobeformscentral.com/?f=ahFfhYU17t5vn6bCzaYFsA#>



Center for  
Causal  
Discovery

## Summer Short Course in... Causal Discovery from Biomedical Data June 8-11, 2015 Registration Form

### Overview:

The CCD Summer Short Course in Causal Discovery from Biomedical Data provides an exciting opportunity for data scientists and biomedical investigators at all levels. Topics and activities will include...

- Central topics in causal graphical modeling
- Machine learning causal models from biomedical data
- Advances in biomedical research through causal discovery

### Important Details:

- Enrollment will be limited to the first 75 registrants.
- Registrations must be received by Friday, May 15.
- Each of the 4 days of the course will consist of 4 sessions of 90 minutes each.
- There is no fee to participate in the CCD Summer School.
- Travel, food, and housing costs are the responsibility of the participants  
Provided options: Low-cost CMU dorm and reduced-rate hotel room  
([Wyndham Pittsburgh University Center](#) - room blocks held until May 17)
- For any questions about the CCD Summer Short Course, please contact Toni Porterfield, [ts18@pitt.edu](mailto:ts18@pitt.edu)
- Please see our website for more details: [www.ccd.pitt.edu/training/summer-short-course-2015/](http://www.ccd.pitt.edu/training/summer-short-course-2015/)

### Personal Information

Last Name\*

First Name\*

E-mail\*

Institution\*

Phone Number

Your Academic Rank\*

What would be your preference for housing? \*



Your primary research area\*



# Next Steps

- Meetings ~ twice a year
- Box collaboration
- Seminars – CCD Colloquia Series
- Suggestions?

# NIH Big Data to Knowledge (BD2K) Initiative

The ability to harvest the wealth of information contained in biomedical Big Data will advance our understanding of human health and disease; however, lack of appropriate tools, poor data accessibility, and insufficient training, are major impediments to rapid translational impact.

To meet this challenge, the National Institutes of Health (NIH) launched the Big Data to Knowledge (BD2K) initiative in 2012.

BD2K is a trans-NIH initiative established to enable biomedical research as a digital research enterprise, to facilitate discovery and support new knowledge, and to maximize community engagement.

# NIH Big Data to Knowledge (BD2K) Initiative



## Major aims:

- Facilitate broad use of biomedical digital assets
- Conduct research and develop the methods, software, and tools needed to analyze biomedical Big Data
- Enhance training in the development and use of methods and tools necessary for biomedical Big Data science.
- Support a data ecosystem that accelerates biomedical knowledge discovery

# NIH BD2K Centers of Excellence

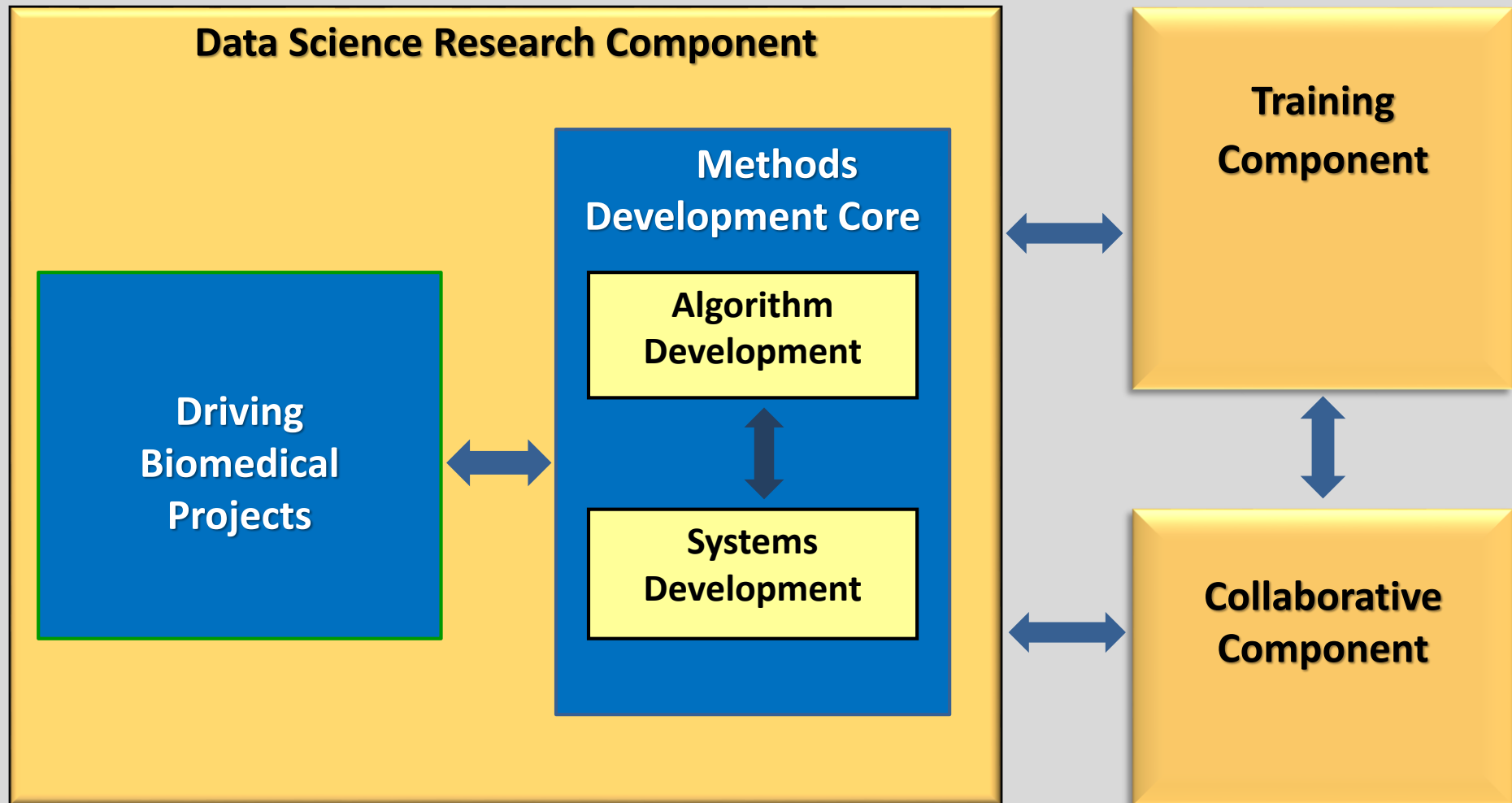


- The Centers of Excellence are part of the overall NIH BD2K initiative.
- The goal is to develop and disseminate computational methods to assist biomedical researchers in using big data to significantly advance biomedical science.
- Project components include research, software development and dissemination, training, and joint Center activities.
- As of September 2014, NIH began funding 12 BD2K Centers of Excellence.
- Funding is about \$2.7M (total) per Center per year for 4 years.
- For more information, see: [www.bd2k.nih.gov](http://www.bd2k.nih.gov)

# Center for Causal Discovery (CCD)



# The Main Components of the Center for Causal Discovery



# Why Establish the CCD?

Science is centrally concerned with the discovery of causal relationships in nature.

- Understanding
- Prediction
- Control

Examples:

- Determine the genes and cell signaling pathways that cause breast cancer
- Discover the clinical effects of a new drug
- Uncover the mechanisms of pathogenicity of a recently mutated virus that is spreading rapidly in the population

# Anticipated Scientific Contributions of the Center for Causal Discovery

- Algorithms that will support the efficient discovery of causal knowledge from big biomedical data
- A computable representation of causal networks that facilitates generating, analyzing, visualizing, comparing, annotating, sharing, and storing such models



# Deliverables

- Software
  - Implements a rich suite of CMD algorithms
  - Available as application programming interfaces (APIs)
  - Open source and free
- An easy-to-use CMD system with a desktop interface, which is open source and free
- Projects with other Centers
- Training

# Training Initiatives

- Train **biomedical researchers** in the use of CMD methods and software applied to big data
- Train **data scientists** to develop new CMD tools and methods for analyzing biomedical big data
- Offer various types of training
  - Online tutorials and courses
  - Summer short courses
  - Postdoctoral, graduate, and undergraduate training (n  $\approx$  20)
  - Data Science workshops
  - Hackathons

# Acknowledgements

- Thanks to the 40+ members of the Center for Causal Discovery for their contributions to the Center activities that are described here.
- The Center for Causal Discovery is supported by grant U54HG008540 awarded by the National Human Genome Research Institute through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative ([www.bd2k.nih.gov](http://www.bd2k.nih.gov)). The content of this presentation is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

# Thank you

[www.ccd.pitt.edu](http://www.ccd.pitt.edu)