Modeling and Simulation in Systems Biology

Transition from Third to Fourth Paradigm- Data Intensive Science

Use of bionetworks to build better maps of disease

Integrated Network Maps of Cancer
Sharing Data, Tools and Models

Stephen Friend MD PhD

Sage Bionetworks (Non-Profit Organization)
Seattle/ Beijing/ San Francisco

PRISME
May 4th, 2011

Alzheimers

Diabetes

Treating Symptoms v.s. Modifying Diseases

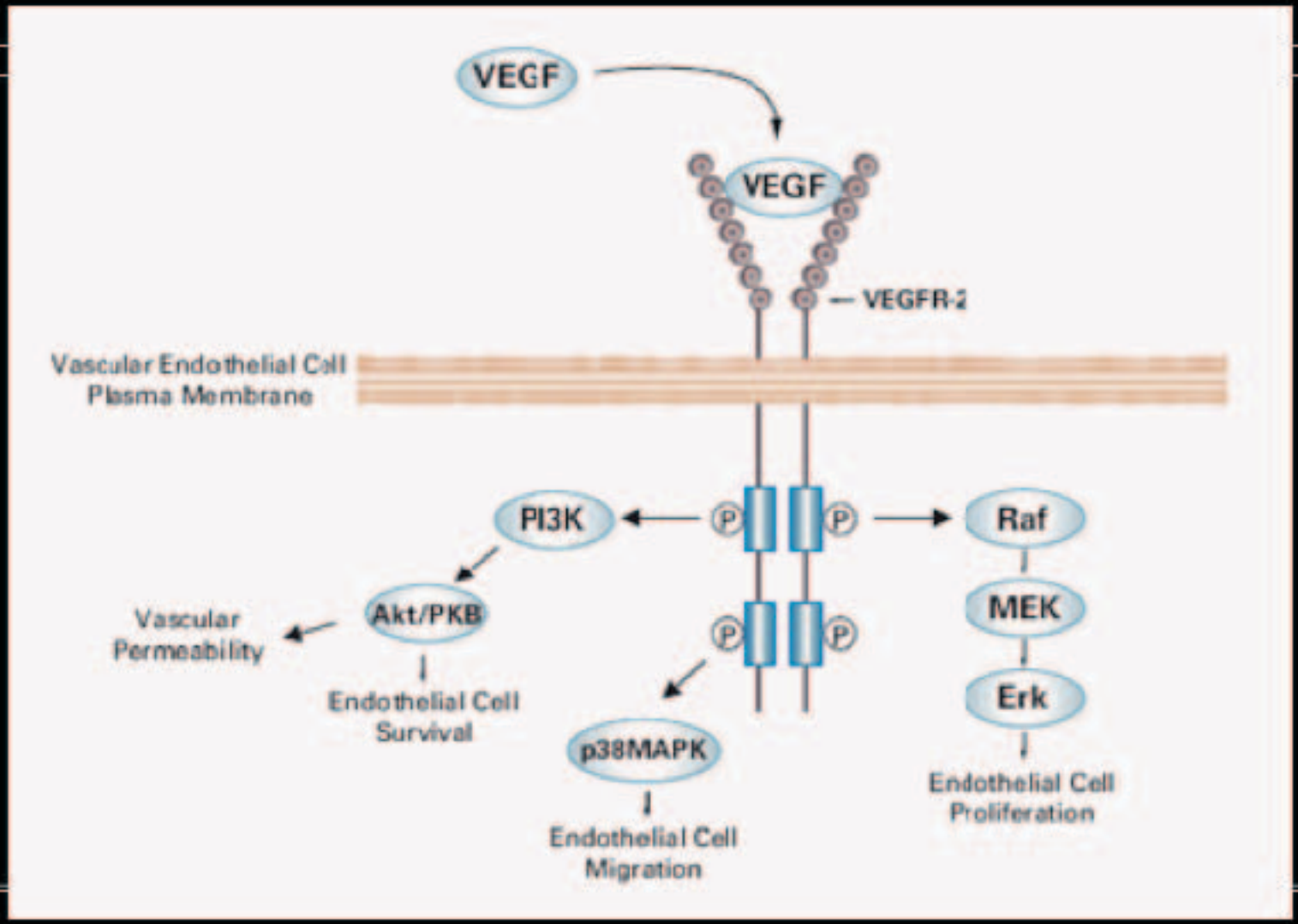
Depression

Cancer

Will it work for me?

Familiar but Incomplete

VEGFR Classical Pathway

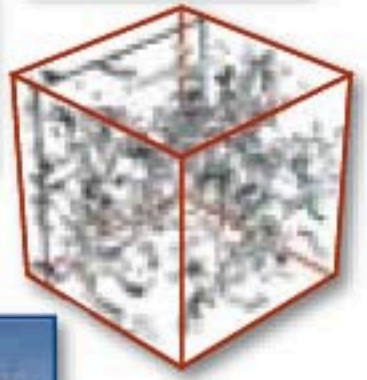


WHY “DATA INTENSIVE” SCIENCE?

Science Paradigms

- Thousand years ago:
  science was **empirical**
  describing natural phenomena
- Last few hundred years:
  theoretical branch
  using models, generalizations
- Last few decades:
  a **computational** branch
  simulating complex phenomena
- Today: **data exploration** (eScience)
  unify theory, experiment, and simulation
  - Data captured by instruments or generated by simulator
  - Processed by software
  - Information/knowledge stored in computer
  - Scientist analyzes database/files using data management and statistics

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G p}{3} - \mathrm{K}\frac{c^2}{a^2}$$



“Data Intensive Science” - Fourth Scientific Paradigm

Equipment capable of generating massive amounts of data

IT Interoperability

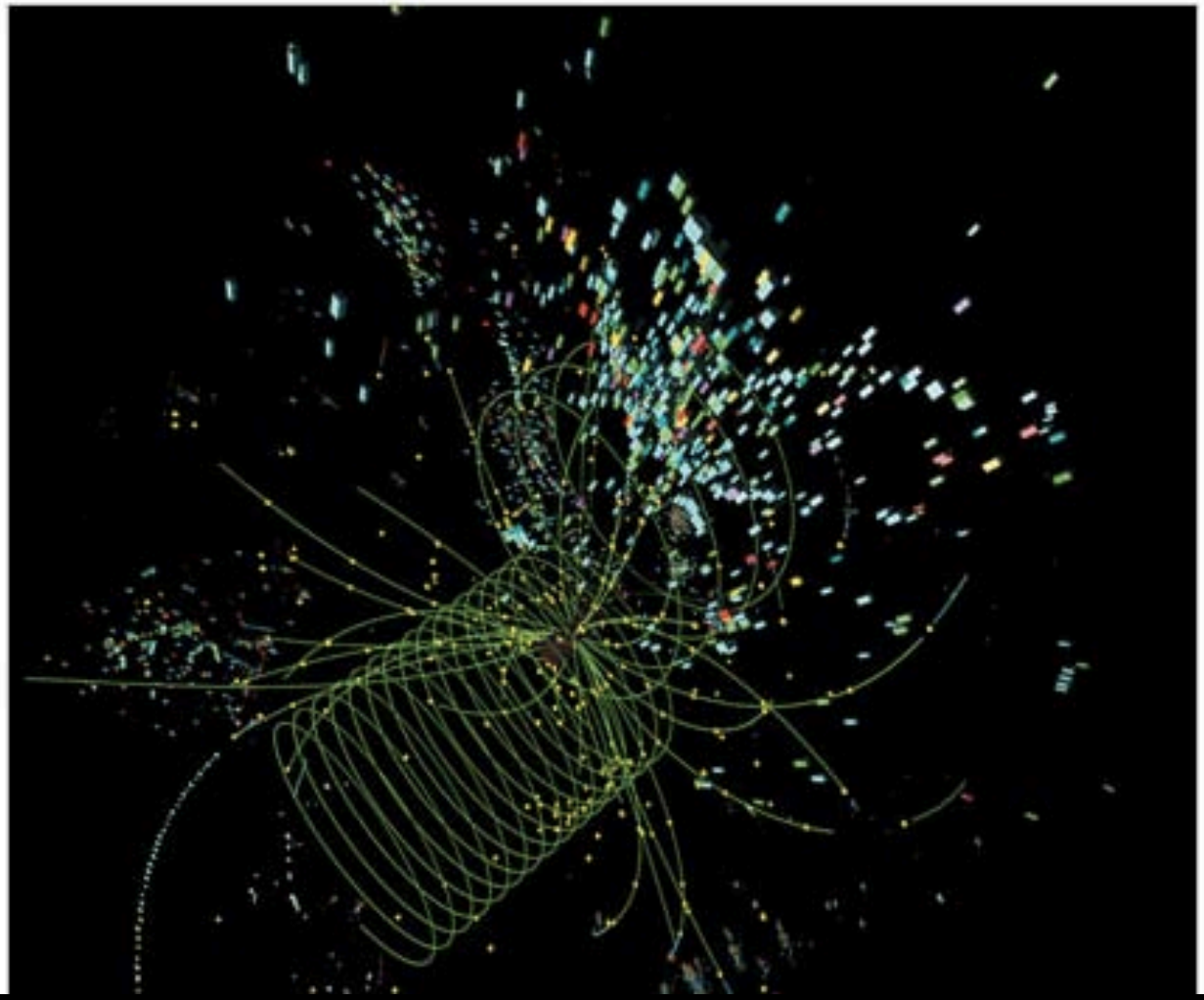
Open Information System

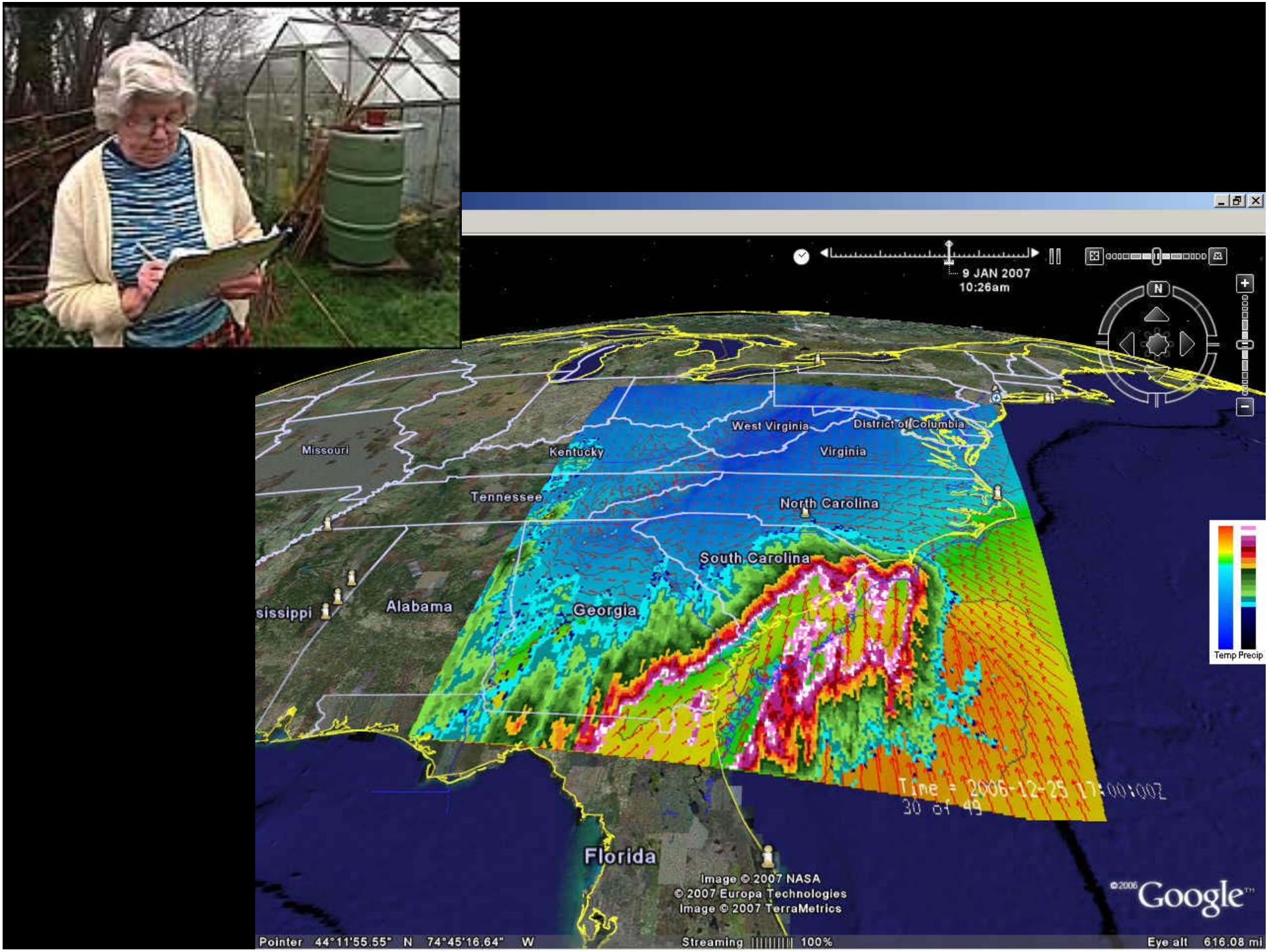
Evolving Models hosted in a Compute Space- Knowledge expert

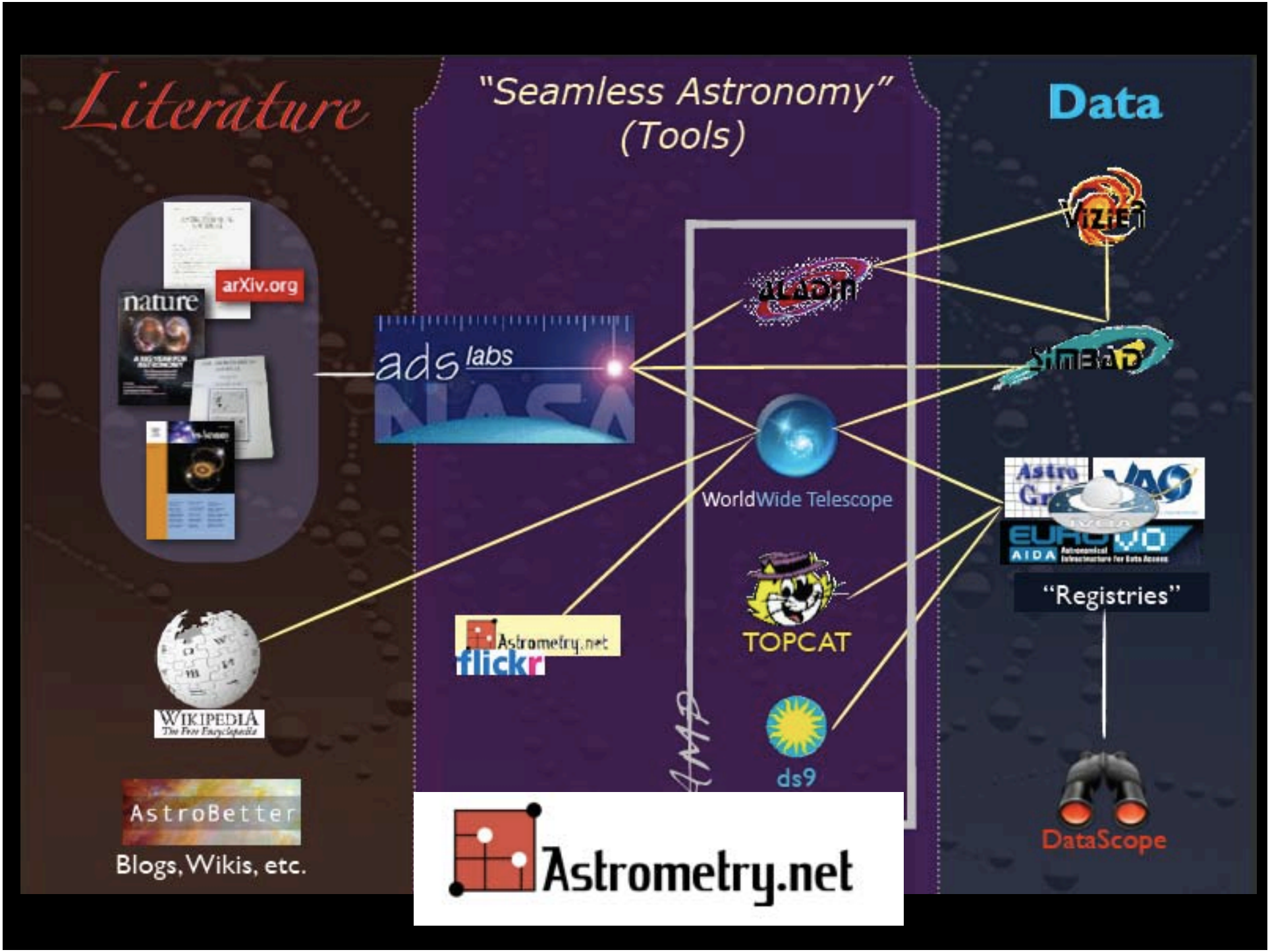
International Linear Collider Matters

Published by Clifford on February 9, 2007 in science. 3 Comments

Yes, it *does* matter. And it does *matter*:




9 JAN 2007
10:26am
N
Missouri
Kentucky
West Virginia
District of Columbia
Virginia
Tennessee
North Carolina
South Carolina
Alabama
Georgia
Florida
Temp Precip
Time = 2006-12-25 17:00:00Z
30 of 49
Image © 2007 NASA
© 2007 Europa Technologies
Image © 2007 TerraMetrics
©2006 Google
Pointer 44°11'55.55" N 74°45'16.64" W
Streaming 100%
Eye alt 616.08 mi


Literature
"Seamless Astronomy"
(Tools)
Data
arXiv.org
nature
ads labs
NASA
ALADIN
VizieR
SIMBAD
WorldWide Telescope
Astro Grid
VO
EURO VO
AIDA
"Registries"
TOPCAT
Astrometry.net
flickr
WIKIPEDIA
The Free Encyclopedia
ds9
AstroBetter
Blogs, Wikis, etc.
Astrometry.net
DataScope

WHY NOT USE
“DATA INTENSIVE” SCIENCE
TO BUILD BETTER DISEASE MAPS?

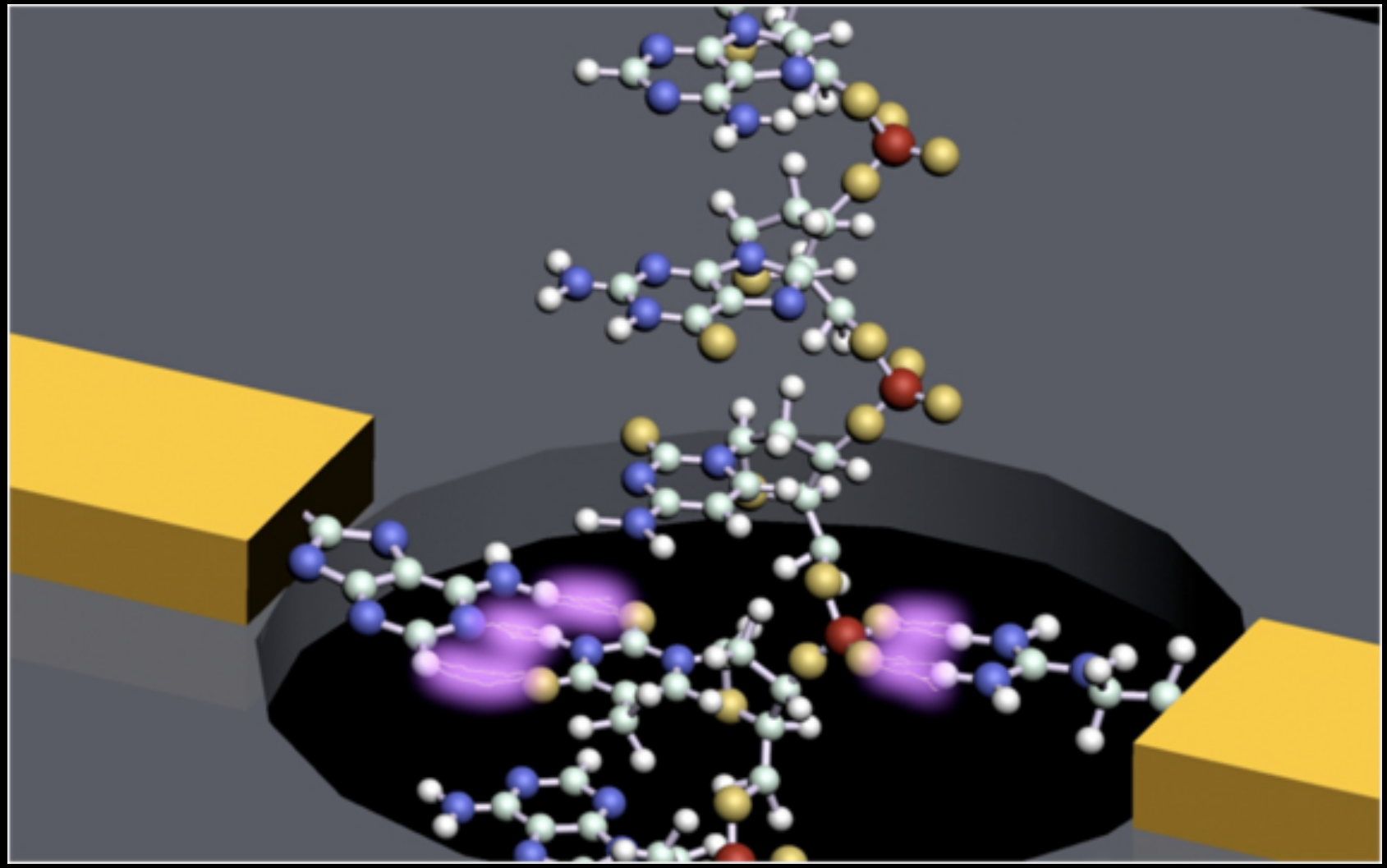
“Data Intensive Science”- “Fourth Scientific Paradigm”
For building: “Better Maps of Human Disease”

Equipment capable of generating massive amounts of data

IT Interoperability

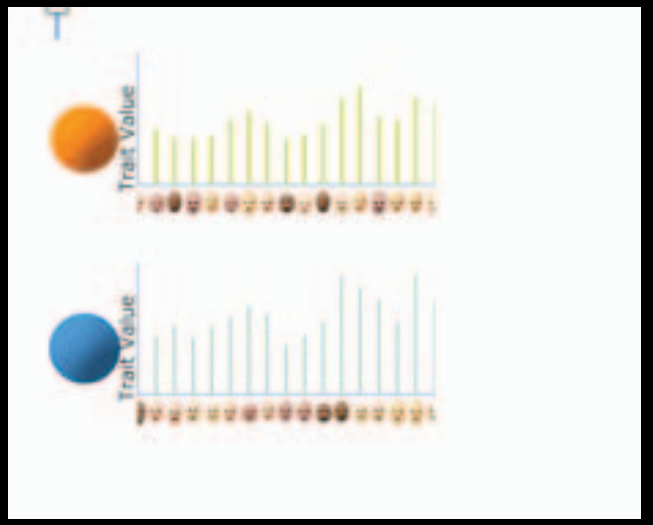
Open Information System

Evolving Models hosted in a Compute Space- Knowledge Expert



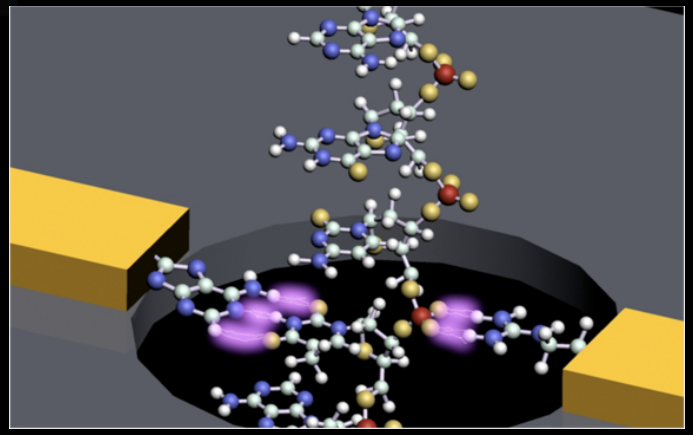
It is now possible to carry out comprehensive monitoring of many traits at the population level

Monitor disease and molecular traits in populations

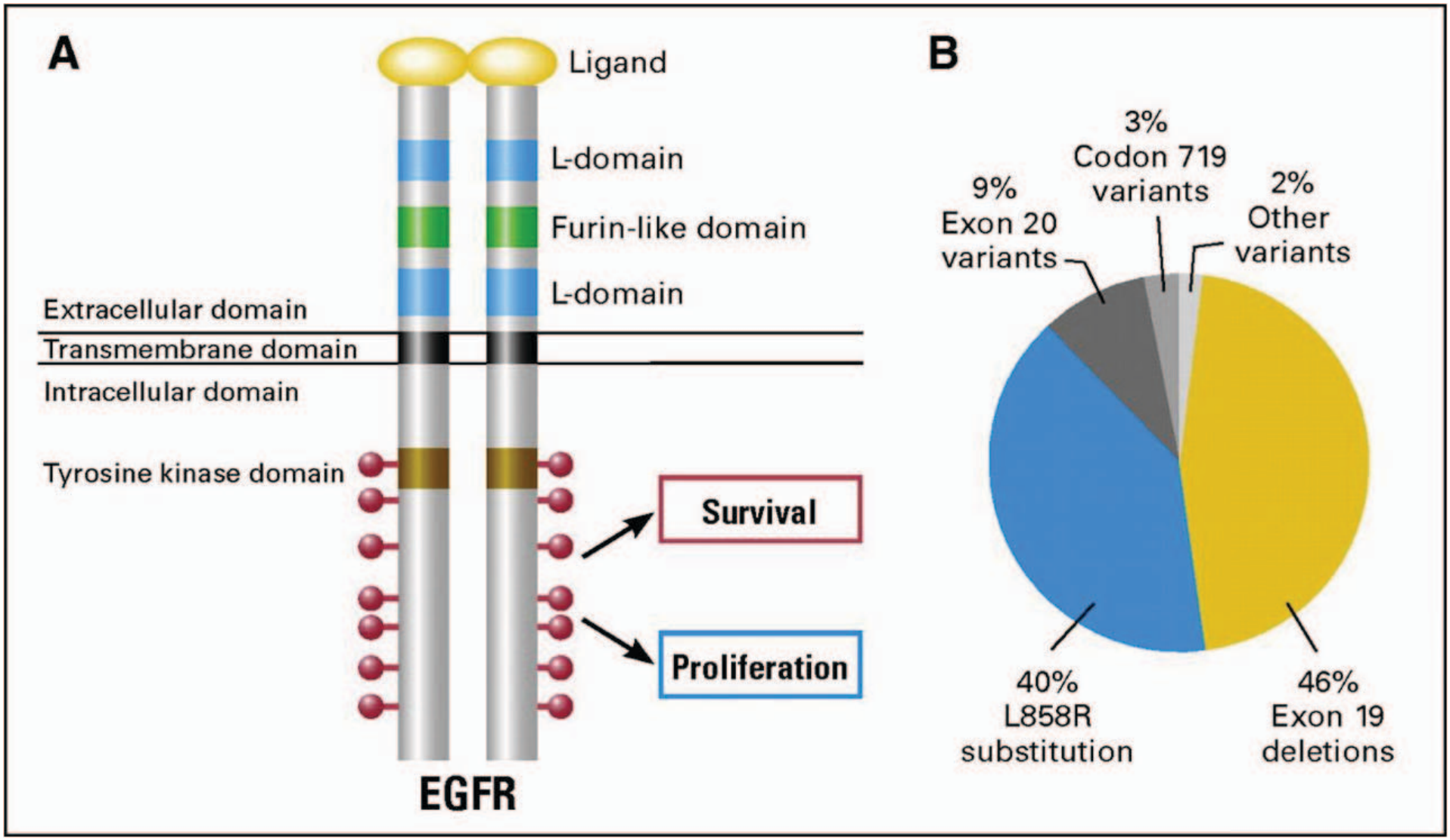


Putative causal gene

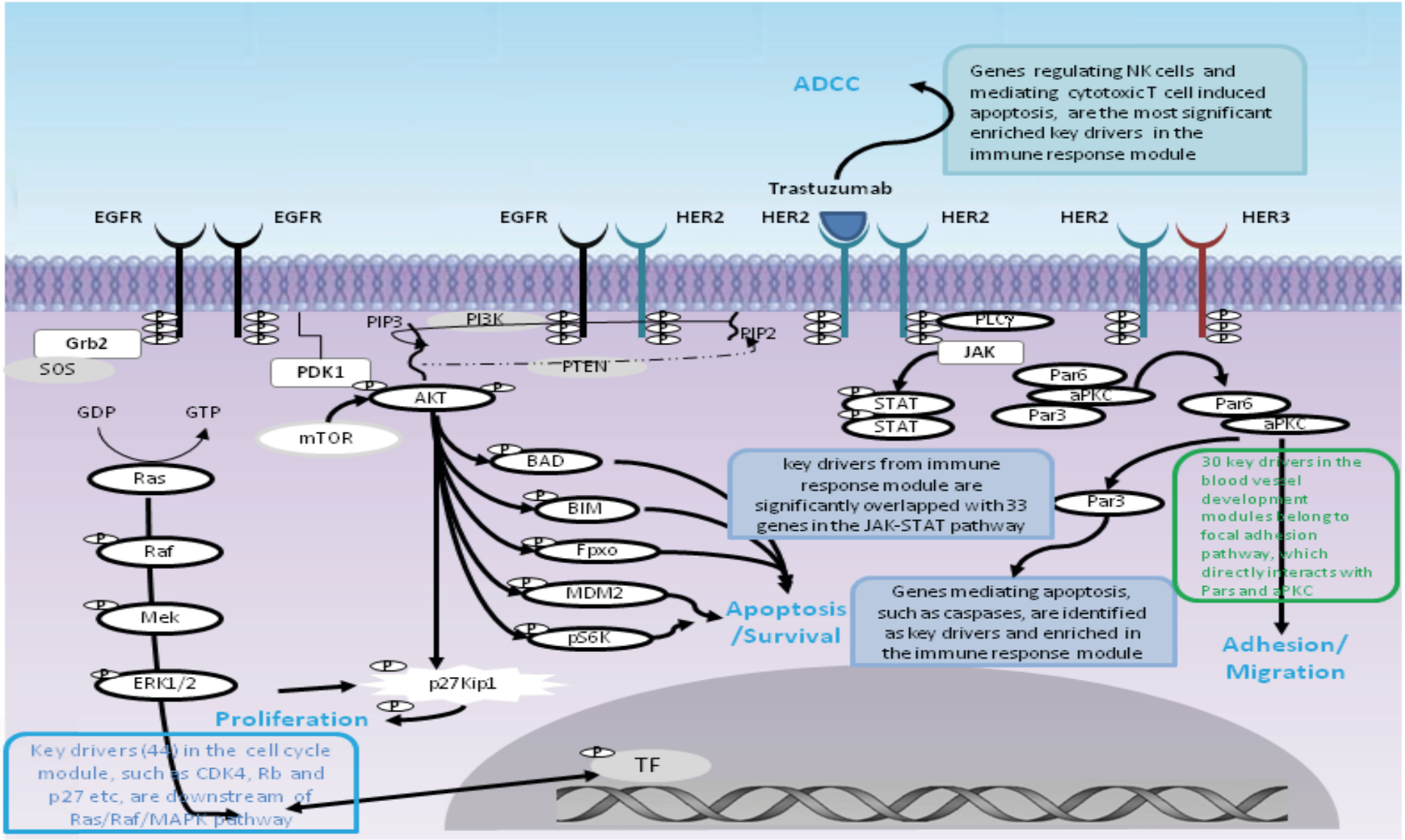
Disease trait

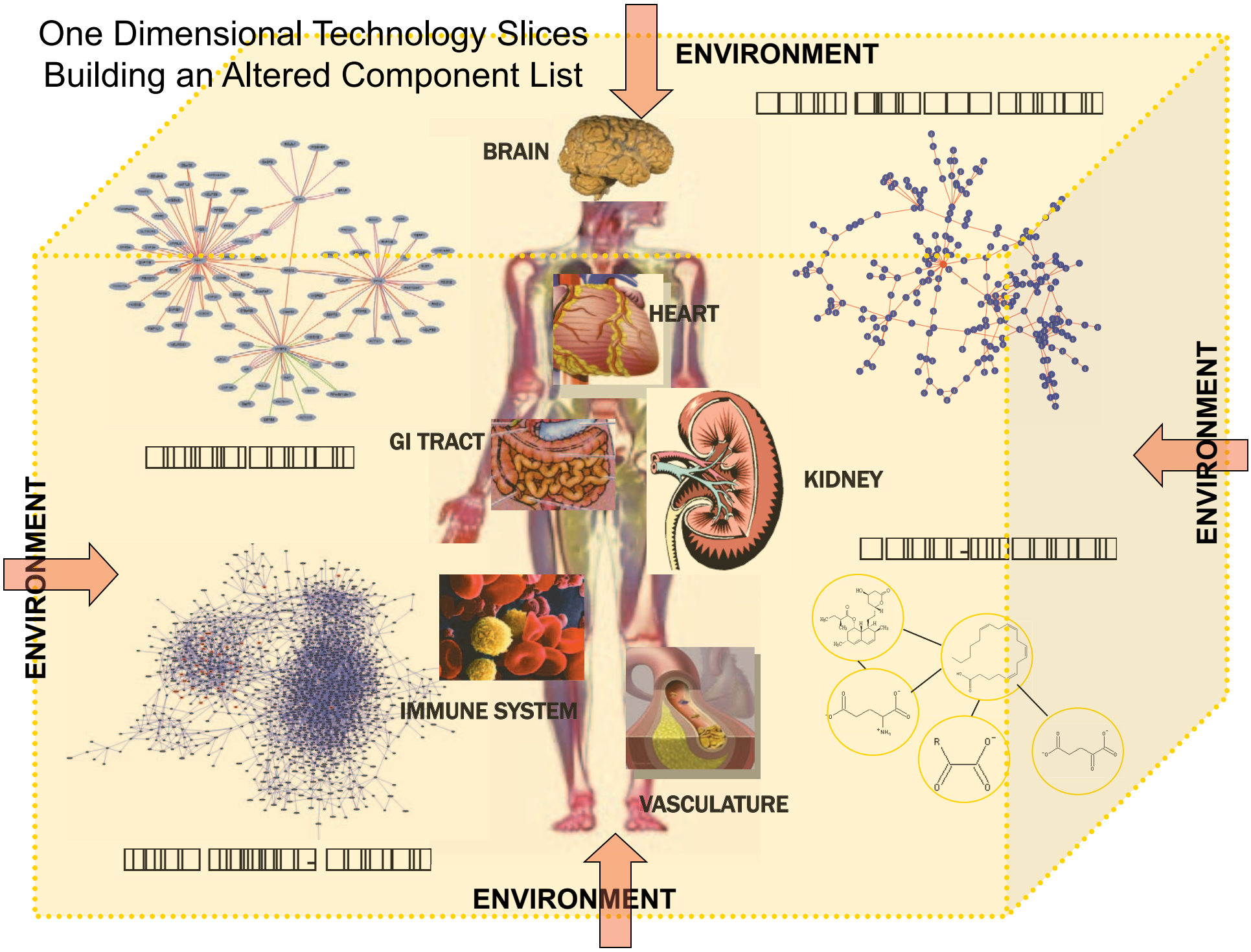


Personalized Medicine 101:
Capturing Single bases pair mutations = ID of responders



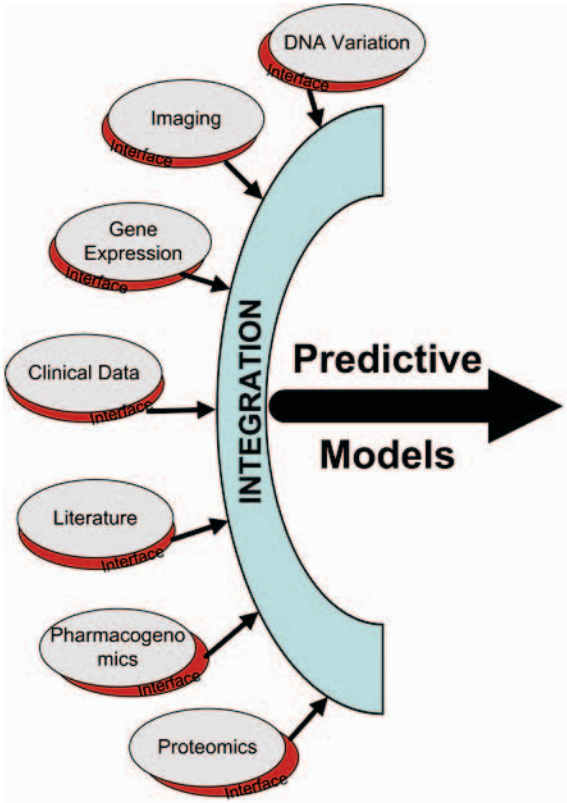
Cancer Complexity: Overlapping of EGFR and Her2 Pathways

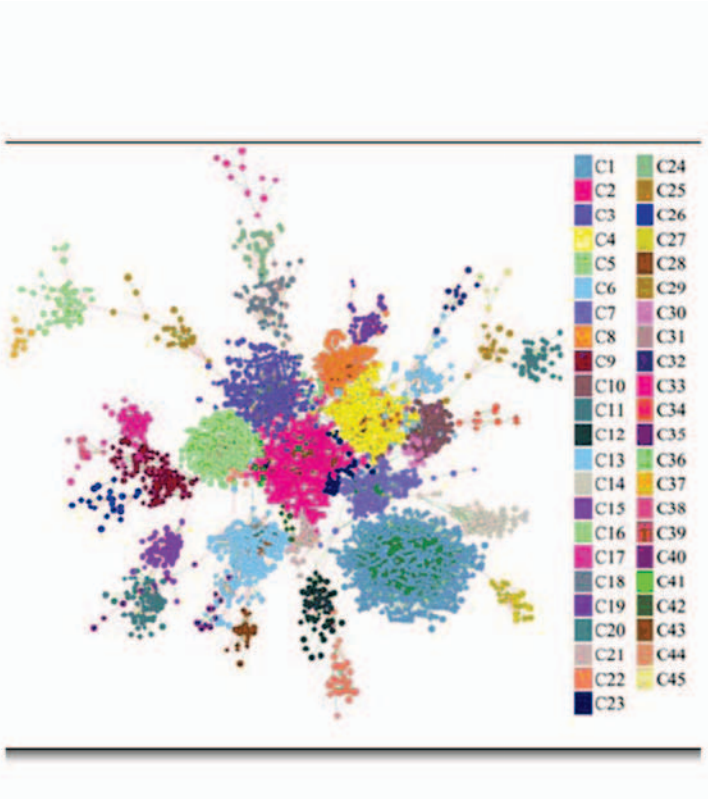



One Dimensional Technology Slices
Building an Altered Component List
ENVIRONMENT
BRAIN
HEART
GI TRACT
KIDNEY
ENVIRONMENT
ENVIRONMENT
IMMUNE SYSTEM
VASCULATURE
ENVIRONMENT

The Rosetta Integrative Genomics Experiment: Generation, assembly, and integration of data to build models that predict clinical outcome

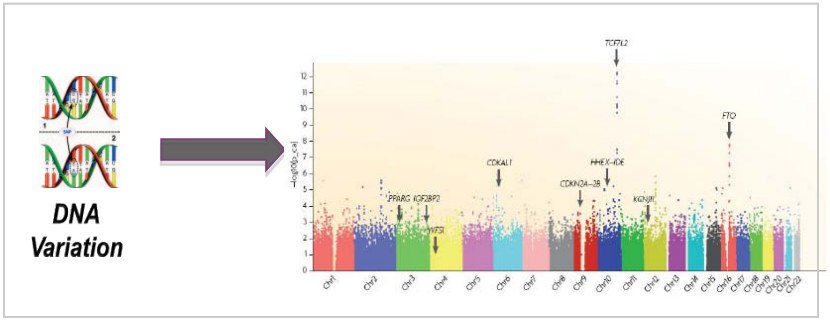
Merck Inc. Co.
5 Year Program
Based at Rosetta
Driven by Eric Schadt
Total Resources
>$150M





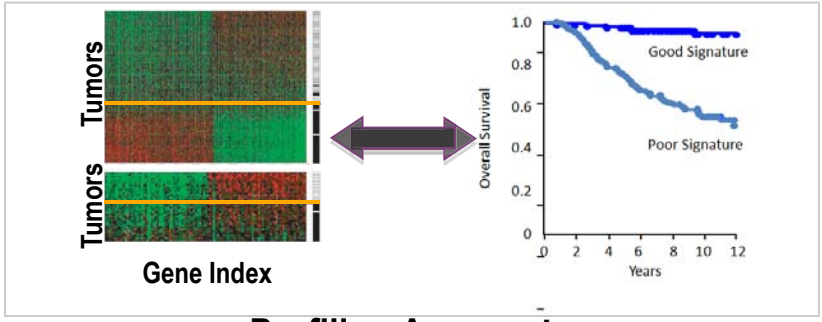
- **Generate data need to build**
- **bionetworks**
- **Assemble other available data useful for building networks**
- **Integrate and build models**
- **Test predictions**
- **Develop treatments**
- **Design Predictive Markers**

How is genomic data used to understand biology?



"Standard" GWAS Approaches

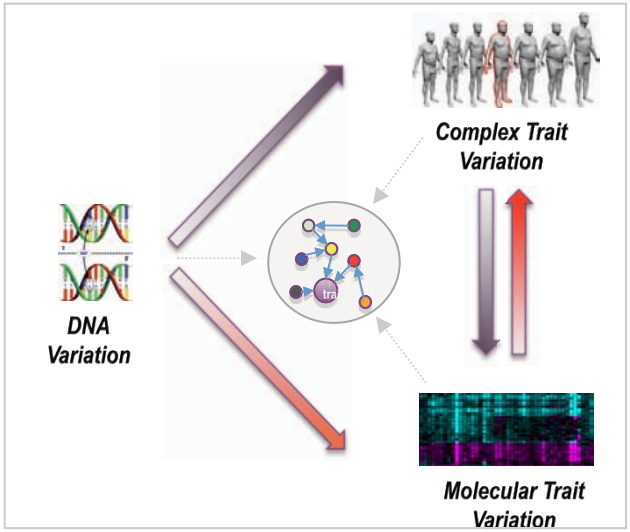
Identifies Causative DNA Variation but provides NO mechanism



Profiling Approaches

Genome scale profiling provide correlates of disease

- Many examples BUT what is cause and effect?

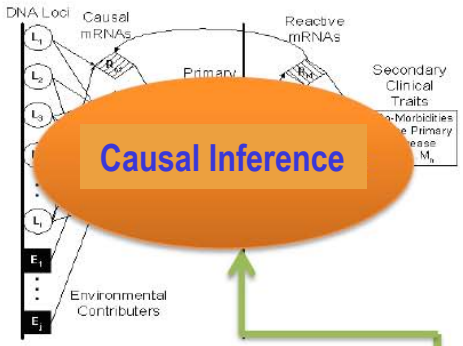


"Integrated" Genetics Approaches

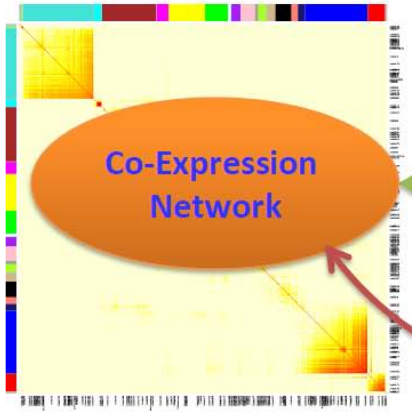
- Provide unbiased view of molecular physiology as it relates to disease phenotypes
- Insights on mechanism
- Provide causal relationships and allows predictions

Integration of Genotypic, Gene Expression & Trait Data

Schadt et al. Nature Genetics 37: 710 (2005)
Millstein et al. BMC Genetics 10: 23 (2009)



"Global Coherent Datasets"
• population based
• 100s-1000s individuals



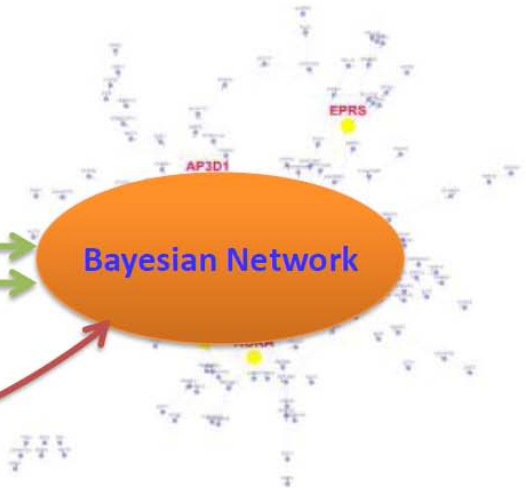
Co-Expression Network

Genotype Data

Gene Expression

Clinical Traits

Other Data



Bayesian Network

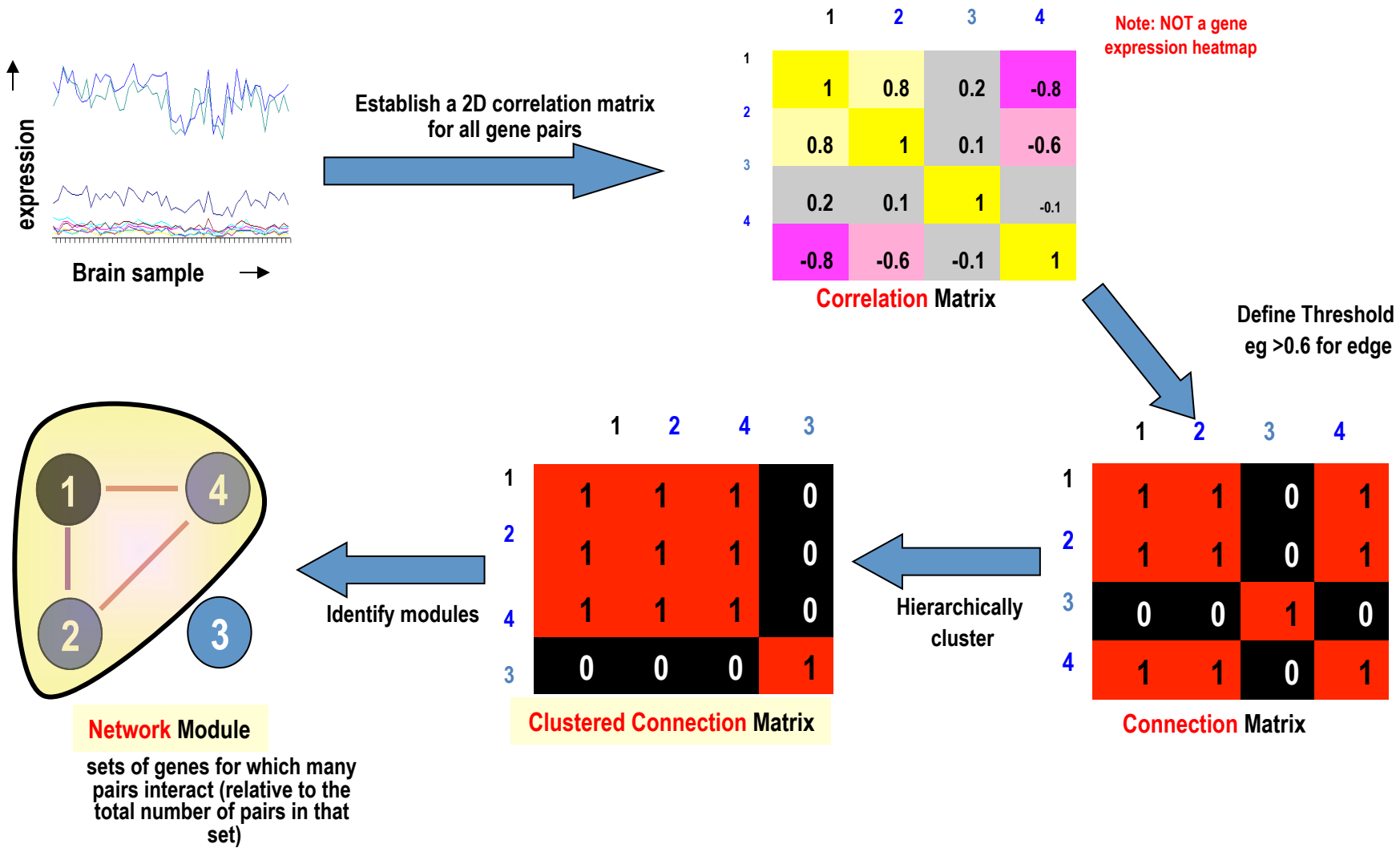
Integration of Coexp. & Bayesian Networks

Chen et al. Nature 452:429 (2008)
Zhang & Horvath. Stat.Appl.Genet.Mol.Biol. 4: article 17 (2005)

Zhu et al. Cytogenet Genome Res. 105:363 (2004)
Zhu et al. PLoS Comput. Biol. 3: e69 (2007)

Constructing Co-expression Networks

Start with expression measures for ~13K genes most variant genes across 100-150 samples



Gene Co-Expression Network Analysis

Define a Gene Co-expression Similarity

▼

Define a Family of Adjacency Functions

▼

Determine the AF Parameters

▼

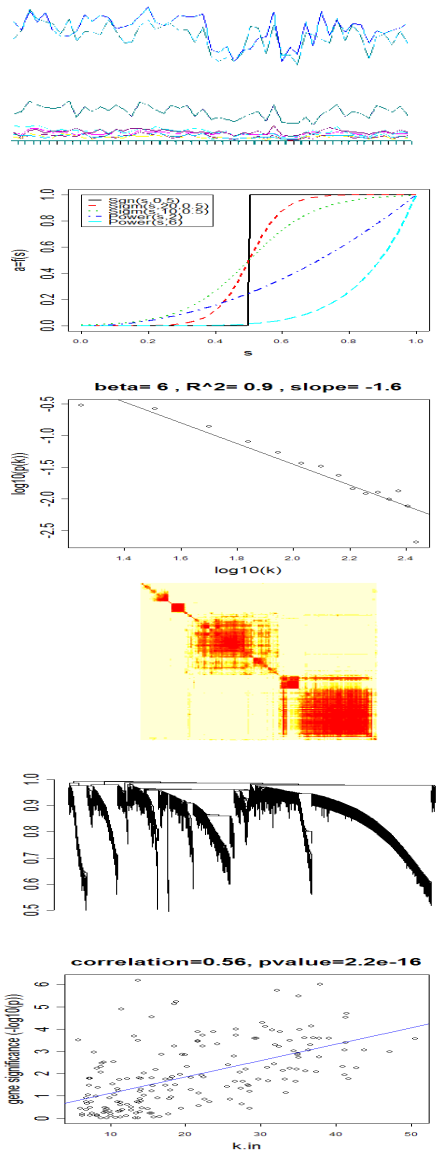
Define a Measure of Node Distance

▼

Identify Network Modules (Clustering)

▼

Relate the Network Concepts to
External Gene or Sample Information

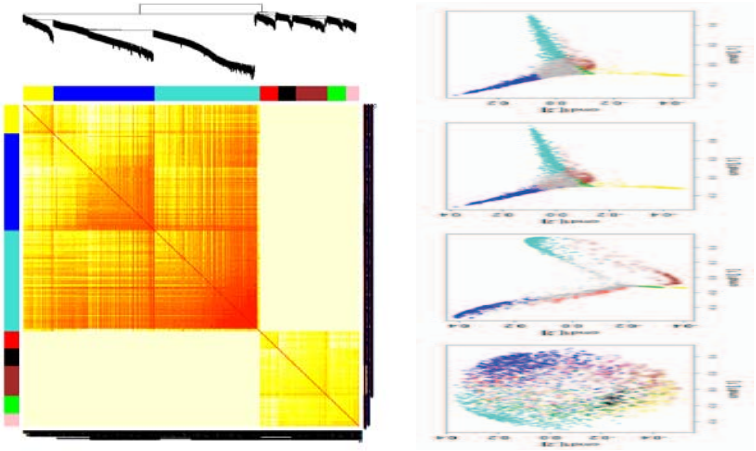


Zhang B, Horvath S. Stat Appl Genet Mol Biol 2005

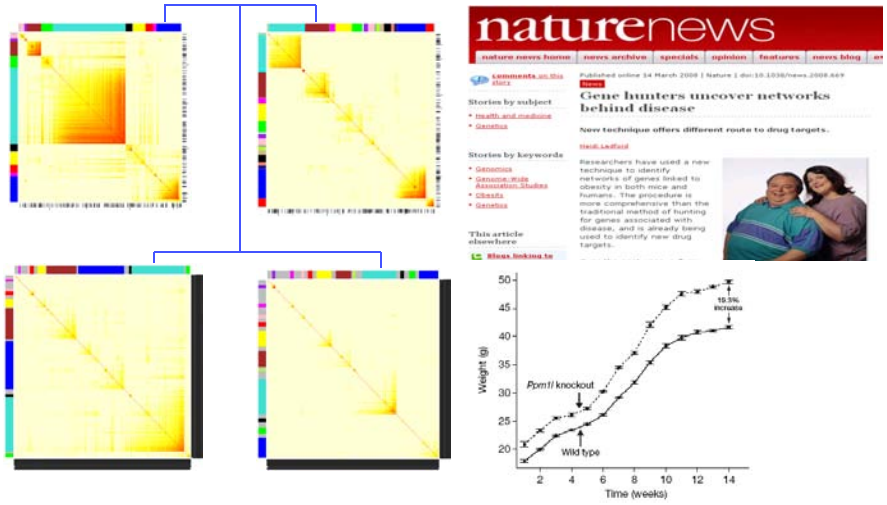
Gene Co-expression Network Analysis

Weighted Gene Network Analysis

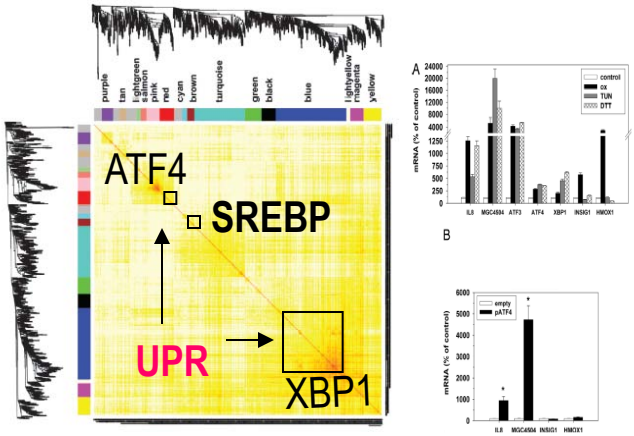
- ❑ >140 citations
- ❑>3400 full-text downloads



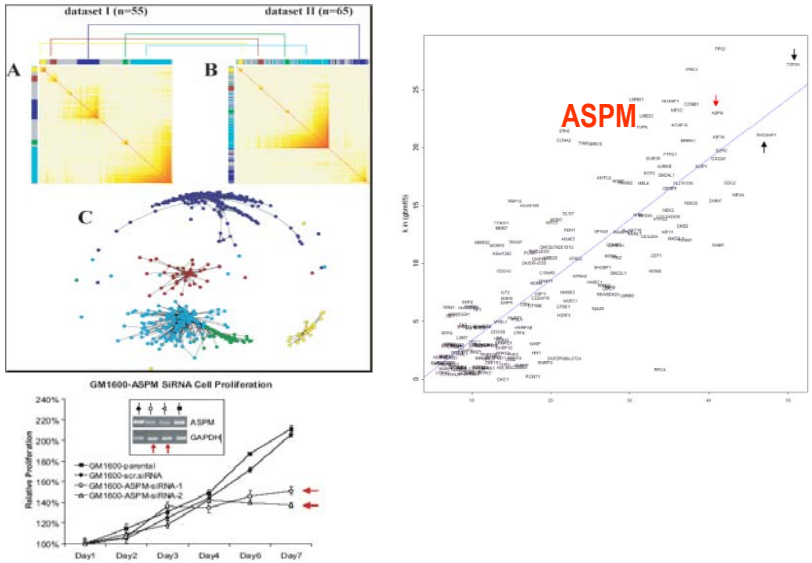
Novel gene network causal for D&O (*Nature, 2008*)



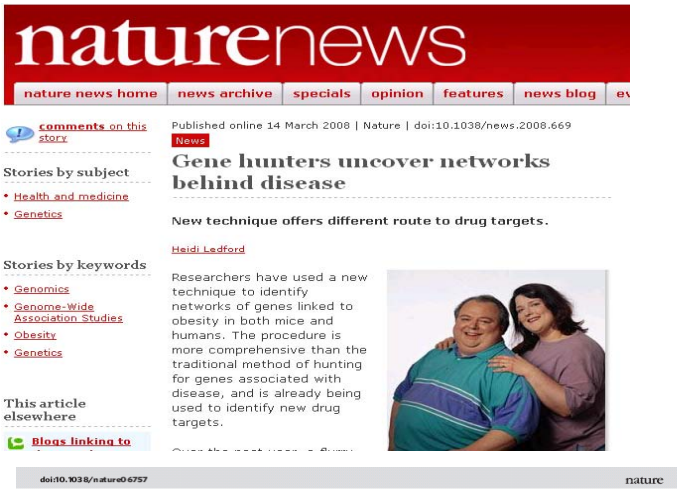
Novel pathways and gene targets in Atherosclerosis (*PNAS, 2006*)



Novel oncogene in brain cancer (*PNAS, 2006*)



A macrophage-enriched metabolic network (MEMN) associated with obesity & diabetes



nature news

nature news home | news archive | specials | opinion | features | news blog

comments on this story

Stories by subject
- Health and medicine
- Genetics

Stories by keywords
- Genomics
- Genome-Wide Association Studies
- Obesity
- Genetics

This article elsewhere
Blogs linking to

Published online 14 March 2008 | Nature | doi:10.1038/news.2008.669

News

Gene hunters uncover networks behind disease

New technique offers different route to drug targets.

Heidi Ledford

Researchers have used a new technique to identify networks of genes linked to obesity in both mice and humans. The procedure is more comprehensive than the traditional method of hunting for genes associated with disease, and is already being used to identify new drug targets.

doi:10.1038/nature06757 nature

ARTICLES

Variations in DNA elucidate molecular networks that cause disease

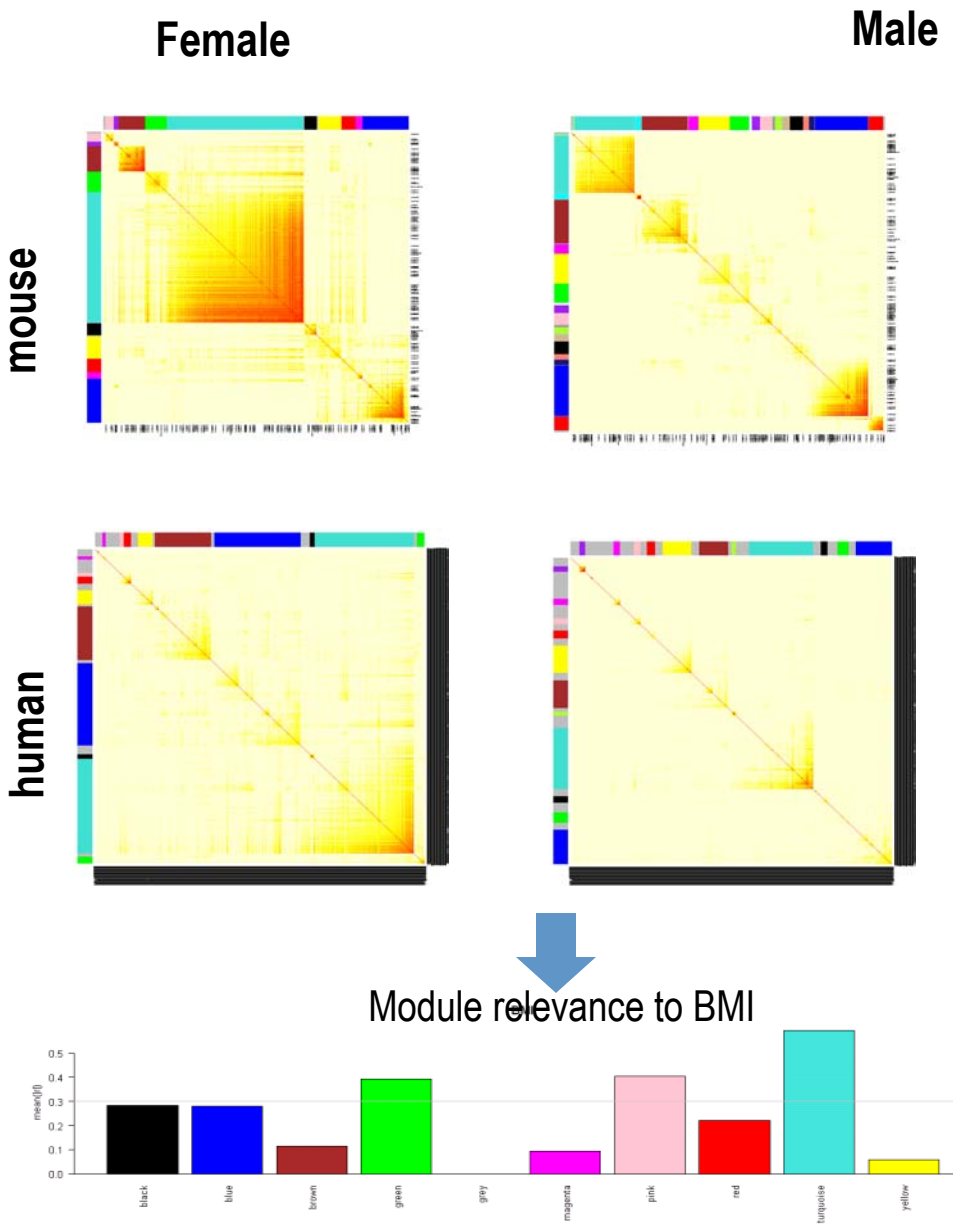
Chen Y, Zhu J et al., Nature 2008

doi:10.1038/nature06758 nature

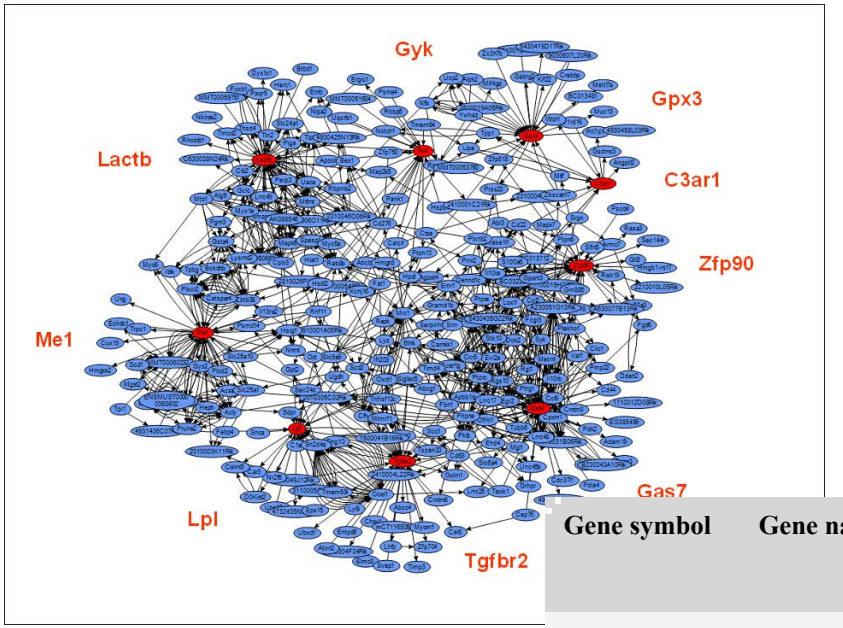
ARTICLES

Genetics of gene expression and its effect on disease

Emilsson V, Thorleifsson G, Zhang B et al., Nature,2008



Preliminary Probabalistic Models- Rosetta /Schadt



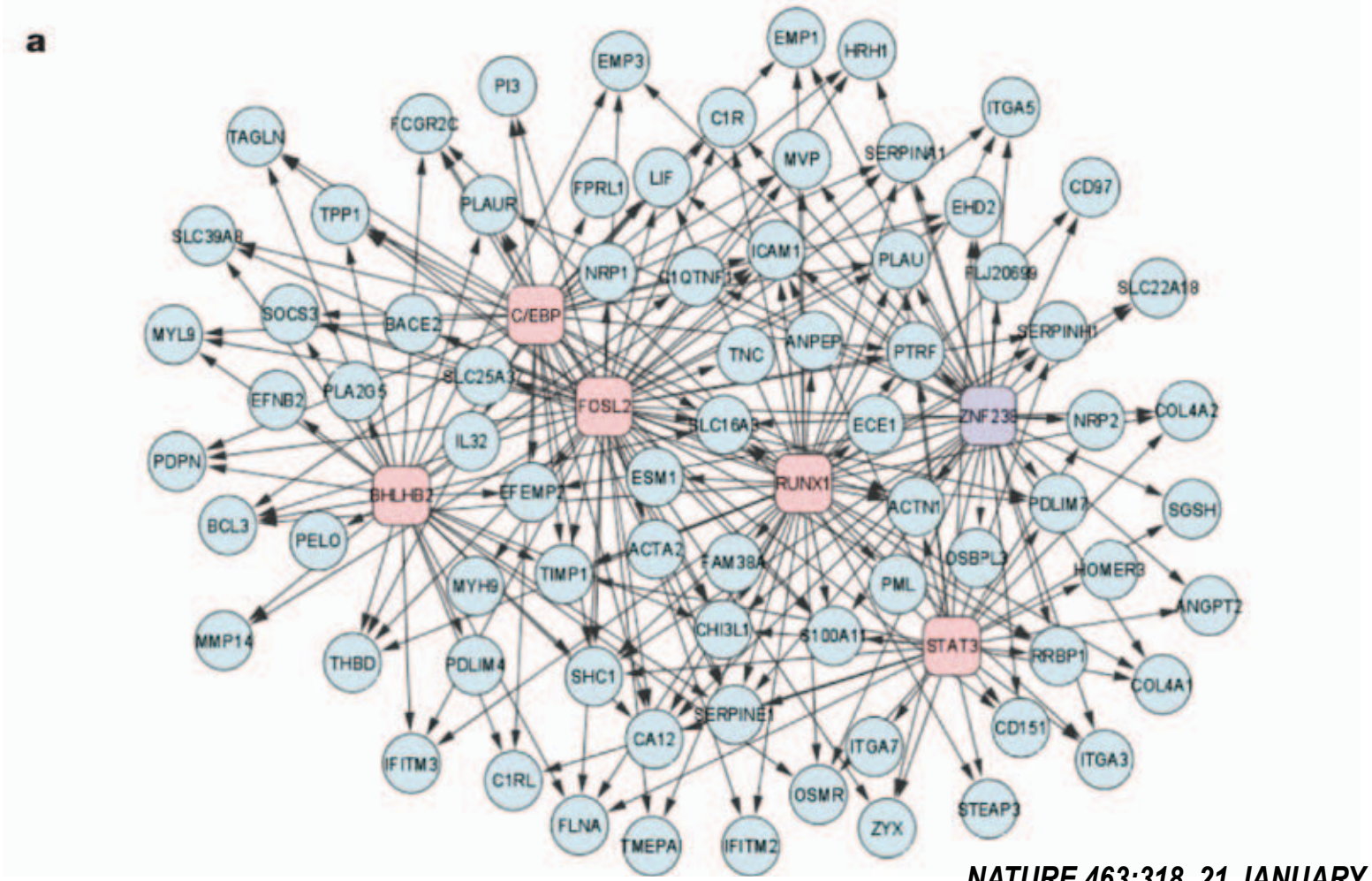
Networks facilitate direct identification of genes that are causal for disease
Evolutionarily tolerated weak spots

Gene symbol	Gene name	Variance of OFPM explained by gene expression*	Mouse model	Source
Zfp90	Zinc finger protein 90	68%	tg	Constructed using BAC transgenics
Gas7	Growth arrest specific 7	68%	tg	Constructed using BAC transgenics
Gpx3	Glutathione peroxidase 3	61%	tg	Provided by Prof. Oleg Mirochnitchenko (University of Medicine and Dentistry at New Jersey, NJ) [12]
Lactb	Lactamase beta	52%	tg	Constructed using BAC transgenics
Me1	Malic enzyme 1	52%	ko	Naturally occurring KO
Gyk	Glycerol kinase	46%	ko	Provided by Dr. Katrina Dipple (UCLA) [13]
Lpl	Lipoprotein lipase	46%	ko	Provided by Dr. Ira Goldberg (Columbia University, NY) [11]
C3ar1	Complement component 3a receptor 1	46%	ko	Purchased from Deltagen, CA
Tgfbr2	Transforming growth factor beta receptor 2	39%	ko	Purchased from Deltagen, CA

Nat Genet (2005) 205:370

The transcriptional network for mesenchymal transformation of brain tumours

Maria Stella Carro[1]*{, Wei Keat Lim[2,3]*{, Mariano Javier Alvarez[3,4]*, Robert J. Bollo[8], Xudong Zhao[1], Evan Y. Snyder[9], Erik P. Sulman[10], Sandrine L. Anne[1]{, Fiona Doetsch[5], Howard Colman[11], Anna Lasorella[1,5,6], Ken Aldape[12], Andrea Califano[1,2,3,4] & Antonio Iavarone[1,5,7]



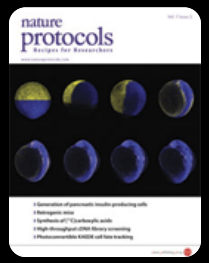
NATURE 463:318, 21 JANUARY 2010

The Evolution of Systems Biology

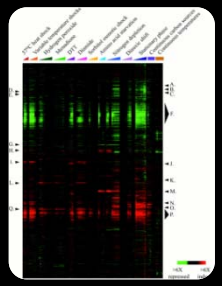
GENERATING HIGH-THROUGHPUT DATA

Genomic

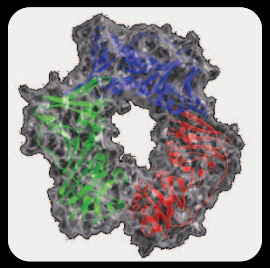
Literature



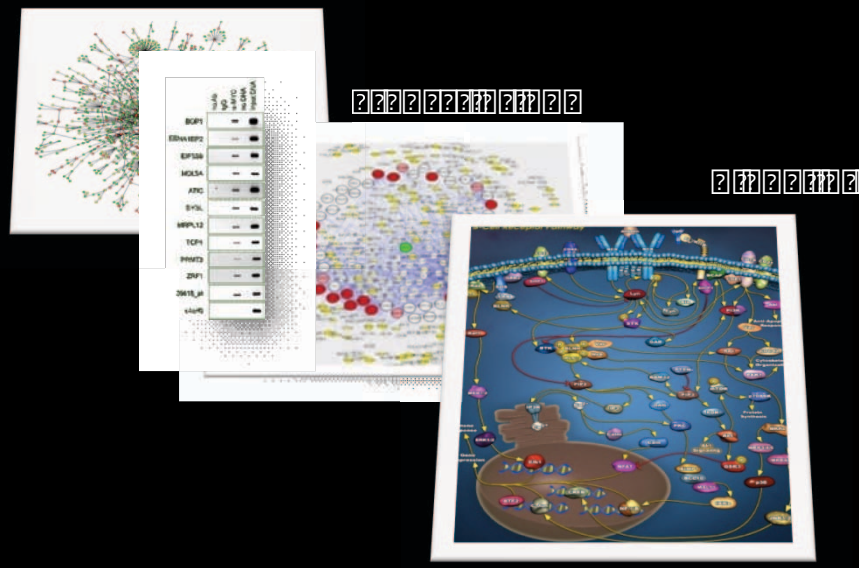
Mol. Profiles



Structure

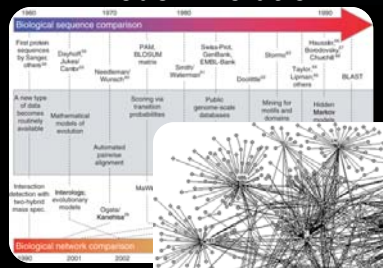


REVERSE ENGINEERING THE MODEL

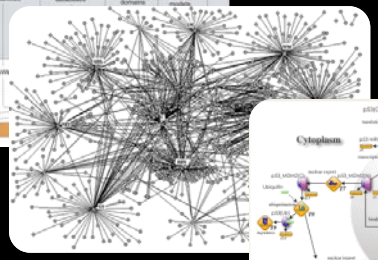


MODELING COMPLEXITY

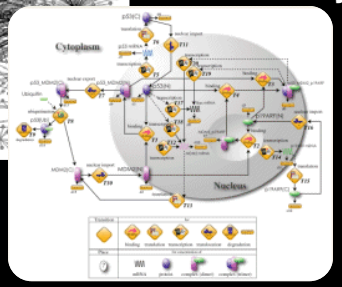
Model Evolution



Model Topology

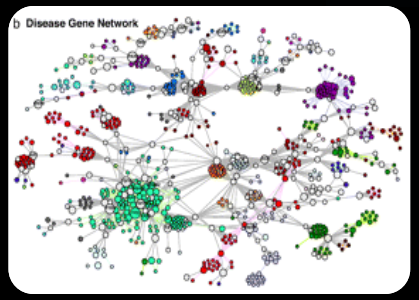


Model Dynamics

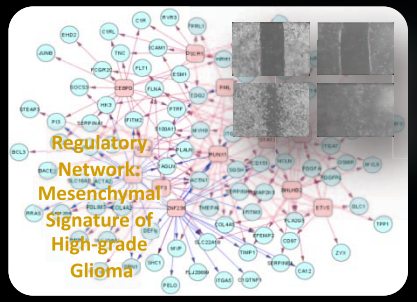


UNDERSTANDING CELL PHYSIOLOGY / DISEASE

Disease Models



Physiologic / Pathologic Phenotype Regulation



Extensive Publications now Substantiating Scientific Approach Probabilistic Causal Bionetwork Models

- **>60 Publications from Rosetta Genetics Group (~30 scientists) over 5 years including high profile papers in PLoS Nature and Nature Genetics**

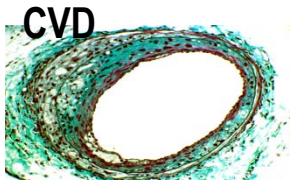
"*Genetics of gene expression surveyed in maize, mouse and man*." **Nature**. (2003)

"*Variations in DNA elucidate molecular networks that cause disease*." **Nature**. (2008)

"*Genetics of gene expression and its effect on disease*." **Nature**. (2008)

"*Validation of candidate causal genes for obesity that affect...*" **Nat Genet**. (2009)

….. Plus 10 additional papers in Genome Research, PLoS Genetics, PLoS Comp.Biology, etc



***"Identification of pathways for atherosclerosis."* Circ Res. (2007)**

"Mapping the genetic architecture of gene expression in human liver." **PLoS Biol.** (2008)

…… Plus 5 additional papers in Genome Res., Genomics, Mamm.Genome

"*Integrating genotypic and expression data …for bone traits…*" **Nat Genet.** (2005)

"..approach to identify candidate genes regulating BMD…" **J Bone Miner Res**. (2009)

"*An integrative genomics approach to infer causal associations* ..." Nat Genet. (2005)

"*Increasing the power to detect causal associations*… "PLoS Comput Biol. (2007)

"Integrating large-scale functional genomic data ..." **Nat Genet.** (2008)

…… Plus 3 additional papers in **PLoS Genet., BMC Genet.**

Society for Pharmacology and Experimental Therapeutics

HE HEPATIC EFFECTS OF FLUTAMIDE IN RATS: A M
WITH CLASSICAL ARYL HYDROCARBON RECEPTOR
ATYPICAL CYP1A INDUCERS[S]

dney D. Ne
Michelle Ca

Chemistry, Un
ubsidiary of Me

Expression profiles of 50 xenobiotic transporter genes in humans and pre-clinical species: A resource for investigations into drug disposition

K. BLEASBY[1,*], J. C. CASTLE[2,*], C. J. ROBERTS[3], C. CHENG[2], W. J. BAILEY[4], J. F. SINA[4], A. V. KULKARNI[2], M. J. HAFEY[1], R. EVERS[1], J.

[1]*Department of D*
[2]*Department of In*
Inpharmatics (a s

OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

PPARα siRNA–Treated Expression Profiles Uncover the Causal Sufficiency Network for Compound-Induced Liver Hypertrophy

Xudong Dai[1*], Angus T. De Souza[2], Hongyue Dai[1], David L. Lewis[3], Chang-kyu Lee[1], Andy G. Spencer[3], Hans Herweijer[3], Jim E. Hagstrom[3], Peter S. Linsley[4], Douglas E. Bassett[1], Roger G. Ulrich[2], Yudong D. He[1*]

1 Informatics, Rosetta
of America, 3 Mirus Bi

Uncovering pat
Developing me
their sufficienc

OPEN ACCESS Freely available online PLoS BIOLOGY

Mapping the Genetic Architecture of Gene Expression in Human Liver

Eric E. Schadt[1☯], Cliona Molony[1☯], Eugene Chudin[1☯], Ke Hao[1], Xia Yang[1], Pek Y. Lum[1], Andrew Kasarskis[1], Bin Zhang[1], Susanna Wang[1], Christine Suver[1], Jun Zhu[1], Joshua Millstein[1], Solveig Sieberts[1], John Lamb[1], Debraj GuhaThakurta[1], Jonathan Derry[1], John D. Storey[1,2,3], Iliana Avila-Campillo[1], Mark J. Kruger[1], Jason M. Johnson[1], Carol A. Rohl[1], Atila van Nas[6], Margarete Mehrabian[4,5], Thomas A. Drake[7], Aldons J. Lusis[4,5,6], Ryan C. Smith[1], F. Peter Guengerich[8,9], Stephen C. Strom[10], Erin Schuetz[11], Thomas H. R

Toxicologic Pathology
http://tpx.sagepub.com

Investigating the Mechanistic Basis for Hepatic Toxicity Induced by an Experimental Chemokin Receptor 5 (CCR5) Antagonist Using a Compendium of Gene Expression Profiles
Paul D. Cornwell and Roger G. Ulrich
Toxicol Pathol 2007; 35; 576
DOI: 10.1080/01926230701383194

List of Influential Papers in Network Modeling

Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks

Xia Yang[1], Joshua L Deignan[1], Hongxiu Qi[1], Jun Zhu[2], Su Qian[3], Judy Zhong[2], Gevork Torosyan[4], Sana Majid[4], Brie Falkard[4], Robert R Kleinhanz[2], Jenny Karlsson[5], Lawrence W Castellani[1], Sheena Mumick[3], Kai Wang[2], Tao Xie[2], Michael Coon[2], Chunsheng Zhang[2], Daria Estrada-Smith[4], Charles R Farber[1], Susanna S Wang[4], Atila van Nas[4], Anatole Ghazalpour[4], Bin Zhang[2], Douglas J MacNeil[3], John R Lamb[2], Katrina M Dipple[4], Marc L Reitman[6], Margarete Mehrabian[1], Pek Y Lum[2], Eric E Schadt[2], Aldons J Lusis[1,4] & Thomas A Drake[5]

Integrative Modeling Defines the Nova Splicing-Regulatory Network and Its Combinatorial Controls

Chaolin Zhang,* Maria A. Frias, Aldo Mele, Matteo Ruggiu, Taesun Eom, Christina B. Marney, Huidong Wang, Donny D. Licatalosi, John J. Fak, Robert B. Darnell*

The transcriptional network for mesenchymal transformation of brain tumours

Maria Stella Carro[1]*†, Wei Keat Lim[2,3]*†, Mariano Javier Alvarez[3,4]*, Robert J. Bollo[8], Xudong Zhao[1], Evan Y. Snyder[9], Erik P. Sulman[10], Sandrine L. Anne[1]†, Fiona Doetsch[5], Howard Colman[11], Anna Lasorella[1,5,6], Ken Aldape[12], Andrea Califano[1,2,3,4] & Antonio Iavarone[1,5,7]

Variations in DNA elucidate molecular networks that cause disease

Yanqing Chen[1]*, Jun Zhu[1]*, Pek Yee Lum[1], Xia Yang[1], Shirly Pinto[2], Douglas J. MacNeil[2], Chunsheng Zhang[1], John Lamb[1], Stephen Edwards[1], Solveig K. Sieberts[1], Amy Leonardson[1], Lawrence W. Castellini[3], Susanna Wang[3], Marie-France Champy[6], Bin Zhang[1], Valur Emilsson[1], Sudheer Doss[3], Anatole Ghazalpour[3], Steve Horvath[4], Thomas A. Drake[5], Aldons J. Lusis[3,4] & Eric E. Schadt[1]

Rewiring of Genetic Networks in Response to DNA Damage

Sourav Bandyopadhyay,[1] Monika Mehta,[2] Dwight Kuo,[3] Min-Kyung Sung,[4] Ryan Chuang,[3] Eric J. Jaehnig,[5] Bernd Bodenmiller,[6] Katherine Licon,[1] Wilbert Copeland,[3] Michael Shales,[7] Dorothea Fiedler,[7,8] Janusz Dutkowski,[1] Aude Guénolé,[9] Haico van Attikum,[9] Kevan M. Shokat,[7,8] Richard D. Kolodner,[5,1,10] Won-Ki Huh,[4] Ruedi Aebersold,[6] Michael-Christopher Keogh,[2*] Nevan J. Krogan,[7*] Trey Ideker[1,3,10*]

Although cellular behaviors are dynamic, the networks that govern these behaviors have been mapped primarily as static snapshots. Using an approach called differential epistasis mapping, we have discovered widespread changes in genetic interaction among yeast kinases, phosphatases, and transcription factors as the cell responds to DNA damage. Differential interactions uncover many gene functions that go undetected in static conditions. They are very effective at identifying DNA repair pathways, highlighting new damage-dependent roles for the Slt2 kinase, Pph3 phosphatase, and histone variant Htz1. The data also reveal that protein complexes are generally stable in response to perturbation, but the functional relations between these complexes are substantially reorganized. Differential networks chart a new type of genetic landscape that is invaluable for mapping cellular responses to stimuli

Cell

Resource

An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man

Timothy Ravasi, Harukazu Suzuki, Carlo Vittorio Cannistraci, Shintaro Katayama, Vladimir B. Bajic, Kai Tan, Altuna Akalin, Sebastian Schmeier, Mutsumi Kanamori-Katayama, Nicolas Bertin, Piero Carninci, Carsten O. Daub, Alistair R.R. Forrest, Julian Gough, Sean Grimmond, Jung-Hoon Han, Takehiro Hashimoto, Winston Hide, Oliver Hofmann, Atanas Kamburov, Mandeep Kaur, Hideya Kawaji, Atsutaka Kubosaki, Timo Lassmann, Erik van Nimwegen, Cameron Ross MacPherson, Chihiro Ogawa, Aleksandar Radovanovic, Ariel Schwartz, Rohan D. Teasdale, Jesper Tegnér, Boris Lenhard, Sarah A. Teichmann, Takahiro Arakawa, Noriko Ninomiya, Kayoko Murakami, Michihira Tagami, Shiro Fukuda, Kengo Imamura, Chikatoshi Kai, Ryoko Ishihara, Yayoi Kitazume, Jun Kawai, David A. Hume, Trey Ideker, and Yoshihide Hayashizaki

The FANTOM Consortium

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Network-Based Elucidation of Human Disease Similarities Reveals Common Functional Modules Enriched for Pluripotent Drug Targets

Silpa Suthram, Joel T. Dudley, Annie P. Chiang, Rong Chen, Trevor J. Hastie, Atul J. Butte

computational BIOLOGY

ANALYSIS

nature biotechnology

Genome-wide identification of post-translational modulators of transcription factor activity in human B cells

Kai Wang, Masumichi Saito, Brygida C Bisikirska, Mariano J Alvarez, Wei Keat Lim, Presha Rajbhandari, Qiong Shen, Ilya Nemenman, Katia Basso, Adam A Margolin, Ulf Klein, Riccardo Dalla-Favera & Andrea Califano

LETTER

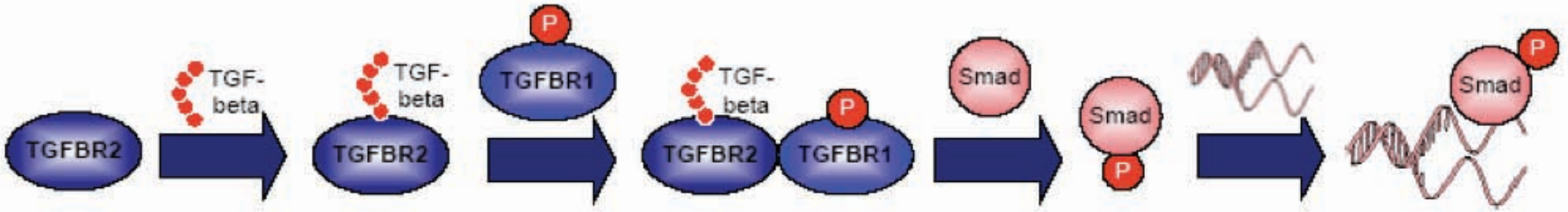
doi:10.1038/nature093

A *trans*-acting locus regulates an anti-viral expression network and type 1 diabetes risk

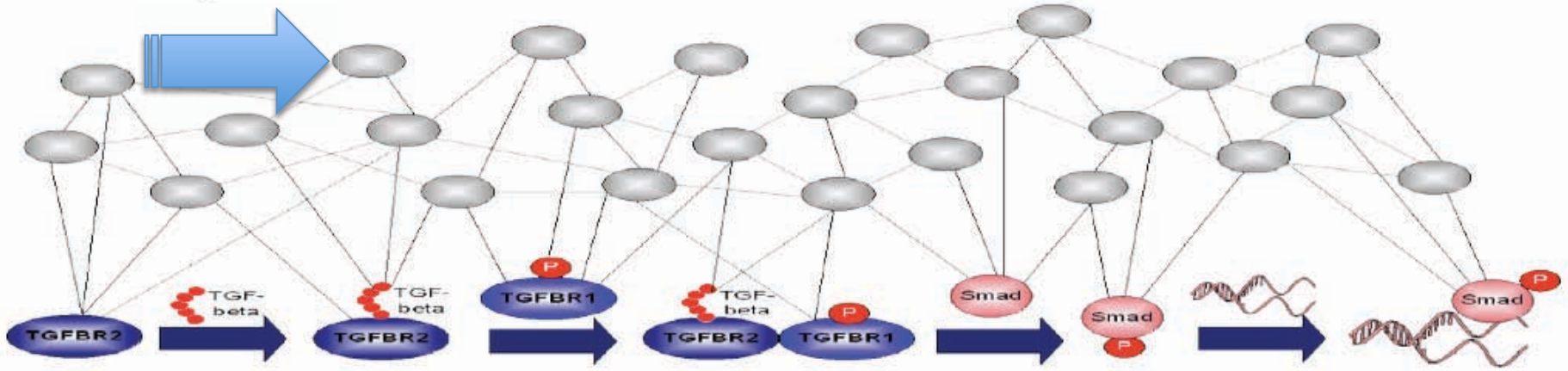
Matthias Heinig, Enrico Petretto, Chris Wallace, Leonardo Bottolo, Maxime Rotival, Han Lu, Yoyo Li, Rizwan Sarwa, Sarah R. Langley, Anja Bauerfeind, Oliver Hummel, Young-Ae Lee, Svetlana Paskas, Carola Rintisch, Kathrin Saar, Jason Cooper, Rachel Buchan, Elizabeth E. Gray, Jason G. Cyster, Cardiogenics Consortium†, Jeanette Erdmann, Christian Hengstenberg, Seraya Maouche, Willem H. Ouwehand, Catherine M. Rice, Nilesh J. Samani, Heribert Schunkert, Alison H. Goodall, Herbert Schulz, Helge G. Roider, Martin Vingron, Stefan Blankenberg, Thomas Münzel, Tanja Zeller, Silke Szymczak, Andreas Ziegler, Laurence Tiret, Deborah J. Smyth, Michal Pravenec, Timothy J. Aitman, Francois Cambien, David Clayton, John A. Todd, Norbert Hubner & Stuart A. Cook

- 50 network papers
- http://sagebase.org/research/resources.php

The way we like to think:



The way it is:

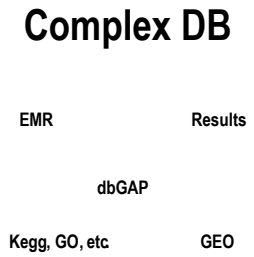


(Eric Schadt)

What's needed to play/compete in the information space?

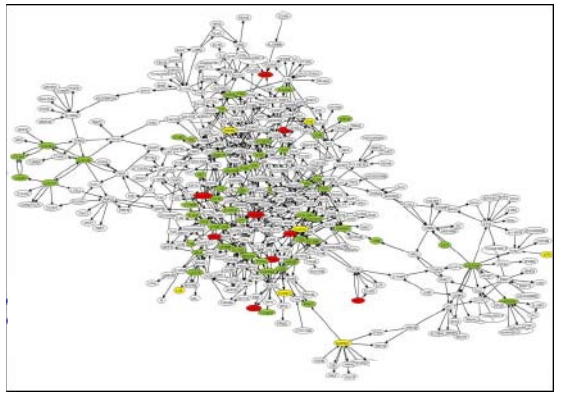
- Three critical components:
  - Big databases, organized (connected) to facilitate integration and model building
    - GE Health, IBM, Microsoft, Google, and so on

1)



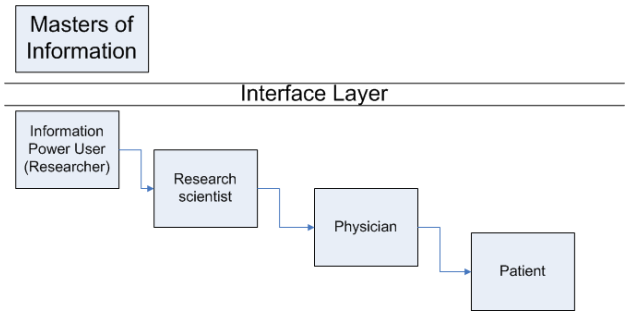
  - Data integration and construction of predictive models
    - Computational, math/stat, high-performance computing, and biological expertise
    - Significant high-performance computing resources

2)



  - Tools and educational resources to translate complex material to a hierarchy of of "users"and ways to cite model-publish

3)



Recognition that the benefits of bionetwork based molecular models of diseases are powerful but that they **require significant resources**

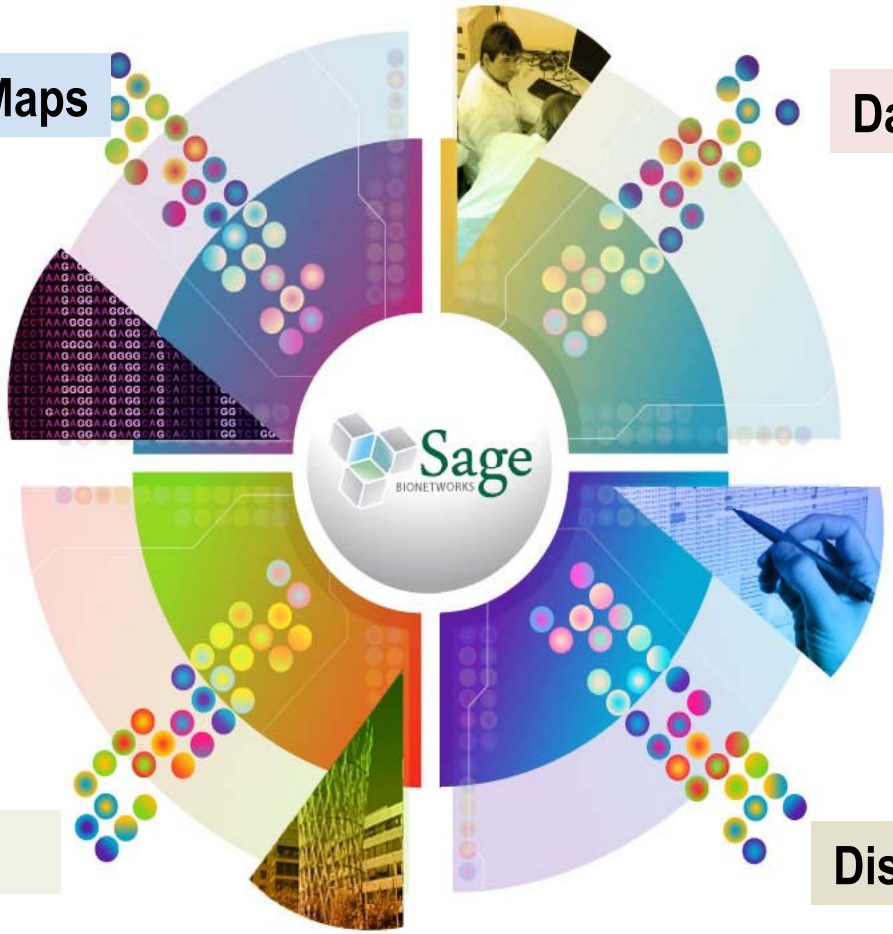
Appreciation that it will **require decades** of evolving representations as real complexity emerges and needs to be integrated with therapeutic interventions

Sage Mission

Sage Bionetworks is a non-profit organization with a vision to create a “commons” where integrative bionetworks are evolved by contributor scientists with a shared vision to accelerate the elimination of human disease

Building Disease Maps

Data Repository

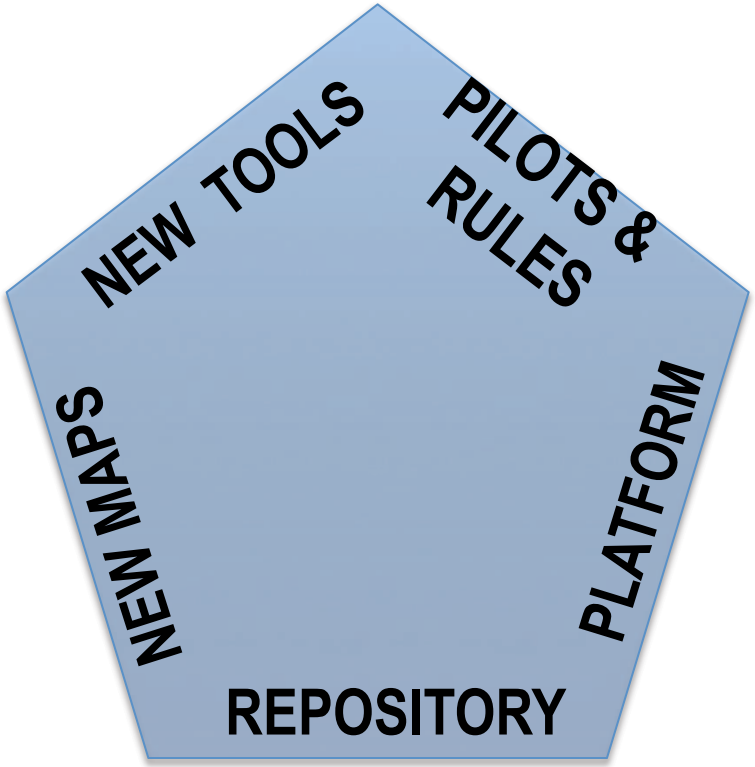


Commons Pilots

Discovery Platform

Sagebase.org

Sage Bionetworks Strategy: Integrate with Communities of Interest



Map Users-
Disease Map and Tool Users-
(Scientists, Industry, Foundations, Regulators...)

Platform Builders –
Sage Platform and Infrastructure Builders-
(Academic Biotech and Industry IT Partners...)

Barrier Breakers-
Data Sharing Barrier Breakers-
(Patients Advocates, Governance and Policy Makers, Funders...)

Data Generators-
Data Tool and Disease Map Generators-
(Global coherent data sets, Cytoscape, Clinical Trialists, Industrial Trialists, CROs…)

Commons Pilots-
Data Sharing Commons Pilots-
(Federation, CCSB, Inspire2Live....)

Sage Bionetworks Functional Organization

Platform

Curation/Annotation

Data Repository

CTCAP
Public Data
Merck Data
TCGA/ICGC

Hosting Data
Hosting Tools Hosting
Models

Discovery Platform

LSDF

Commons

Outposts
Federation
CCSB

Commons Pilots

LSDF-WPP
Inspire2Live
POC

Research

Cancer
Neurological Disease
Metabolic Disease

Building Disease Maps

Pfizer
Merck
Takeda
Astra Zeneca
CHDI
Gates
NIH

Bayesian Models
Co-expression Models

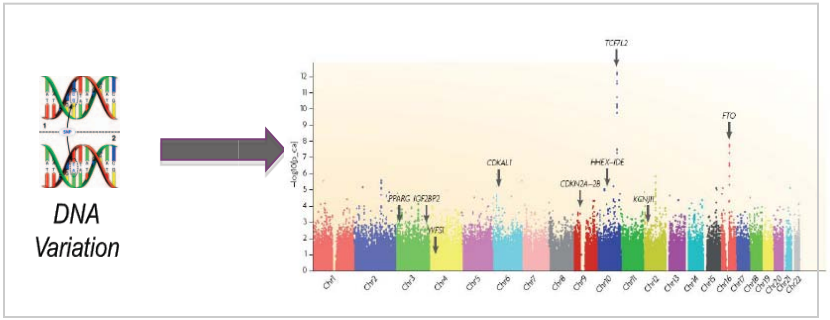
Tools & Methods

KDA/GSVA

Sage Bionetworks Collaborators

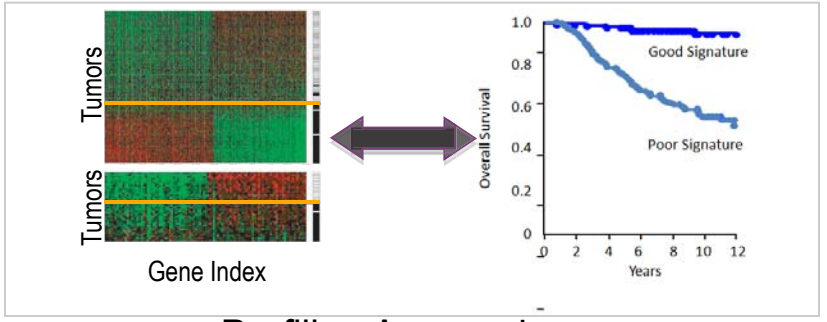
- Pharma Partners
  - Merck, Pfizer, Takeda, Astra Zeneca, Amgen
- Foundations
  - CHDI, Gates Foundation
- Government
  - NIH, LSDF
- Academic
  - Levy (Framingham)
  - Rosengren (Lund)
  - Krauss (CHORI)
- Federation
  - Ideker, Califarno, Butte, Schadt

Building Integrated Models



"Standard" GWAS Approaches

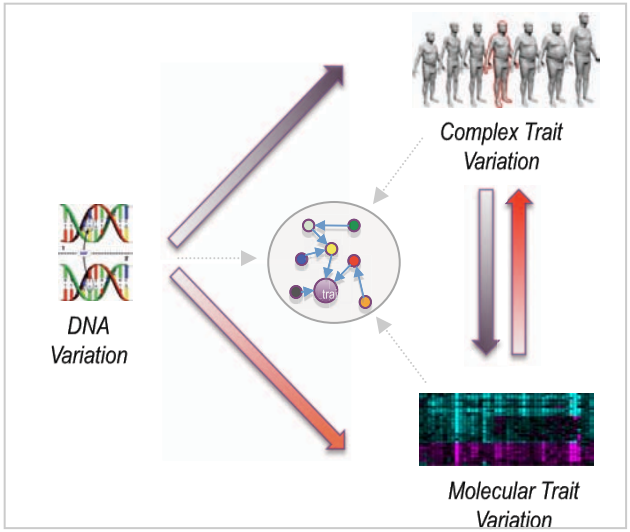
Identifies Causative DNA Variation but provides NO mechanism



Profiling Approaches

Genome scale profiling provide correlates of disease

- Many examples BUT what is cause and effect?

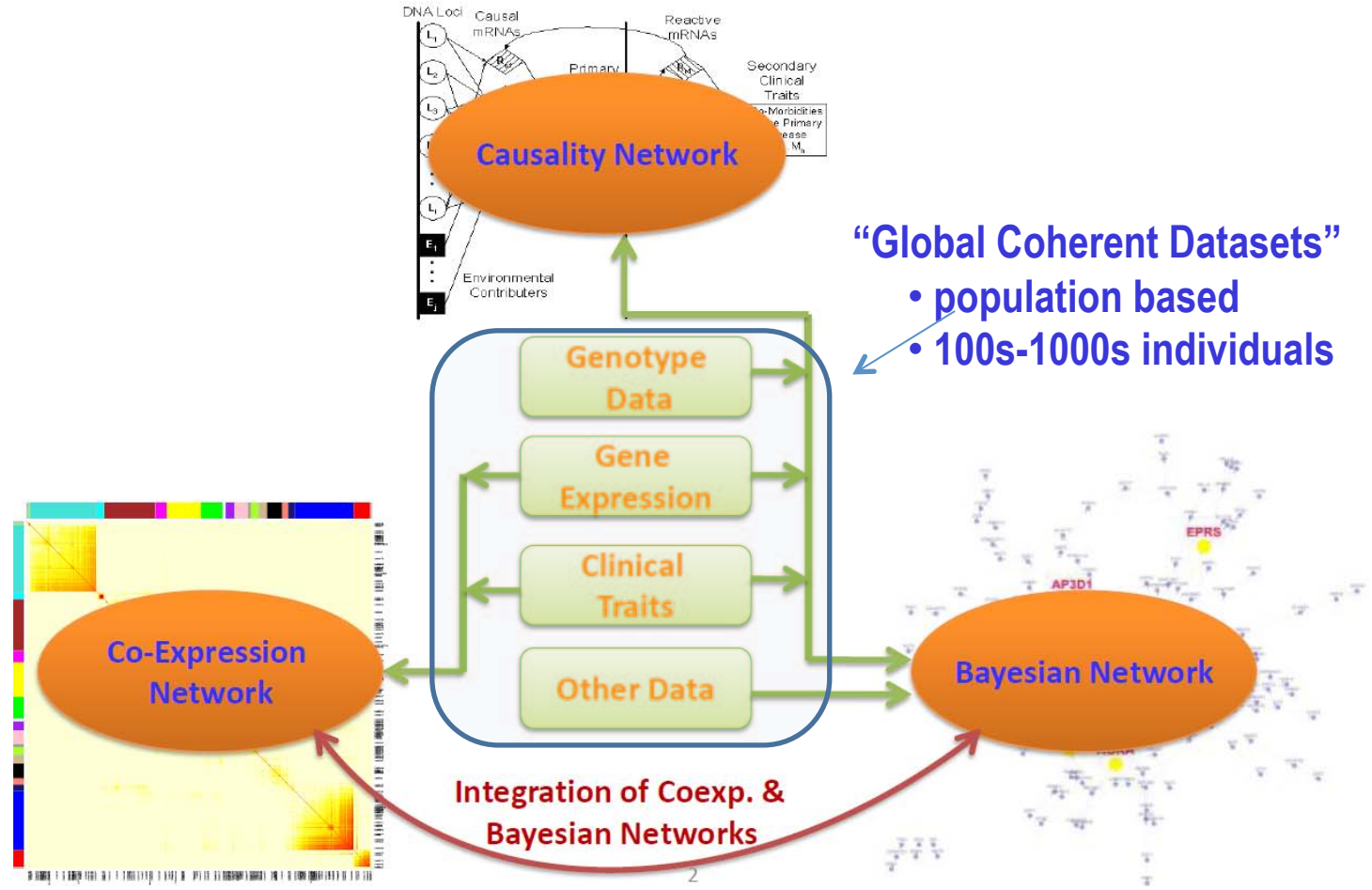


"Integrated" Genetics Approaches

- Provide unbiased view of molecular physiology as it relates to disease phenotypes
- Insights on mechanism
- Provide causal relationships and allows predictions

Methods for Integration of Genotypic, Gene Expression & Clinical Data

Schadt et al. Nature Genetics 37: 710 (2005)
Millstein et al. BMC Genetics 10: 23 (2009)



"Global Coherent Datasets"
- population based
- 100s-1000s individuals

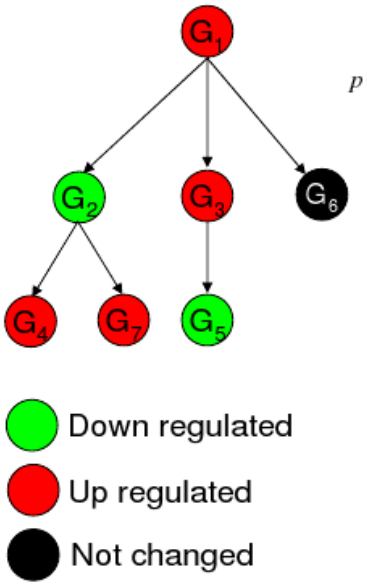
Chen et al. Nature 452:429 (2008)
Zhang & Horvath. Stat.Appl.Genet.Mol.Biol. 4: article 17 (2005)

Zhu et al. Cytogenet Genome Res. 105:363 (2004)
Zhu et al. PLoS Comput. Biol. 3: e69 (2007)

Bayesian Networks

(does Gene A control Gene B?)

- Captures the stochastic nature of biological system
- Probability based
- Can include priors such as causality information
- PREDICTIVE: *CAN be used to predict outcomes or perturbations*



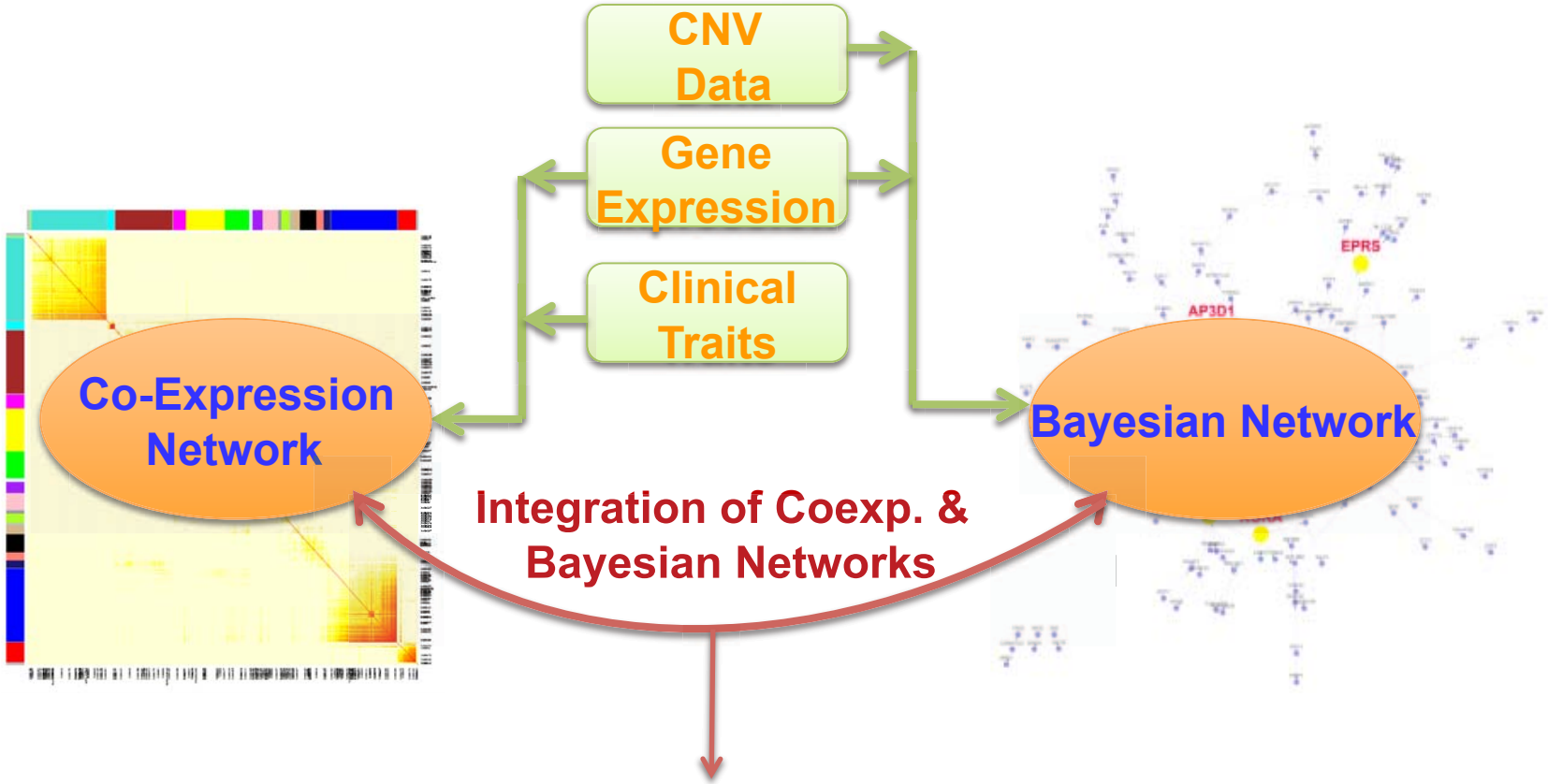
- BN method provides a way to decompose a joint probability distribution based on conditional independence

$$p(Network) = p(G_1, G_2, G_3, G_4, G_5, G_6, G_7)$$
$$= p(G_1)p(G_2 \mid G_1)p(G_3 \mid G_1)p(G_6 \mid G_1)p(G_4 \mid G_2)p(G_7 \mid G_2)p(G_5 \mid G_3)$$

- For a given network, we find the maximum likelihood of the network given the observed data *D*, *p(D|Network)*
- Training Bayesian Networks
  - We want to search the space of all networks to find the optimal one
  - Calculate probability tables associated with the networks
- To find the best network we perform the search 1,000 times using random seeds
  - Computationally intense procedure
  - Presently runs on a 6000+ CPU (IBM Blade) Cluster
- Common features are then extracted (e.g., connections seen in > 30% of the networks are extracted) and probability tables are updated

Integration of Multiple Networks for Pathway and Target Identification

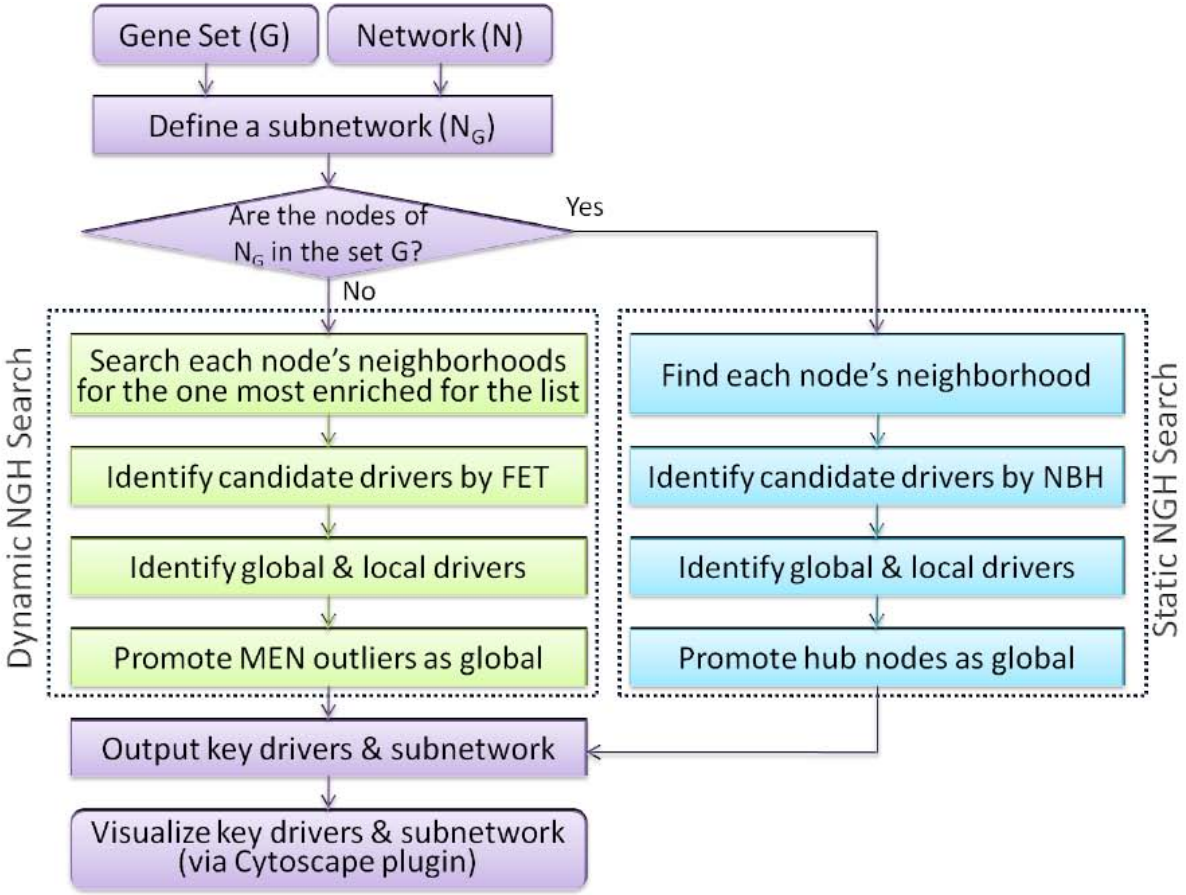
Bin Zhang
Jun Zhu

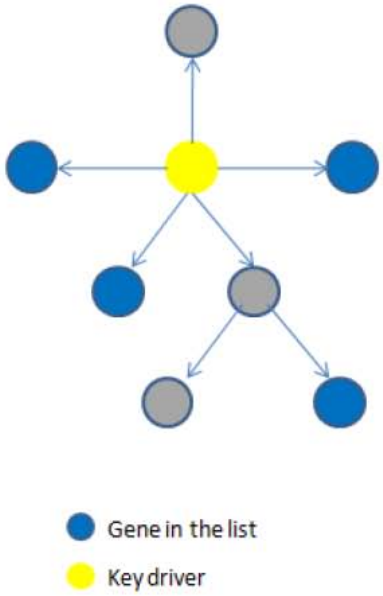


Key Driver Analysis

Key Driver Analysis

Bin Zhang
Jun Zhu
Justin Guinney



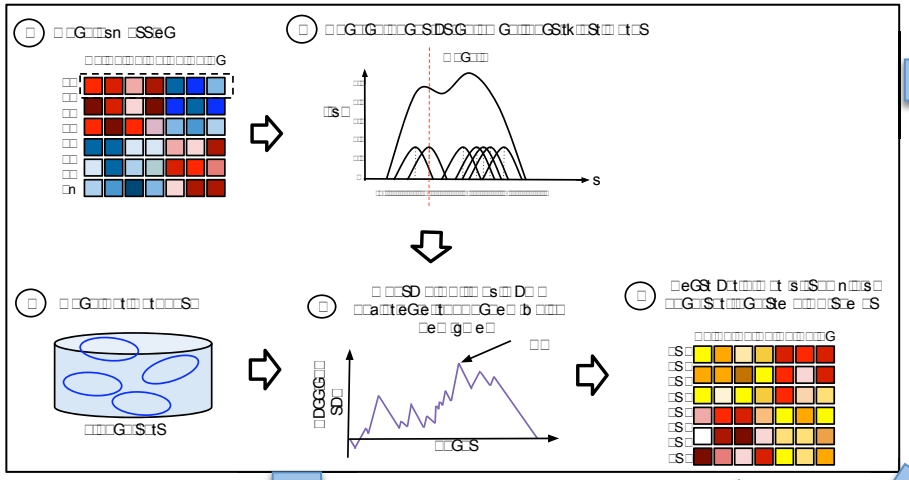


3

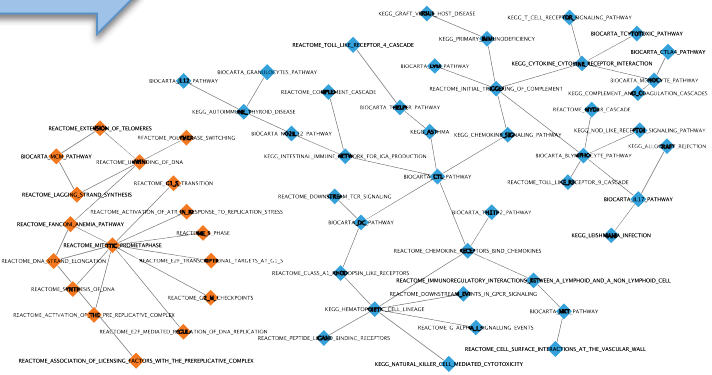
http://sagebase.org/research/tools.php

Gene Set Variation Analysis (GSVA)

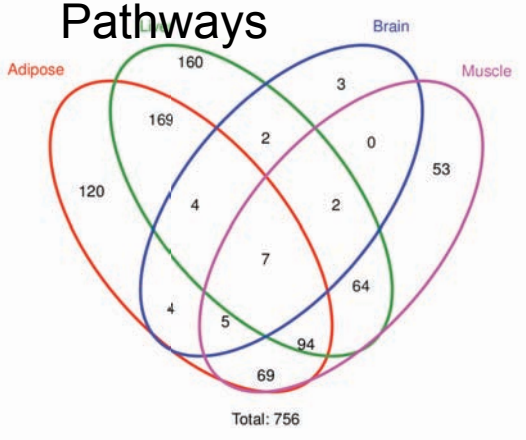
Justin Guinney
Sonja Haenzelmann



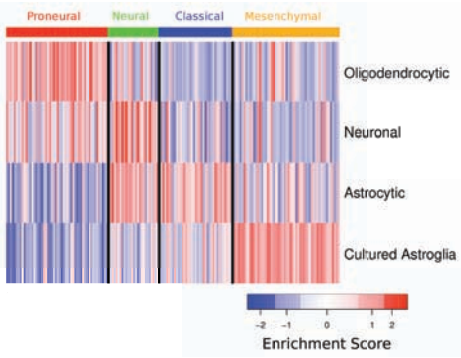
Meta-pathways



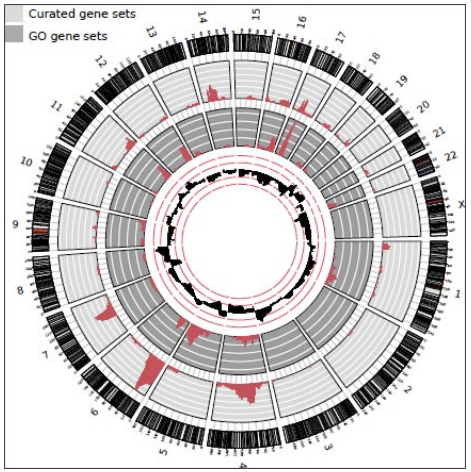
Cross-tissue Pathways



Pathway Clustering



Pathway CNV



Model of Breast Cancer: Co-expression

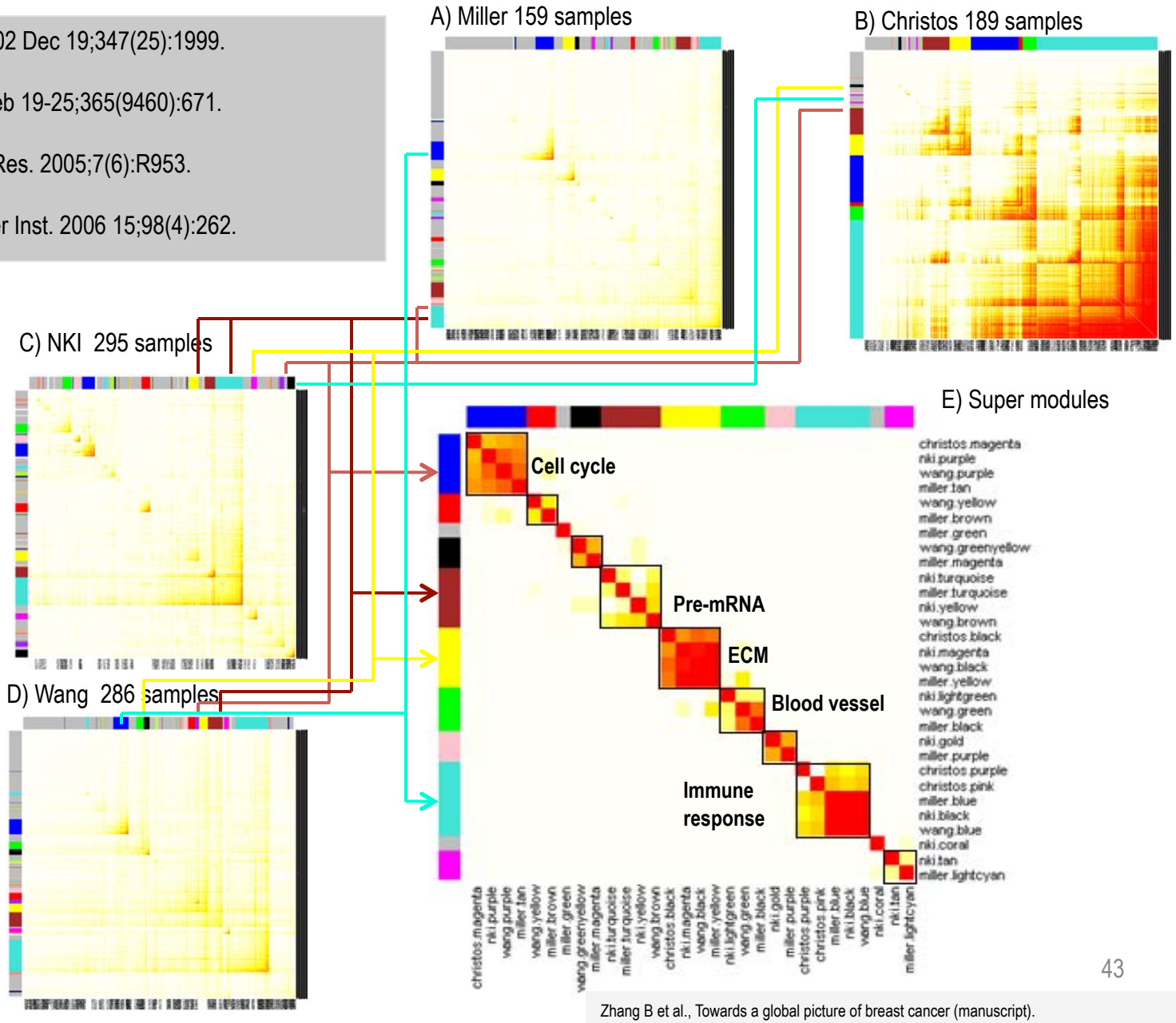
Bin Zhang
Xudong Dai
Jun Zhu

NKI: N Engl J Med. 2002 Dec 19;347(25):1999.

Wang: Lancet. 2005 Feb 19-25;365(9460):671.

Miller: Breast Cancer Res. 2005;7(6):R953.

Christos: J Natl Cancer Inst. 2006 15;98(4):262.



Zhang B et al., Towards a global picture of breast cancer (manuscript).

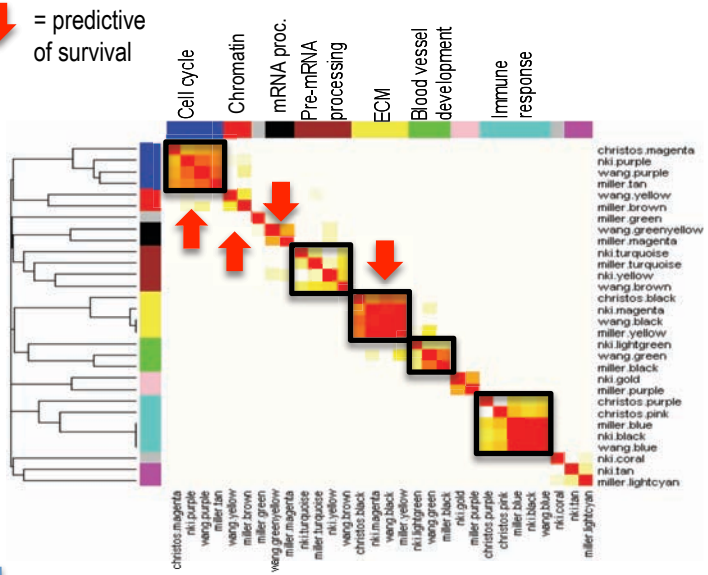
Model of Breast Cancer: Integration

Bin Zhang
Xudong Dai
Jun Zhu

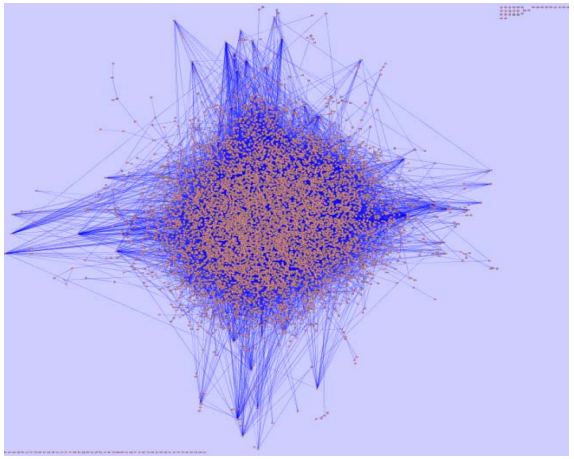
Sage BIONETWORKS

Conserved Super-modules

= predictive of survival



Breast Cancer Bayesian Network

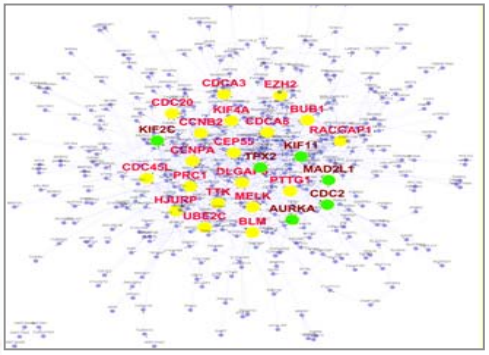


Extract gene:gene relationships for selected super-modules from BN and define Key Drivers

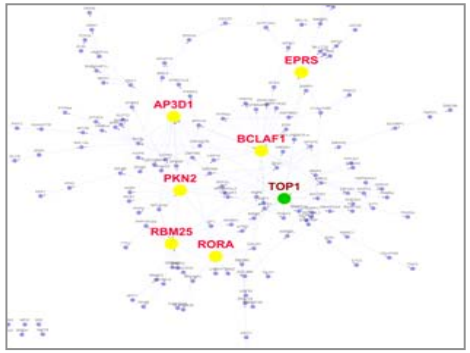
Pathways & Regulators

(Key drivers=yellow; key drivers validated in siRNA screen=green)

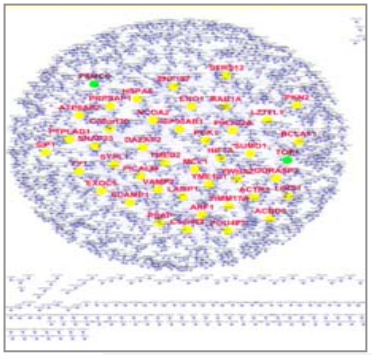
Cell Cycle (Blue)



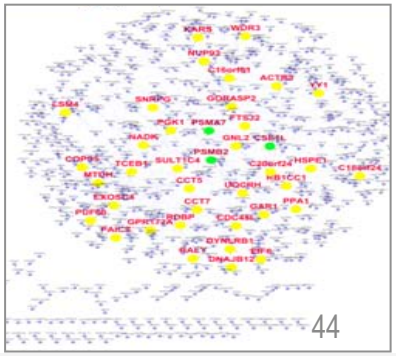
Chromatin Modification (Black)



Pre-mRNA proc. (Brown)



mRNA proc. (red)

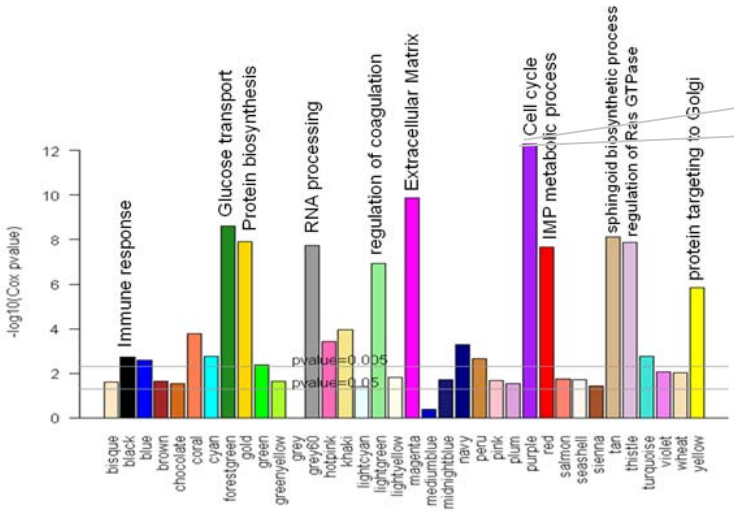


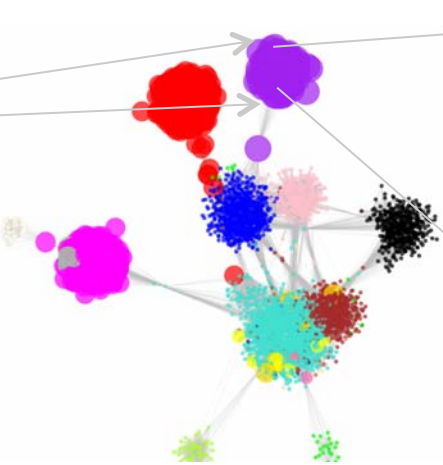
Zhang B et al., Key Driver Analysis in Gene Networks (manuscript)

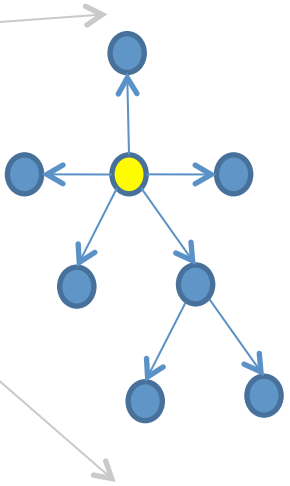
Model of Breast Cancer: Mining

Bin Zhang
Xudong Dai
Jun Zhu

Co-expression sub-networks predict survival; KDA identifies drivers







Co-expression modules correlate with survival

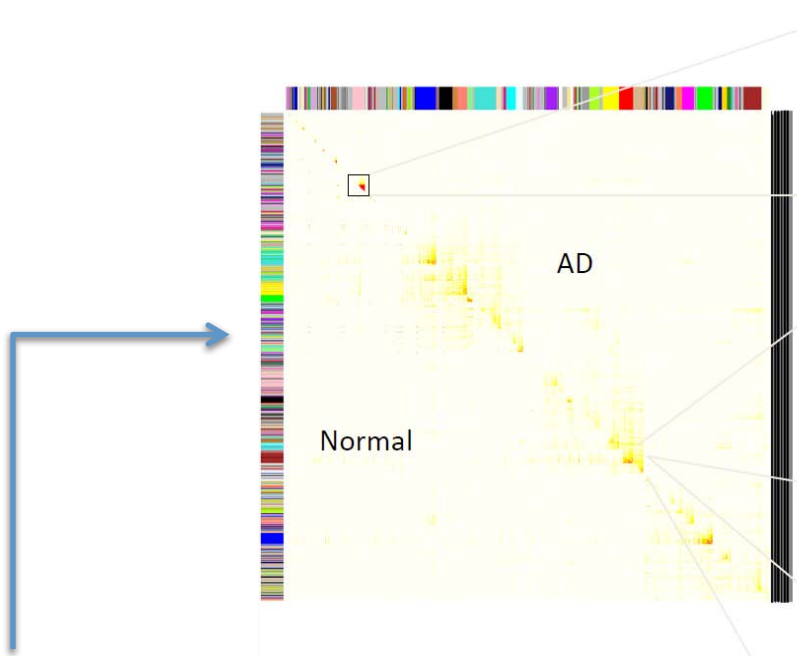
Map to Bayesian Network

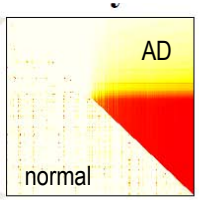
Define Key Drivers

Model of Alzheimer's Disease

Bin Zhang
Jun Zhu

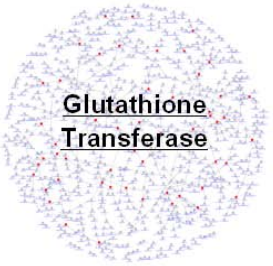
Sage BIONETWORKS

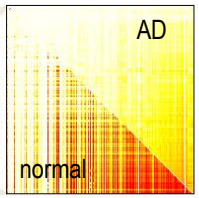




Glutathione transferase

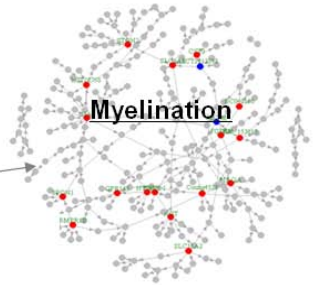
GOC (91 fold)

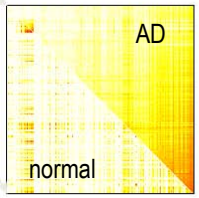




nerve ensheathment

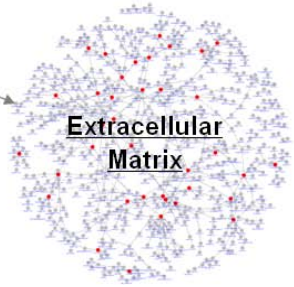
LOC (40%)





extracellular matrix

GOC (1.9 fold)



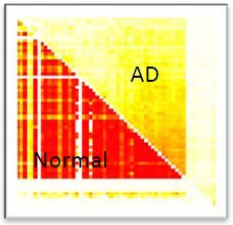
Harvard Brain Tissue Resource Center

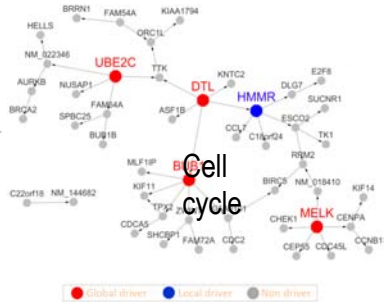
Species: Human
Tissue: Brain visual cortex,Brain cerebellu
Disease: Neurological Disease
Investigator: Francine Benes/ Eric Schadt
Institution: Harvard Brain Tissue Resource Ce
Approximate Number Subjects: 803

Abstract:

This study aims at identifying functional variation in the human genome (especially as it pertains to brain expressed RNAs) and elucidate its relationship to disease and drug response. The ~800 individuals in this dataset are composed of approximately 400 Alzheimers disease (AD) cases, 230 Huntington's Disease and 170 controls matched for age, gender, and post mortem interval (PMI). The tissue specimens for this study were provided by Harvard Brain Tissue Resource Center (HBTRC).

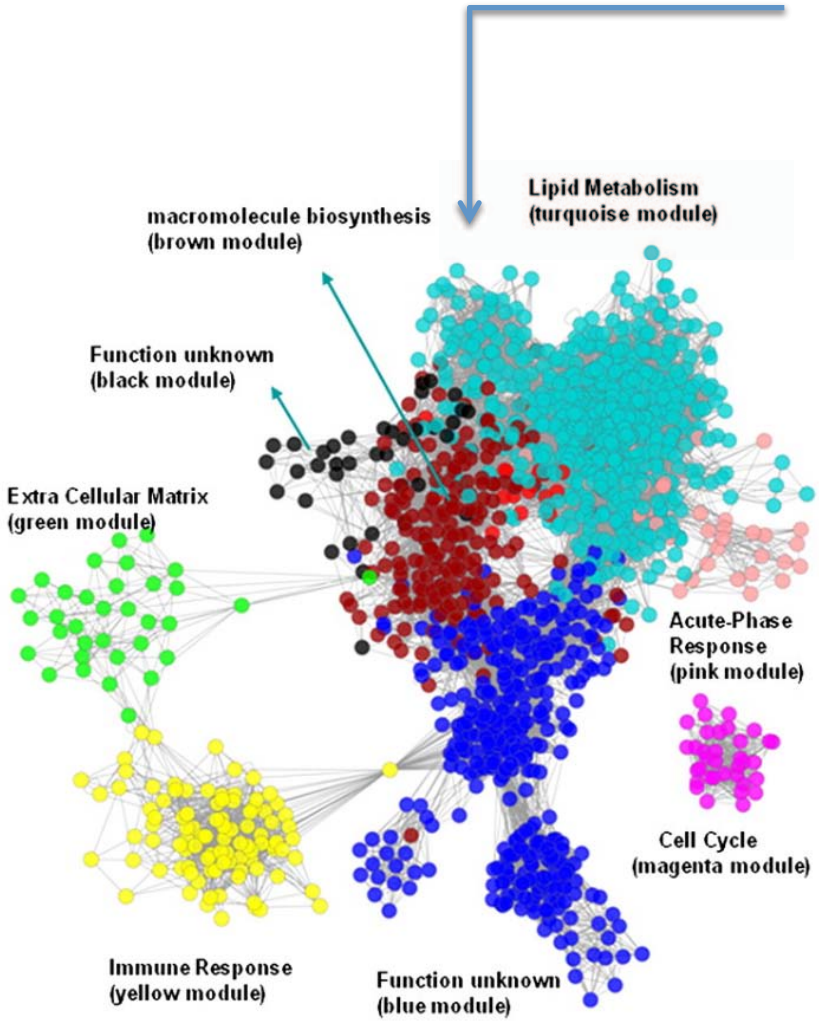
http://sage.fhcrc.org/downloads/downloads.php

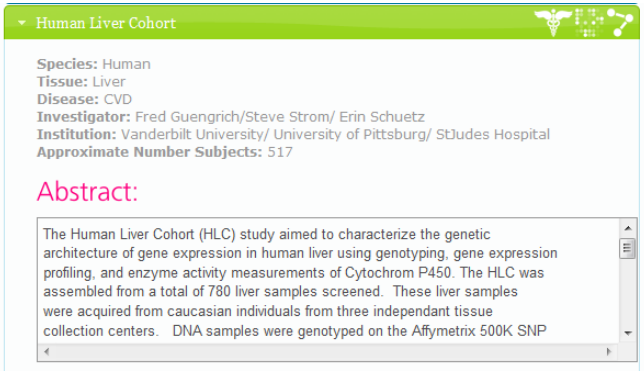




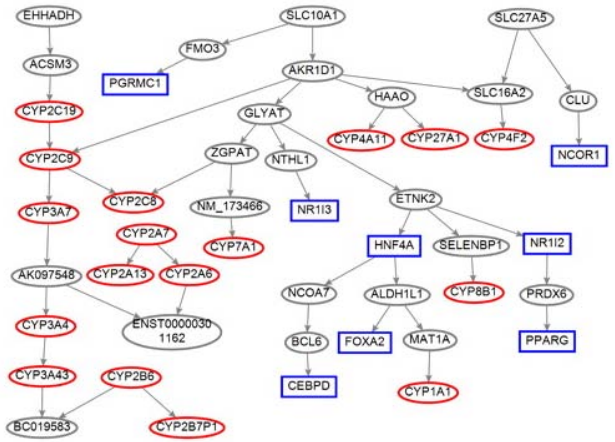
Liver Cytochrome P450 Regulatory Network Models

Xia Yang
Bin Zhang
Jun Zhu





http://sage.fhcrc.org/downloads/downloads.php



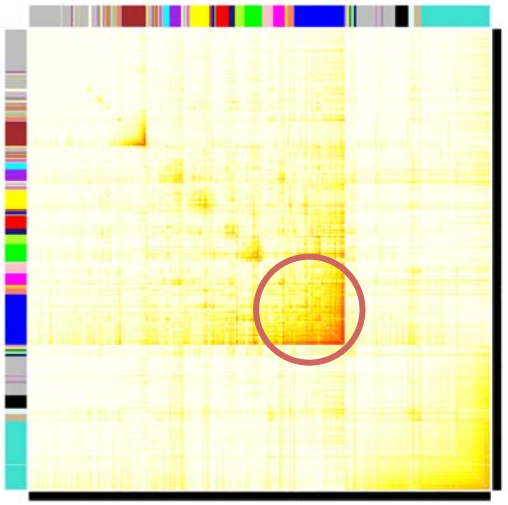
Regulators of P450 network

Yang et al. Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver. 2010. Genome Research 20:1020.

New Type II Diabetes Disease Models

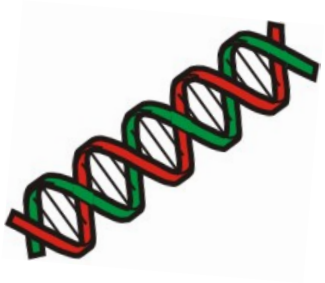
Anders Rosengren
LUND UNIVERSITY

Global expression data
from 64 human islet donors



Blue module: 3000 genes
Associated with
Type 2 diabetes
Elevated HbA1c
Reduced insulin secretion

340 genes in islet-specific
open chromatin regions



168 overlapping genes, which have

- Higher connectivity
- Markedly stronger association with
  - Type 2 diabetes
  - Elevated HbA1c
  - Reduced insulin secretion
- Enrichment for beta-cell transcription factors and exocytotic proteins

New Type II Diabetes Disease Models

Anders Rosengren
Lund University

- **Search across 1300 datasets in MetaGEO at Sage for similar expression profiles**
  Top hit: Islet dedifferentiation study where the 168 genes were upregulated in mature islets and downregulated in dedifferentiated islets (*Kutlu et al., Phys Gen 2009*)

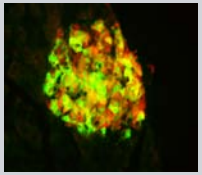
- Analyses of expression-SNPs and clinical SNPs as well as Causal Inference Test

↓

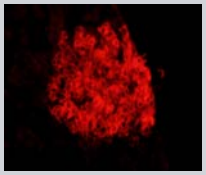
- Identification of candidate key genes affecting beta-cell differentiation and chromatin

Working hypothesis:

Normal beta-cell: open chromatin in islet-specific regions, high expression of beta-cell transcription factors, differentiated beta-cells and normal insulin secretion



Diabetic beta-cell: lower expression of beta-cell transcription factors affecting the identified module, dedifferentiation, reduced insulin secretion and hyperglycemia



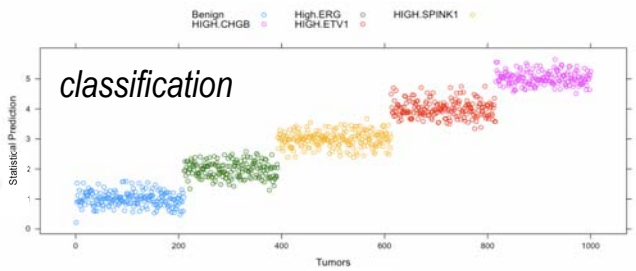
↓

Next steps: Validation of hypothesis and suggested key genes in human islets

Validating Prostate Cancer Models

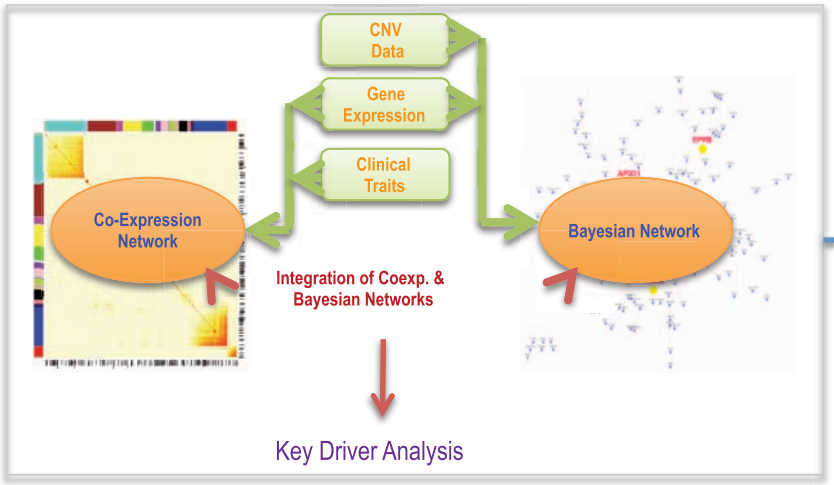
Brig Mecham
Xudong Dai
Pete Nelson
Rich Klingoffer

Gene Expression Data on >1000 prostate cancer samples (GEO)



Gene Expression & CNV Data ~200 prostate cancers (Taylor et al)

Gene Expression & CNV Data ~120 rapid autopsy Mets (Nelson)



Gene Expression & CNV Data ~30 prostate xenografts (Nelson)

siRNA Screen Data (Nelson)

Integrated network analysis

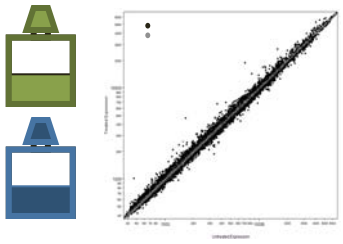
Key Drivers Matched to Xenografts for validations with Presage Technology

Systems biology approach to pharmacogenomics

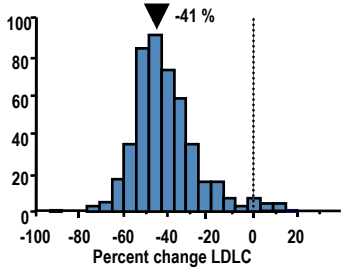
Lara Mangravite

Ron Krauss

Molecular simvastatin response

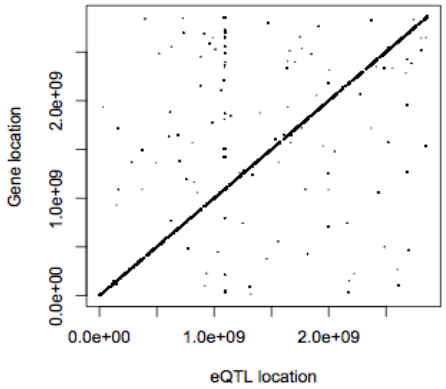


Clinical simvastatin response



Simon et al, Am J Cardiol 2006

Integrative Genomic Analysis



Ongoing:

Cellular validation of novel genes and SNPs involved in statin efficacy and cellular cholesterol homeostasis

Examples: The Sage Non-Responder Project in Cancer

Purpose:

- To identify Non-Responders to approved drug regimens so we can improve outcomes, spare patients unnecessary toxicities from treatments that have no benefit to them, and reduce healthcare costs

Leadership:

- Co-Chairs Stephen Friend, Todd Golub, Charles Sawyers & Rich Schilsky

Initial Studies:

- AML (at first relapse)
- Non-Small Cell Lung Cancer
- Ovarian Cancer (at first relapse)
- Breast Cancer
- Renal Cell
- Multiple Myeloma

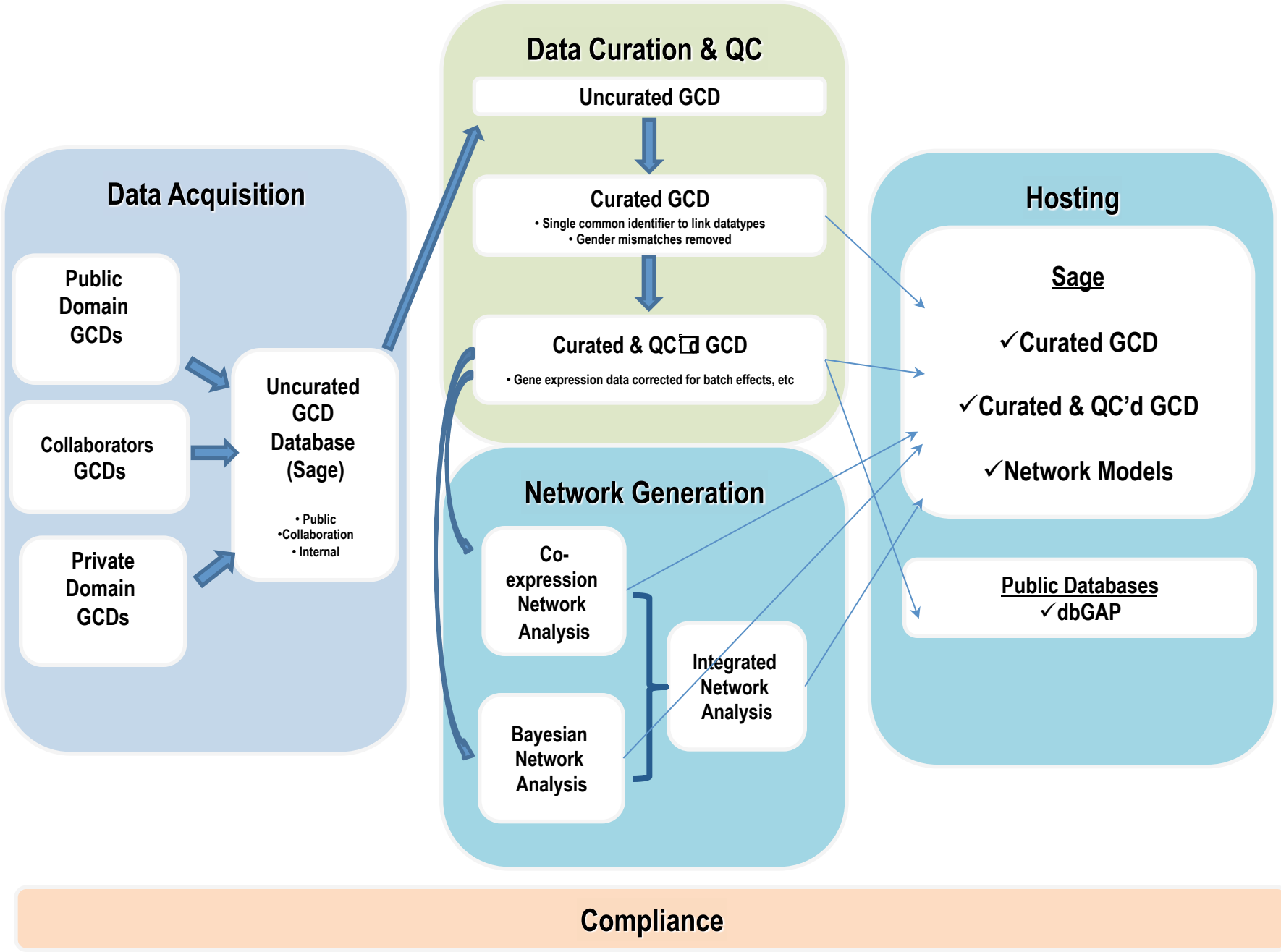
Clinical Trial Comparator Arm Partnership (CTCAP)

Partnering for Cures | A FasterCures Meeting

Bridging the Chasm Between Microscope and Marketplace

- **Description**: Collate, Annotate, Curate and Host Clinical Trial Data with Genomic Information from the Comparator Arms of Industry and Foundation Sponsored Clinical Trials: Building a Site for Sharing Data and Models to evolve better Disease Maps.
- **Public-Private Partnership** of leading pharmaceutical companies, clinical trial groups and researchers.
- **Neutral Conveners**: Sage Bionetworks and Genetic Alliance [nonprofits].
- **Initiative to share existing trial data** (molecular and clinical) from non-proprietary comparator and placebo arms to create powerful new tool for drug development.

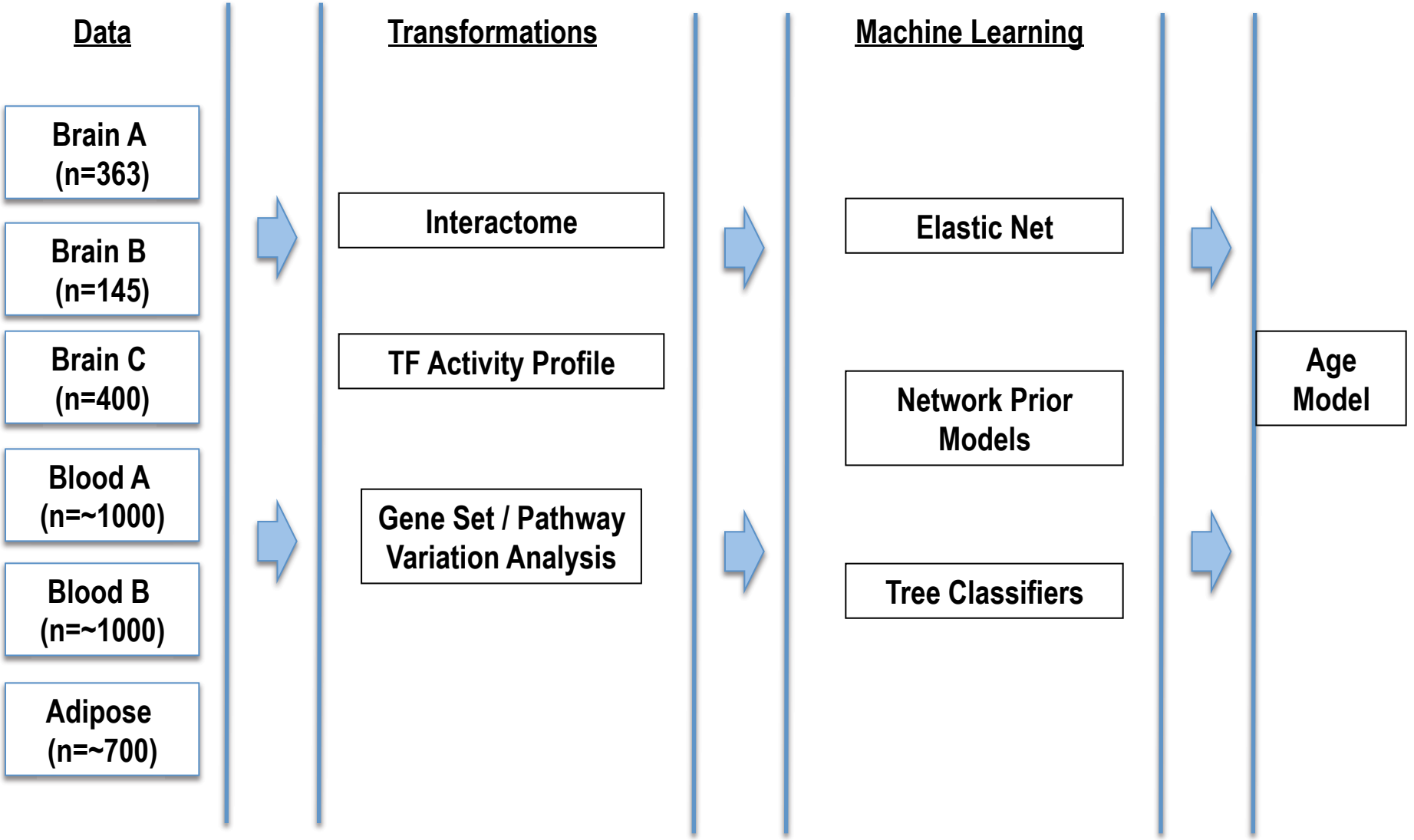
CTCAP Workstreams



Examples: The Sage Federation

- Founding Lab Groups
  - Seattle- Sage Bionetworks
  - New York- Columbia: Andrea Califano
  - Palo Alto- Stanford: Atul Butte
  - San Diego- UCSD: Trey Ideker
  - San Francisco: UCSF/Sage: Eric Schadt
- Initial Projects
  - Aging
  - Diabetes
  - Warburg
- Goals: *Share all datasets, tools, models*
  Develop interoperability for human data

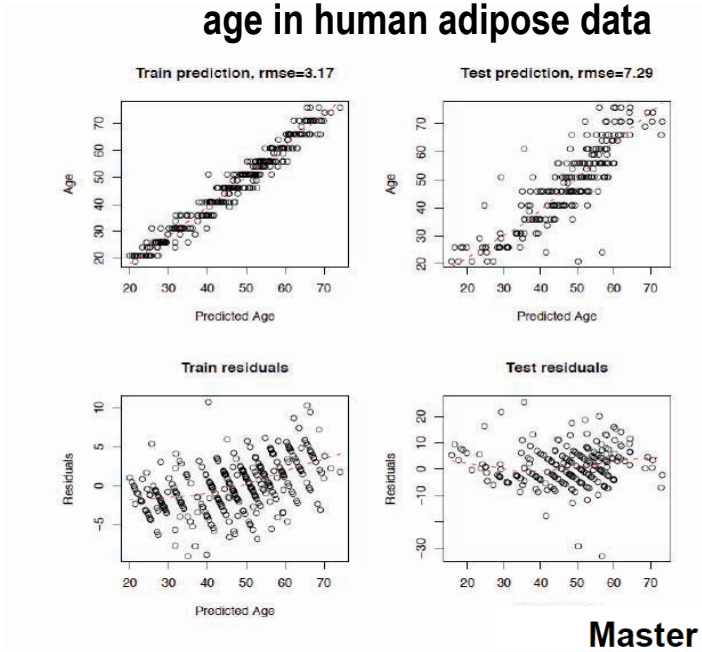
Human Aging Project



Preliminary Results

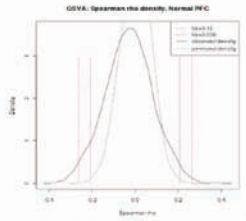
Adipose Age Prediction

multivariate logistic regression model predicting age in human adipose data



GSVA: Prefrontal cortex, normal (neg)

MSigDB (n=2871)

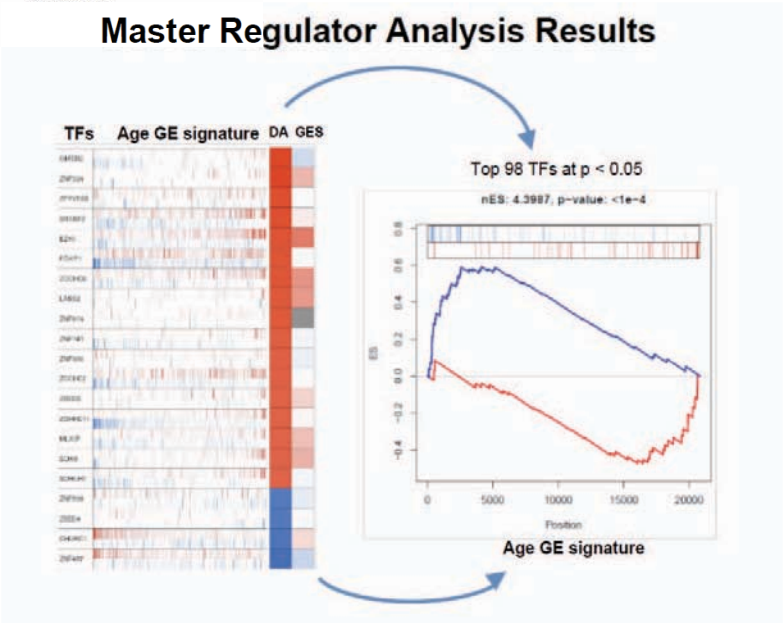


empirical FDR
.12 @ alpha=.01
.04 @ alpha=.001

+ Terpenoids shown to improve blood flow (contained in Ginkgo)
* Decreased mTOR activity has been found to slow aging in S. cervisiae, C. elegans, & D. melangoster

Rank	Geneset name	Spearman rho
1+	KEGG_TERPENOID_BACKBONE_BIOSYNTHESIS	-.36
2	BIOCARTA_CDC42RAC_PATHWAY	-.36
4	BIOCARTA_ACTINY_PATHWAY	-.33
8	KEGG_ALZHEIMERS_DISEASE	-.31
12	ELVIDGE_HYPOXIA_BY_DMOG_DN	-.31
16	WEINMANN_ADAPTATION_TO_HYPOXIA_UP	-.30
19	BIOCARTA_AKAP95_PATHWAY	-.29
20	BIOCARTA_EIF_PATHWAY	-.29
23	KEGG_PARKINSONS_DISEASE	-.29
27	ELVIDGE_HIF1A_TARGETS_UP	-.28
30*	BIOCARTA_MTOR_PATHWAY	-.28
31	ELVIDGE_HYPOXIA_DN	-.27
40	WONG_MITOCHONDRIA_GENE_MODULE	-.27
42	BIOCARTA_ETC_PATHWAY	-.27
43	**LU_AGING_BRAIN_UP**	-.27
44	REACTOME_ELECTRON_TRANSPORT_CHAIN	-.26
48	BIOCARTA_PROTEASOME_PATHWAY	-.26

Master Regulator Analysis (MARINa) from Califano's lab.



Warburg Effect Studied by the Federation's Genome-wide Network and Modeling Approach

Warburg effect: the association of aerobic glycolysis, an inefficient way for ATP generation, with cancer cell and their progression. Linked with rapidly dividing cells.

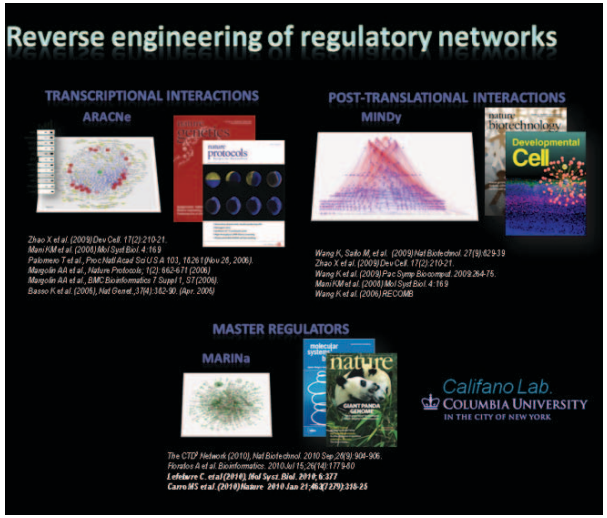
Two Key Questions:

1. **Are cancer cells genetically decoupled from the altered metabolism that is seen in rapidly dividing cells?**

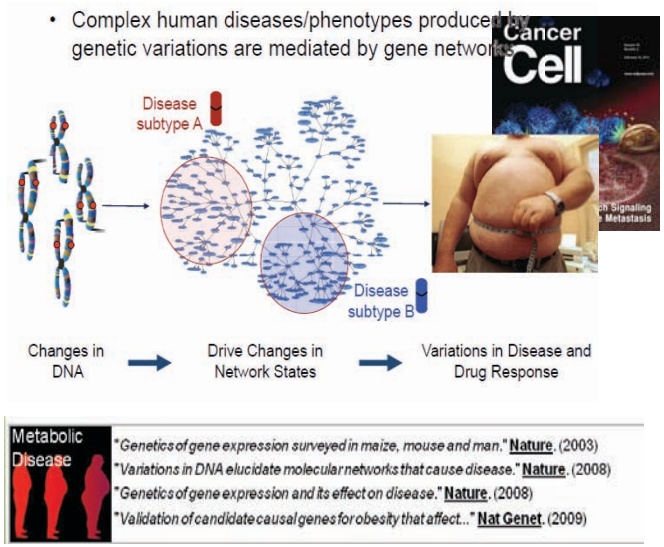
2. **Is there evidence that cancer outcomes are associated with altered metabolic circuits?**

Federation's Genome-wide Network and Modeling Approach

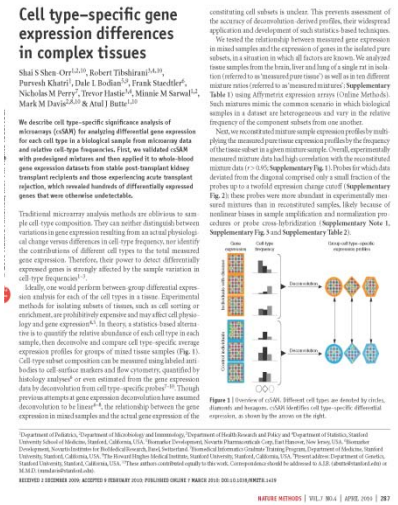
Califano group at Columbia



Sage Bionetworks

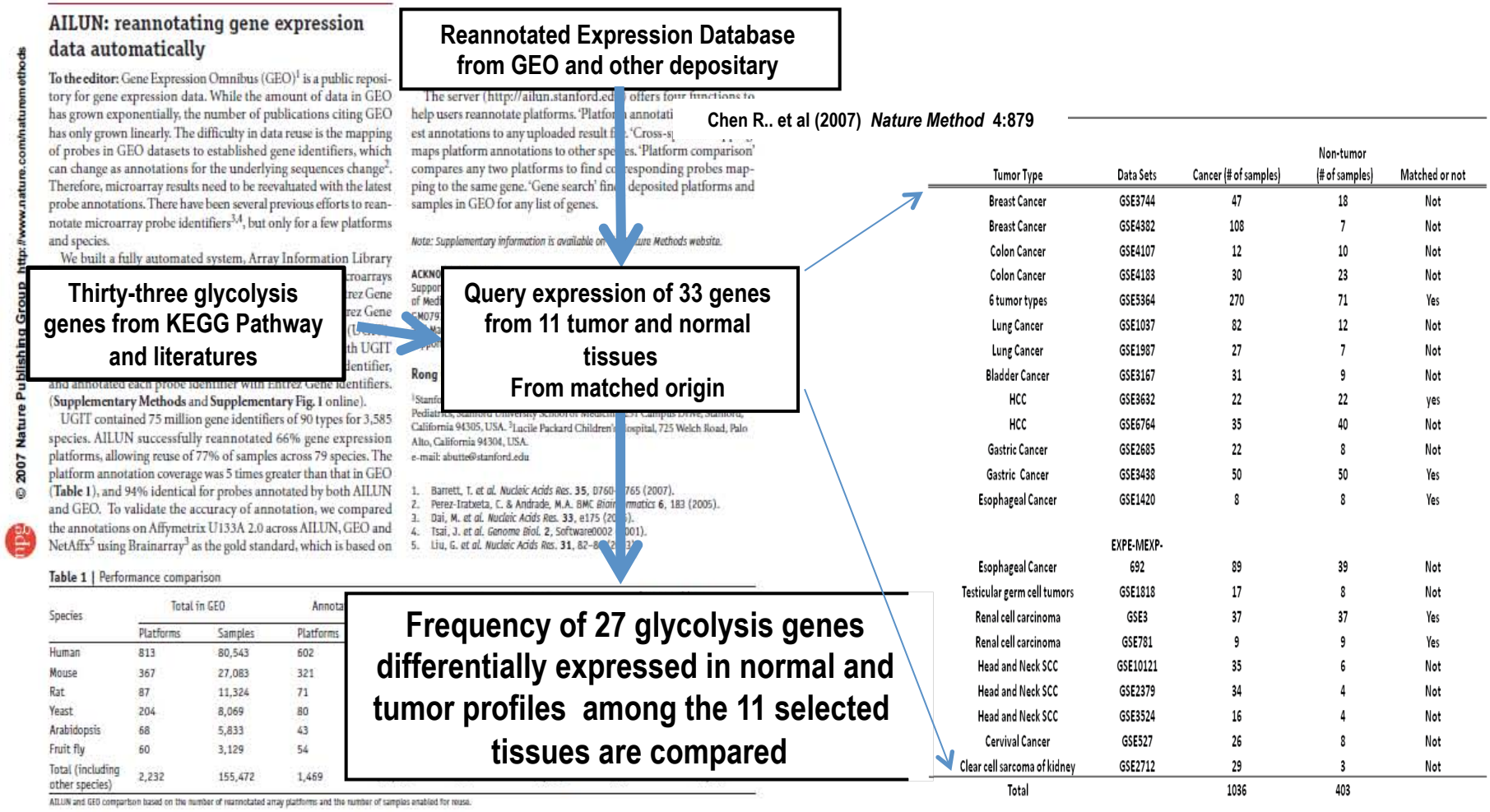


Butte group at Stanford



Characterizing Pattern of Glycolysis Genes in both Tumor and Normal Tissues of the Matched Origin

Andy Beck's approach (Butte Lab)



AILUN: reannotating gene expression data automatically

To the editor: Gene Expression Omnibus (GEO)[1] is a public repository for gene expression data. While the amount of data in GEO has grown exponentially, the number of publications citing GEO has only grown linearly. The difficulty in data reuse is the mapping of probes in GEO datasets to established gene identifiers, which can change as annotations for the underlying sequences change[2]. Therefore, microarray results need to be reevaluated with the latest probe annotations. There have been several previous efforts to reannotate microarray probe identifiers[3,4], but only for a few platforms and species.

We built a fully automated system, Array Information Library

and annotated each probe identifier with Entrez Gene identifiers. (**Supplementary Methods** and **Supplementary Fig. 1** online).

UGIT contained 75 million gene identifiers of 90 types for 3,585 species. AILUN successfully reannotated 66% gene expression platforms, allowing reuse of 77% of samples across 79 species. The platform annotation coverage was 5 times greater than that in GEO (**Table 1**), and 94% identical for probes annotated by both AILUN and GEO. To validate the accuracy of annotation, we compared the annotations on Affymetrix U133A 2.0 across AILUN, GEO and NetAffx[5] using Brainarray[3] as the gold standard, which is based on

The server (http://ailun.stanford.edu) offers four functions to help users reannotate platforms. 'Platform annotation' finds annotations to any uploaded result file. 'Cross-species' maps platform annotations to other species. 'Platform comparison' compares any two platforms to find corresponding probes mapping to the same gene. 'Gene search' finds deposited platforms and samples in GEO for any list of genes.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

Support of Medical Scientist Training Program GM079...

Rong ...

[1]Stanford University School of Medicine, Department of Pediatrics, Stanford University School of Medicine, 251 Campus Drive, Stanford, California 94305, USA. [2]Lucile Packard Children's Hospital, 725 Welch Road, Palo Alto, California 94304, USA.
e-mail: abutte@stanford.edu

1. Barrett, T. *et al. Nucleic Acids Res.* **35**, D760–765 (2007).
2. Perez-Iratxeta, C. & Andrade, M.A. *BMC Bioinformatics* **6**, 183 (2005).
3. Dai, M. *et al. Nucleic Acids Res.* **33**, e175 (2005).
4. Tsai, J. *et al. Genome Biol.* **2**, Software0002.1–0002.1001 (2001).
5. Liu, G. *et al. Nucleic Acids Res.* **31**, 82–86 (2003).

Table 1 | Performance comparison

Species	Total in GEO		Annotated
	Platforms	Samples	Platforms
Human	813	80,543	602
Mouse	367	27,083	321
Rat	87	11,324	71
Yeast	204	8,069	80
Arabidopsis	68	5,833	43
Fruit fly	60	3,129	54
Total (including other species)	2,232	155,472	1,469

AILUN and GEO comparison based on the number of reannotated array platforms and the number of samples enabled for reuse.

Reannotated Expression Database from GEO and other depositary

Thirty-three glycolysis genes from KEGG Pathway and literatures

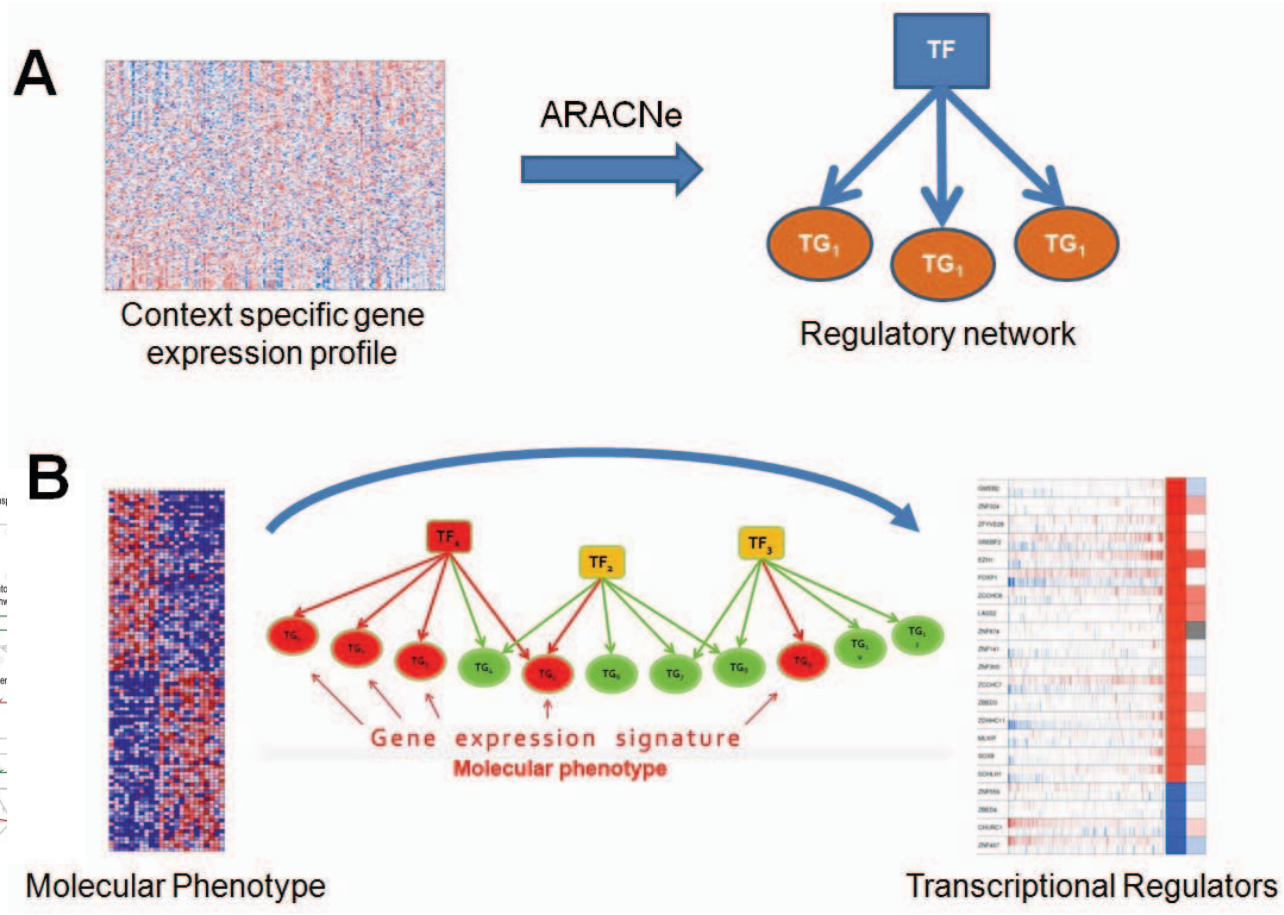
Query expression of 33 genes from 11 tumor and normal tissues From matched origin

Frequency of 27 glycolysis genes differentially expressed in normal and tumor profiles among the 11 selected tissues are compared

Chen R.. et al (2007) *Nature Method* 4:879

Tumor Type	Data Sets	Cancer (# of samples)	Non-tumor (# of samples)	Matched or not
Breast Cancer	GSE3744	47	18	Not
Breast Cancer	GSE4382	108	7	Not
Colon Cancer	GSE4107	12	10	Not
Colon Cancer	GSE4183	30	23	Not
6 tumor types	GSE5364	270	71	Yes
Lung Cancer	GSE1037	82	12	Not
Lung Cancer	GSE1987	27	7	Not
Bladder Cancer	GSE3167	31	9	Not
HCC	GSE3632	22	22	yes
HCC	GSE6764	35	40	Not
Gastric Cancer	GSE2685	22	8	Not
Gastric Cancer	GSE3438	50	50	Yes
Esophageal Cancer	GSE1420	8	8	Yes
	EXPE-MEXP-			
Esophageal Cancer	692	89	39	Not
Testicular germ cell tumors	GSE1818	17	8	Not
Renal cell carcinoma	GSE3	37	37	Yes
Renal cell carcinoma	GSE781	9	9	Yes
Head and Neck SCC	GSE10121	35	6	Not
Head and Neck SCC	GSE2379	34	4	Not
Head and Neck SCC	GSE3524	16	4	Not
Cervival Cancer	GSE527	26	8	Not
Clear cell sarcoma of kidney	GSE2712	29	3	Not
Total		1036	403	

Deriving Master Regulators from Transcription Factors Regulatory Networks Glycolysis & Glycogenesis



Inferred Transcriptional Factors Regulating GGMSE in Tumors from Variety of Tissues

TFs for genes of interest (in Prostate cancer)

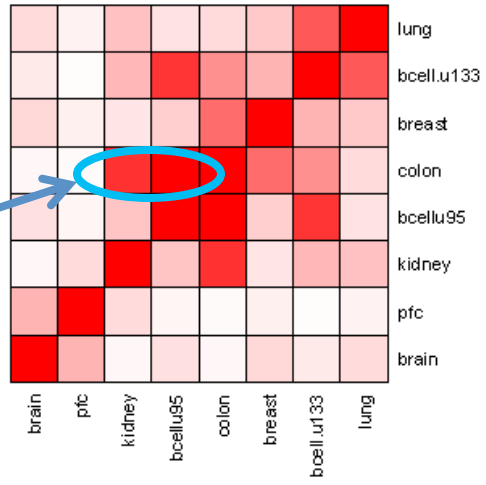
Gene of Interest	TFs that regulate this gene						
ALDOB	DMRTC2	RFX4	ST18	VSX1	TFAP2D	FOXR1	ISX
	NEUROD6		ESRRB	OTX2	HSF5	FOXG1	AIRE
PGK1	CARHSP1	ILF2	MTA2	TFDP1	ZNF135	CEBPG	
	ZNHIT3	NME1-NME2					
ALDOA	ZC3H7B	NFE2L1	THRA	GATAD2B	ZNF454	ZBTB4	NR1H2
	STAT6	SUPT6H	STAT5B	FOXJ2	CUX2	ZSCAN23	MTA2
	MSRB2	ETV6	ZMAT3	TCFL5	PSMC3IP	MRPL28	FOXK2
ENO1 (TF)	TRIM25	ZFP91	TCF21	ZNF135	RFX5	MTA2	ZFR
	NOTCH2	HNRNPAB		ILF2	ZNF789	NME1-NME2	
	PA2G4	NME2	NANOG	NFE2L3	PRKDC		

Overlap between transcriptional regulators of metabolism

The following tables shows $-\log_{10}(p-value)$ from FET.

Glycolysis

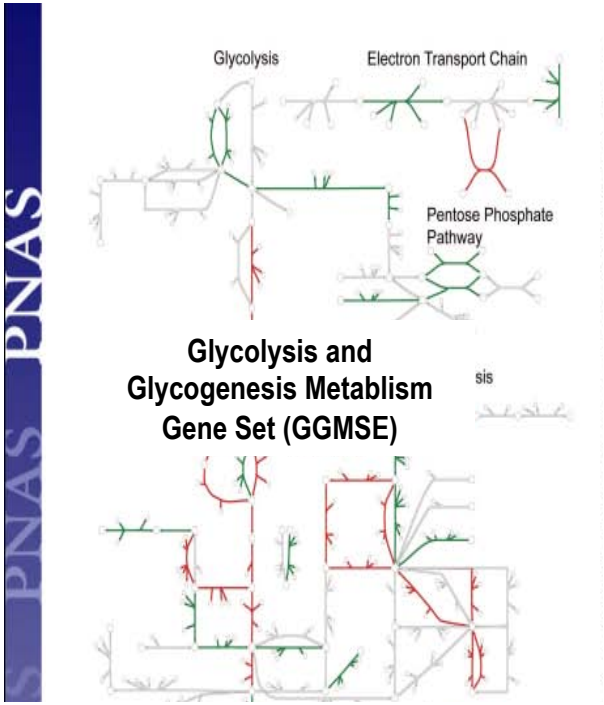
	brain	pfc	kidney	bcellu95	colon	breast	bcell.u133	lung
brain		1.28	0.14	0.54	0.17	0.67	0.37	0.62
pfc	1.28		0.65	0.21	0.09	0.28	0.07	0.24
kidney	0.14	0.65		1.04	3.51	0.47	1.25	1.09
bcellu95	0.54	0.21	1.04		4.35	0.85	3.47	0.52
colon	0.17	0.09	3.51	4.35		2.50	1.87	0.65
breast	0.67	0.28	0.47	0.85	2.50		1.30	0.92
bcell.u133	0.37	0.07	1.25	3.47	1.87	1.30		2.85
lung	0.62	0.24	1.09	0.52	0.65	0.92	2.85	



(a) glycolysis

Inferring Prostate Cancer Regulatory Modules for Glycolysis &Glycogenesis Metabolism Pathway

Sage bionetworks' approach



Glycolysis and Glycogenesis Metablism Gene Set (GGMSE)

Duarte N. et al (2006) *PNAS* 107(6):1777-1782

Prostate cancer global coherent data set (GSE21032)

Taylor BS. et al (2010) *Cancer Cell* 18(1):11-22

Integrated Bayesian Approach

Zhu J. et al (2008) Nature Genetics 40(7): 854-61

ARTICLES

nature genetics

Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks

Jun Zhu[1], Bin Zhang[1], Erin N Smith[2,3], Becky Drees[4], Rachel B Brem[5], Leonid Kruglyak[2], Roger E Bumgarner[4] & Eric E Schadt[1]

Inferred Transcriptional Regulatory Network in Prostate Cancer

Prostate Cancer Regulatory Modules for GGMSE and Other Metabolism Pathways

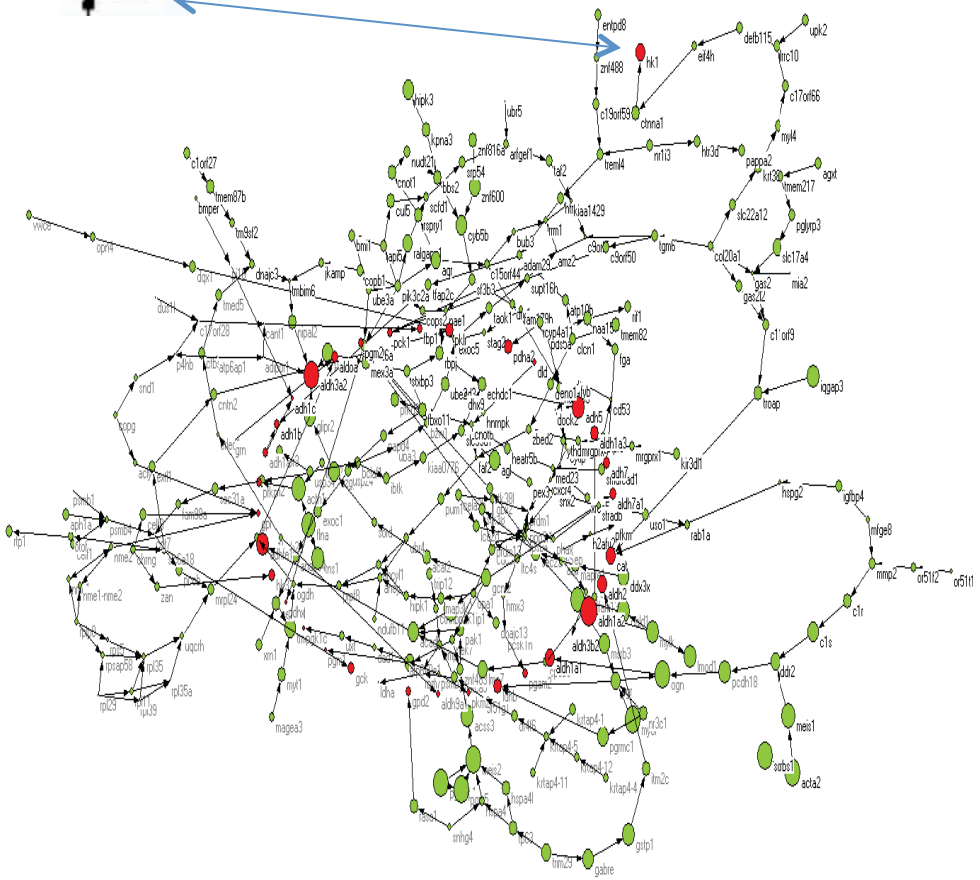
Cox Proportional-Hazards Regression model based on individual gene for recurrence free survival

Metabolism pathways with regulatory modules enriched by poor prognosis genes for prostate cancer

Genes Associated with Poor Prognosis are disproportionally found among the networks regulating the "glycolysis" Genes

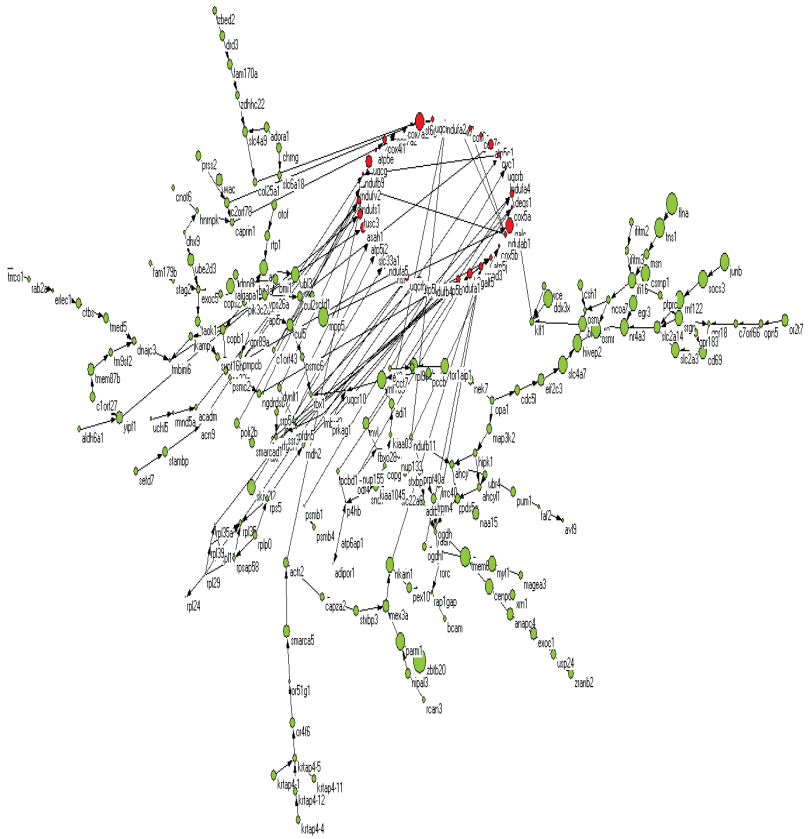
P-Value<0.005

Size of the node proportional to -log10 P value for recurrence free survival.



Inferred regulatory module for GGMSE

>5 fold enrichment of recurrence free prognostic genes with the Glycolysis BN module than random selection (p<1e-100)



Inferred regulatory module for Oxidative Phosphorylation and Sphingolipid Metabolism genes

THE FEDERATION

Butte Califano Friend Ideker Schadt

vs

Sage Bionetworks 22 publications in last year

Reproducible science==shareable science

Sweave: combines programmatic analysis with narrative

Dynamic generation of statistical reports using literate data analysis

```
297
298<<>>=
299data(leukemia)
300leukemia_eset
301head(pData(leukemia_eset))
302table(leukemia_eset$subtype)
303@
304Let's examine the variability of the expression profiles across samples by
305plotting the cumulative distribution of IQR values as shown in Figure~\ref{figIQR}.
306About 50\% of the probesets show very limited variability across samples
307and, therefore, in the following non-specific filtering step we will filter
308out this fraction from further analysis.
309
310<<figIQR, echo=FALSE, results=hide>>=
311png(filename="GSVA-figIQR.png", width=500, height=500, res=150)
312IQRs <- esApply(leukemia_eset, 1, IQR)
313plot.ecdf(IQRs, pch=".", xlab="Interquartile range (IQR)", main="Leukemia data")
314abline(v=quantile(IQRs, prob=0.5), lwd=2, col="red")
315dev.off()
316@
317\begin{figure}[ht]
318\centerline{\includegraphics[width=0.5\textwidth]{GSVA-figIQR}}
319\caption{Empirical cumulative distribution of the interquartile range (IQR) of
320expression values in the leukemia data. The vertical red bar is located at the
32150\% quantile value of the cumulative distribution.}
322\label{figIQR}
323\end{figure}
```

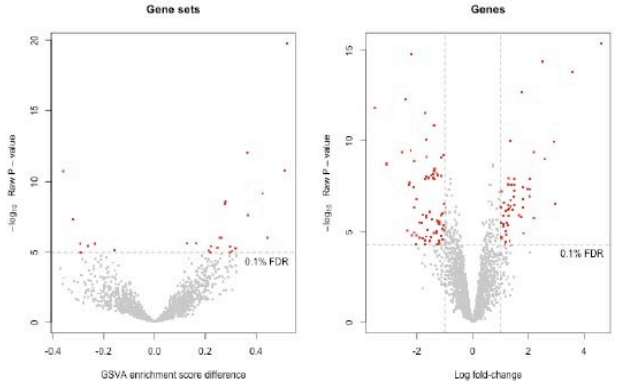


Figure 2: Volcano plots for differential pathway activation (left) and differential gene expression (right) in the leukemia data set.

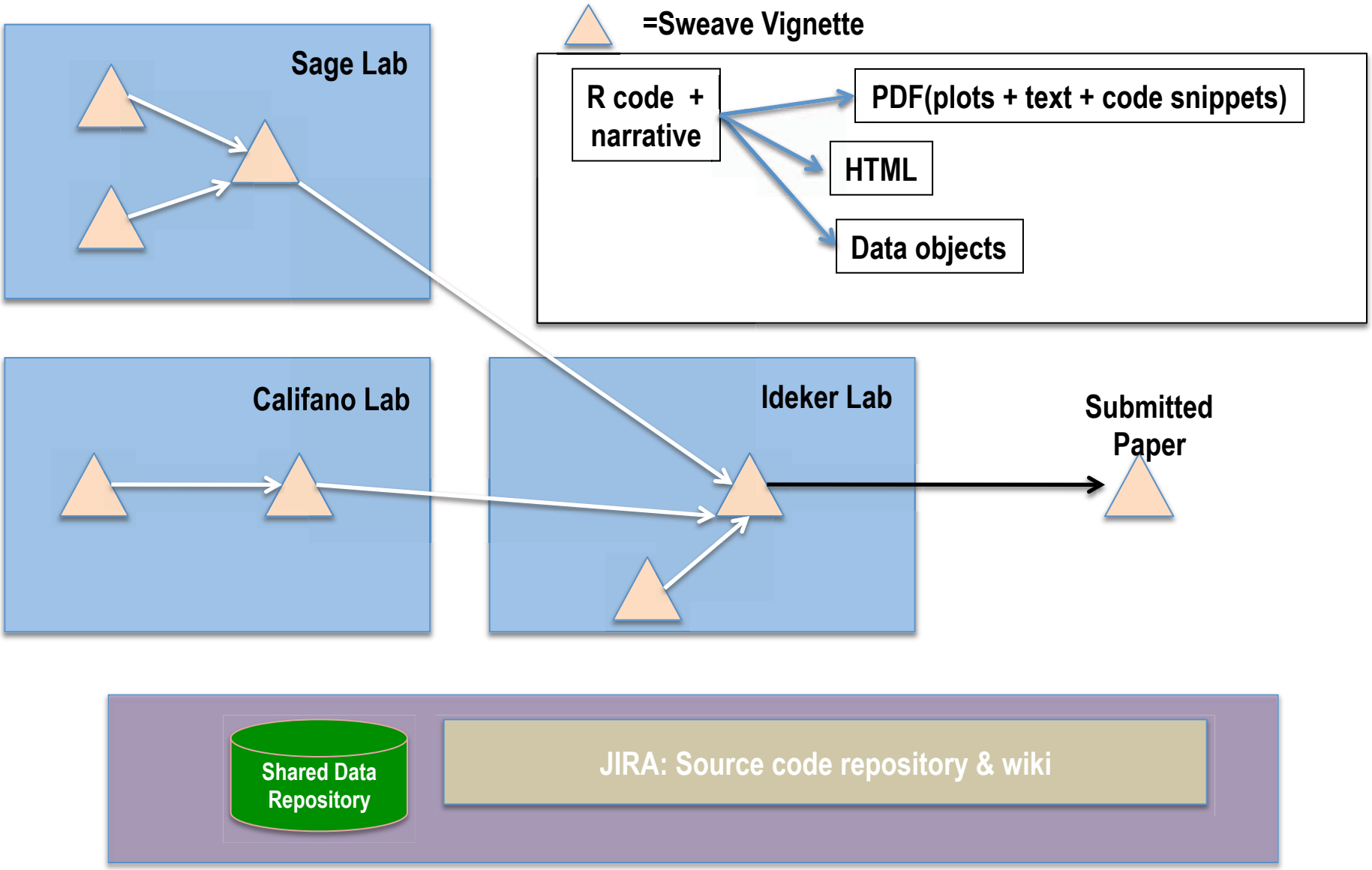
```
    ALL MLLvsALL
-1    2       7
0  2027    2006
1     4      20
```

Thus, there are 27 MSigDB C2 curated pathways that are differentially activated between MLL and ALL at 0.1% FDR. When we carry out the corresponding differential expression analysis at gene level:

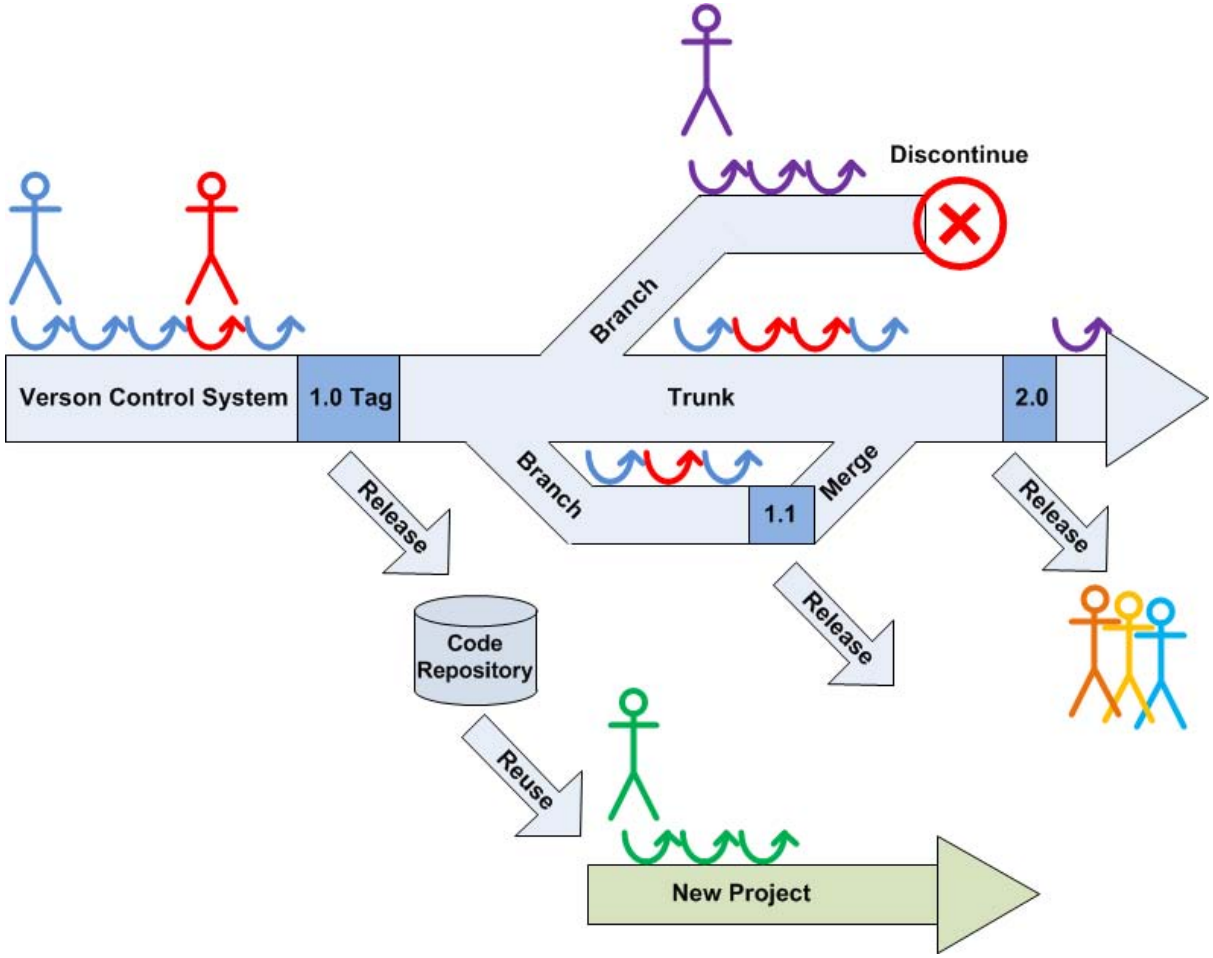
```
> logFCcutoff <- log2(2)
> design <- model.matrix(~factor(leukemia_eset$subtype))
> colnames(design) <- c("ALL", "MLLvsALL")
> fit <- lmFit(leukemia_filtered_eset, design)
> fit <- eBayes(fit)
> allGenes <- topTable(fit, coef = "MLLvsALL", number = Inf)
> DEgenes <- topTable(fit, coef = "MLLvsALL", number = Inf,
+     p.value = adjPvalueCutoff, adjust = "BH",
+     lfc = logFCcutoff)
```


Sweave.Friedrich Leisch. Sweave: Dynamic generation of statistical reports using literate data analysis. In Wolfgang Härdle and Bernd Rönz,editors, Compstat 2002 – Proceedings in Computational Statistics,pages 575-580. Physica Verlag, Heidelberg, 2002. ISBN 3-7908-1517-9

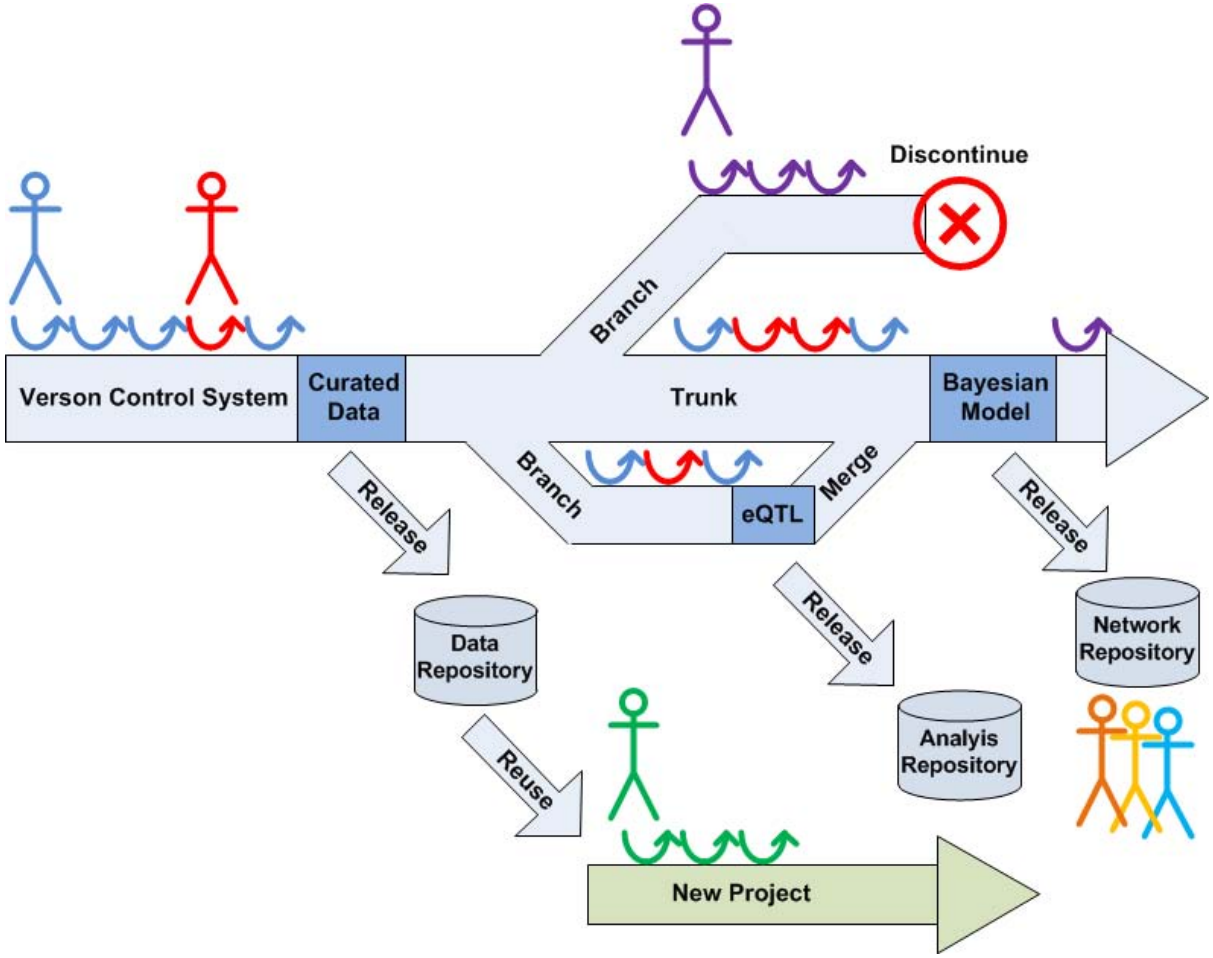
Federated Aging Project :
Combining analysis + narrative



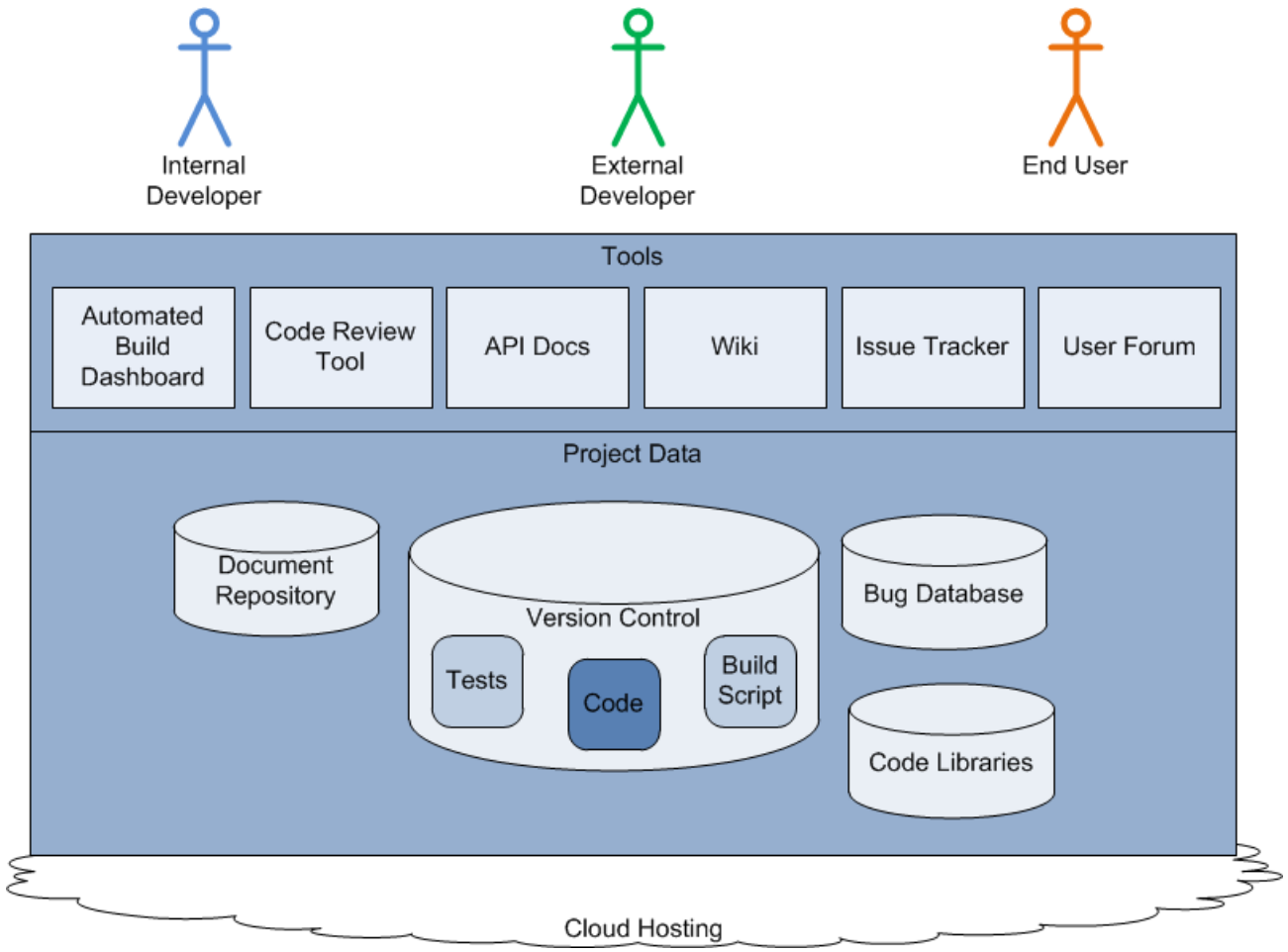
Evolution of a Software Project



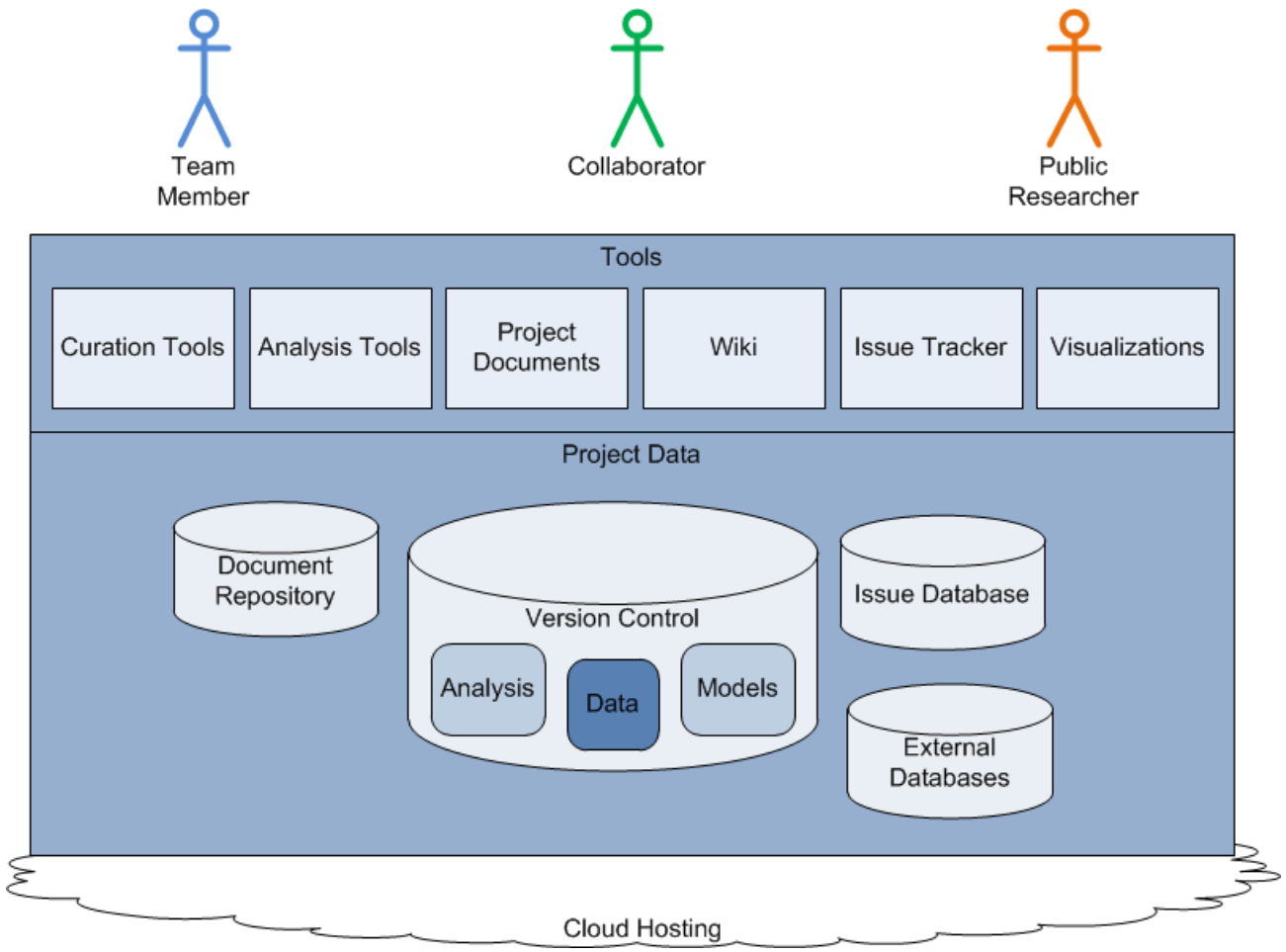
Evolution of a Biology Project



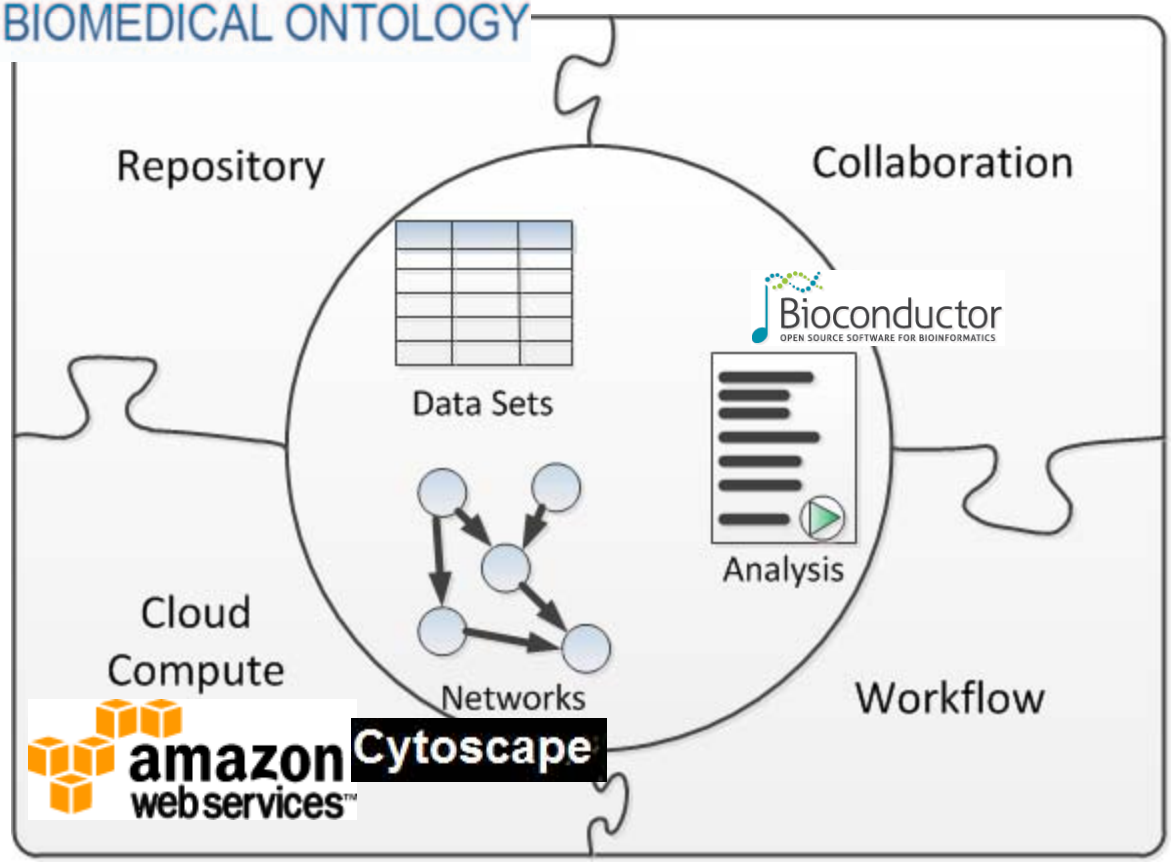
Software Tools Support Collaboration



Biology Tools Support Collaboration



Potential Supporting Technologies



i2b2

biomart

Microsoft Amalga

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

Microsoft Office

INSTITUTE FOR Systems Biology

Addama

Google Apps

NEXTBIO

ELIXIR

Google

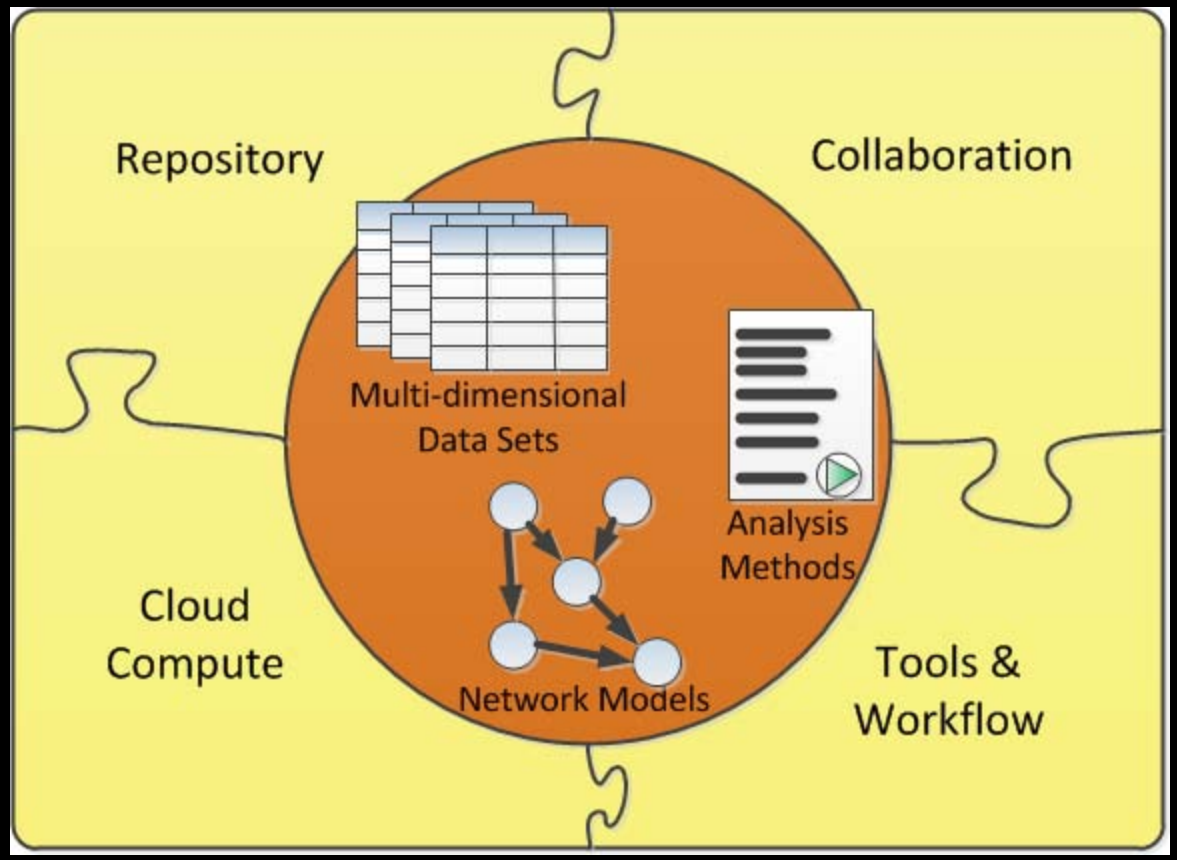
EMBL

Windows Azure

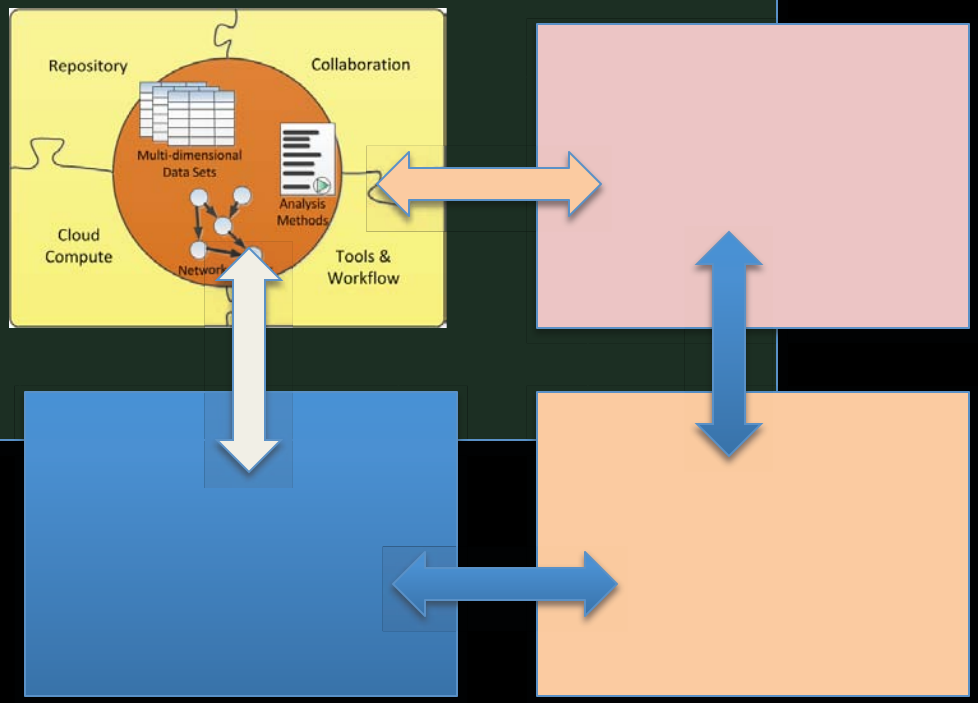
Taverna

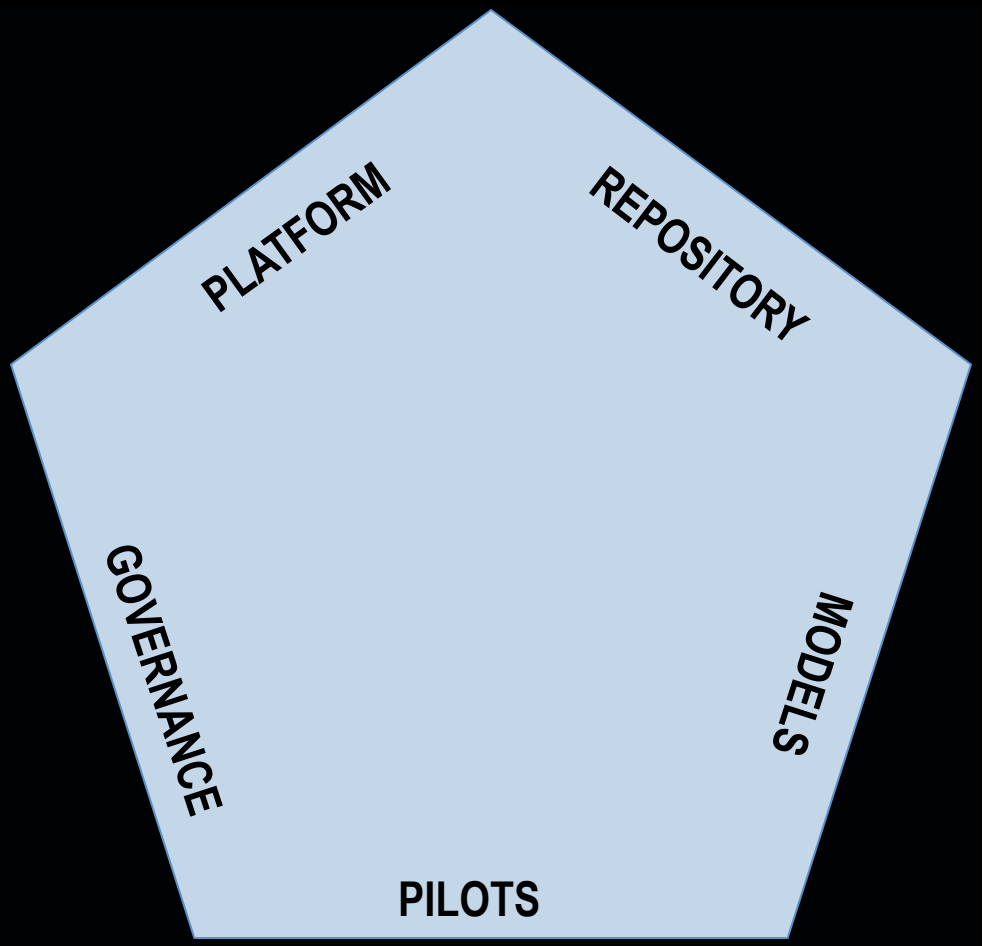
tranSMART

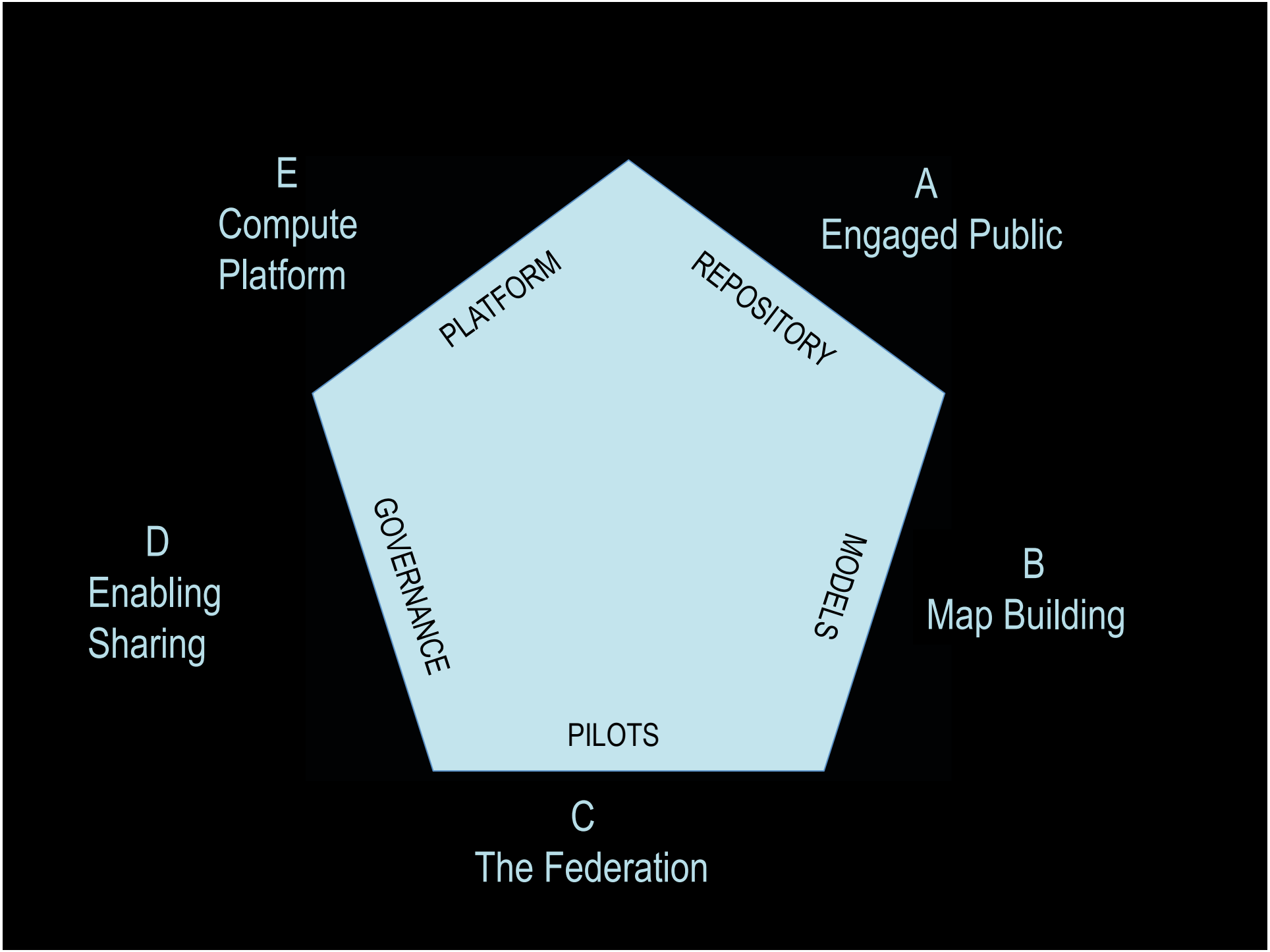
A Platform Node for Modelling



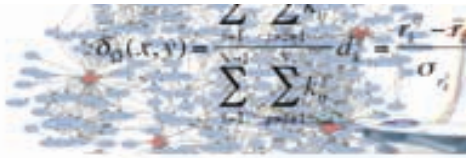
INTEROPERABILITY




PLATFORM
REPOSITORY
MODELS
PILOTS
GOVERNANCE


E
Compute
Platform
A
Engaged Public
PLATFORM
REPOSITORY
GOVERNANCE
MODELS
PILOTS
D
Enabling
Sharing
B
Map Building
C
The Federation

Sage Commons Congress



News 2010 Congress 2011 Commons Congress Announcement 2011 Congress Venue Prelimina

2011 CONGRESS VENUE

The 2011 Sage Bionetworks Commons Congress will be held at the Mission Bay Conference Center at UCSF an exciting new meeting and event destination centered in the University of California, San Francisco's new 43-acre life sciences campus for teaching and research. Located just south of downtown San Francisco, and convenient to Oakland and Berkeley, this stunning new building offers a world-class setting in the heart of Mission Bay.

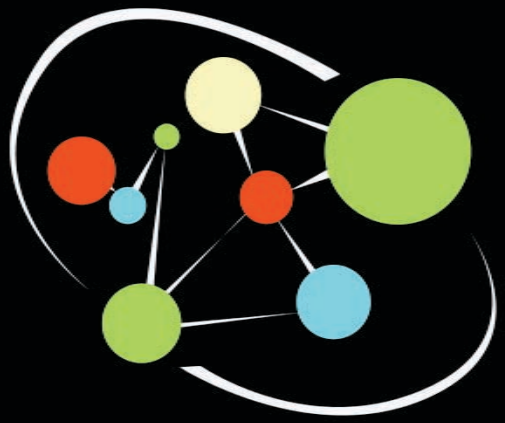
Accommodation information will be available shortly.

Mission Bay Conference

SAGE BIONETWORKS TWEETS

- RT @genomicslawyer: In case yo missed it, @dgmacarthur has ported Genetic Future to @wiredsciblogs http://bit.ly/dZ6vpt Absolute mus read . 14 hours ago
- Cancer Commons: a new molecular model-based, patient-centric paradigm for translational medicine http://bit.ly/fefPVH . 2011/01/18
- RT @GeneticAlliance: Should we all contribute DNA for research? Check out Biobank Bulletin http://ow.ly/3Bbe9 and tweet us your thoughts! . 2011/01/13
- Biomarker Blues; G. Poste on "dismal patchwork of fragmented research on disease-associated

http://sagecongress.org



Open Network Biology

Arch2POCM

COMMENTARY

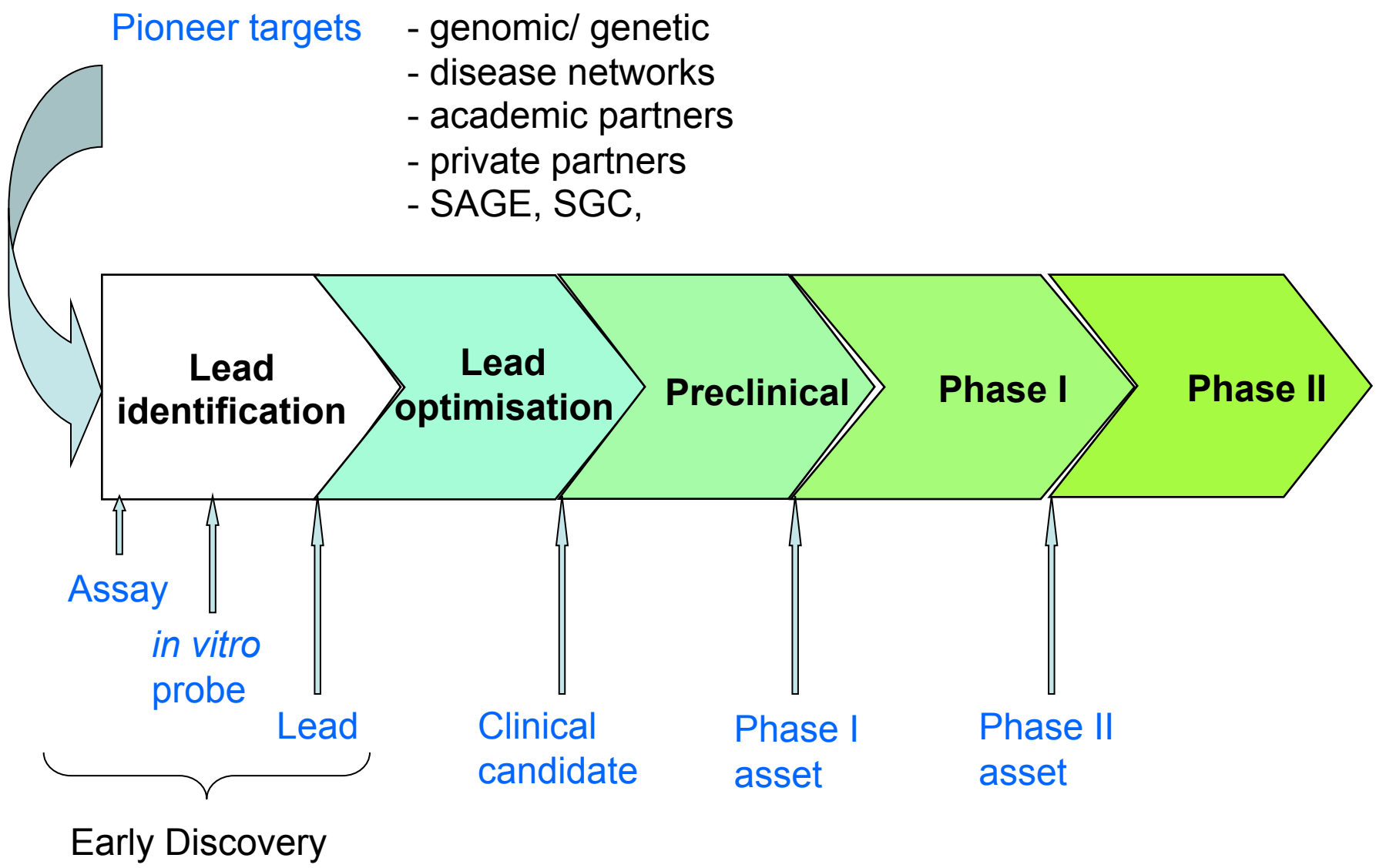
THERAPEUTICS DISCOVERY

The Precompetitive Space: Time to Move the Yardsticks

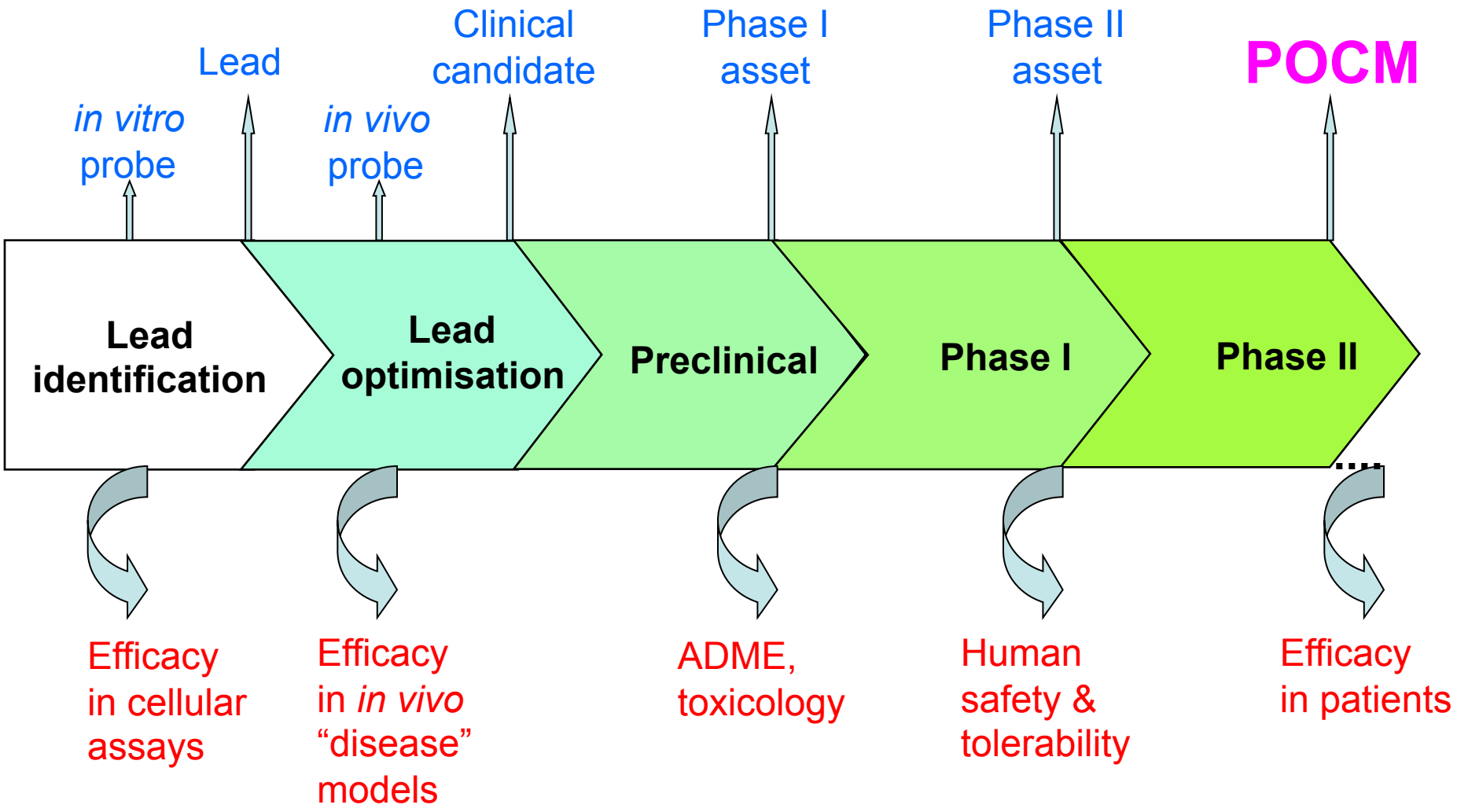
Thea Norman,[1] Aled Edwards,[2] Chas Bountra,[3] Stephen Friend[4]*

Industry, government, patient advocacy groups, public funders, and academic thought leaders met in Toronto, Canada, to set into motion an initiative that addresses some of the scientific and organizational challenges of modern therapeutics discovery. What emerged from the meeting was a public-private partnership that seeks to establish proof of clinical mechanism (POCM) for selected "pioneer" disease targets using lead compounds—all accomplished in the precompetitive space. The group will reconvene in April 2011 to create a business plan that specifies the generation of two positive POCM results per year.

Entry points



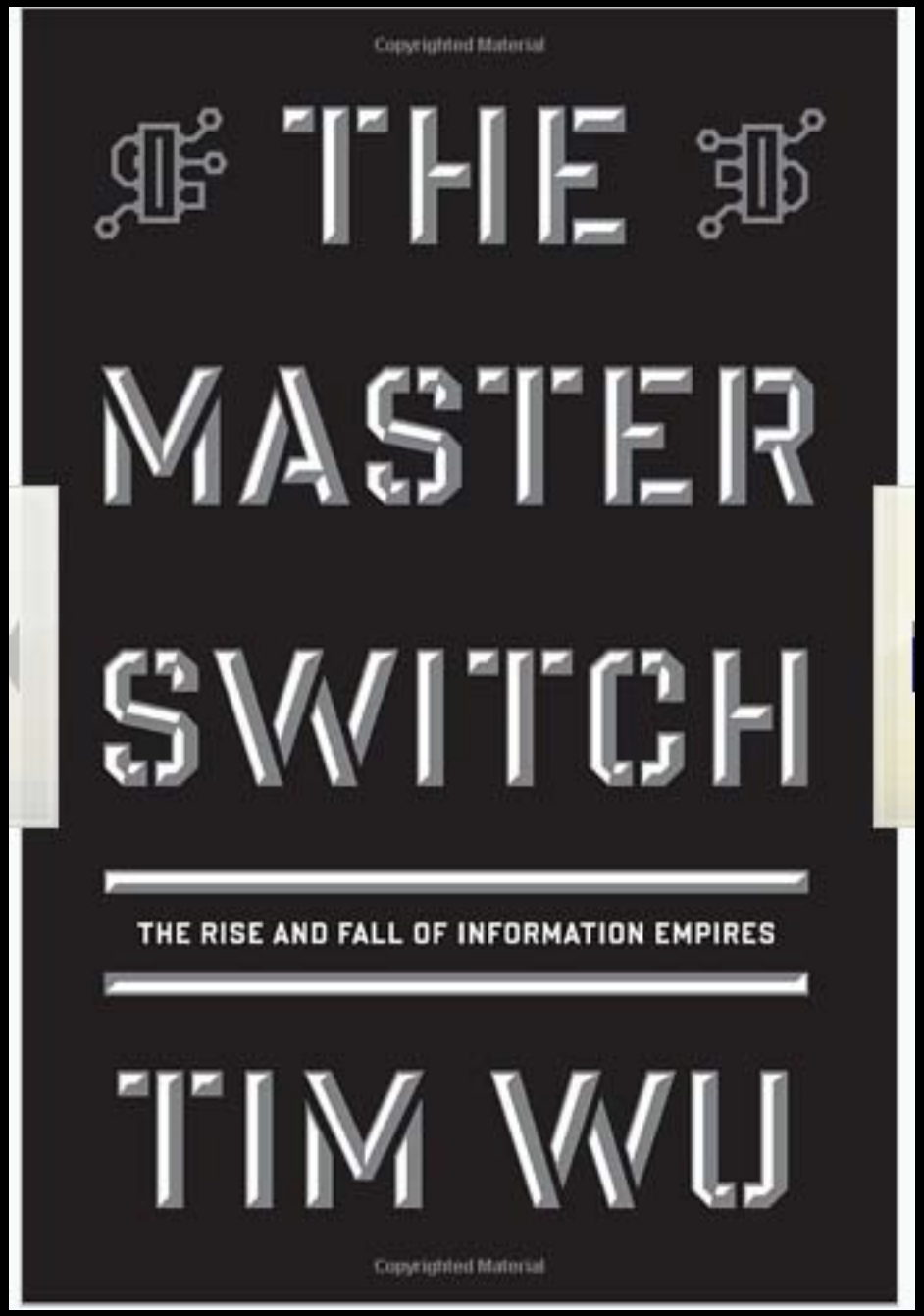
Reagents and publications will facilitate collaboration, more leveraged funds, improved disease maps and target discovery



BETTER MAPS OF DISEASE USING
DATA INTENSIVE SCIENCE

NOT JUST WHAT WE DO BUT HOW WE DO IT

POWER OF BUILDING A PRE-COMPETITIVE
COMMONS FOR EVOLVING
GENERATIVE MODELS OF DISEASE
USING A PUBLIC PRIVATE PARTNERSHIP


Copyrighted Material
THE
MASTER
SWITCH
THE RISE AND FALL OF INFORMATION EMPIRES
TIM WU
Copyrighted Material

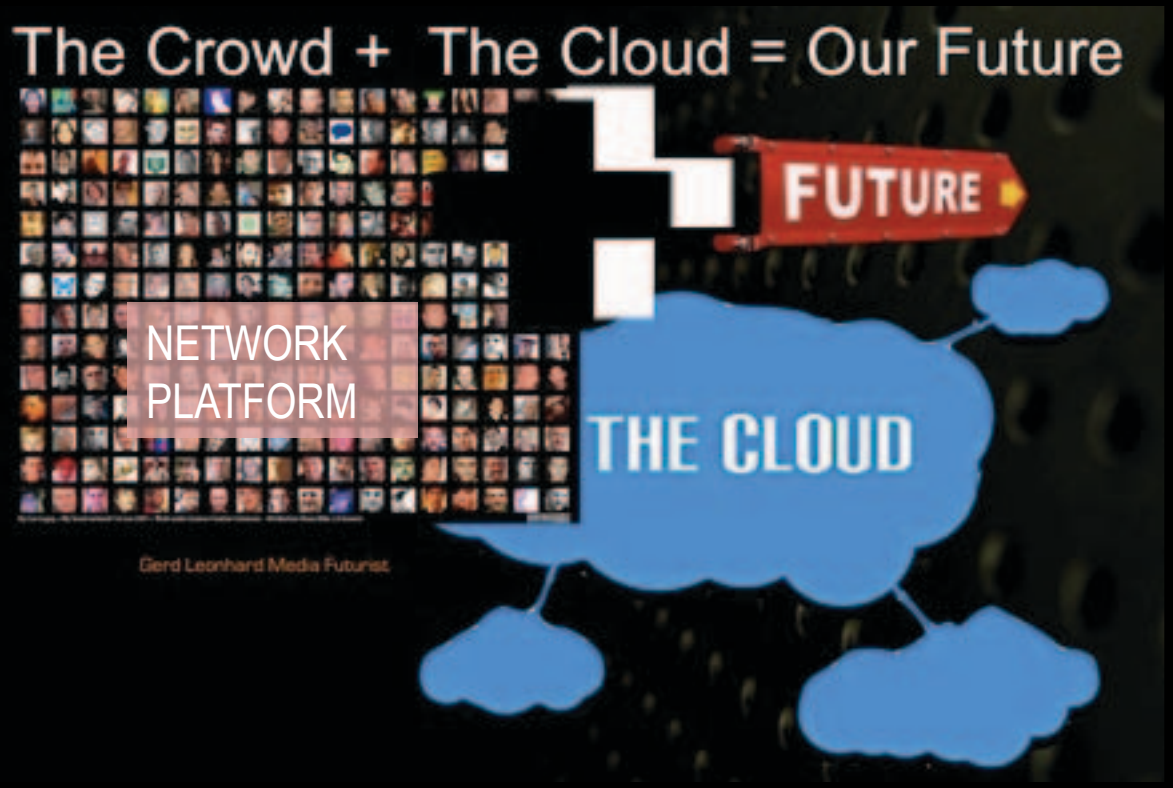
Who will build the datasets/ models capable of providing powerful safety and efficacy insights?

Academia

PIs

Post Docs

Grad Students



PHARMA

Therapy Areas

Scientists

Patients Physicians Citizens "Knowledge Expert"