Computational Medicine and Bioinformatics at the University of Michigan

Gilbert S. Omenn, MD, PhD

Director, Center for Computational Medicine and Bioinformatics

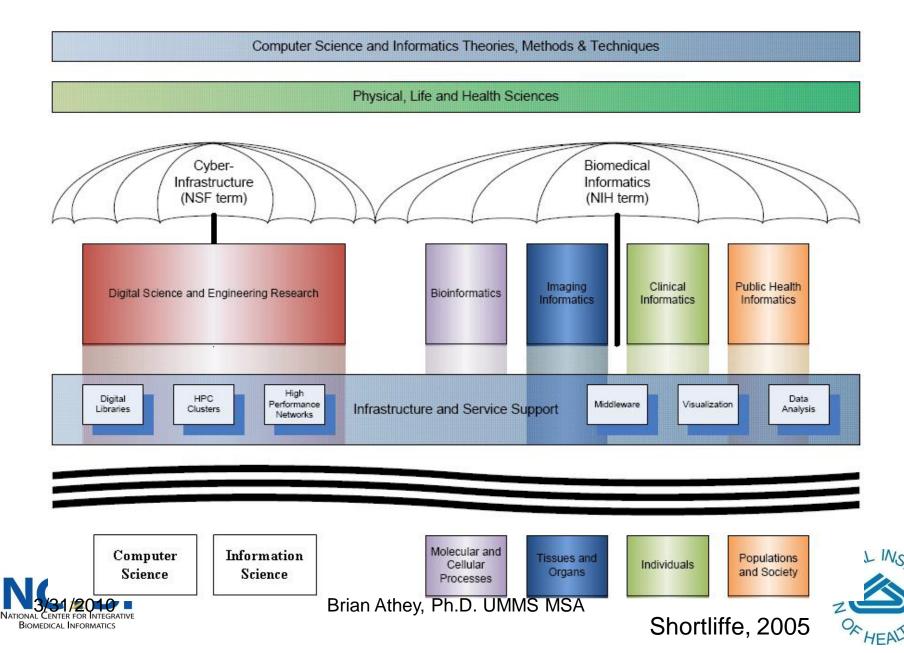
Components: Bioinformatics Graduate Program Collaborative Computing & Data Unit Interdisciplinary Research, led by NCIBI

Peking Union Medical College 23 March, 2010





Scope of Biomedical Informatics and Cyberinfrastructure



NCIBI: National Center for Integrative Biomedical Informatics An NIH NCBC Resource to Enhance Basic and Translational Research

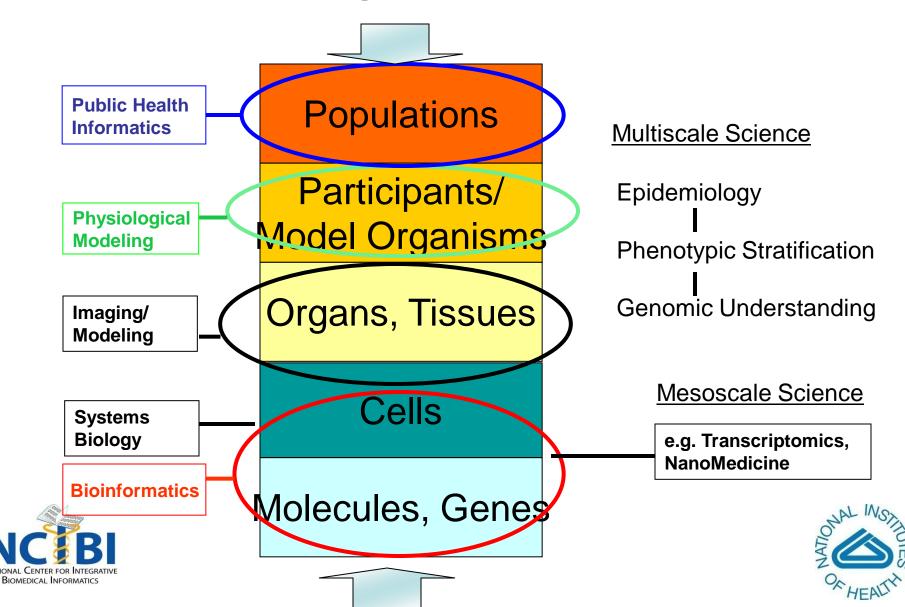
http://NCIBI.org, Brian Athey, PI

Center for Computational Medicine and Biology University of Michigan

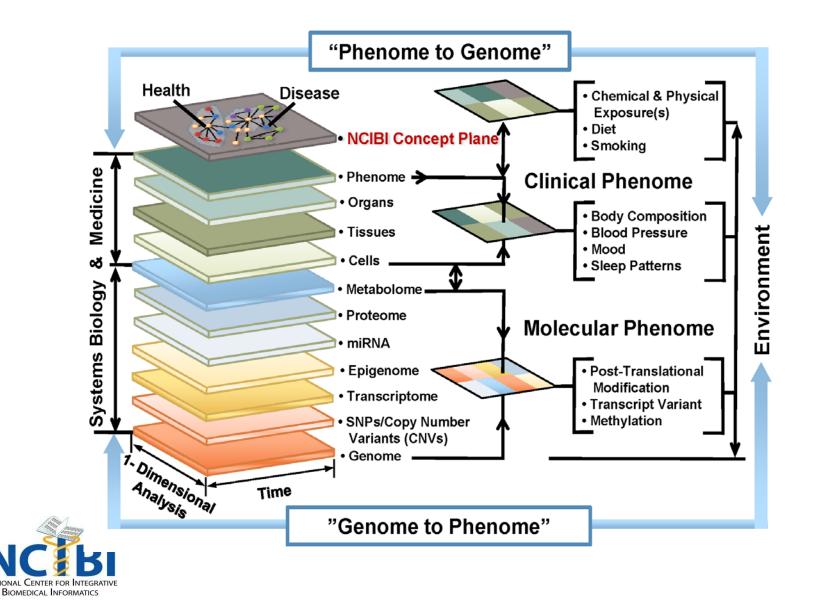
NIGMS/NIDA U54-DA-021519



Integrative Informatics Enables Synthesis of Knowledge at Multiple Levels

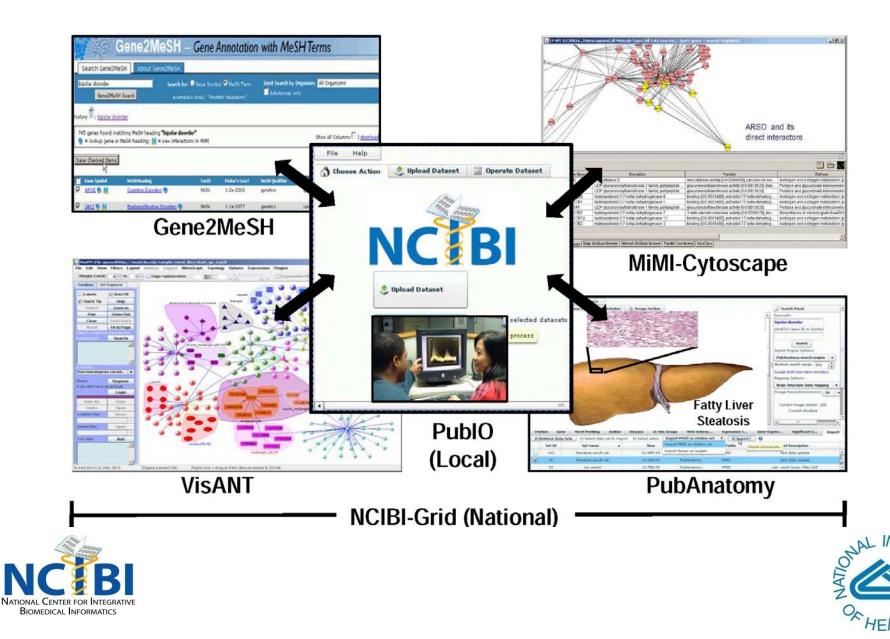


Integrating High-Throughput Measurements with the Phenotype is Key

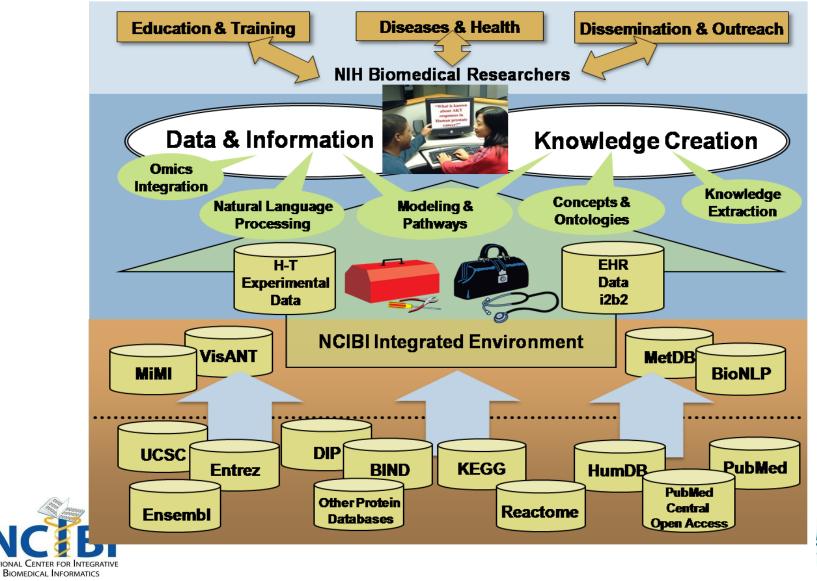




Integrated Tools that Track the Levels

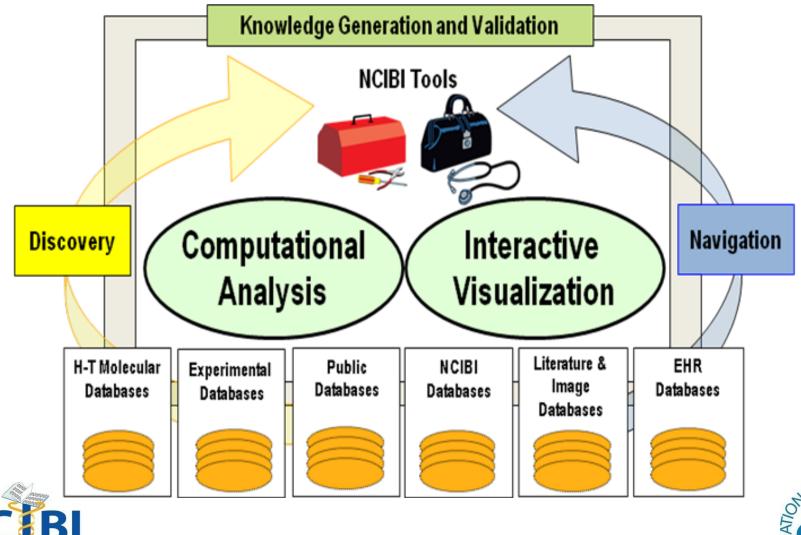


Overview of the NIH National Center for Integrative Biomedical Informatics (NCIBI)





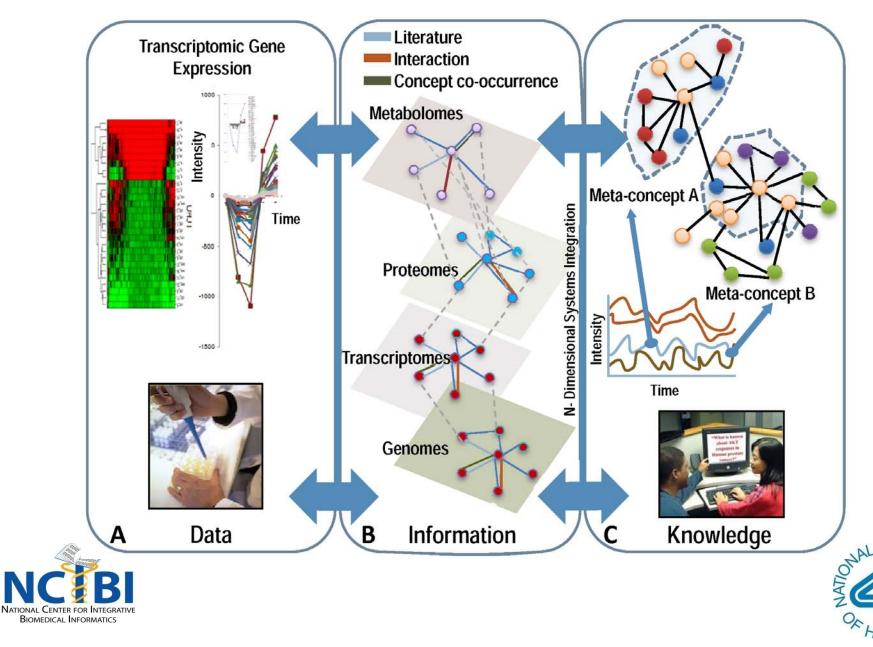
NCIBI Computational Science Core Abstract View



BIOMEDICAL INFORMATICS



Theme: From Data to Knowledge



NCIBI Senior Leadership Team



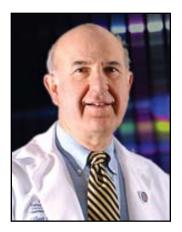
H.V. "Jag" Jagadish, Ph.D. Senior Scientific Director Core 1 Director



James D. Cavalcoli, Ph.D.



Brian D. Athey, Ph.D. PI and Core 2, Core 7 Chair Senior Scientific Director



Gilbert S. Omenn M.D., Ph.D. Senior Scientific Director Core 3 Director



Barbara Mirel, PhD. Cores 5 & 6 Co-Director

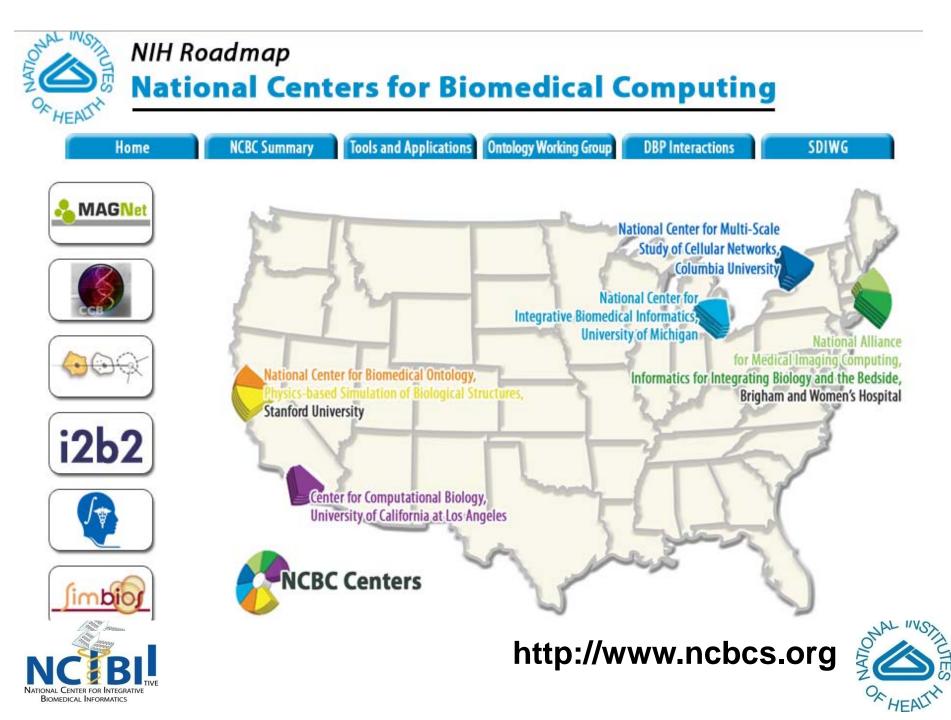


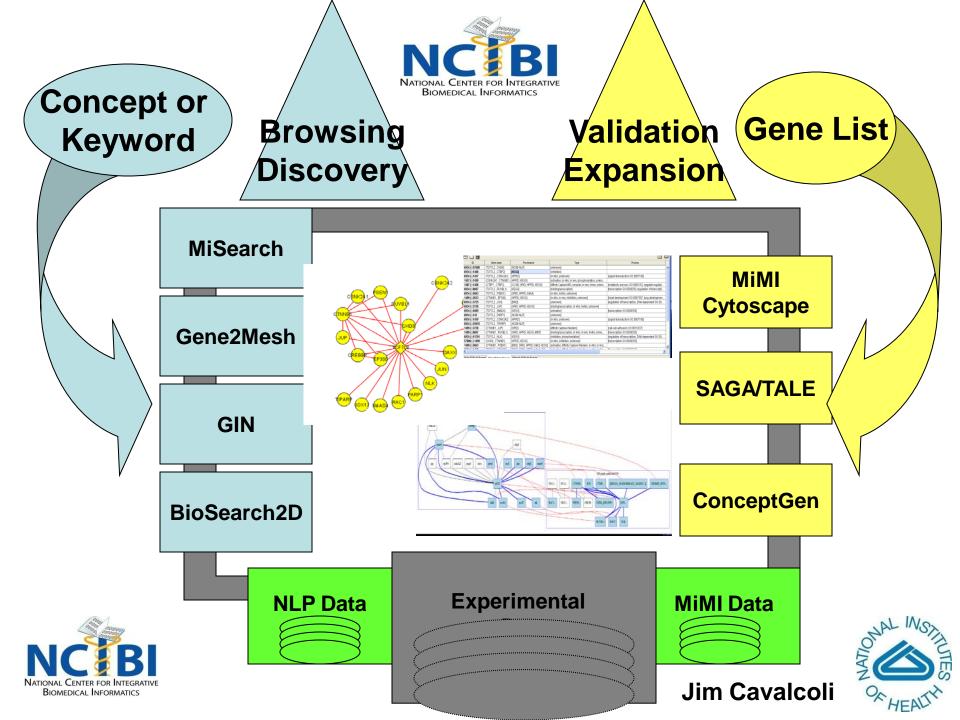
Matthias Kretzler, M.D. Core 3, Diabetes complications



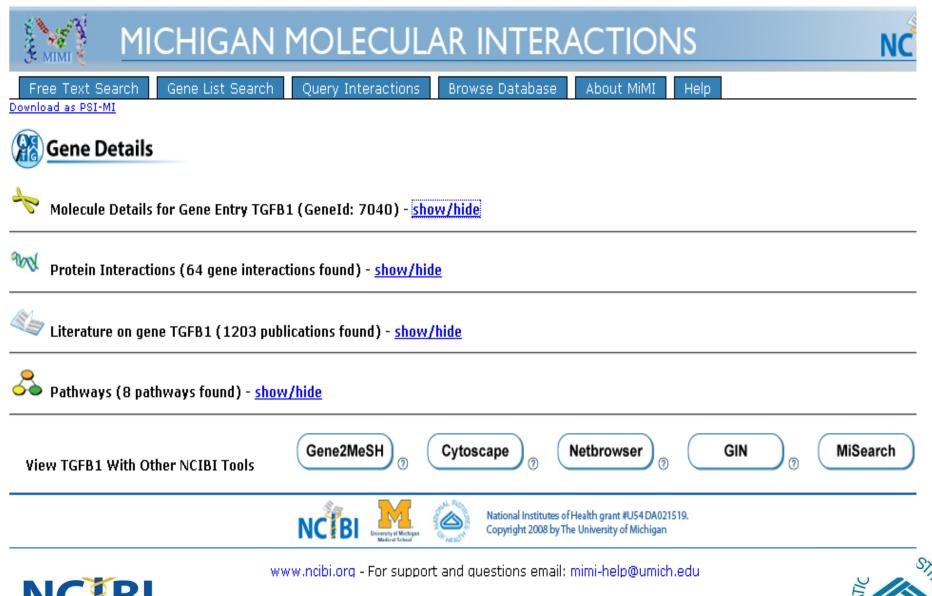
NCIBI Collaborative National Partners





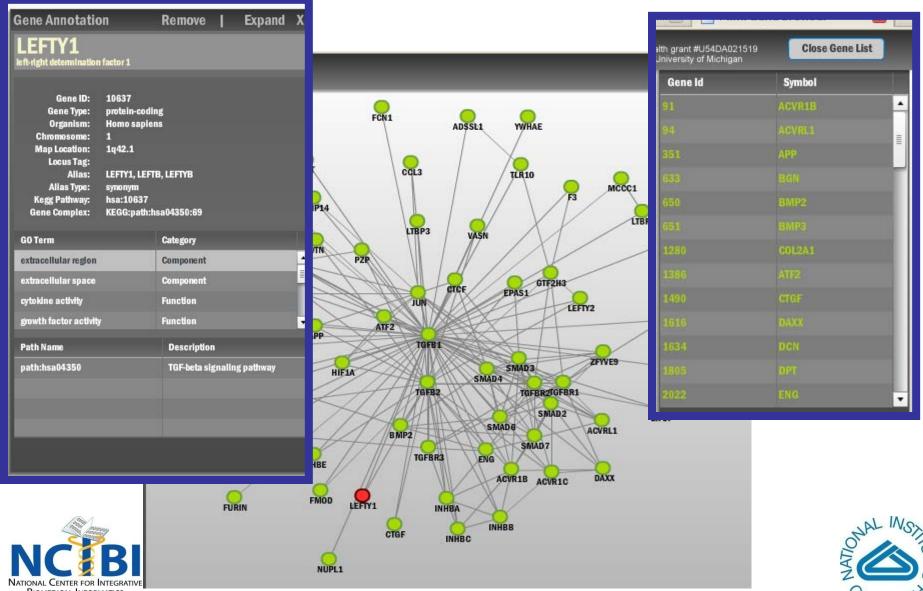


http://mimi.ncibi.org





NCIBI ConceptGen NetBrowser



BIOMEDICAL INFORMATICS

Tian and Patel

SAGA and TALE: Effective and Efficient Graph Matching

- Graph databases are large and growing rapidly in size. e.g. KEGG has grown by ~1000x since 1999.
- Data is noisy or incomplete (or both)
- Need "approximate" graph matching
 - SAGA: when the query graph is small (< 100 nodes)
 - TALE: when the query graph is large (1000s of nodes)
- Shown to be:
 - More effective able to find results other tools miss
 - Orders of magnitude faster than existing tools
- Publications:

BIOMEDICAL INFORMATICS

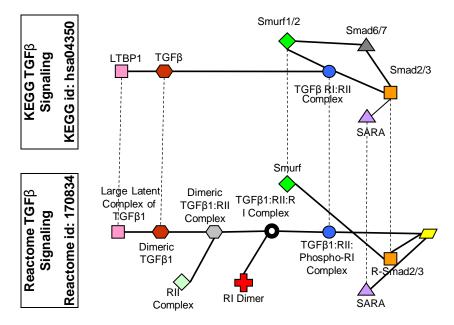
• SAGA: Tian et al. Bioinformatics'07

TALE: Tian and Patel, IEEE Data Engineering'08



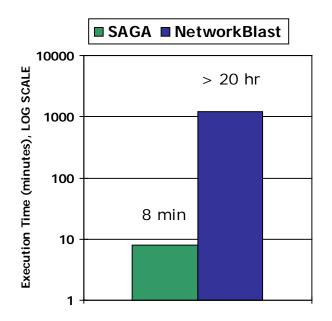
SAGA Example: Integrating Different Pathways Databases

KEGG Database vs. Reactome Database



Potential: Integrate graphs from different sources.

BIOMEDICAL INFORMATICS



Orders of magnitude faster than existing tools



http://saga.ncibi.org

Match No.	Match Graph Name (#Nodes, #Edges)	Graph Dis
Match #1	path:hsa04330 (17,16) [Notch signaling pathway]	29.00
<u>Match #2</u>	path:hsa04310 (59,70) [Wnt signaling pathway]	36.00

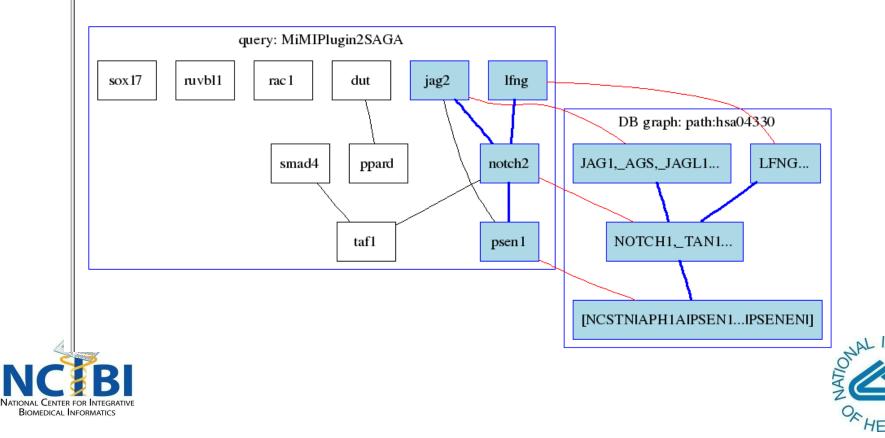
Details of the Matches:

[Go Back to Matches Overview]

Match #1: path:hsa04330 (17 nodes, 16 edges) [Notch signaling pathway]

Graph Distance 29.00 (4 out of 11 nodes match)

[Link to KEGG Picture] (with the matching nodes highlighted)



Gene2Mesh: Automated Literature-Based Genome Annotation Using MeSH

- GO
 - Linked to genes and genomes
 - Only describes normal physiology
- OMIM
 - Heritable disorders
 - Hard to search
- MeSH
 - Broad coverage of biomedical concepts
- Need links to genes

- Statistical association
 - All of PubMed
 - All MeSH terms
 - All papers referring to genes
- Fisher's exact test
 - Significant associations
 - Informative ranking
- http://gene2mesh.ncibi .org



Mesh and the OMIM MorbidMap

Genes associated with "prostatic neoplasms"

<u>Symbol</u>	Description		
AR	androgen receptor		
RNASEL	ribonuclease L		
ELAC2	elaC homolog 2 (E. coli)		
PTEN	phosphatase and tensin homolog		
CD82	CD82 molecule		
PCA3	prostate cancer antigen 3		
MSR1	macrophage scavenger receptor 1		
KLF6	Kruppel-like factor 6		
TMEM16G	transmembrane protein 16G		

<u>OMIM</u>

Prostate Cancer

Prostate Cancer, Hpc1 ;Prca1

Prostate Cancer

Prostate Cancer

- Prostate Cancer
- Prostate Cancer Antigen 3; Pca3

Prostate Cancer

Prostate Cancer

Prostate Cancer-Associated Protein 5;

74 Additional prostate cancer associated genes identified using MeSH analysis KLK3 (PSA), NKX3, FOLH1 (PSMA1), AMACR, ACPP, SRD5A2, TMPRSS2...

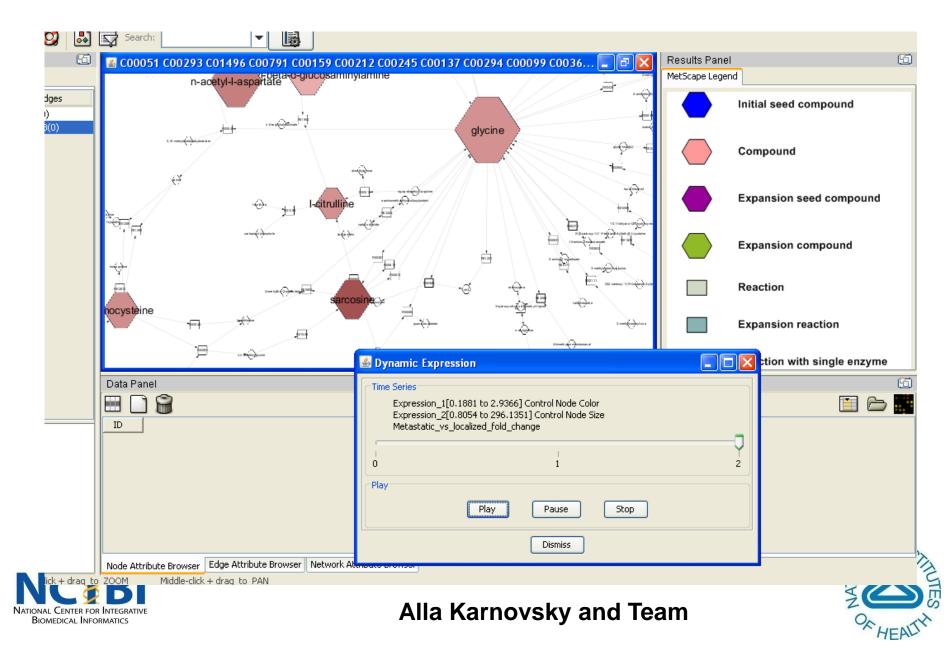
11 Genes in OMIM, not found by MeSH analysys BRCA2 and CHEK2 missed threshold, 9 genes with minimal literature or not enriched

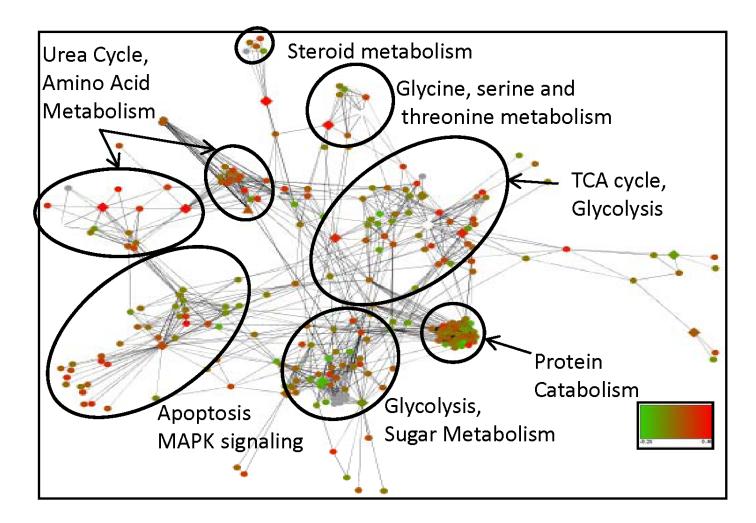


http://gene2mesh.ncibi.org



New NCIBI Tool--Metscape





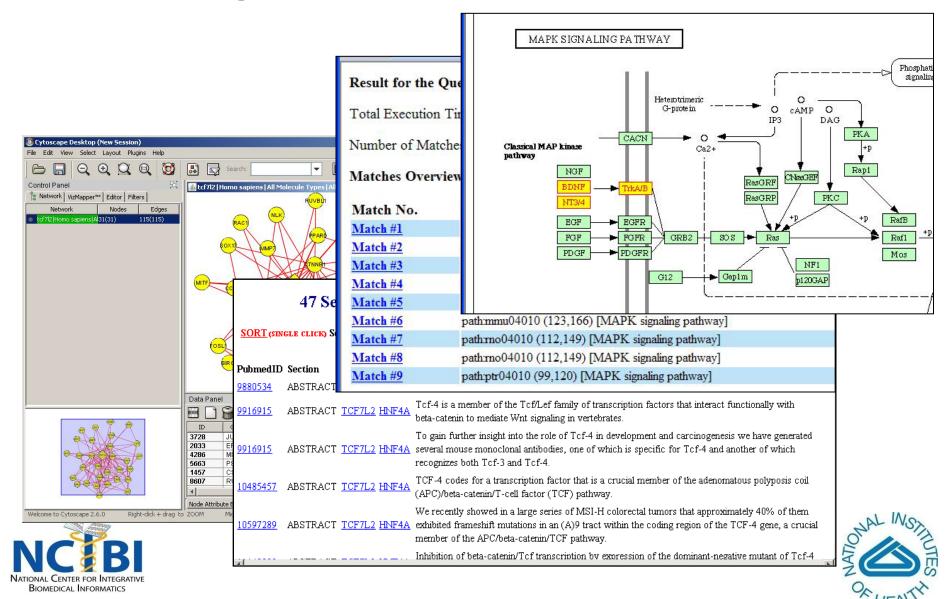
MetScape visualization of liver mRNA levels (nodes) whose expression pattern correlates positively (red) or negatively (green) with liver glutamine levels

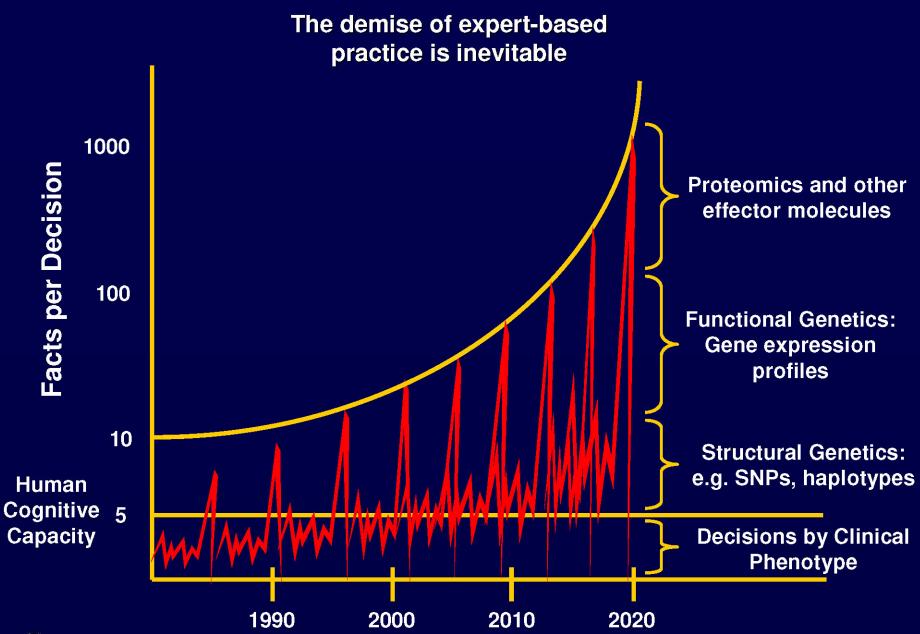


d Attie

Burant, Karnovsky, and Attie

Visualization tools: MiMI Plugin for Cytoscape Integrated with BioNLP and SAGA







NCIBI Driving Biological Problems (DBPs) Y1-Y5

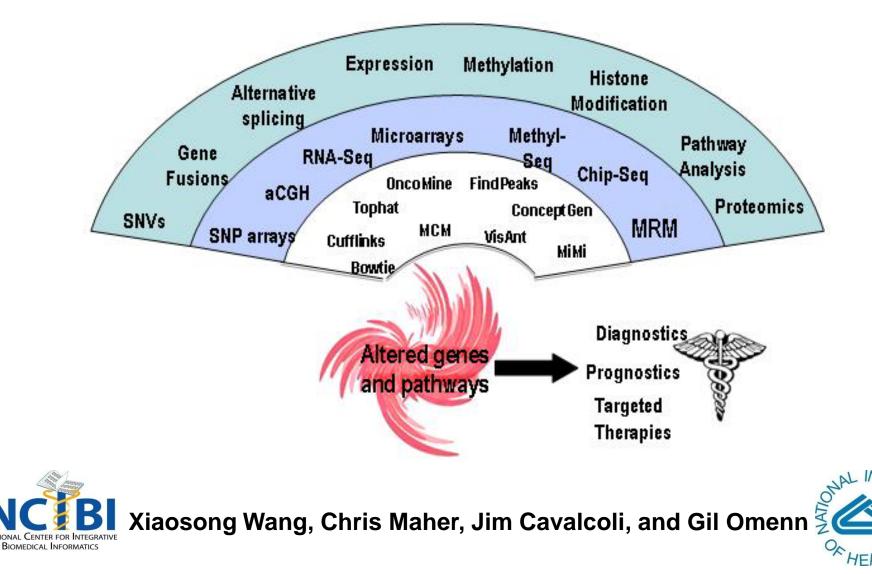
Year 1	Year 2	Year 3	Year 4	Year 5	
Prostate Cancer Progression			Gene Fusions In Cancers		
Type 1 Diabetes Complications Neuropathy		ons	Type 1 Diabetes Complications Nephropathy		
Type 2 Diabetes Genetic Heterogeneity		geneity	Type 2 Diabetes Metabolomics		
Bipolar Susceptibility Genetics		tics	Bipolar Disease Co-Morbidities With Drug Abuse		

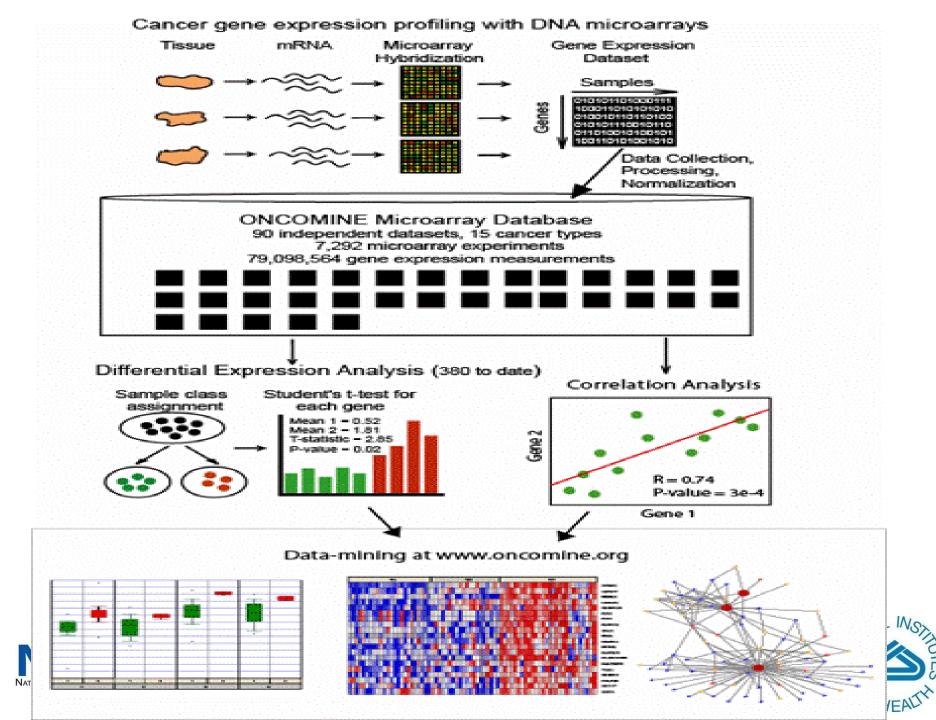
Gil Omenn MD, PhD, Senior Scientific Director



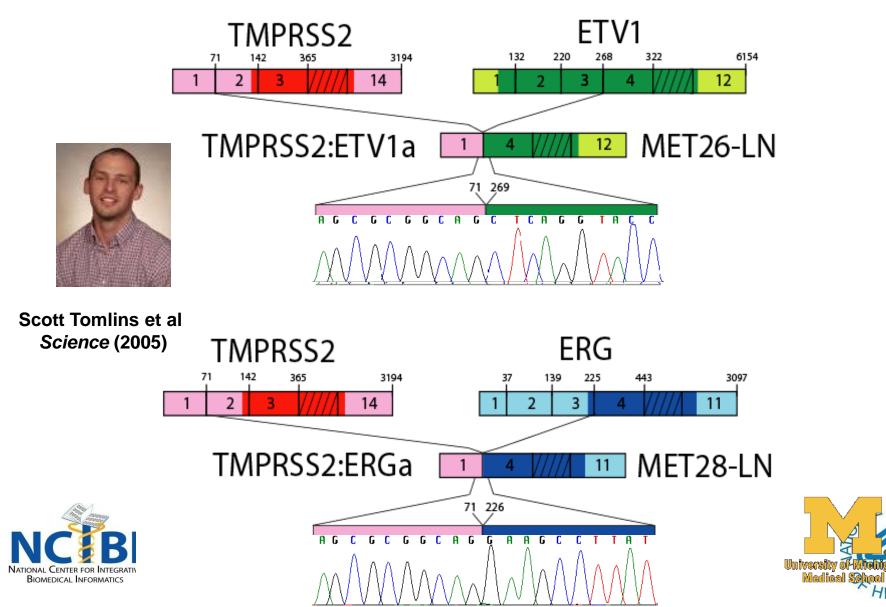


Prostate Cancer Progression DBP Integration Strategy

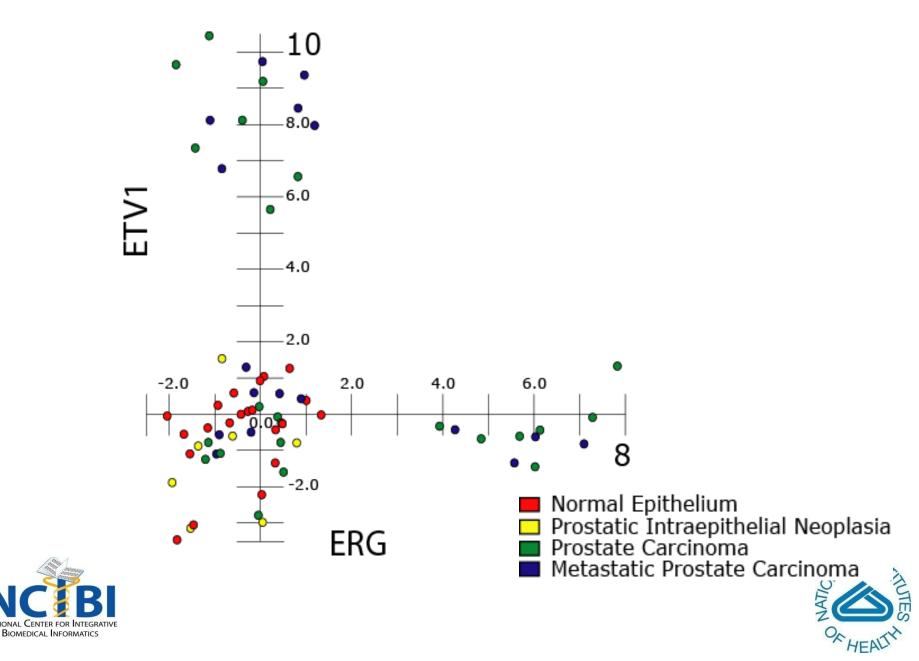




Fusions of TMPRSS2 to the ETS Family of Transcription Factors

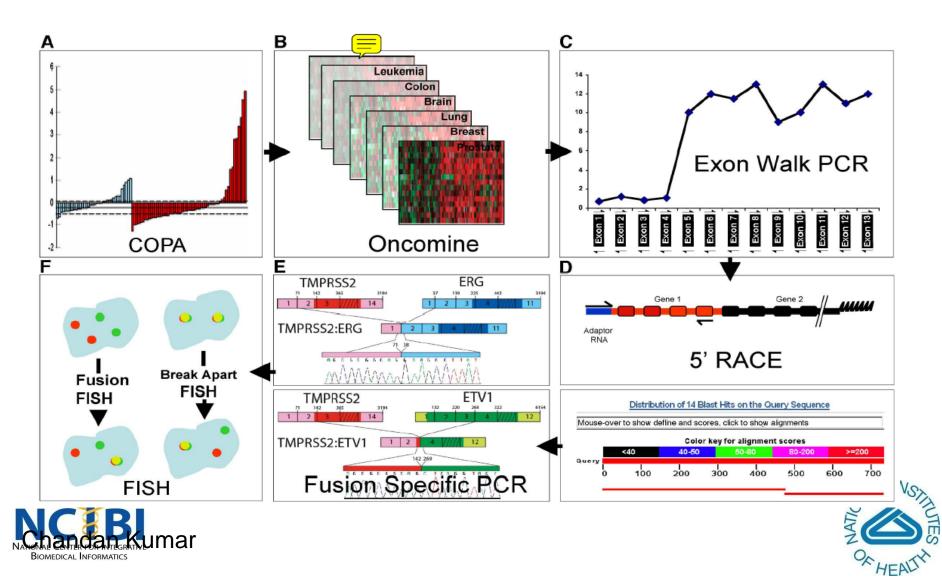


Exclusive Outlier Expression of ETV1 and ERG

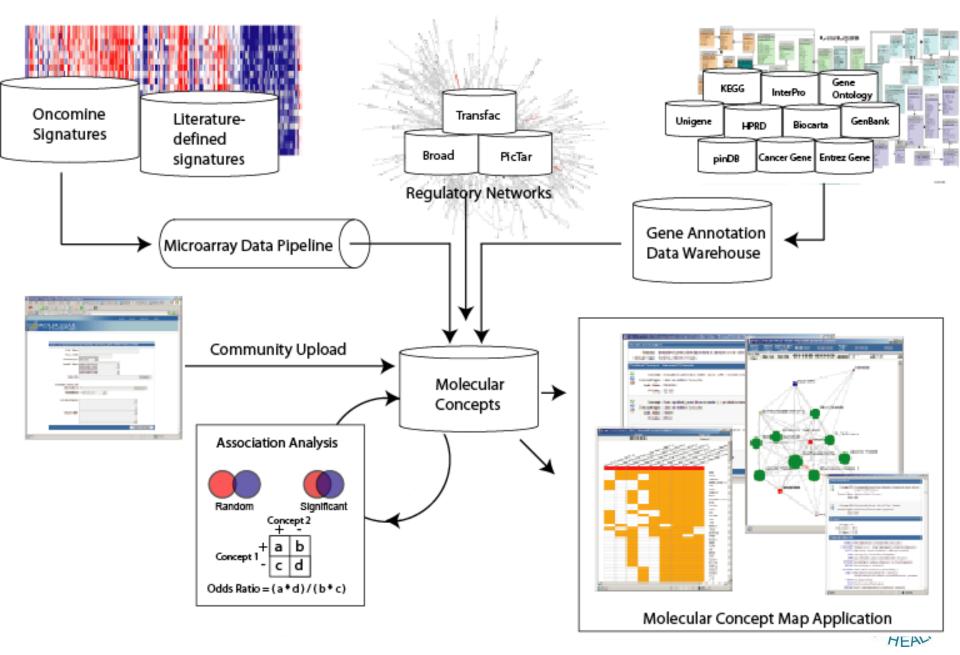


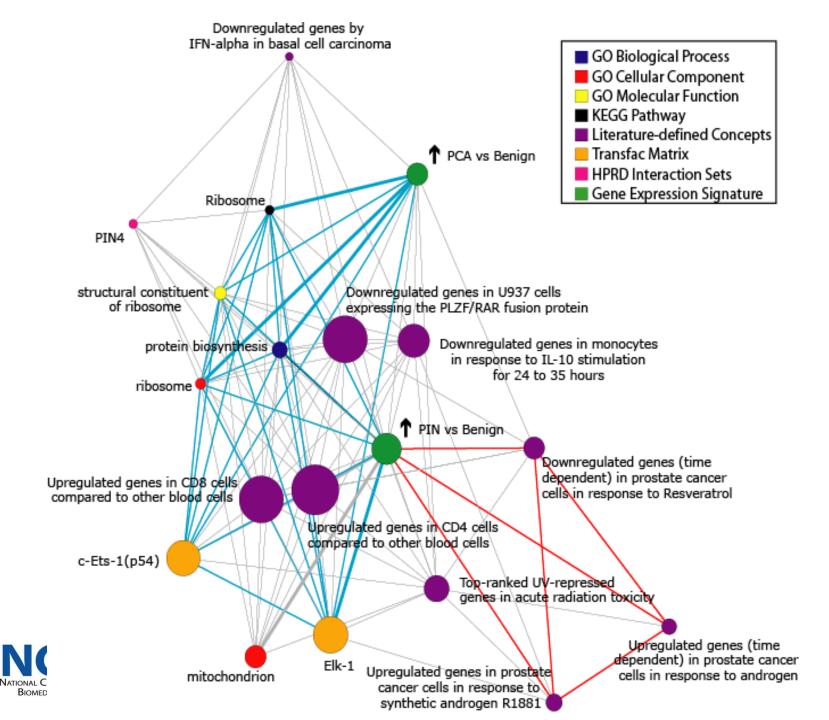


A Bioinformatics Approach Leads to Discovery of Gene Fusions in Prostate Cancer



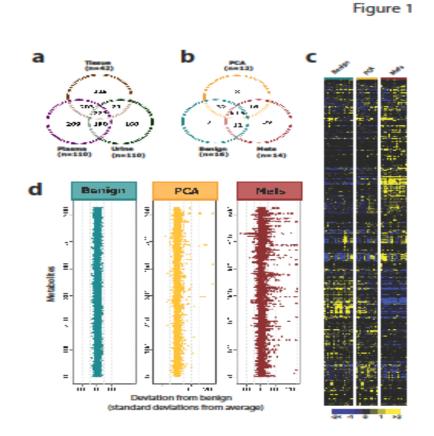
The Molecular Concept Map Project







Metabolomic and Bioinformatic Analysis of Prostate Cancers Sreekumar et al, Nature 2009







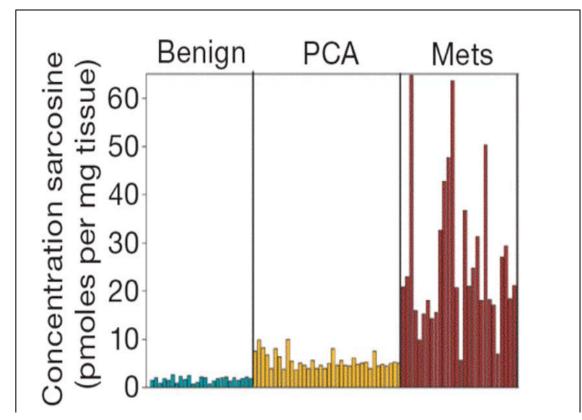


Figure Sarcosine concentration is greatly increased in metastatic prostate cancers, compared with localized tumors or benign tissue.





Biomarker Discovery from Tumor Tissues and Plasma: Strategies

- 1. Start with microarray or next-gen sequencing evidence for carcinogenic pathway mechanisms in tumor and track corresponding protein biomarker candidates to the plasma; e.g., TMPRSS2/ETS fusion and sarcosine in prostate cancers.
- 2. Identify alternative splice isoforms of biologically meaningful proteins in cancers and in plasma of humans and mouse models.
- 3. Perform targeted proteomics with SRM/MRM to identify and quantitate these candidates.
- 4. Detect auto-antibodies in plasma as a biological amplification of tumor protein signals, confirm in the second second



A New Class of Biomarker Candidates, from Alternative Splicing

Generates protein diversity without increasing genome size

Most genes produce alternative transcript isoforms

Alternative Splice Isoforms

- Contribute to diseases, especially cancers
- Potentially useful as biomarkers for cancer

Greatly improved MS/MS instrumentation enables confident identification of peptides from proteins coded by mRNA transcript sequences expressed at quite low levels.





Summary of Total Number of Alternative Splice Variants Identified

	Breast		Colon		Pancreatic	
Distinct proteins	Known	Novel	Known	Novel	Known	Novel
	540	68	461	28	328	92



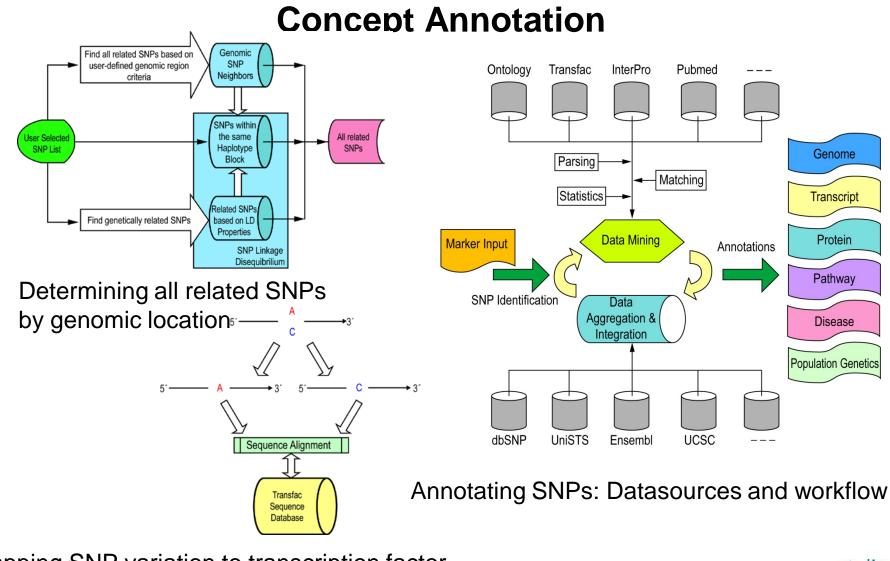


Type 2 Diabetes: Genetic and Phenotypic Heterogeneity

- Exploit ongoing FUSION study datasets
- Implement SNP analysis workflow for whole genome association study (GWAS)
 - For each SNP (and then haplotype), screen databases to identify: coding/noncoding, effect on protein, evolutionary conservation, splice variant, transcriptional regulation, microRNA binding, CpG island, DNAse hypersensitivity, disease associations (adding epigenomic marks)
 - -Devise algorithms with much higher throughput



NCIBI Innovations in SNP Analysis and Genome



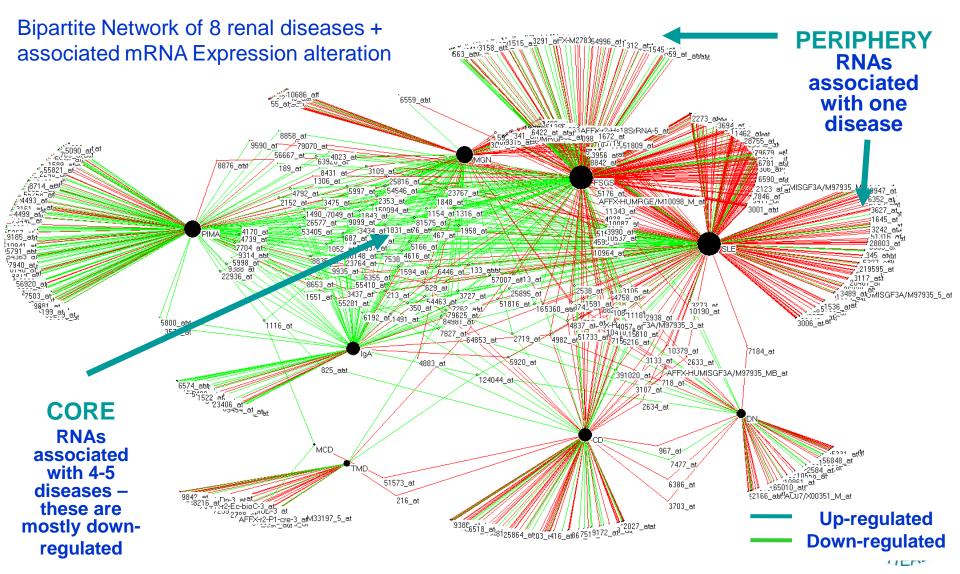
Mapping SNP variation to transcription factor

NATIONAL CENTER FOR INTEGRATIVE BIOMEDICAL INFORMATICS binding sites

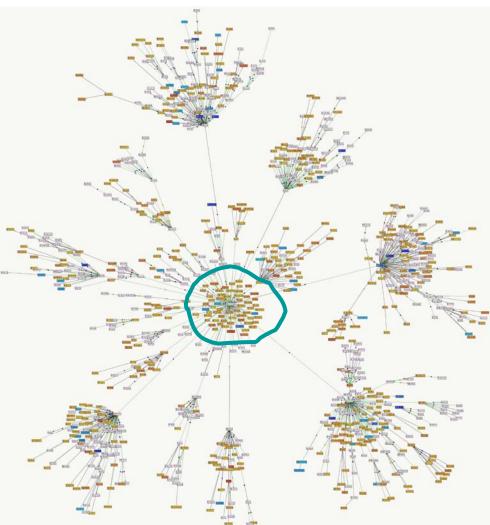
www.ncibi.org/resources



Global Molecular Network View of Kidney Diseases

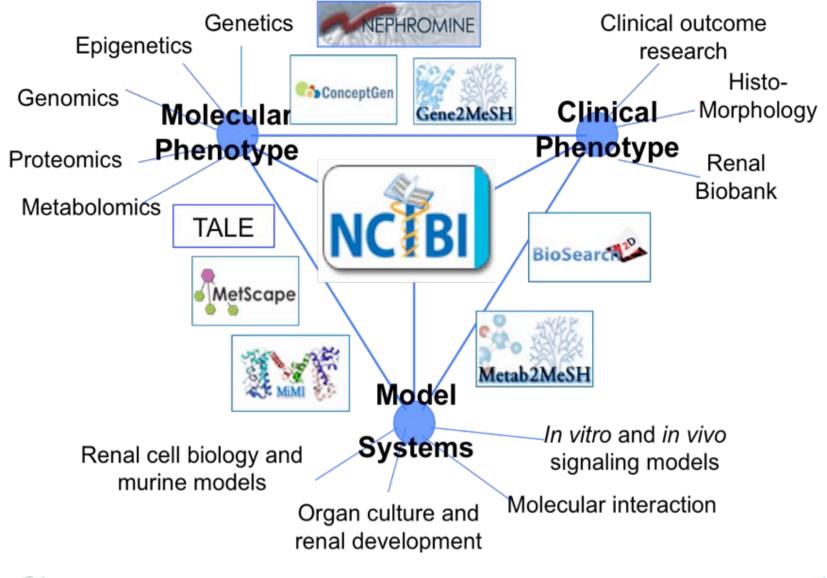


Defining Cross-species Conserved Networks of Early Diabetic Glomerulopathy



BIOMEDICAL INFORMATICS

- 24 early DN glomerular biopsies (NIDDK trial "Renoprotection in Early Diabetic Nephropathy in Pima Indians, OH95-DK-N037) versus 24 living donor transplants.
 - Nodes: differentially regulated mRNA
 - Blue nodes are repressed;
 - Orange nodes induced genes;
 - White: non-regulated transcription factors
 - Black edges: co-citation of mRNAs
 - Green edges: respective transcription factor binding sites in promoter regions of regulated mRNAs
 - Stat1 dependent genes are induced and form a key node in the network, consistent with JAK/STAT pathway activation in human DN.







New NCIBI NIDA Genetics Consortium Pilot Project

"Genetic Predisposition to Co-Morbidity of Bipolar Disorder and Substance Abuse in African-American Women"



Dr. Sharon Lewis Langston University





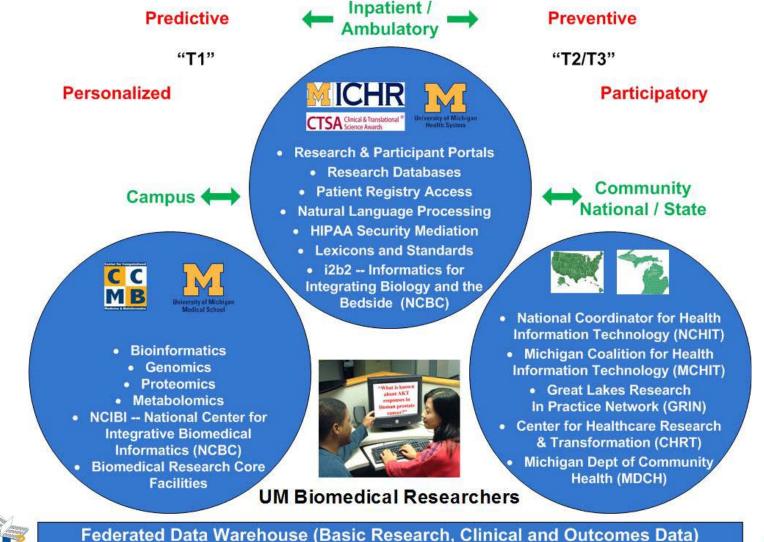
Dr. Tonya Gerald NC Central University



Dr. Raphael D. Isokpehi Dr. Wellington K. Ayensu Jackson State Jackson State University University



The Scope of Biomedical Informatics Striving Toward Desired Future P4 Medicine

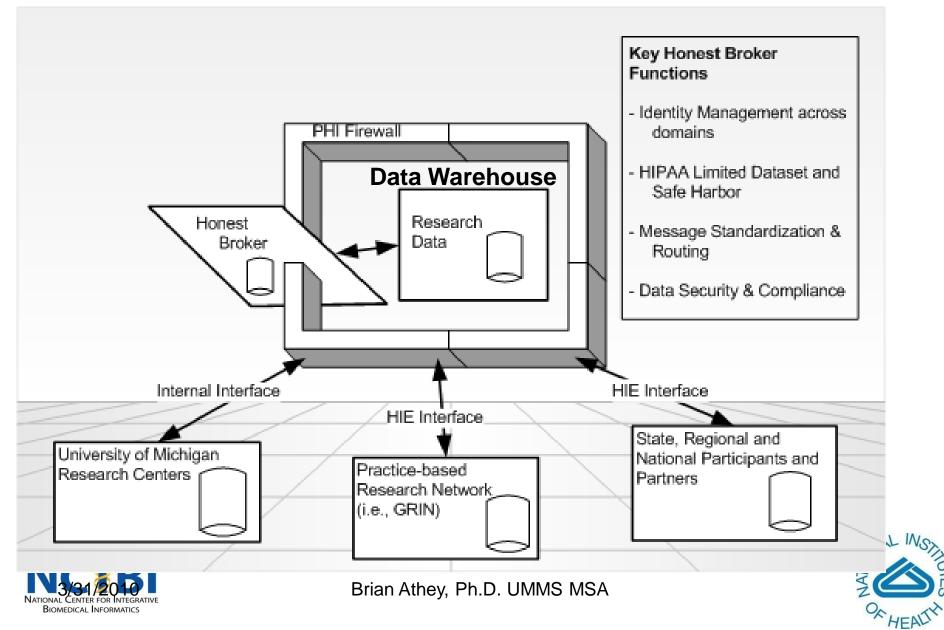




Federated Data Warehouse (Basic Research, Clinical and Outcomes Data) Spans – Genotype, Phenotype, Populations, and Comparative Effectiveness



Honest Broker: Critical Component to Implement into our Production IT Fabric



Necessary Details

Patient Cohort Search Cohort Chart Review	Clinical Rese Data Da Extraction Manage	ta Data	End user Applications	
	XML	Middleware Service - Transport		
A	ccess and Security L	Middleware Service - Security		
	SOA API	Middleware Service - Transport		
	Application Serve	Server		
Clinical Data Warehouse	Biospecimen Database	Research Databases	Application data	
	Master Person Inde	Middleware Service - data		
e.g. IC	Semantic Layer D9, CPT, SNOMED	Knowledge Model		
RIN	I-Based Data Model	Knowledge Model		
	Oracle 11G Databa	Server		

Stanford STRIDE Architecture Stack

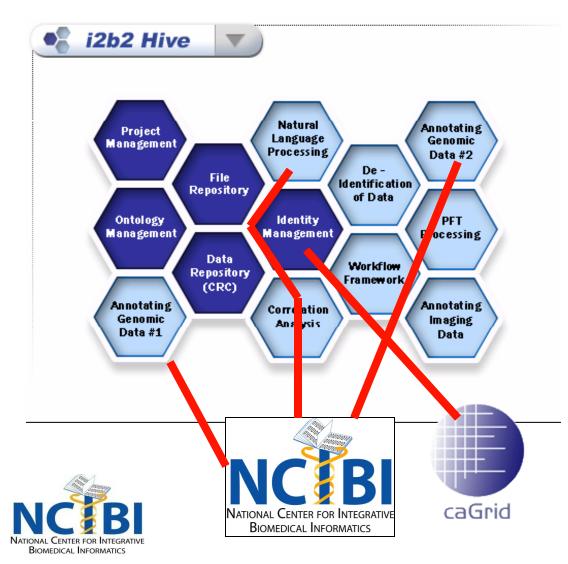
adapted from H. J. Lowe, AMIA 2009 Symposium Proceedings, p 391.



Brian Athey, Ph.D. UMMS MSA

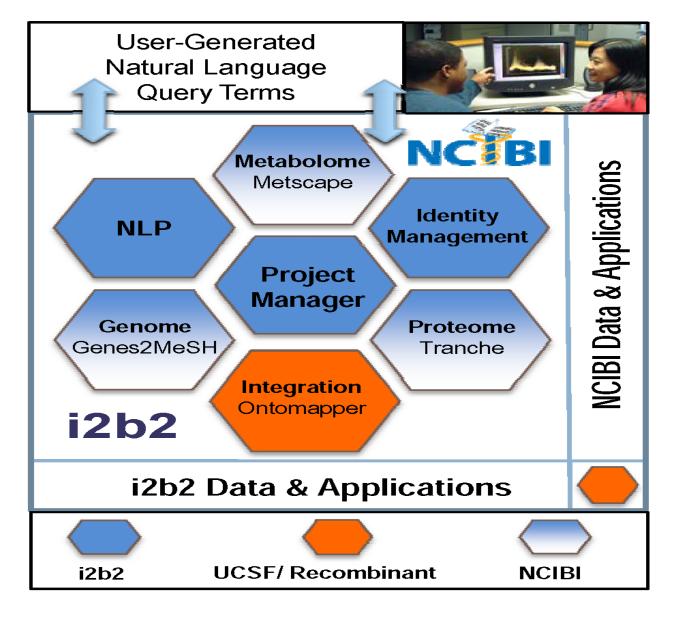


Our "Next-Gen" i2b2 Hive Vision In cooperation with Harvard University



- The i2b2 Hive is centered around two concepts –
- Services applications are "wrapped" into functional units, with functionality exposed as messages that travel to and from the cells of the hive
- Persistent data storage managed by the Clinical Research Chart



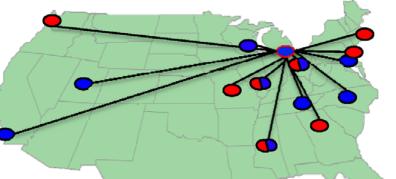






National NCIBI/CTSA Partner Network Emerging

NCIBI (University of Michigan, Ann Arbor, MI)



Subcontracting Site

- Boston University, Boston, MA
- Emory University, Atlanta, GA
- University of Illinois, Urbana-Champaign, IL
- Institute of Systems Biology (ISB), Seattle, WA
- Princeton University, Princeton, NJ
- Jackson State University, Jackson, MS

Collaborator Site

- Albert Einstein Medical School, Yeshiva University, NYC
- Duke University, Durham, NC
- Institute of Systems Biology (ISB), Seattle, WA
- Jackson State University, RTRN DTCC
- Johns Hopkins University Medical School, Baltimore, MD
- Mt. Sinai School of Medicine, NYC
- National Center for Supercomputing Applications (NCSA)
- Oak Ridge National Laboratory (ORNL)
- Ohio State University Medical Center, Columbus, OH
- Ohio Supercomputing Center (OSC)
- Rockefeller University, NYC
- University of California, San Diego
- University of Utah
- University of Wisconsin, Madison, WI
- Washington University in St. Louis, MO





NCIBI Training the Next Generation of Informaticians







Junior Faculty Actively Engaged in NCIBI Enhanced Paths to Promotion and Tenure



Steve Qin, Ph.D. Biostatistics







James D. Cavalcoli, Ph.D. Computational Medicine <u>& Bioinformatics</u>





TIONAL CENTER FOR INTEGRATIVE JIGNESH Patel, Ph.D. Peter Woolf, Ph.D. Drago Radev, Ph.D. Biomedical Information Science Chemical Engineering Information Science HENT

First Annual RCMI/NCIBI Translational Bioinformatics Summer Institute July 29 - 30th, 2009

