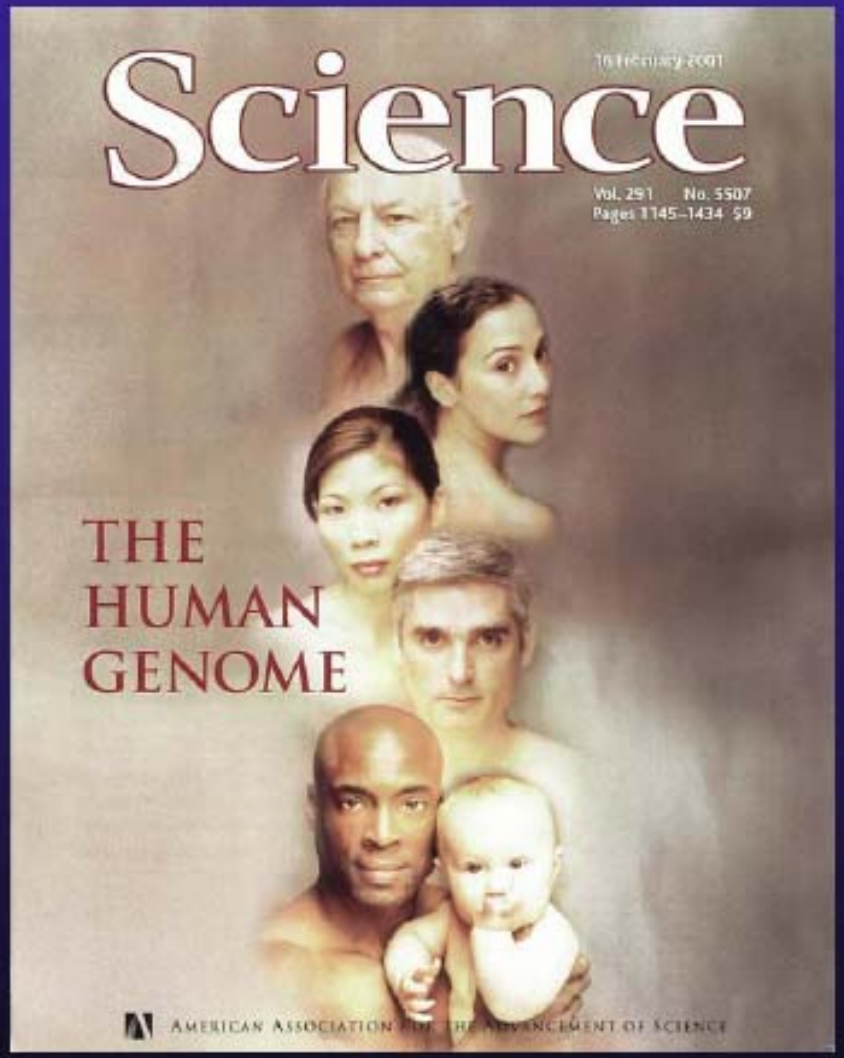
DRIVING BIOLOGICAL PROBLEMS FOR COMPUTATIONAL AND INFORMATICS APPLICATIONS

Gilbert S. Omenn, M.D.,Ph.D.

University of Michigan

NIH Roadmap/NCBC All-Centers Meeting, 19 July, 2006

National Center for Integrative Biomedical Informatics


Science
16 February 2001
Vol. 291 No. 5507
Pages 1145–1434 $9
THE HUMAN GENOME
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

15 February 2001
nature
www.nature.com
the human genome
Nuclear fission
Five-dimensional energy landscapes
Seafloor spreading
The view from under the Arctic ice
Career prospects
Sequence creates new opportunities

It’s a New World

- New Biology---New Technology
- Genome Expression MicroArrays
- Comparative Genomics
- Proteomics
- Bioinformatics

Path to predictive, personalized, preventive healthcare

FINANCIAL TIMES WEDNESDAY FEBRUARY 21 2001

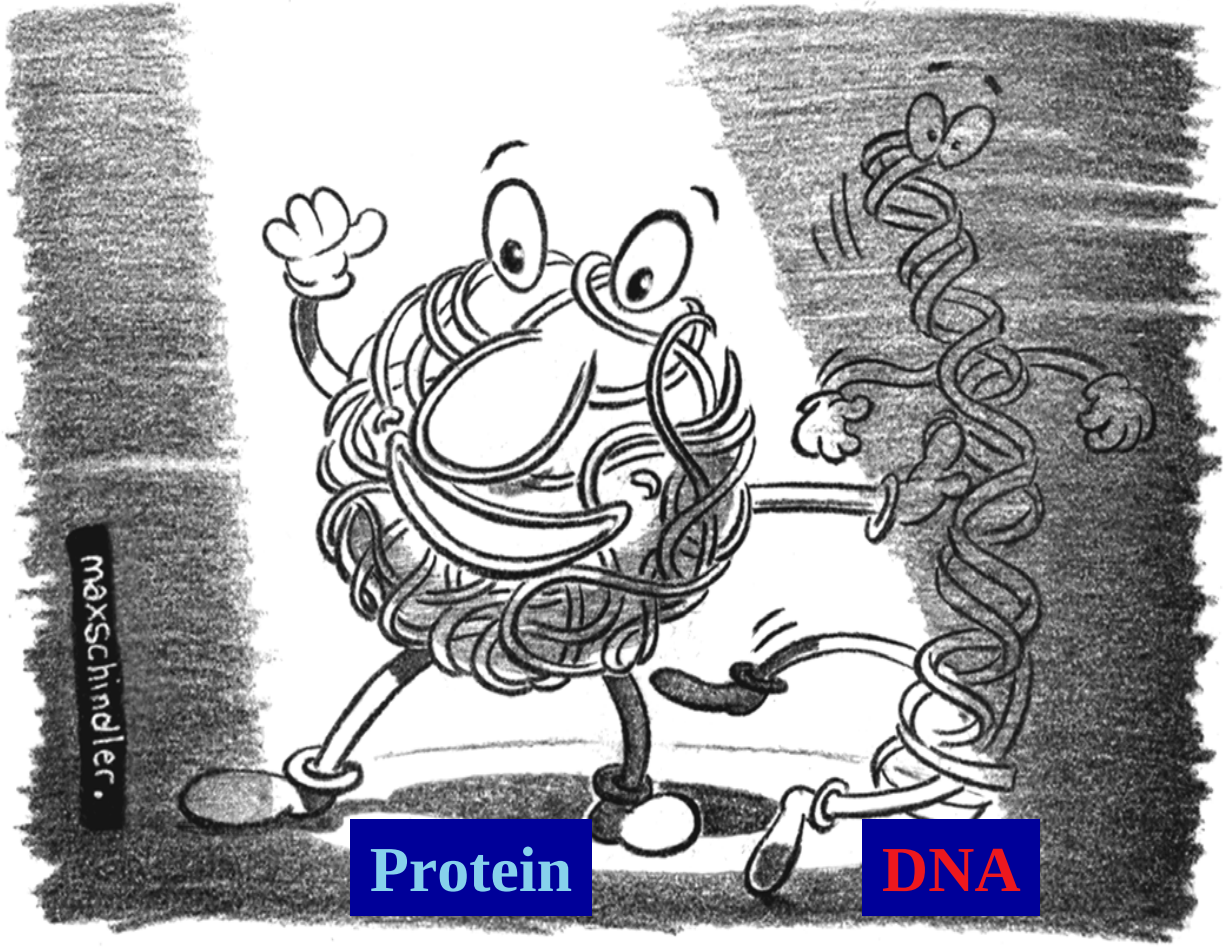
INSIDE TRACK

Strategy, Management, Technology & People

PROTEOMICS

Searching for the real stuff of life

The discovery that humans have fewer genes than expected has thrust proteins into the research spotlight, says Victoria Griffith



www.businessweek.com
BusinessWeek
APRIL 10, 2000
A PUBLICATION OF THE McGRAW-HILL COMPANIES

NEW TOOL: *Faster ways to isolate individual proteins are here*

BIOTECH'S NEXT HOLY GRAIL

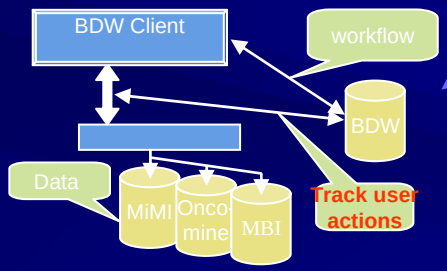
Now, companies are racing to decipher the human protein set

Granular Overview of NCIBI Activities Underway: Cores 1-3

First User Environment

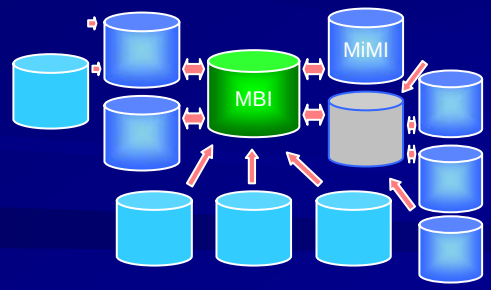
User Environment Analyses: Mirel, Ackerman labs

BDW: Biomedical Data Workbench prototype: Weymouth



Database Technologies for Deep Integration of Biological Information

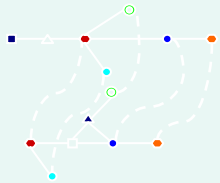
MiMI: Michigan Molecular Interactions DB: Jagadish lab, Tarcea



MBI: Molecular Biology Integration DB: States lab, Phillips

Tools for large-Scale analysis

SAGA: Patel Lab



Bayesian Nets: Woolf Lab

Workflows: Weymouth, McEachin, & Programming team

Natural Language Processing: States, Meng, Radev Labs

Concept Mapper: Rhodes, Patel, Woolf

Driving Biological Problems

Feldman Lab: Diabetes Type 1, NRF2 signaling T1DM pathways

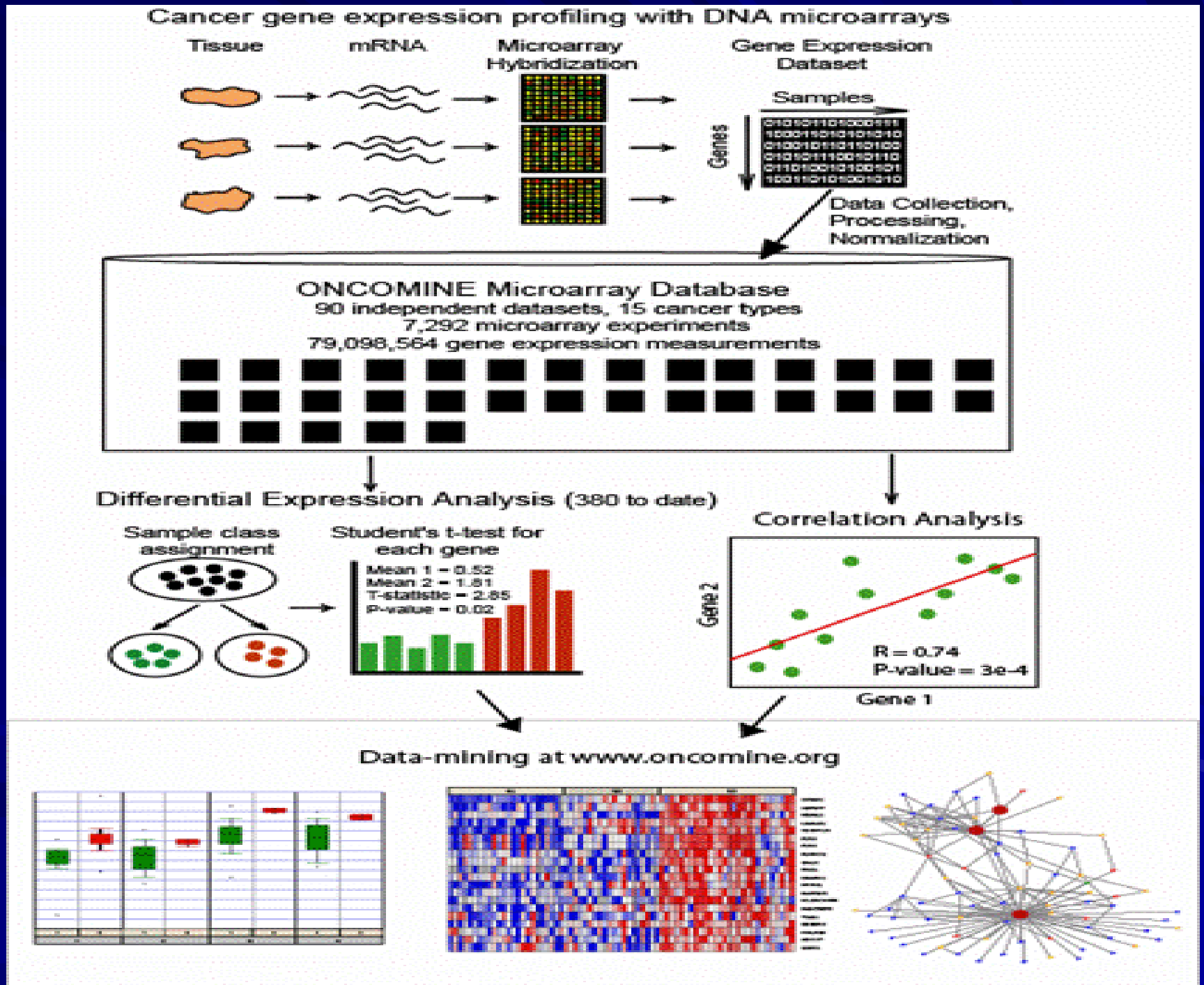
McInnis lab: Bipolar Disorder Interactions Between Genetic Linkage Peaks, WNT signaling

Boehnke Lab: Diabetes Type II SNP WGA workflow Genetic Interaction linkage workflow

Chinnaiyan and Hood Labs: Prostate Cancer Bayesian, NLP, Oncomine, COPA

Impetus for Computational Analyses of Cancers (NCIBI-DBP)

- Multi-level systems analysis of cancer biology mechanisms: genes, epigenetics, transcripts, proteins, metabolites, pathways, modules, and networks
- Recognition of heterogeneity of tumors and patients
- Earlier diagnosis of specific cancers
- Discovery & validation of therapeutic targets
- Prediction of treatment (preventive) outcome


Cancer gene expression profiling with DNA microarrays
Tissue
mRNA
Microarray Hybridization
Gene Expression Dataset
Samples
Genes
Data Collection, Processing, Normalization
ONCOMINE Microarray Database
90 independent datasets, 15 cancer types
7,292 microarray experiments
79,098,564 gene expression measurements
Differential Expression Analysis (380 to date)
Sample class assignment
Student's t-test for each gene
Mean 1 = 0.52
Mean 2 = 1.51
T-statistic = 2.85
P-value = 0.02
Correlation Analysis
Gene 2
R = 0.74
P-value = 3e-4
Gene 1
Data-mining at www.oncomine.org

CANCER PROFILING DATABASE
ONCOMINE™

Home About Support Login Register

Statistics

Studies - Catalog: 958
Studies - Analyzed: 206
Microarrays: 19913
Data points: 420010520
Cancer Types: 50
Registered Users: 9454

Highlights

Mining for Regulatory Programs

Oncomine recognized in JAMA

Oncomine highlighted in Discover Magazine

Links

Human Interactome Map

Chinnaiyan Laboratory

Adobe SVG Viewer download

About Oncomine

Welcome to Oncomine, a resource for examining gene expression in cancer. The goal of the project is to collect, standardize, analyze, and deliver published cancer gene expression data to the research community. Probe the expression of a gene across thousands of cancer samples or explore genes, processes, and pathways deregulated in a particular type of cancer. Oncomine pre-computes cancer profiles, clusters, and gene set modules so you can focus on discovery. Read more here.

Oncomine was developed by physicians, scientists, and software engineers at the University of Michigan. Direct questions and comments to oncomine@umich.edu

Oncomine Features

- Microarray datasets from many labs
- Sort by gene, cancer type
- Test for differential expression
- Test for co-expression, map to pathways & molecular concepts
- Estimate biological & statistical significance
- Link to other transcript datasets (MPSS), proteomics, cellular localization of proteins (Protein Atlas), all kinds of clinical stratification

Oncomine Utility - Gene

- Which normal tissues express my gene?
- Which cancer types over-express my gene?
- Which clinical, pathological and molecular subtypes of cancer differentially express my gene?
- Which other genes are co-expressed with my gene in a given cancer type?

Oncomine Utility - Cancer

- Which genes are most over-expressed in my cancer type of interest?
- Which kinases / membrane bound receptors / secreted proteins are most over-expressed?
- Which pathways, processes, chromosome arms and regulatory motifs are most disproportionately over-expressed?
- Which gene products are most likely to be detectable in plasma?

Two Aims of NCIBI (UM/ISB) Prostate Cancer DBP

- **Create an integrated model for prostate cancer progression using microarray gene expression, MPSS transcript, proteomics, and protein–protein interaction data and text.**

  Perform integrative analysis of proteomics (ICAT/MS/MS) and transcriptomics data of early androgen response program and its relationship to prostate progression.

Prostate Cancer DBP (cont'd)

- **Explore at a systems level the roles of androgen receptor (AR) regulation in androgen-dependent and androgen-independent states of prostate cancers, and the relationship of androgens to tumor progression.**

  Apply natural language processing (NLP) searches and Bayesian network modeling by Core 2 to build androgen response network to describe and predict perturbations by androgen, anti-androgens, and specific siRNA knockdown experiments.

The Crucial Influence of Heterogeneity

- The need to speak of "cancers" in the plural: the illusion of "The War on Cancer" (1970)
- The emergence of new signaling pathways, receptors, molecular targets, and mRNA and protein expression patterns
- The importance for research strategies on carcinogenic mechanisms, diagnosis, treatments, and preventive interventions
- The ramifications of subpopulation analysis and pharmaco-genomics

The Crucial Role of Heterogenity in the Biology and Treatment of Cancers

- Breast Cancers (plural)

  Her2/neu amplification, in 15-20% of patients; basis for Herceptin targeted Rx
- CML, cytogenetic abnormality, t9:22, Philadelphia chromosome: reflects bcr-abl translocation; target for Gleevec
- Lung Cancers: EGFR mutations (5-10% of cases)

Cancer Molecular Analyses

(Microdissected) Tumors

Profiling

Genomic

Transcriptomic

Proteomic

Goals

Molecular Classification

Novel Biomarkers

Model Cell Lines

Profiling

Secreted proteins

Membranes & Antigenic proteins

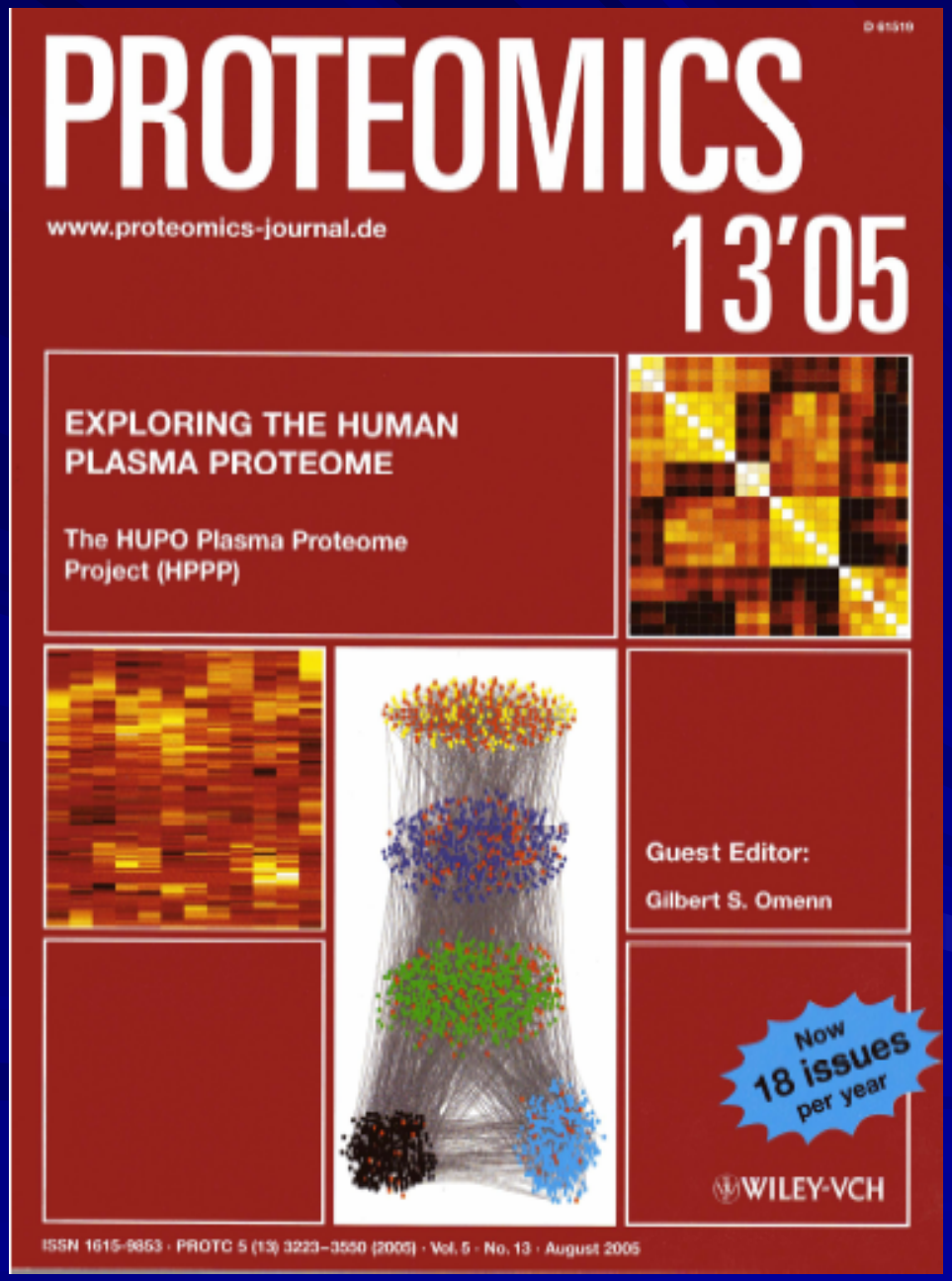
Goals

Molecular Classification

Novel Biomarkers: Proteins and Auto-Antibodies

Rationale for Plasma Profiling in Patients

- Most accessible, least invasive specimen from humans.
- Feasible to monitor over time
- May capture proteins released from all organs and tissues in health and disease---detect differences in concentration, as well as protein modifications
- Multiple complementary methods available
- HUPO Plasma Proteome Project findings and database are a resource to facilitate these applications: www.bioinformatics.med.umich.edu/hupo/ppp; www.ebi.ac.uk/pride; www.peptideatlas.org.

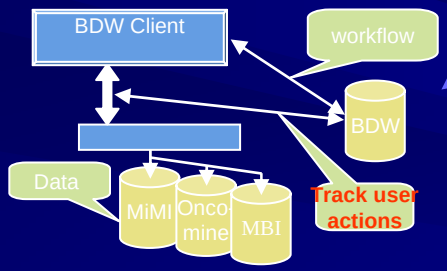

PROTEOMICS
www.proteomics-journal.de
13'05
EXPLORING THE HUMAN PLASMA PROTEOME
The HUPO Plasma Proteome Project (HPPP)
Guest Editor:
Gilbert S. Omenn
Now 18 issues per year
WILEY-VCH
ISSN 1615-9853 · PROTC 5 (13) 3223–3550 (2005) · Vol. 5 · No. 13 · August 2005

Granular Overview of NCIBI Activities Underway: Cores 1-3

First User Environment

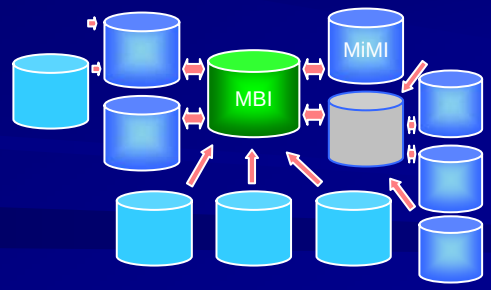
User Environment Analyses: Mirel, Ackerman labs

BDW: Biomedical Data Workbench prototype: Weymouth



Database Technologies for Deep Integration of Biological Information

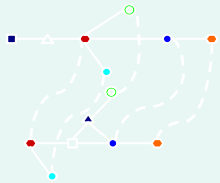
MiMI: Michigan Molecular Interactions DB: Jagadish lab, Tarcea



MBI: Molecular Biology Integration DB: States lab, Phillips

Tools for large-Scale analysis

SAGA: Patel Lab



Bayesian Nets: Woolf Lab

Workflows: Weymouth, McEachin, & Programming team

Natural Language Processing: States, Meng, Radev Labs

Concept Mapper: Rhodes, Patel, Woolf

Driving Biological Problems

Feldman Lab: Diabetes Type 1, NRF2 signaling T1DM pathways

McInnis lab: Bipolar Disorder Interactions Between Genetic Linkage Peaks, WNT signaling

Boehnke Lab: Diabetes Type II SNP WGA workflow Genetic Interaction linkage workflow

Chinnaiyan and Hood Labs: Prostate Cancer Bayesian, NLP, Oncomine, COPA