Show me the evidence!

Discovery through exhaustive search of the evidence

Isaac S. Kohane, MD, PhD

Associate Professor of Health Sciences and Technology and Pediatrics, Harvard Medical School

Co-Director HMS Center for Biomedical Informatics

Director, Countway Library of Medicine, Harvard Medical School

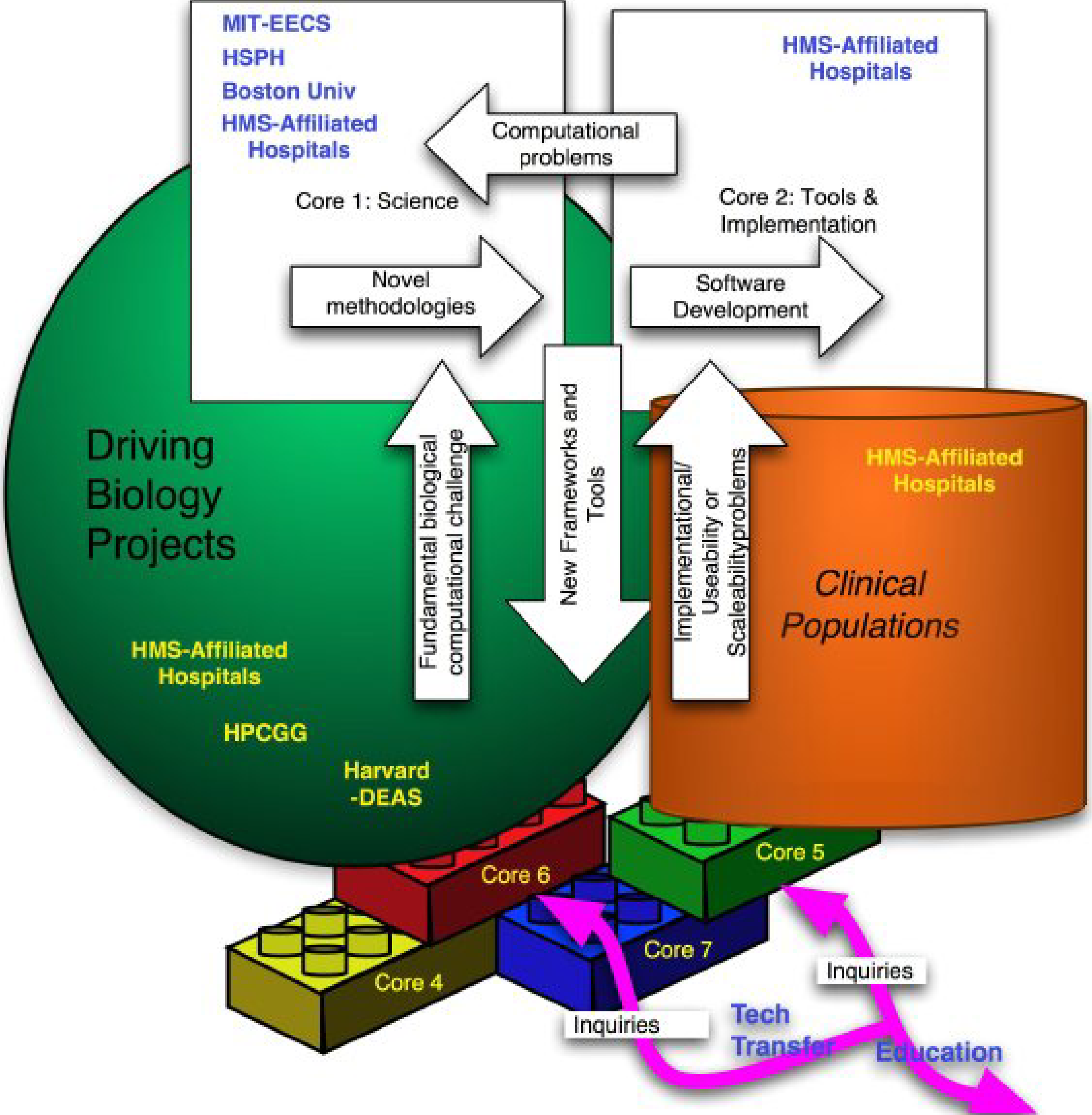
Director, Children’s Hospital Informatics Program?

One Very Old Biomarker:
A Prismatic Example of the Failure of EBM

- **1410 original articles on PSA screening for cancer**
- **179 review articles**
- **With 1000’s of gene products, how do we systematically address this problem?**

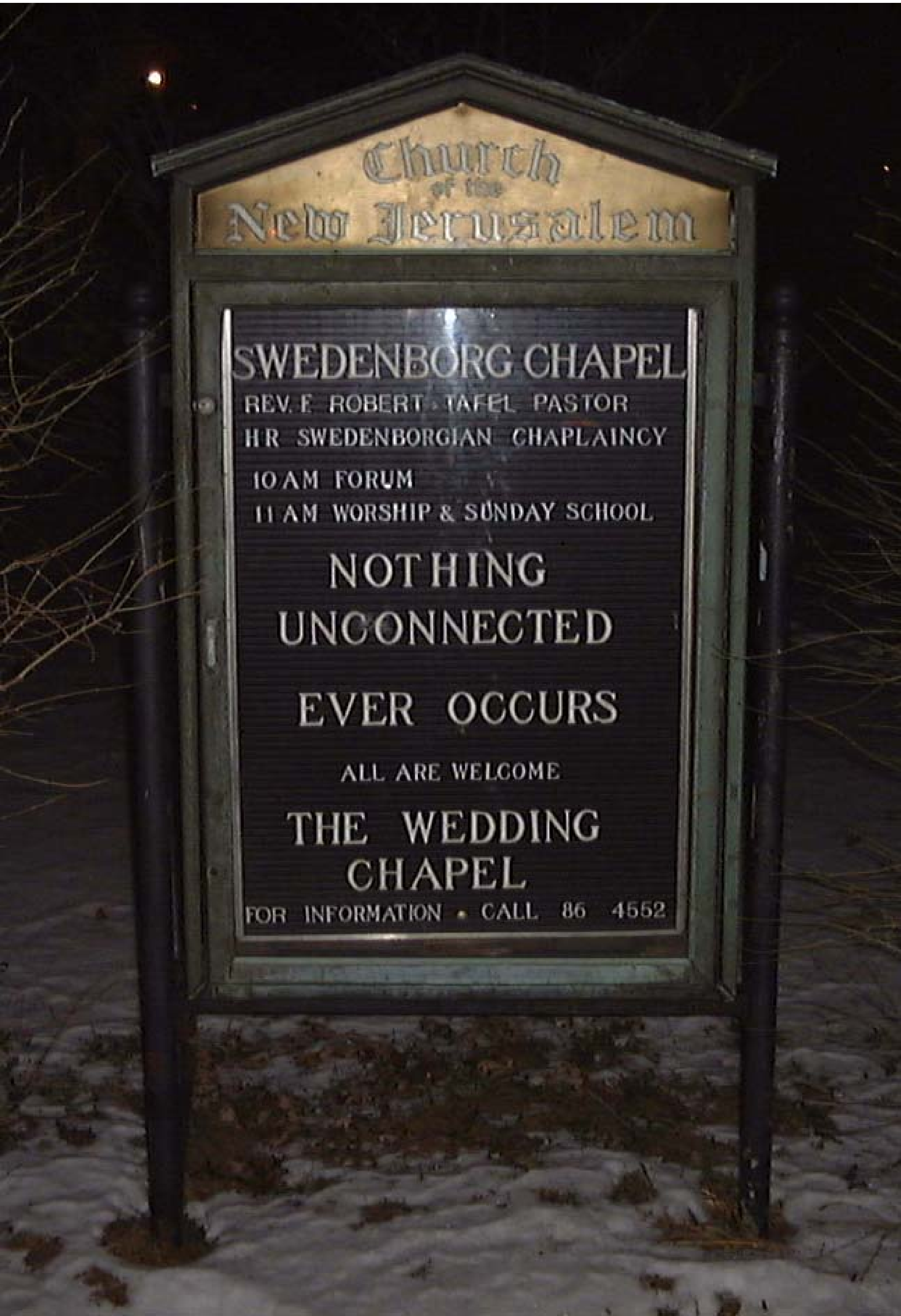
QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.



i2b2

[Informatics for I](#)



Relevance Networks & Computational Challenges

Construction of Relevance Networks 1

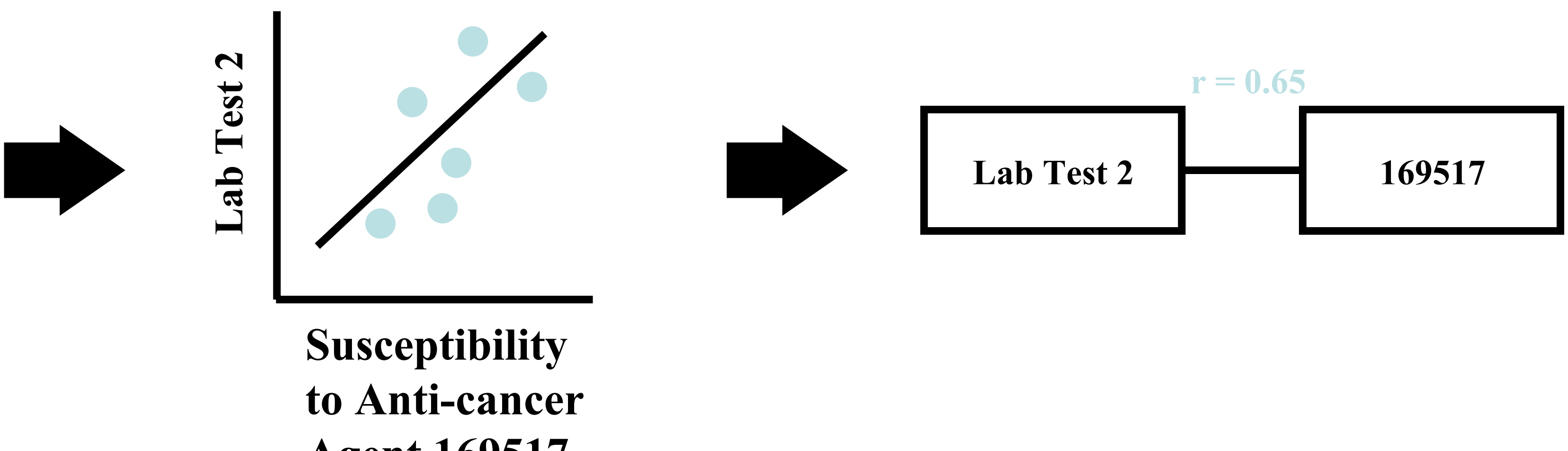
- Patients and cell lines are analyzed as cases
- Clinical parameters, laboratory tests, RNA expression, and susceptibility to anti-cancer agents are all example features of those cases

Patient, Cell Line, Time, etc.

Lab Test 1	**Lab Test 2**	**Clinical Param 1**	**RNA Expr J02923**	**Susceptibility to Anti-cancer Agent 169517**
138	3.7	105	0.7	8.1
134	4.5	99		2.1
132	5.3	102	7.4	3.3

Construction of Relevance Networks 3

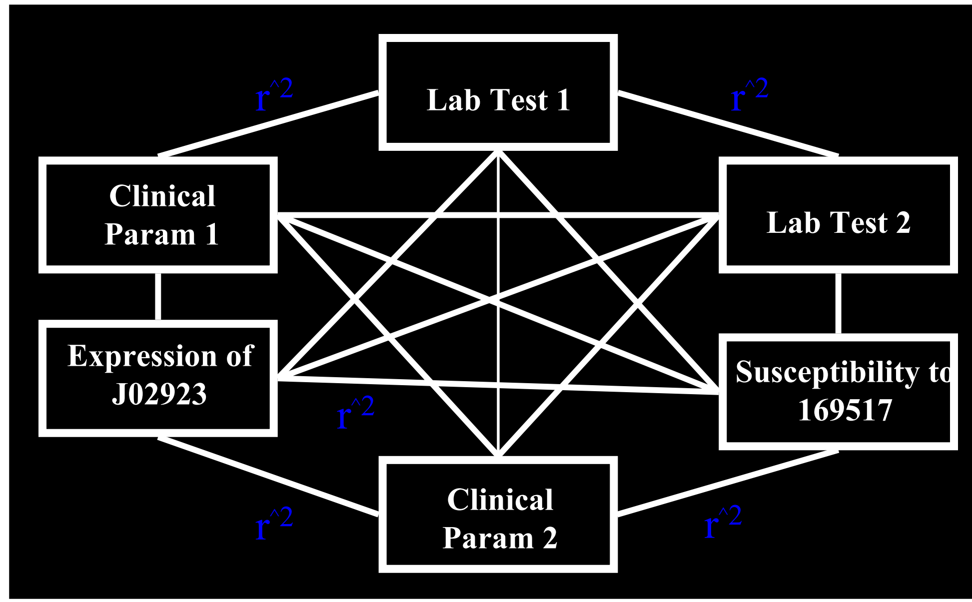
- Perform a pairwise comparison between all features
- For each scatter plot, we fit a linear model and stored
  - Correlation coefficient r



- Every feature is completely connected to every other feature by a linear model of varying quality

Construction of Relevance Networks 4

- $r^{\wedge 2} = r^2 * r / abs(r)$
- Choose a threshold $r^{\wedge 2}$ to split the network
- Drop links with $r^{\wedge 2}$ under threshold

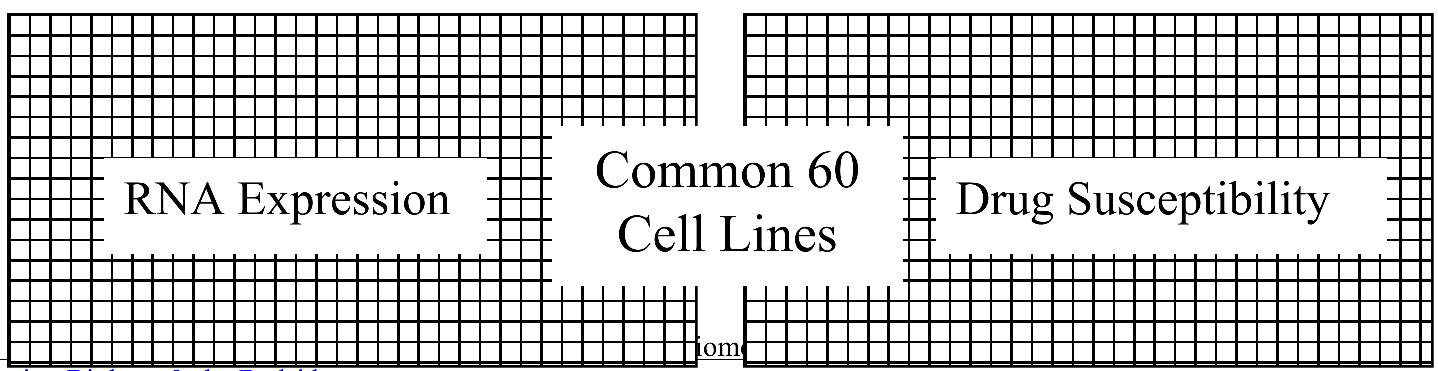


- Breaks the completely connected network into islands where connections are stronger than threshold
- Islands are what we call "relevance networks"
- Display graphically, with thick lines representing strongest links

The New Pharmacology

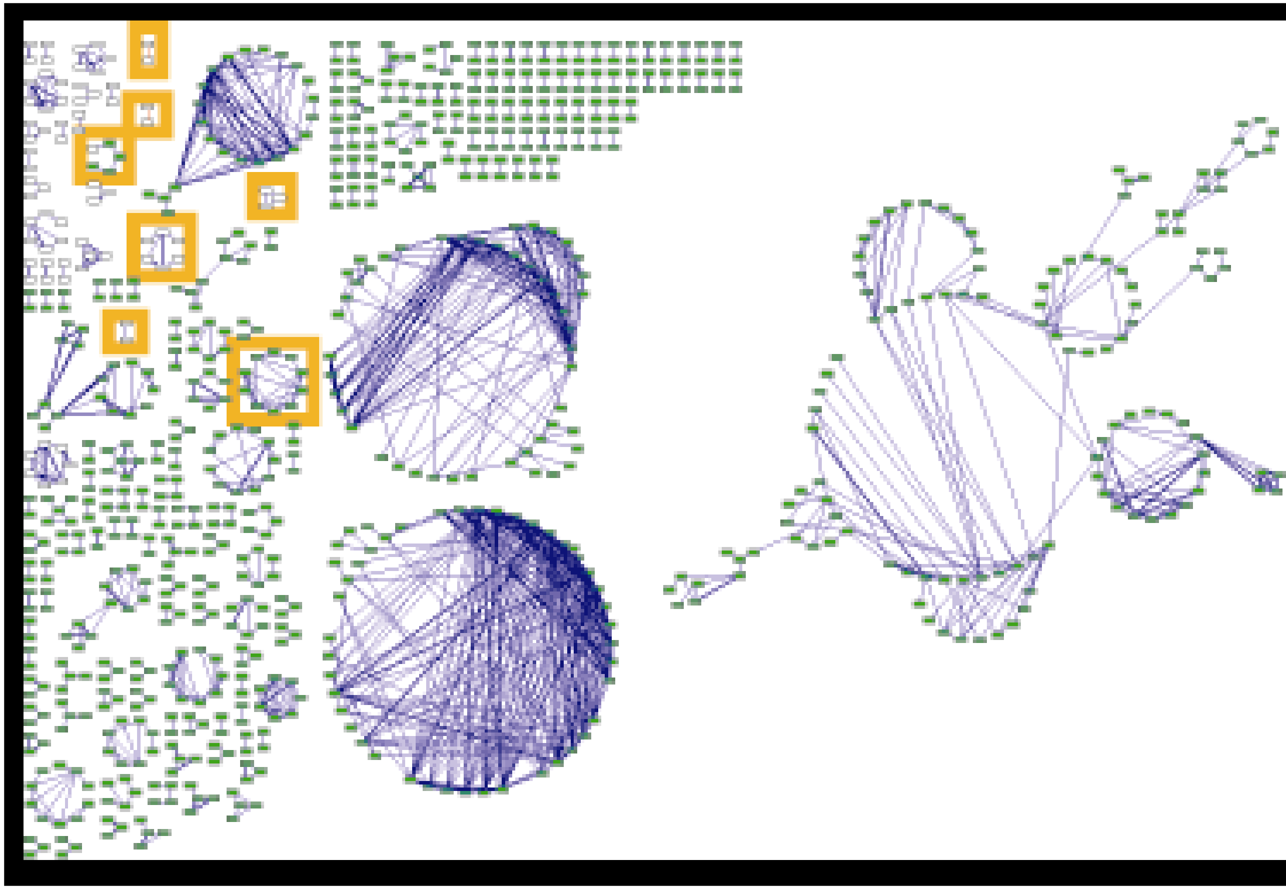
- RNA expression in NCI 60 cell lines was determined using Affymetrix HU6000 arrays
  - 5,223 known genes
  - 1,193 expressed sequence tags

- The RNA expression data set and Anti-cancer susceptibility data set were merged, using the 60 cell li vo table



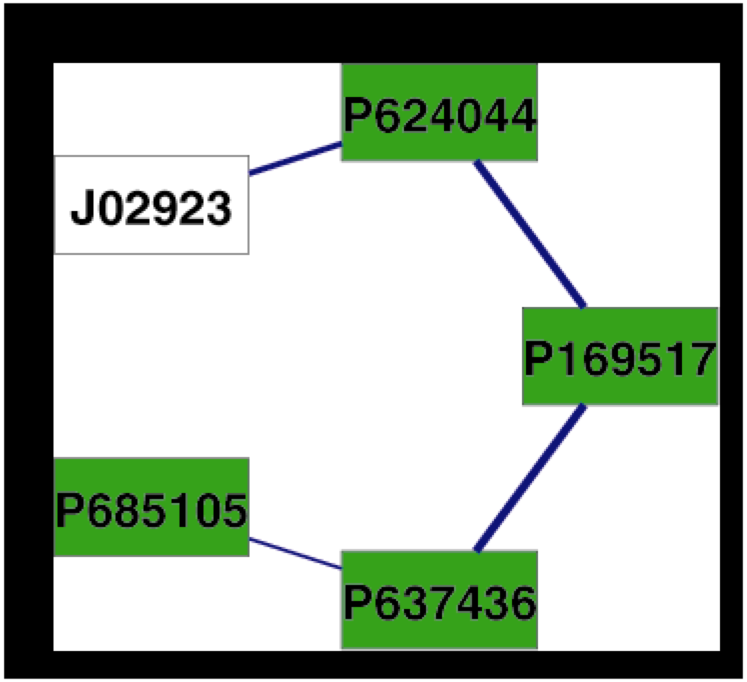
Genes and Anti-Cancer Agents

- Threshold r^{2} was 0.8
- 202 networks
- 834 features out of 11,692 (7.1%)
- 1,222 links out of 68,345,586 (.0018%)
- Only one link between a gene and anti-cancer agent



Genes and Anti-Cancer Agents

- Elevated levels of J02923 (lymphocyte cytosolic protein-1, LCP1, L-plastin, pp65) is associated with increased sensitivity to 624044
- Agent 624044 is 4-Thiazolidinecarboxylic acid, 3-[[6-[2-oxo-2-(phenylthio)ethyl]-3-cyclohexen-1-yl]acetyl]-2 thioxo-, methyl ester, [1R-[1a(R*),6a]]- (9CI))
- LCP1 is an actin-binding protein involved in leukocyte adhesion
- A role for LCP1 in tumorogenicity has been previously postulated
- Low level expression of LCP1 is thought to occur in most human cancer cell lines
- Other thiazolidine carboxylic acid derivatives are known to inhibit tumor cell growth



Butte et al. PNAS 2000

RelNet: Live Patient ‘Phenotypic’ Data

- After patient cohorts are identified through the electronic medical records analysis and they agree to participate in the ‘genomics part’ of the study (see Globe story)

- RelNet methodology is easily expandable to non-numeric, patient demographics and other phenotypic data variables identified from the electronic medical records part of the analysis

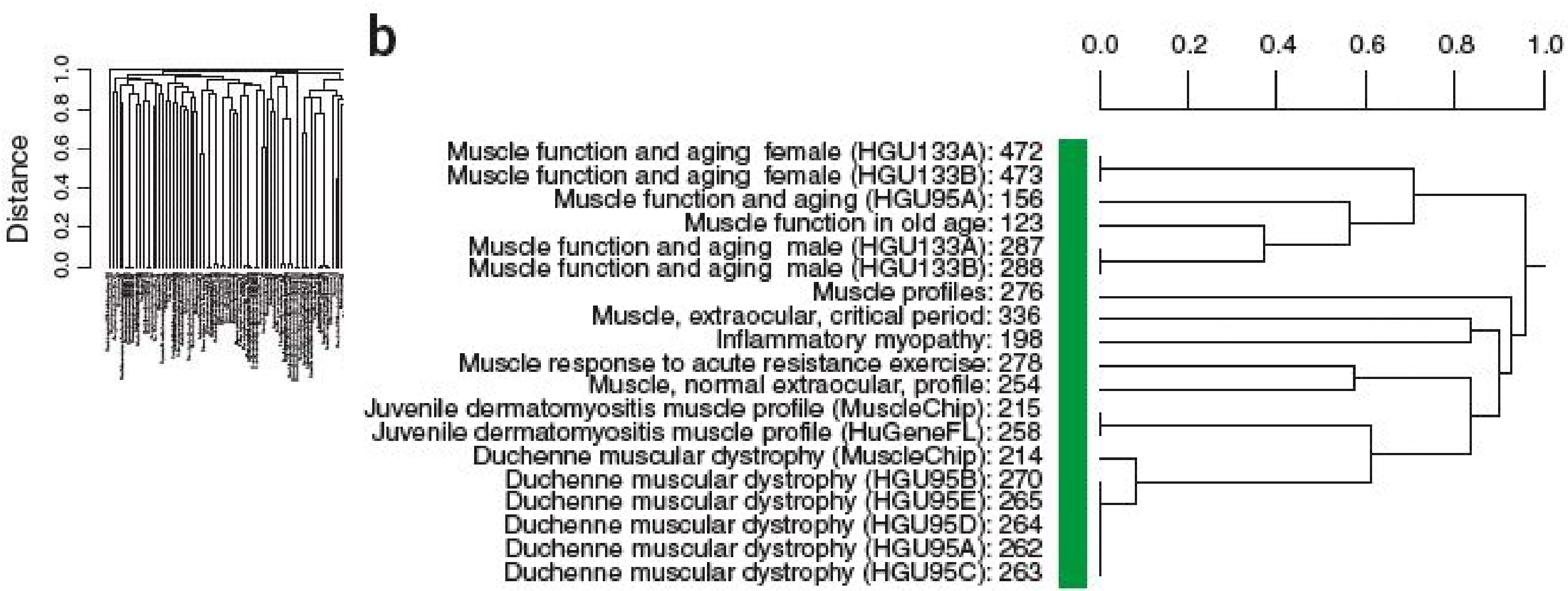
- This will potentially increase the number of variables with thousands that will make the problem even more challenging

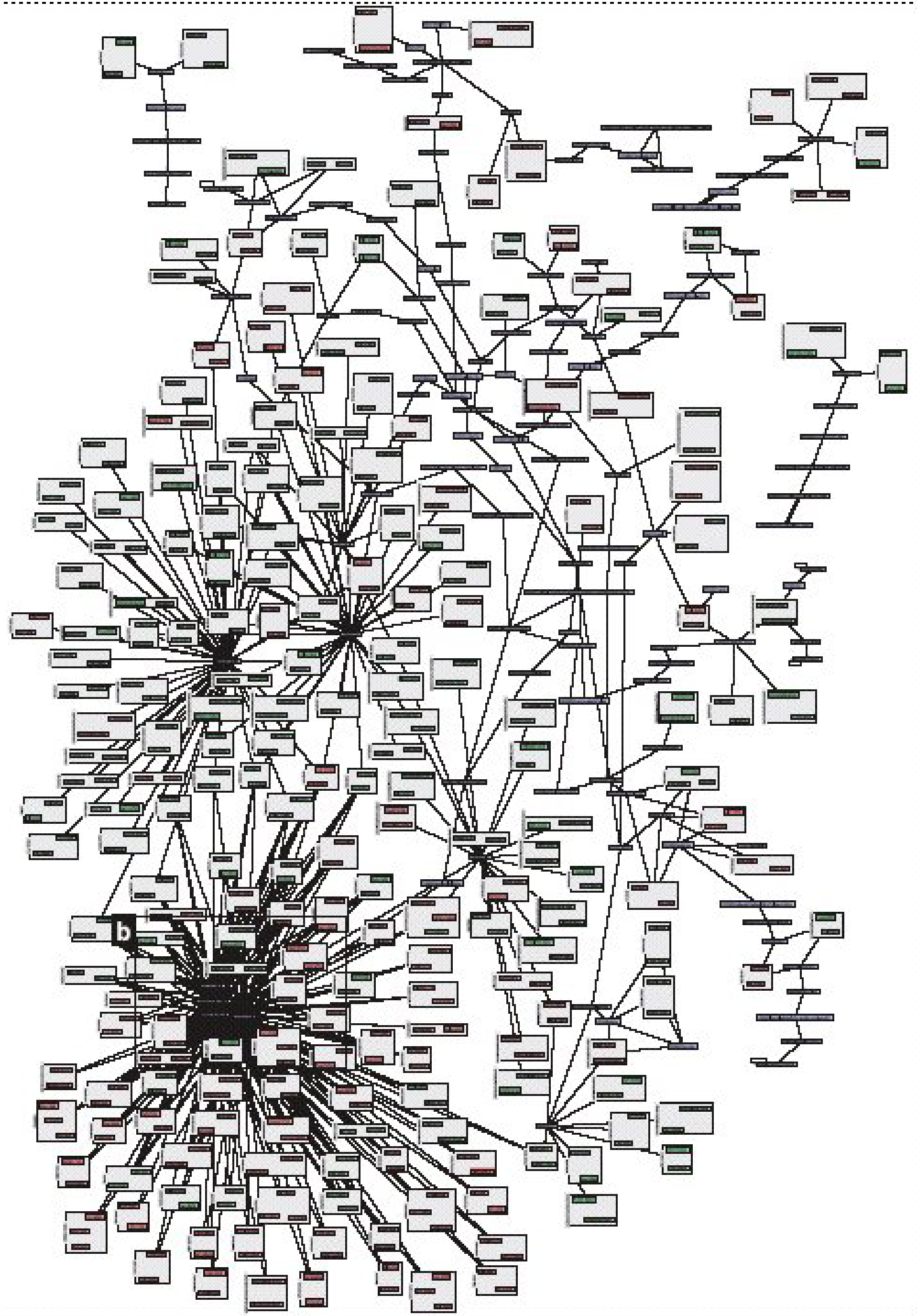
_computational BIOLOGY

ANALYSIS

Creation and implications of a phenome-genome network

Atul J Butte[1] & Isaac S Kohane[2]



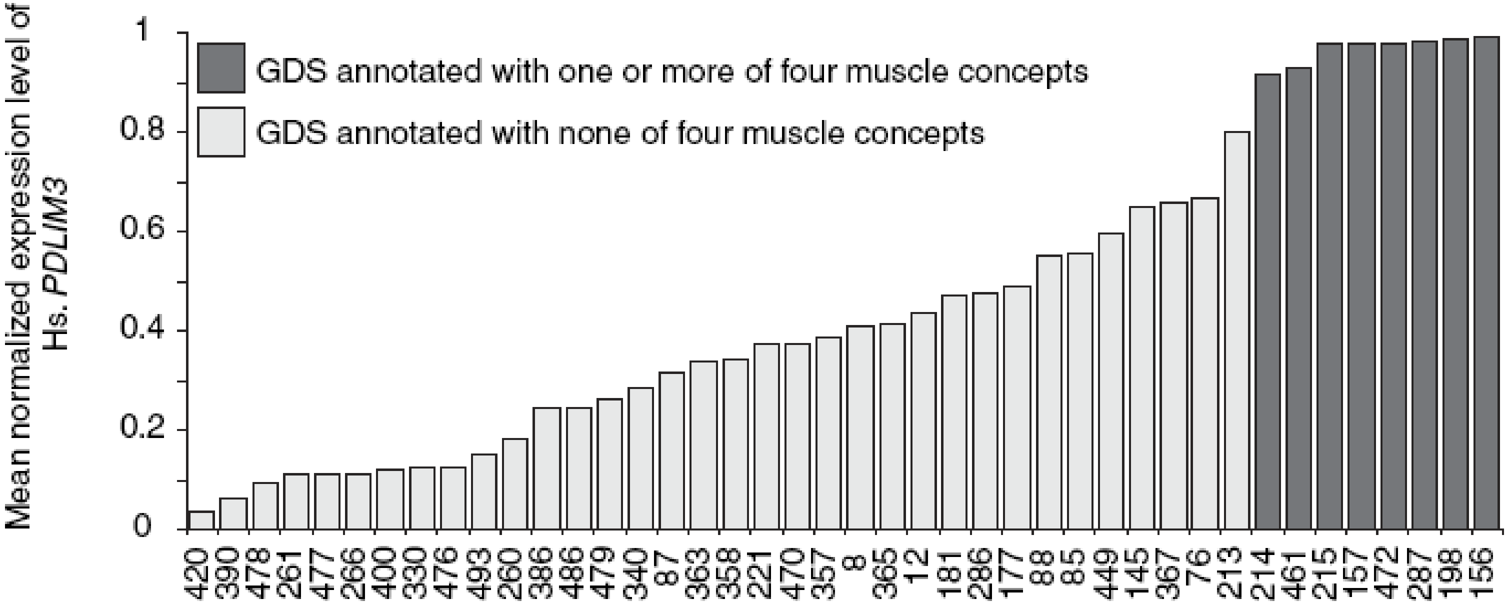


New taxonomy

Discoveries samples:

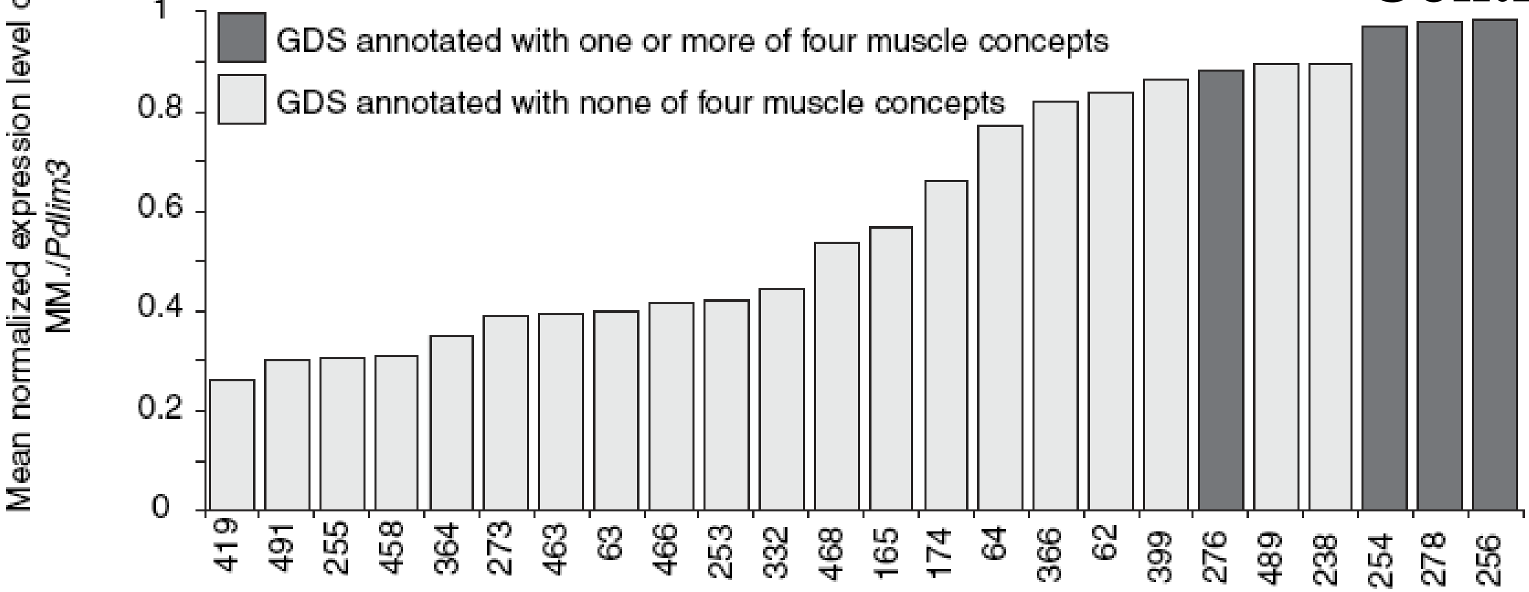
BDNF expressed at lower levels in data sets associated with aging

BDNF previously shown to be decreased in aging skin. 9 other genes associated with aging.



d

Biomarkers as Continuum



The Incidentalome

A Threat to Genomic Medicine

Isaac S. Kohane, MD, PhD
Daniel R. Masys, MD
Russ B. Altman, MD, PhD

GENOMIC MEDICINE IS POISED TO OFFER A BROAD ARray of new genome-scale screening tests. However, these tests may lead to a phenomenon in which multiple abnormal genomic findings are discovered, analogous to the "incidentalomas" that are often discovered in radiological studies. If practitioners pursue these unexpected genomic findings without thought,

There is a rich literature in radiology on the "incidentaloma," which is a finding (most commonly a mass) found on computed tomography or magnetic resonance imaging studies ordered for symptoms or concerns totally unrelated to the gland in which the mass is found. The workup of an incidentaloma is complicated by concerns that it may be associated with malignant disease and, at least initially, the lack of good data on the prevalence of malignant disease in the general population. Incidentalomas occur because imaging modes do not only report on the areas of direct clinical concern but, incidentally, report on all organs in the field of view.[1]

The single gene case

- Single genomic test that is 99.9% sensitivity (true positive rate) and 99.9% specificity (true negative rate) for a rare treatable disease X
  - Works well in patient group that is suspected of disease (high prior)
- In the general population (disease 1 in 100,000)
  - In 10,000,000 patients
  - 100 true positives, 10,000 false positives
  - Less than 1 in 1000 positives are true.

Impanelling the Incidentalome

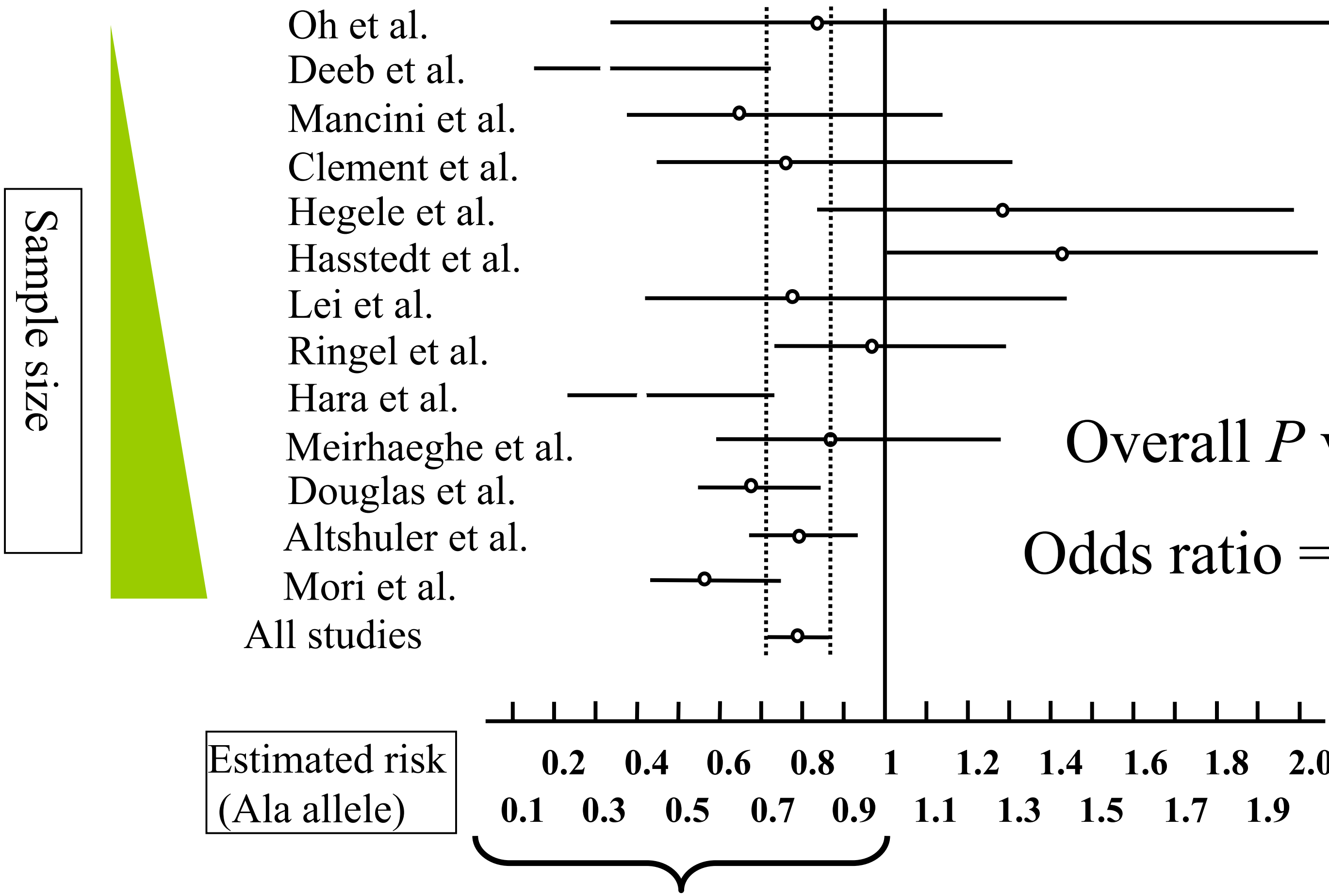
- The economic case
- The comprehensiveness case
- Using a much much better test
  - sensitivity of 100% and a specificity of 99.99%
  - Only 10 in 100,000 false positives
- A chip with 10,000 independent tests

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Averting the Incidentalome's Threat

- Education of clinicians
- Informatics for decision support of sequencing of clinical testing
- Real evidence, real probabilities.

Example: PPARγ Pro12Ala and diabetes



Overall *P* value = 2 x 10^{-7}

Odds ratio = 0.79 (0.72-0.86)

Ala is protective

Courtesy J. Hirschhorn

Airways Disease Research at *Brigham and Women's Hospital*

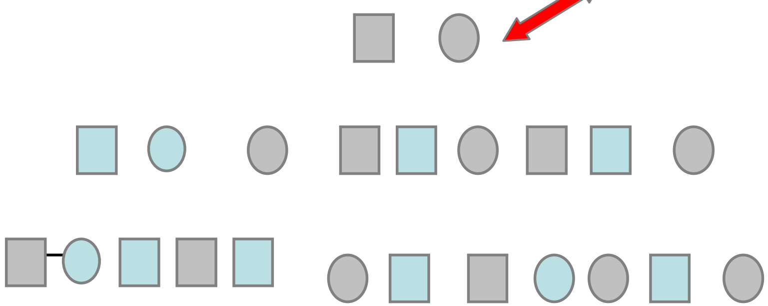
China - 800
ASP

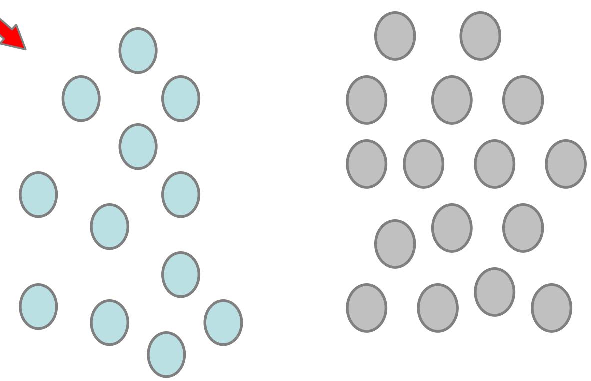
Linkage

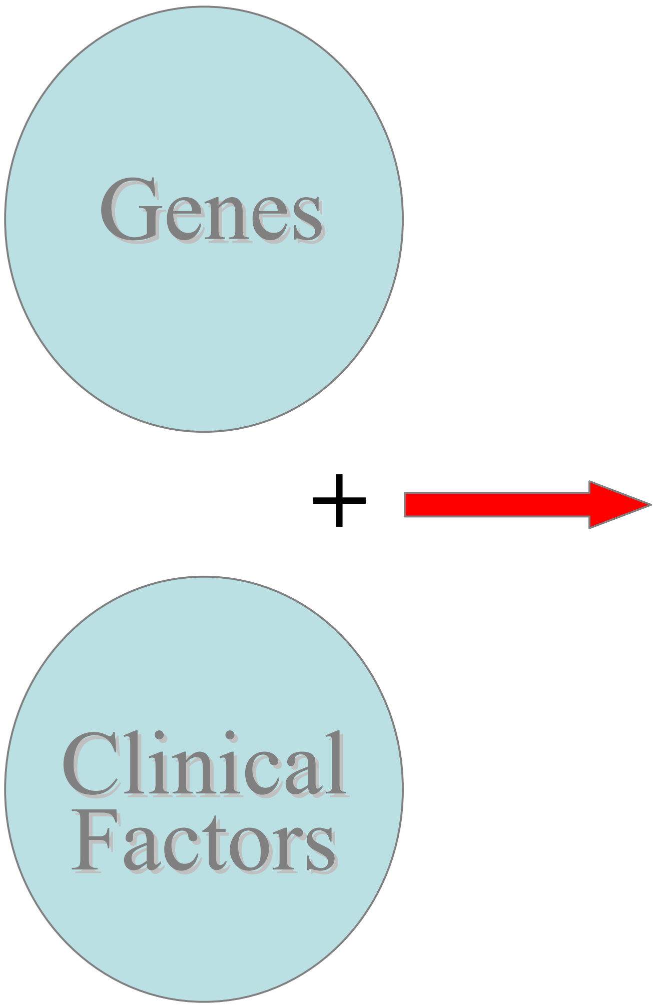
Fine Mapping

Candidate Genes

Pharmacogenetics







Who?

Health Care Utilization

(Hospitalization, ED Visits)

96,000 patients identified through i2b2 datamart

Thank you

http://www.i2b2.org