



Finding the Glue to Fasten Clinical to Genomic Databases

Isaac “Zak” Kohane

Division of Health Sciences Technology of
Harvard and MIT



Bioinformatics

Clinical Applications

Children's

Hospital

Informatics

Program

Progeria
Progeria is a rare condition affecting 1 in 6 million births, now associated with a mutation of the Lamin A gene. To further explore the molecular pathogenesis of progeria, in collaboration with the Progeria Research Foundation, CHIP has obtained and is analyzing gene expression patterns of fibroblasts derived from progeria patients.

Newborn Infant Genomics
Newborn genomics refers to use of DNA to assess that may have long term effects, including chronic lung disease, neural tube and gastrointestinal injury. Functional genomics (FG) profiling will be used to correlate gene expression patterns with functional markers of lung inflammation and clinical health status. Genes profiles will support identification and treatment of high risk premature infant populations.

Cancer & Development: Shared and Divergent Mechanisms. CHIP is actively collaborating with researchers at the Dana-Farber Cancer Institute to understand the molecular programs underlying normal somatic cell and lung cancer through comparative analysis of developmental.

SNP-based analysis of the human genome
Single Nucleotide Polymorphisms (SNPs) are an increasingly popular tool for genetic research, since they are very common and easy to detect and characterize. We are developing computational tools and techniques to evaluate the effects of SNPs on our genome. We are interested in studying the consequences of SNPs on protein domains, on transcription factor binding sites, on evolutionarily conserved regions and on other functional elements of our genes. Our efforts will lead to a greater understanding of the organization of our genome, as well as an increased ability to correlate genetic variation with diseases.

Technology Linking Parents and Providers
We developed a patient centered technology that facilitates collection of medical data directly from parents acting as nurse practitioners. The system uses the expertise of parents working in their own multi-media based structure and gathers data regarding a patient's symptoms, care requirements, and needs in greater detail. A clinical trial of the ActiveVoice (AV) System (2004) will investigate its ability to produce improved rates to improve level of adherence to care as well as to foster better communication between parents and providers.

Patient Safety & Information Technology
As Children Hospital Ontario seeks to improve the clinical care and patient outcomes that fully address the processes of care for children in both inpatient and ambulatory settings, we will study the impact of such technologies and control on patient safety, especially in relation to patients, who will have research that have previously done. This will include prospectively studying the effect of Computerized Physician Order Entry (CPOE) including data-based alerts, automated Clinical Practice Guidelines (CPG), and practice decision aids, in a randomized patient environment.

Pediatric Renal Disease, Organ Transplantation, and Immunogenetics
Pediatric renal disease includes a spectrum of renal diseases that are most common among causes of end stage renal disease requiring dialysis or transplantation in children. CHIP faculty are studying genetic and proteomic factors to facilitate accurate classification of pediatric renal disease, with a clinical research team research project. Collaborations with the Southwestern Children's Transplant Center and Brigham and Women's Hospital also include transplant immunogenetics analyses in adult renal patients, liver, heart/lung, and pancreas transplantation.

Pharmacist: Killing Unsuccessful Drugs Sooner
Failed new chemical entities cost the pharmaceutical industry billions of dollars per year. Reducing this rate to less than 20% would save hundreds of millions of dollars per drug. CHIP researchers from the Ontario EPRIM/INFORM application called Pharmacist that leverages a specially programmed AI company's data to evaluate new molecule candidates for their chemical entities as they go to development/publication.

Data Integration
The Medical Informatics Application and Transmission Database (MEDITE) system provides access to multiple information sources by connecting the number of connected imaging databases and performing data-mapping between systems. Cancer based information systems of database integration.

Personal Interconnected Notary & Guardian (PING) Individual electronic healthcare records are increasingly difficult to share by placing the control in control of their own health care and not providing the correct information. PING enables providers and researchers to access critical information electronically, including financial, policy and consent.

MMSEH CHIP has developed a tool to assist with the creation of hospital level clinical information and patient health records. It enables data from health information systems to be used in decision support systems. Using predictive and decision support systems, MMSEH provides clinical and operational systems across various data sources on the MMSEH network in partnership with other research and public health authorities to address all patients with similar diseases, close together in time or space.

Geotemporal methods for real time outbreak detection
Researchers at CHIP are developing methods to enable detection of potential patterns of disease in populations. This includes a thorough understanding and model of the evidence "temporal" patterns of disease as well as a sensitive detection to find public health authorities to diseases of patients with similar diseases, close together in time or space.

Building Tools for Biosurveillance and Biopreparedness
CHIP researchers apply tools from machine learning, complex systems, epidemiology, and bioinformatics to the problem of public health and preparedness. The ADIOS (Advanced Distributed Information Systems for Biological and Operational) system performs automated, near real time capabilities for laboratories as well as publicly accessible systems. The research group is developing the multi-scale web-based surveillance system for the Commonwealth of Massachusetts national detection algorithm, mathematical models of disease spread, decision support systems to assist health clinicians in the event of a biological attack, as well as multi-scale health information systems to provide clinical and public health coordination.

Linking Genomic Data to Patient Survival
Development of statistical techniques for correlating high-throughput gene expression and proteomic data to various phenotypes, especially patient survival time. Predictive models according to the large number of covariates have the potential to be used for prognostic purposes as well as for biological understanding.

Neuromuscular Disease
Gene expression profiles of muscle biopsy specimens from patients with inflammatory myopathies are studied to improve disease classification and our understanding of pathogenesis.

Diabetes Genomics
Now epidemic in the U.S., type 2 diabetes is marked by an inability of the body's cells to use insulin properly. CHIP is providing key information support for the Diabetes Genome Anatomy Project, a five-year, five-institution project that has begun to unravel the complex genetics and biology behind insulin resistance, type 2 diabetes, and prediabetes. Part III, (Burt AJ, et al. PNAS 100:8666-71 (2003))

Shared Pathology Informatics Network (SPIN)
Researchers from a host of pathology centers across the province could enable many important phenotypic-genomic correlations. CHIP researchers are leading efforts with other research informatics to create a province-wide distributed network that allows data to be analyzed in a distributed manner. This network currently operates across several Toronto hospitals as well as across the province of Ontario.

Cardiogenomics: Dissecting the Functional Genomics of Heart Disease
Studies of the genomes of the developing heart and mature ventricular tissues reveal the genetic architecture of heart failure using expression microarrays, and population studies in collaboration with cardiologists at several Toronto-affiliated hospitals. CHIP researchers are investigating the "chain of command" of transcription factors using novel transcriptional techniques and high resolution analysis of transcription factor binding sites. Through these studies, possible novel drug sites could include including HDAC2/HDAC9 (http://chipgenomics.org) that allow the substrate's release of several "cardiac" marker genes: data have access the web.

Microarray Time Series
Temporal microarray experiments allow us to observe the genome in action. We have developed methods to reduce these experiments to uncover the behavioral similarities of genes (http://chipgenomics.org/asp.asp)

Asthma, Genomics and Innate Immunity
In collaboration with the Ontario Laboratory and Genomics of Asthma we are studying predictors of asthmatic disease using large scale association studies. In doing so, we have generated several genes with striking links.

- Shiu Chung Au, MEng
- Atul Butte, MD, MS
- Zhaohui Cai, MD, PhD
- Sara Dempster, PhD
- Sangeeta B. English, PhD
- Daniel Fleisher, BS
- Hamish Fraser, MBChB, MSc, MRCP
- Steven A. Greenberg, MD, MS
- Rolf Hanson, BS
- Albert Hong, BS
- Alvin Kho, PhD
- Isaac S. Kohane, MD, PhD
- Winston P. Kuo, DDS, MS
- Kenneth D. Mandl, MD, MPH
- Maung Zaw Min, MS
- Daniel J. Nigam, MD, MS
- Ashish Ningaonkar, MD, MTECH

- Karen L. Olson, PhD
- Peter J. Park, PhD
- Stephen Porter, MD, MPH
- Marco Ramoni, PhD
- Alberto Riva, PhD
- Ulrich Sax, PhD
- Asher Schachar, MD, MMSc, MS
- Ronald O. Sievert, BS, MSEE
- Bill Simons, MS
- Yao Sun, MD, PhD
- Sumil Saluja, MD
- Zoltan Szallasi, MD
- Jennifer Sun, MD
- Eric Tsang, PhD
- Alexander Turchin, MD
- Cecily J. Wolfe, PhD

Public Health / Biosurveillance



Outline

- Professional context and problems
- Why this problem is at the center of genomic medicine
- How **distributed** and standardized solutions can be used to address the problem.



Clinical Informatics vs. Bioinformatics

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

QuickTime
compressed
needed to see

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

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

Quick
Uncompre
needed to



Bioinformatics: A House Divided

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

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

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

bioinformatics "is a tool, not a
discipline, and tools have a way
of getting absorbed into
science." -L. Stein

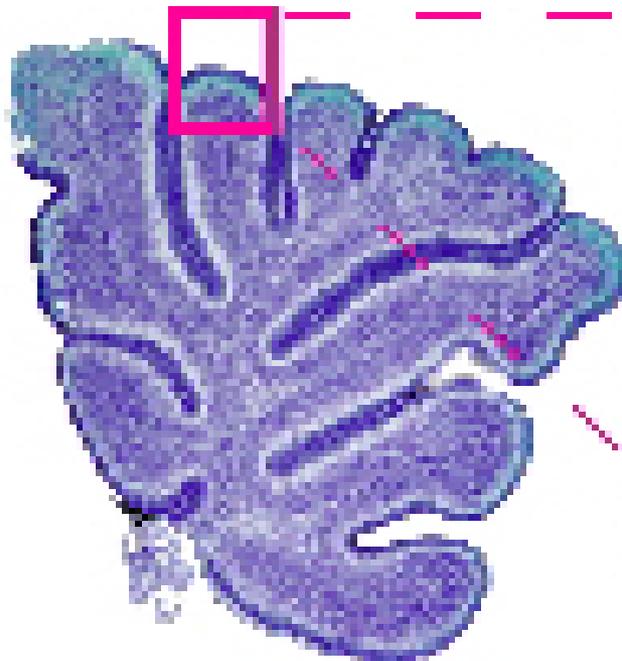


Combining Genomics from Different Species: Sequence and Expression Based Comparative Genomics

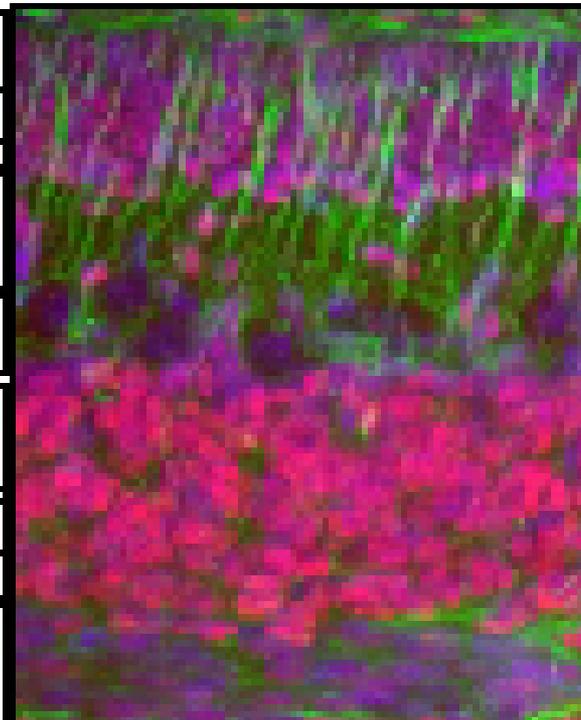
- The hypothesis that conservation of sequence across species has borne many fruit
- The harvest from the conservation of functional genomic dependencies (in expression or proteomics) has been less fruitful.
- We provide an example

Functional genomics of complex tissue

- Cerebellum has pivotal roles in the coordination of posture and locomotion
- Laminar organization of the cerebellar cortex has facilitated understanding its basic circuitry, functions and ontogeny

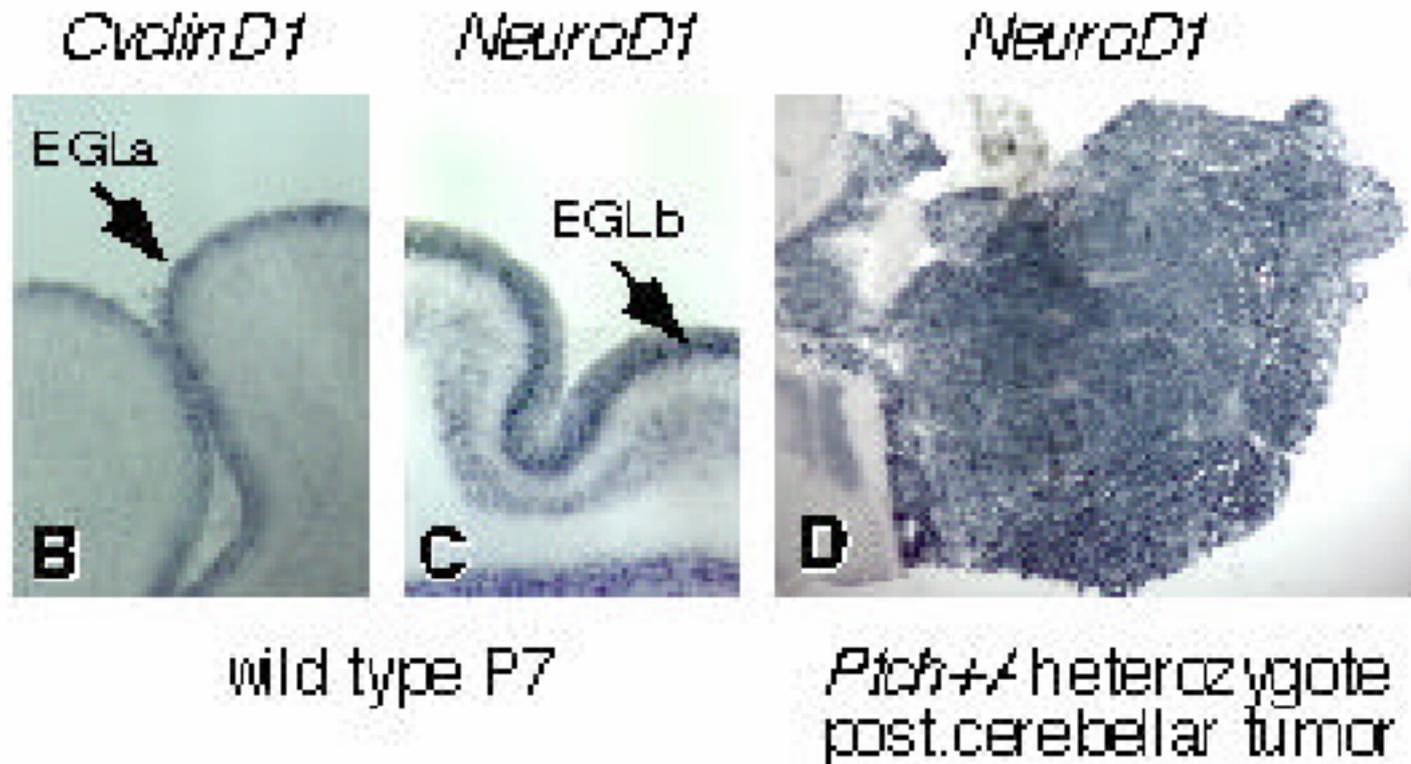


EGL a
EGL b
MOL
Purkinje
IGL a
IGL b
CWM



BrDU+, *Math1* ↑
BrDU-, *NeuroD2* ↑
Zic-1 (low)
Wnt3, *Shh*
Zic-1 (high)
NeuN
Olig1

Sonic Hedgehog (Shh), development and tumorigenesis



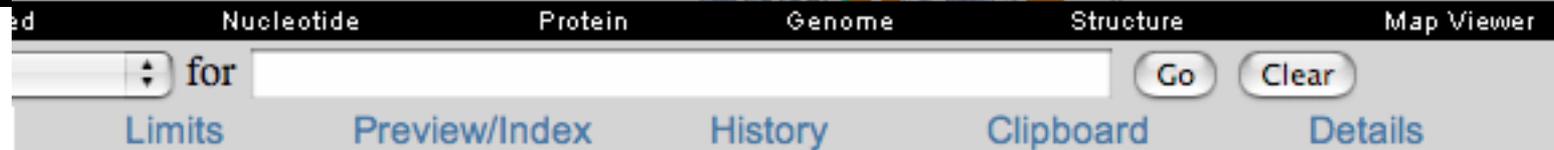
Zhao et al. PNAS 2002



Integrative Approach to Comparative Genomics

- We studied time-series of day 1 to day 60 of life of the mouse
 - ✓ Each time point, cerebellum measured with a microarray
 - ✓ PNAS Zhao et al. 2002
- How can we leverage this developmental view of the mouse?
 - ✓ Does it matter for clinical medicine?
- Human medulloblastoma microarray data
 - ✓ Pomeroy et al, Nature 2002
- Principal Component Analysis to find the main sources of variance in the developmental time series

HomoloGene



B HomoloGene is a system for automated detection of homologs among the annotated genes of several completely sequenced eukaryotic genomes.

HomoloGene Release Statistics

Initial numbers of genes from complete genomes, numbers of genes placed in a homology group, and the numbers of groups for each species

Species	HomoloGene Build 36		HomoloGene groups
	Number of genes Input	Grouped	
H.sapiens	22,827	18,055	16,782
M.musculus	24,019	19,996	18,036
R.norvegicus	20,913	17,429	16,042
D.melanogaster	12,918	8,717	7,683
A.gambiae	12,012	8,543	7,577
C.elegans	19,109	6,502	5,260
S.pombe	4,947	3,625	3,359
S.cerevisiae	5,863	3,612	3,146
N.crassa	10,079	6,156	6,049
M.grisea	11,109	6,307	6,028
A.thaliana	26,281	8,022	4,791
P.falciparum	5,222	1,770	1,589

Last updated on: 05/25/2004

Related Resources

Entrez Genome

A collection of complete genome sequences that includes more than 1000 viruses and over 100 microbes

- [Archaea](#)
- [Bacteria](#)
- [Eukaryota](#)
- [Viruses](#)

Tax Plot

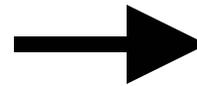
Three-way view of genome similarities



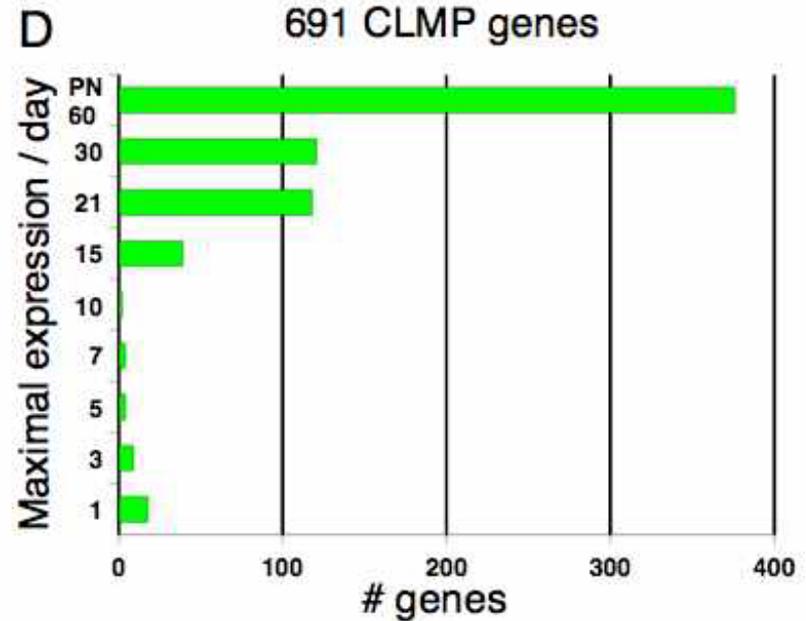
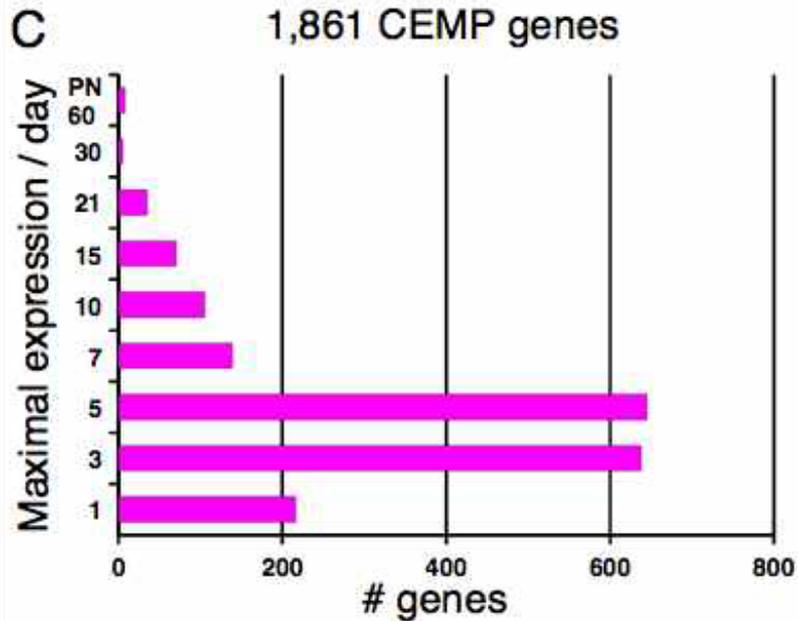
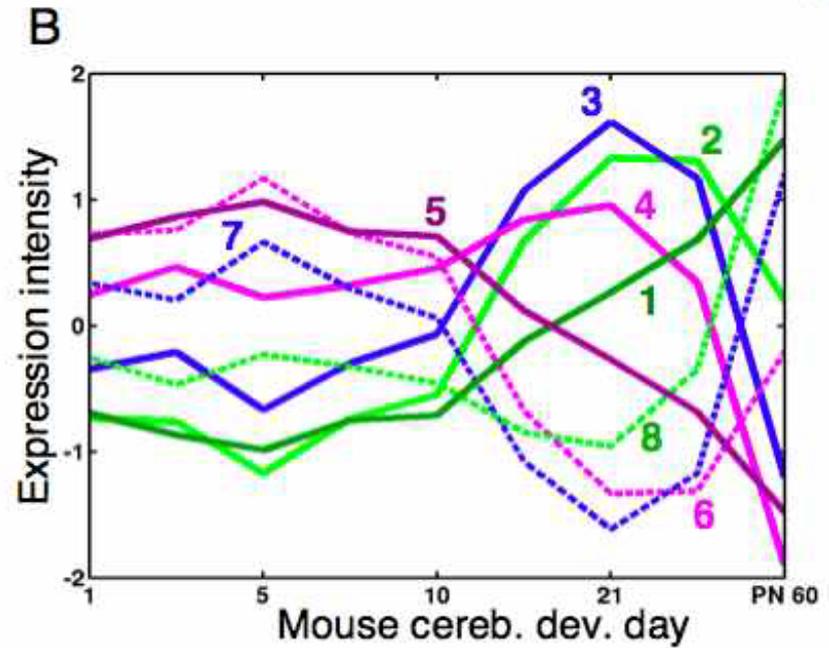
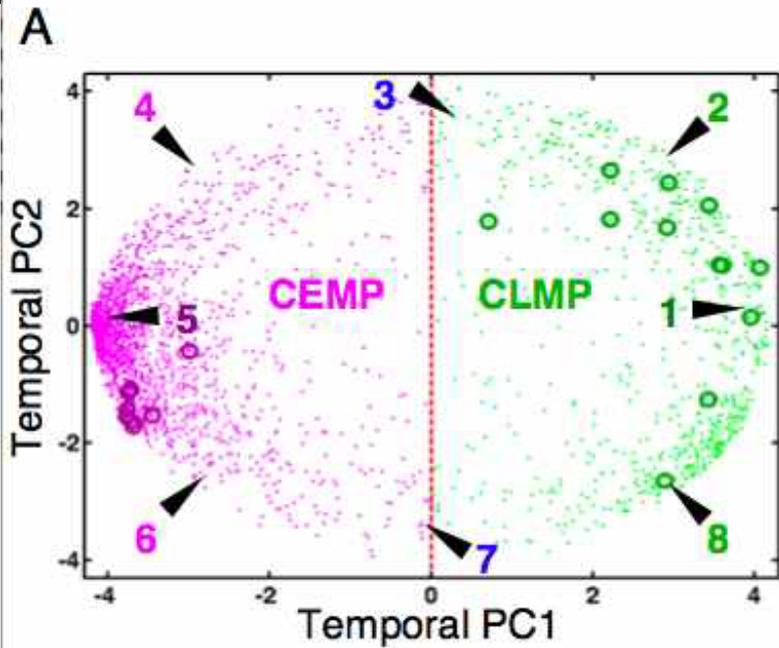


Principal Components

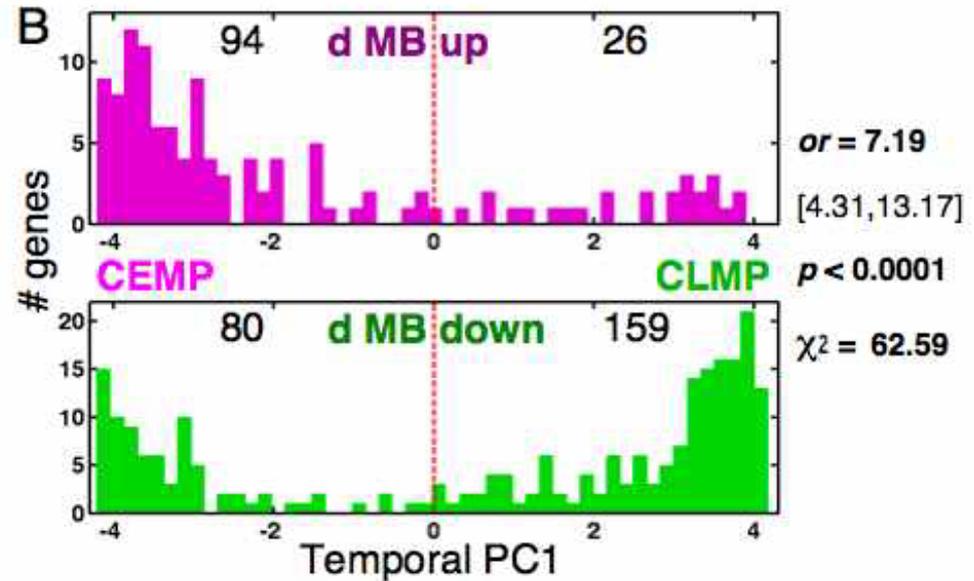
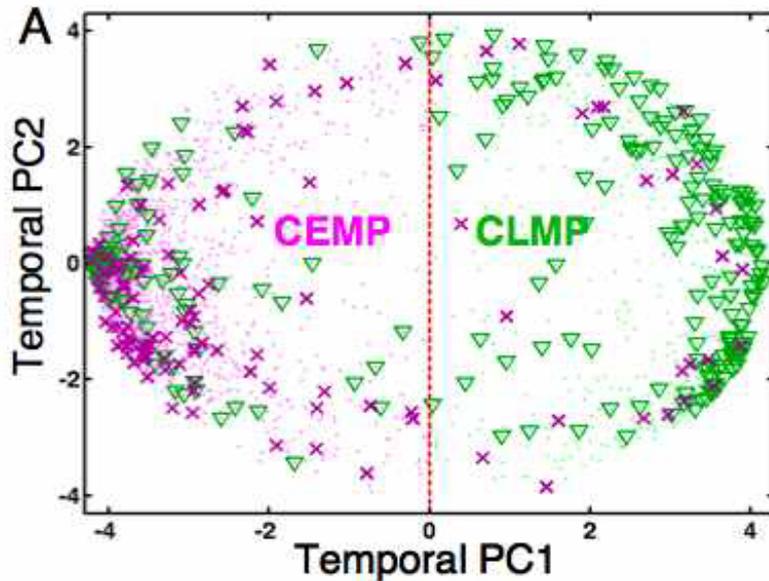
QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.



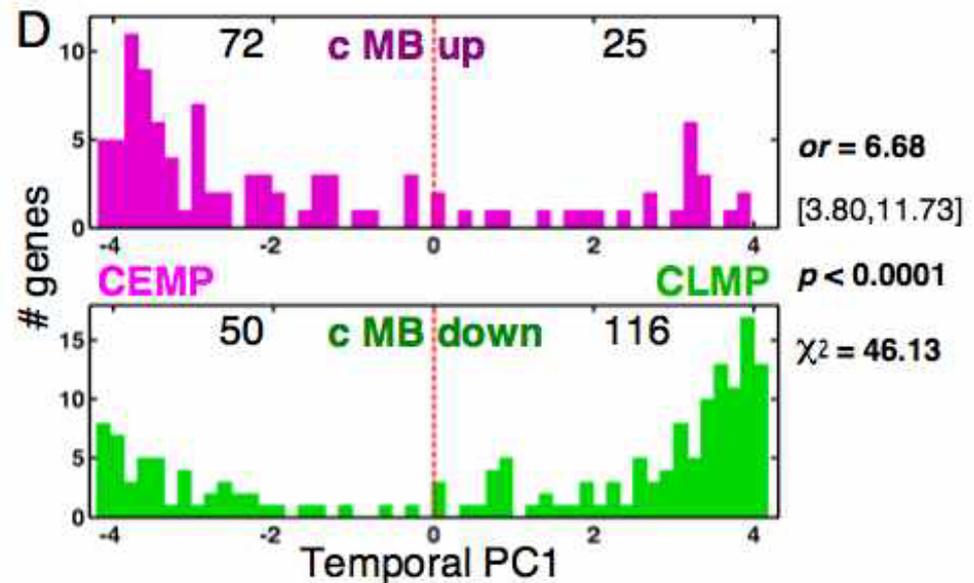
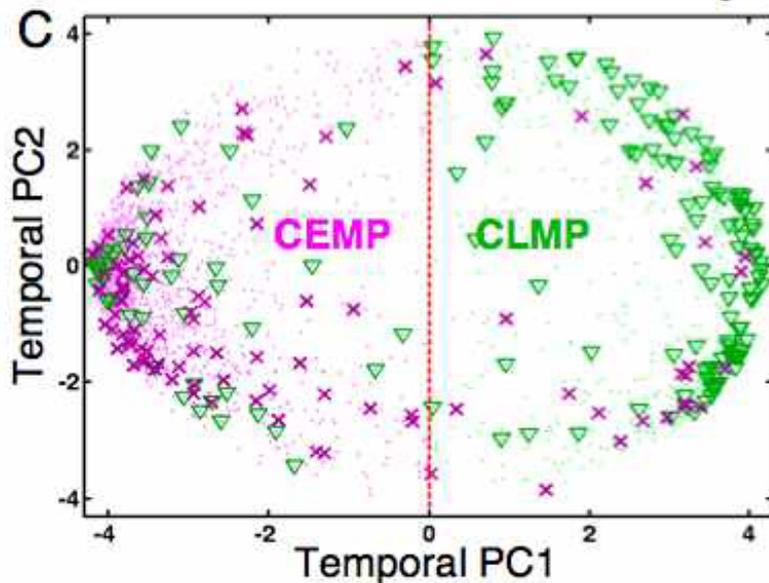
QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.



Human *d* MB genes in mouse cereb. dev.

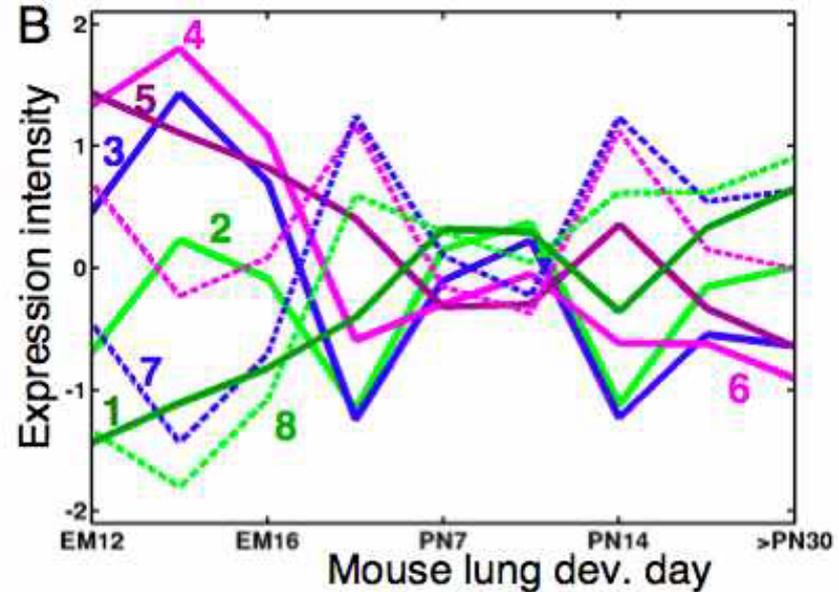
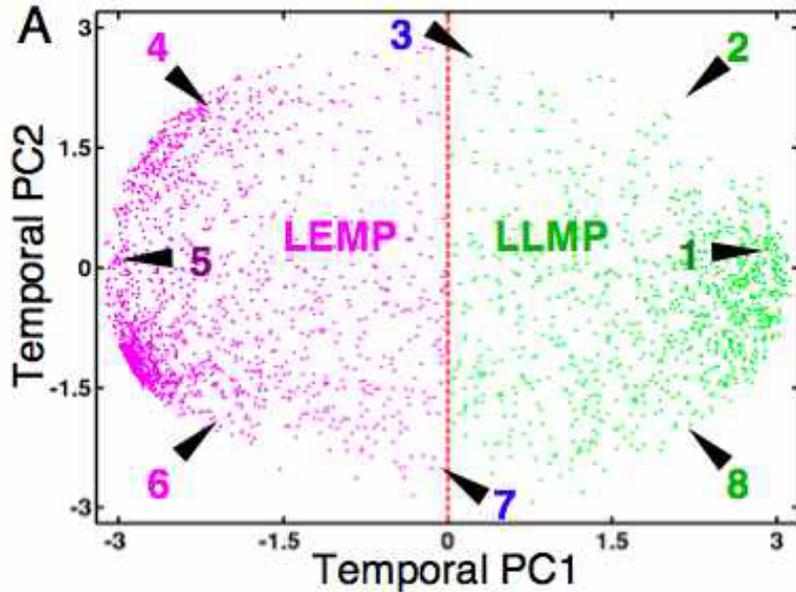


Human *c* MB genes in mouse cereb. dev.

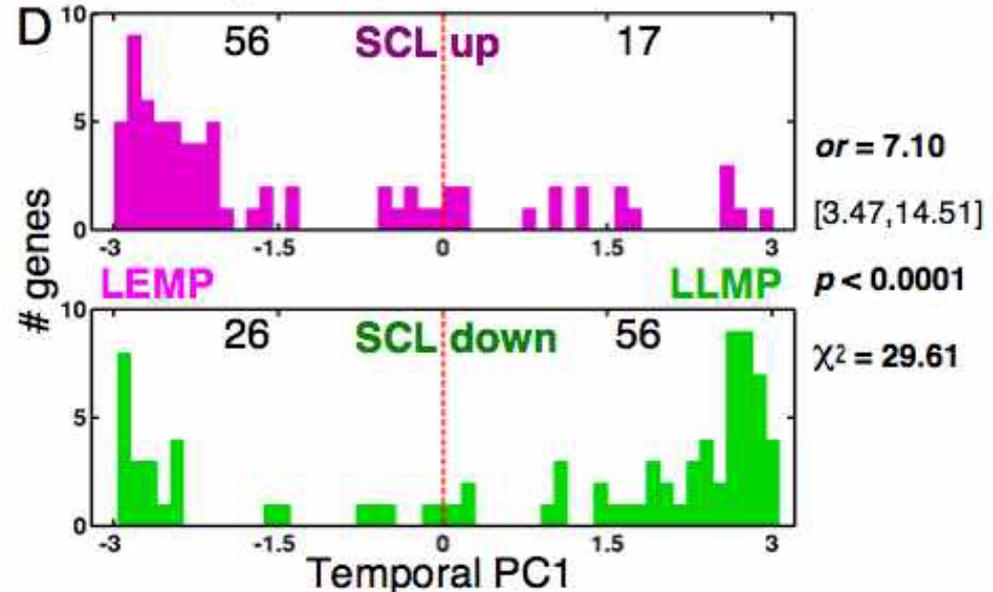
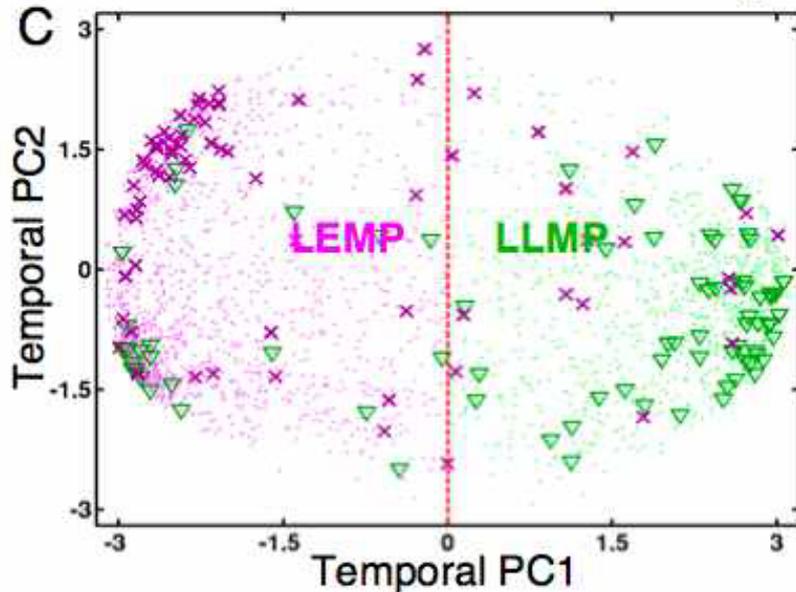




Does this pattern generalize to other tissues?



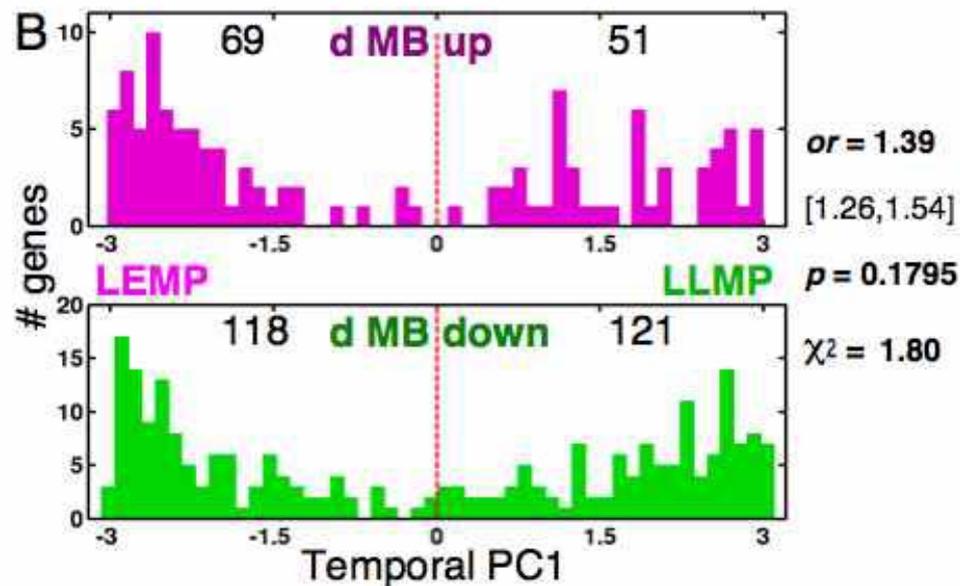
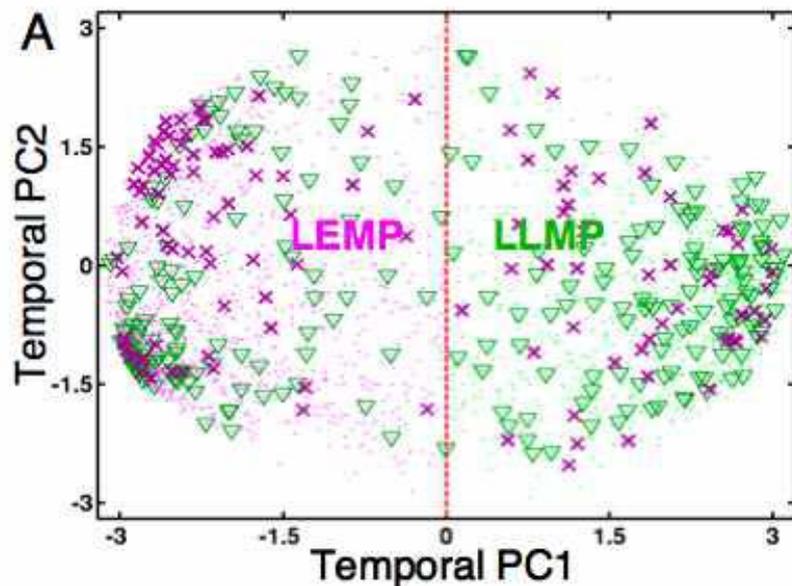
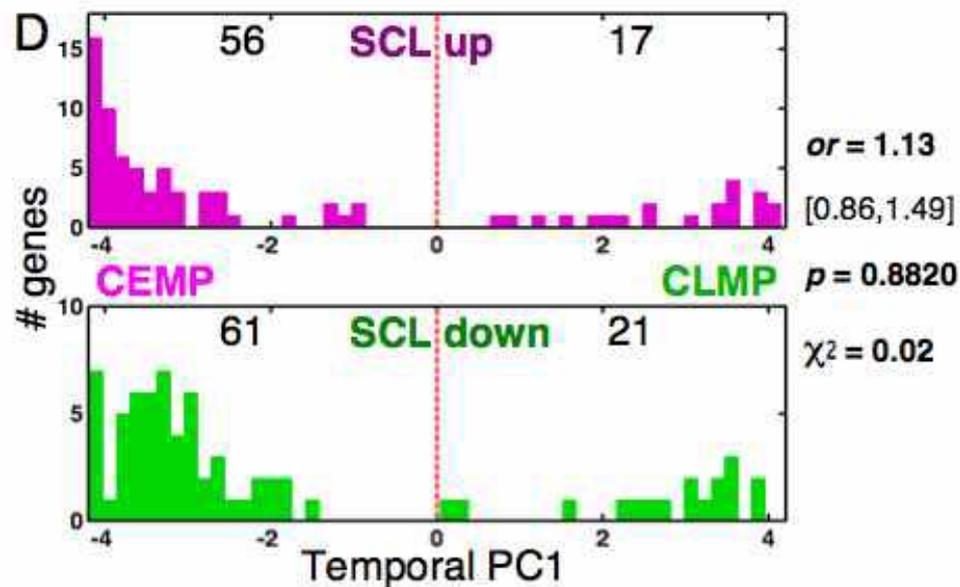
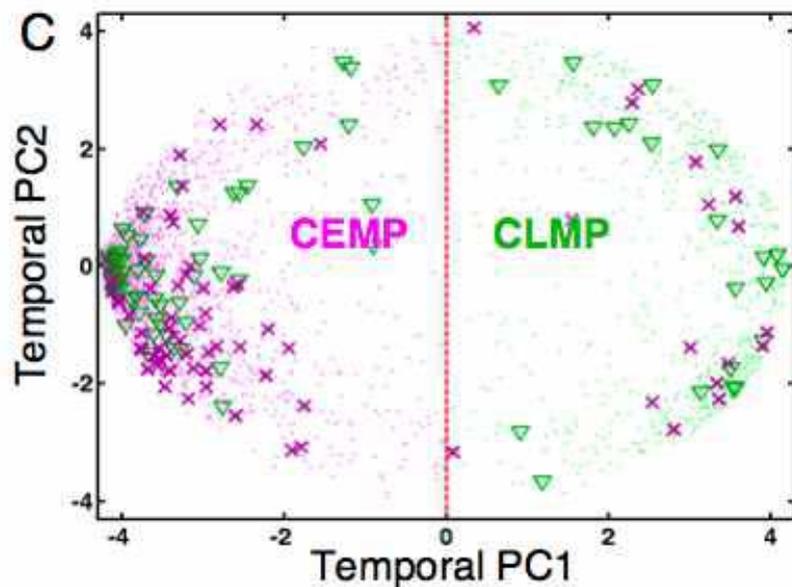
Human SCL genes in mouse lung dev.





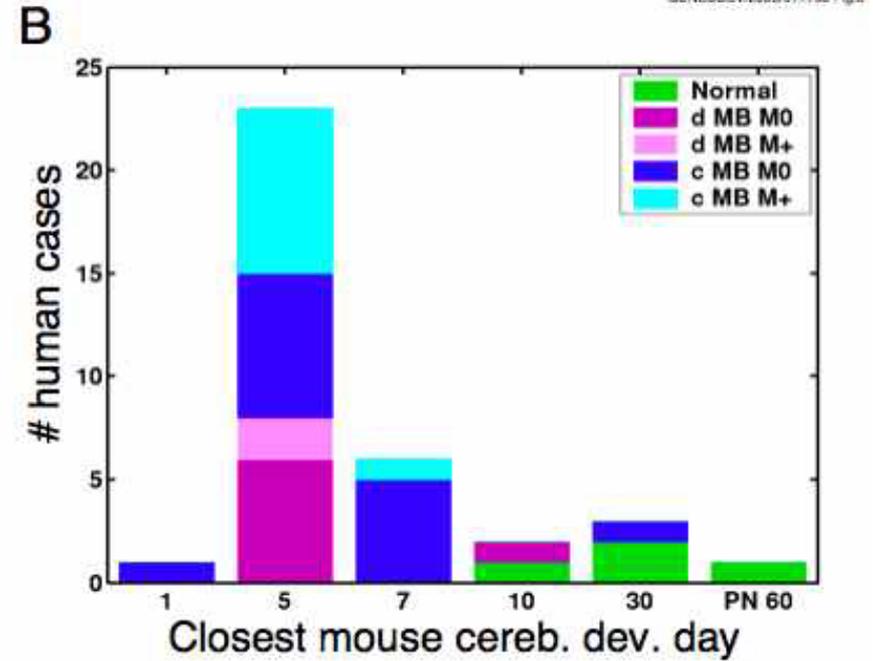
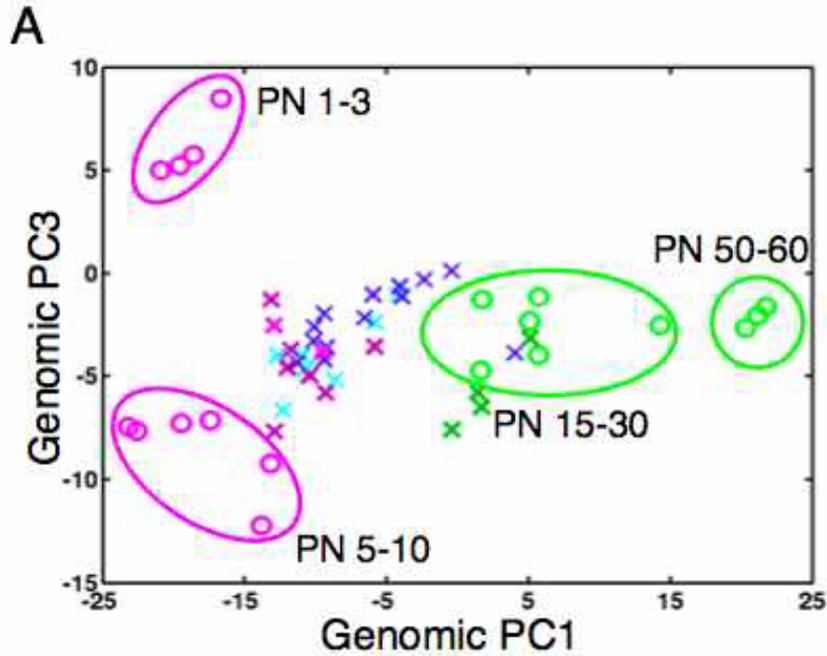
Is this phenomenon, merely reporting proliferation?

Or, does it reflect the tissue-specific developmental program?

Human *d* MB genes in mouse lung dev.Human *SCL* genes in mouse cereb. dev.



What about a macro view?





Interim Conclusion (I)

- Revisiting an old idea with the quantitation and precision of the genomic era.
- Lobstein (1829) and Cohnheim (1887) were amongst the first to theorize similarities between human embryogenesis and the biology of cancer cells .
- The brain tumor classification system of Bailey and Cushing (1926),
 - ✓ from which modern taxonomies derive,
 - ✓ emphasizes the histologic resemblance to cells of the developing central nervous system

Kho et al. *Genes and Development*, March 2004



Conclusion (II)

- Projecting human solid tumors against a background of mouse models provides insight
 - ✓ Into diagnostic staging
 - ✓ Into biological process that characterize the tumors
 - ✓ May have biologically ground prognostic value
- We demonstrate that this may be generalizable to many other systems of human disease
- Demonstrate that comparative **functional** genomic data sets **DEPENDS** crucially: on a **COMMON VOCABULARY** and cross-species mapping.



Integrative Biology

- With the multiplicity of data sets, increasing efforts at integration to learn more about the underlying biology.
- Integration however present fundamental methodological problems.



Example: Alizadeh 2000

Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling

Ash A. Alizadeh^{1,2}, Michael B. Eisen^{2,3,4}, R. Eric Davis⁵, Chi Ma⁵, Izidore S. Lossos⁵, Andreas Rosenwald⁵, Jennifer C. Boldrick¹, Hajeer Sabet⁵, Truc Tran⁵, Xin Yu⁵, John L. Powell⁷, Liming Yang⁷, Gerald E. Marti⁸, Troy Moore⁹, James Hudson Jr.⁹, Lisheng Lu¹⁰, David B. Lewis¹⁰, Robert Tibshirani¹¹, Gavin Sherlock⁴, Wing C. Chan¹², Timothy C. Greiner¹², Dennis D. Weisenburger¹², James O. Armitage¹³, Roger Warnke¹⁴, Ronald Levy⁶, Wyndham Wilson¹⁵, Michael R. Grever¹⁶, John C. Byrd¹⁷, David Botstein⁴, Patrick O. Brown^{1,18} & Louis M. Staudt⁵

Departments of ¹Biochemistry, ³Genetics, ¹⁴Pathology, ⁶Medicine, ¹⁰Pediatrics and ¹¹Health Research & Policy and Statistics, and ¹⁸Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA

⁵Metabolism Branch, Division of Clinical Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

⁷Bioinformatics and Molecular Analysis Section, CBEL, CIT, NIH, Bethesda, Maryland 20892, USA

⁸CBER, FDA, Bethesda, Maryland 20892, USA

⁹Research Genetics, Huntsville, Alabama 35801, USA

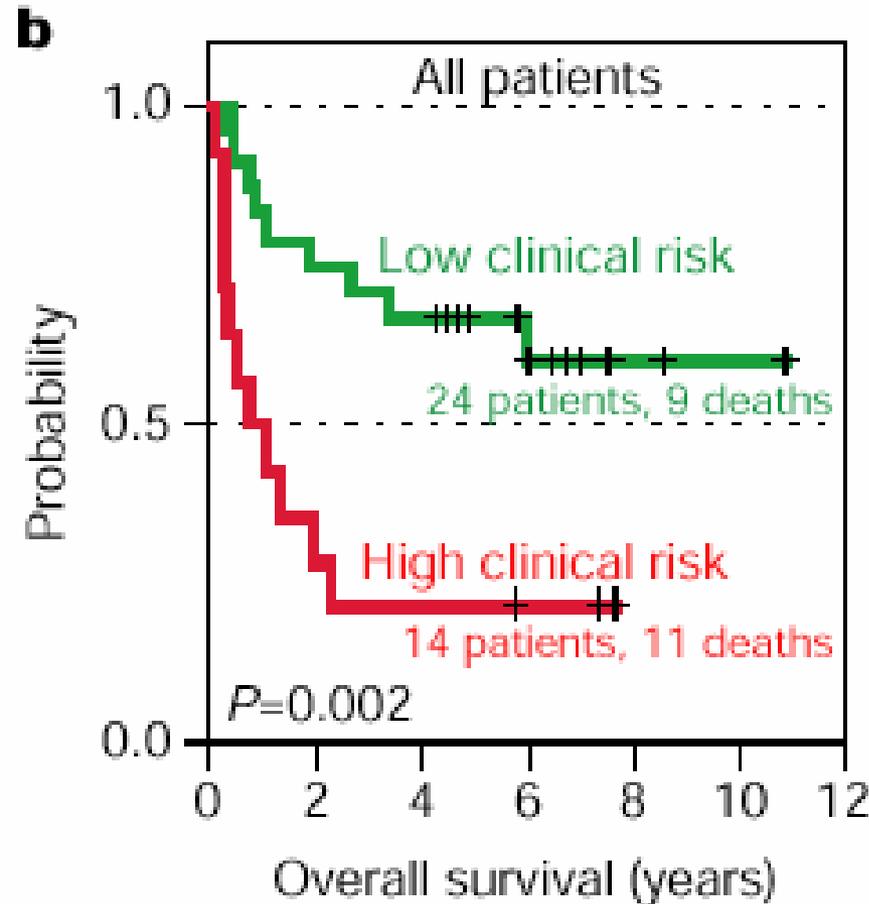
Departments of ¹²Pathology and Microbiology, and ¹³Internal Medicine, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA

¹⁵Medicine Branch, Division of Clinical Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

¹⁶Johns Hopkins Oncology Center, Johns Hopkins School of Medicine, Baltimore, Maryland 21287, USA

¹⁷Walter Reed Army Medical Center, Washington, DC 20307, USA

²These authors contributed equally to this work



Alizadeh et al.



Example: Shipp 2002

Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning

MARGARET A. SHIPP¹, KEN N. ROSS², PABLO TAMAYO², ANDREW P. WENG³, JEFFERY L. KUTOK³, RICARDO C.T. AGUIAR¹, MICHELLE GAASSENBEK⁴, MICHAEL ANGELO², MICHAEL REICH⁵, GERALDINE S. PINKUS⁵, TANE S. RAY⁶, MARGARET A. KOVAL¹, KIM W. LAST⁴, ANDREW NORTON⁷, T. ANDREW LISTER⁴, JILL MESIROV², DONNA S. NEUBERG¹, ERIC S. LANDER^{2,7}, JON C. ASTER⁸ & TODD R. GOLUB^{1,2}

¹Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA

²Whitehead Institute for Biomedical Research/Massachusetts Institute of Technology Center for Genome Research, Cambridge, Massachusetts, USA

³Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

⁴ICRF Medical Oncology Unit and ⁵Pathology Unit, St. Bartholomew's Hospital, London, UK

⁶Department of Computer Science, Maths and Physics, University of West Indies, Bridgetown, Barbados

⁷Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

K.N.R. and P.T. contributed equally to this study.

Correspondence should be addressed to M.A.S.: email: margaret_shipp@dfci.harvard.edu, or T.R.G.: email: golub@genome.wi.mit.edu

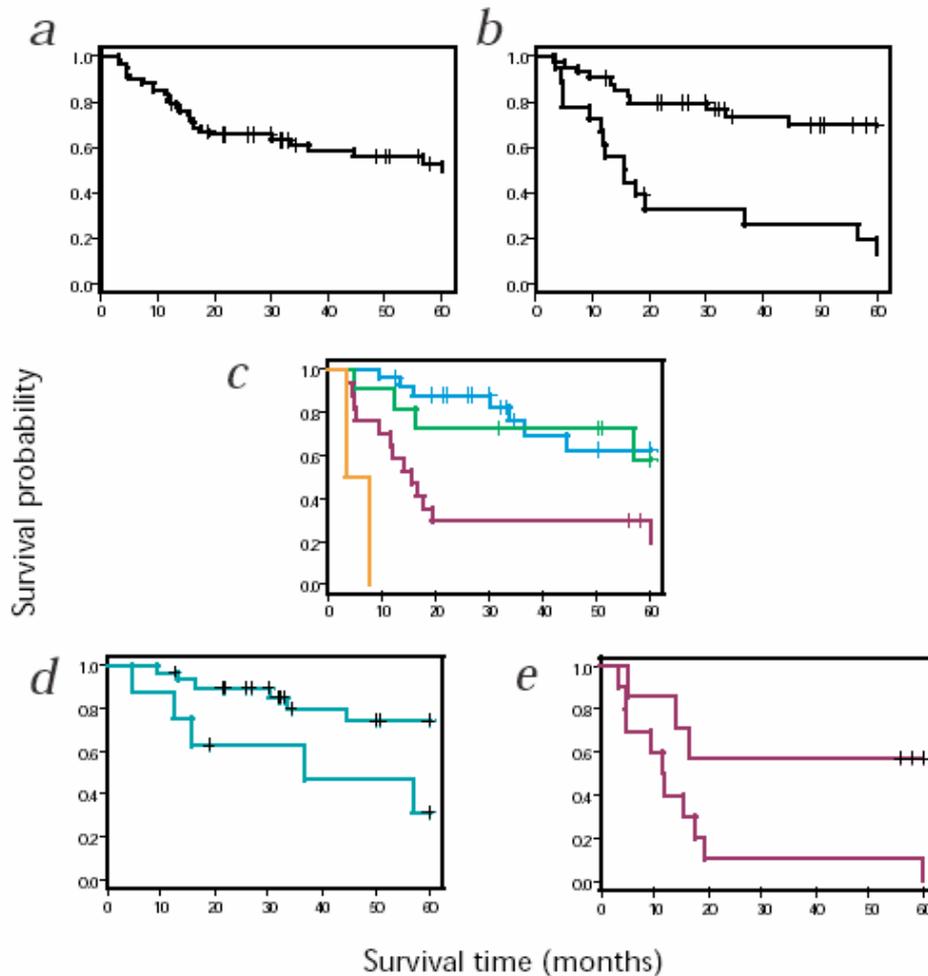
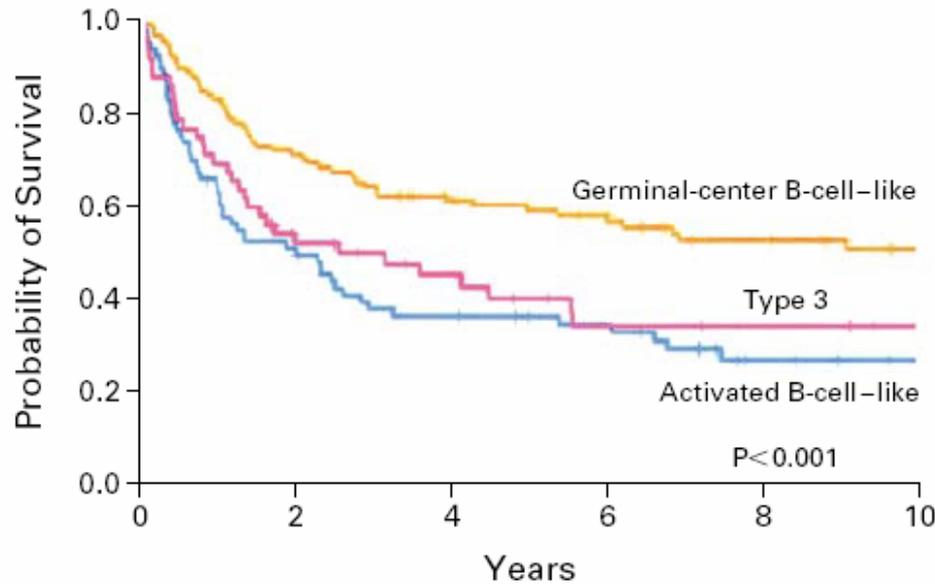


Fig. 4 Overall survival predictions for DLBCL study patients. **a**, 5-year OS for the entire study group. 33 of 58 DLBCL study patients remained alive at a median of a 58-month follow-up. The predicted 5-year OS for the group as a whole was 54%. **b**, 5-year OS for favorable and unfavorable risk groups defined by the 13-gene model (70% versus 12%, $P = 0.00004$). Top line, cured; bottom, fatal/refractory. **c**, 5-year OS for patients in L-risk (green line), LI-risk (blue line), HI-risk (red line) and H-risk (orange line) categories as defined by the IPI: L, 26 pts; LI, 11 pts; HI, 17 pts; H, 2 pts. **d**, 5-year OS for combined L/LI-risk patients with favorable or unfavorable disease as defined by the molecular model (75% versus 32%, nominal $P = 0.02$) Top line, cured; bottom, fatal/refractory. **e**, 5-year OS for HI-risk patients with favorable or unfavorable disease as defined by the molecular model (57% versus 0%; nominal $P = 0.02$). Top line, cured; bottom, fatal/refractory.



Example: Rosenwald

C



No. AT RISK

Germinal-center B-cell-like	115	81	60	46	32	19
Type 3	52	24	18	10	8	5
Activated B-cell-like	73	35	23			

The New England
Journal of Medicine

Copyright © 2002 by the Massachusetts Medical Society

VOLUME 346

JUNE 20, 2002

NUMBER 25



THE USE OF MOLECULAR PROFILING TO PREDICT SURVIVAL
AFTER CHEMOTHERAPY FOR DIFFUSE LARGE-B-CELL LYMPHOMA

ANDREAS ROSENWALD, M.D., GEORGE WRIGHT, PH.D., WING C. CHAN, M.D., JOSEPH M. CONNORS, M.D.,
ELIAS CAMPO, M.D., RICHARD I. FISHER, M.D., RANDY D. GASCOYNE, M.D., H. KONRAD MULLER-HERMELINK, M.D.,
ERLEND B. SMELAND, M.D., PH.D., AND LOUIS M. STAUDT, M.D., PH.D.,
FOR THE LYMPHOMA/LEUKEMIA MOLECULAR PROFILING PROJECT



Interim Conclusion

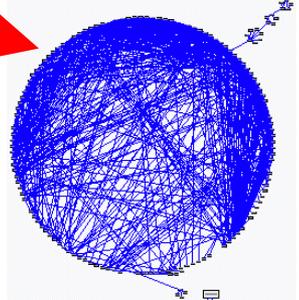
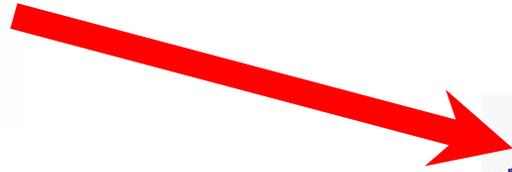
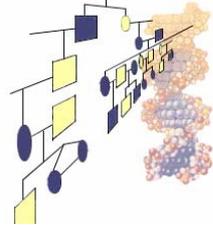
- Less than 100 patients is small in any clinical study
- With thousands of genomic variables, 100's is even smaller with respect to the dangers of:
 - ✓ Over-fitting
 - ✓ Multiple-hypothesis testing
- Many methodological ills can be forgiven by large numbers of cases



So, how do we get enough samples?



Family Hx



Address	Contact Number	SEARCH	Program notes	To do list
Phone	Emergency text	Feedback	Message	Template notes
Gender	Communication	Link	Feedback	Notes
				Chat/Share

Prescriptions	Drugs	Strength	Frequency	Route	Start Date	End Date	Refill	Notes
Aspirin	81 mg	qd		PO	11/15/2000	11/15/2000	1	
Acetaminophen	325 mg	prn		PO	11/15/2000	11/15/2000	1	
Amoxicillin	500 mg	tid		PO	11/15/2000	11/15/2000	1	

Medication	Start Date	End Date	Refill	Notes
Aspirin	11/15/2000	11/15/2000	1	
Acetaminophen	11/15/2000	11/15/2000	1	
Amoxicillin	11/15/2000	11/15/2000	1	

Consent

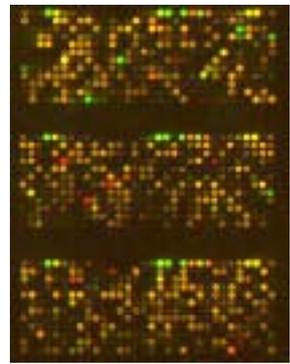
Annotation

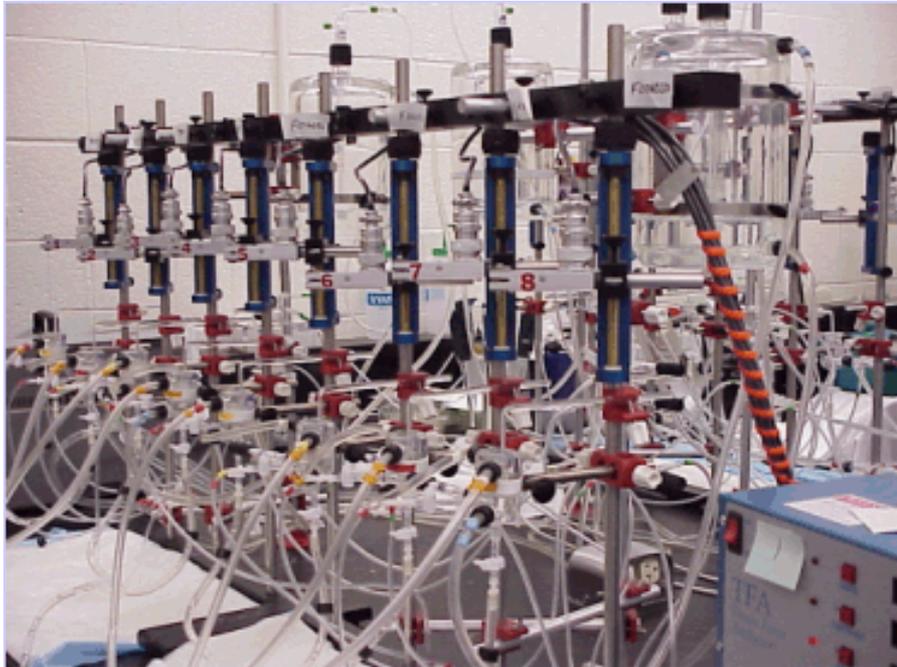


Banking



Genomics





High-throughput phenotyping at MCW





Not Registered? [ENTER HERE](#) ▶

[REQUEST ACCOUNT](#) ▶

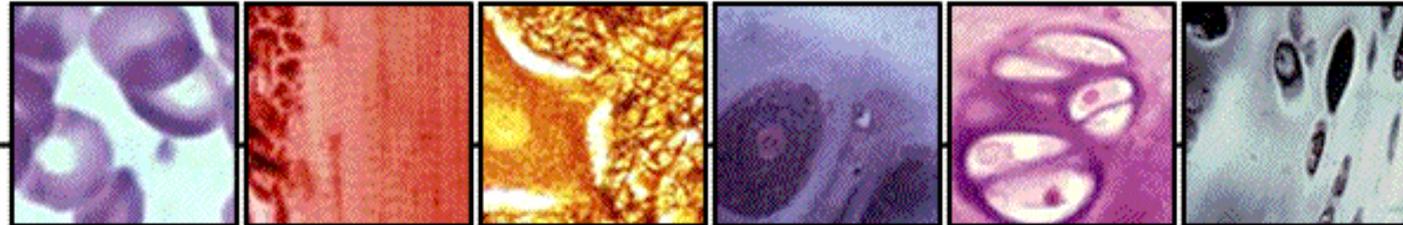
username:

password:

[REGISTERED USER LOGIN](#) ▶

National Institutes of Health
National Cancer Institute

Brigham & Women's Hospital
Beth Israel Deaconess Medical Center
Cedars-Sinai Medical Center
Children's Hospital
Dana-Farber Cancer Institute
Massachusetts General Hospital
Olive View Medical Center
UCLA Medical Center
VA Greater LA Healthcare System
University of Pittsburgh Medical Center

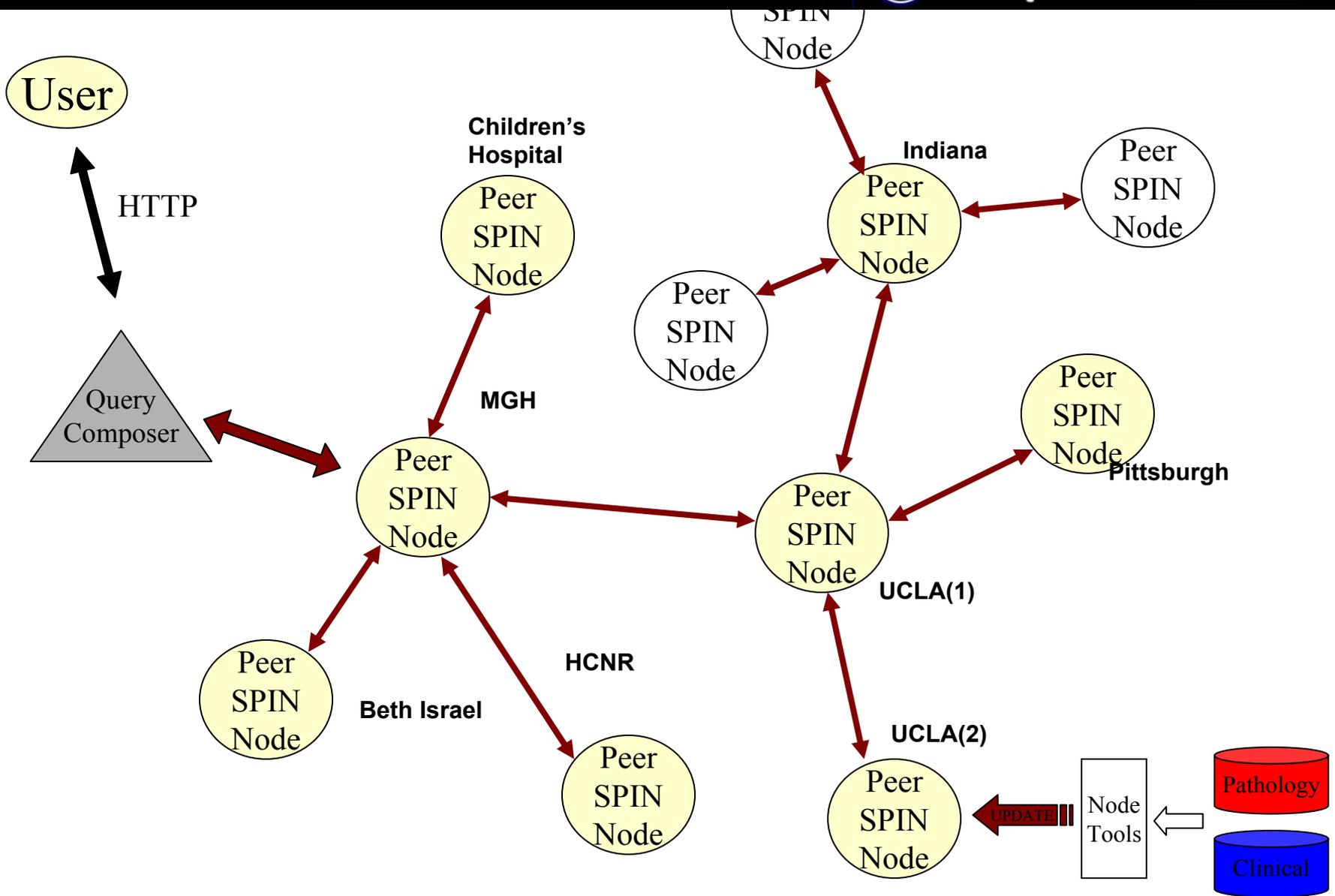


Shared *p*athology *i*nformatics *n*etwork

The objective of this initiative is to use state-of-the-art informatics techniques to establish an Internet-based virtual database that will allow investigators to locate appropriate human tissue specimens for their research.

The SPIN software will allow approved researchers access to data that describe archived tissue specimens across multiple institutions while still allowing those institutions to maintain local control of the data. The need for this capability has been fueled by the growing use of tissues, diagnostic specimens, and their related clinical data in modern biomedical research.

[\[terms & conditions\]](#)





SEARCH **ADVANCED SEARCH** **RESULTS** **LOGIN** [home] [about]

diagnosis ▼ ▲

Text: Code(s):

gender ▼ ▲

Female Male Transgender Unknown

age at specimen collection ▼ ▲

BETWEEN year(s) ▼ AND year(s) ▼

topology ▼ ▲

date of specimen collection ▼ ▲

results format ▼ ▲

search name:

search type: CHIRPS Only SPIN Network

statistics: ▼

detailed table fields:

Age at specimen collection

Date of specimen collection

Gender

sort by

clear search ▶

[\[terms & conditions\]](#)

National Institutes of Health
National Cancer Institute

Brigham & Women's Hospital
Beth Israel Deaconess Medical Center
Cedars-Sinai Medical Center
Children's Hospital
Dana-Farber Cancer Institute
Massachusetts General Hospital



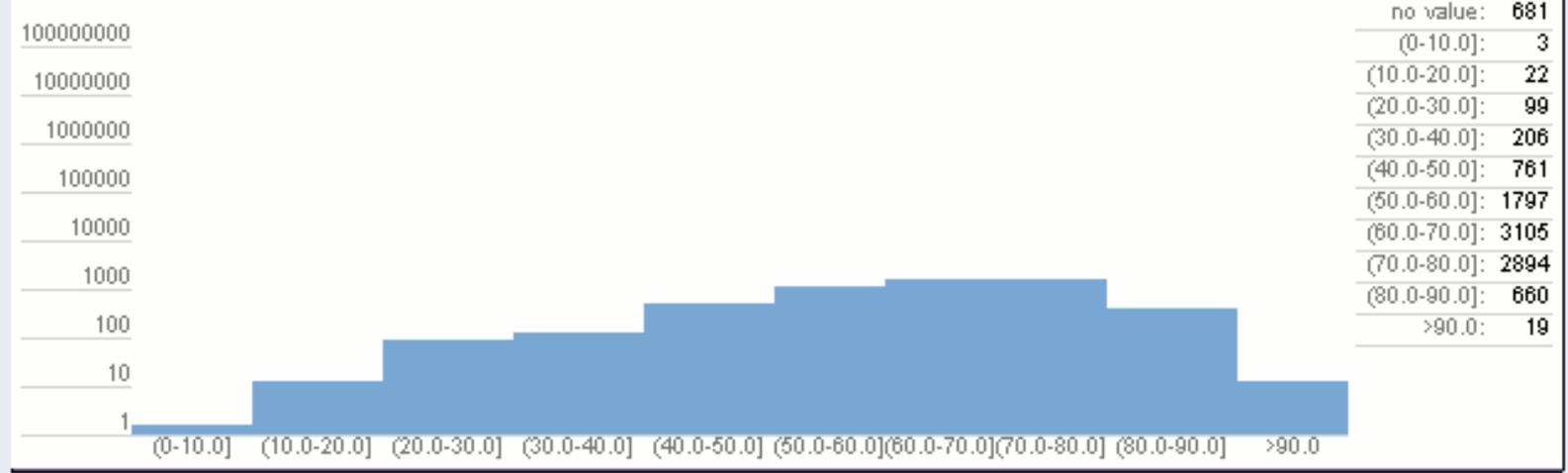
Select ALL

delete selected results

RESULTS SUBQUERY NODES Fetching. [red] [down] [up]

name: records: 10247 nodes: 7 sent: 2004-05-27 23:49:38

age at sample collection



RESULTS SUBQUERY NODES [down] [up]

name: records: 0 nodes: 6 sent: 2004-05-27 23:48:55

RESULTS SUBQUERY NODES [down] [up]

name: detail records: 990 nodes: 7 sent: 2004-05-27 23:45:23

RESULTS SUBQUERY NODES [down] [up]

name: records: 0 nodes: 7 sent: 2004-05-27 23:43:57

delete selected results



Shared *pathology* informatics *network*

[SEARCH](#)[ADVANCED SEARCH](#)[RESULTS](#)[LOGIN](#)[\[home\]](#) [\[about\]](#) Select ALL[delete selected results](#) [RESULTS](#)[SUBQUERY](#)[NODES](#)**Fetching**

name: records: 990 nodes: 7 sent: 2004-05-27 23:32:25

nodes that responded to your query:

- (1) CHIRPS - Beth Israel Deaconess Medical Center
- (1) CHIRPS - Brigham & Women's Hospital
- (1) CHIRPS - Children's Hospital Boston
- (1) CHIRPS - Massachusetts General Hospital
- (2) CHIRPS - University of California at Los Angeles Medical Center
- (1) Other - unidentifiable

[delete selected results](#)



Shared Pathology Informatics Network



SEARCH

ADVANCED SEARCH

RESULTS

LOGIN

[\[home\]](#) [\[about\]](#)

diagnosis ▼ ▲

Text: Code(s):

gender ▼ ▲

Female Male Transgender Unknown

age at specimen collection ▼ ▲

BETWEEN year(s) ▼ AND year(s) ▼

topology ▼ ▲

date of specimen collection ▼ ▲

results format ▼ ▲

<p>search name:</p> <input type="text" value="detail"/> <p>search type:</p> <input type="radio"/> CHIRPS Only <input checked="" type="radio"/> SPIN Network	<p>statistics:</p> <input type="text" value="bin on age"/> ▼
	<p>detailed table fields:</p> <input checked="" type="checkbox"/> Age at specimen collection <input checked="" type="checkbox"/> Date of specimen collection <input type="checkbox"/> Gender
	<input type="text" value="date of specimen collection"/> ▼



SEARCH

ADVANCED SEARCH

RESULTS

LOGIN

[home] [about]

Select ALL

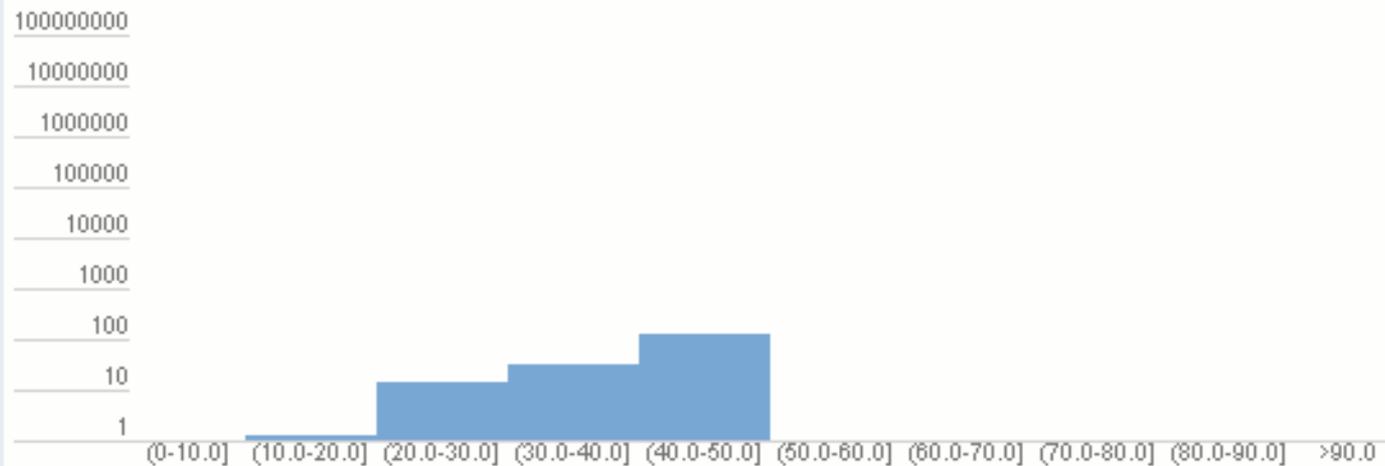
delete selected results

RESULTS SUBQUERY NODES Fetching



name: detail records: 301 nodes: 6 sent: 2004-05-27 23:54:58

age at sample collection



no value:	Count
(0-10.0]:	0
(10.0-20.0]:	2
(20.0-30.0]:	24
(30.0-40.0]:	57
(40.0-50.0]:	218
(50.0-60.0]:	0
(60.0-70.0]:	0
(70.0-80.0]:	0
(80.0-90.0]:	0
>90.0:	0

detail

Age at Specimen Collection	Tissue Acquisition Date
50	1988-03-04
50	1989-08-10
46	1989-09-28
45	1989-10-26
46	1989-11-30
48	1990-01-11
34	1990-03-09
36	1990-04-11



Interim Conclusion

- Tasteful delegation of control and access enables large data sharing at the scale required of the genomic age.
- A **COMMON VOCABULARY** across our clinical systems is essential.
- A common set of **STANDARDIZED PROTOCOLS** across distributed systems is the only way to scale to the national level.
- Taking seriously the building of **LEVELS of ABSTRACTION** and **ABSTRACTION BARRIERS** is the key to large system construction



But if medical records are incomplete, how do we get the entire patient history/phenotype?

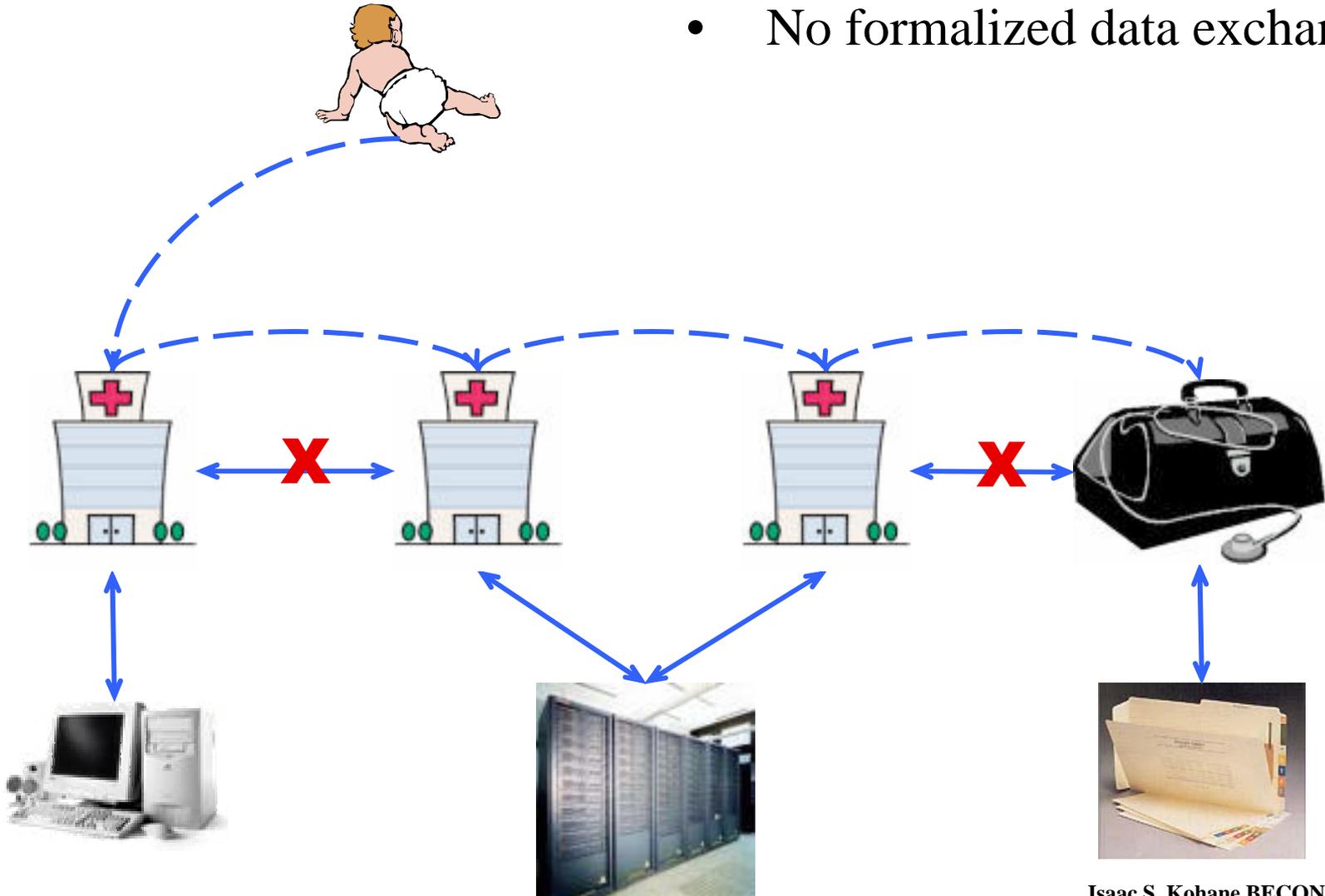
- ✓ Daily medications
- ✓ Exercise level
- ✓ Adverse events with over the counter medications
- ✓ Absence from work
- ✓ ...
- Given persistent failure to get institutional buy-in
- Can we harness the patient?

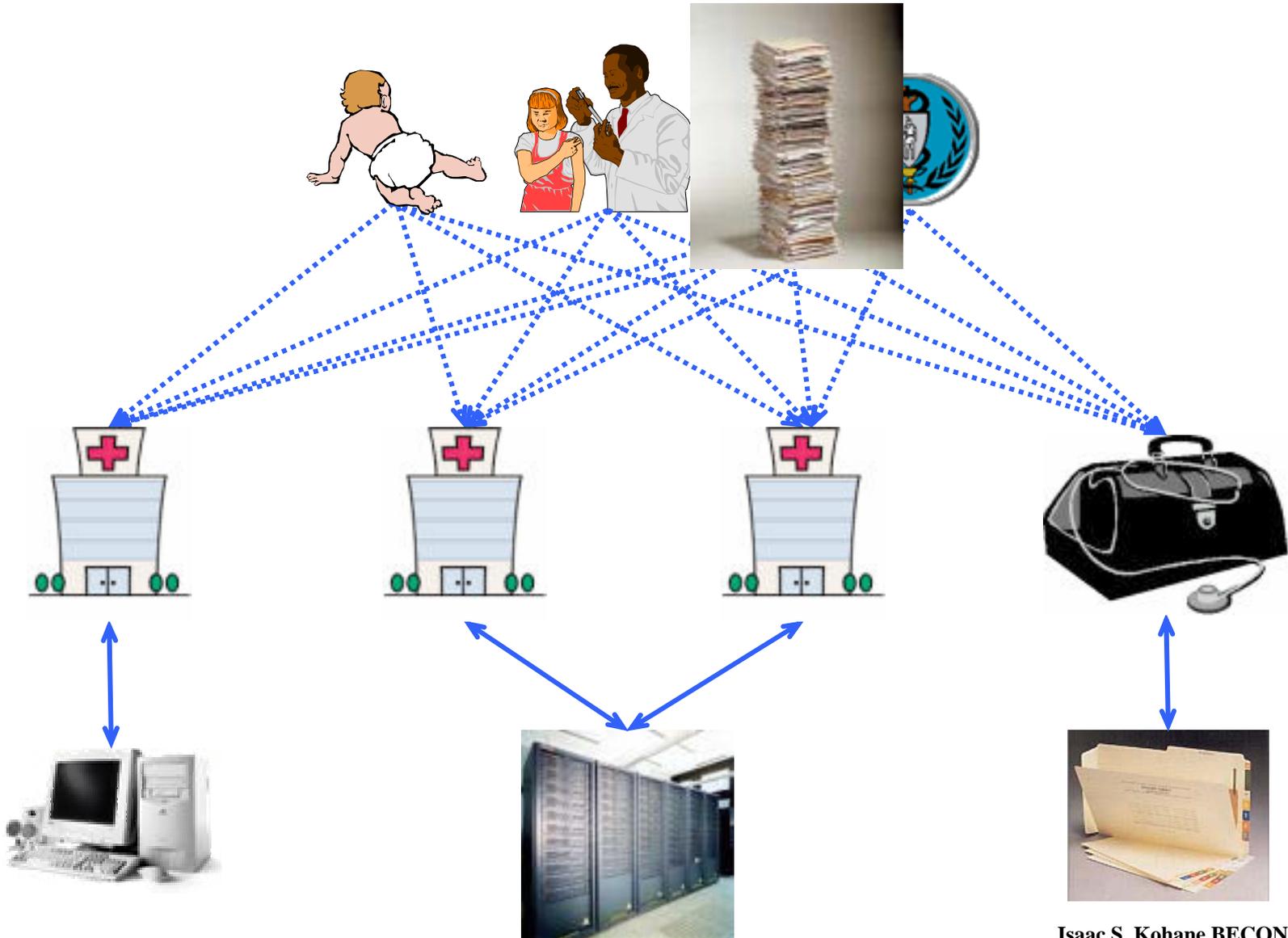


The Problem

- Medical records are fragmented across multiple institutions
- There's no unified view of a patient's record
- Patients have difficulty accessing their medical information

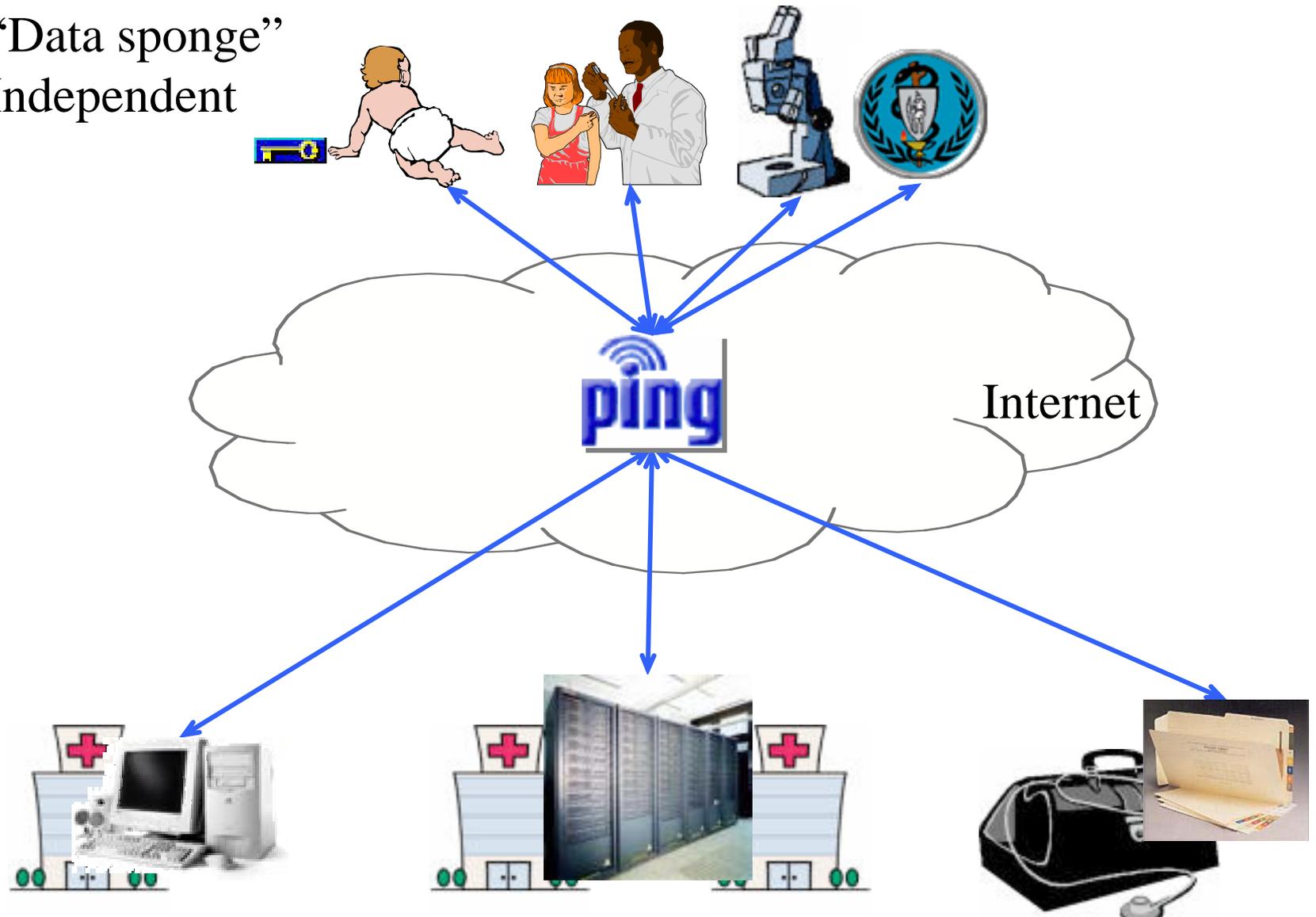
- Highly mobile patients
- No formalized data exchange

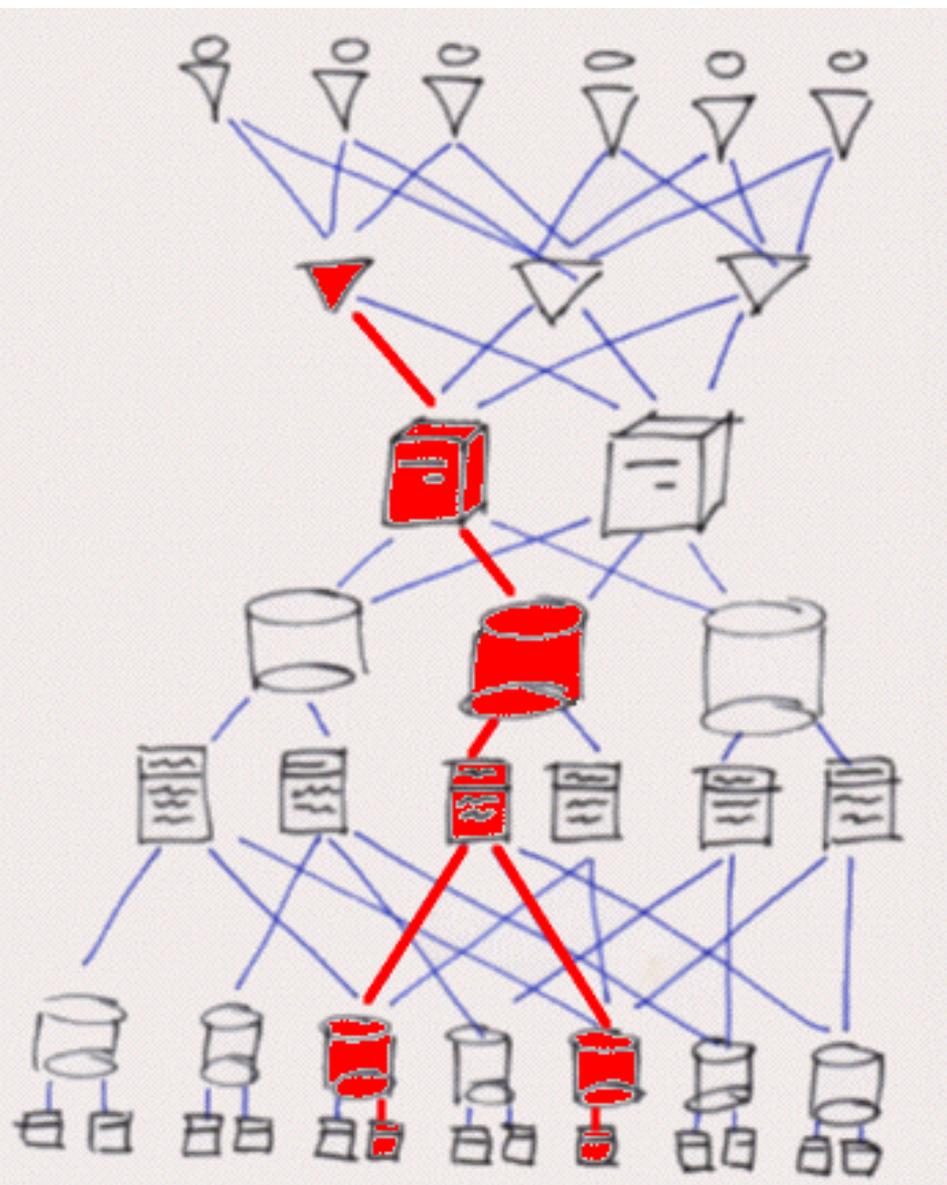






- “Data sponge”
- Independent





- Clients access the PING server, which in turn, accesses the patient's record
- The PING server authenticates the agent and can perform the following four atomic operations:
 - ✓ Create
 - ✓ Read
 - ✓ Modify
 - ✓ Annotate
- The server manages the PING records which may be stored by *service bureaus* or on any server of the patient's choice
- Medical record data and images may be stored with the PING record, or the record may contain pointers to information and images stored elsewhere.

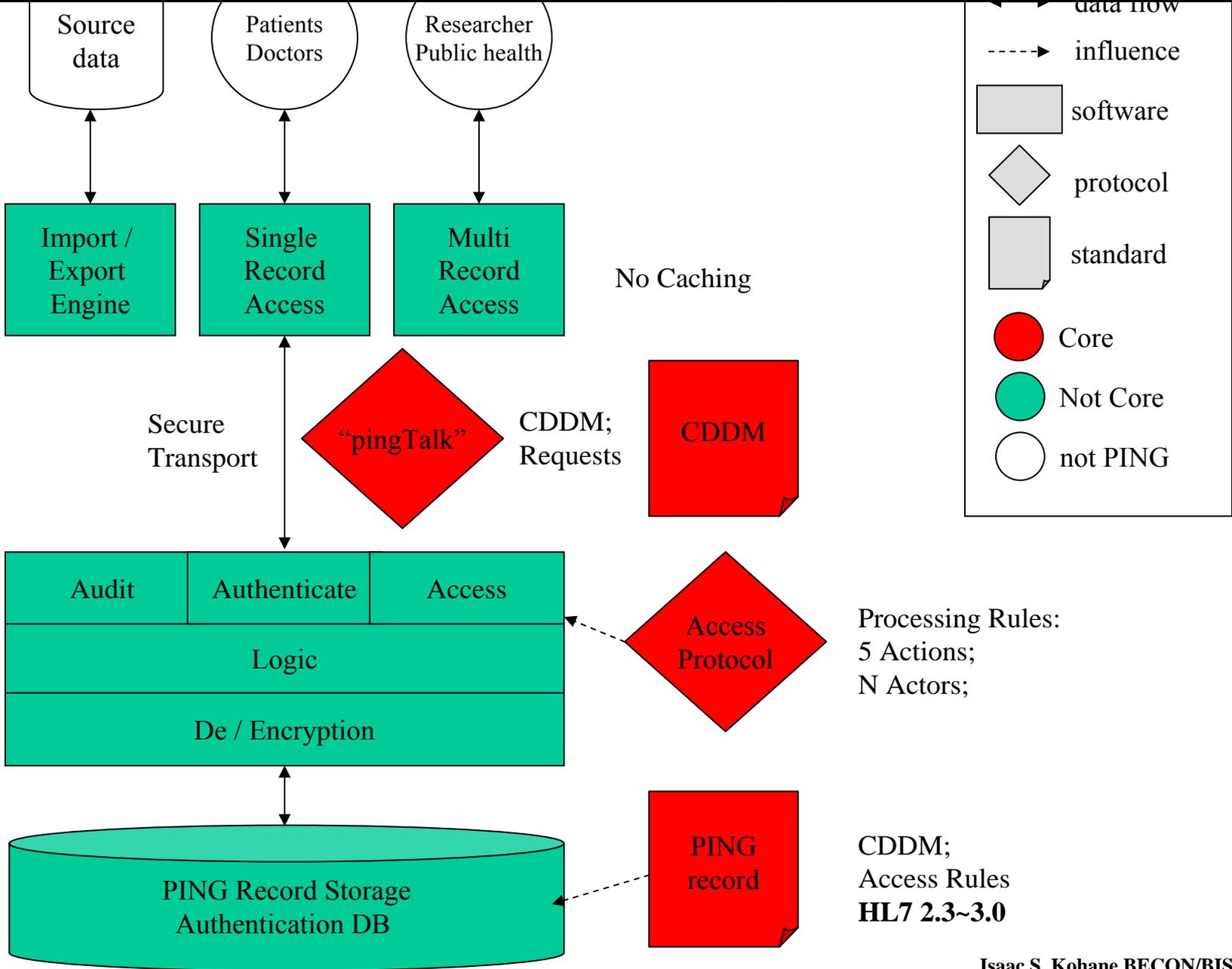


lients

Gateways

Server

Storage





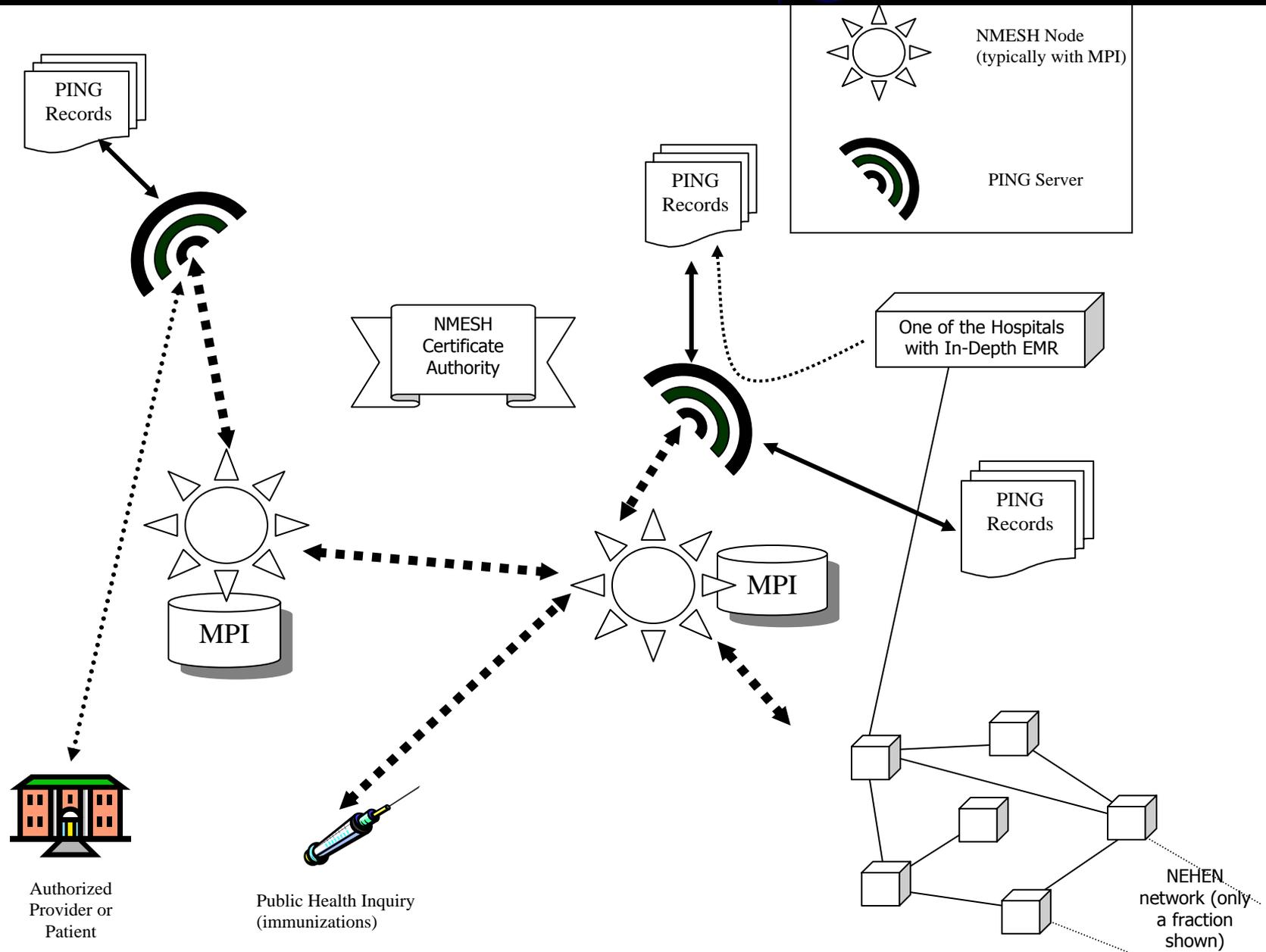
Personal Internetworked Notary and Guardian

- For large-scale disasters (PING-Response)
- For patients poorly tied into the healthcare system.
 - ✓ PING-Citizen (Canada)
- For immunization record
 - ✓ PING 'baby book'
- ***PING Genome***



And Let's Bring it All Together

- The National Multi-Protocol Ensemble for Self-scaling Systems for Health (NMESH)
- Cover an entire region (Northeast)
- Use HL7 data flow from multiple hospitals into PING records
- Use SPIN to provide a peer-peer mechanism to query all the PING records
- Apply for research, clinical care and disaster management.



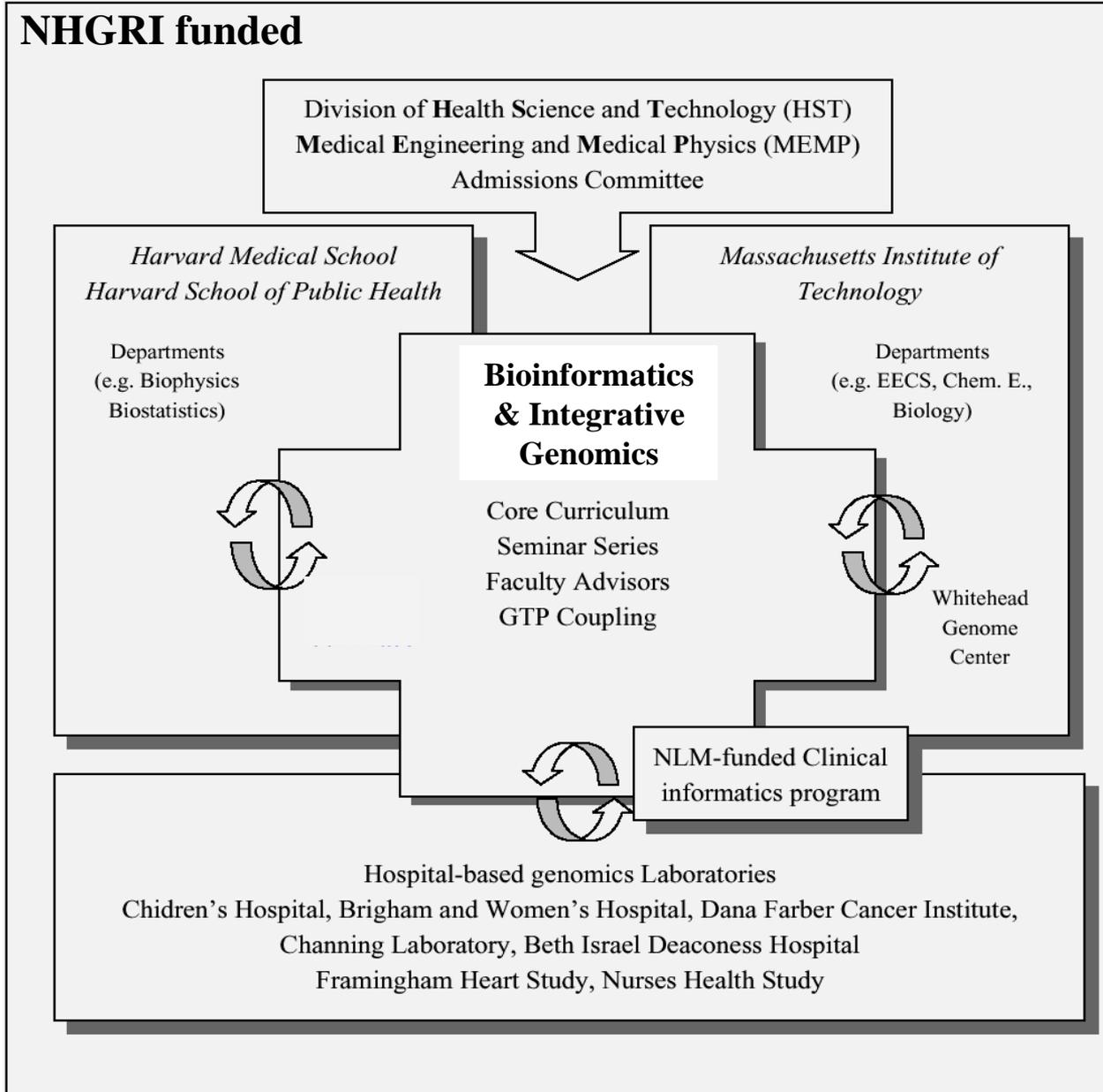


Summary

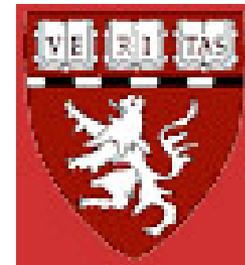
- **STANDARDIZED VOCABULARIES** and **PROTOCOLS** are the essential **glue** to allow
 - ✓ Linking of disparate data types and sources
 - ✓ Leveraging existing methodologies and tools
 - ☞ E.g. homology maps and phenotypic classification
- **Lightweight social engineering is necessary**
 - ✓ Preserve local autonomy
 - ✓ Enable local curation
 - ✓ Represent the first and most important steps to unifying clinical informatics to bioinformatics and the tool builders and discoverers.
- **Do not try to solve all data representations problems**
 - ✓ **Tasteful, partial solutions will incrementally bring you to your goal**
 - ✓ **All out effort will bring you nowhere.**



NHGRI funded



Bioinformatics and Integrative Genomics (BIG)





Thank you

Isaac_Kohane@harvard.edu