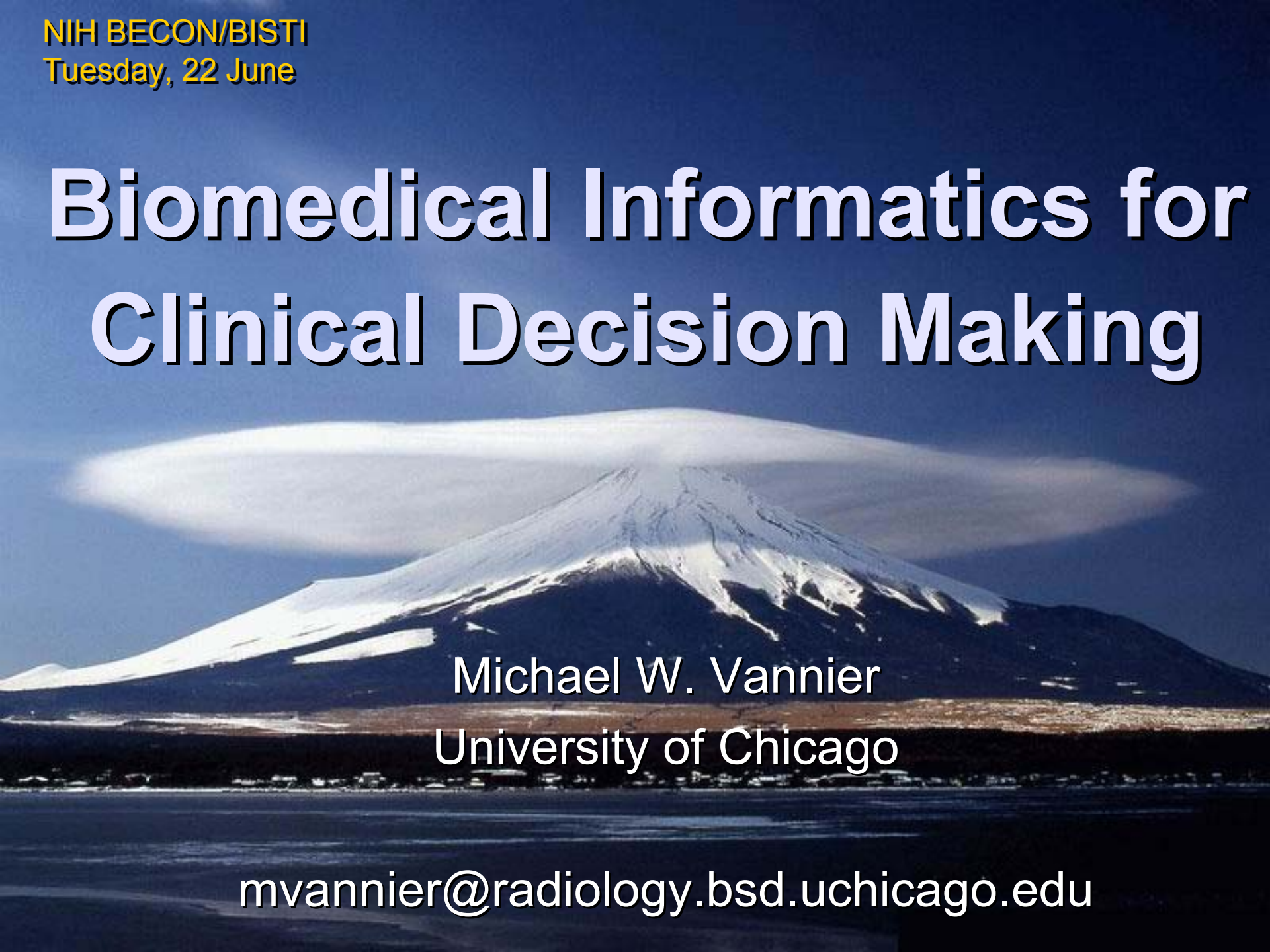
NIH BECON/BISTI
Tuesday, 22 June

Biomedical Informatics for Clinical Decision Making



Michael W. Vannier
University of Chicago

mvannier@radiology.bsd.uchicago.edu

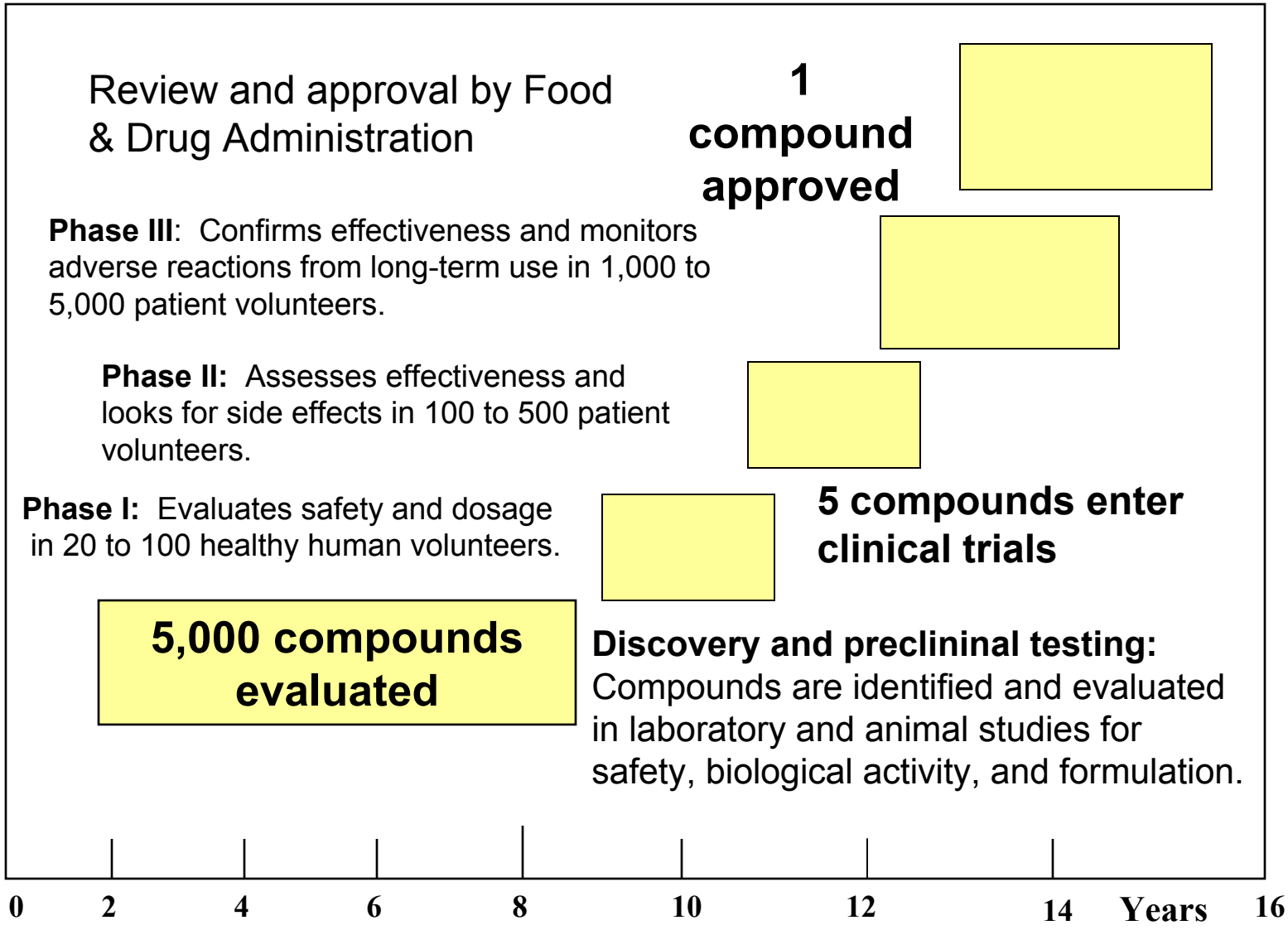
Objectives

- Statement of the problem(s)
- Data management: databases & digital libraries
- Enabling technologies: modeling, software tools & techniques

Outline

- Why do we make decisions?
- What do we need?
- What's wrong? And what can we do to fix it?
  - Integration of diverse sources
  - Infrastructure
  - Challenges and potential solutions
- Conclusion

Bringing a New Drug to Market



Source: Tufts Center for the Study of Drug Development

Current Status

- Clinical decision making is important but errors are common;
  - We need to reuse prior experience and augment human capabilities
- There are several barriers to progress:
  - Reluctance to share primary data (clinical records and images)
  - Islands of excellence in a sea of incompatible systems and data sets

Problem

- High cost
  - (duplication of effort, impaired ability to learn from mistakes, systemic inefficiency)
- Low performance
  - (avoidable errors are common, misdirected effort such as treatments that don't match individual needs)
- System is resistant to change
  - (poor coordination of effort due to Babelization and Balkanization)

Common themes

- Human-centric:
  - Patient-oriented, observer-based, standard-of-care, subject to social / ethical norms
- Database, repository, archive, biobank, data warehouse, registry, …
- Standards: CDISC, HL7, DICOM, …
- Stakeholders: investigators, individual patients, sponsors, institutions, …

- Phenotype – genotype

- Prediction, diagnosis, treatment selection

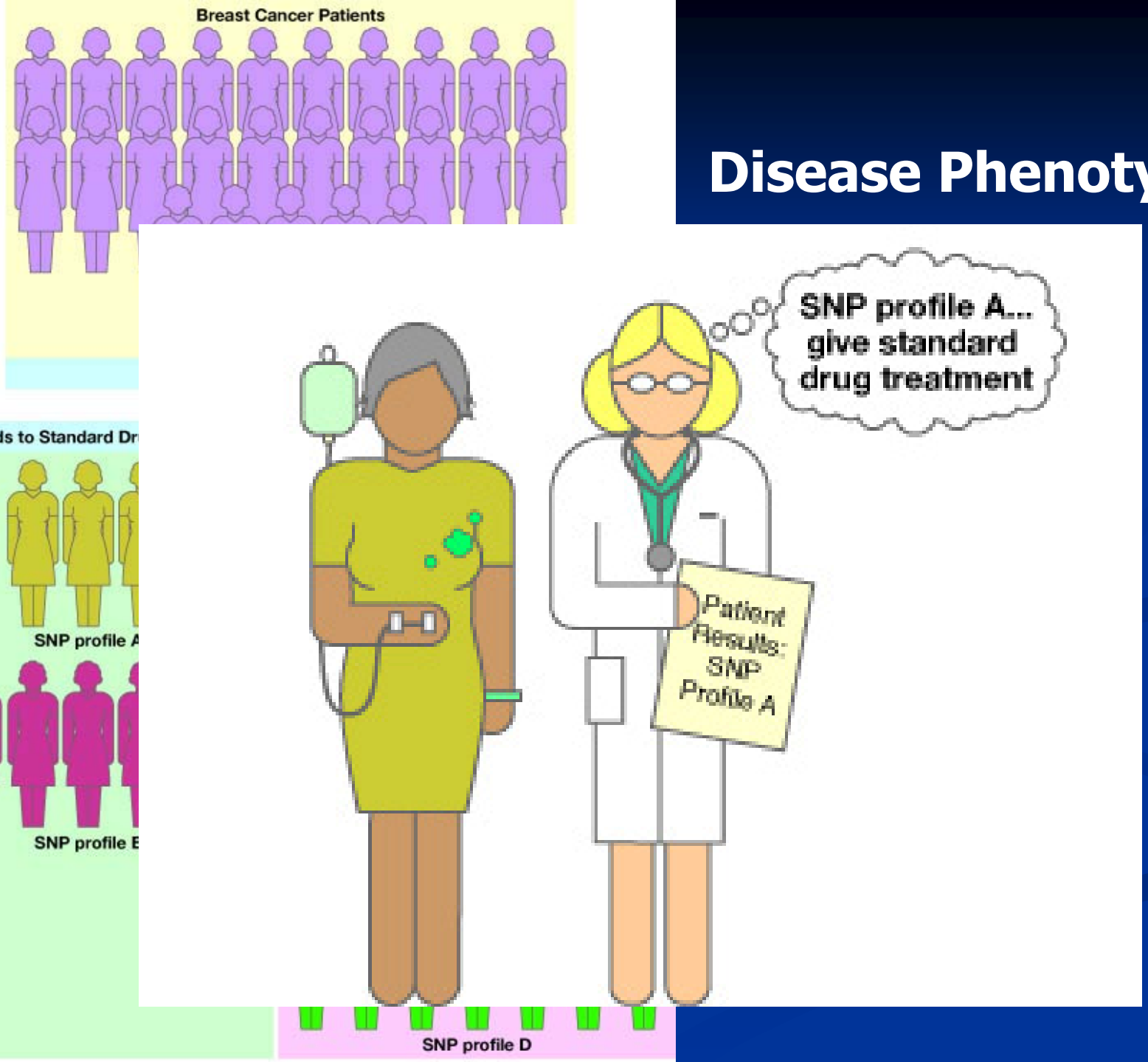
Why do we make images?
Why do we collect clinical data?

- Medical Record
  - Lab and diagnostic testing
  - Prior procedures; drugs
- **To answer important questions and aid decision-making..**
- Diagnosis
- Staging
- Therapy
  - Select best alternative(s)
  - Plan and guide interventions

Decisions

- Which diagnostic test(s) to use?
- Which therapy is best for this patient?
  - Implies that alternatives exist and the outcome is different for patients depending on which one is selected.
  - Common problem: patients receive treatment that confers little or no benefit.
- What's the right dose? Schedule of treatment?
- Will a combination of therapies be more beneficial than one?

Disease Phenotyping



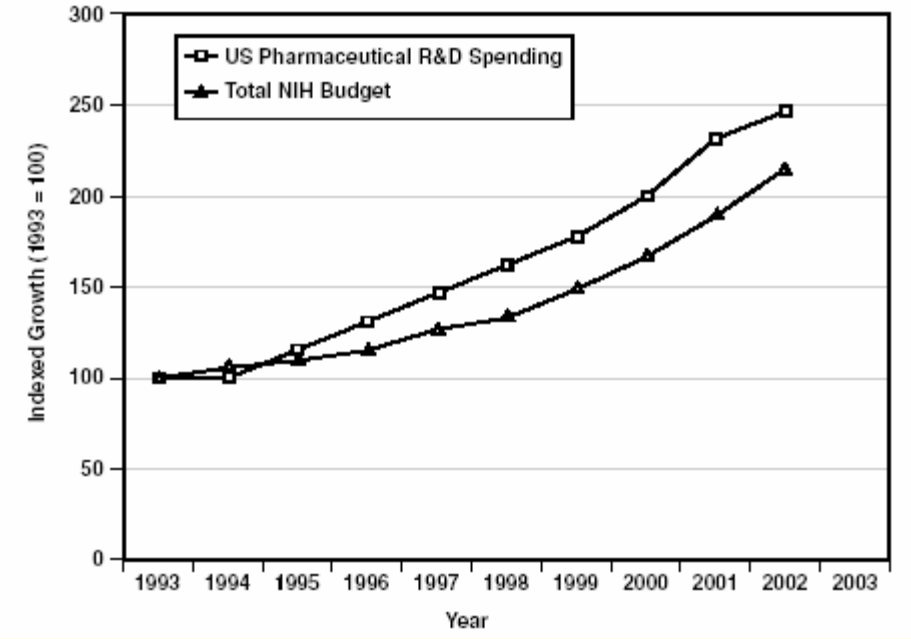
Challenge and Opportunity on the Critical Path to New Medical Products

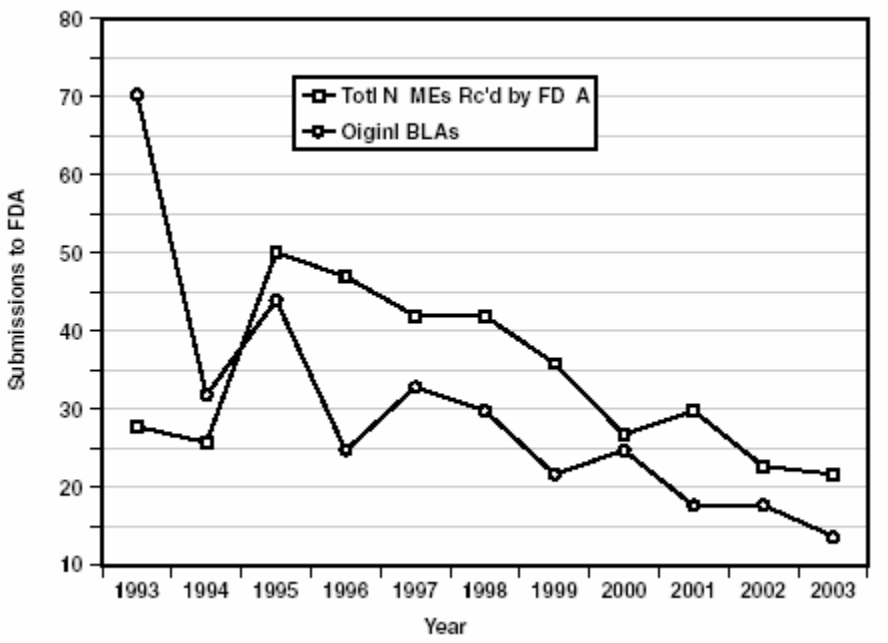
U.S. Department of Health and Human Services
Food and Drug Administration

March 2004

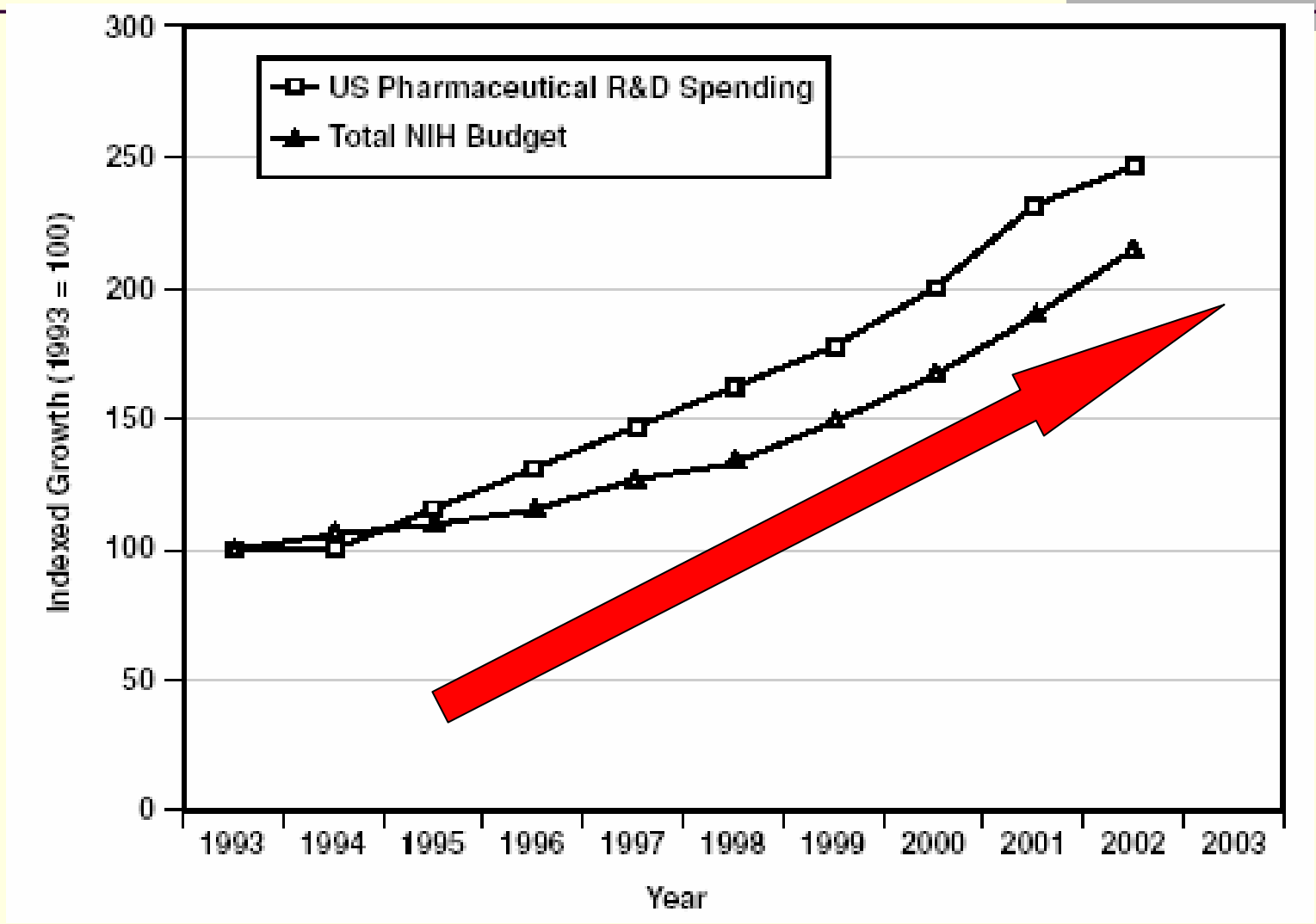
FDA: March 2004 report

The medical product development process is no longer able to keep pace with basic scientific innovation. Only a concerted effort to apply the new biomedical science to medical product development will succeed in modernizing the critical path.

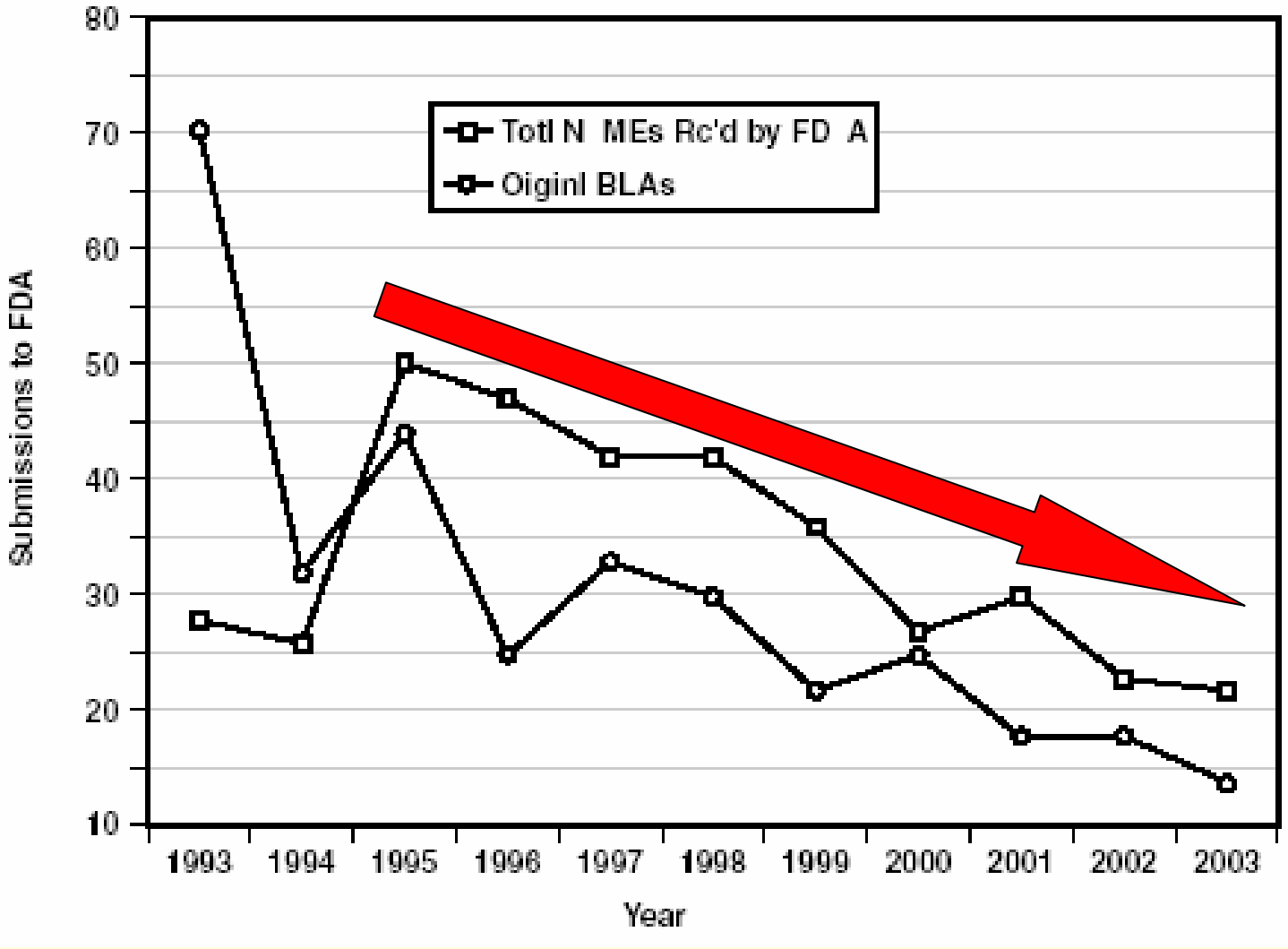




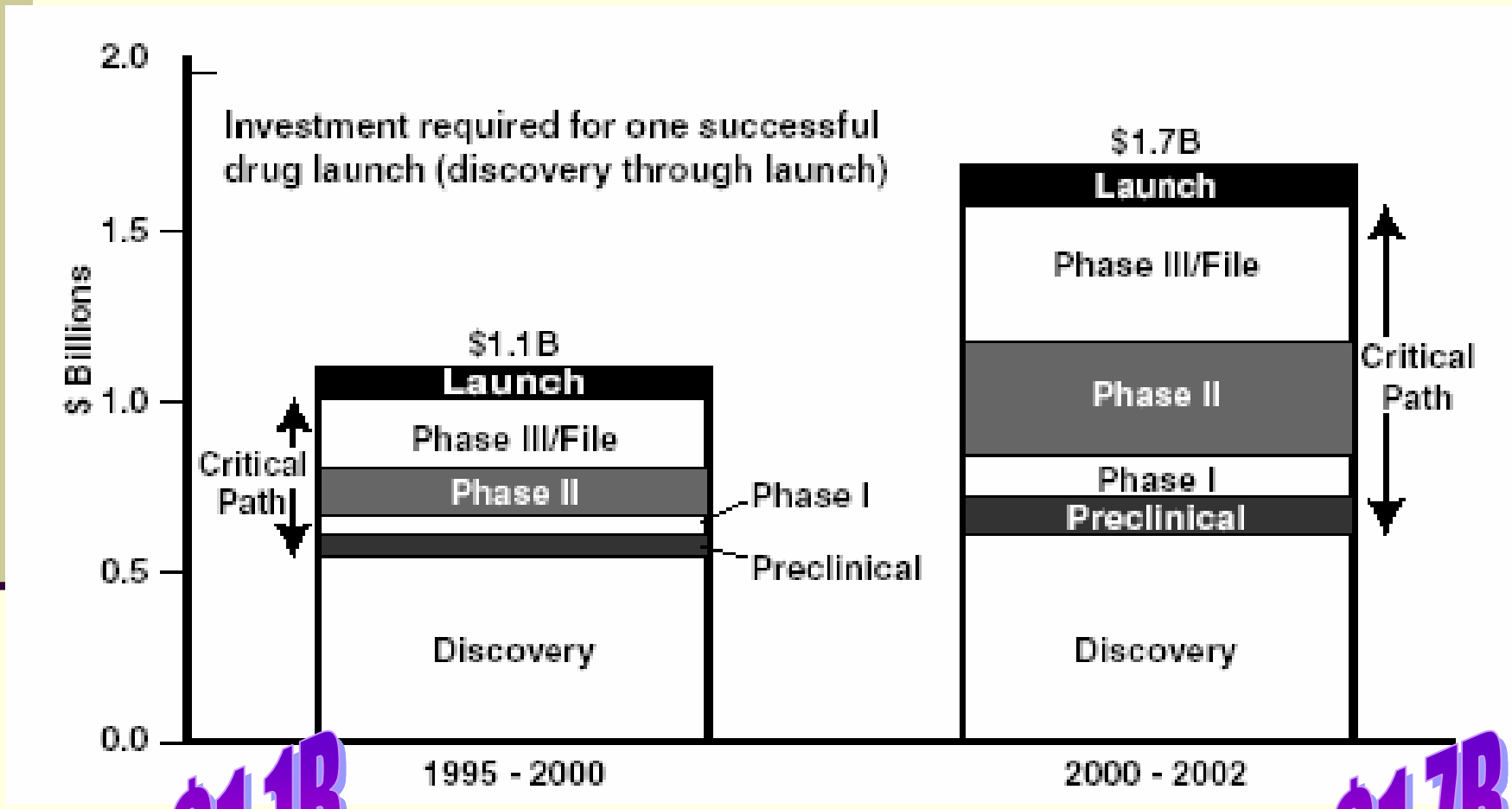
10-Year Trends in Biomedical Research Spending



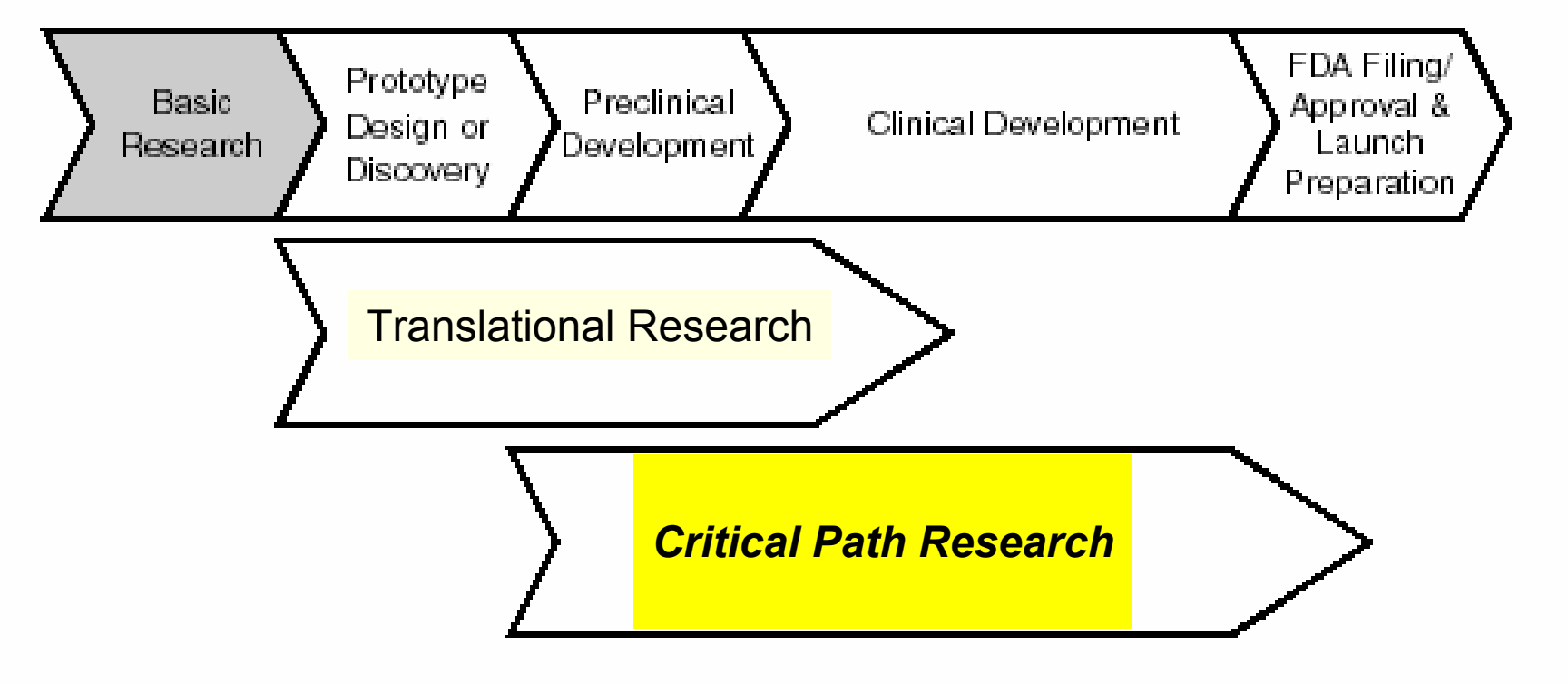
10-Year Trends in Major Drug and Biological Product Submissions to FDA



Investment Escalation per Successful Compound

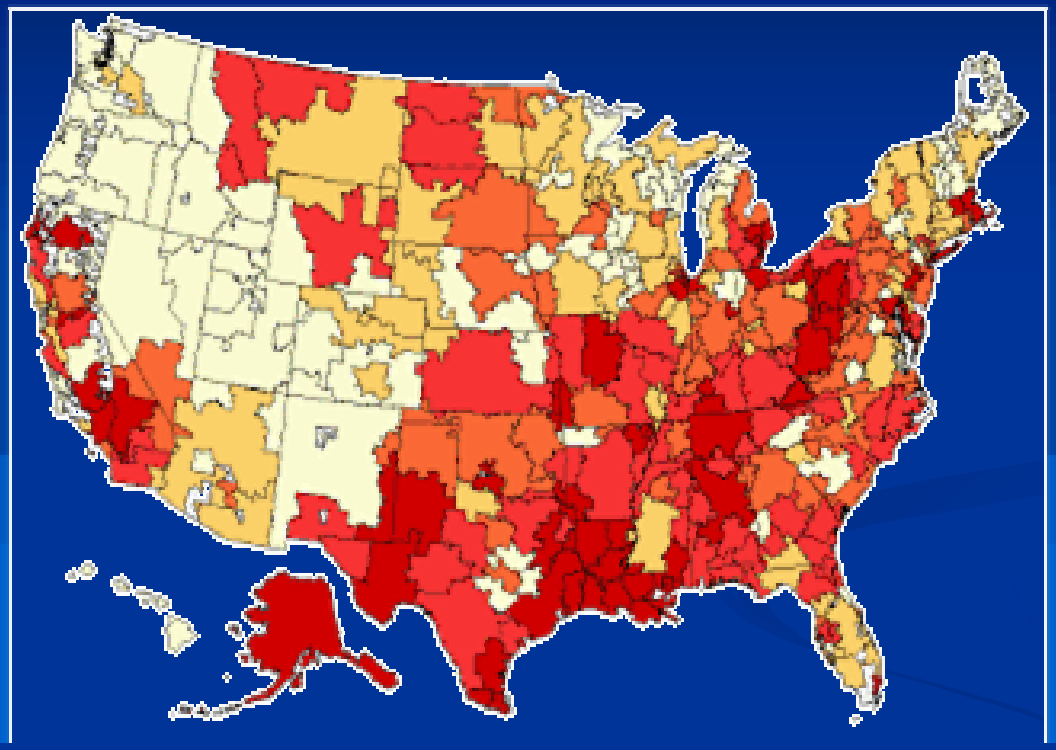


Research Support for Product Development



Medical Errors

The Dartmouth Atlas of Health Care



Inpatient Hospital Services per Medicare Enrollee

By Hospital Referral Region (1995)

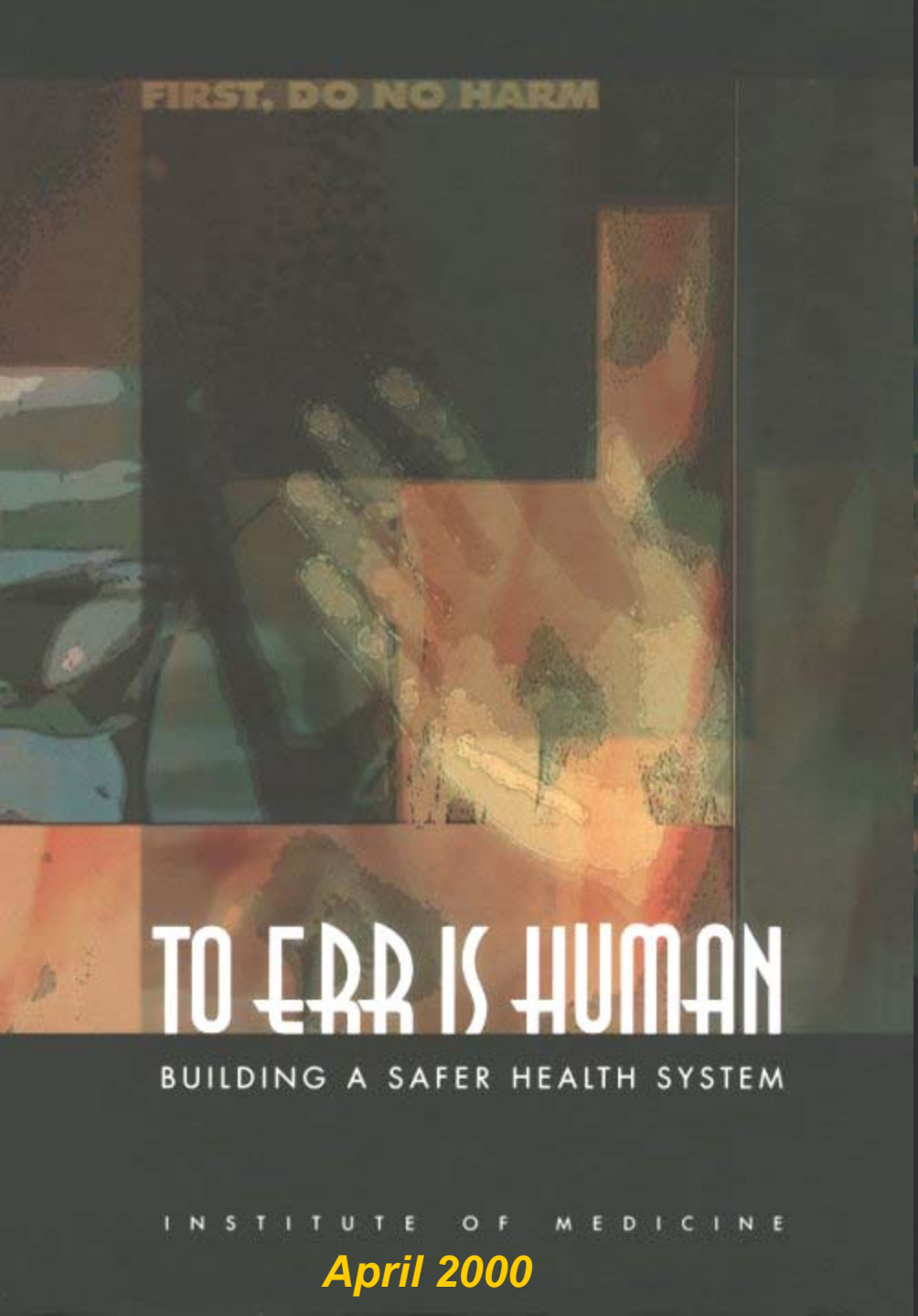
From ~$1500 to $3750 per individual

(More than 2X variation !!)

Clancy

AHRQ

Examples of Popular Press Headlines



April 2000

INSTITUTE OF MEDICINE

July 2001

Institute of Medicine Report
To Err is Human: Building a Safer Health System

Preventable medical errors

- 44,000 to 98,000 Americans die each year
- Eighth leading cause of death in the United States
- Annual cost as much as $29 billion annually
- IOM conclusion: The majority of these problems are systemic, not the fault of individual providers

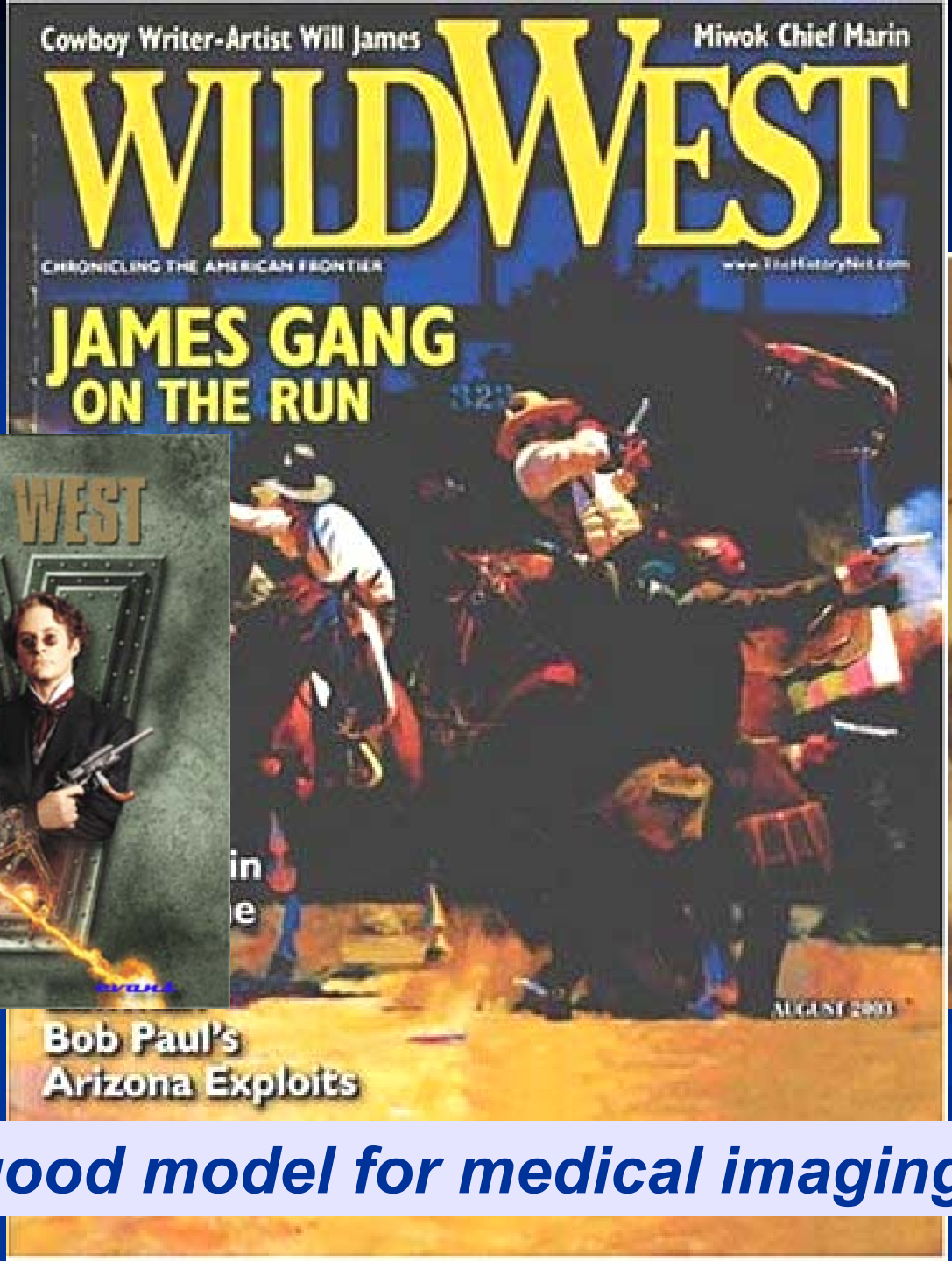
Variability is a major source of errors

Sources of Variability in Medical Imaging

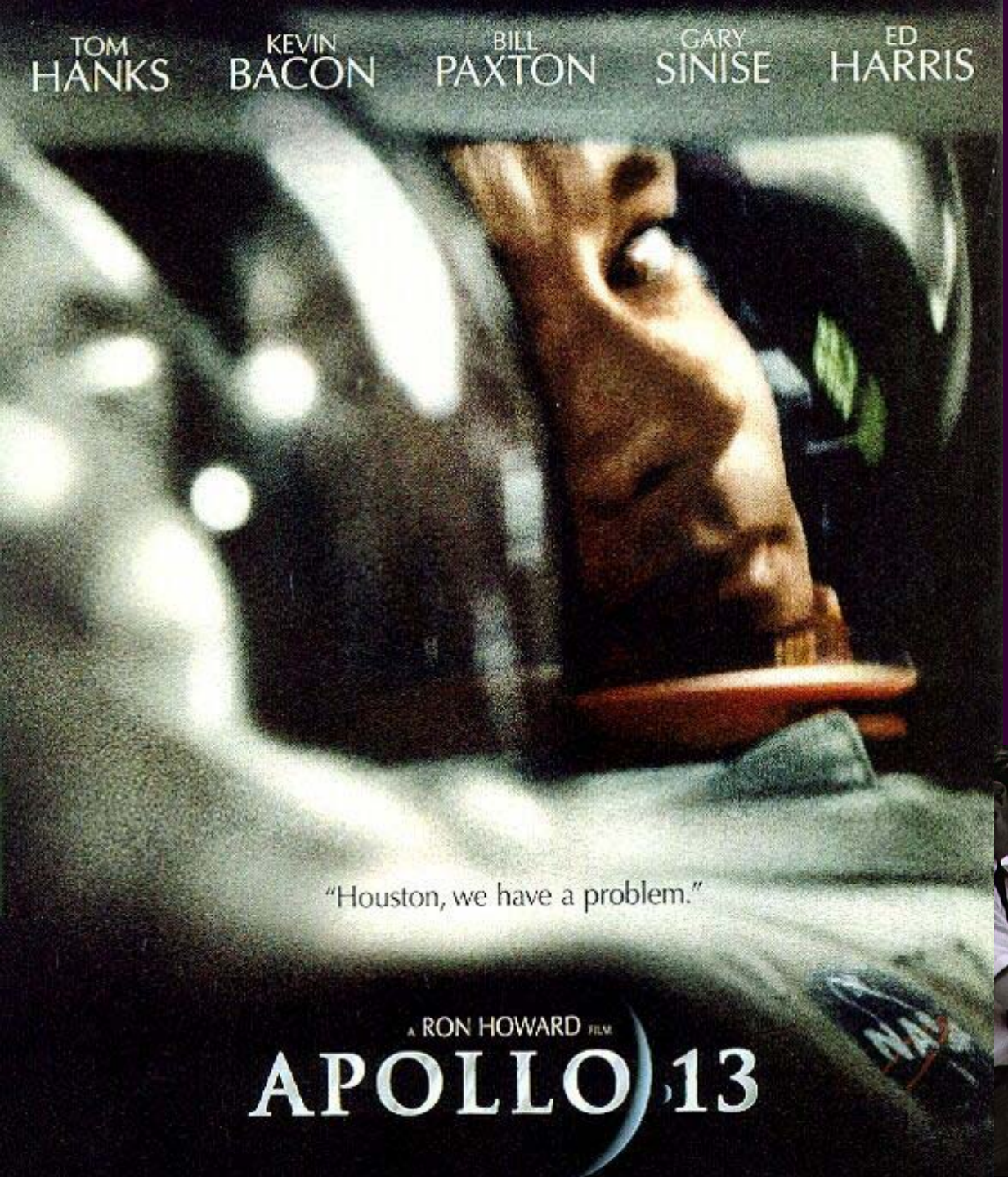
- Instrumentation differences
- Absence of meaningful comparisons
- Missing standards or lack of adherence to them
- Inconsistent acquisition protocols
- Small sample size; selection bias
- Few controls
- Frequent use of human observers and subjective judgments (even when objective measurements are possible)
- AND – a pervasive lack of sharing (data, software, resources)

Wild West of Medical Imaging

- Do your own thing
  - Use different equipment, protocols, formats
  - Unselected or poorly documented subjects
  - Minimal, if any, controls
- Hide the data and don't let anyone else use it
- Conceal the source code used in the analysis tools
  - Create your own tools and keep them to yourself



Is this a good model for medical imaging science?



"Houston, we have a problem."

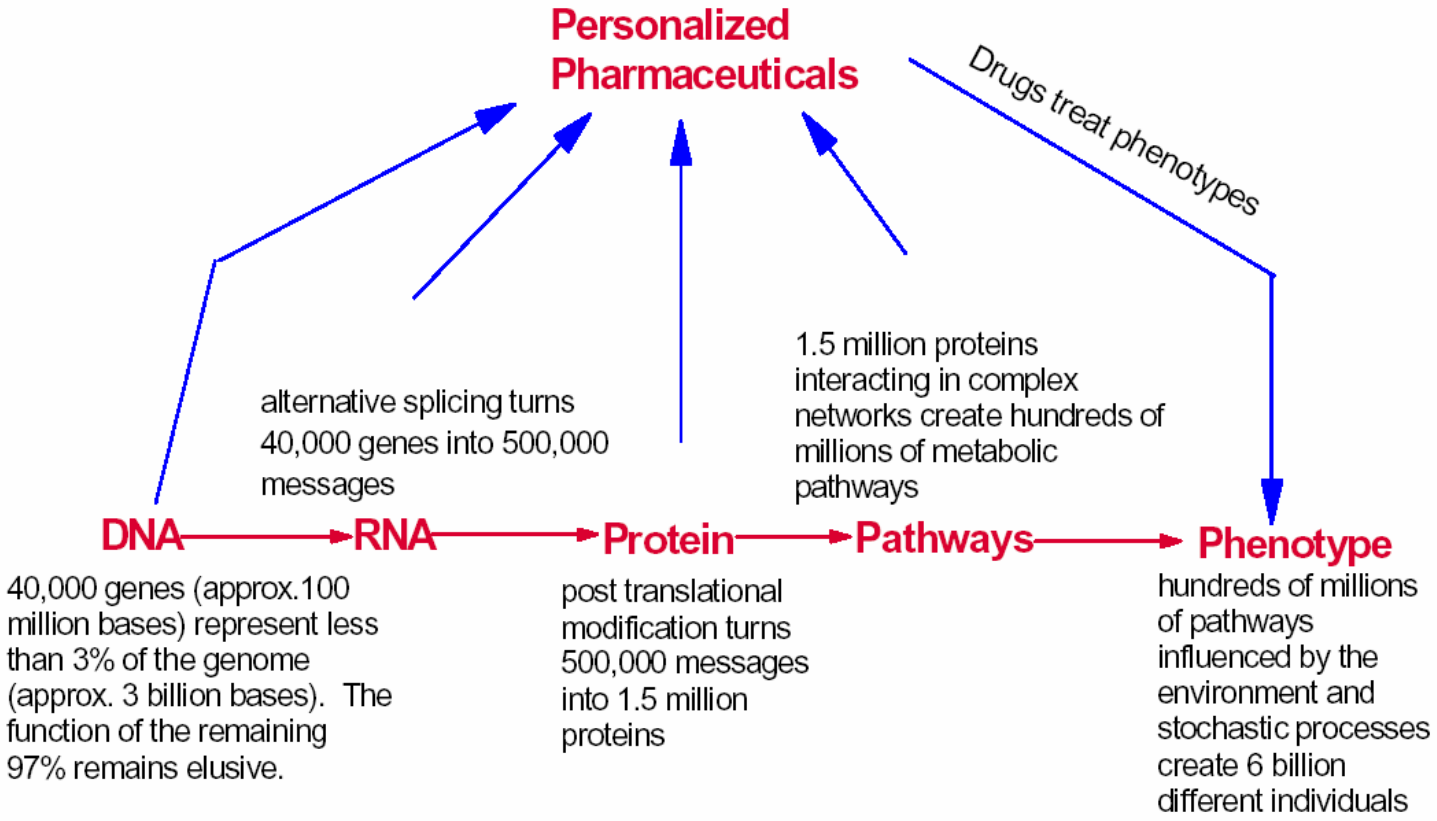
Genetics is Moving

From
A week of usually forgotten medical school lectures

To
The subject that will impact nearly every clinical decision health care providers make

New Sources of Value Creation

Rational drug development requires managing enormous complexity. Pharmaceutical companies are beginning to differentiate themselves on the power of their information technology platforms. IT Platform intellectual property is likely to be more valuable than content (gene sequences, metabolic pathways, protein structures, etc.)



Historically, 220 targets have generated $3trillion of value. Industrialized genome sequencing has created a target rich, lead poor environment that will slowly reverse over the next several years as in-silico biology drives the discovery of new lead compounds.

IBM Life Sciences

New opportunities

- Genome Project
  - Interest for biologists
  - One gene at a time
  - Monogenic diseases
  - Tedious genotyping
  - DNA level
  - Bioinformatics explosion
- Post-Genomics
  - ***Clinical interest***
  - Hundreds or thousands of genes simultaneously
  - ***Complex diseases***
  - High throughput genotyping
  - DNA, RNA, Proteins
  - ***Integration of clinical and genetic information***

Overview

Human Genetic Variation

Technologies | Data | Applications

Genotyping

Haplotyping

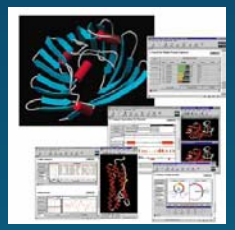
Individual genomics (SNPs and mutations)

Diagnosis

Pharmaco-genetics

Individualised healthcare

Genome



BIOINFORMATICS & MEDICAL INFORMATICS

Information

Gene Expression DNA arrays MS, 2D ef

Functional genomics

proteomics

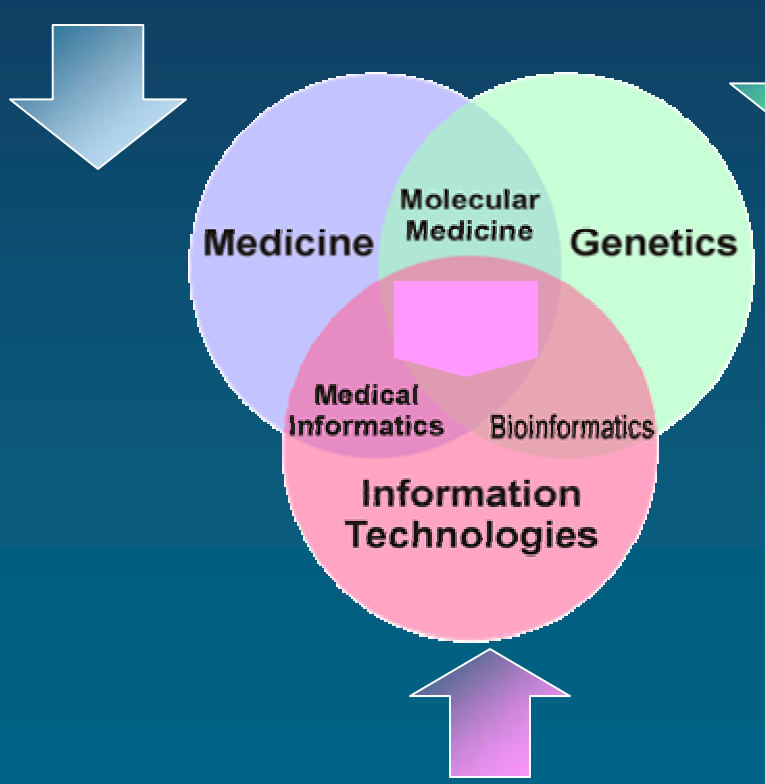
Disease classification Pharmaco-genomics

Molecular medicine

Molecular causes of diseases

A model for studying interactions

To foster the application of bioinformatics in health



To adapt medical informatics systems to the genetics paradigm

To apply IT to facilitate molecular medicine

Assets

- Computer and network technology, imaging systems, fundamental biosciences are improving at a rapid rate
- There are already many tools for biomedical informatics that can substantially improve clinical decision making
- We know how to build new tools and improve existing ones that can address persistent problems in clinical research and practice

Barriers

- Ownership of information by multiple stakeholders
  - Individual, investigator, institution, sponsor; IRB, HIPAA
- Lack of consensus on risk-benefit
  - Imposition of a general solution doesn't work
- Variability and quality
  - Intrinsic variability in human populations and disease processes
  - Lack of consistency in vocabulary, data formats, instrumentation and medical practice norms
- Few examples of mature successful systems
  - How do you measure progress?

Evaluation

- Technical benefits
  - Faster, easier, more reliable, less expensive, …
- Reduce errors
  - Better outcomes
- More and better tools and therapies
  - New drugs and devices become available sooner
  - Reduce time-to-market and increase ROI
- Vanguard projects
  - Reuse of clinical trial data; combination of multiple trials; better trial designs
  - Combination therapies
  - Persistent infrastructure (that transcends individual trials)

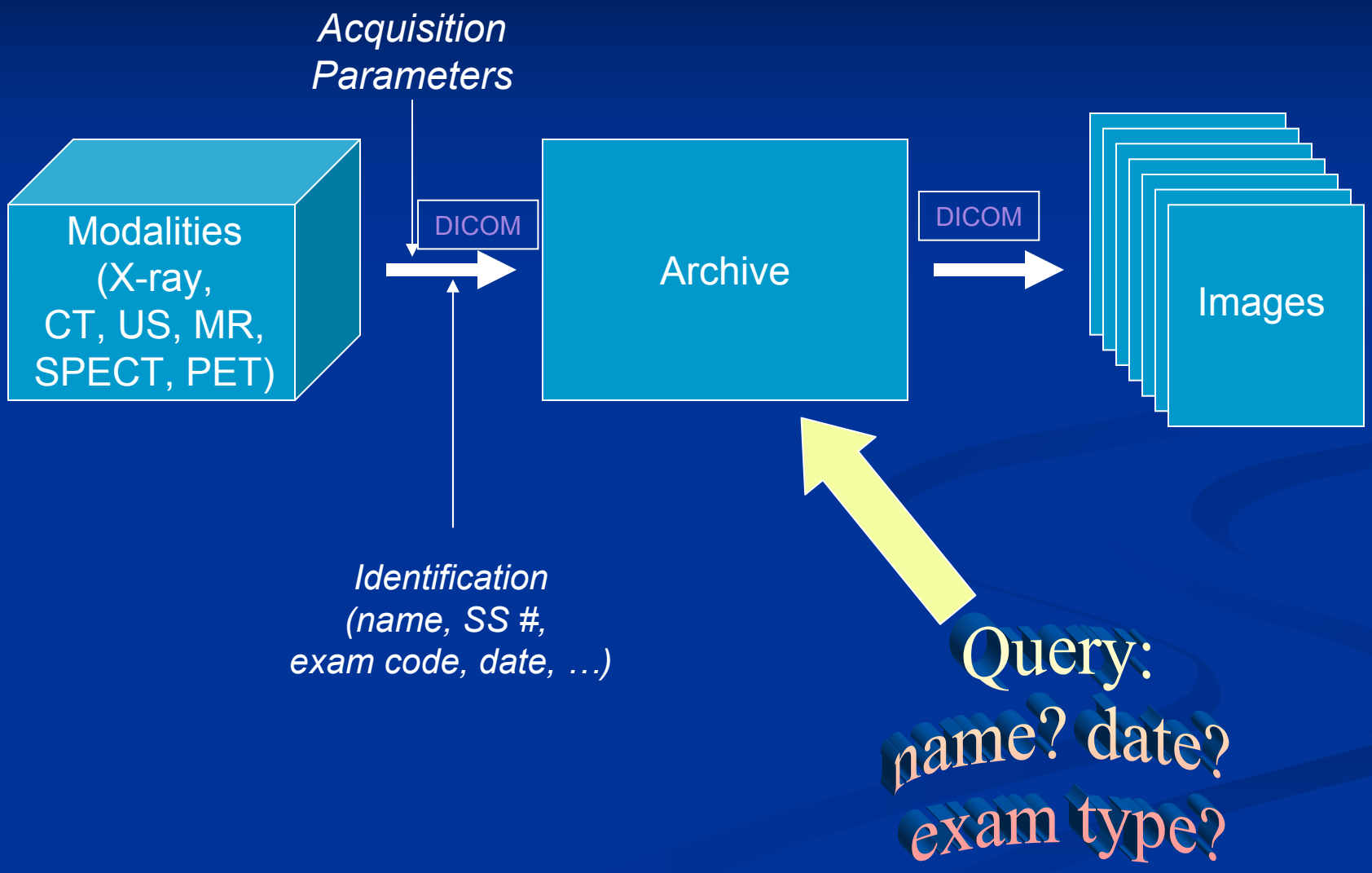
Solutions

- Enable and encourage use of information-based tools to improve decisions (reduce errors, optimize results)
  - Reuse experience
  - Dependence on human observers; data overload
  - Share information
  - Integrate sources

Ongoing Efforts

- At NIH:
  - NCI = caBIG, caCORE (Cancer)
  - NCRR = BIRN (Neuroscience)
  - NECTAR (Roadmap)
- Not yet fully developed
- Operational experience (especially in a clinical setting with integrated records and images) is minimal
- These consortia will guide development and expansion of infrastructure, definition of needs, and provide proof of benefit

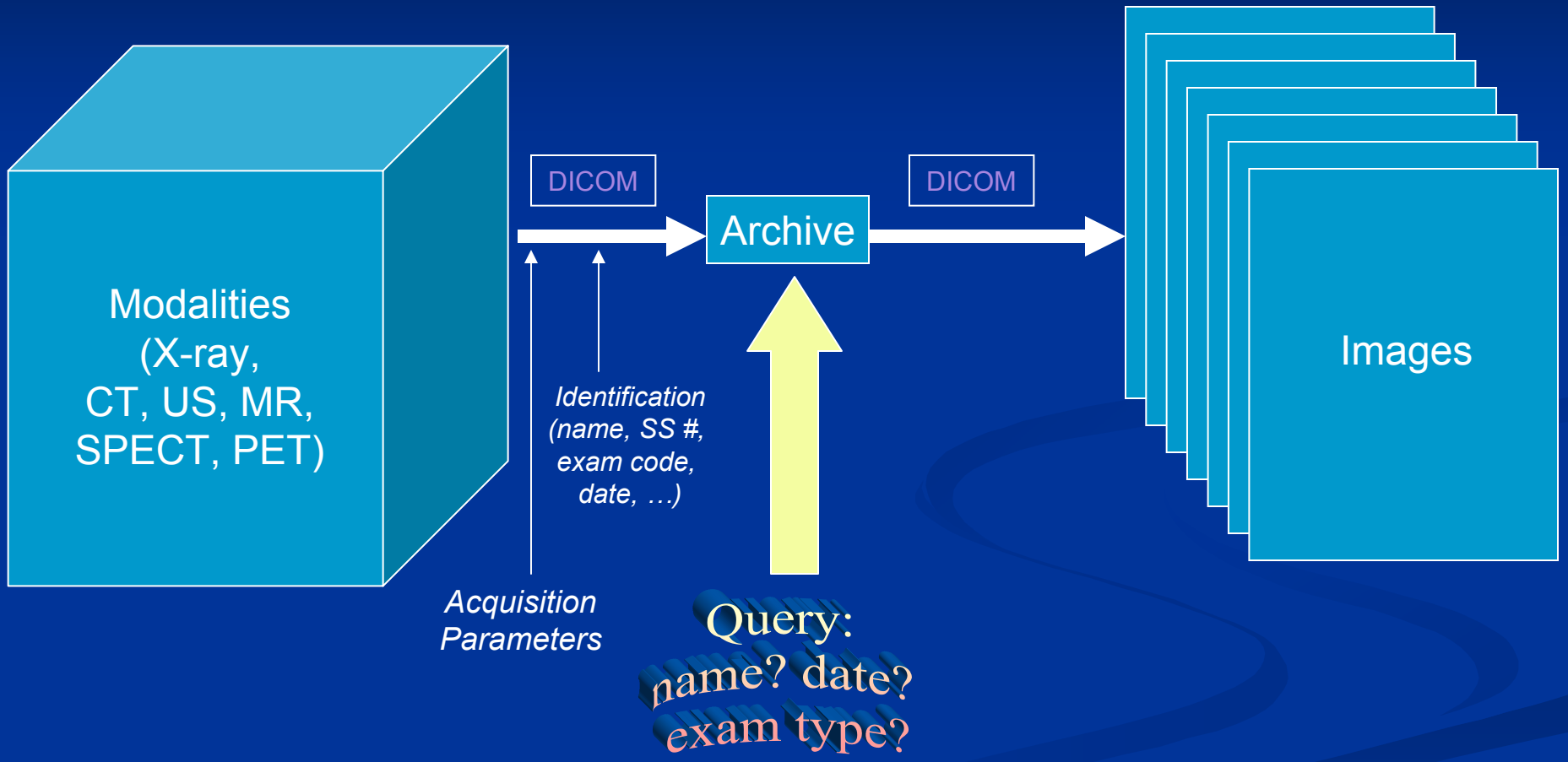
Today

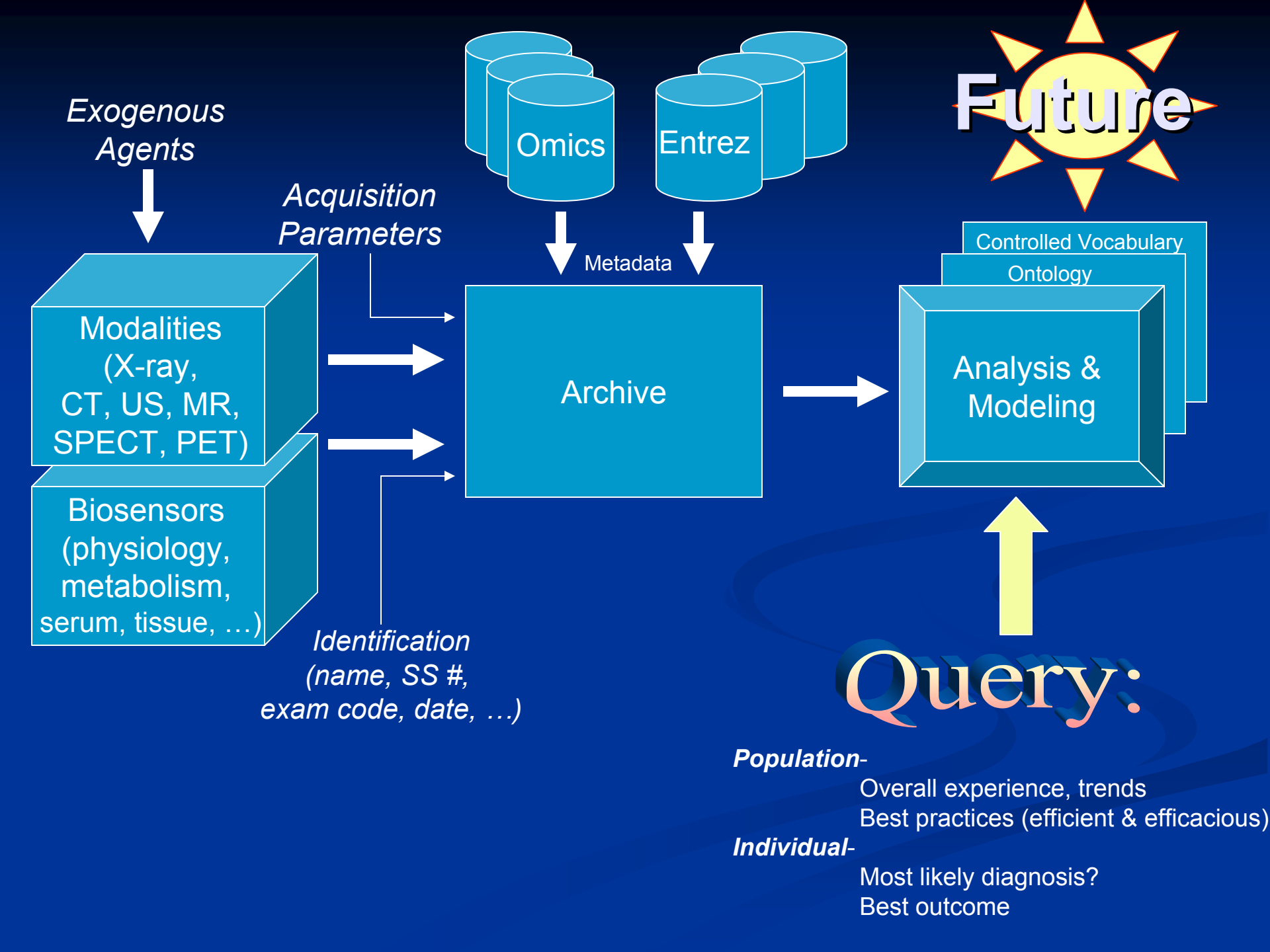

Acquisition Parameters
Modalities (X-ray, CT, US, MR, SPECT, PET)
DICOM
Archive
DICOM
Images
Identification (name, SS #, exam code, date, …)
Query: name? date? exam type?

What's wrong?

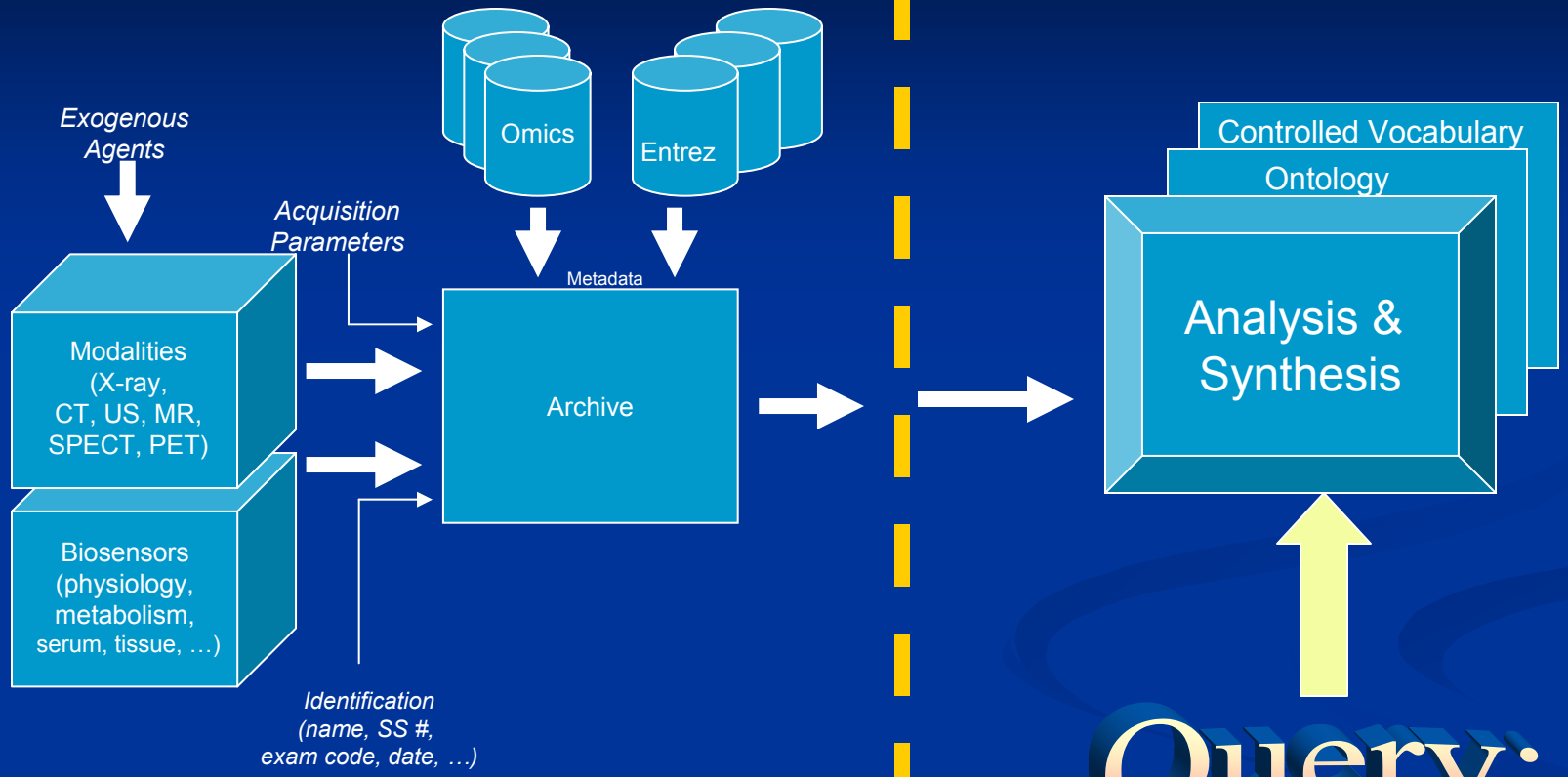
- Too much variation
- Inability to collect and reuse experience in a beneficial manner
- Information is incomplete, contradictory, or misinterpreted
- Avoidable mistakes are made
- Experts are few, subspecialized and not always available

An Image-Centric World View




Exogenous Agents
Acquisition Parameters
Omics
Entrez
Future
Metadata
Controlled Vocabulary
Ontology
Modalities (X-ray, CT, US, MR, SPECT, PET)
Biosensors (physiology, metabolism, serum, tissue, …)
Archive
Analysis & Modeling
Identification (name, SS #, exam code, date, …)
Query:
Population-
Overall experience, trends
Best practices (efficient & efficacious)
Individual-
Most likely diagnosis?
Best outcome

Future



Hidden

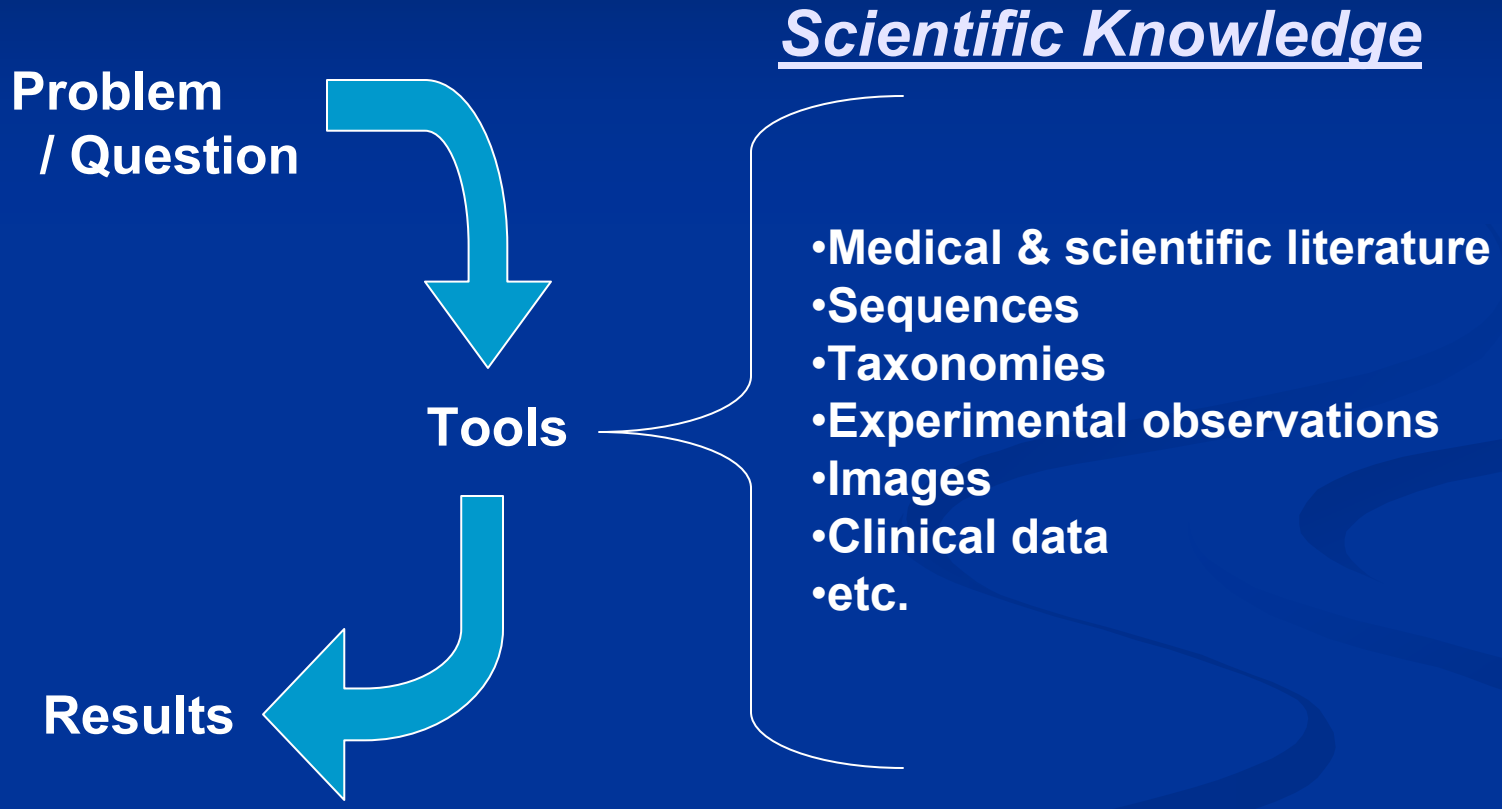
Population-

Overall experience, trends
Best practices (efficient & efficacious)

Individual-

Most likely diagnosis?
Best outcome

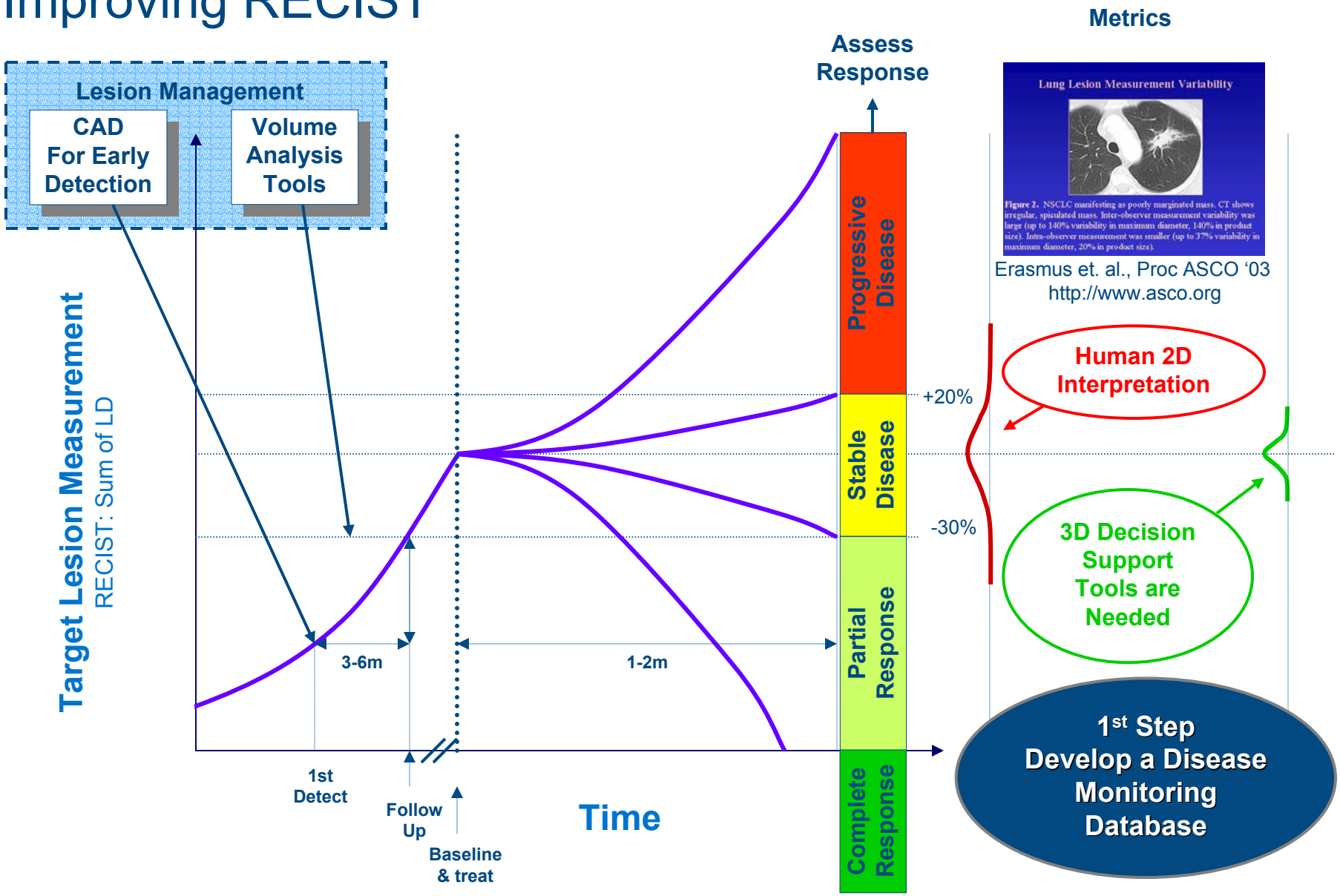
Investigation



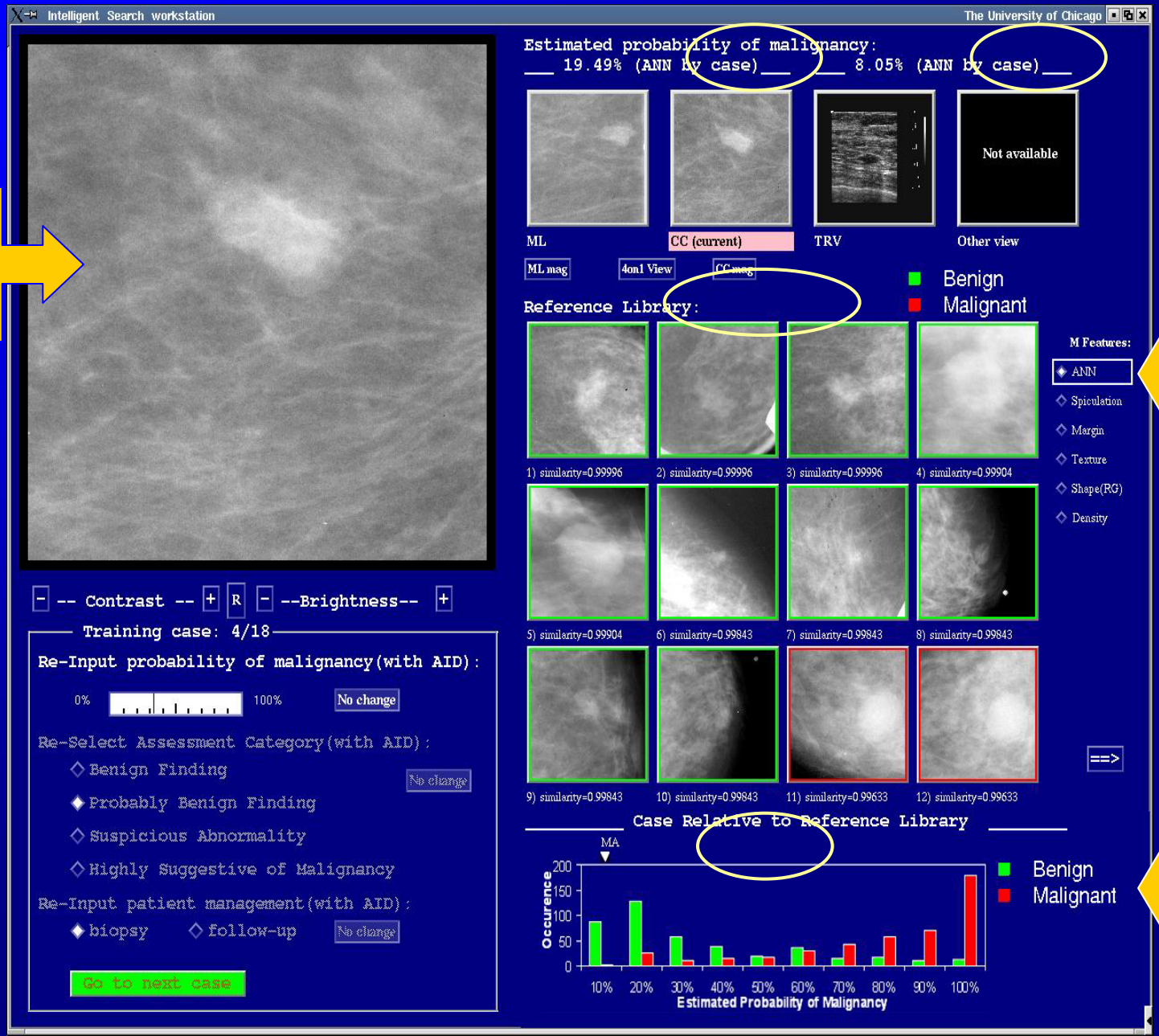
Response Criteria for Imaging of Solid Tumors (RECIST)

- Unidimensional size of tumor from CT scans or other images
- Mandatory for NCI clinical trials
- Measurements are made by human operator
- We can do a lot better…

Improving RECIST



Intelligent Workstation



Current Case

Matched Cases from Archive

(with known results)

Similarity Measure

The New England Journal of Medicine

CYCLIN E AND SURVIVAL IN PATIENTS WITH BREAST CANCER

KHANDAN KEYOMARSI, PH.D., SUSAN L. TUCKER, PH.D., THOMAS A. BUCHHOLZ, M.D., MATTHEW CALLISTER, M.D., YE DING, PH.D., GABRIEL N. HORTOBAGYI, M.D., ISABELLE BEDROSIAN, M.D., CHRISTOPHER KNICKERBOCKER, M.S., WENDY TOYOFUKU, B.S., MICHAEL LOWE, B.S., THADDEUS W. HERLICZEK, M.D., AND SARAH S. BACUS, PH.D.

ABSTRACT

Background Cyclin E, a regulator of the cell cycle, affects the behavior of breast-cancer cells. We investigated whether levels of cyclin E in the tumor correlated with survival among patients with breast cancer.

Methods Tumor tissue from 395 patients with breast cancer was assayed for cyclin E, cyclin D1, cyclin D3, and the HER-2/*neu* oncogene with the use of Western blot analysis. Full-length, low-molecular-weight, and total cyclin E were measured. Immunohistochemical assessments of cyclin E were also made of 256 tumors. We sought correlations between levels of these molecular markers and disease-specific and overall survival.

Results The median follow-up was 6.4 years. A high level of the low-molecular-weight isoforms of cyclin E,

THE prognosis in patients with newly diagnosed breast cancer is determined primarily by the presence or absence of metastases in draining axillary lymph nodes.[1] However, in approximately one third of women with breast cancer who have negative lymph nodes, the disease recurs, and about one third of patients with positive lymph nodes are free of recurrence 10 years after local–regional therapy.[2,3] These data highlight the need for more sensitive and specific prognostic indicators.

A number of biologic factors have been used to refine risk categories in breast cancer. We have focused on the role of cyclin E in determining the virulence and metastatic potential of tumor cells.[4-8] In normal dividing cells, cyclin E regulates the transition from

N Engl J Med, Vol. 347, No. 20, **1566-1575** November 14, 2002

overall survival. Total cyclin E levels and the level of low-molecular-weight forms of cyclin E as measured by Western blotting but not by immunohistochemical analysis proved to be strongly associated with survival among patients with breast cancer.

METHODS

Tissue Samples and Study Patients

Tumor tissue was obtained from a centralized reference laboratory (Quantitative Diagnostic Laboratories). A total of 430 samples consisting of a minimum of 100 mg of breast-cancer tissue were available. Each patient had received a diagnosis of breast cancer between 1990 and 1995 at 1 of 12 hospitals in the Chicago area. Specimens were shipped to the Wadsworth Center research laboratories for Western blot analysis. This study was approved by the institutional review board of the Wadsworth Center.

The reference laboratory also provided base-line pathological and demographic data (obtained from the individual hospitals), as well as the steroid-receptor status, the DNA index, and the proliferation index (as described below). Information concerning clinical staging and survival was obtained from the tumor registries of each hospital. Patients whose death was clearly documented to be due to breast cancer were considered to have died of breast cancer; other deaths were considered not to have been caused by breast cancer. The data presented here are from 395 patients for whom data on outcome were available.

Hormone-Receptor, DNA, and Proliferation Assays

The procedures for the hormone-receptor and proliferation assays

values obtained from normal tissues. To
pression was scored as high if the value w
than the highest value for normal breast
normal tissue samples were examined.
weight and total cyclin E, specimens wi
were classified as high. All normal-cell
cyclin D3 and HER-2/*neu*. On Western
ues for these proteins clustered into thre
as negative, low level, or high level. Den
were used to standardize for equal protei
ples assayed. The Western blot analysis an
tioned biologic markers were performed
unaware of the patients' outcomes.

Immunohistochemical Studies

A subgroup of 256 samples of tumo
immunohistochemical analysis with the
polyclonal antibody to cyclin E.[5] We us
corresponding to amino acids 381 to 4
tigen in the affinity purification. This po
the same epitope as monoclonal HE12 a
Western blots to detect both the full-l
weight isoforms of cyclin E.[5,7,8] Snap-froz
in Optimal Cutting Temperature compo
tervals, placed on coated slides, fixed, and
viously described.[22-24] At least two represe
each patient with breast cancer were ex
from 0 to 10 on the basis of the intensi
centage of tumor cells stained. In 15 case
were tested along with tumor tissue. Sc
controls ranged from 0 to 2. The tumo
ignated as having eithe

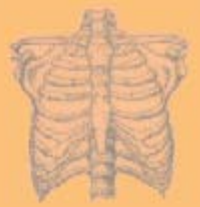
http://www.rand.org/publications/MR/MR954/

Handbook of Human Tissue Sources

A National Resource of Human Tissue Samples

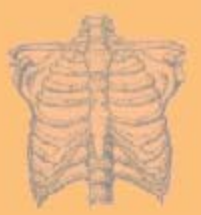
Elisa Eiseman
Susanne B. Haga

Science and Technology Policy Institute
RAND



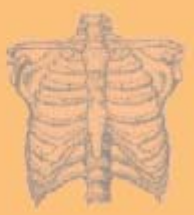
INFORMATION IN THE HANDBOOK

- Where are tissues stored?
- How many tissues are stored in each repository?
- Who are the sources of stored tissue samples?
- Why were the tissue samples originally collected?
- For what purposes have the tissues been used?
- Who has access to the samples?
- How are the tissue samples stored?
- What identifying information is kept with the tissues?



LARGE TISSUE BANKS, REPOSITORIES, AND CORE FACILITIES

(~120 million specimens)



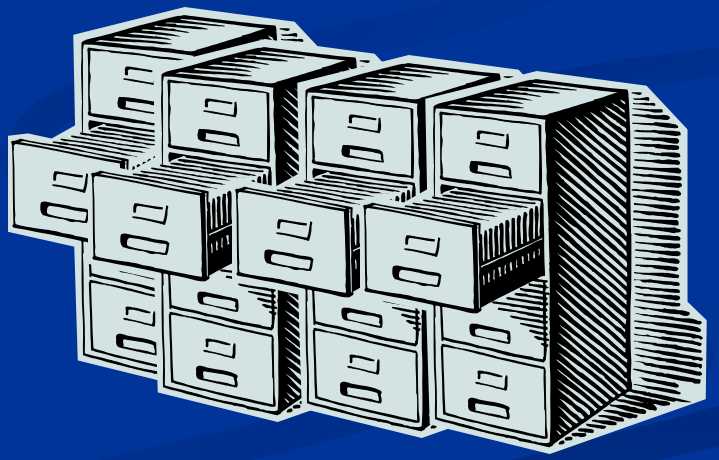
- Military Facilities
- National Institutes of Health
- National Institutes of Standards and Technologies
- Environmental Protection Agency
- Research Universities and Academic Medical Centers
- Commercial Enterprises
- Nonprofit Organizations

Image Repositories

- Open access to image archives is rare
- Image archives are not organized for research queries
- Image archives are not linked to other forms of biological data
- In general, there is no equivalent of a “specimen repository” for images
- This is a major problem for imaging research

Silo of Data

=



VIEWPOINT: NEUROSCIENCE

Neuroimaging Databases

The Governing Council of the Organization for Human Brain Mapping (OHBM)

These are comments written by the Governing Council of the Organization for Human Brain Mapping (OHBM), the primary international organization dedicated to neuroimaging research. The purpose of these comments is to identify and frame issues concerning data sharing within the neuroimaging community. Data sharing has become an important issue in most fields of science. The neuroimaging community is no exception, and it clearly perceives potential benefits in such efforts, as have been realized in other fields such as genomics. At the same time, such efforts can be costly (both in time and expense), and there are important factors that differentiate brain imaging from other fields and that pose specific challenges to the generation of useful neuroimaging databases. These include the rapid pace of change in brain imaging technologies; the complexity of the variables that must be specified to meaningfully interpret the results (such as the method of image acquisition, behavioral design, and subject characteristics); and concerns about participant confidentiality. These issues are outlined with the goal of framing and promoting a public discussion of the benefits and risks of data sharing, which can inform the field of neuroimaging as well as others that face similar challenges.

that obscures their full complexity. The data themselves take a variety of forms and typically are not accessible for widespread sharing and use. Making neuroimaging data more accessible for sharing would facilitate the comparison of findings across laboratories, to allow better assessment of the reliability of methods and reproducibility of results; encourage meta-analyses that explore phenomena that are not apparent in individual data sets; and give investigators who do not have access to neuroimaging facilities the opportunity to conduct research using existing data. All of these are more efficient uses of neuroimaging data, which are relatively expensive to collect.

Some challenges. These potential benefits and the success of data sharing in other com-

Essay

Neuroscience Networks

Data-sharing in an Information Age

Thomas R. Insel,* Nora D. Volkow, Ting-Kai Li, James F. Battey, Jr., Story C. Landis

NIH Institute Directors
NIMH
NIDA
NIAAA
NIDCD
NINDS

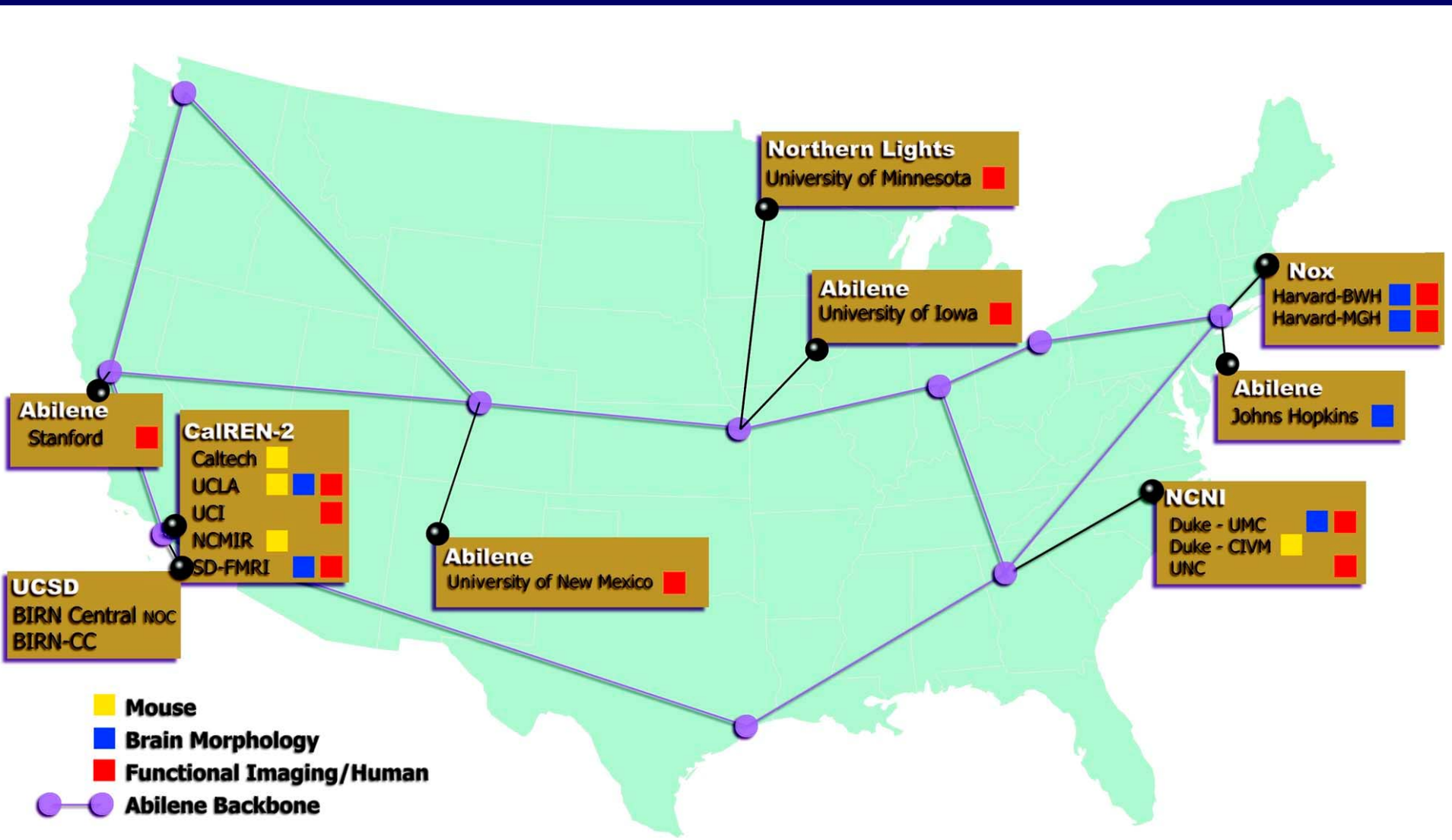
As we emerge from the "decade of the brain," we are entering a decade for which data-sharing will be the currency for progress in neuroscience.

PLoS Biology | http://biology.plosjournals.org

1(1):9-11; 2003

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

BIRN Network



IT Infrastructure to hasten the derivation of new understanding and treatment of disease through use of distributed knowledge

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

What is BIRN?

- Testbed for a biomedical knowledge infrastructure
- Creation and support federated bioscience databases
- Data integration
- Interoperable analysis tools
- Datamining software
- Scalable and extensible
- Driven by research needs pull, not technology push

http://www.nbirn.net

The Value of Tissue Banks to Drug and Dx Developers

Barbara L. Handelin, Ph.D.

Conflicts of Interest, Privacy/Confidentiality, and Tissue Repositories: Protections, Policies, and Practical Strategies

From Columbia University Bioethics Workshop, May 2004

Tissue Banks for Tx and Dx Developers: What is the need?

- Basic research: the biological revolution in the medicinal chemical industry
  - D
  - M
  - B
- Clin
  pharmacogenetic/genomic profiling
  - Toxicity
  - Drug responsiveness
  - Rescuing failed drugs

PGx will drive the collection and use of stored tissues into common practice involving millions of subjects

Examples: Tissue Repositories

- Cooperative Groups (such as the Cooperative Human Tissue Network)
- Ardais
- GeneLogic
- International Genomics Consortium
- IMPATH
- Integrated Lab Services

Ardais

A Framework for Bioethical Standards

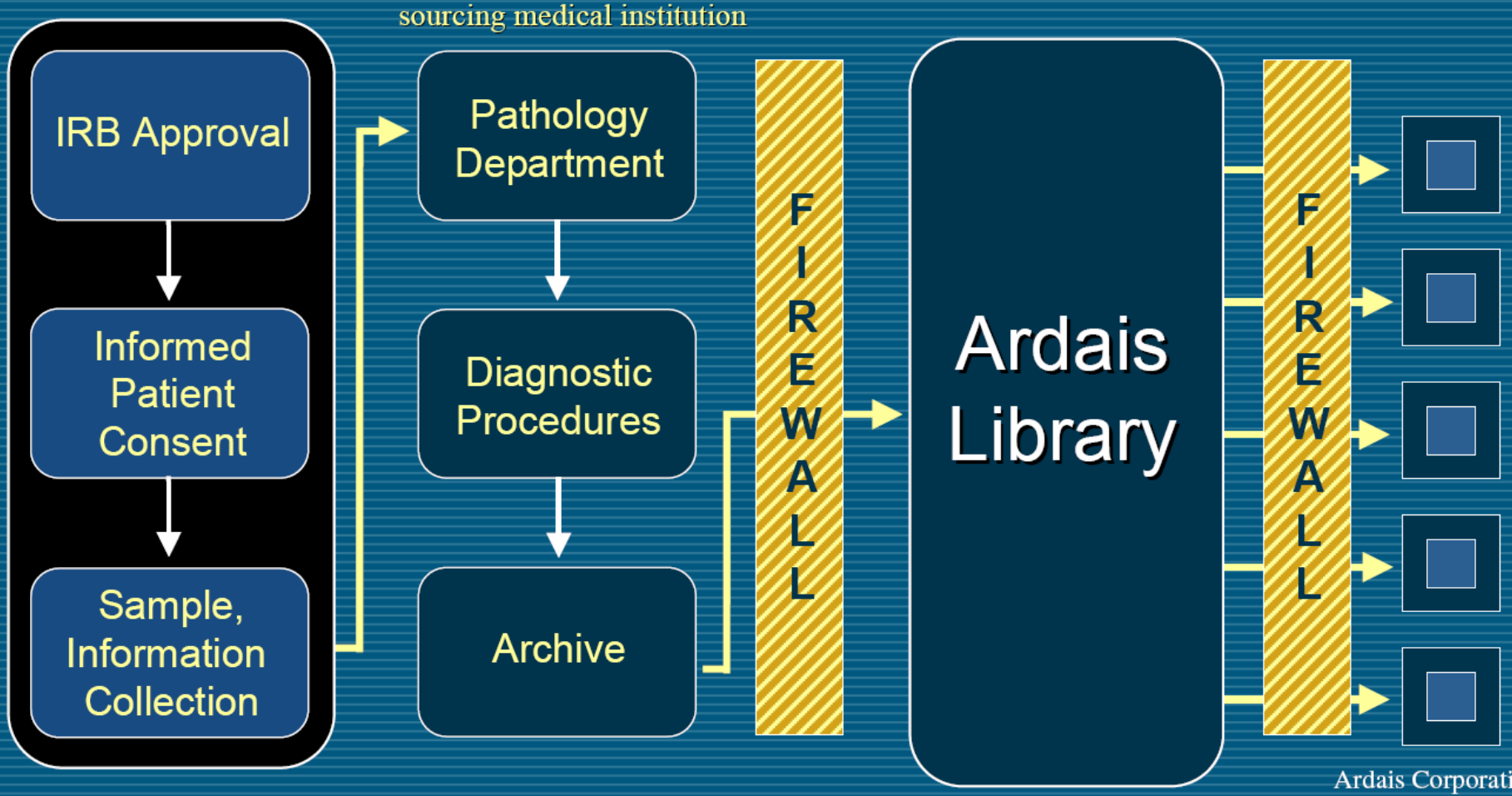
Presentation to ISBER Annual Meeting

May 7, 2002

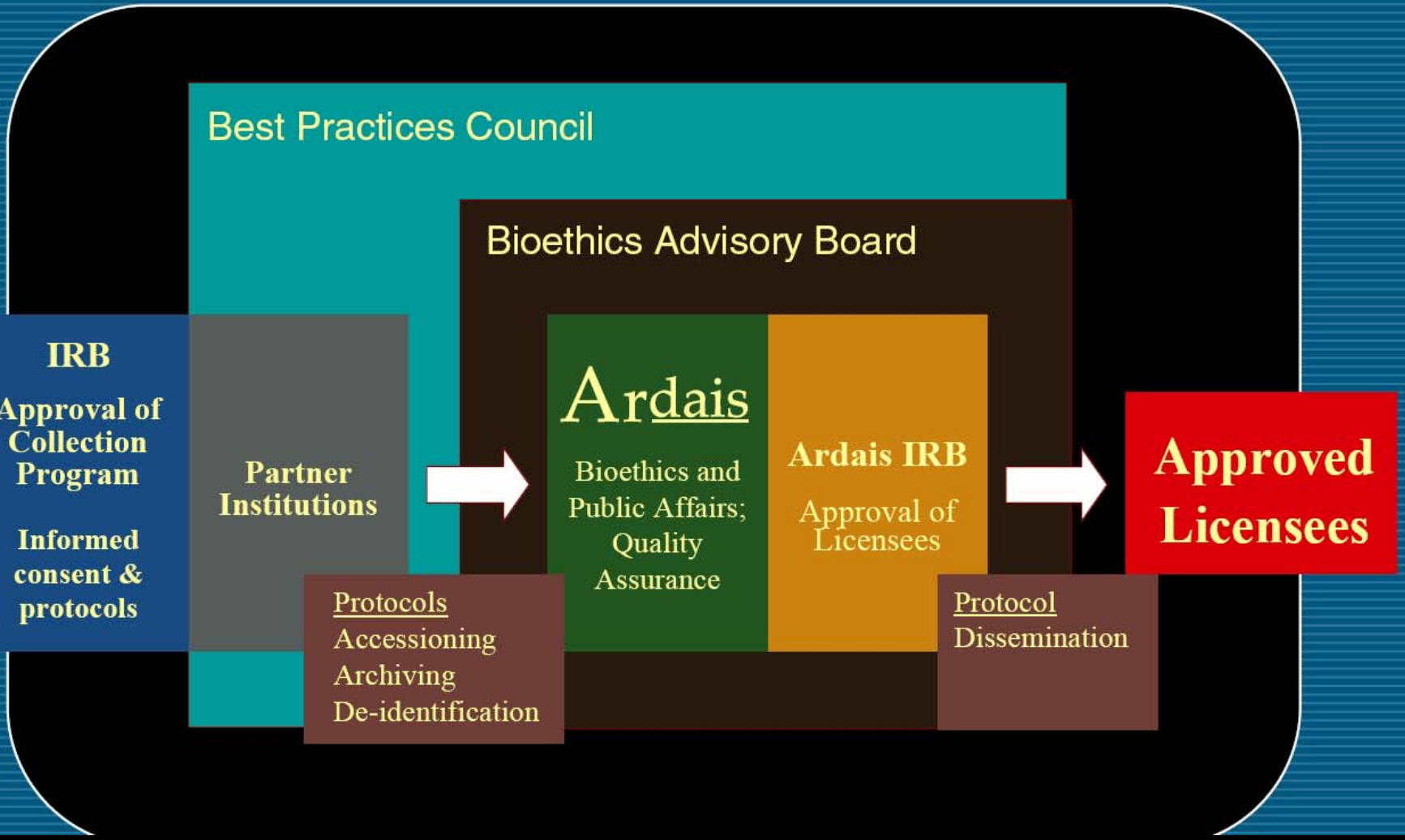
* Establishment of Double Firewall Model

Double Protection:

1. Patient identity never leaves medical institution
2. Licensee does not know patient identity nor the sourcing medical institution



* Establish Ethics Oversight and Mechanism for Continuous Improvement



The New England Journal of Medicine

VOLUME 346 MAY 16, 2002 NUMBER 20

HIGH BONE DENSITY DUE TO A MUTATION IN LDL-RECEPTOR–RELATED PROTEIN 5

LYNN M. BOYDEN, PH.D., JUNHAO MAO, PH.D., JOSEPH BELSKY, M.D., LYLE MITZNER, M.D., ANITA FARHI, R.N., MARY A. MITNICK, PH.D., DIANQING WU, PH.D., KARL INSOGNA, M.D., AND RICHARD P. LIFTON, M.D., PH.D.

ABSTRACT

Background Osteoporosis is a major public health problem of largely unknown cause. Loss-of-function mutations in the gene for low-density lipoprotein receptor–related protein 5 (*LRP5*), which acts in the Wnt signaling pathway, have been shown to cause osteoporosis–pseudoglioma.

Methods We performed genetic and biochemical analyses of a kindred with an autosomal dominant syndrome characterized by high bone density, a wide

OSTEOPOROSIS is a major public health problem, and its prevalence is increasing.[1-3] In the United States, nearly 1 million fractures occur annually in people over the age of 65 years, the majority of which are due to osteoporosis.[1,4] Osteoporotic fractures are associated with substantial morbidity, and the estimated rate of death in the first year after a hip fracture is 25 to 30 percent.[5,6]

Bone mass, a major determinant of the risk of os-

es an autosomal dominant disorder characterized by high bone density, torus palatinus, and a wide, deep mandible.

Our in vitro and in vivo studies show that the $LRP5_{V171}$ mutation increases Wnt signaling. The mutation impairs antagonism of Wnt signaling by Dkk-1 in vitro, and the levels of fibronectin, a downstream target of Wnt signaling, are increased in vivo in patients with this mutation. These findings indicate that unopposed Wnt signaling due to loss of action of a

gle propeller of the low-density lipoprotein (LDL) receptor in humans, mice, rats, pigs, hamsters, and rabbits. Moreover, glycine is also found at this position in the first propeller of the *Drosophila melanogaster* LDL-receptor–related protein homologue, *arrow*. In addition, glycine is present at this position in a wide range of other YWTD propellers, including those in other LDL-receptor–related proteins, as well as those in the epidermal growth factor precursor, the very-low-density lipoprotein receptor, and the vitellogenin receptor in fruit flies and mosquitos (protein sequences are available at http://www.ncbi.nlm.nih.gov/entrez). The evolutionary conservation of this glycine residue is strong evidence of the functional importance of its mutation in our kindred.

Protein sequences are available at

ENTREZ

Molecular Studies

If this mutation indeed causes gain of LRP5 function and increased Wnt signaling, downstream target genes in the Wnt signaling pathway should show increased expression in vivo. A direct transcriptional target of Wnt signaling is the extracellular matrix protein fibronectin.[31] Fibronectin levels were markedly elevated in the affected members of our kindred, with

It is striking that the same mutation is associated with nonsyndromic high bone mass in one family and syndromic high bone mass in the other. These findings suggest that alleles of other genes or environmental factors influence phenotypic manifestations of the mutation and that other phenotypes in kindreds with autosomal dominant high bone mass may also arise from

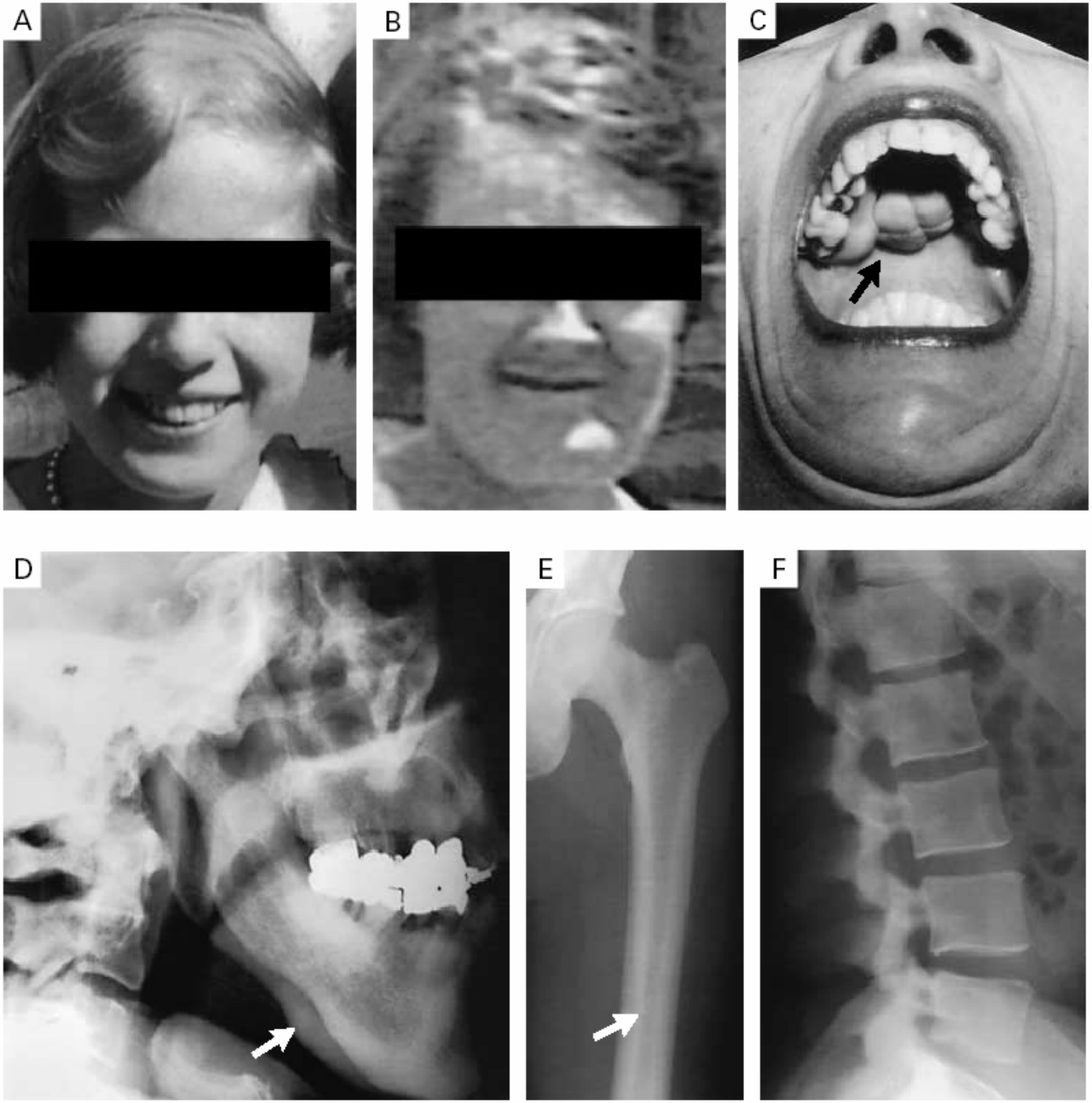
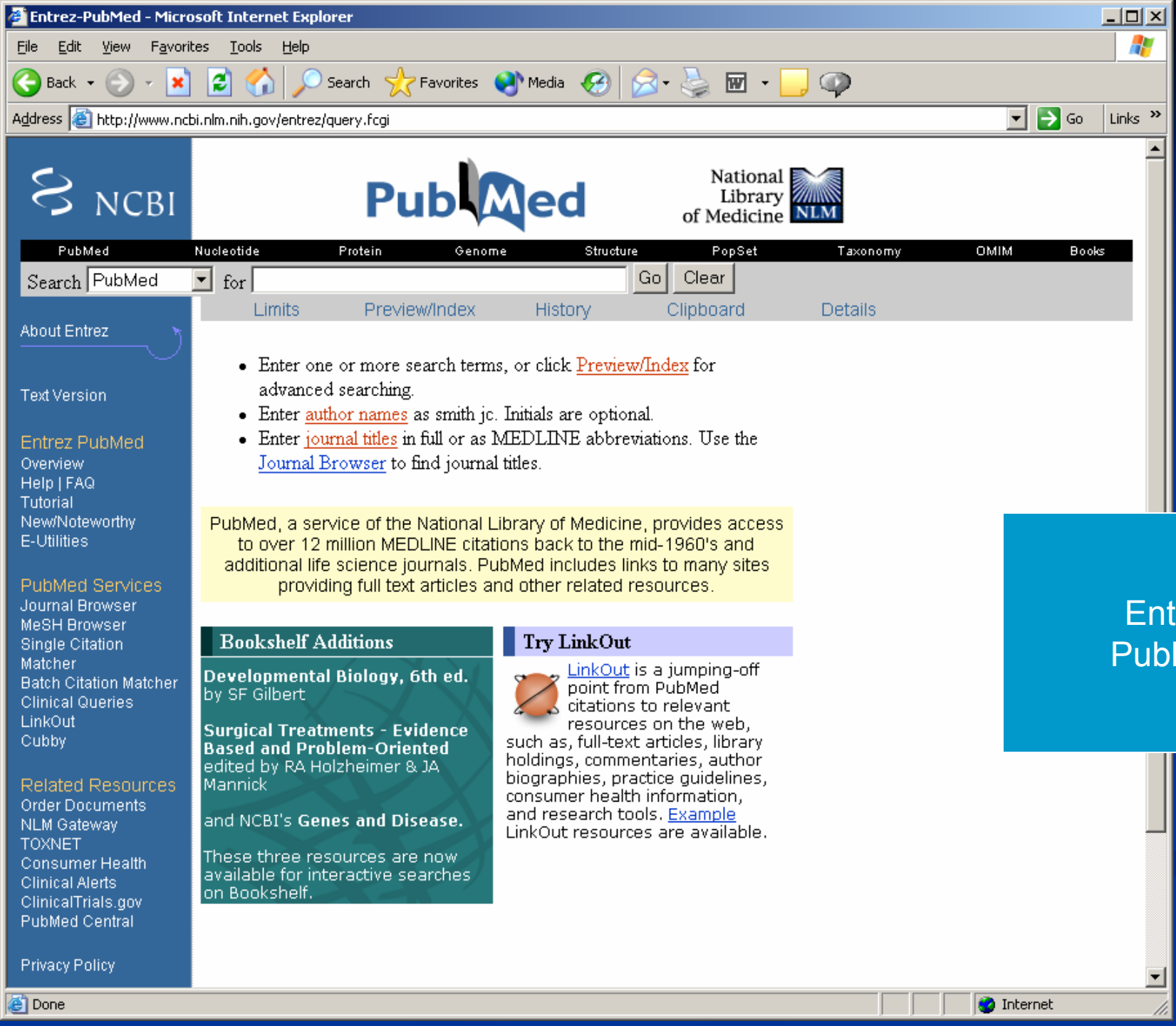


Figure 1. Clinical and Radiographic Features of Affected Members of the Kindred.

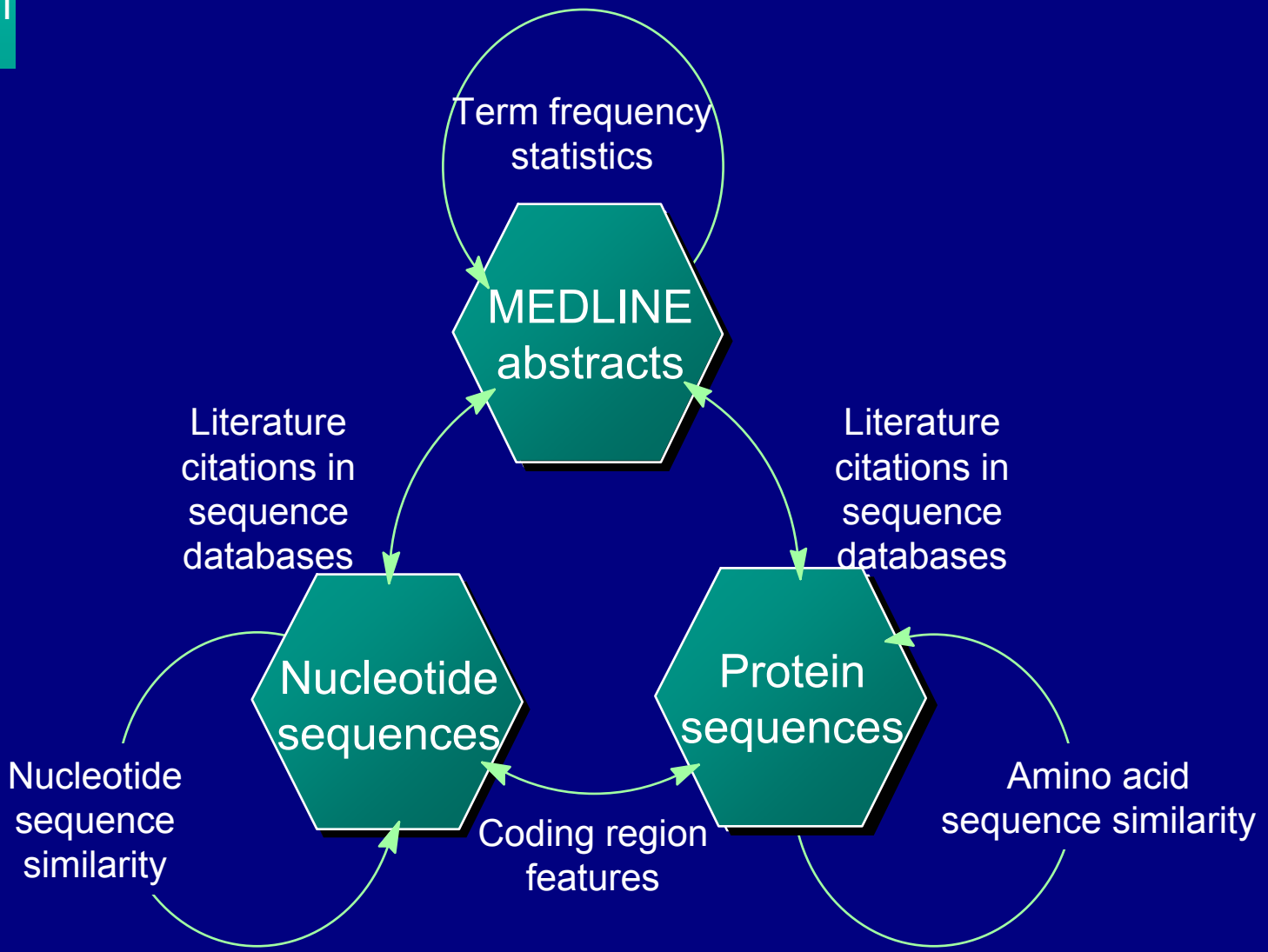
Figure 1

Page 1515

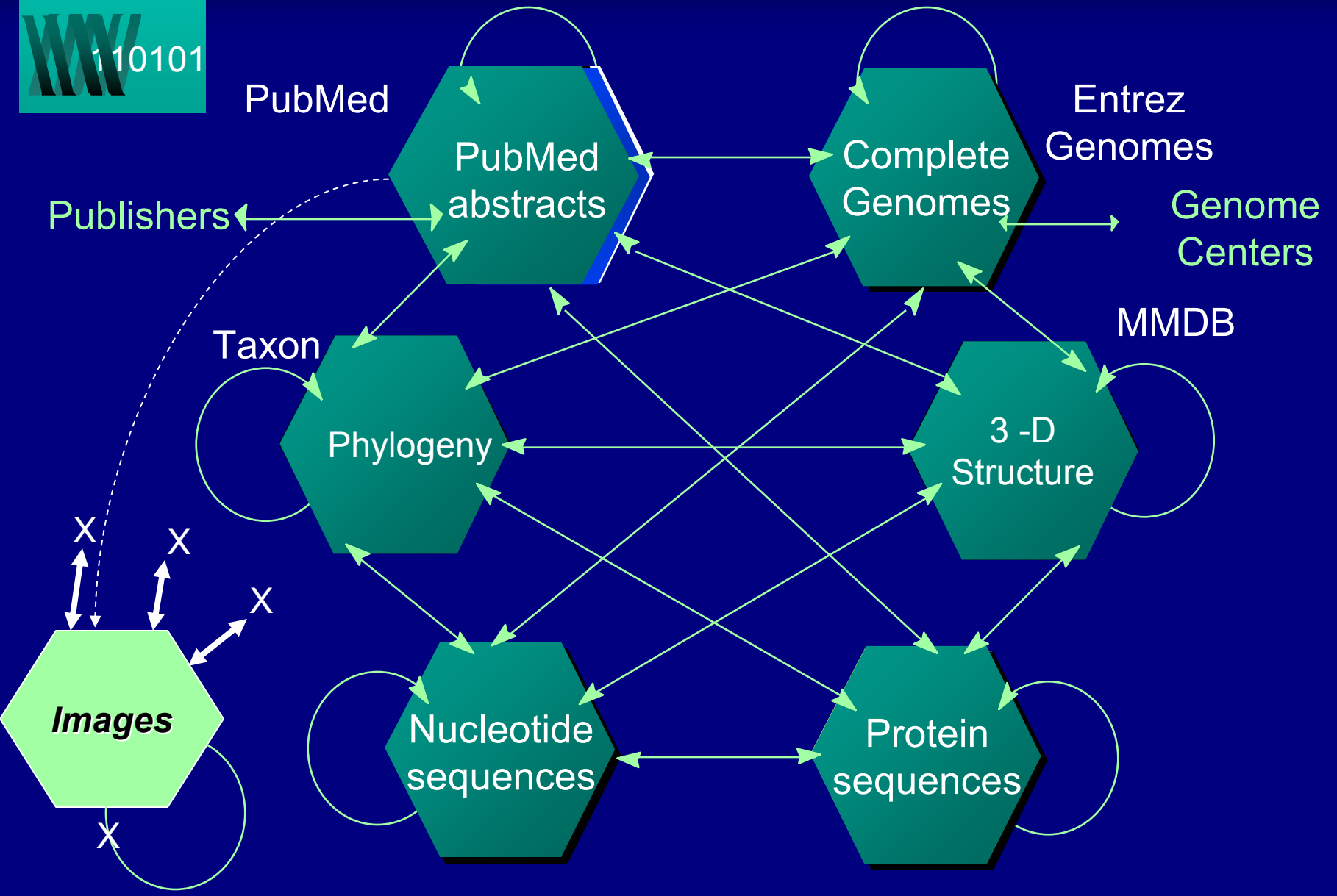


Entrez PubMed

Entrez: Pathway to Discovery



Entrez Increases Discovery Space



Issues

- Clinical trials sponsored by NIH and industry (for FDA review) are intrinsically different
- Inconsistencies in microarray studies of gene expression patterns across laboratories
- Population studies are essential, but USA is very diverse and resistant to central collection of specimens and medical data
- Linking diverse sources of clinical and biological information on a patient-by-patient basis is new
- Key standards organizations such as CDISC and DICOM are not linked
- Public-private collaboration may accelerate progress with “open” systems

Challenges

- Preservation of clinical record collections (long term databases)
- Access to archived clinical records, imaging, genomic/proteomic and related data
- Tools that augment human performance (and provide measurable benefit), informed by current and archived data
- Transparency; ease of use; quality of service; validity; reproducibility

Recommendations

- Persistence and enthusiasm toward a focused community (? Clinical decision making ?) is essential
  - Akin to the cell signaling community, for example
- Ingredients are available (caBIG, BIRN, NECTAR, DICOM, CDISC, external consortia)
  - These and other demonstration projects must succeed to ensure future support
  - Emphasize benefits to individuals
- Get buy-in of industry (medical instruments / imaging, pharma), gov't agencies (FDA, NSF, NIST, DHHS, …)
- Humans are central: augment observers, human-computer interaction, ethics and incentives to participants, patient advocates