open.michigan

Author(s): Brian Athey

License: Unless otherwise noted, this material is made available under the terms of the **Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License**: http://creativecommons.org/licenses/by-nc-sa/3.0/

We have reviewed this material in accordance with U.S. Copyright Law **and have tried to maximize your ability to use, share, and adapt it.** The citation key on the following slide provides information about how you may share and adapt this material.

Copyright holders of content included in this material should contact **open.michigan@umich.edu** with any questions, corrections, or clarification regarding the use of content.

For more information about **how to cite** these materials visit http://open.umich.edu/education/about/terms-of-use.

Any **medical information** in this material is intended to inform and educate and is **not a tool for self-diagnosis** or a replacement for medical evaluation, advice, diagnosis or treatment by a healthcare professional. Please speak to your physician if you have questions about your medical condition.

Viewer discretion is advised: Some medical content is graphic and may not be suitable for all viewers.

Citation Key

for more information see: http://open.umich.edu/wiki/CitationPolicy

Use + Share + Adapt

{ Content the copyright holder, author, or law permits you to use, share and adapt. }

PD-GOV **Public Domain – Government**: Works that are produced by the U.S. Government. (17 USC § 105)

PD-EXP **Public Domain – Expired**: Works that are no longer protected due to an expired copyright term.

PD-SELF **Public Domain – Self Dedicated**: Works that a copyright holder has dedicated to the public domain.

CC ZERO **Creative Commons – Zero Waiver**

CC BY **Creative Commons – Attribution License**

CC BY-SA **Creative Commons – Attribution Share Alike License**

CC BY-NC **Creative Commons – Attribution Noncommercial License**

CC BY-NC-SA **Creative Commons – Attribution Noncommercial Share Alike License**

GNU-FDL **GNU – Free Documentation License**

Make Your Own Assessment

{ Content Open.Michigan believes can be used, shared, and adapted because it is ineligible for copyright. }

PD-INEL **Public Domain – Ineligible**: Works that are ineligible for copyright protection in the U.S. (17 USC § 102(b)) *laws in your jurisdiction may differ

{ Content Open.Michigan has used under a Fair Use determination. }

© FAIR USE **Fair Use**: Use of works that is determined to be Fair consistent with the U.S. Copyright Act. (17 USC § 107) *laws in your jurisdiction may differ

Our determination **DOES NOT** mean that all uses of this 3rd-party content are Fair Uses and we **DO NOT** guarantee that your use of the content is Fair.

To use this content you should **do your own independent analysis** to determine whether or not your use will be Fair.

University of Michigan
Health System

How Bioinformatics is Transforming Biomedical Research and Practice

Brian Athey
Professor and Chair
Department of Computational Medicine and Bioinformatics
Professor of Psychiatry
Associate Director, Michigan Institute for Clinical and Health Research
University of Michigan Medical School

Disclosure Information:

Clinical Research Forum IT Roundtable 2012

Name of Speaker: Brian D. Athey, PhD

I have the following financial relationships to disclose:

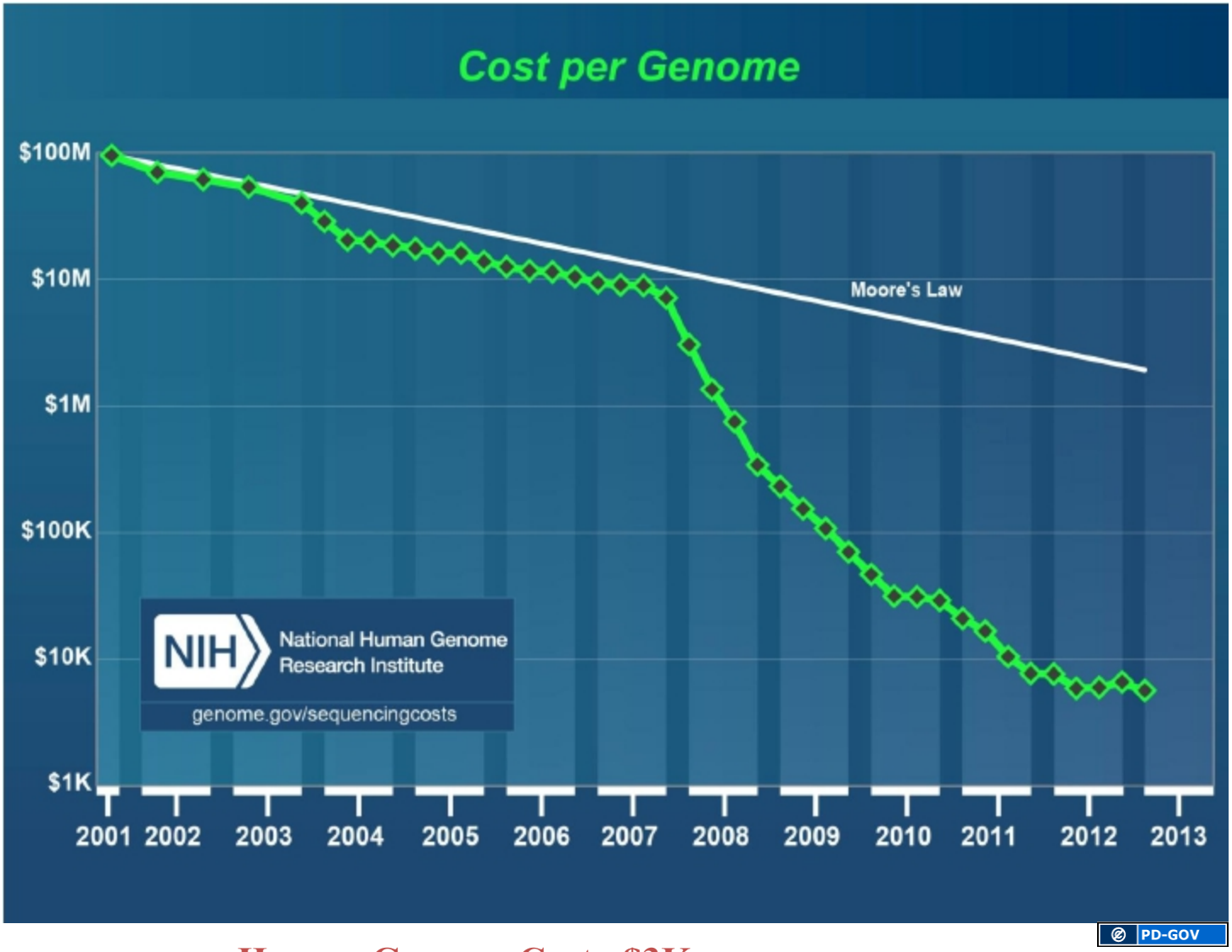
Employee: University of Michigan
Board of Directors: tranSMART Foundation (NFP); Scientists and Engineers for America (NFP)
Consultant for (Scientific Advisory Board): Appistry, Inc. (St. Louis, MO); Biovest International (Tampa, Fl.); AssureRx Health (Mason, Ohio)
Speaker's Bureau for: none
Grant/Research support from: National Institutes of Health
Stockholder in: All for profit companies named above
Honoraria from: none

X **I will not discuss off label use and/or investigational use in my presentation.**

Vision of Biology as an Information Science: Key Components to Discuss (Omenn & Athey, 2010)

- An avalanche of molecular information: NGS sequence data, validated SNPs, haplotype blocks, candidate genes/ alleles, exome sequences, microarray data, epigenomics data, proteins, and metabolites—to be associated with disease risks
- Powerful computational methods
- Effective linkages with better environmental, dietary, and behavioral datasets for eco-genetic analyses
- Credible privacy and confidentiality protections in research and clinical care
- Breakthrough tests, vaccines, drugs, behaviors, and regulatory actions to reduce health risks and cost-effectively treat patients globally
- Novel data integration methods to understand and address health disparities and emerging public health threats

The Cost of DNA Sequencing is Dropping



Human Genome Cost ~$3K

http://www.genome.gov/

Human genome will be a part of every patient's medical record in 10 years

- A human genome will cost perhaps $100 in 5 years.
- Genomes should be sequenced in families
- Societal and scientific objections will fade driven by actionable gene variants
- Your genome will be checked yearly for newly discovered actionable gene variants

Lee Hood IOM February 27, 2012

Institute for Systems Biology
Revolutionizing Science. Enhancing Life.

Personal "Omics" Profiling (POP)

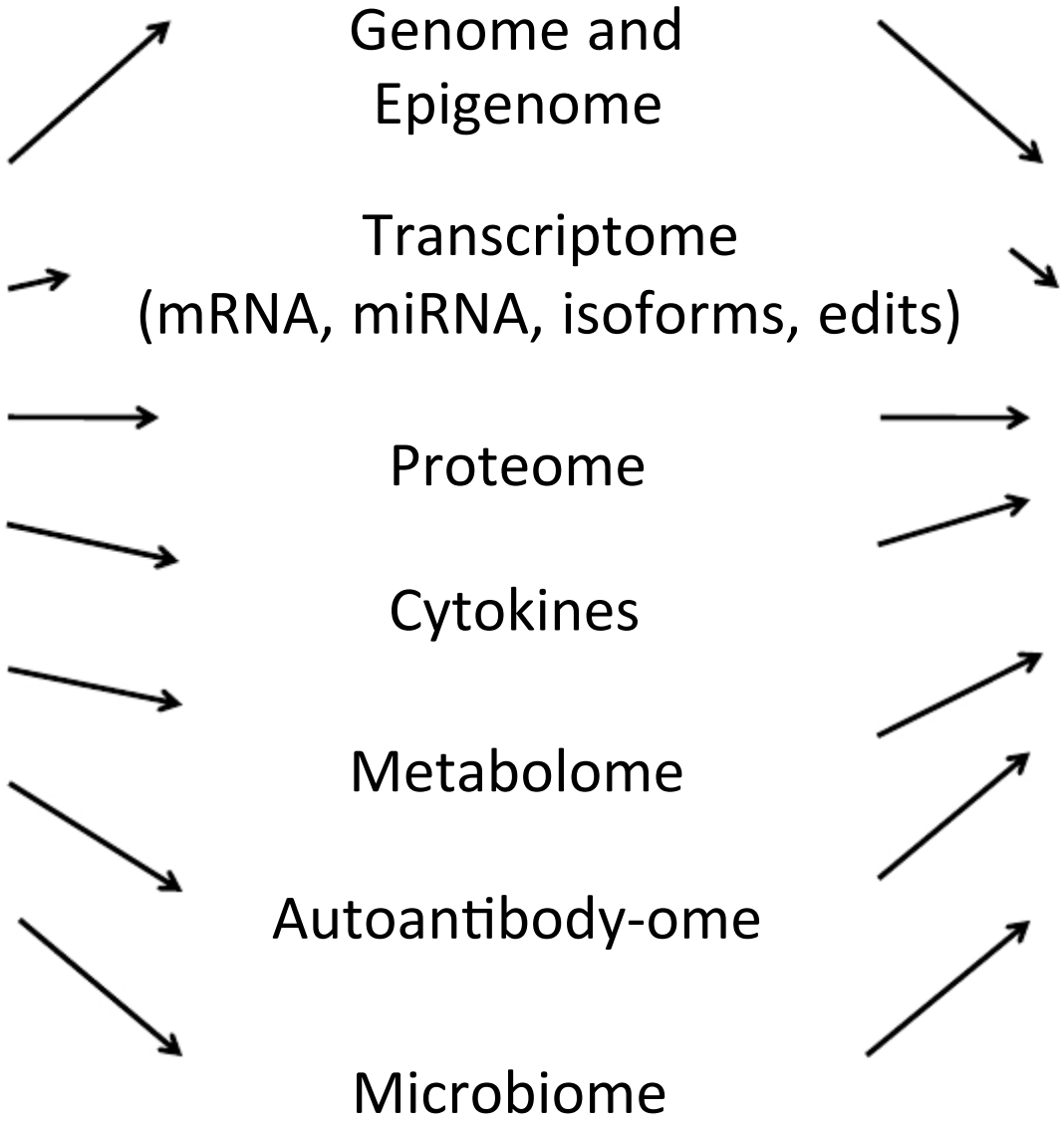
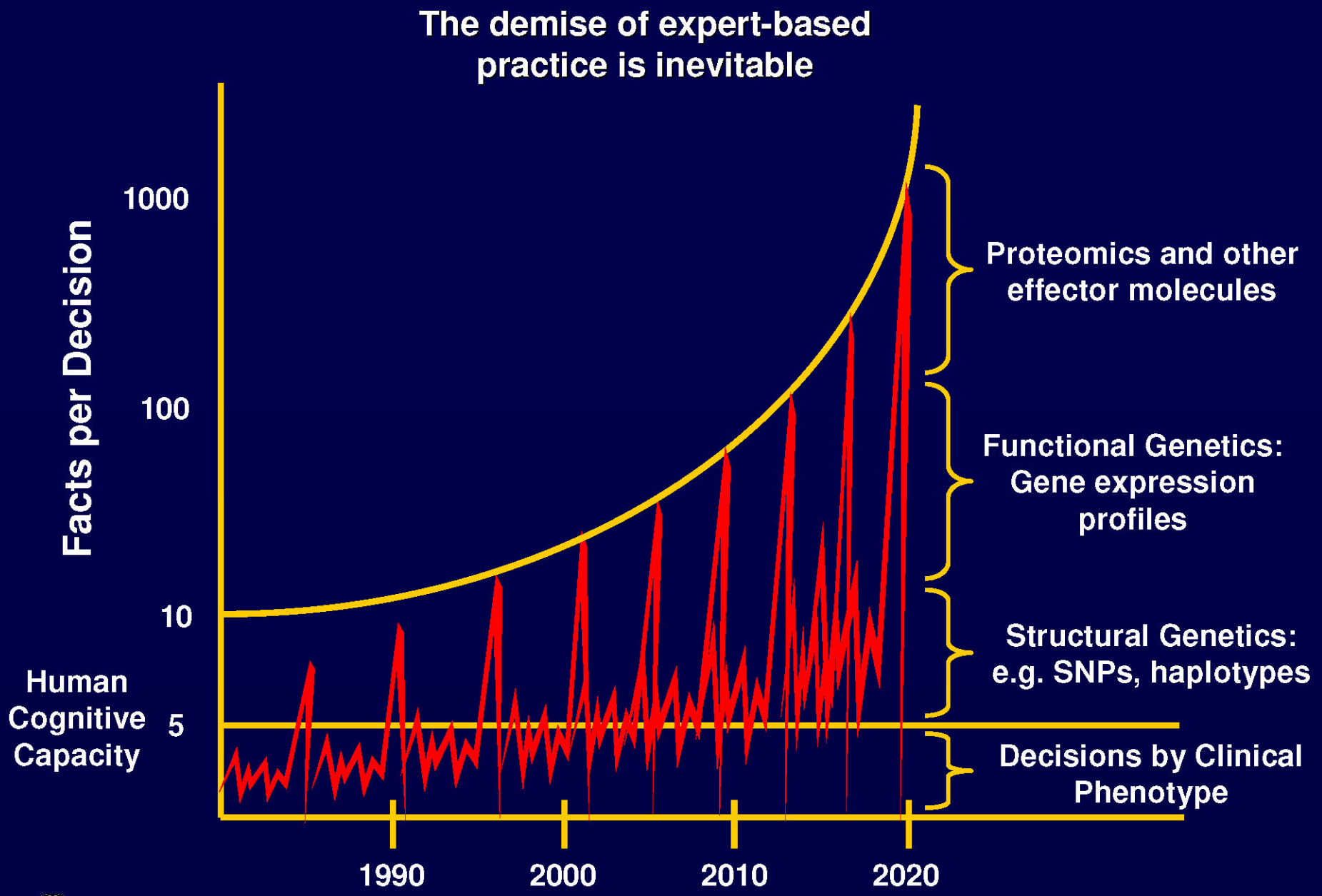


Image Removed - Copyright

Personal Omics Profile


The demise of expert-based practice is inevitable
Facts per Decision
1000
100
10
5
Human Cognitive Capacity
1990
2000
2010
2020
Proteomics and other effector molecules
Functional Genetics: Gene expression profiles
Structural Genetics: e.g. SNPs, haplotypes
Decisions by Clinical Phenotype

PD-INEL

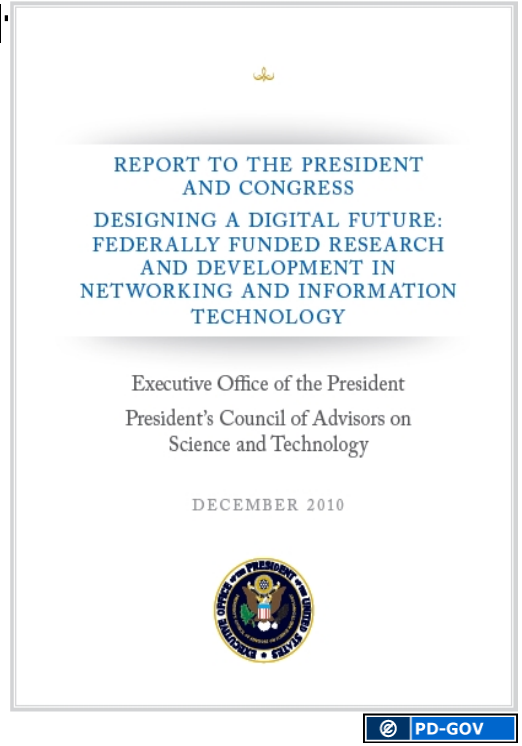
White House PCAST Dec 2010
NITRD Recommendation 3

University of Michigan Health System

"It is recommended that a Dynamic 'Omics Analytics and Data Management Infrastructure for enhanced analysis and standardized interoperability with a Longitudinal Patient-Centric Electronic Health Record (EHR)/Personal Health Record (PHR) be created. This will enable Integration between 'mul
Patient/Research

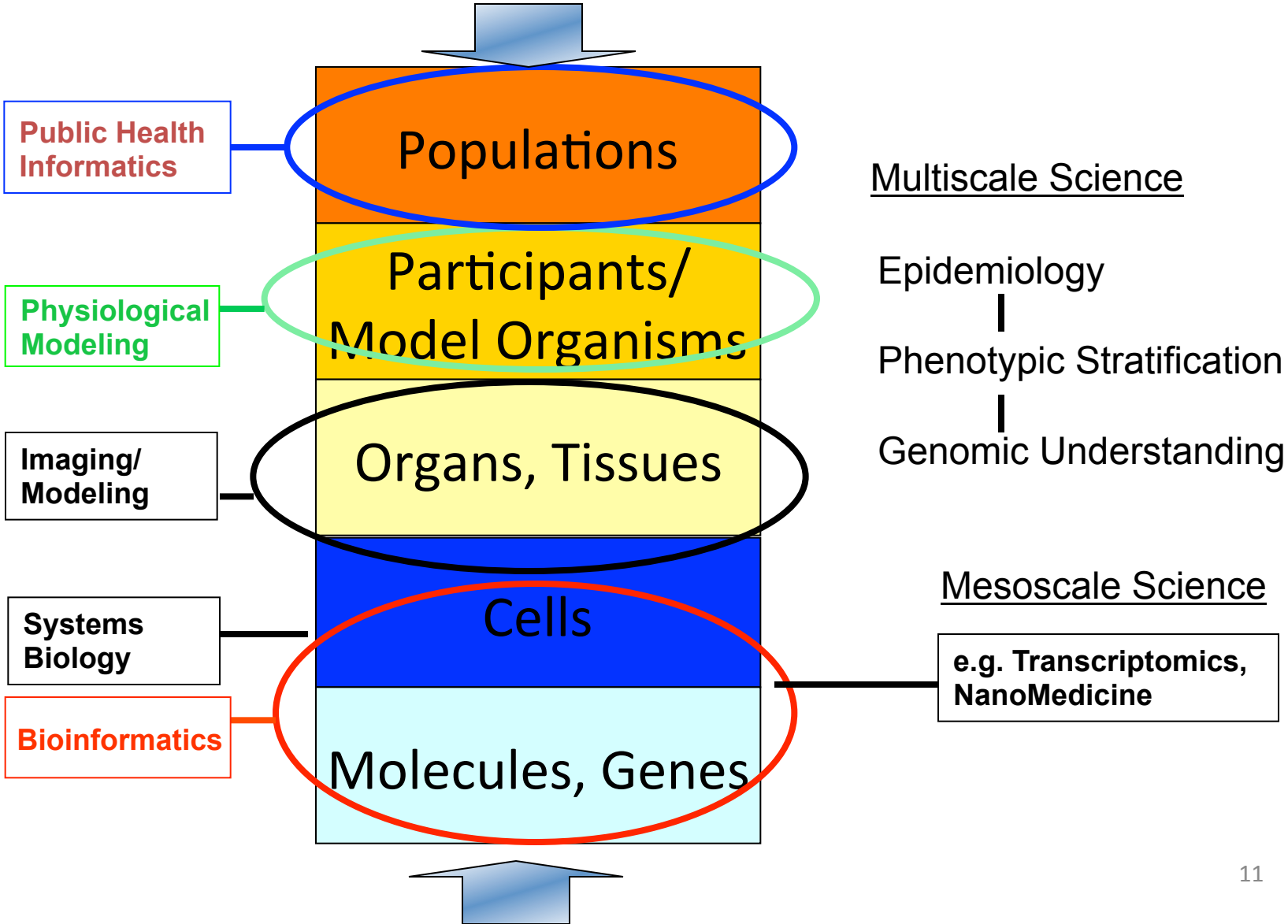
Participant level in EHR:

- **Genomics; Epigenomics; Proteomics; Metabolomics**
- **Pharmacogenomics; Toxicogenomics**
- **Imaging; Cognitive and Behavioral measures; Environmental measures**
- **Secure links to Patient Data in EHR/PHR**
- **Socio-economic measures"**

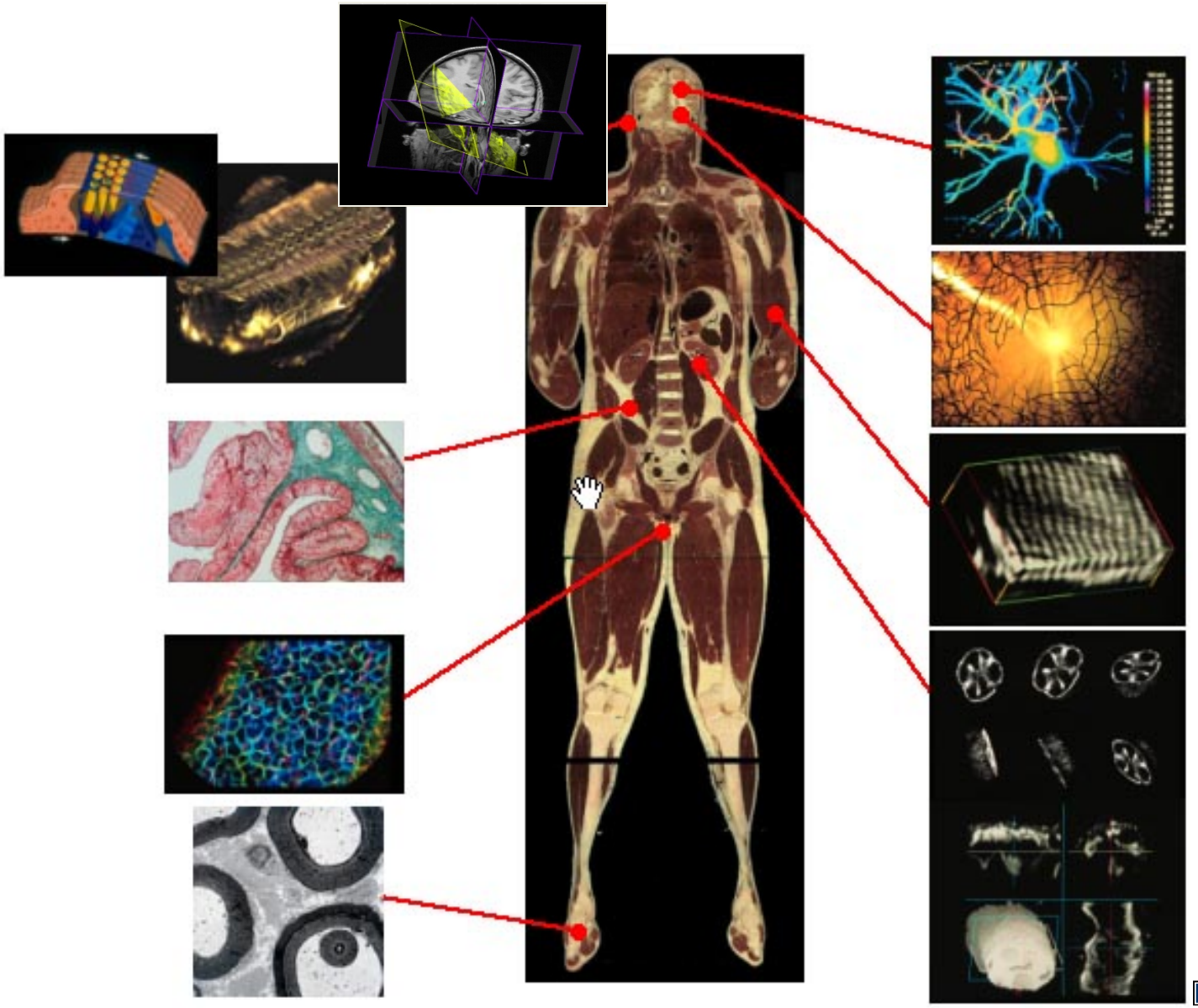


PD-GOV

Integative Informatics Enables Synthesis of Knowledge at Multiple Levels



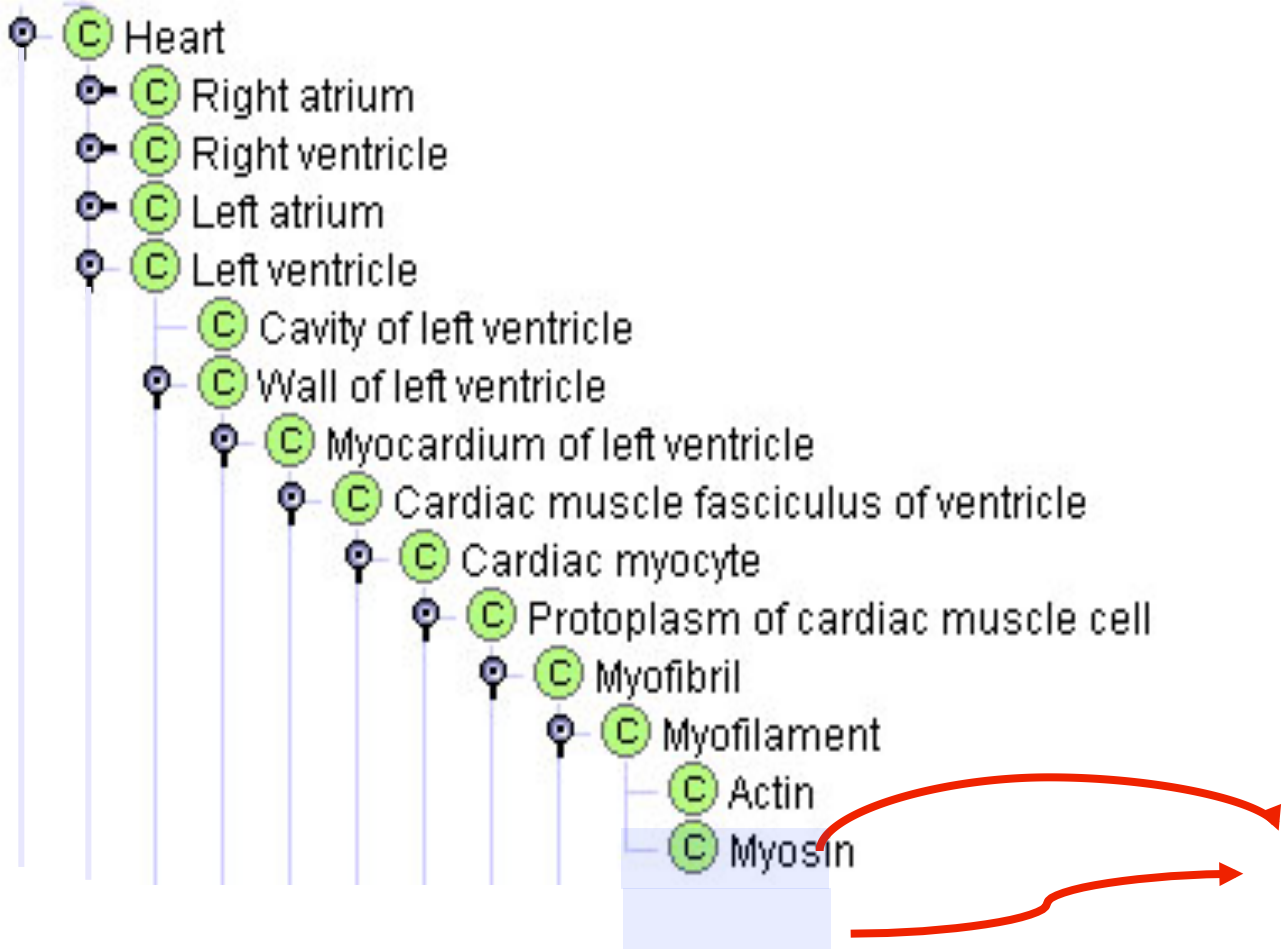
Human Systems Biology is an Emerging Field to Address the Enormous Complexity of the "Physiome"



Foundational Model of Anatomy (Cornelius Rosse)

Multi-scale Human Anatomy

University of Washington



Images Removed - Copyright

Cellular Systems Biology--Overview of the Science

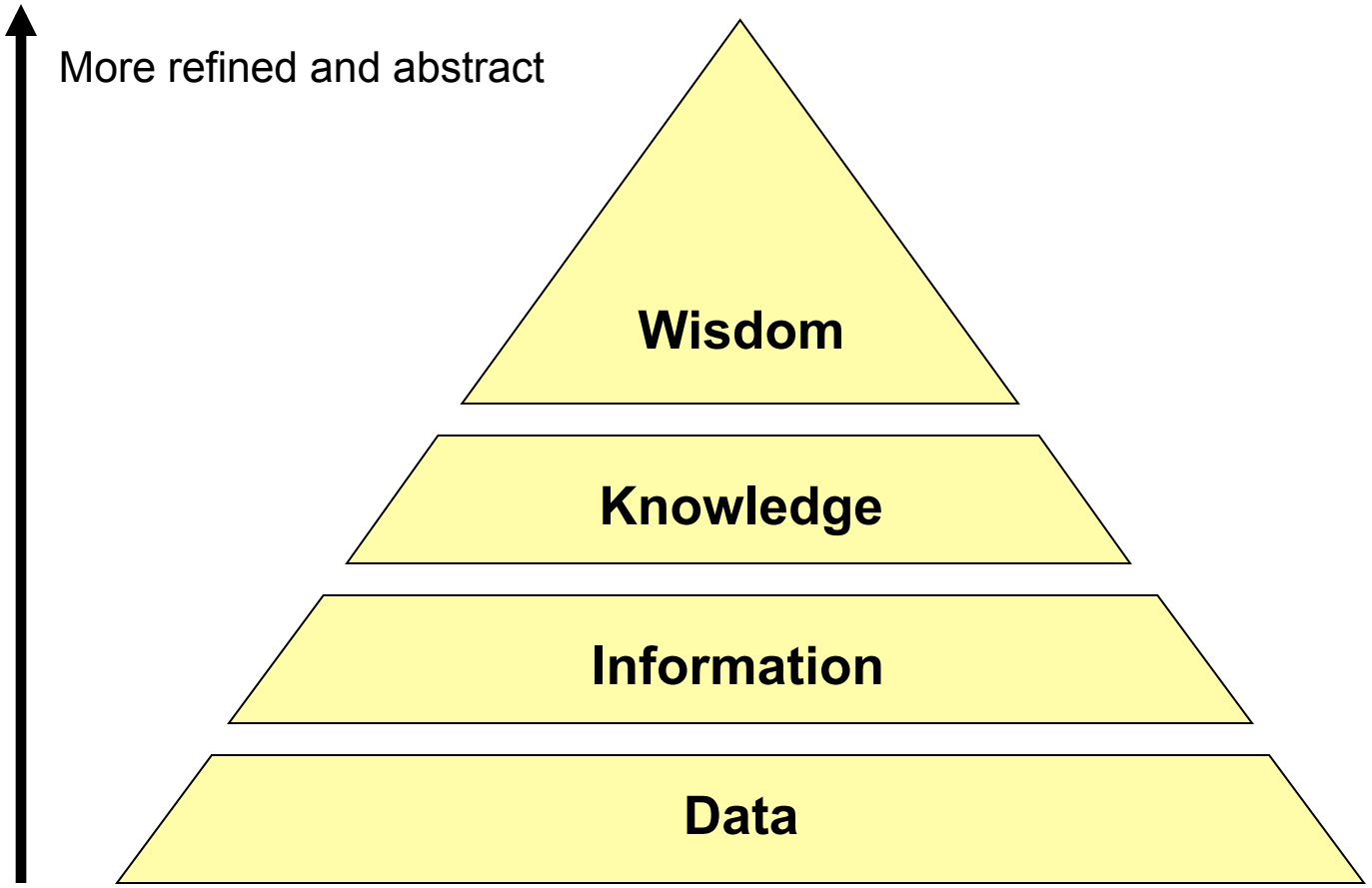
- We are developing a multiscale concept of **Integrated Informatics Framework** to enable **Cellular Systems Biology**
- We seek to **integrate systems at multiple levels:**

  - **Nuclei/Molecular**—Genome/ Epigenome, the "Archive": Sequence, Structure, SNPs, Haplotype, Copy Number Variation, chromatin, epigenomic "marks"--- GWAS (Technique)
  - **Nuclei/Process Regulation**—Transcriptome, a process against the Archive: mRNAs, Global gene expression, transcription factors, splice variants, siRNAs
  - ---
  - **Cytoplasm/Protein Synthesis and Regulation**--Translationome: microRNAs, ribosomal substrate, t-RNAs, Proteome Synthesis
  - ---
  - **Cell(s)**—Organelle, Pathways and Interactome(s), Proteome Localization Metabolome, Lipome
  - ---
  - **Tissue/Environment**, Cellular organization and Tissue Ultrastructures, Environments—e.g. Metabolome, Lipome, Plasma Proteome; Host-Pathogen/symbiotic environments (e.g. Microbiome)
  - ---
  - **Cellular Phenotype(s)—"Physiologic Signatures" Spatial/Temporal/Functional**

Bioinformatics and Computational Biology Transforming Basic Biomedical Science

- Genetics→Genomics
- Biological Chemistry→Pathway Analysis
- Physiology→Multiscaler Computational Systems Biology
- Pharmacology→Computational Systems Pharmacology
- Microbiology→Human Microbiome
- Anatomy→Digital Humans, Digital Histology
- Neuoroscience/Psychiatry→Pharmacogenomics
- Cell and Develomental Biology→Epigenomic Regulation

Information Hierarchy



PD-INEL

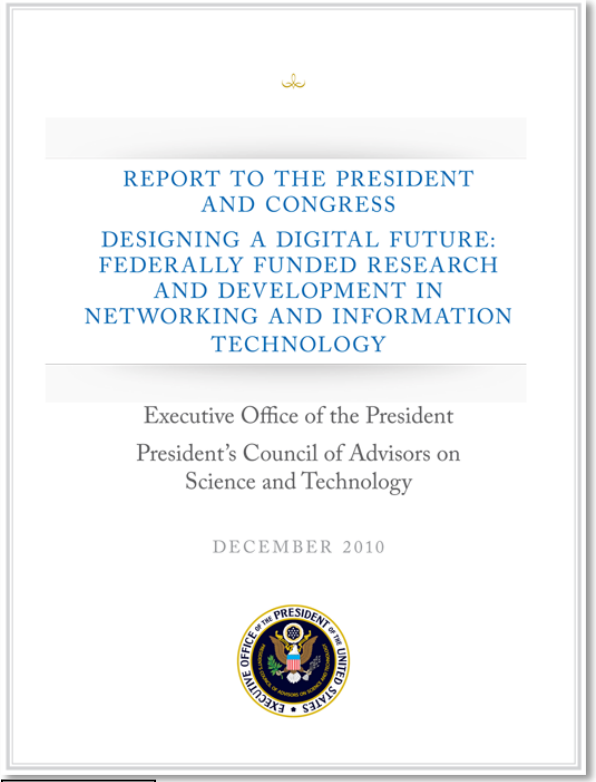
Bruce Schatz, *Telesophy* **1985**

Digital Informatics Hierarchy

- Data
  - The raw material of information
- Information
  - Data organized and presented in a particular manner
  - Metadata
- Knowledge
  - “Justified true belief”
  - Information that can be acted upon
- Wisdom
  - Distilled and integrated knowledge
  - Demonstrative of high-level “understanding”

PCAST NITRD "Big Data"

Strategy Directive



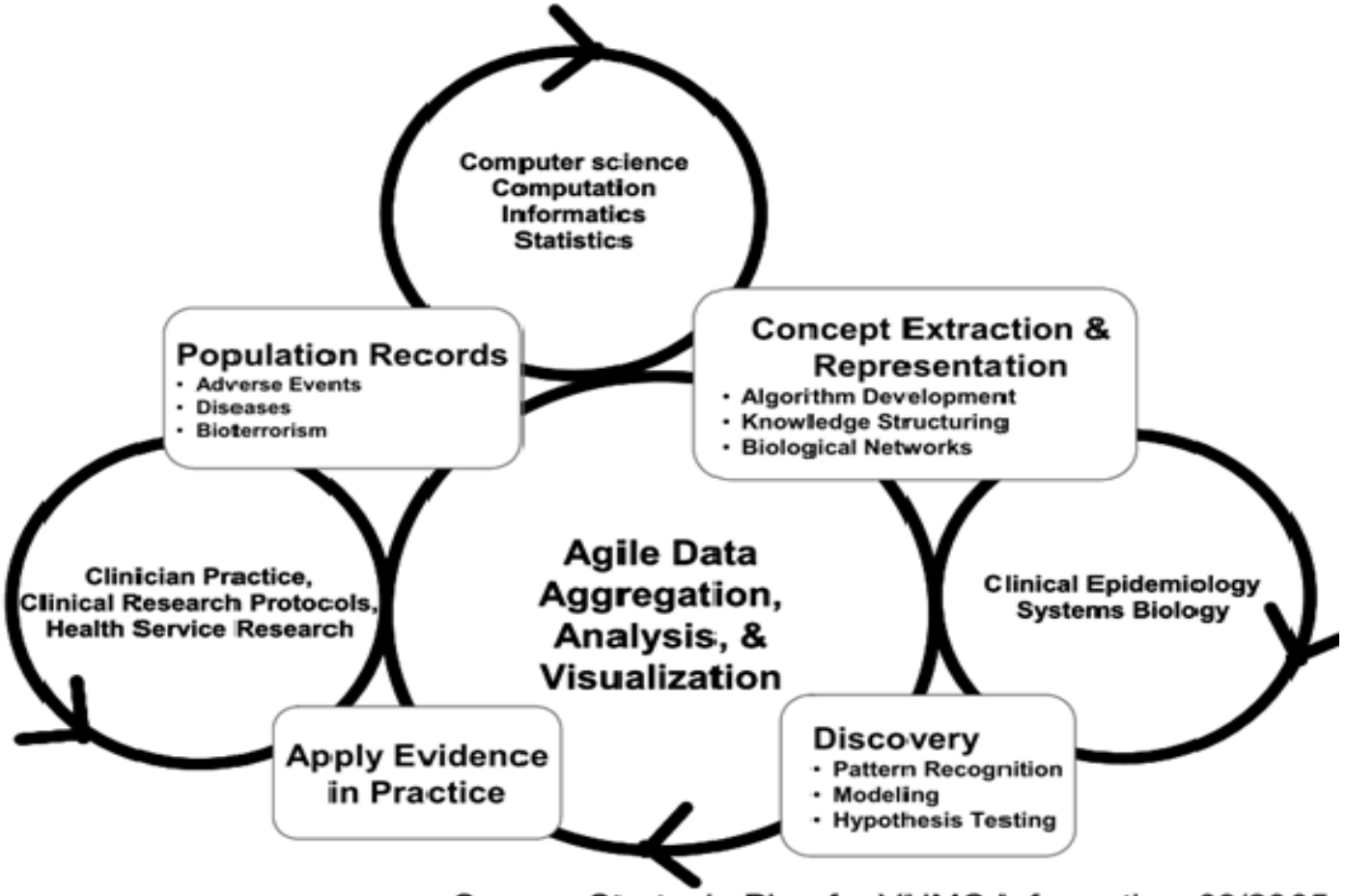
PD-GOV

"Data volumes are growing exponentially"

- **There are many reasons for this growth:**
  - **the creation of nearly all data today in digital form**
  - **a proliferation of sensors (e.g. Next-Generation Sequencing)**
  - **new data sources such as high-resolution imagery and video.**
- **The collection, management, and analysis of data is a fast-growing concern of NIT research.**
- **Automated analysis techniques such as data mining and machine learning facilitate.**
- **Transformation of data into knowledge, and of knowledge into action.**

"Every federal agency needs to have a 'big data' strategy"

Data Aggregation, Integration, Analysis, and Visualization as a Creativity Engine for Biomedical Research and Practice



Source: Strategic Plan for VUMC Informatics, 09/2005

PD-INEL

Routine CWGA:
Gateway to Genomic Medicine

Sample Collection

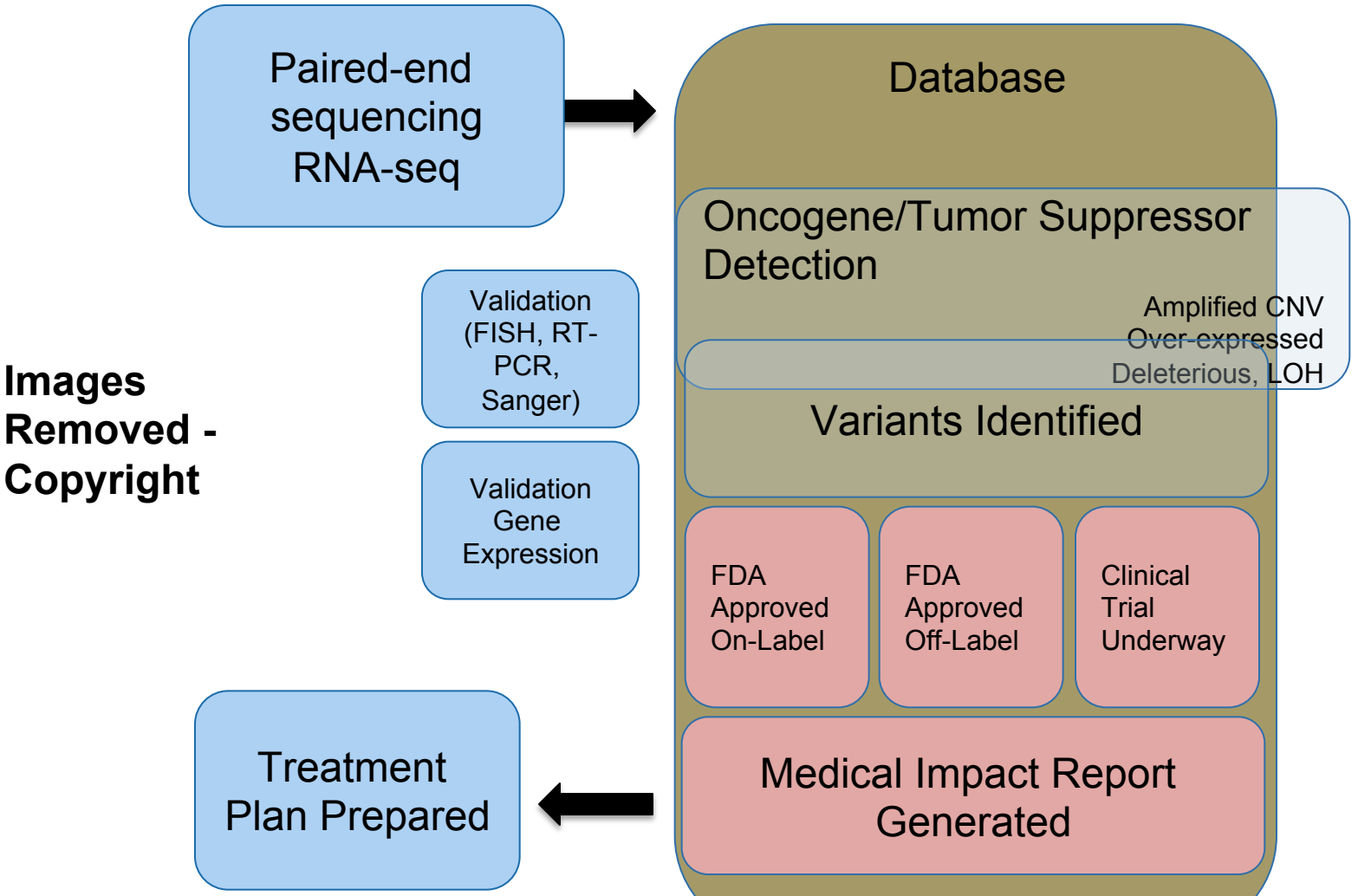
Sequencing

Analysis

Clinical Action

P. Tonnellato, CBMI, HMS

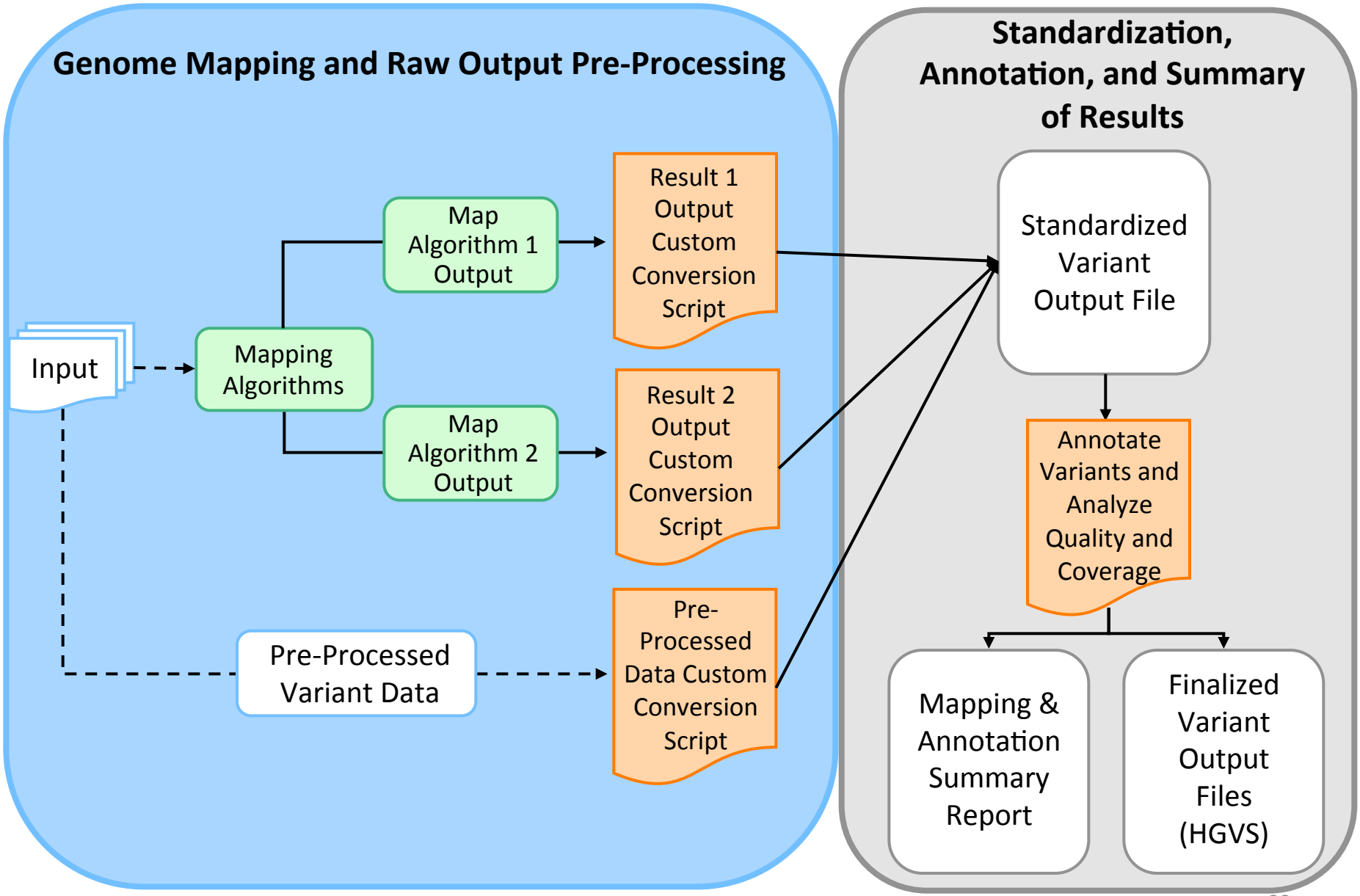
From WGA to Clinical Annotation



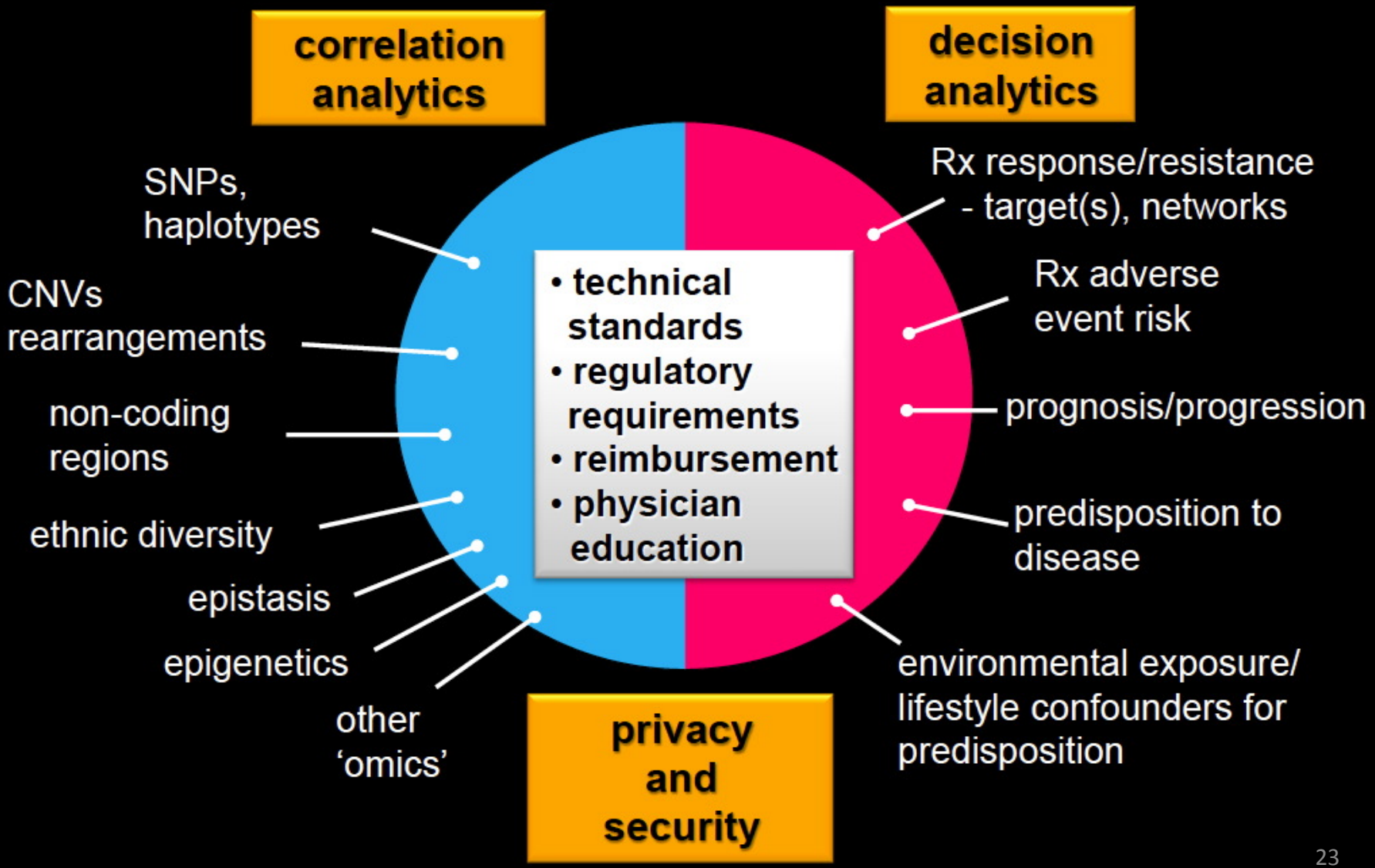
Images Removed - Copyright

D. Wall, CBMI, HMS

Whole Genome Mapping and Variant Annotation Pipeline



Low Cost Whole Genome Sequencing and Molecular Medicine: Dependency on Large Scale (Massive) Data Annotation and Analytics

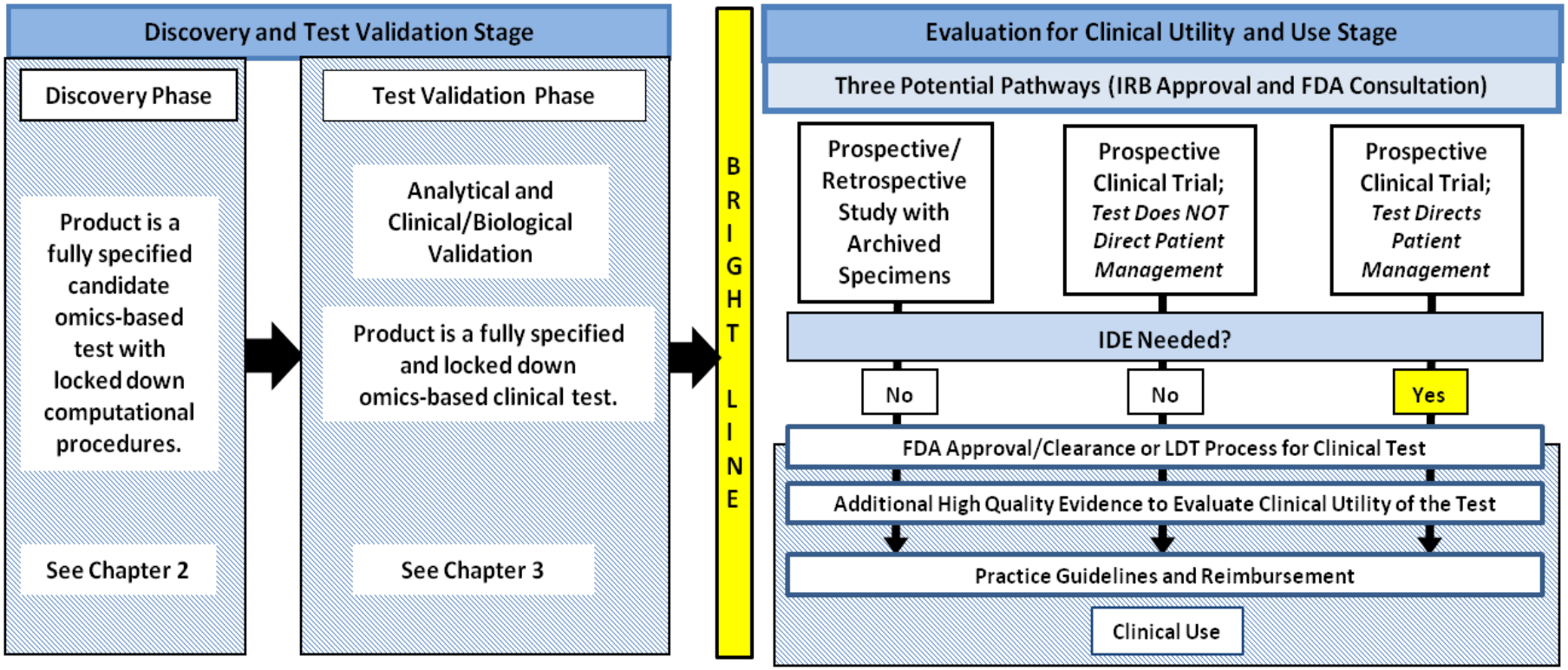


PD-INEL

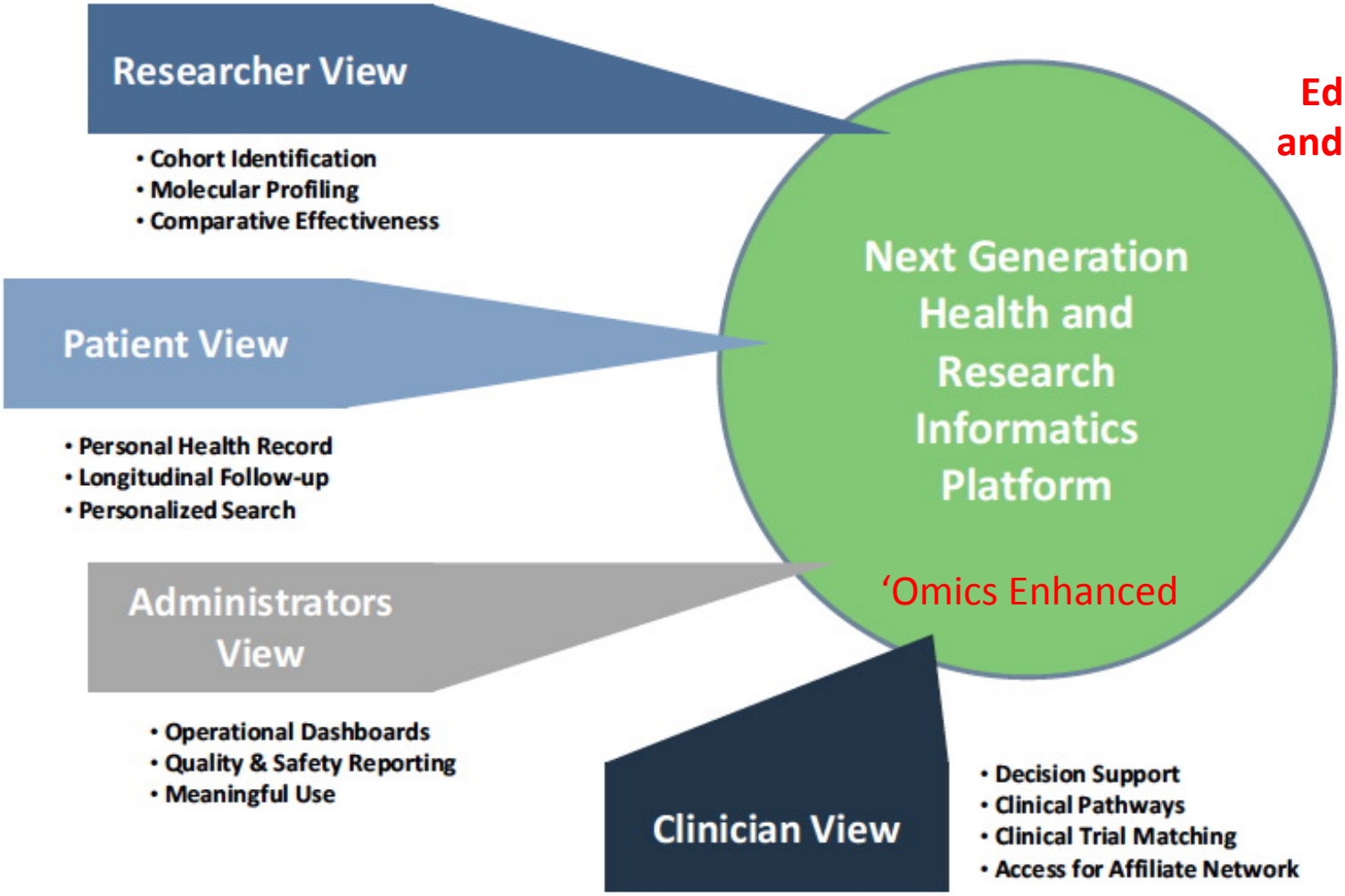
'Omics-Based Test Development Framework IOM March 2012 "Omenn Report"

Image Removed - Copyright

University of Michigan Health System



Five ~~Four~~ Portals to Health/Research Informatics Platform

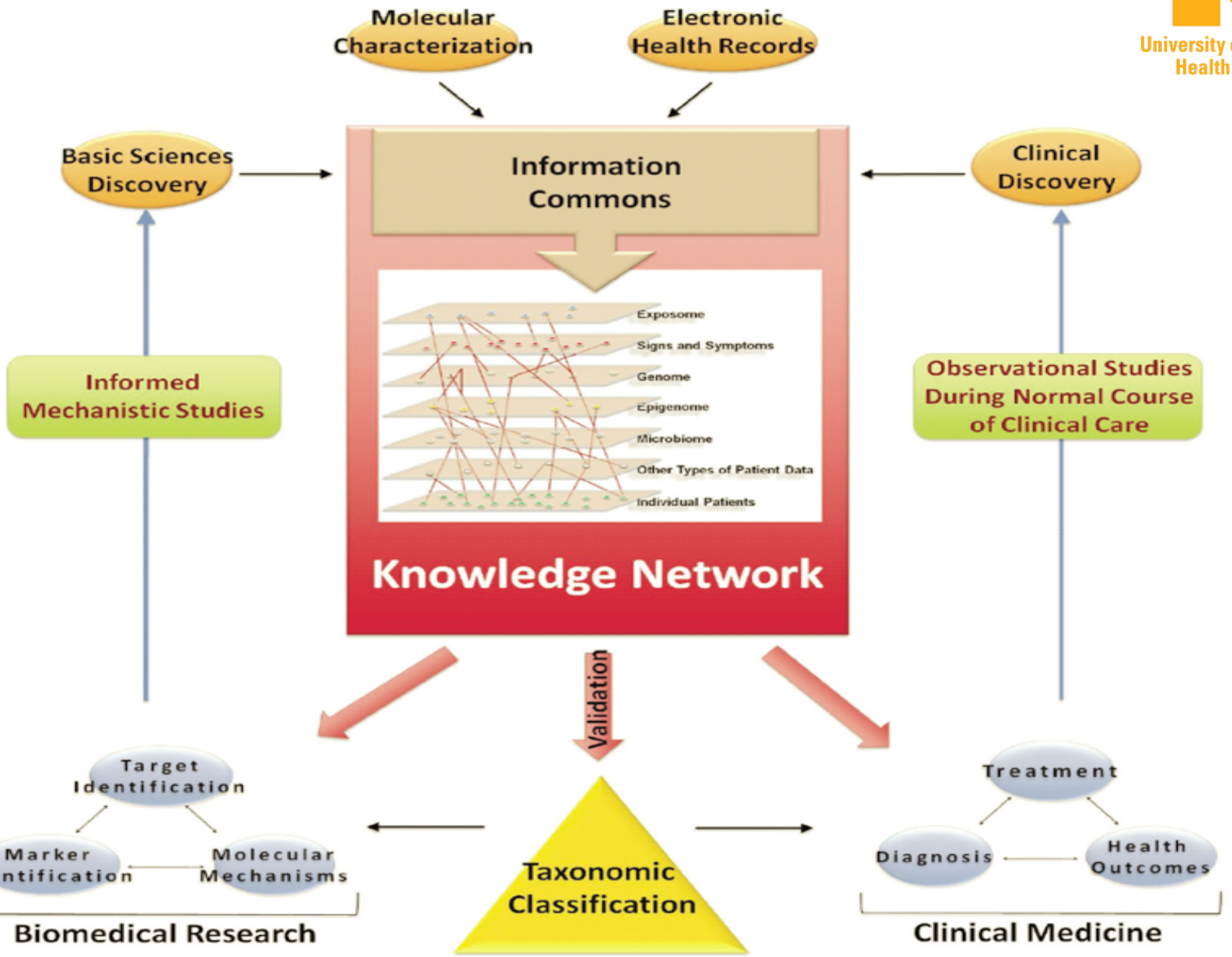


William S. Dalton; Moffitt Cancer Center; IOM Feb. 27, 2012

PD-INEL

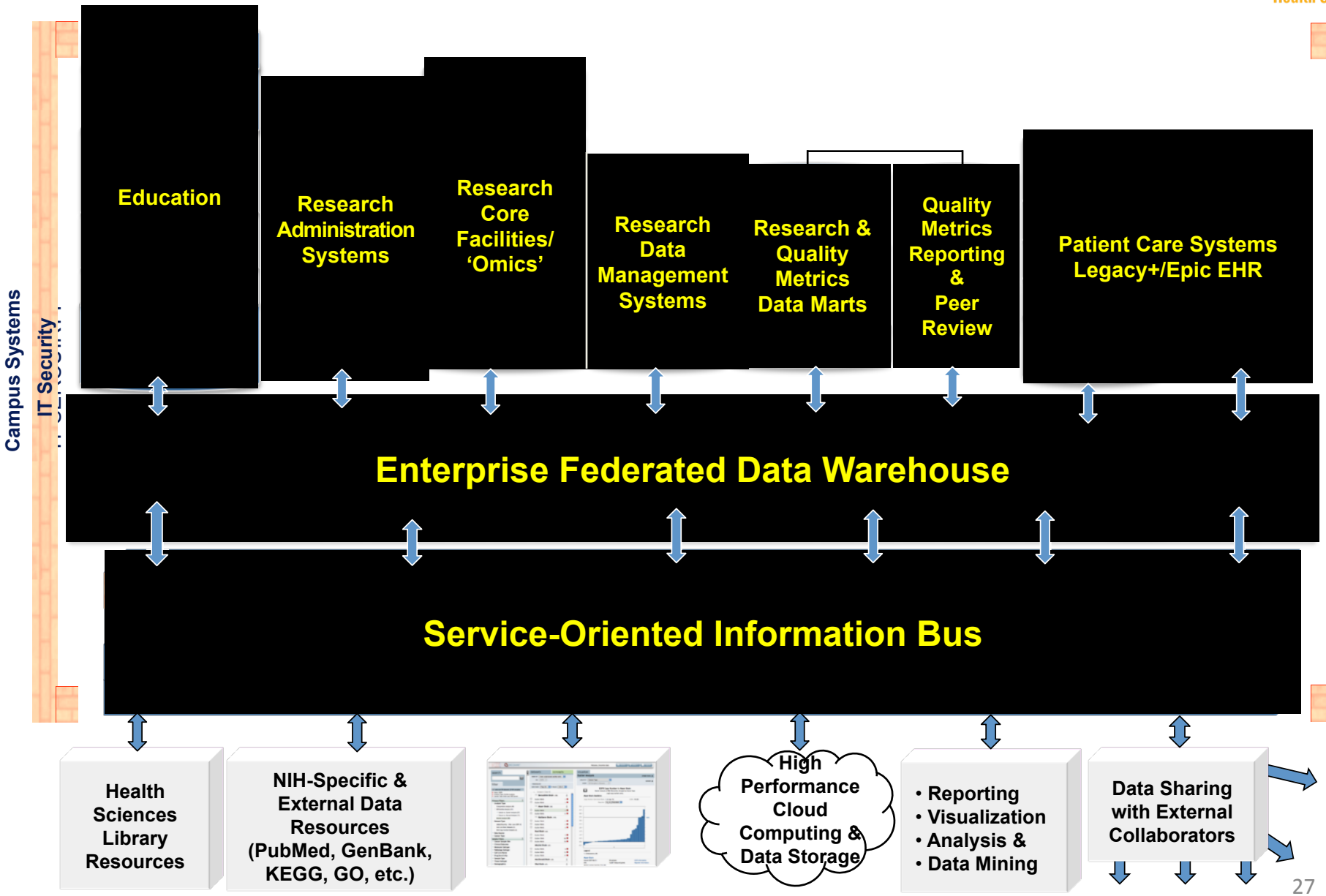
"Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease" (National Academies Report)

University of Michigan Health System

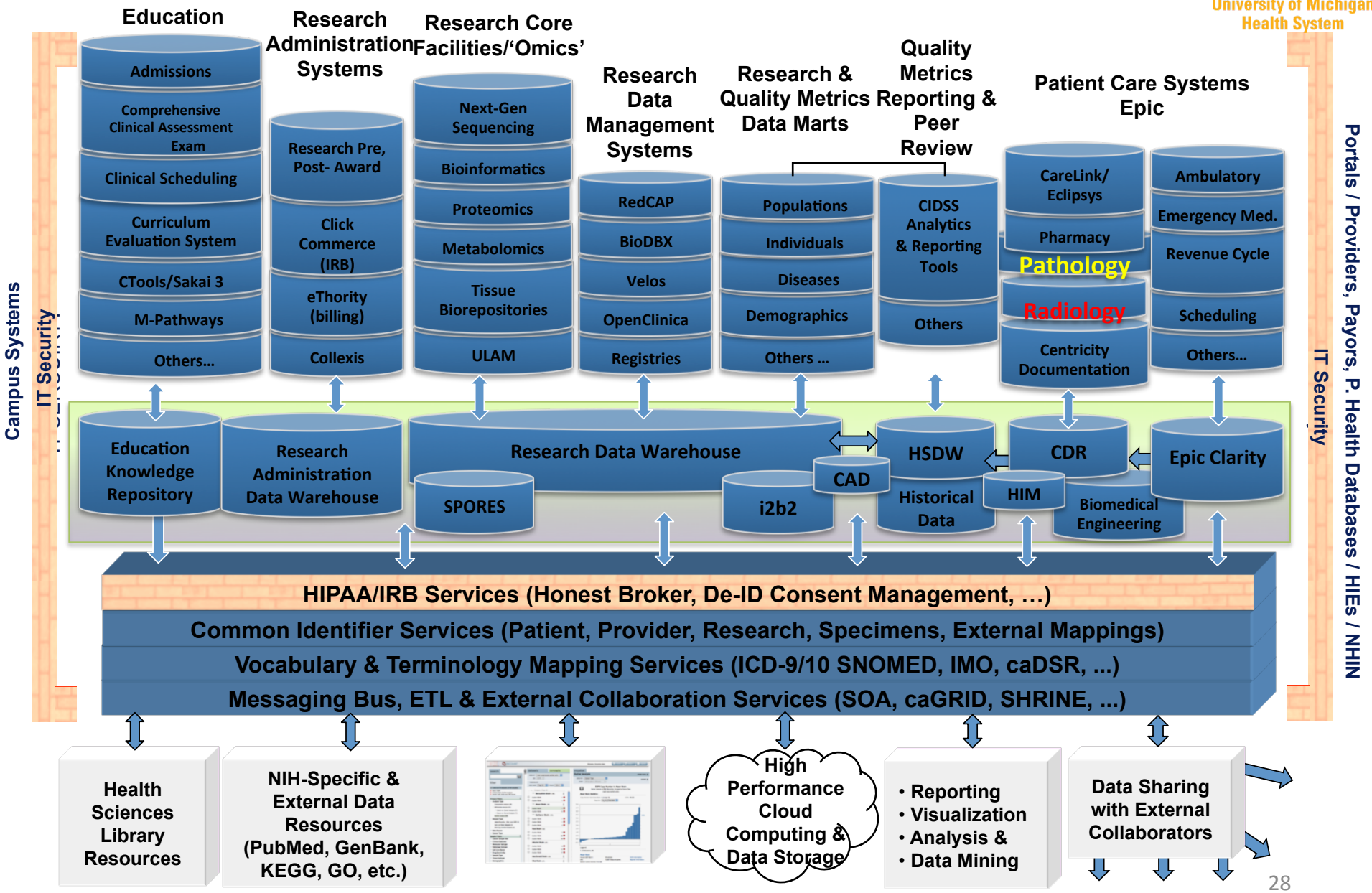


UMHS Data Architecture Unifying the Three Missions: *Education, Research, & Patient* Care

University of Michigan Health System

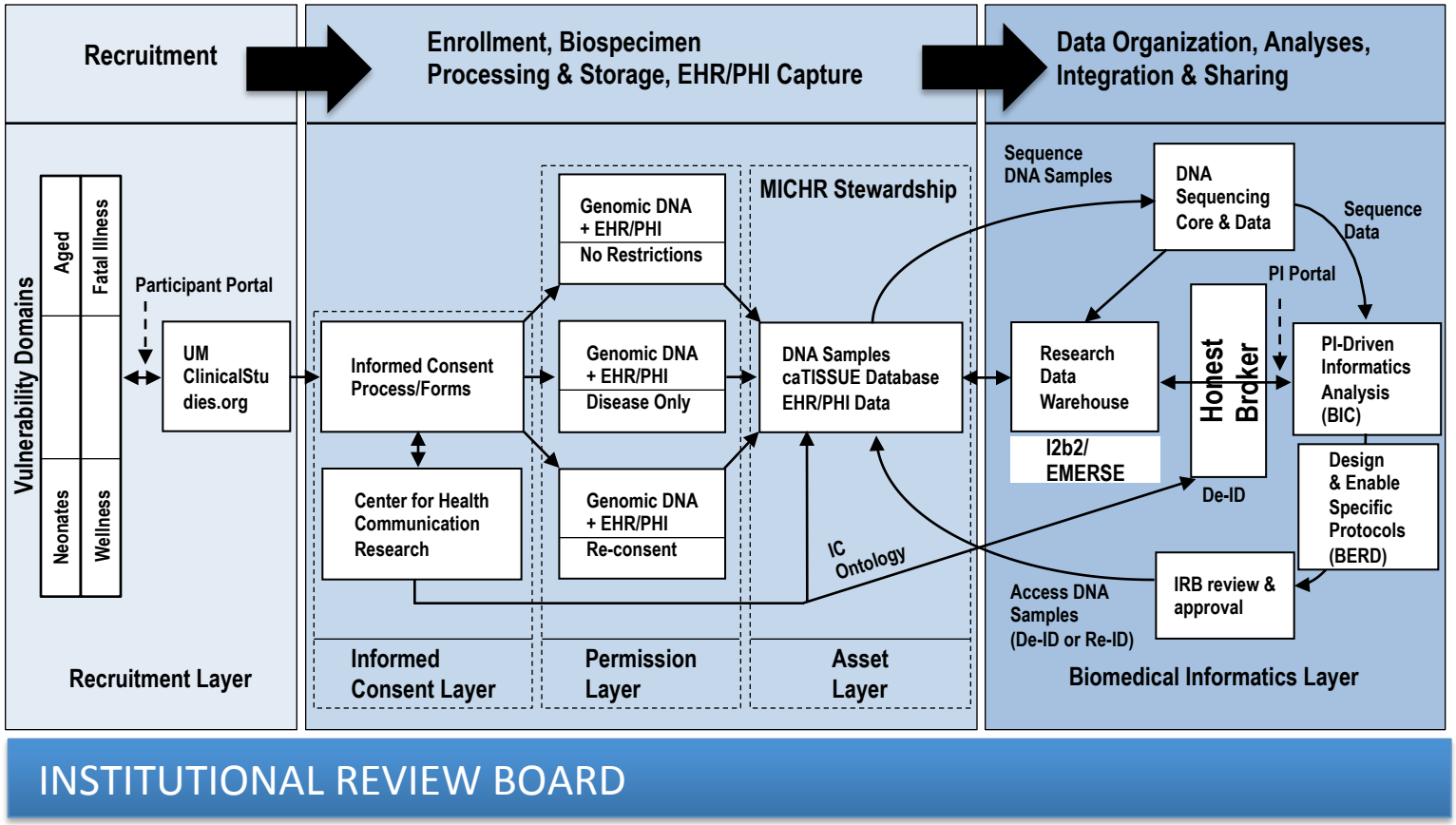


UMHS Data Architecture Unifying the Three Missions
Education, Research, & Patient Care



Process Overview of Michigan Genomic DNA BioLibrary

University of Michigan Health System



INSTITUTIONAL REVIEW BOARD

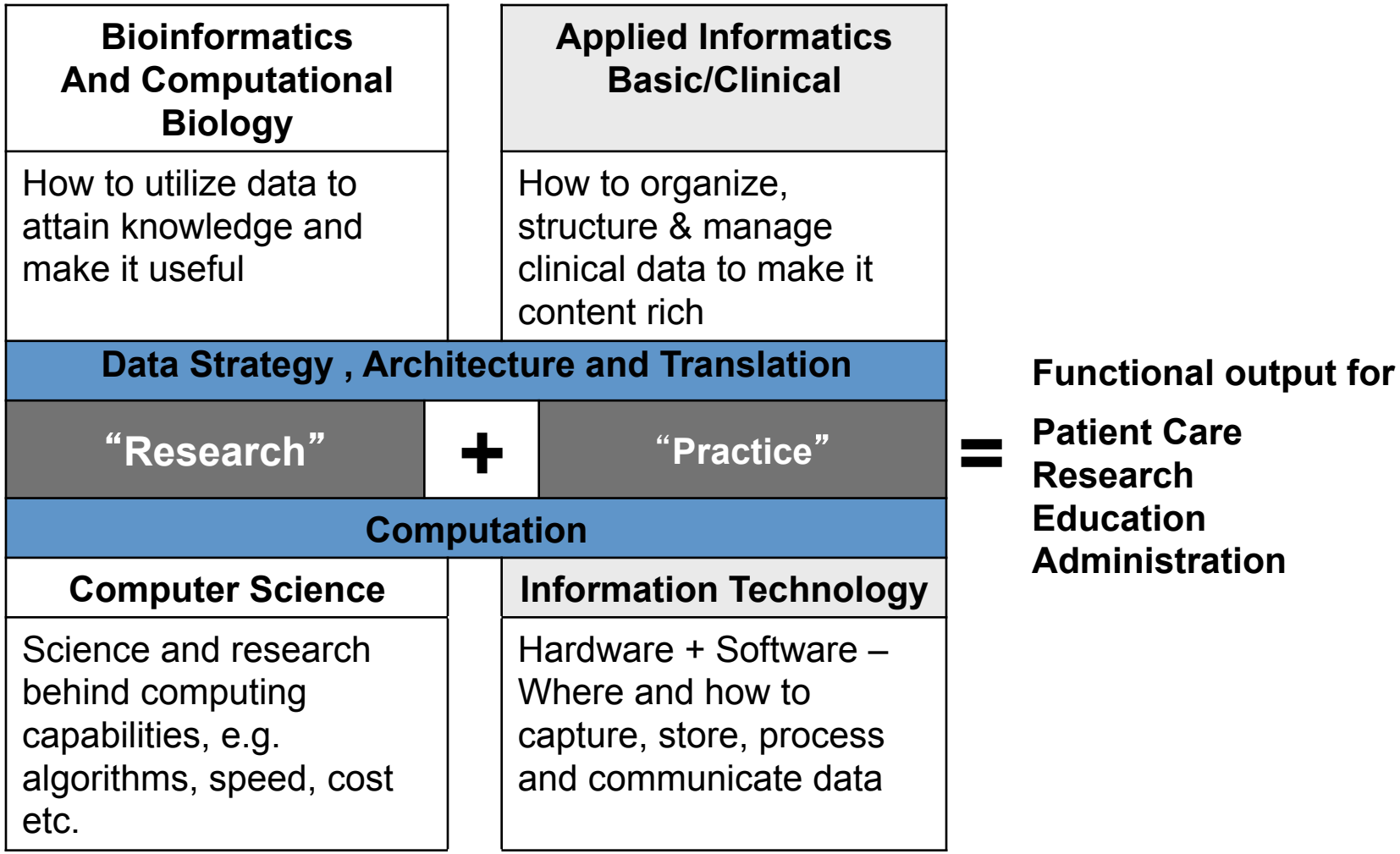
"Technical desiderata for the integration of genomic data into electronic health records" (Masys et al., J. Biomed. Inf., 2012)

Goal: Understand how genomic data differs from other health data in the medical record and how to 'handle it'.

Conclusions:

- ✓ Maintain separation of primary data and observations
- ✓ Support lossless compression
- ✓ Link observations to lab methods
- ✓ Compactly represent clinical actionability
- ✓ Support human and machine-readable formats
- ✓ Anticipate changes in our understanding of variation
- ✓ Support both clinical care and discovery science

Terminology – IT meets Informatics



UMMS Research IT: Current Foci

University of Michigan Health System

- Federated Research Data Warehouse Structure
- Honest Broker System (Rules and People)
- Enterprise Clinical Research Data Management
- Integrated Biorepository
- Centralized and Affordable Data Storage
- IT Infrastructure for Research Cores
- Enhanced Support for the Clinical Research Interface with Epic through ECRIT Team (Committee)

Research IT Strategic Team
UMMS Office of Research

Information Technology for Cancer and Healthcare

Core Requirements

- IT infrastructure: sufficient cycles and storage, open source, extensible, interoperative, integrated with technology platforms and analytical tools
- Top down enterprise solutions dangerous—domain-expertise driven, data-driven (bottom up), user acceptable—how to integrate software packages
- Gold standard for internet medical information
- Handle conventional medical records and histories
- Handle molecular, cellular and phenotypic data
- Identify actionable gene variants in the individual genome sequences
- Handle the comparative and subtractive analyses of billions of genomes and their attendant phenotypic data
- Handle the digitization of personal data
- Handle extensive personalized imaging data
- Handle social network data
- Handle longitudinal data gathered on individual patients
- Derive and analyze the dynamical "network of networks" for the individual
- Integrate data/networks and generate predictive and actionable models

Institute for Systems Biology
Revolutionizing Science. Enhancing Life.