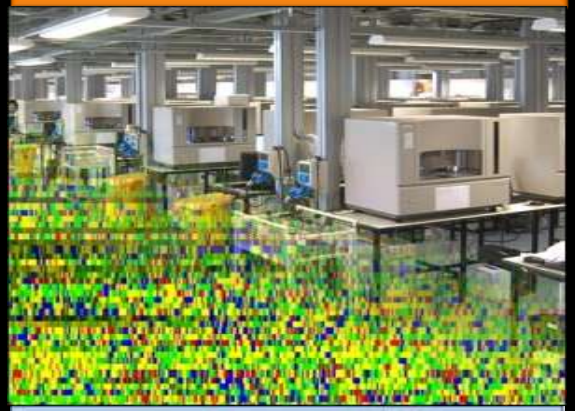
What's Wrong with Biomarker Discovery? Just About Everything!

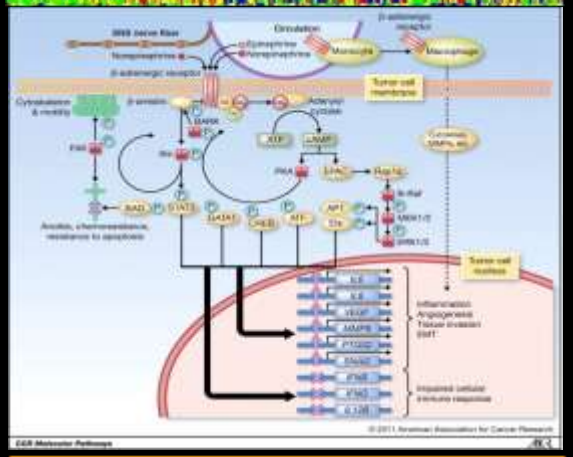
Dr. George Poste
Chief Scientist, Complex Adaptive Systems Initiative
and Del E. Webb Chair in Health Innovation
Arizona State University
george.poste@asu.edu
www.casi.asu.edu

NBDA Workshop V, The Phoenician, Scottsdale AZ
July 14-15, 2014

The Quest for Precision Medicine:
A New Era of Massive Expansion of Molecular Profiling Data (panOmics)

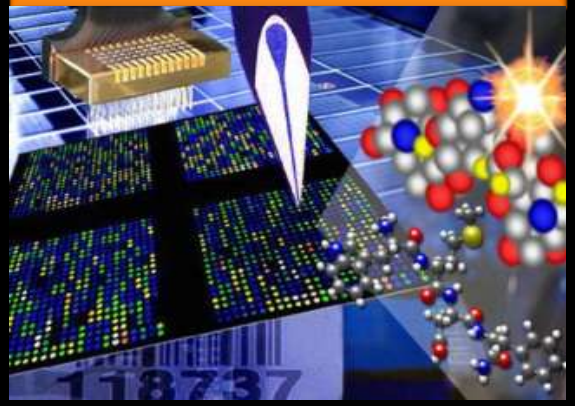
(Epi)Genomics

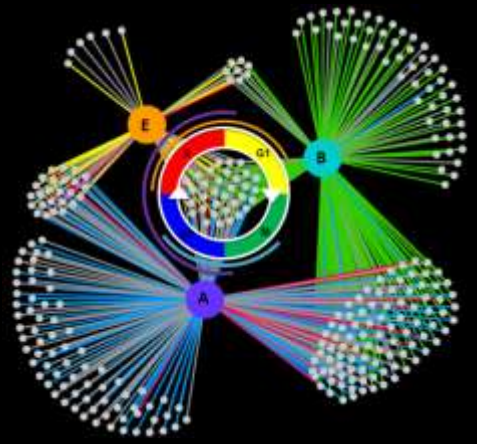




Molecular Pathway Analysis

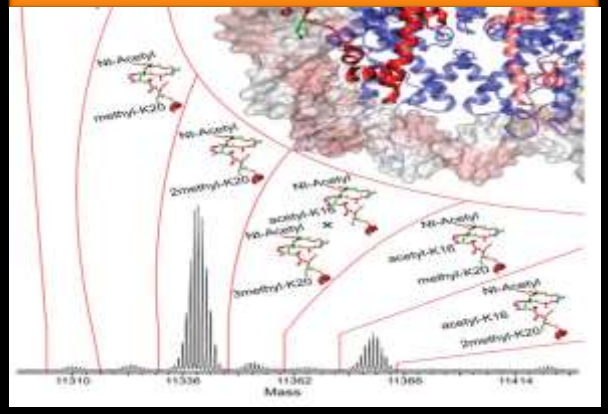
Transcriptomics

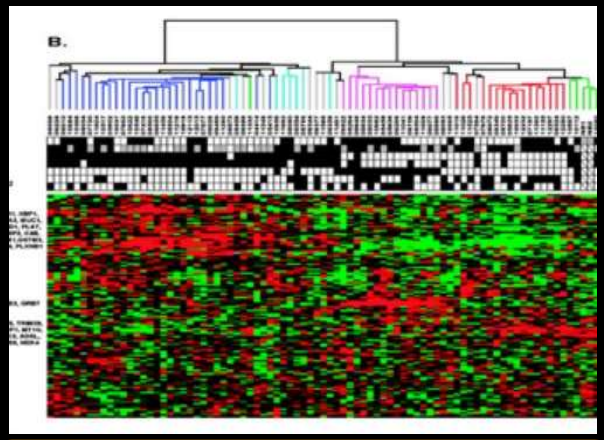




Network Topology and Architecture

Proteomics





Network Perturbation(s) and Disease Subtypes

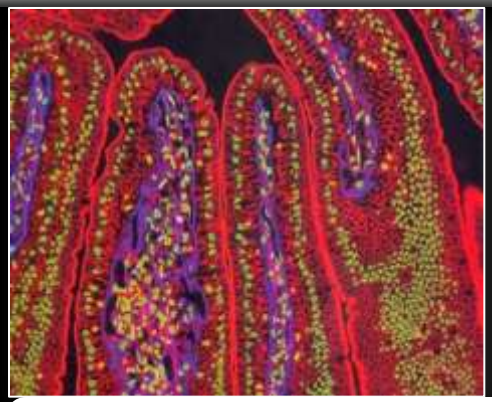
Biomarkers, Molecular Diagnostics (MDx) and the Promise of Precision Medicine

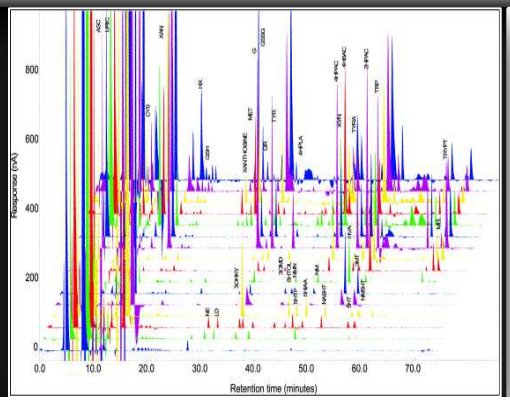
- the potential economic and health benefits from biomarkers transcend any other current category of healthcare innovation
  - increased diagnostic accuracy
  - rational treatment selection
  - monitoring treatment efficacy
  - earlier detection of treatment resistance
  - identification and mitigation of disease predisposition risk
  - health monitoring and optimized wellness
- key driver for improved care at lower cost

The Design of the Biomarker Discovery and Validation Process

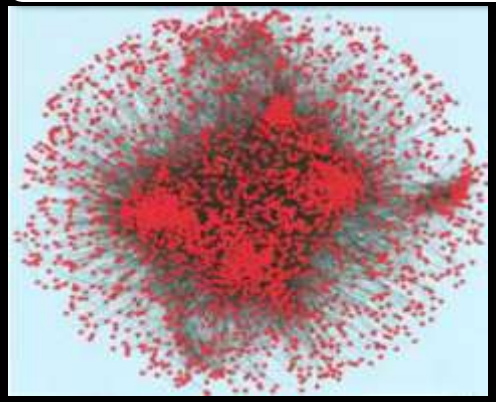
- demonstration of robust correlation and/or causality with specific traits/phenotypes requires sophisticated ***multi-dimensional integration of diverse datasets***
  - characterization of regulation of complex molecular networks (panOmics)
  - mapping system states: cells, organs, individuals
- understanding the complexity of genomic: phenotypic relationships and varied impact of environmental factors (exposome)

Identification and Validation of Disease-Associated Biomarkers: Obligate Need for a Systems-Based Approaches

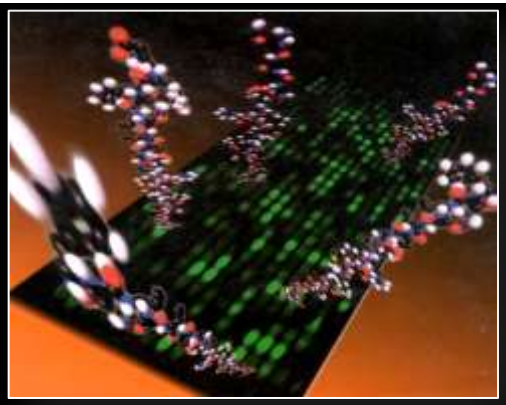




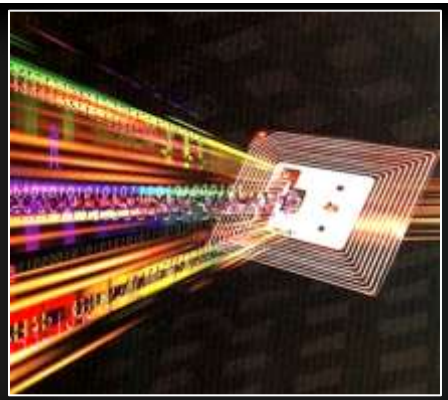
End-to-End Integration Logistics



Biospecimens and Analysis of Molecular Pathways and Networks



Multiplex Assays and New Analytical Platforms



Novel Instrumentation, Informatics Algorithms and Complex Signal Deconvolution

New Regulatory Requirements and Medical Education

The Design of Biomarker Discovery and Validation Process: A Complex Multi-Dimensional, Multi-Disciplinary Exercise

Comparable Technical, Logistical and Regulatory Complexity to (Bio)Pharmaceutical R&D

Failure to Acknowledge and Address This Reality is the Principal Cause of the Dismal Historical Productivity of Biomarker R&D

The Design of the Biomarker Discovery and Validation Process

- adopt systems-based approaches to resolve biological complexity
- from isolated (siloed) uni-dimensional, single discipline efforts to sophisticated integration of multi-dimensional, multi-disciplinary efforts (systems engineering)
- 3M networks: multi-investigator, muti-institution, multi-million
- learning from other complex multi-disciplinary domains
  - engineering, computing, experimental physics
  - logistics and management of large scale projects

The Design of the Biomarker Discovery and Validation Process

performance requirements: endpoints and outcomes

- unmet needs
- merits of competing (alternate) approaches

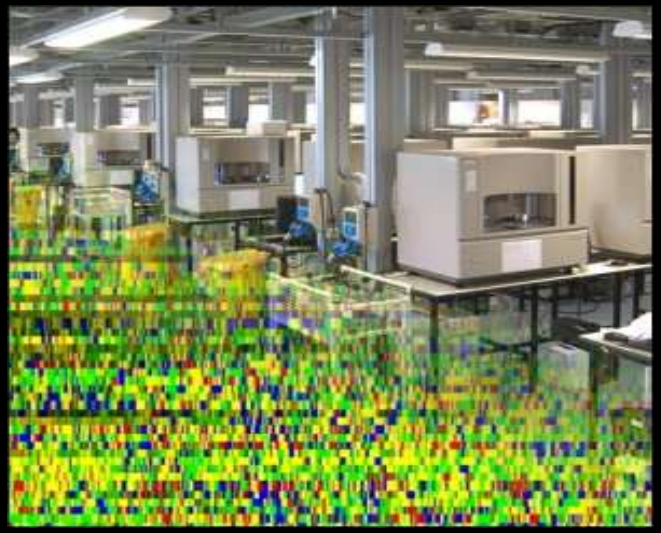
execution requirements: R&D

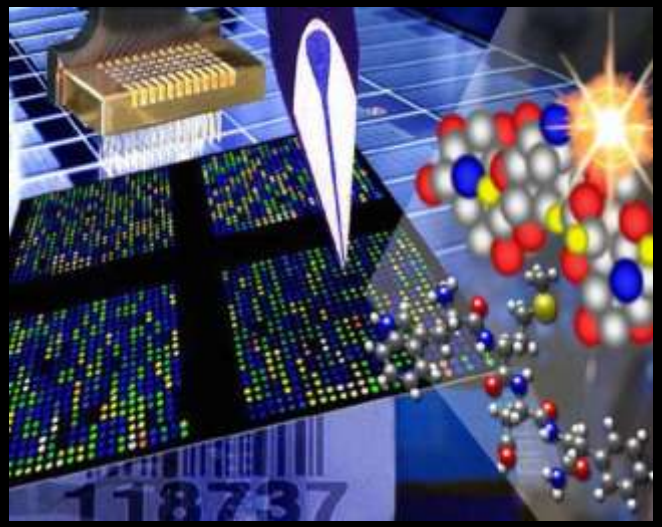
- conceptual, technical, logistical and financial
- collaboration networks and clinical trials
- intellectual property
- new regulatory and reimbursement requirements

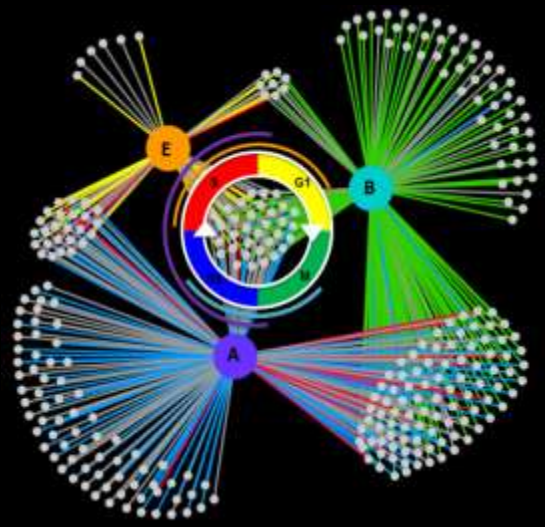
market adoption: return on investment

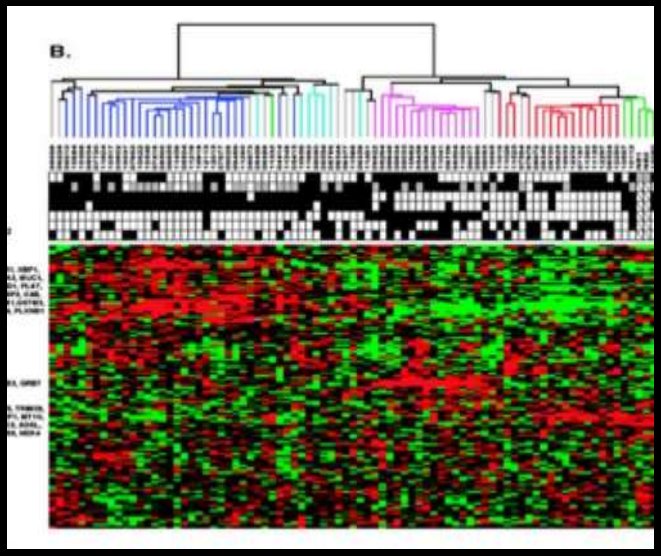
- value creation

The Challenge of Translation of Burgeoning panOmics Data Into Clinically Relevant (Actionable) Knowledge

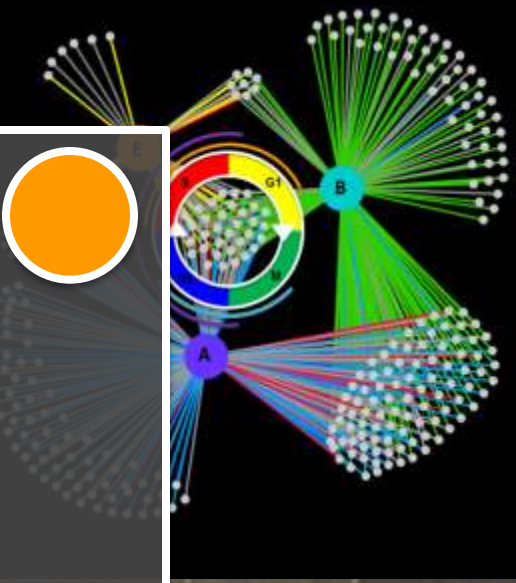






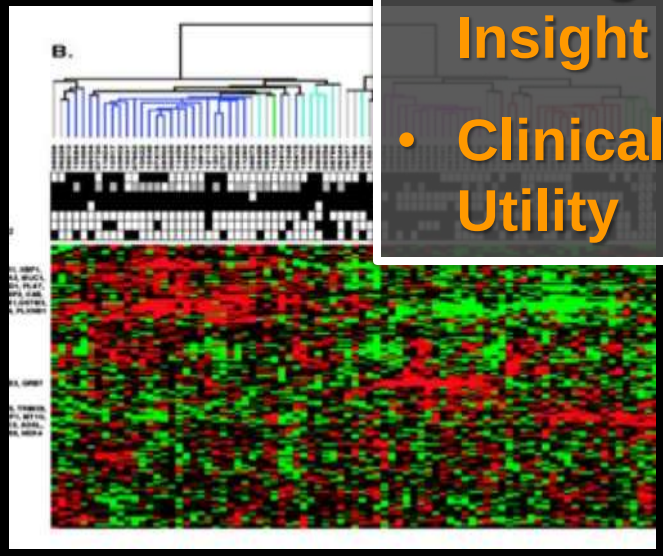


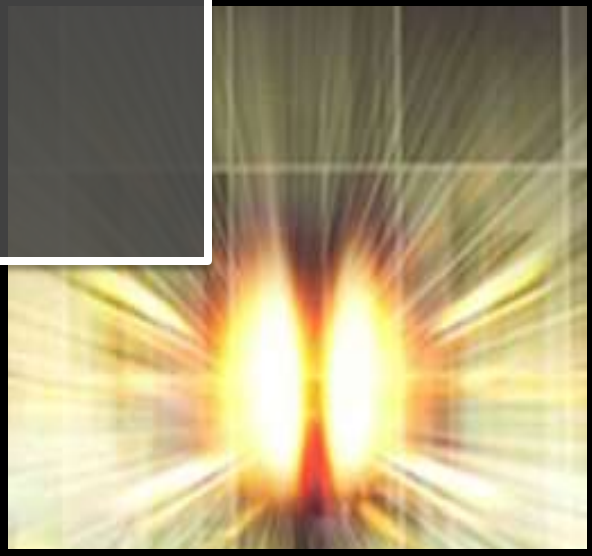
The Challenge of Translation of Burgeoning panOmics Data Into Clinically Relevant (Actionable) Knowledge



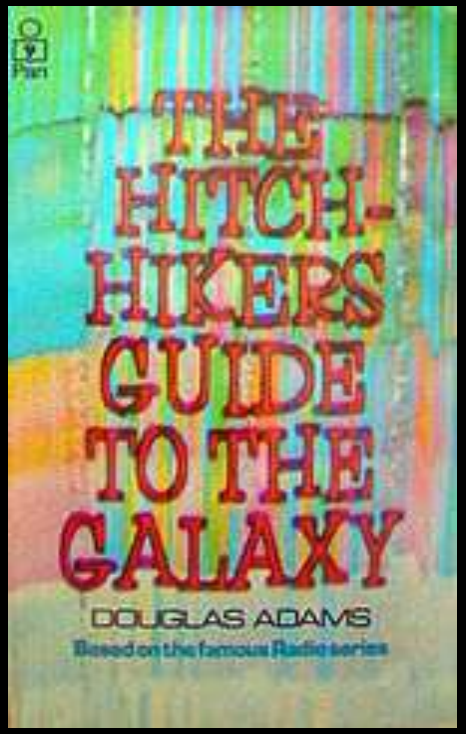
- Data
- Reliability and Robustness
- Biological Insight
- Clinical Utility

?





Defining the Right Question in an Era of Data Overload



"O Deep Thought Computer
….what is the answer?"
asked the expectant questioners

"42" said Deep Thought with infinite
calm and majesty

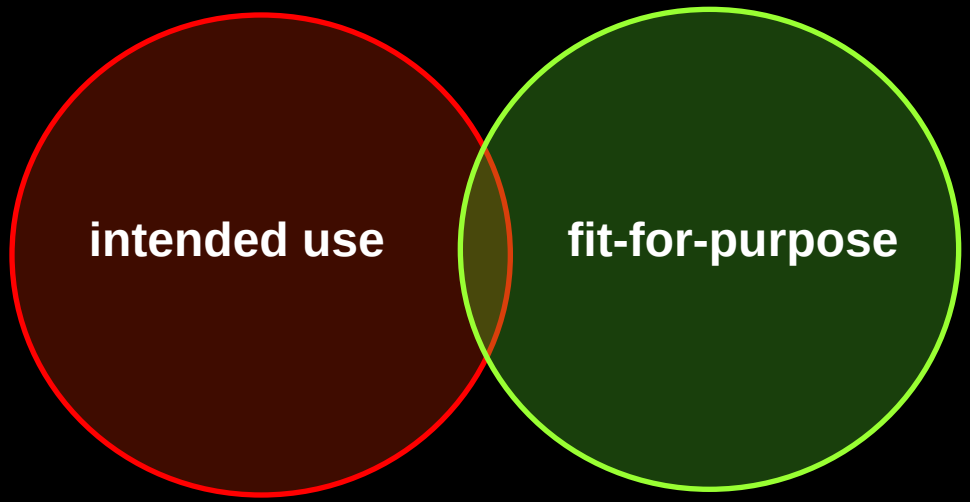
"But what does 42 mean?"

"Once you know what the question actually is,
you'll know what the answer means,"
said Deep Thought with the air of one who
suffers fools gladly (chapter 28)

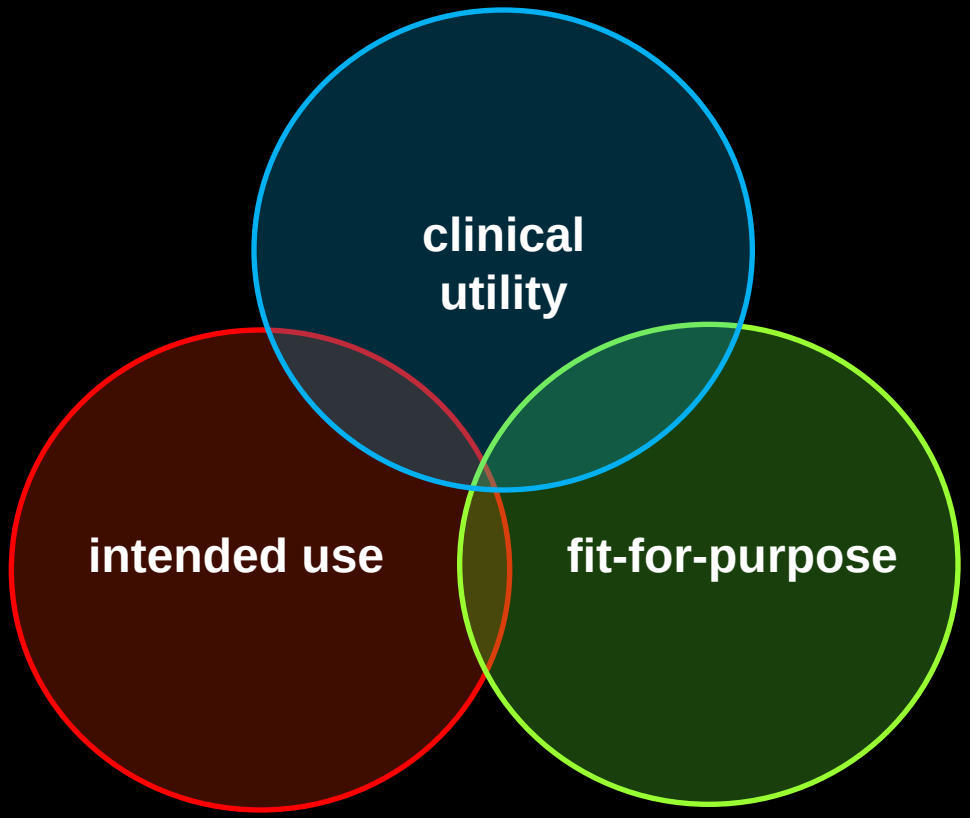
“What people think of as the moment of discovery is really the discovery of the question.”

Jonas Salk

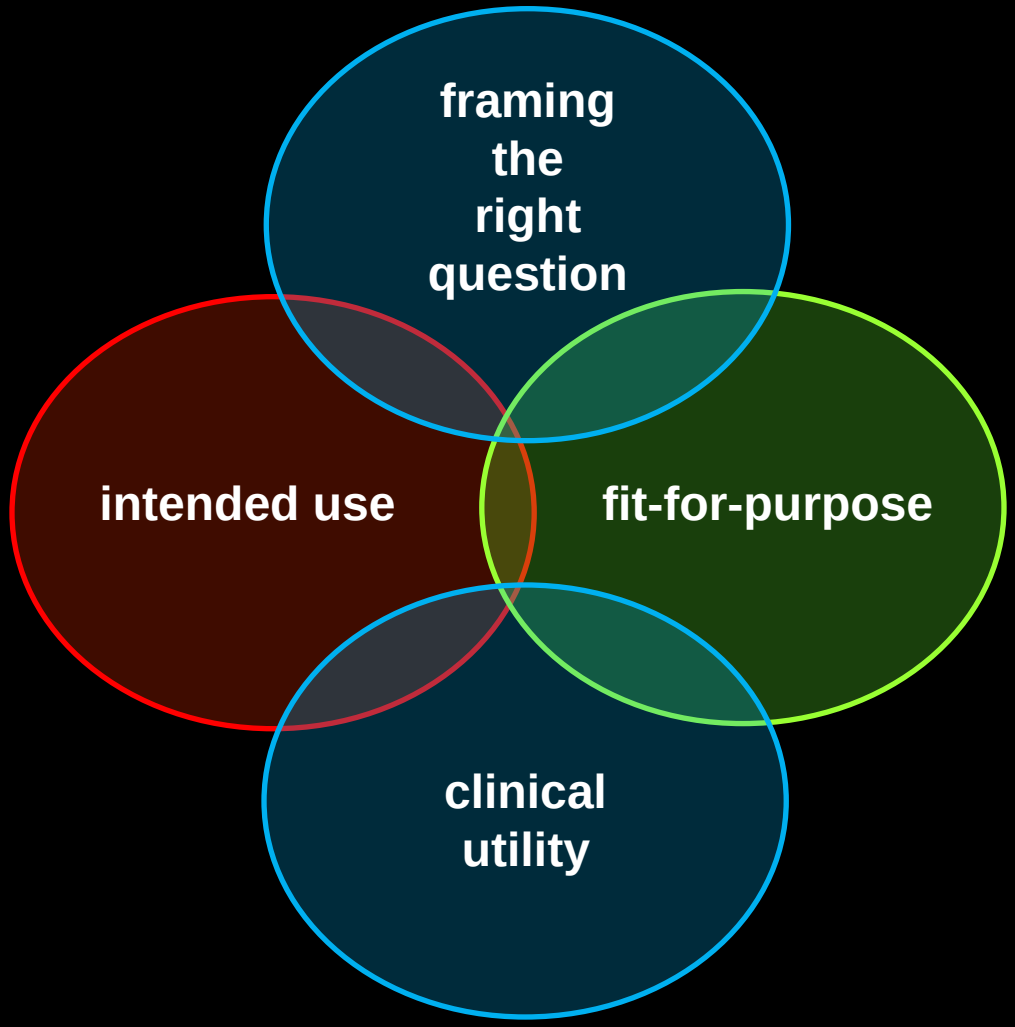
Biomarker Discovery and Validation: Overarching Endpoints



Biomarker Discovery and Validation: Overarching Requirements



Biomarker Discovery and Validation: Process Design



Criteria for Establishment of Causality Association for Biomarkers (Modified Bradford Hill's Principles of Causality)*

- strength of association
  - sensitivity, specificity, PPV, NPV
- consistency
  - across the clinical populations studied
  - across the methods used
- temporality
  - relationship to disease progression and/or Rx response
- biological gradient
  - dose-response relationship and identification of thresholds for Rx use/discontinuation
- coherence and plausibility
  - coherence = does not conflict with substantive knowledge
  - plausibility = consistent with substantive knowledge
  - both assume current concepts of pathophysiology are correct

*A. Bradford-Hill (1965) Proc. R. Asoc. Med 58, 295-300

Rethinking and Redesigning Biomarker Discovery and Validation

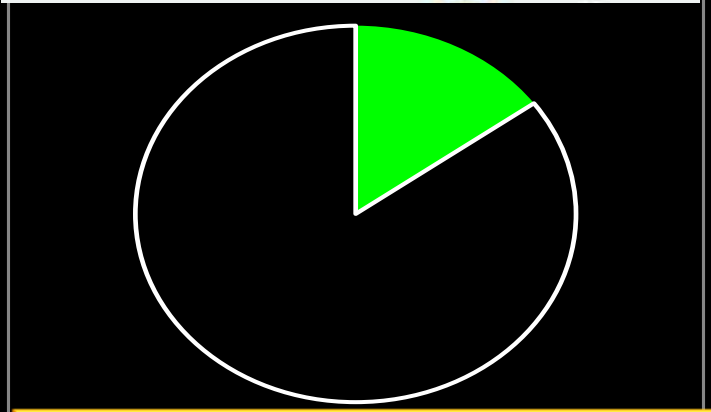
NBDA
National Biomarker Development Alliance

WORKSHOP V

- formulating a clinically important question
- experimental design (translatable robust discovery)
- biospecimens
- technology standards
- data collection, quality and management
- data analysis and analytics

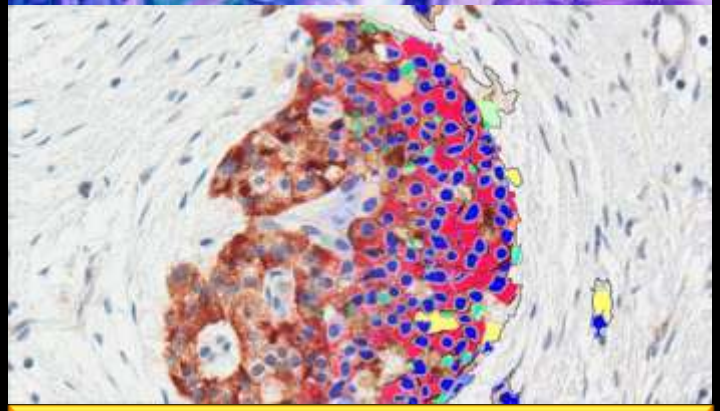
Step One in Biomarker Discovery and Development: Error(s) in the Most Proximal Activities Will Cascade and Contaminate Downstream Efforts

Inadequate and Erratic Clinical Phenotyping



Statistically Underpowered Sample Sets
n<<p

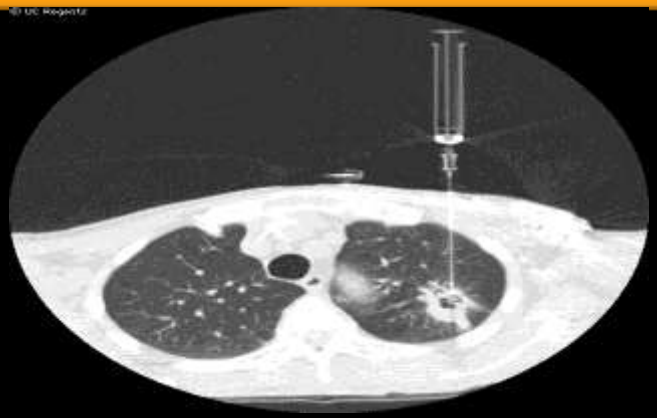
Specimens of Convenience: Poor Annotation and Uncertain Provenance



Sampling Bias: Intra- and Inter-lesional Heterogeneity (Oncology)

The Primacy of Biospecimen Selection, Annotation and Curation for Robust Biomarker Discovery

Intended Use
Sampling Protocols



Pre-analytical Variables
and
Stringent Documentation

Analytical Variation
and
Lack of Reference Standards

Idiosyncratic Lab-Specific
Methods and Data Formats

Sloppy and Unstandardized Science:
The Growing Problem of Poor Reproducibility in Biomedical Publications

How Not to Be a Bioinformatician

M. Corpas et al. (2012) Source Code for Biol. and Med. 7:3

- claim to be open source without being open
- make tools that make no sense to biologists
- make sure the output of your application is unreadable, unparseable and does not comply to any known standards
- never maintain your databases, web services or any information that you may provide at any time
- do not ever share your results and do not reuse
- make your algorithm or analysis method irreproducible

Statistical Analysis of Multiplex Assays and Large Scale Data: A Growing Knowledge Void for Many Researchers, Clinicians, IRBs and Journal Reviewers

STATISTICAL ERRORS

P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume.

BY REGINA NUZZO

Nature (2014) 506, 150

"Readers must have confidence in the conclusions published in our journal."

Science (2014) 345, 6192

Core Tenets for Robust Biomarker R&D

- **Scale**
- **Scalability**
- **Standards**

Building Large Scale Biomedical Collaborative Research Networks and Commitment to Standardization of Experimental Platforms, Models and Data Reporting

Standardized

- calibration is traceable to a reference measurement procedure that is in turn calibrated with an appropriate reference material

Where Are the Reference Materials for Standardization/Harmonization of panOmics Profiling?

Deployment of WGS in Clinical Care

- because we can?
- because it is useful?

Meeting the ‘Fit-for-Purpose’ Standard

The Urgent Imperative to Define Analytical and Interpretation Standards for Clinical Grade Genome Sequencing

Obligate Reading for Anyone Who Wants to Deploy WGS in a Clinical Setting

JAMA. 2014;311(10):1035-1044.
doi:10.1001/jama.2014.1717

Original Investigation

Clinical Interpretation and Implications of Whole-Genome Sequencing

Frederick E. Dewey, MD; Megan E. Grove, MS; Cuiping Pan, PhD; Benjamin A. Goldstein, PhD; Jonathan A. Bernstein, MD, PhD; Hassan Chaib, PhD; Jason D. Merker, MD, PhD; Rachel L. Goldfeder, BS; Gregory M. Enns, MB, ChB; Sean P. David, MD, DPhil; Neda Pakdaman, MD; Kelly E. Ormond, MS; Colleen Caleshu, MS; Kerry Kingham, MS; Teri E. Klein, PhD; Michelle Whirl-Carrillo, PhD; Kenneth Sakamoto, MD; Matthew T. Wheeler, MD, PhD; Atul J. Butte, MD, PhD; James M. Ford, MD, PhD; Linda Boxer, MD; John P. A. Ioannidis, MD, PhD; Alan C. Yeung, MD; Russ B. Altman, MD, PhD; Themistocles L. Assimes, MD, PhD; Michael Snyder, PhD; Euan A. Ashley, MRCP, DPhil; Thomas Quertermous, MD

Current Issues Related to the Accuracy and Quality of WGS for Clinical Applications

- error rate
- sequence completeness
- sequencing depth
- phasing
- instrument platform variation
- base calling algorithms
- aligning algorithms
- adequacy of reference genomes
- annotation, analysis and curation of large scale data

DNA Variations Not Well Detected or Undetectable by WES and WGS

- repetitive DNA including trinucleotide repeats (and phasing challenges)
- copy number variants
- long insertion-deletion variants
- structural variants and chromosomal translocations
- aneuploidy
- epigenetic alterations

Ignoring Biological Complexity

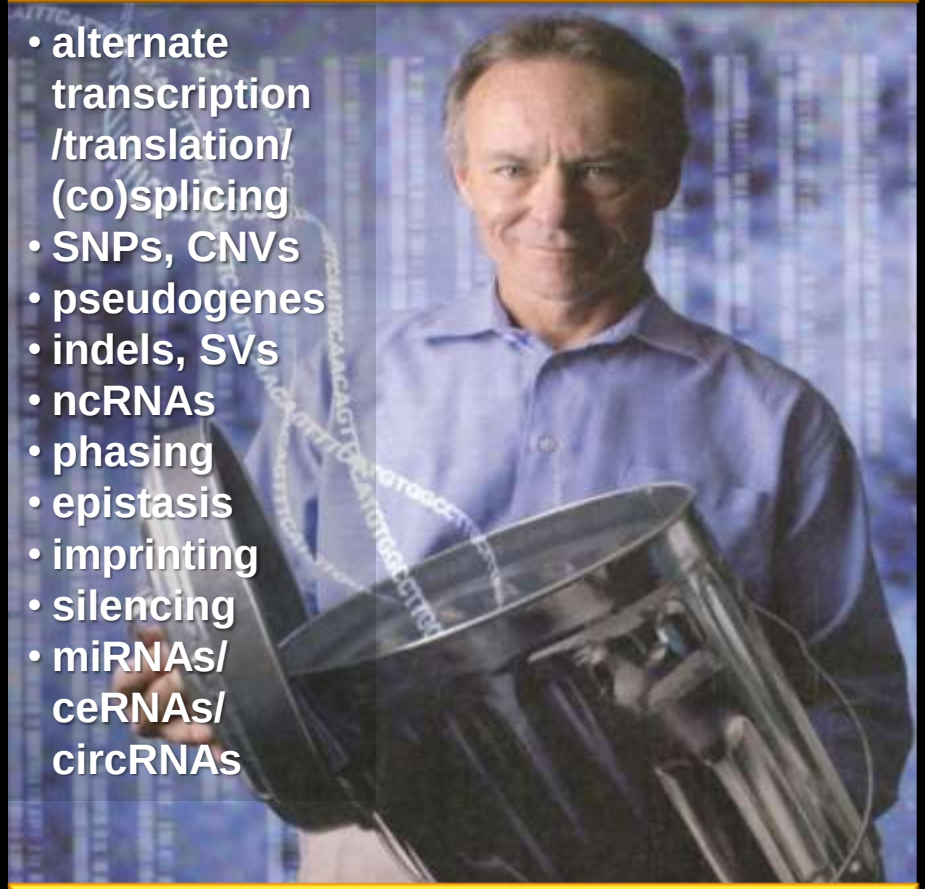
Genes For ….
The Overly Simplistic and Deterministic Dangers of a Genome-Sequence Centric Perspective

Biology is More Than the Germ Line and Somatic Genomes

The Over-Simplified Perspective That While Exome-and Whole Genome-Sequencing Will Reveal the Full Etiology of Disease Pathogenesis

Individual Variation, Genome Complexity and the Challenge of Genotype-Phenotype Predictions

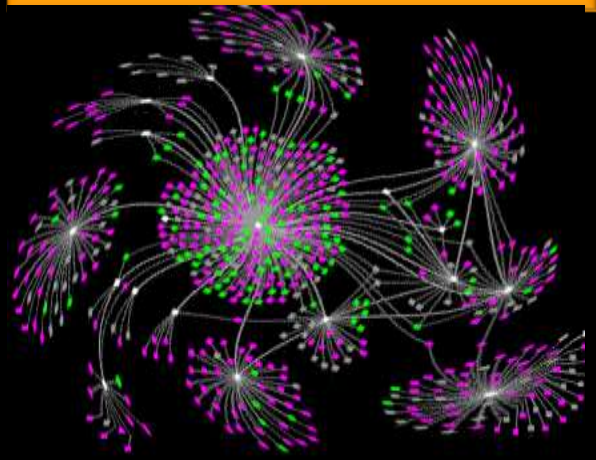
Junk No More: Pervasive Transcription



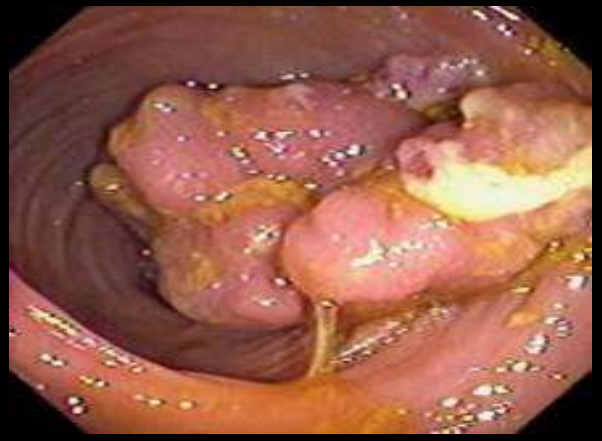
- alternate transcription /translation/ (co)splicing
- SNPs, CNVs
- pseudogenes
- indels, SVs
- ncRNAs
- phasing
- epistasis
- imprinting
- silencing
- miRNAs/ ceRNAs/ circRNAs

recognition of genome organizational and regulatory complexity

Cell-specific Molecular Interaction Networks



Perturbed Networks and Disease



Nature (2014) 508, 469
doi:10.1038/nature13127

Guidelines for investigating causality of sequence variants in human disease

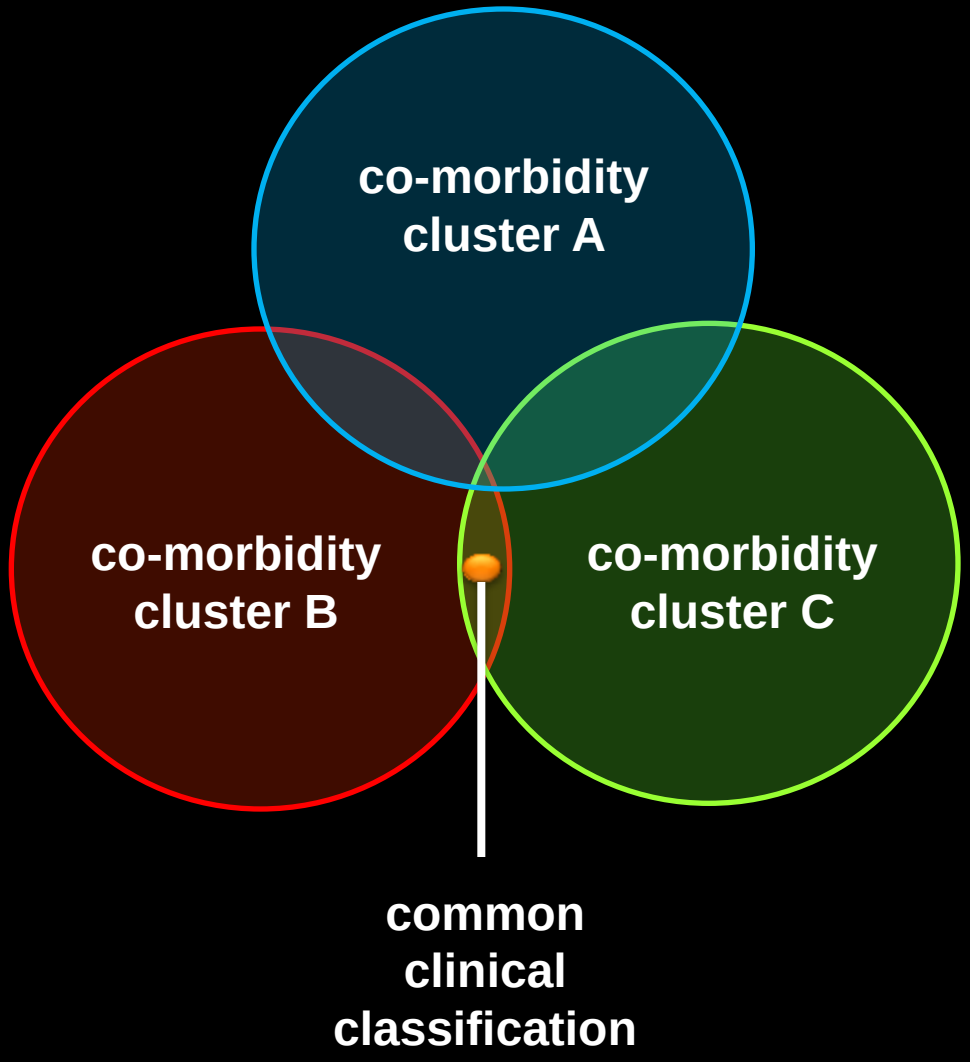
D. G. MacArthur[1,2], T. A. Manolio[3], D. P. Dimmock[4], H. L. Rehm[5,6], J. Shendure[7], G. R. Abecasis[8], D. R. Adams[9,10], R. B. Altman[11], S. E. Antonarakis[12,13], E. A. Ashley[14], J. C. Barrett[15], L. G. Biesecker[16], D. F. Conrad[17], G. M. Cooper[18], N. J. Cox[19], M. J. Daly[1,2], M. B. Gerstein[20,21], D. B. Goldstein[22], J. N. Hirschhorn[2,23], S. M. Leal[24], L. A. Pennacchio[25,26], J. A. Stamatoyannopoulos[27], S. R. Sunyaev[28,29], D. Valle[30], B. F. Voight[31], W. Winckler[2]† & C. Gunter[18]†

The discovery of rare genetic variants is accelerating, and clear guidelines for distinguishing disease-causing sequence variants from the many potentially functional variants present in any human genome are urgently needed. Without rigorous standards we risk an acceleration of false-positive reports of causality, which would impede the translation of genomic research findings into the clinical diagnostic setting and hinder biological understanding of disease. Here we discuss the key challenges of assessing sequence variants in human disease, integrating both gene-level and variant-level support for causality. We propose guidelines for summarizing confidence in variant pathogenicity and highlight several areas that require further resource development.

From Genotype to Phenotype: Understanding Genetic Architecture and Biological Networks in Health and Disease

- which alleles and their variants drive disease phenotypes?
- what are the network interactions between causal and modifying genes that define expressivity, penetrance and ultimate phenotypic impact?
  - genesis of a likely continuum of clinical phenotypes
  - from subtle to severe disease generated by graded perturbations within- and between-molecular pathways and networks

Deep Phenotyping and Longitudinal Clinical Monitoring

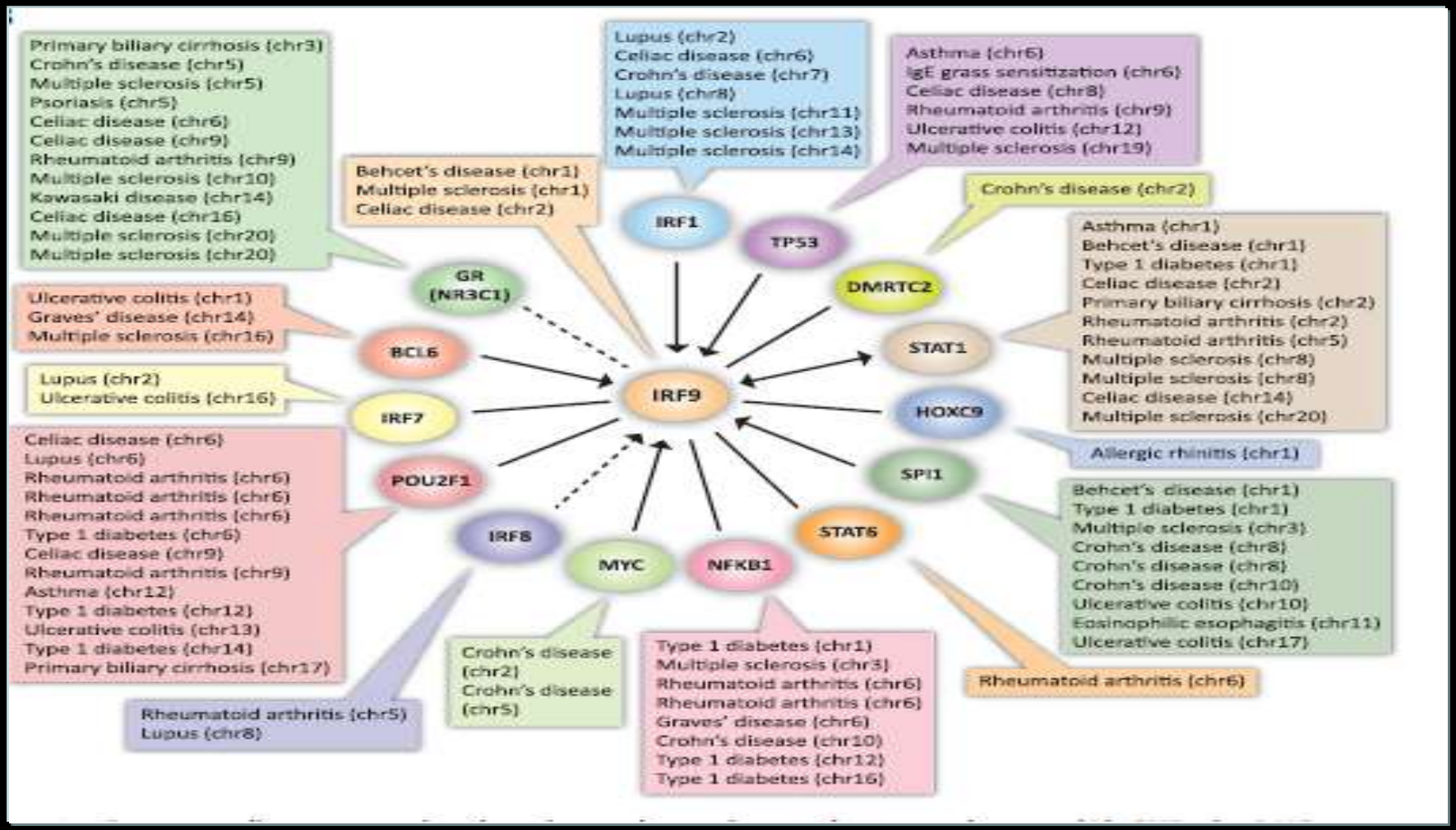


Value of NLP of EMRs to ID Cohorts with Disease Pathotypes with Clustered Co-Morbidities

- Doshi-Velez et al. (2014) Pediatrics 133 e.54-e.63
  - autism spectrum disorder cohorts with distinct co-morbidities
  - seizures
  - high prevalence of infections/autoimmunity
  - variety of additional neuropsychiatric disorders

Variants for Common Disease in Related Disease Categories Cluster in Shared Regulatory Networks

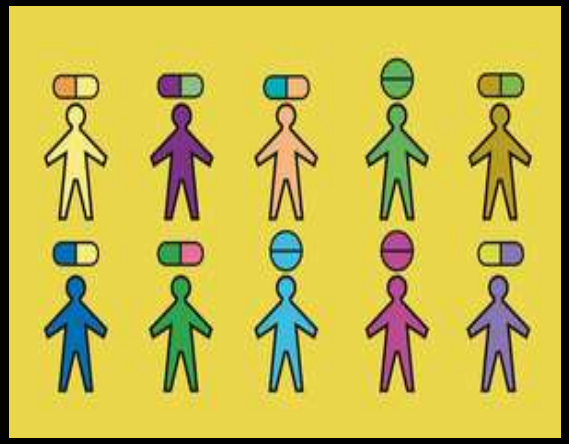
Autoimmune Disorders



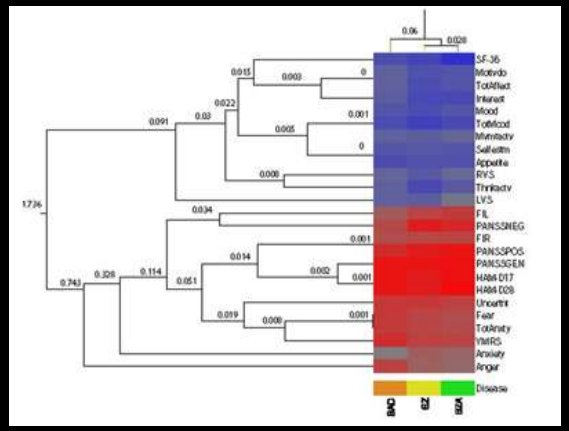
From: M. T. Maurano et al. (2012) Science 337, 1190

Clinical Validation of Biomarkers and a Key Element in Novel Clinical Trial Designs

Clinical Trial Recruitment (Stratified Trials)



New Clinical Trial Designs (Adaptive; Basket)



Dynamic Informed Consent Protocols

Large Scale Profiling of Cancer Patients to Identify Cohorts Expressing Low Frequency Rx Target(s) for Phase II Trials

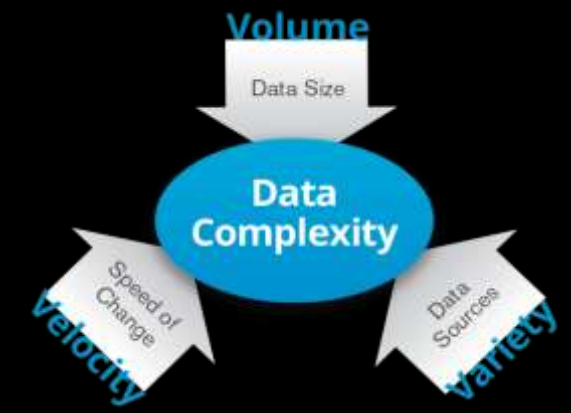
Target	# Patients Screened	# Eligible Patients	# Centers	# Countries
EML4 ALK^{+}: lung cancer*	1500	82	9	1
HER2^{+}: gastric cancer**	3803	549	122	24

* E.L. Kwak et al. (2010) NEJM 363, 1693

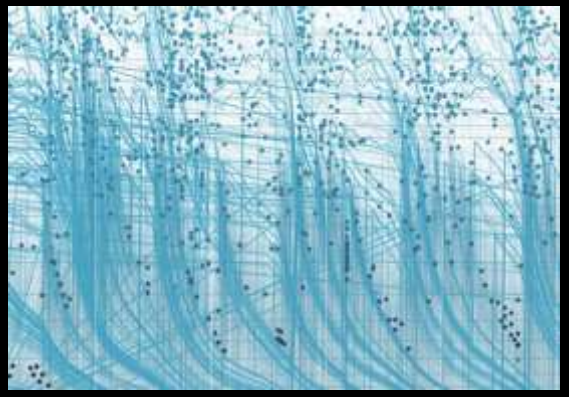
** Y. Bang et al. (2010) Lancet 376, 687

The Unavoidable Data-Intensive Nature of Biomarker Discovery and Validation

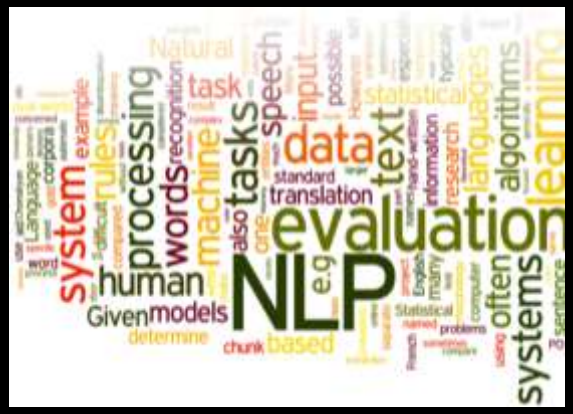
PB and TB Data Streams



Ontologies and Formats for Data Integration



Longitudinal Data Migration and Inter-operable Dbases



New Data Analytics, Machine Learning, NLP Methods

Infrastructure, Storage and Privacy

Informaticians

Biomedical R&D and Clinical Medicine: An Unavoidable Transition to Data-and Computation-Intensive Methods

- “silos” of research/clinical activities
- descriptive, subjective , qualitative data versus reproducible quantitative data
- proliferation of poorly standardized and fragmented data, semantic anarchy and incompatible databases

The Need for Standards in the Production, Publication, and Analysis of Large Biological Datasets

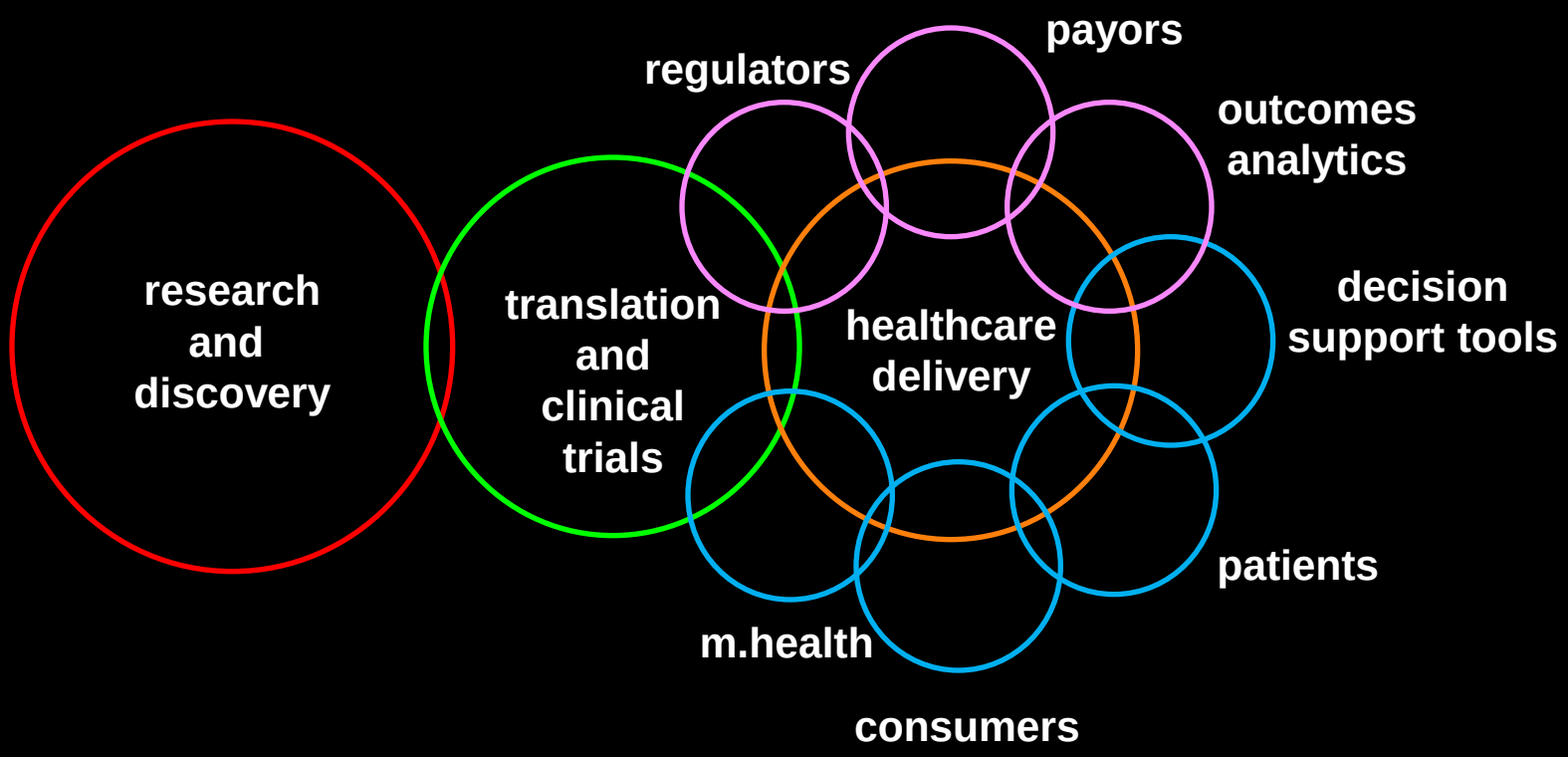
- few incentives for academic research and funding agencies to develop standards or share data
- failure of journals and funders to reinforce minimum reporting guidelines and standards and/or public deposition requirements for raw data/computer codes to facilitate replication and meta-analysis
- commercial software vendors protect/fragment markets via incompatible proprietary formats/interfaces

The Pending Zettabyte Era
1,000,000,000,000,000,000,000

Managing Big Data in Biomedicine is Not a Simple Extrapolation from Current Practices

Current Institutional Structures and Competencies Are Ill-Prepared for Pending Disruptive Change

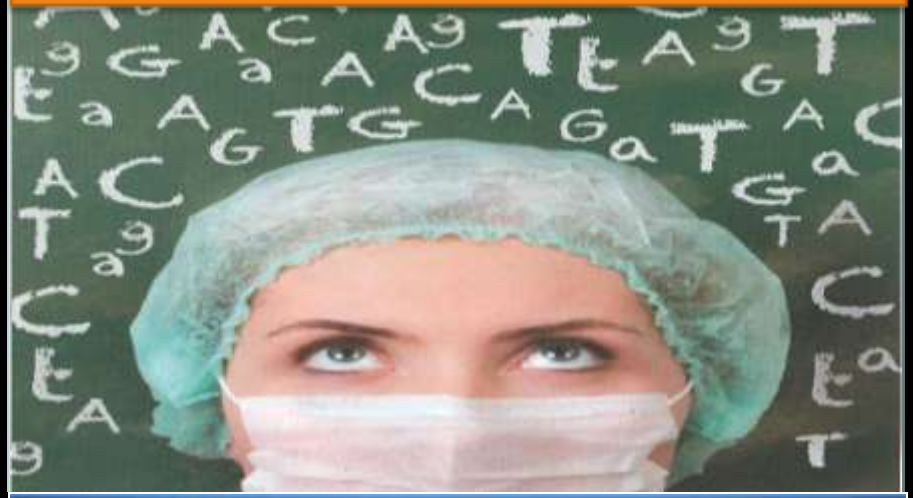
Design of Facile Exchange Formats for Data Assembly, Curation and Use Across the Continuum from Discovery to Healthcare Delivery

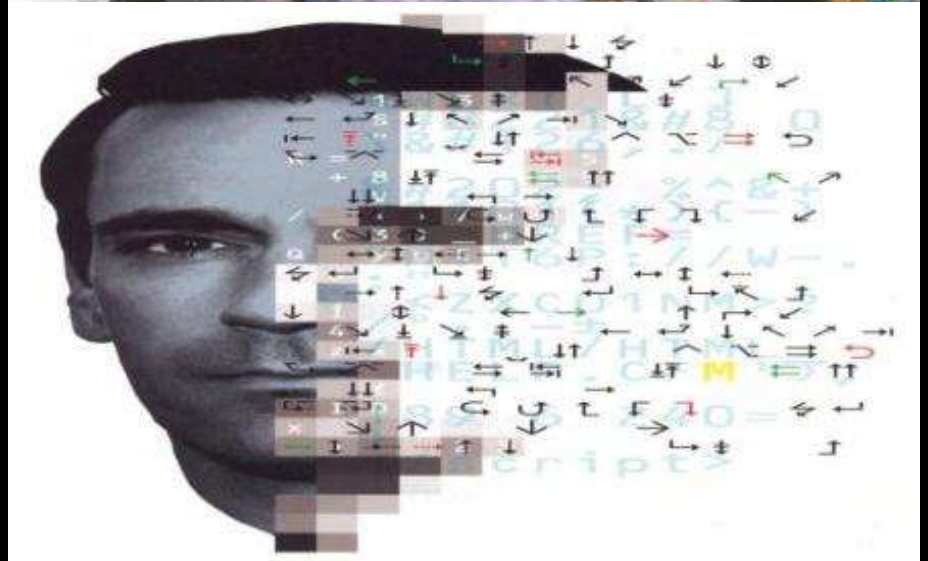


Technology Acceleration and Convergence: The Escalating Challenge for Professional Competency, Decision-Support and Future Education Curricula

Data Deluge

New Science and Cognitive Bandwidth



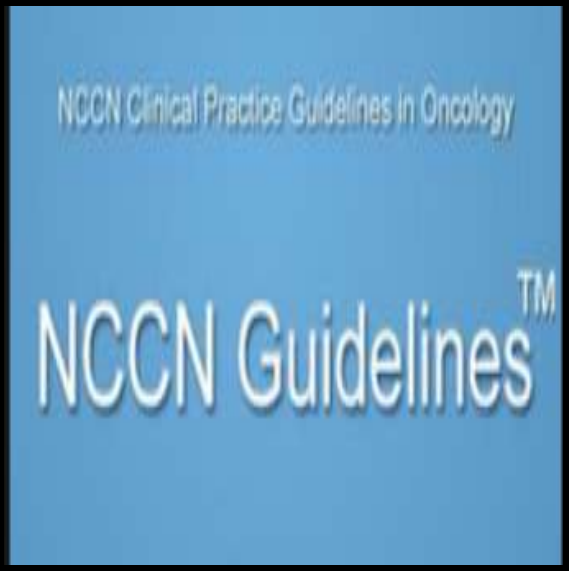


Automated Analytics and Decision Support

Facile Formats for Actionable Decisions

Barriers to Clinical Adoption of Biomarkers

Anachronistic Clinical Guidelines



Variable Clinical Practice and Guideline Adoption

Financial Risk by Displacement of Current Revenue Sources

Pervasive Problems in Biomarker R&D: Reimbursement

- payor resistance: FDA approval plus evidence of ‘value’ in clinical setting (fit-for-purpose)
- regulatory ambiguities for approval of multiplex assays and NGS/WES/WGS
- a growing dilemma for MDx investment and corporate viability
- higher R&D cost and risk of multiplex assays versus traditional unianalyte LTDs

What Must Change?

Biomarker R&D: Silos Versus Systems

- dismal historical productivity of biomarker R&D
- classic case-study of the conceptual, technical organizational and cultural deficits pervasive in contemporary biomedical R&D created by siloed specialities and reductionism
- reinforcement of fragmented efforts and lack of standards
  - lack of leadership by public funding agencies
  - publish or perish academic tenure provisions
  - journal policies and competition for ‘hot’ topics

“The perfect uselessness of knowing the answer to the wrong question.”

Ursula K. LeGuin
The Left Hand of Darkness

The Core Tenets in the Successful Design of Biomarker R&D

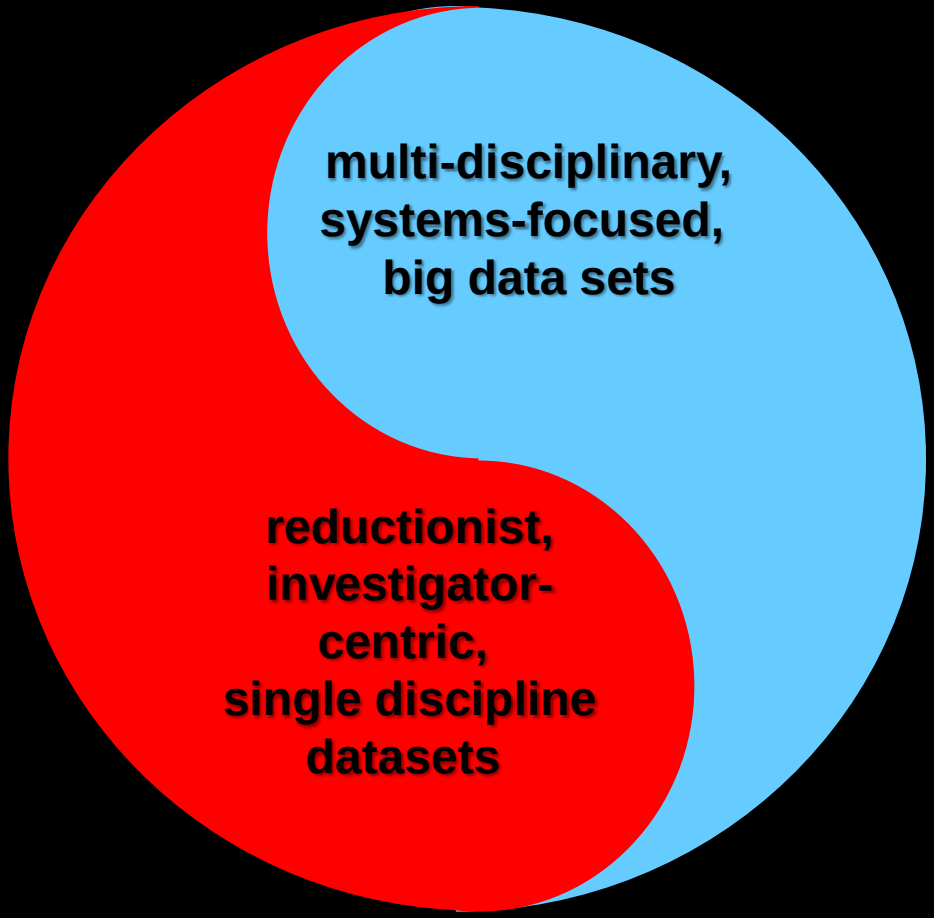
- Specimens
- Standards
- Scale and Scalability
- Systems-Based Approaches: End-to-End Integration
- Skills

Robust Processes

- prospectively defined checkpoints for every step
- ‘locked in’ methods and protocols
  - preanalytical handling and analytical assays
  - statistical and computational algorithms
  - clinical validation endpoints and confirmation of fit-for-purpose

Cross-Domain Convergence, Complexity in Biomedicine and Increasing Dependency on Data-Intensive Methods and Large Scale Collaboration

hypothesis driven research



unbiased datasets and new analytics for pattern mining

Defining An Optimum Balance

From Silos to Systems

• single discipline, single investigators	→	• multi-disciplinary teams
• single institution activities/resources	→	• large scale collaboration networks
• academic isolation	→	• academia-industry-healthcare provider networks
• fragmented qualitative data	→	• quantitative data
• fragmented data	→	• integrated data
• incompatible data formats	→	• data interoperability from discovery to clinical care

Managing the V5 Data Challenge: Volume, Variety, Velocity, Veracity and Value

- difficulty and expense of gaining access to "Big Data" will produce "have" and "have not" research cultures
- automating research and elements of decision-making/authority changes the definition of knowledge, its acquisition and the nature of learning
- data-intensive research and care provision will change the dominant intellectual skills and competencies
  - the looming talent gap in biomedical informatics and large scale data analytics
- disruptive change: new participants; new organizational frameworks; new business models

The Imperative for Redesign of Biomarker Discovery and Validation: Adoption of Integrated Systems-Based Approaches

- biomarkers as a key intellectual foundation for rational clinical decisions (precision medicine)
- contrarian courage to declare that major change is needed versus safe refuge of herd mentalities and status quo
- withstand denial and backlash from entrenched defenders of the status quo
- change is not easy: major cultural and economic disruptions are difficult
- commitment to patients (current and future)
- training a new generation of researchers and clinicians and informaticians able to participate in systems-based efforts