ASU Complex Adaptive Systems Initiative
Arizona State University

Burrill & Company
Life Sciences: Venture Capital, Private Equity, Merchant Banking, Media

Integrative Omics (iOmics) and the Evolution of Molecular Medicine

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

Keynote Presentation 8th Annual Burrill Personalized Medicine Meeting
September 13, 2012 ● San Francisco, CA

The Healthcare Environment: Access, Cost, Efficiency and Outcomes

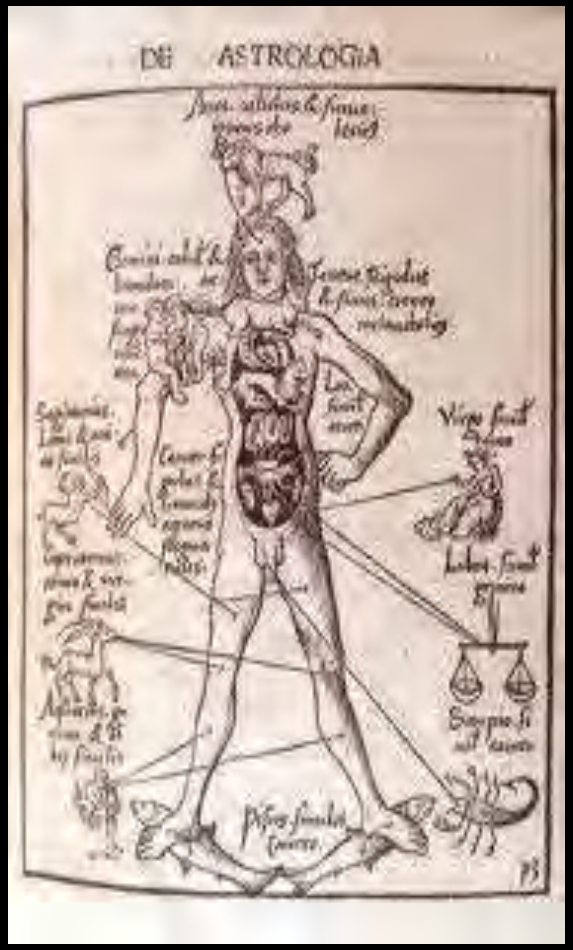
Healthcare: An Expensive Menu Without Prices

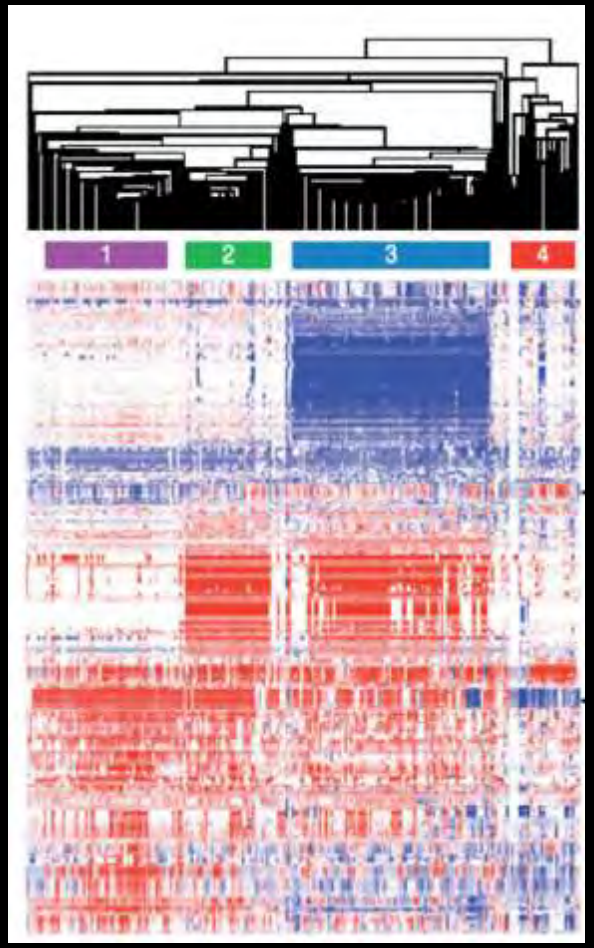
Sustainable Health: Societal (Economic) and Individual (Wellness)

Managing the Demands of an Aging Society and Chronic Disease Burden in an Era of Economic Constraint

From a "Do More, Bill More" Healthcare System to Managing Individual Risk to Improve Health Outcomes and Control Cost

Medical Progress:
From Superstitions to Symptoms to Signatures

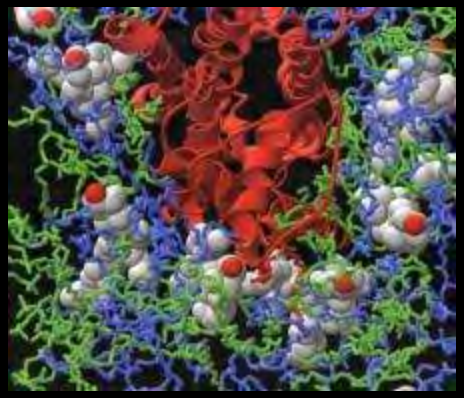




Mapping The Molecular Signatures of Disease: The Intellectual Foundation of Rational Diagnosis and Treatment Selection

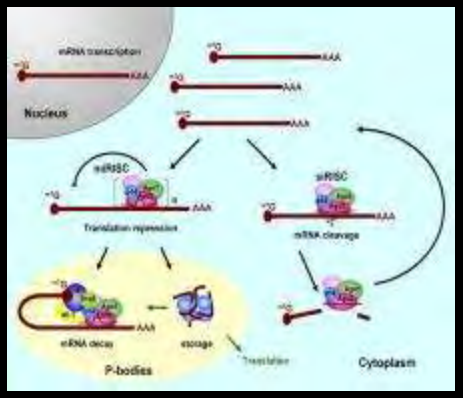
Genomics

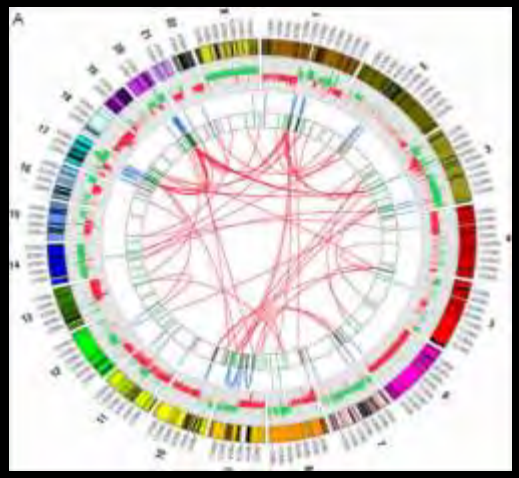
Proteomics



Molecular Pathways and Networks

Network Regulatory Mechanisms

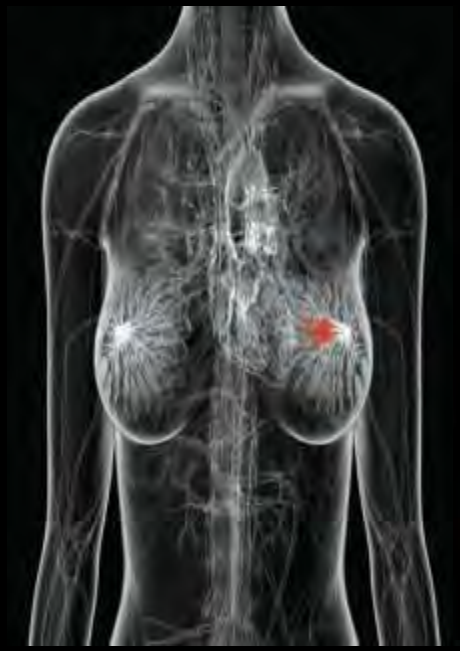




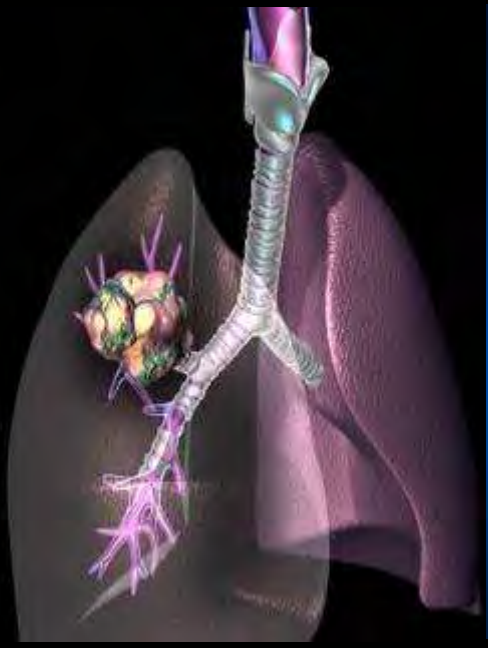
ID of Causal Relationships Between Network Perturbations and Disease

Patient-Specific Signals and Signatures of Disease or Predisposition to Disease

Mapping the Molecular Signatures of Disease, Disease Subtyping and Targeted Therapy: Companion Diagnostics and the Right Rx for the Right Disease (Subtype)

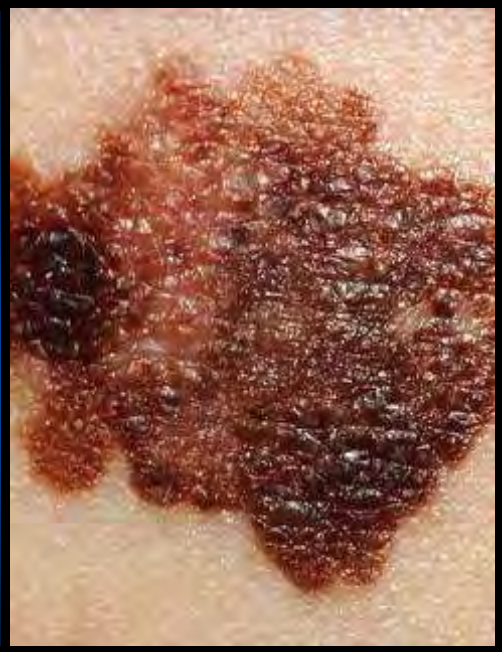


Her-2+
(Herceptin)



EML4-ALK
(Xalkori)

KRAS
(Erbitux)
(Vectibix)



BRAF-V600
(Yervoy)
(Zelboraf)

The Evolution of Molecular Medicine

Integrated Personal Omics Profiles (iPOP)

Personalized Medicine

Precision Medicine

Targeted Therapeutics

P4 Medicine

Genomic Medicine

Individualized Medicine

Disease Subtyping

Systems Medicine

Pharmacogenomics

Patient Stratification

Whole Genome Sequencing

Personal Genomics

Predisposition Risk Profiling

m.Health

Quantified Self

Microbiomics

New Standards of Care

New Value Propositions, Delivery Channels and Business Models

Personalized Medicine: Competing Claims

- many physicians believe they already do it
- personalized medicine is just hype
- personalized medicine threatens many established constituencies in the current healthcare ecosystem
- personalized medicine will be so expensive as to be unaffordable
- personalized medicine is the logical, inevitable outcome of understanding disease at the level of alterations in molecular information networks
- personalized medicine represents the intellectual foundation for rational care, improved outcomes and cost control

Reactive, Episodic and Fragmented Healthcare Delivery: Symptoms-Based, Erratic Paper Records and Medical Paternalism

**"Sorry For the Wait.
How Have You Been Since Your Last Visit?
Any stiffness?"**

New Yorker 8/13/12

Integrative Omics (iOmics) and Personal Omics Profiling (the iPOP): Dynamic Tracking of Molecular and Medical Phenotypes

Dr. M. Snyder
Stanford Univ.

Cell (2012)
148, 1293

- profiling genomic, transcriptomic, proteomic, metabolic and autoantibody responses over 14 months
- extensive dynamic heteroallelic changes in health, two viral disease episodes and onset of T2D

The Integrative Personal Omics File and Longitudinal Tracking of Health Status and Risk

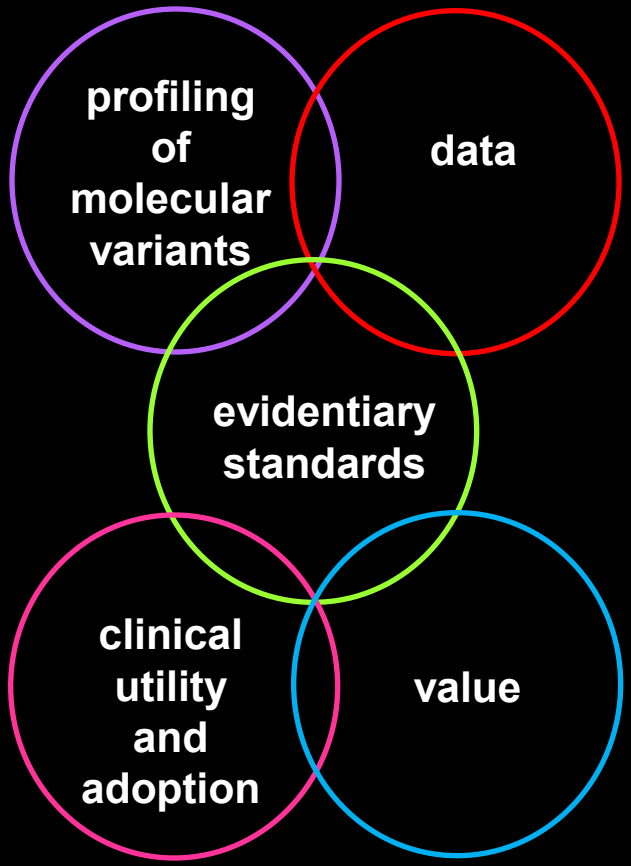
Newborn Baseline Profile

HARD TO HEAL: Screening infants' blood for signs of disease may not make sense if effective treatment does not exist.

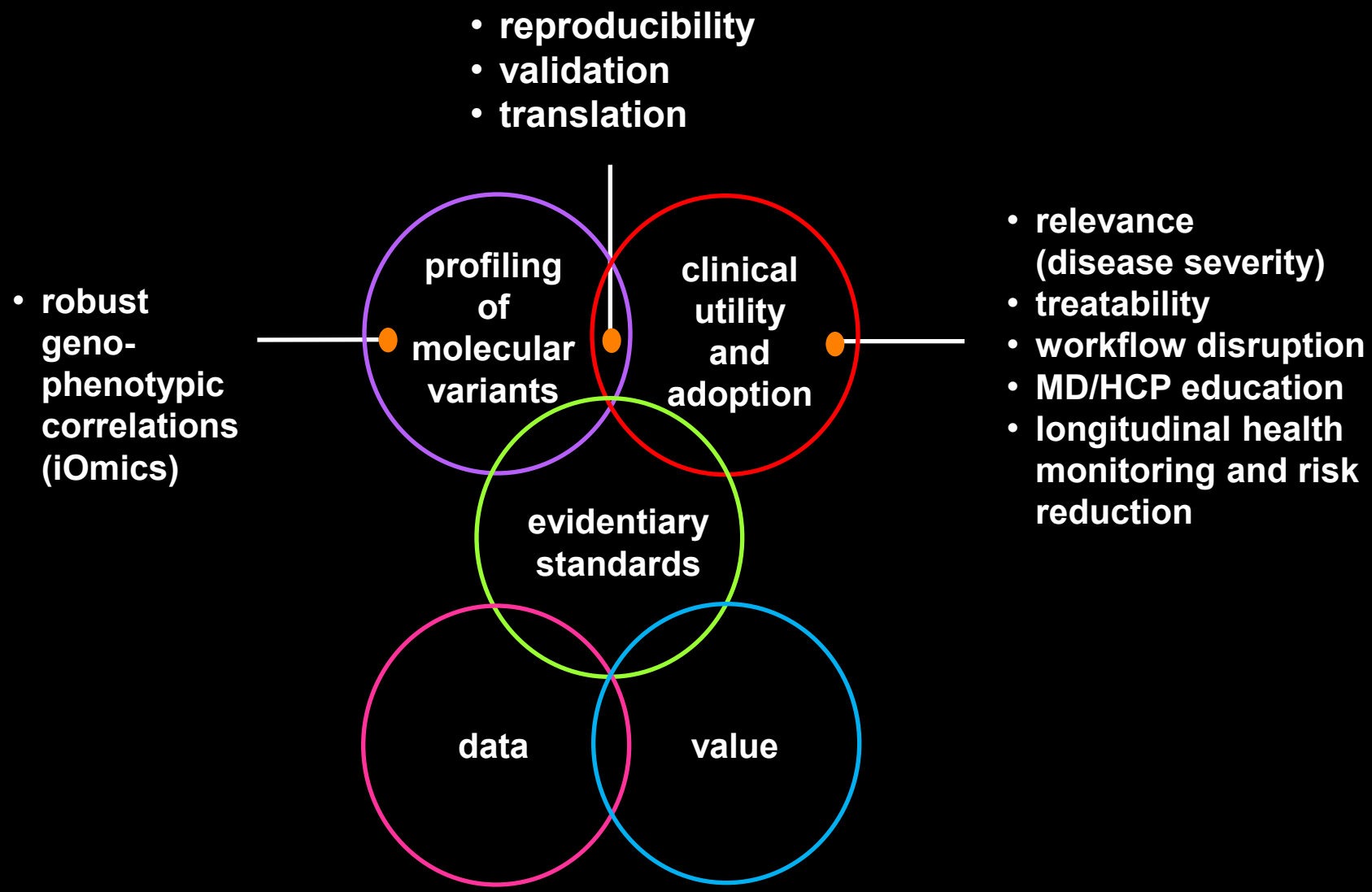
Each Individual as Own Control

Risk Mitigation and Sustaining Health (Wellness)

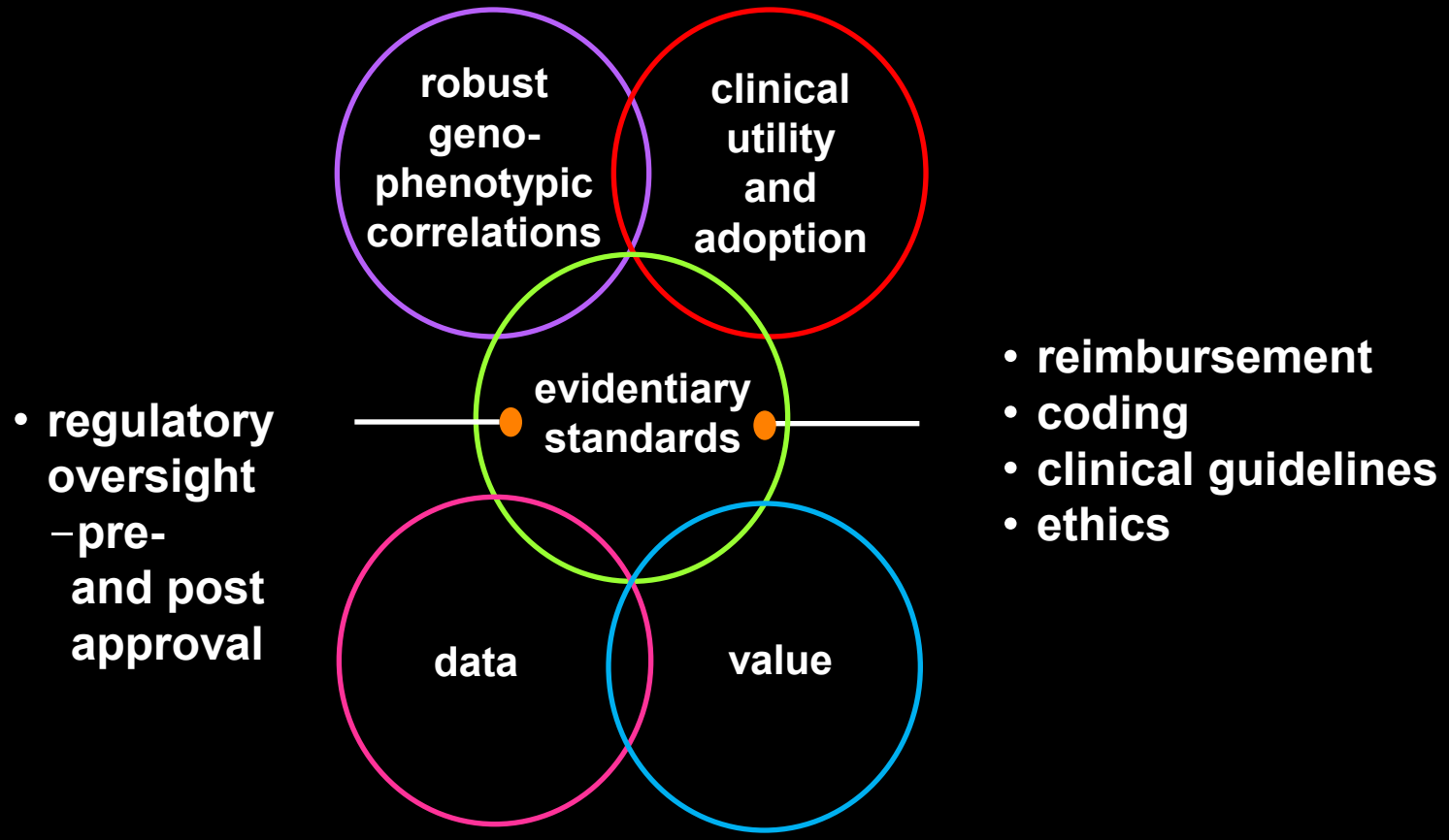
Analytical and Clinical Validation of Molecular Determinants of Disease, Treatment Options and Predisposition Risk



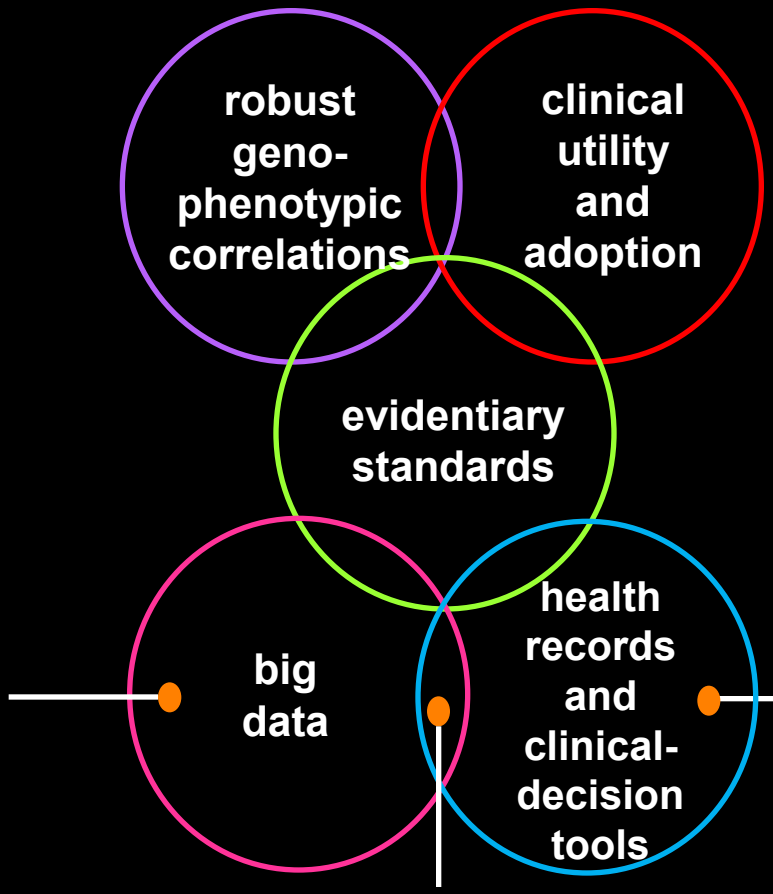
Analytical and Clinical Validation of Molecular Determinants of Disease, Treatment Options and Predisposition Risk



Regulation, Reimbursement and Rapidity in Updating Clinical Guidelines Will Define the Adoption Trajectory for Molecular Medicine



Data:
The Foundation of Molecular Medicine, Risk Reduction and Optimizing Wellness



- V3: volume, velocity, variety
- standards
- integration
- curation
- security

- automated analytics

- diagnostic precision/ Rx selection
- improved outcomes
- risk mitigation
- consumer/patient engagement

The New ROI

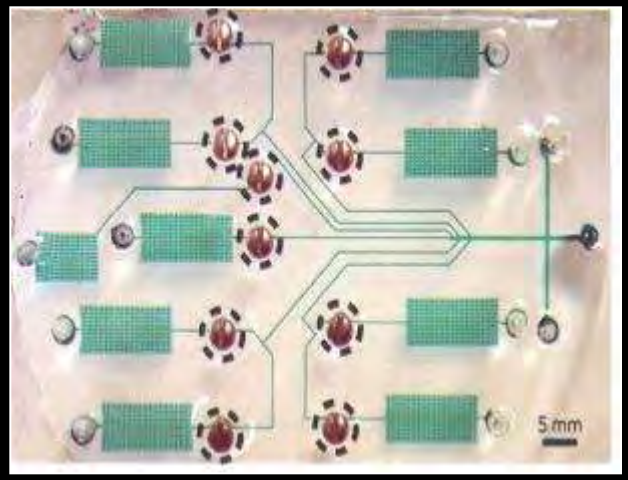
Return on Actionable Information

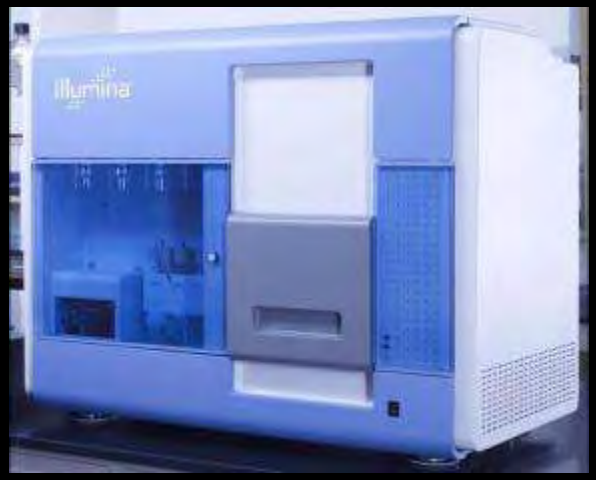
Multiplex Biomarkers, Sequenced Genomes and Integrated Personal Omics Profiles

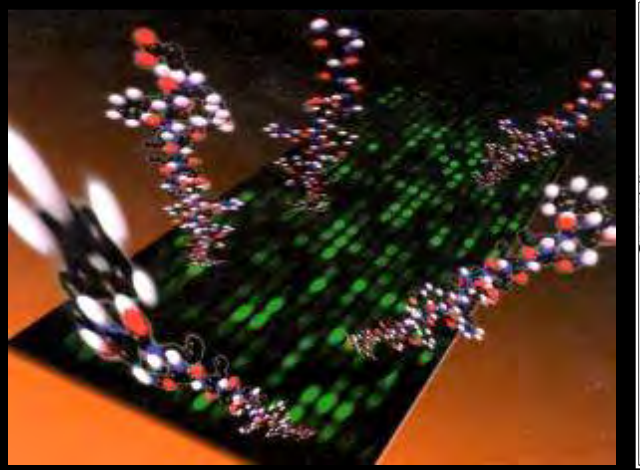
Promises, Pitfalls and Realizing the Potential

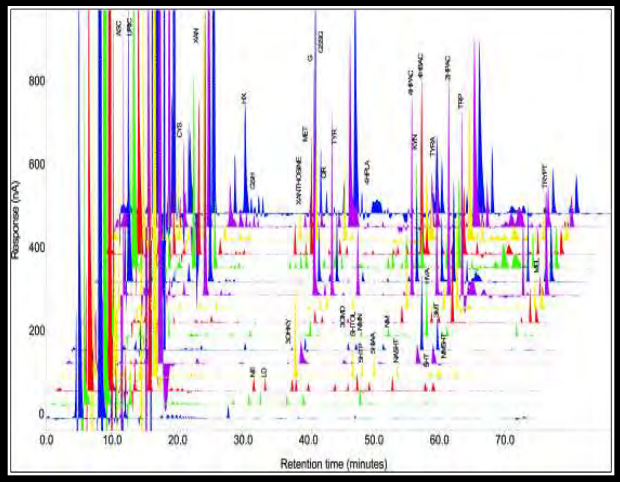
Technology Platforms for Integrative Omics Profiling: Integration of Multiple High Dimensionality Datasets

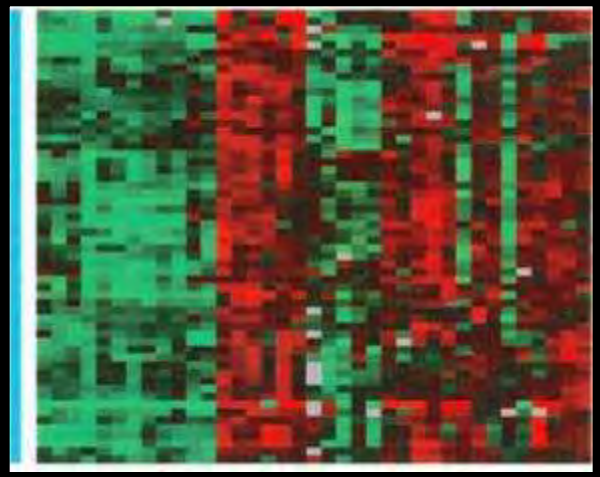
Miniaturization, Massive Parallelism, Automation and High Throughput Measurements of Diverse Analytes





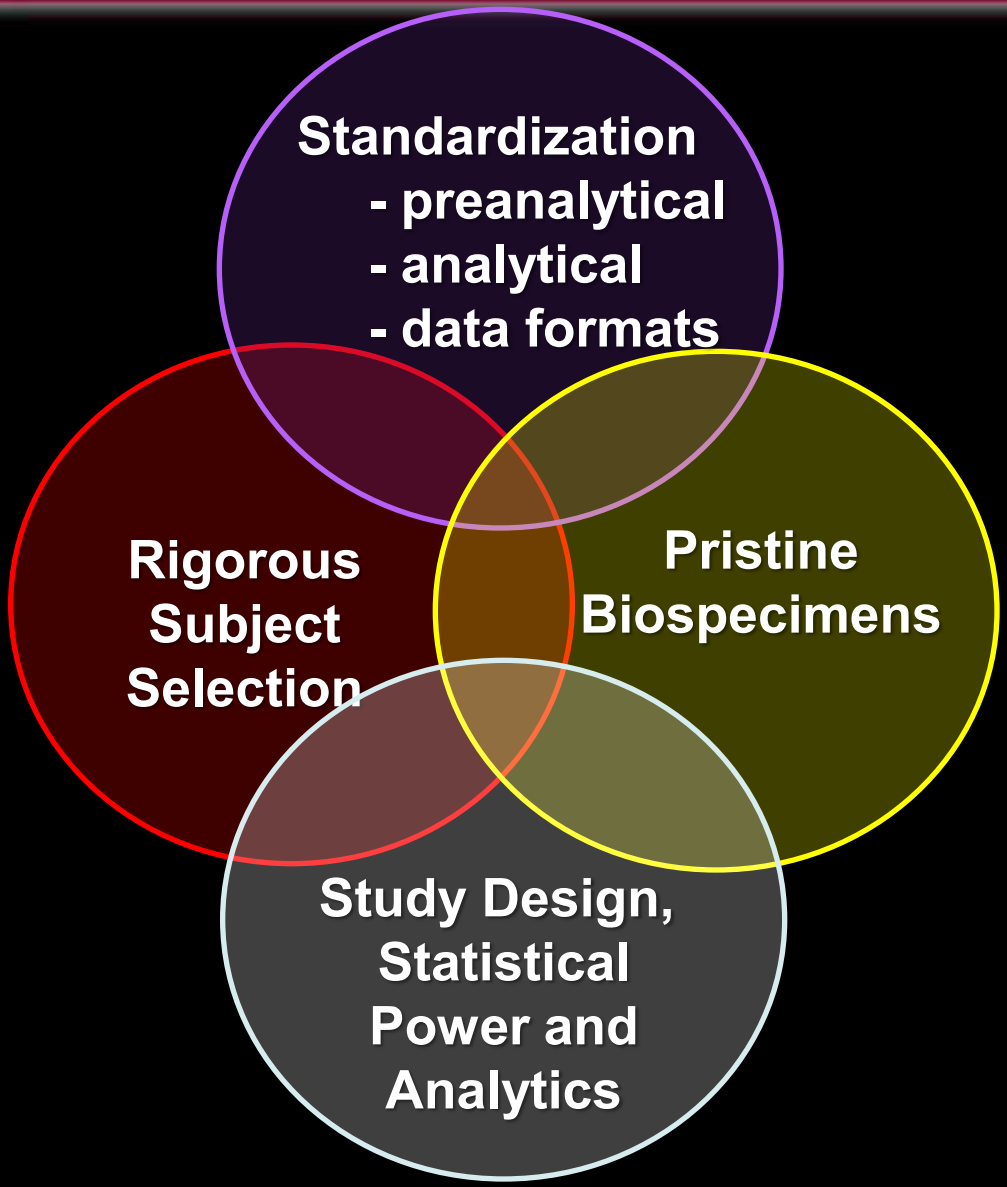






Complex Multiplex Signals, Deconvolution, Multivariate Analysis = Big Data

Integrated Omics Profiling to Identify Causal Determinants of Disease or Disease Risk Demands a Rigorous Systems-Based Approach

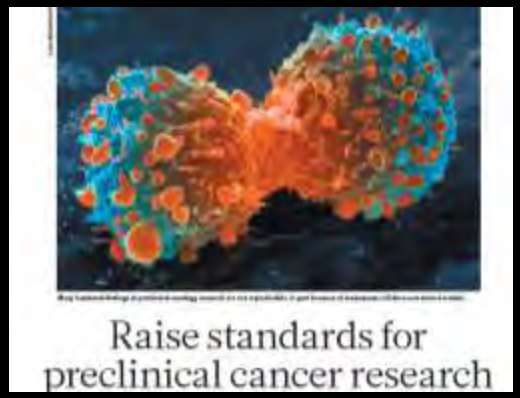


Lack of Standards and Shoddy Science: Increasingly Pervasive Problems in Academic Biomedical Research

Poor Replication and Reproducibility

Slow Adoption of Standards

Nature (2012) 483,531



Failure of Academia to Work to Industry Standards

Nature Rev. Drug Disc. (2011) 10, 643

Reliability of 'new drug target' claims called into question

The Small 'N' Problem

JAMA (2011) 305, 2200

Comparison of Effect Sizes Associated With Biomarkers Reported in Highly Cited Individual Articles and in Subsequent Meta-analyses

Statistical Flaws and Bias

Nature (2012) 485, 149

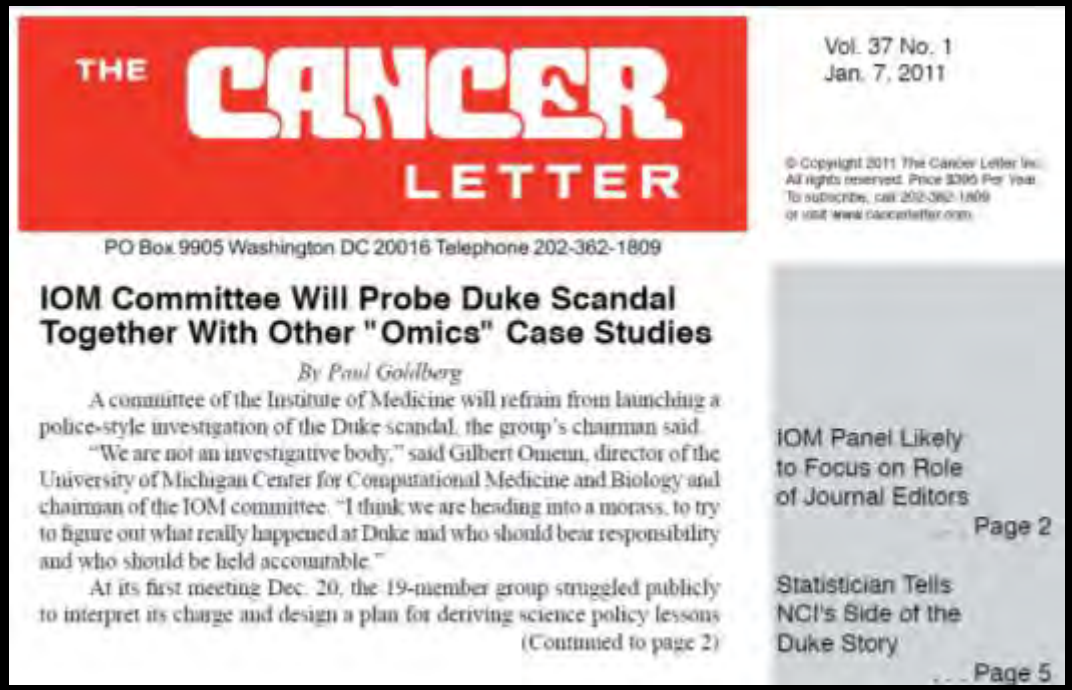
Inefficient Translation

Nature 5 April 2012

Duke
UNIVERSITY

NATIONAL® CANCER INSTITUTE

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES



THE CANCER LETTER

Vol. 37 No. 1
Jan. 7, 2011

PO Box 9905 Washington DC 20016 Telephone 202-362-1809

IOM Committee Will Probe Duke Scandal Together With Other "Omics" Case Studies

By Paul Goldberg

A committee of the Institute of Medicine will refrain from launching a police-style investigation of the Duke scandal, the group's chairman said.

"We are not an investigative body," said Gilbert Omenn, director of the University of Michigan Center for Computational Medicine and Biology and chairman of the IOM committee. "I think we are heading into a morass, to try to figure out what really happened at Duke and who should bear responsibility and who should be held accountable."

At its first meeting Dec. 20, the 19-member group struggled publicly to interpret its charge and design a plan for deriving science policy lessons

(Continued to page 2)

Evolution of Translational Omics: Lessons Learned and the Path Forward

Released: March 23, 2012
Type: Consensus Report
Topics: Biomedical and Health Research, Health Services, Coverage, and Access
Activity: Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials
Board: Board on Health Care Services

Genotyping Comprehensiveness and Tamoxifen Benefit by Exlcusion of Metabolically Inactive CYP2D6 Isoforms in ER$^+$ Breast Cancer

Findings	Study	Number of SNPs in analysis	Alleles interrogated in analysis	Note
Negative	Okishiro et al. [28]	1	*10	
	Toyama et al. [65]	1	*10	
	Wegman et al. [59]	1	*4	
	Stingl et al. [66]	1	*4	
	Wegman et al. [12]	1	*4	
	Nowell et al. [29]	1	*4	*3 and *6 assayed but excluded
	Lash et al. [23]	1	*4	Quantitative bias analysis
	Park et al. [60]	3	*5, *10, *41	
	Abraham et al. [32]	6	*4, *5, *6, *9, *10, *41	Tag SNPs included
	Kiyotani et al. [27]	6	*4, *5, *10,*21, *36, *41	Gene duplications included
	Ramón et al. [67]	29	Roche AmpliChip™	Gene duplications included
Positive	Xu et al. [45]	1	*10	
	Bijl et al. [30]	1	*4	
	Newman et al. [24]	4	*3, *4, *5, *41	
	Schroth et al. [14]	5	*3, *4, *5, *10, *41	Gene duplications included
	Kiyotani et al. [13]	7	*4, *5, *10, *14, *21, *36, *41	Gene duplications included
	Thompson et al. [31]	29	Roche AmpliChip™	Gene duplications included

Abbreviation: SNP, single nucleotide polymorphism.

From: D. L. Hertz et al. (2012) The Oncologist 17, 620

Sloppy Science and Inadequate Reproducibility

"There is no 'bad ingredient central clearinghouse'
where one can report analyte failures.
Maybe we need an 'Angie's List'
for antibodies and cell lines."

"Underlying everything else:
a system that encourages getting published first
rather than getting the science right
In short, a system based on shortcuts".

Dr. George W. Sledge, Jr.
Indiana University Simon Cancer Center,
Past President, ASCO
Oncology Times, 25 May 2012, p. 28

REPRODUCIBILITY INITIATIVE

"The N of One: The Large N" Dilemma in High Dimensionality Biomarker Profiling

N of One

- individualized patient biomarker profiling for clinical diagnostic subtyping and/or Rx selection for *validated* biomarker

Large N

- validation of *putative* marker(s) requires large, statistically powered sample sets
  - high dimensionality markers (10^3 - 10^6)/WGS using very small sample sets (10^1 - 10^2) results in inevitable overfitting
  - large N of 10^3 - 10^4 samples needed for robust validation cohorts
  - logistics and cost of screening candidate pool for low frequency markers (e.g., ALK, ROS in NSCLC)

Access to High Quality Biospecimens, Biobanks and DNA Repositories: An Obligate Prerequisite to Productive Validation of Putative Causal Disease Markers

requisite
scale
and
audited
QA/QC
standards

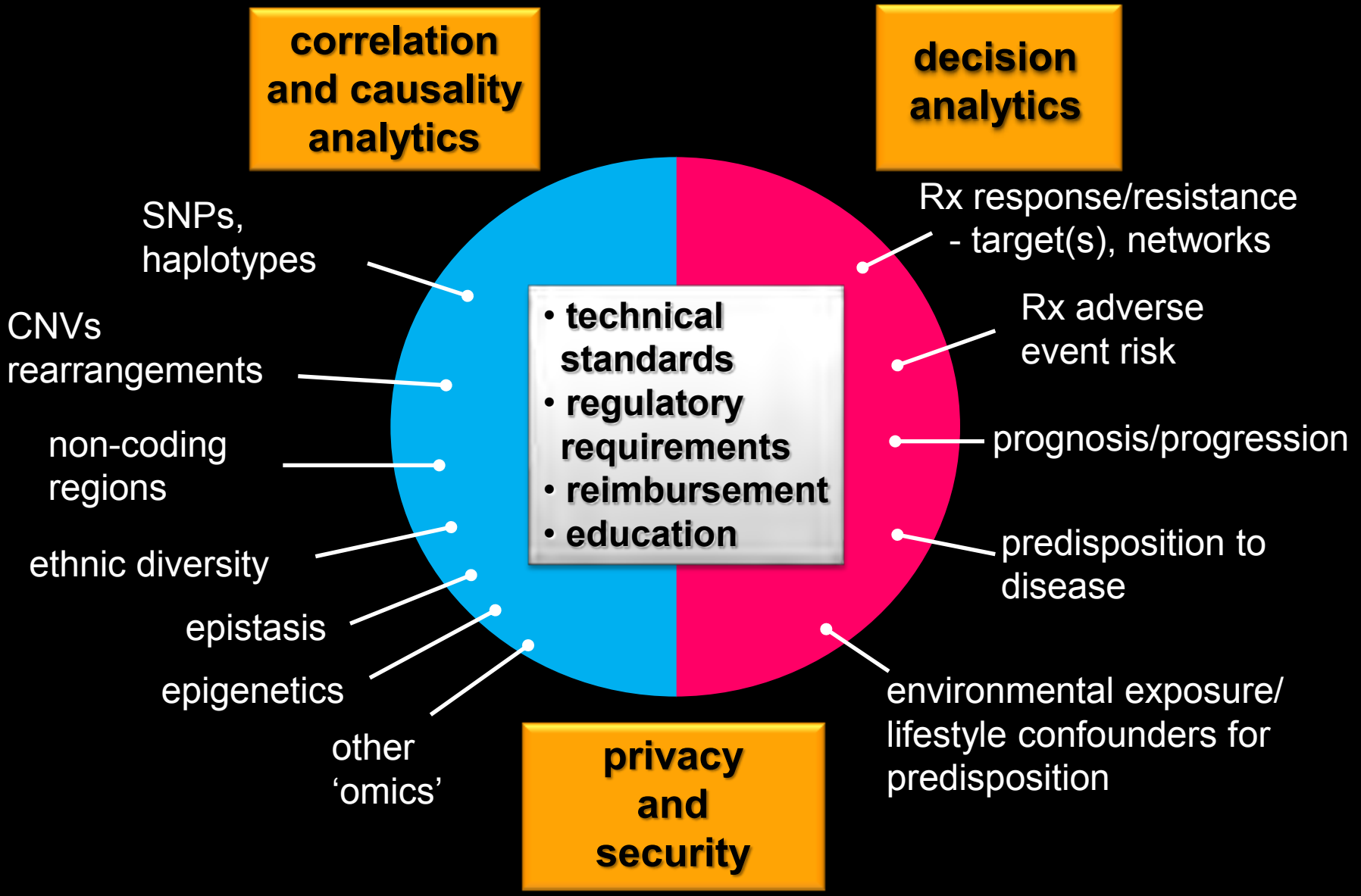
or

mere
academic
anecdotes
and
wasted
investment

Large Scale Genome Sequencing: Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS)

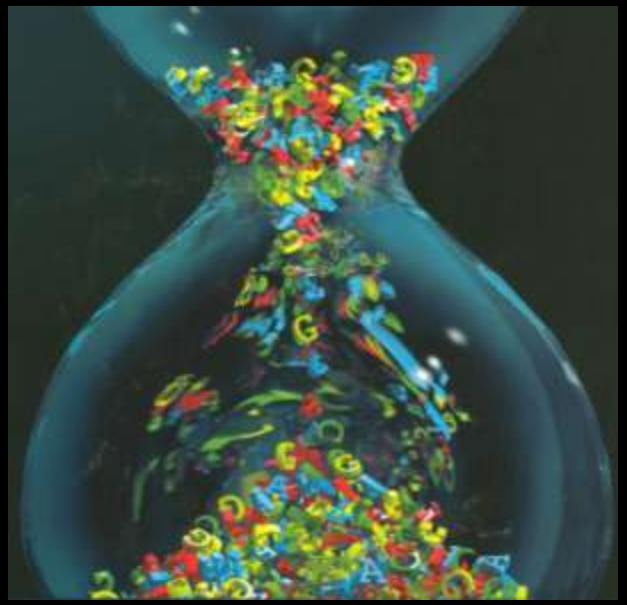
- from research odysseys to a routine clinical tool ("just another lab value!")
- early clinical applications
  - rare diseases of suspected genetic etiology
  - hereditary cancers and individual risk assessment
  - HLA profiling for transplant matching
  - oncology
  - infectious diseases
  - cardiomyopathy
  - X-linked intellectual disability
  - congenital muscular dystrophy
  - mitochondrial disorders

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



What Is A Complete and Accurate Analysis of Genome Sequence, Architecture, Topology and Genomic Regulatory Networks?

Current Chokepoints and Challenges in Adoption of Personal Omics Profiling Data for Clinical Decisions



- production of sequencing data outstripping interpretational capacities
- CLIA compliance and other regulatory requirements for clinical decisions
- confusing maze of base calling, alignment, assembly and analysis tools
- many software tools insufficiently robust and/or customized for one type of data or sequencing platform
- variation in clinical significance predictions from different algorithms using well known algorithms (SIFT, PolyPhen, LRT, MAPP, VarioWatch)
- (comparable data standardization/ validation problems in large scale proteomics)

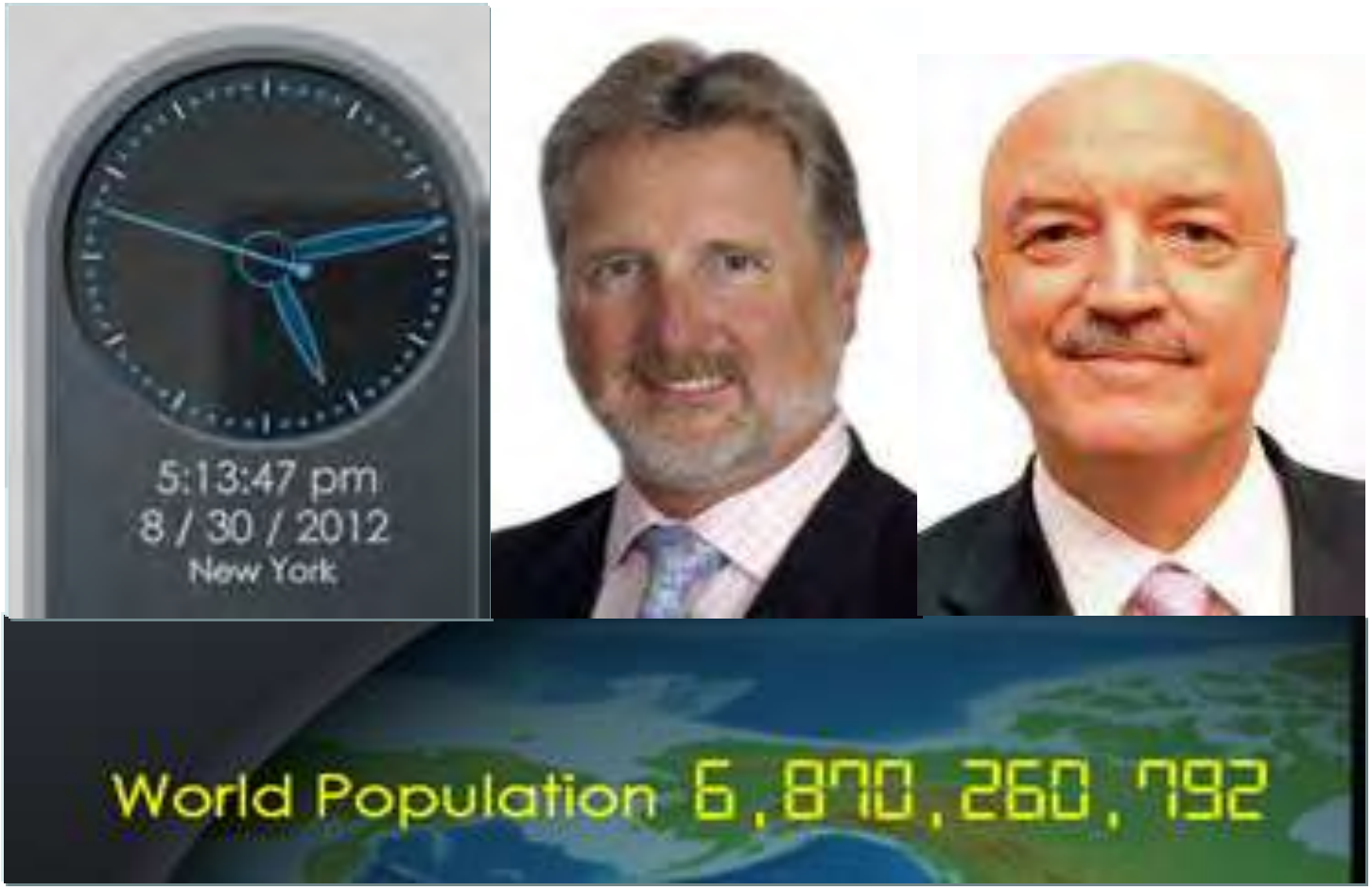
The Imperative for Regulatory Clarity Regarding Test Classification and Analytical Standards for Molecular Diagnostics and WGS

- regulatory classification as LDTs vs. 510(k)/PMA submission?
- sequencers as Class III devices?
- FDA enforcement of Quality Systems Regulations (QSR; 21CFR820)
  - laboratories and suppliers
  - already imposed on medical device industry and FDA-cleared IVD products
- concern over QA for RUO reagents not manufactured using QSRs
  - action to forbid RUO materials when QSR-grade available
- need for Certified Reference Materials with high-level QA and traceability
  - Genome in a Bottle Consortium (2012)

IHE-LAW
A Major Advance for Integration of Diagnostic Laboratory Automation, Information Systems and Electronic Health Records

- partnership between Integrating the Healthcare Enterprise (IHE) and In Vitro Diagnostics Connectivity Consortium (IICC)
- IHE-LAW (Laboratory Analytical Workflow) standard
  - uniform IT connectivity standards for LIS, automation systems, middleware, CPOE and EHR
  - use of ISO HL7 messaging
- participation of leading instrument manufacturers
  - Abbott, BD, Beckman Coulter, BioMerieux, Ortho, Roche, Siemens
- projected final standard in collaboration with The Clinical and Laboratory Standards Institute (CLSI) in 2013

Mapping Individual Genome Variation and Patterns of Disease Risk and Progression

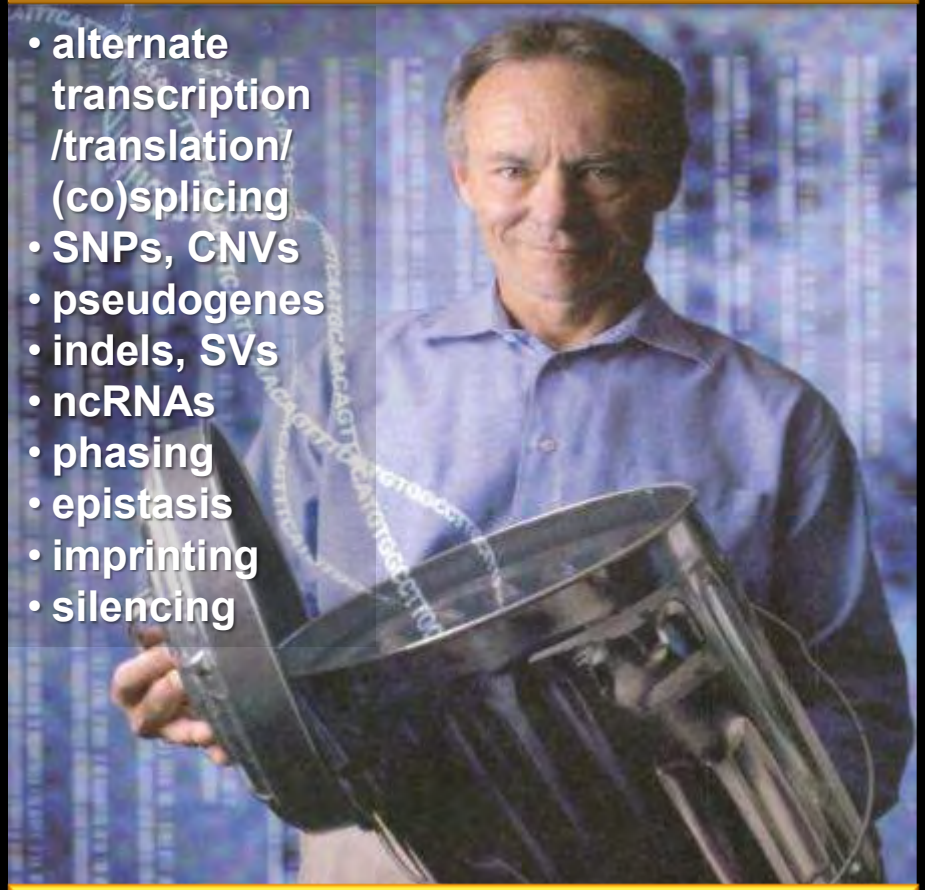


Genes For …..

The Overtly Simplistic and Deterministic Dangers of a Genome-Sequence Centric Perspective

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

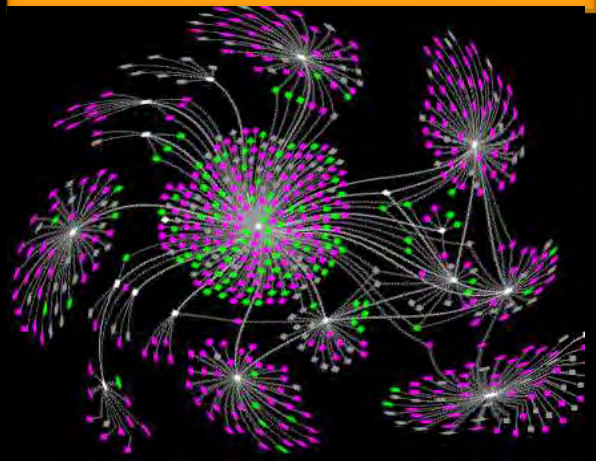
Junk No More: Pervasive Transcription



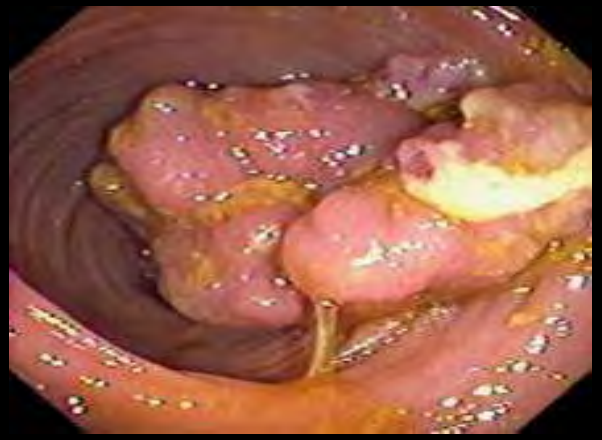
- alternate transcription /translation/ (co)splicing
- SNPs, CNVs
- pseudogenes
- indels, SVs
- ncRNAs
- phasing
- epistasis
- imprinting
- silencing

recognition of genome organizational and regulatory complexity

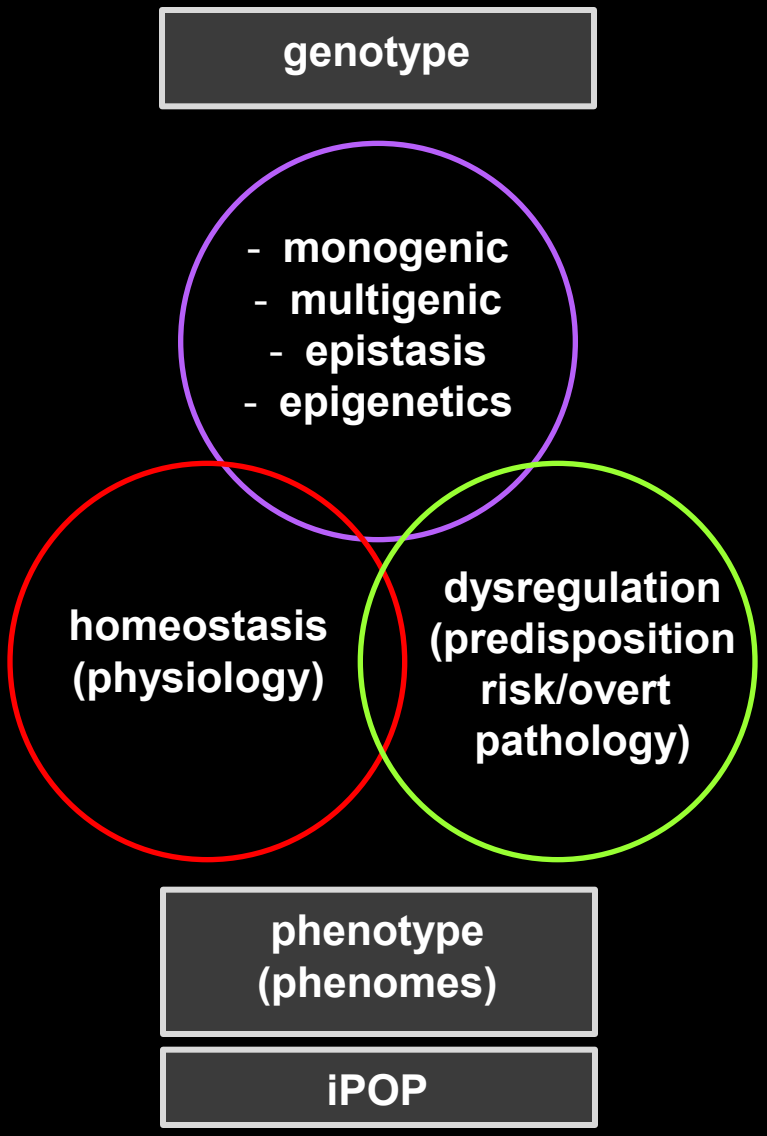
Cell-specific Molecular Interaction Networks



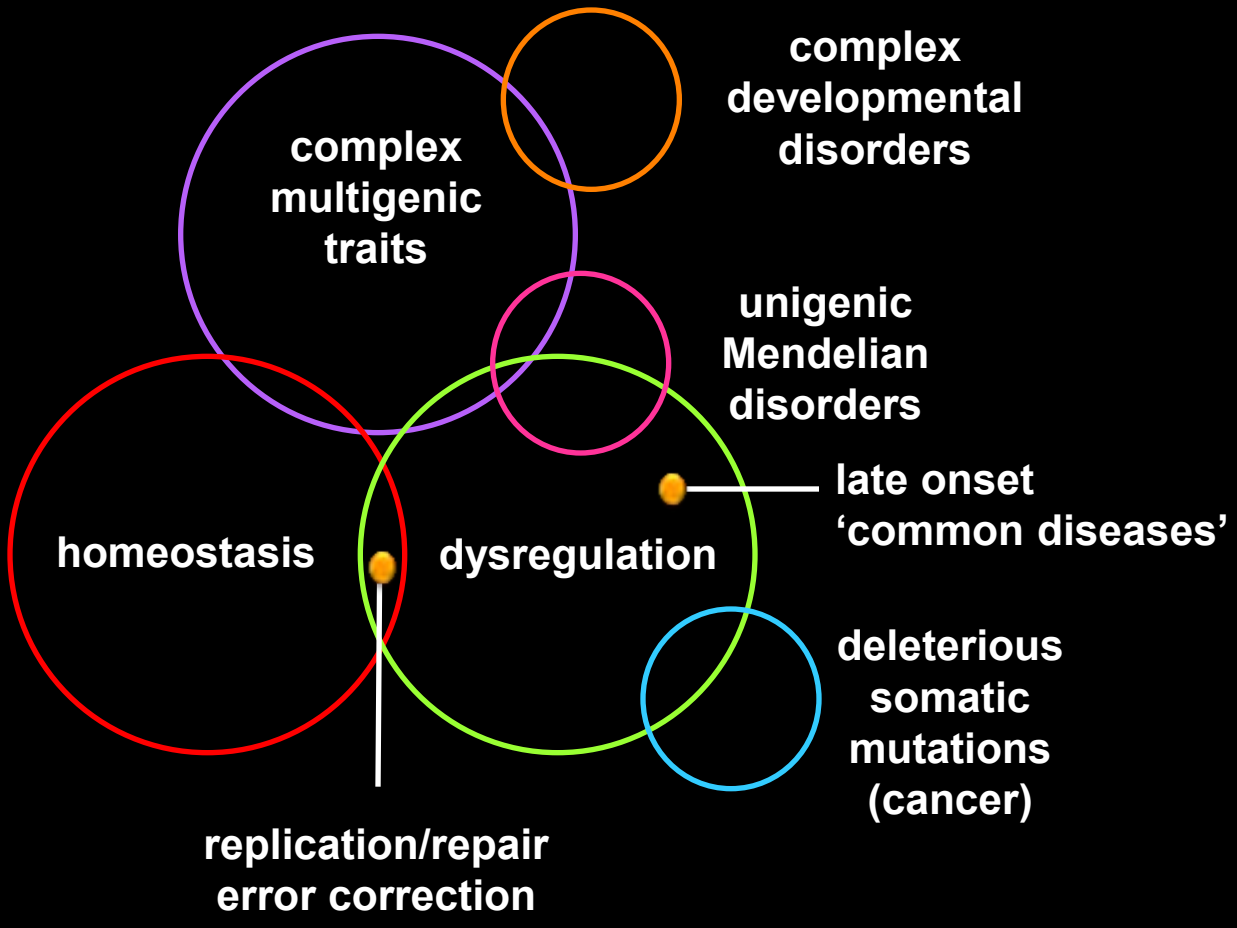
Perturbed Networks and Disease



Genotype-Phenotype Relationships: Integrative Personal Omics Profiling

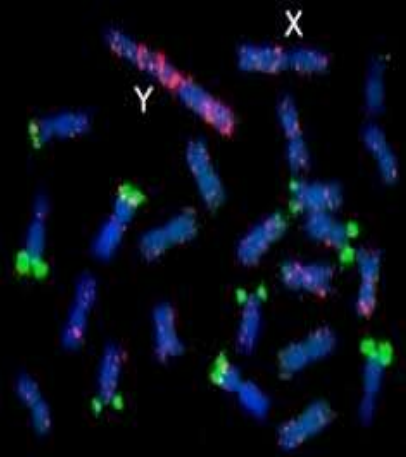


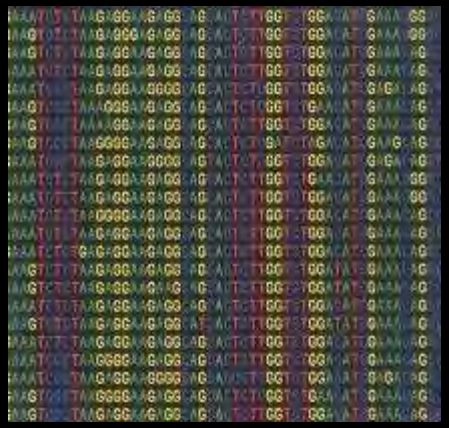
Phenotypes


complex developmental disorders
complex multigenic traits
unigenic Mendelian disorders
late onset 'common diseases'
homeostasis
dysregulation
deleterious somatic mutations (cancer)
replication/repair error correction
epigenetic changes

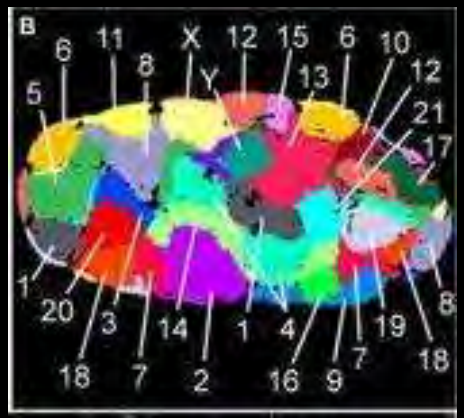
Genomes, Genetics and Individual Variation: From Elements to Networks

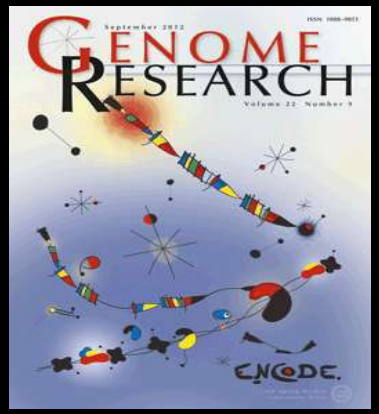
Unidimensional toponymy and local regulation of linearly ordered genes (flat genome)





topology of precision chromosome localization and folding (nuclear space) and global choreography via extensive networks of long-range *cis*-interactions (3-D genome)





The Human Genome (ENCODE 2012): Daunting Complexity

- protein-coding DNA = c.1.5% gnome= 20,687 protein-coding genes
- pervasive transcription
  - 93% bases transcribed into RNA
  - 18,400 non-coding RNA genes
  - 70,000 promoter regions, 400,000 enhancer regions
  - diverse transcription and (co)splicing processing patterns
- 11,244 DNA pseudogene regions with variable transcription
- 42% DNA accessible at 3.0 million sites for interaction with regulatory elements
- complex 3-D topology
  - average 3.9 distal (long range) DNA regions link with beginning of each gene

The Scale and Complexity of Human Genome Variation

- individual genomes on average carry:
  - 3.5 -4.0 million SNV, 1000 CNVs (>450bp)
  - 3-4 hundred indels
  - 200-500,000 private SNV
  - 20-400 loss-of-function variants
- estimated up to 60 new inherited mutations/generation
  - gender dependent transmission: maternal 15/paternal 25-45
  - impact of paternal age at fertilization

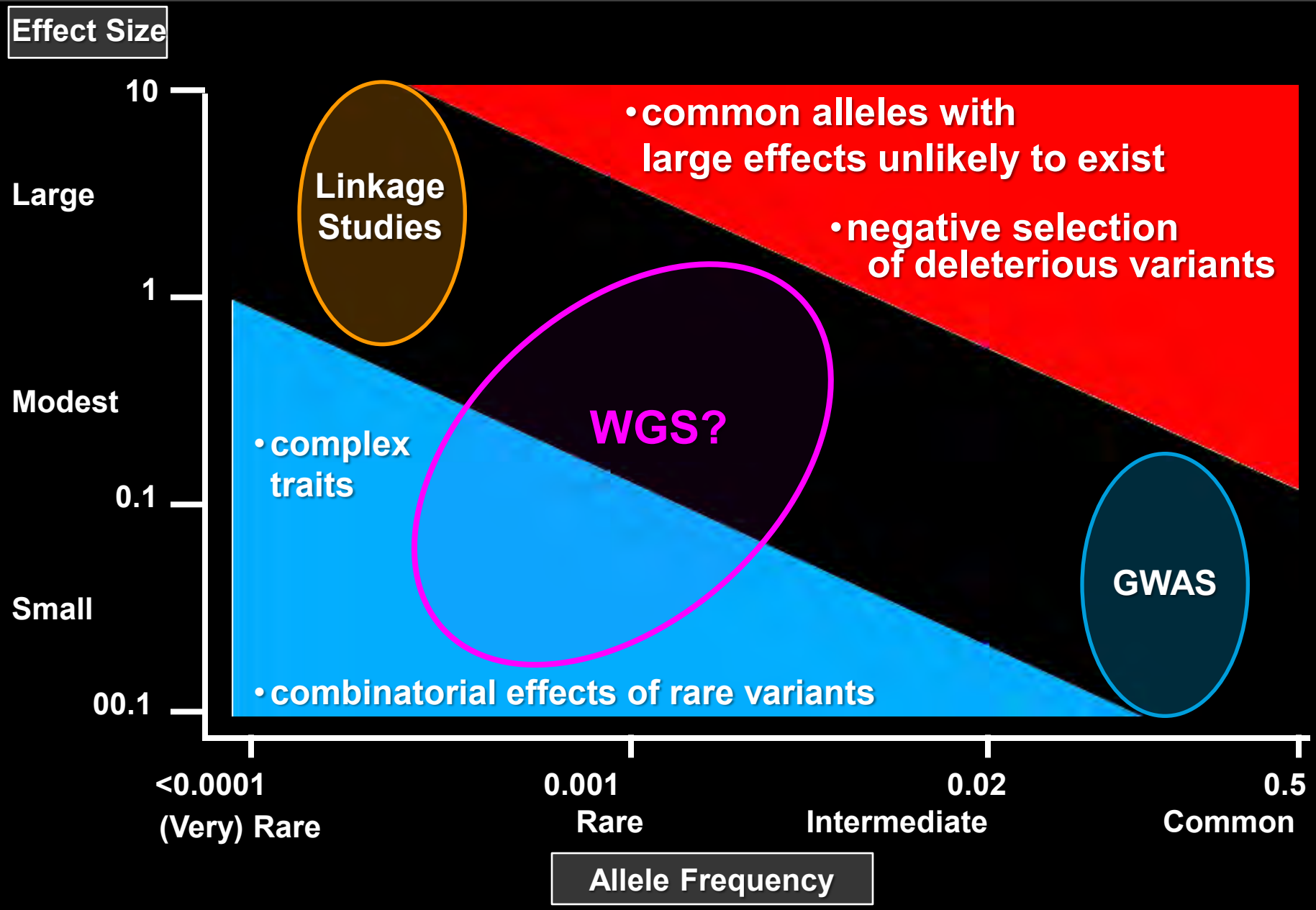
Mapping Human Genome Variation and Identification of Causal Variants for Disease

- both causal and protective alleles
- hypotheses
  - small number of common variants with large effects
  - large number of common variants with small effects
  - large number of rare variants with small effects

plus

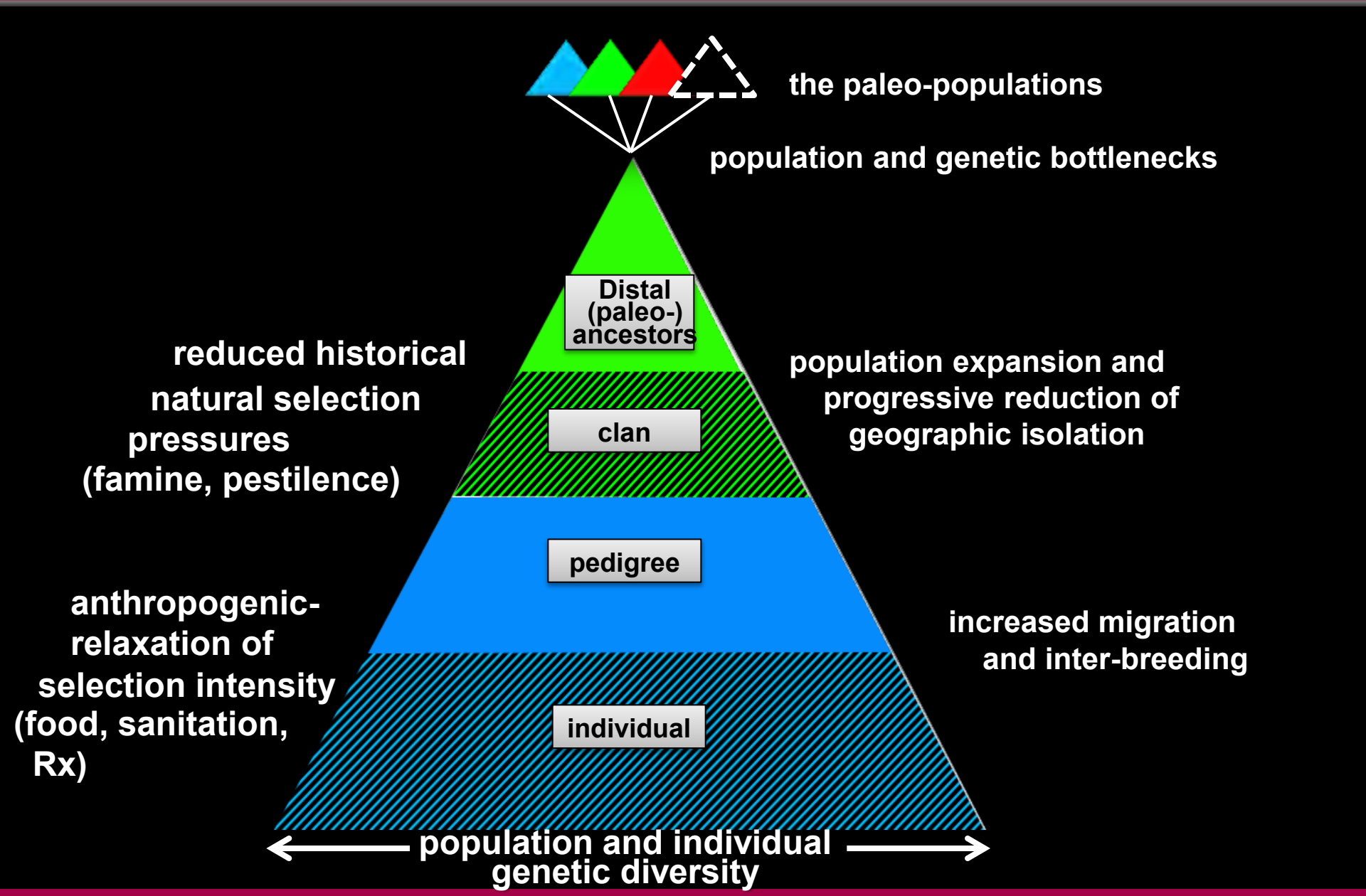
- role of environmental and epigenetic influences

The Relationship Between Allele Frequency and Effect Size and The Lack of Intermediate Frequency Associations

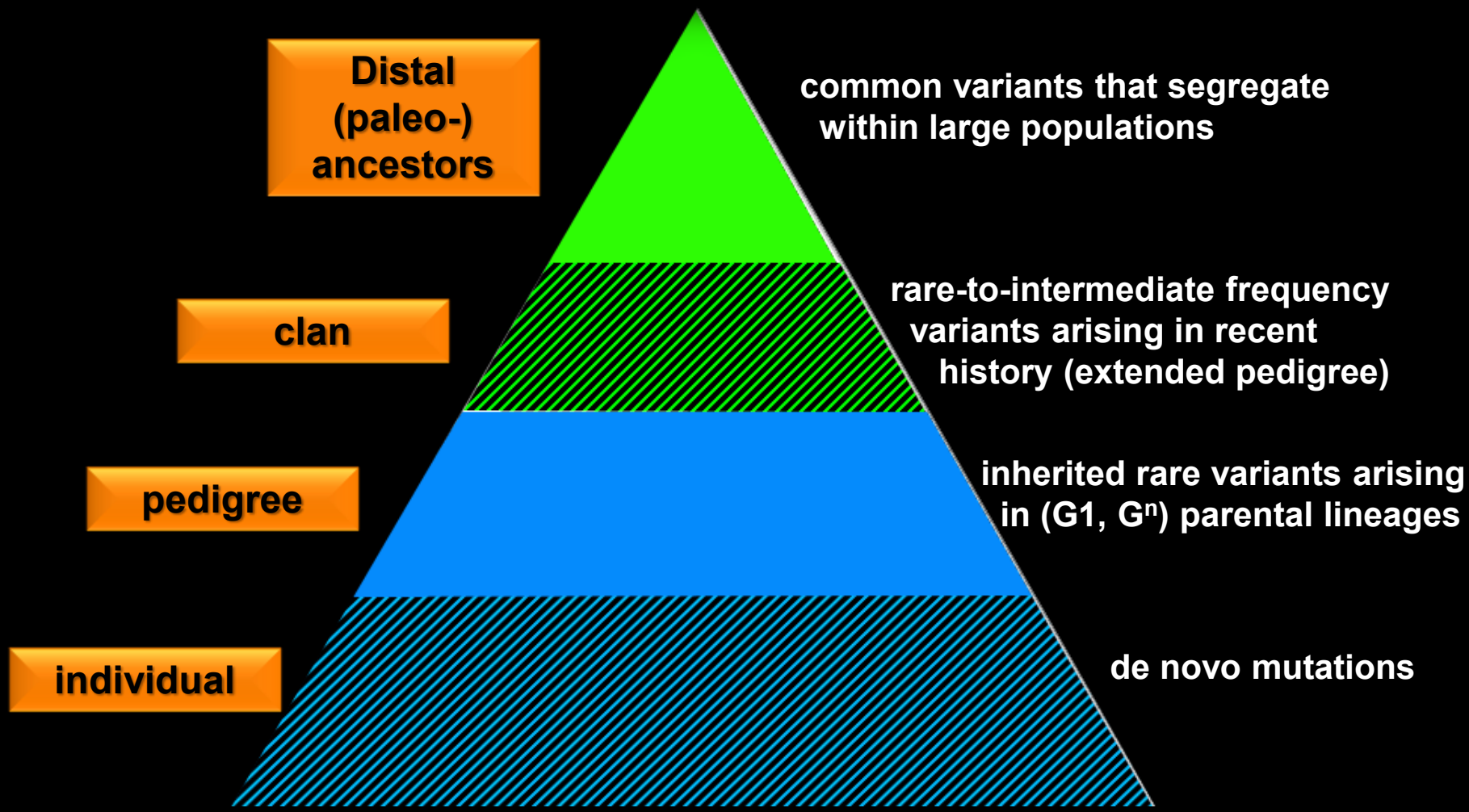


Are Humans Still Evolving?

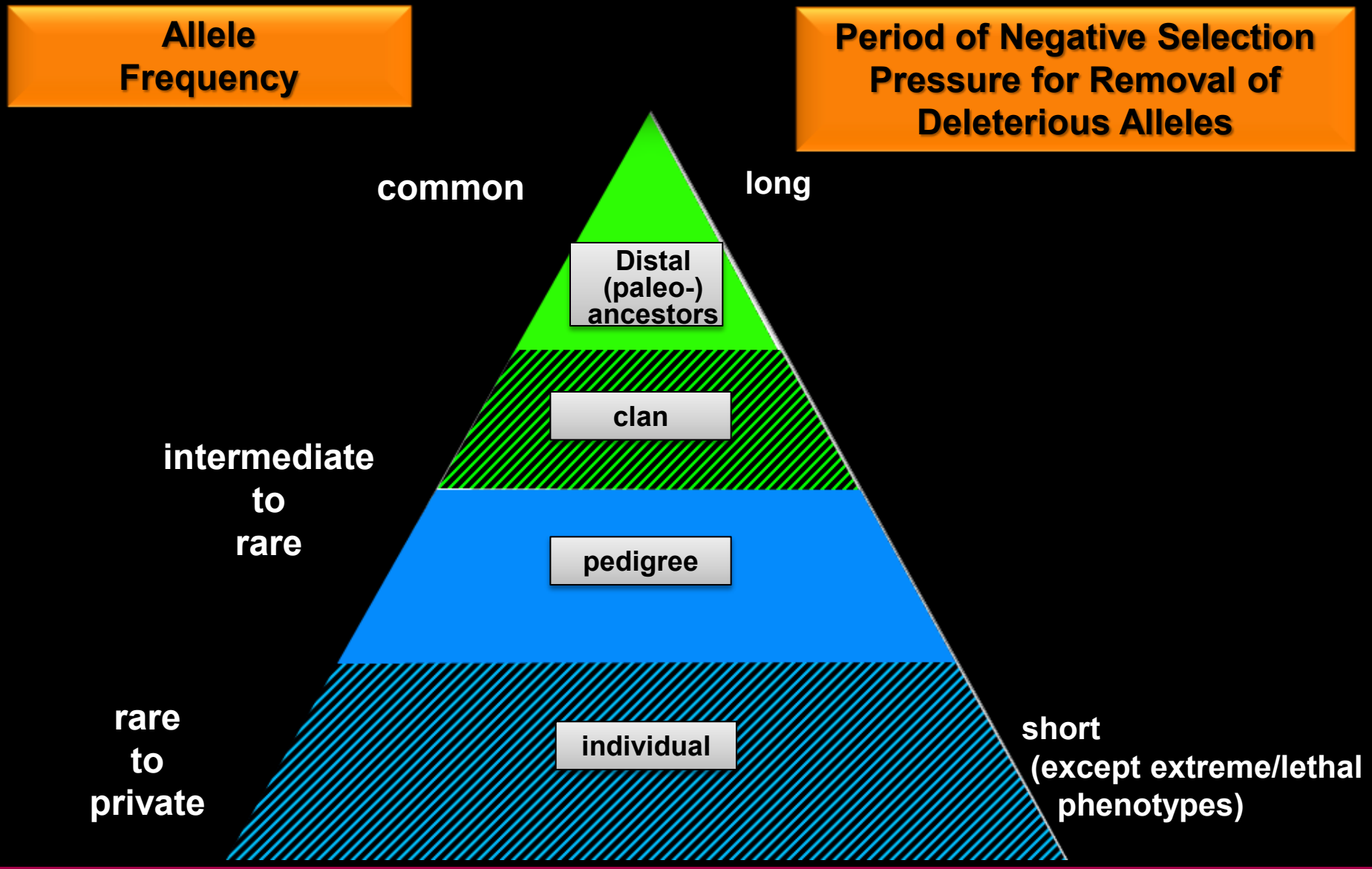
Human Genetic Diversity and Evolutionary History



Defining Genetic Variation in Human Populations: The Skewing of the Allele Frequency Spectrum Towards Rare and Private Genetic Variants



Defining Genetic Variation in Human Populations: The Skewing of the Allele Frequency Spectrum Towards Rare and Private Genetic Variants



Mapping The Spectrum of Human Genetic Variation: The Under-Explored Epigenome and the Chromatin Landscape

Distal (paleo-) ancestors

clan

pedigree

individual

epigenetics

rare/private

in utero and post-natal environmental effects "exposome"

parental age at reproduction

Regulatory DNA Domains as Major Source of Minor Frequency Causal Disease Variants

- disease-and trait associated SNP variants are concentrated in regulatory DNA
  - modify transcription factor recognition
  - alter allelic chromatin states
- impact of early gestational exposure and post-natal environmental insults on chromatin landscape and regulatory DNA
  - fetal origins of disease hypothesis
  - depletion of ‘fetal-type’ DHSs in aging-related disease, cancer, inflammatory disorders

Implications of Role of Rare/Private Variants in Disease for Identification and Validation Studies

- very large sample sizes (logistics, cost)
- replication of findings across different populations (ethnicity, geographic history) will be limited
- renewed focus on clan: pedigree cohorts to identify "recent" disease causal variants not yet purged by negative selection
- large scale data analytics on random cohorts may be less productive
- cancer presents unique challenges
  - extravagant scale of causal somatic mutations plus rapid progression of intra-and inter-lesional heterogeneity in advanced disease

Mapping Human Genetic Diversity: Transcending Political Correctness and Denial of Biology

- non-trivial genetically-based biological variation exists in individuals and groups
- ignoring such variations is illogical, poor science, poor clinical medicine and potentially dangerous
- mapping group genetic diversity is fundamental knowledge
  - human evolution and trait acquisition
  - interplay of genomes and environment in determining outcomes
  - variations in disease susceptibility, xenogeneic metabolism and clinical decisions for optimum treatment

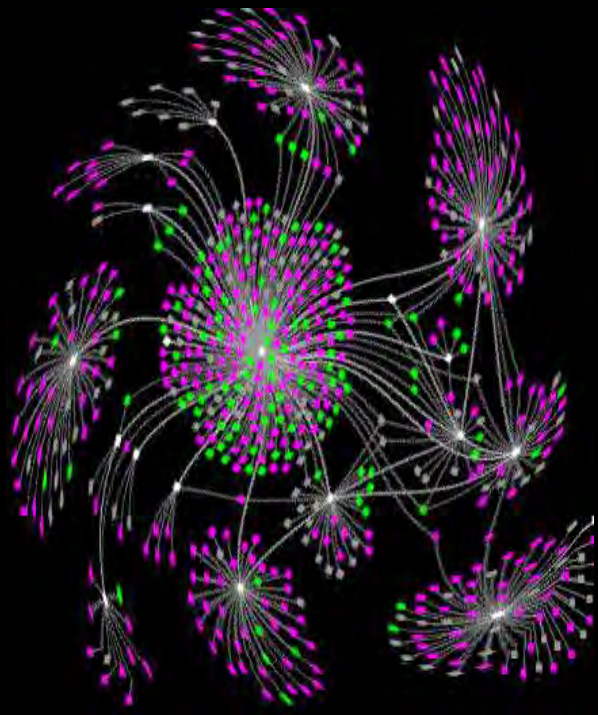
The Design and Validation of Facile Browsers for iPOP Data

Science (2012) 337, 1190

Systematic Localization of Common Disease-Associated Variation in Regulatory DNA

Matthew T. Maurano,[1*] Richard Humbert,[1*] Eric Rynes,[1*] Robert E. Thurman,[1] Eric Haugen,[1] Hao Wang,[1] Alex P. Reynolds,[1] Richard Sandstrom,[1] Hongzhu Qu,[1,2] Jennifer Brody,[3] Anthony Shafer,[1] Fidencio Neri,[1] Kristen Lee,[1] Tanya Kutyavin,[1] Sandra Stehling-Sun,[1] Audra K. Johnson,[1] Theresa K. Canfield,[1] Erika Giste,[1] Morgan Diegel,[1] Daniel Bates,[1] R. Scott Hansen,[4] Shane Neph,[1] Peter J. Sabo,[1] Shelly Heimfeld,[5] Antony Raubitschek,[6] Steven Ziegler,[6] Chris Cotsapas,[7,8] Nona Sotoodehnia,[9,9] Ian Glass,[10] Shamil R. Sunyaev,[11] Rajinder Kaul,[4] John A. Stamatoyannopoulos[1,12]†

Genome-wide association studies have identified many noncoding variants associated with common diseases and traits. We show that these variants are concentrated in regulatory DNA marked by deoxyribonuclease I (DNase I) hypersensitive sites (DHSs). Eighty-eight percent of such DHSs are activ during fetal development and are enriched in variants associated with gestational exposure–related phenotypes. We identified distant gene targets for hundreds of variant-containing DHSs that may expl phenotype associations. Disease-associated variants systematically perturb transcription factor recogniti sequences, frequently alter allelic chromatin states, and form regulatory networks. We also demonstra tissue-selective enrichment of more weakly disease-associated variants within DHSs and the de novo identification of pathogenic cell types for Crohn's disease, multiple sclerosis, and an electrocardiogram trait, without prior knowledge of physiological mechanisms. Our results suggest pervasive involvemen regulatory DNA variation in common human disease and provide pathogenic insights into diverse disord



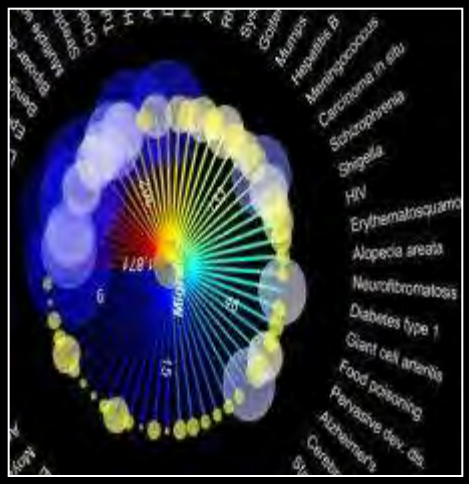
Causal Disease Variant Association (What)

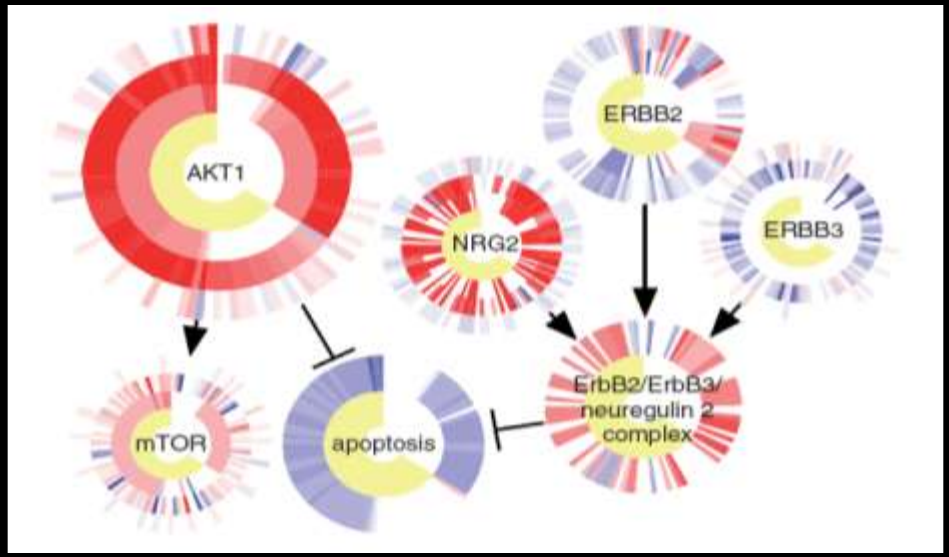
Variants to Network Perturbation (How)

The Design and Validation of Facile Browsers for iPOP Data

- V3: variety, volume, velocity
- data standards, quality and provenance
- interoperable formats and integration of diverse data feeds
  - from omics catalogs (research/clinical trials) to outcomes (clinical, epidemiological) and optimum care decisions
- new visualization tools for mapping and interactive analytics
  - from 1-D to 3-D genomes to dynamic networks
  - dynamic time series data to track individual health status and population disease risk burden

Integration of Omics Data Into Electronic Health Records and Clinical Decisions





Informatics and High Performance Computing

- interactive modeling and simulation of hierarchical omics networks of complexity
  - point-and-click analytics
  - super-scalable

- development of new mathematical, statistical and computing tools for analysis and modeling of non-linear phenomena in complex networks

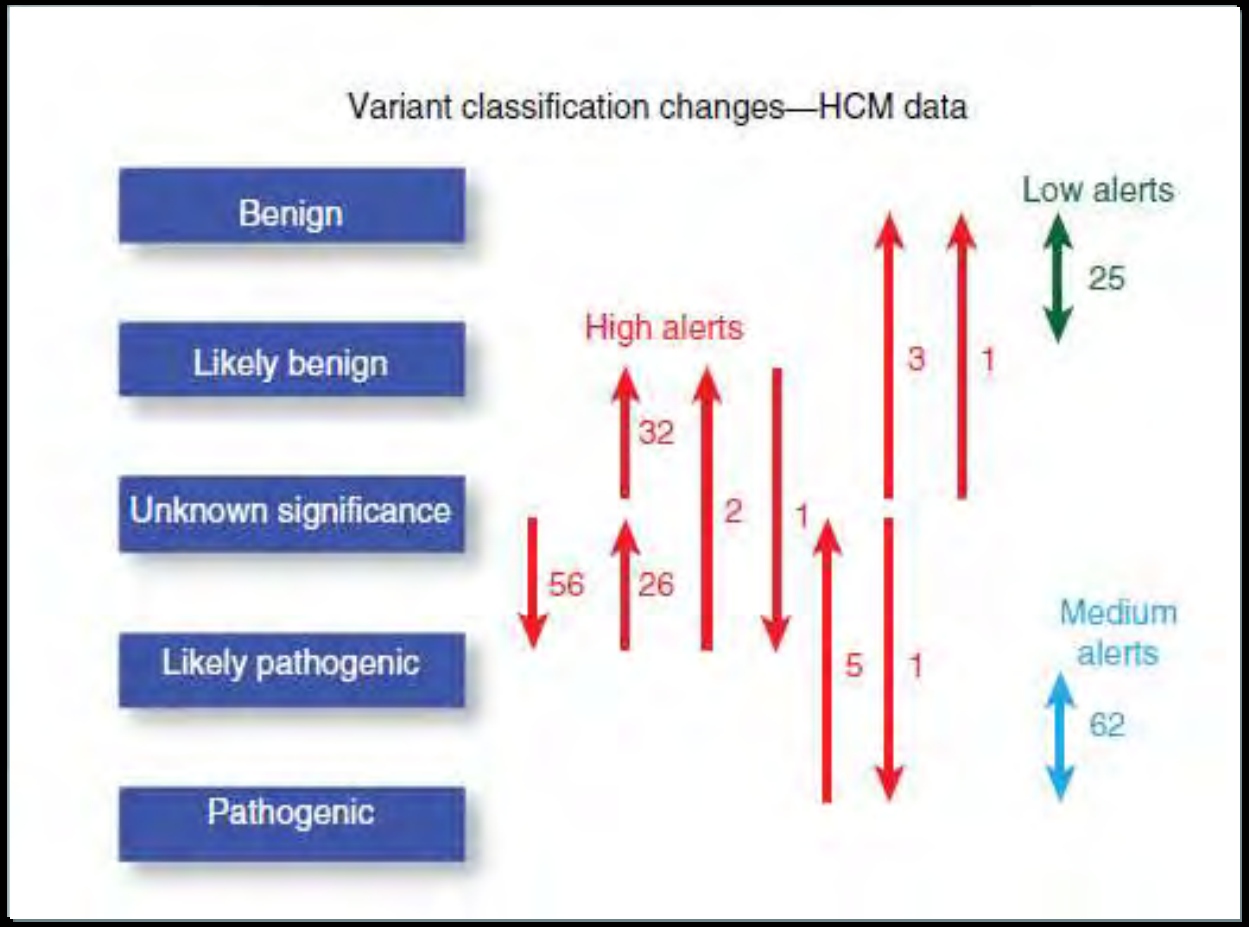
- application of advanced machine learning tools, avatars, robots and automated data production suites for customized data to promote optimum decision-actions

Molecular Medicine:
Genotype-Phenotype Correlations for Improved Clinical Care

Managing "The Incidentalome"

- identification of incidental disease risk factors during clinical omics profiling for a different purpose
- evidentiary standards and decision thresholds for follow-up/recontact research participants
- duties/obligations to recontact/reprofile based on new knowledge?
- consented vs. non-consented follow-up
- obligations to inform extended biological pedigree of serious risk(s)
- declining guarantees for anonymity, privacy and confidentiality

214 Changes Over Seven Years in Risk Classification for Hypertrophic Cardiomyopathy (HCM) Risk Variants on 11 Genes on HCM CardioChip Test*

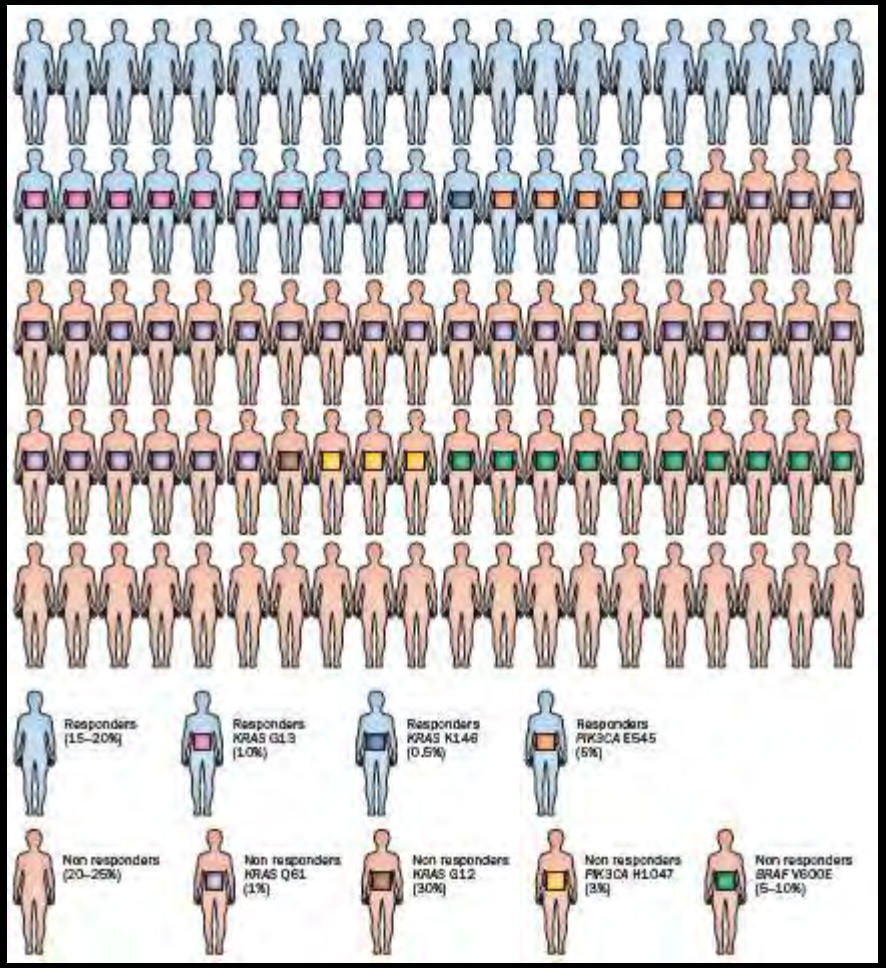


*S. J. Aronson et al. (2012) Genetics in Medicine 14, 713
Partners Health Care HCM Knowledge base: 1472 variants, 2279 family members, 4923 tests

Individual Genetic Variation, Disease Subtypes and Prospect of New Categories of 'Orphan Diseases'

Common Diseases: Are There Any?

Frequencies of Molecular Alterations in CRC and Responsiveness to Cetuximab or Panitumumab



From: M. Martini et al. (2012) Nature Rev. Clin. Oncol.

Molecular Diagnostics and Identification of Responder/Non-Responder Patients for Rational Rx

"The problem with all these (MDx subtyping) tests,
soon I'll have nothing
I can offer my patients"

"Eminent Oncologist" (journal designation)
Drug Discovery World. Spring 2011, p. 61.

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

Pharmacogenetic Profiling (PGx) for Rx Response/Safety

- higher Rx costs for omics- segmented markets to ensure adequate ROI?
- premium pricing for higher efficacy (guaranteed efficacy)?
- inadequate/erratic adoption of PGx testing to date
  - professional and payer knowledge gaps
- predictive value of some PGx tests may be insufficient for widespread clinical utility
- physician obligations to offer PGx test and obligation to use results?
- future liabilities: an evolving legal landscape (see G. Marchant, ASU)
  - physicians, pharmacists, companies, payers

Reimbursement for Molecular Profiling Tests

- uncertainty regarding data requirements
- CMS still relies on the same (and undefined) "reasonable and necessary" coverage threshold
- increasing payor adoption of non-payment "refuge" on the need for "practice-based evidence"
  - observational studies, pragmatic trials
- asynchrony between FDA approval and CMS coverage (CED solution?)
- preliminary decision by CMS (9/4/12) to place MDx in clinical laboratory fee schedule versus physicians pathology services classification (20% co-pay problem)

CPT (Current Procedural Terminology) Manual of the AMA

- CPT 5-digit classifier codes: established 1970, updated annually
- 2010 CPT handbook only contained 20 codes for genetic tests and process-focused (PCR, FISH….)
- new 2012 handbook
  - 100 gene-specific codes
  - 9 Tier 2 ‘buckets’ for other tests
  - difficulty for payors to price differentiate tests of different complexity binned in same bucket
  - currently no allowance for high throughput sequencing
- CPT editorial panel proposal for 2013 manual
  - Multi-Analyte Assays with Algorithms (MAAAs)

The Growing Education and Knowledge Gaps in Comprehension of Molecular Medicine Concepts Among Healthcare Professionals

Overcoming Gaps in Physician Knowledge of Molecular Medicine and a Paper-Centric Healthcare System

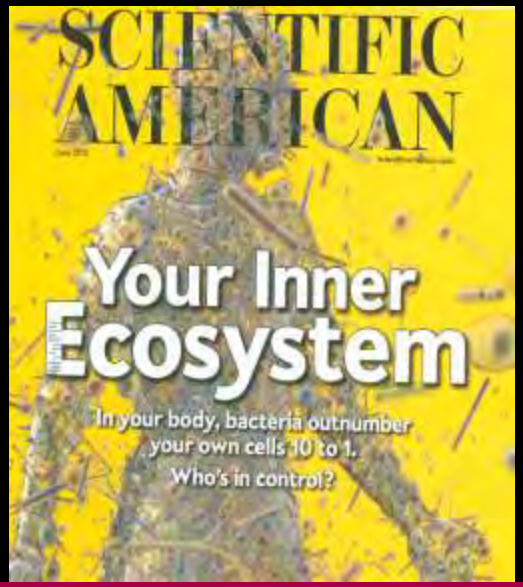
- 90% of Americans lack confidence in their clinicians ability to understand and use genetic information
  - http://www.cogentresearch.com/news/Press%20Releases/CGAT_2010
- professional cultural vulnerability/reluctance to acknowledge
- refuge in outdated SOC/guidelines that fail to integrate much new molecular profiling data
- protracted deliberations by professional societies/boards
- less than 4% of 8967 ACGME programs relate to genetic expertise (JAMA 2011 306, 1015)
- MD curriculum/CME challenges
- generational gap in IT use/facileness and resistance to computerized decision-support tools

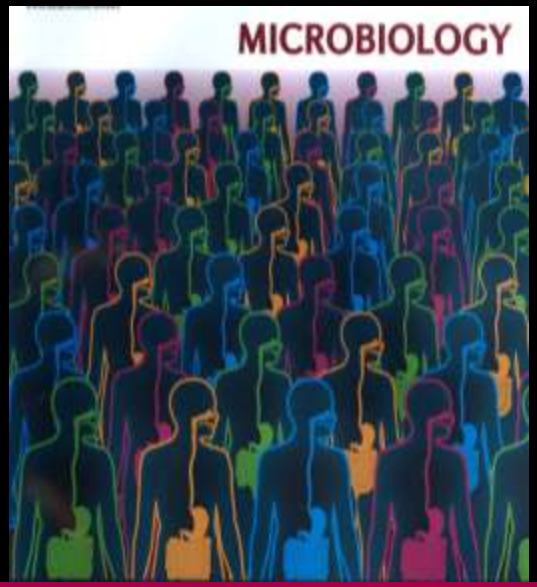
The Clinical Void in Understanding Laboratory Diagnostic Tests

"We don't teach (medical) students how to interpret lab results or how to pick them.
We're spending 61 to 302 hours in anatomic pathology and nine hours teaching laboratory medicine.
To pass anatomic pathology you've got to pass a test. There are no tests for lab. medicine."

Dr. M. Laposta MD. Ph.D.
Executive Vice-Chair of Pathology, Microbiology and Immunology
Vanderbilt Univ. School of Medicine
Member, CDC Clinical Laboratory Integration Into Healthcare Collaborative (CLIHC)
Clin. Lab. News. (2012) Sept. p. 2

We Are Not Alone: Variation in the Human Microbiome as a Potential Factor in Health and Disease

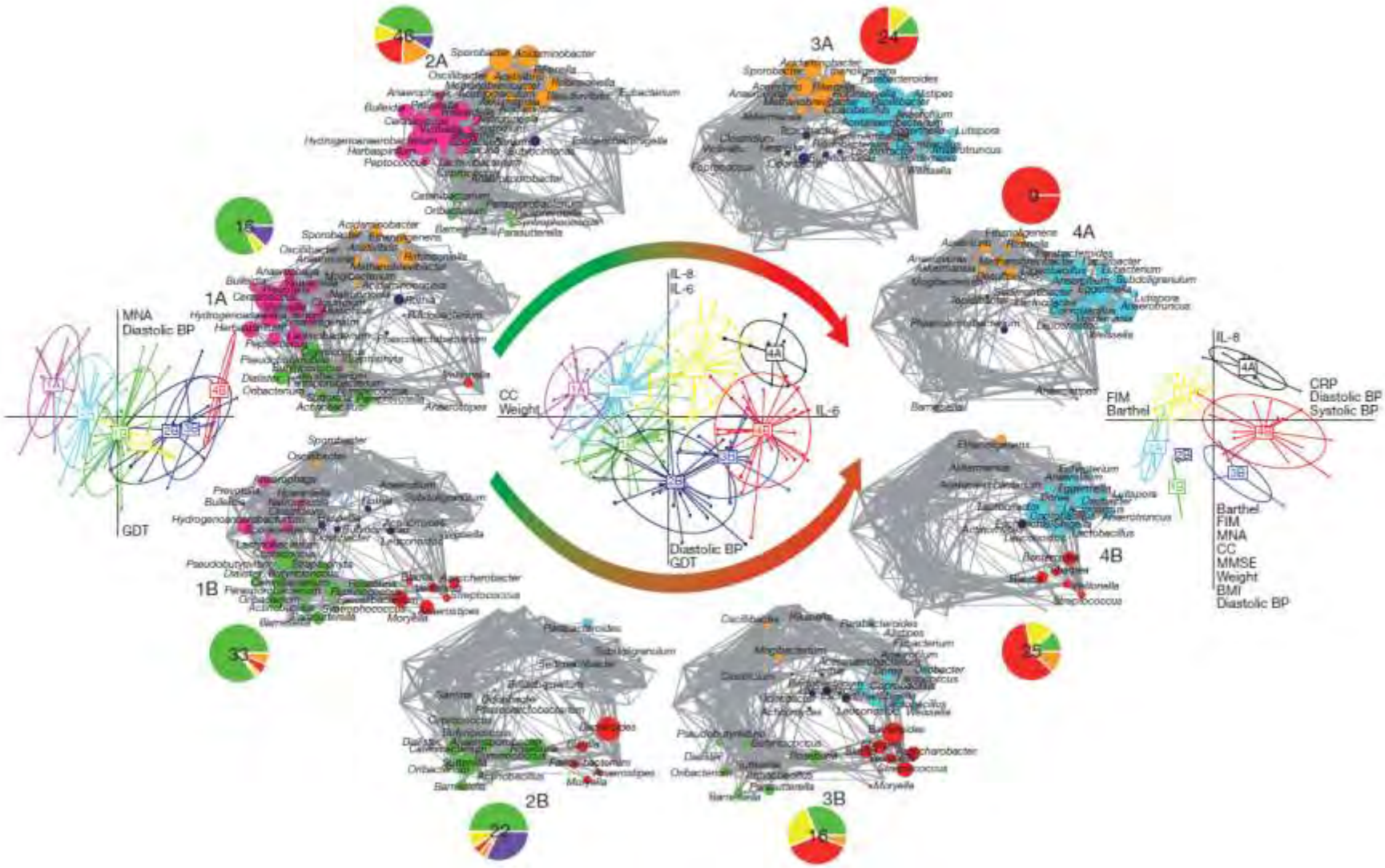




The Metagenomics and Complex Ecology of the Human Microbiome

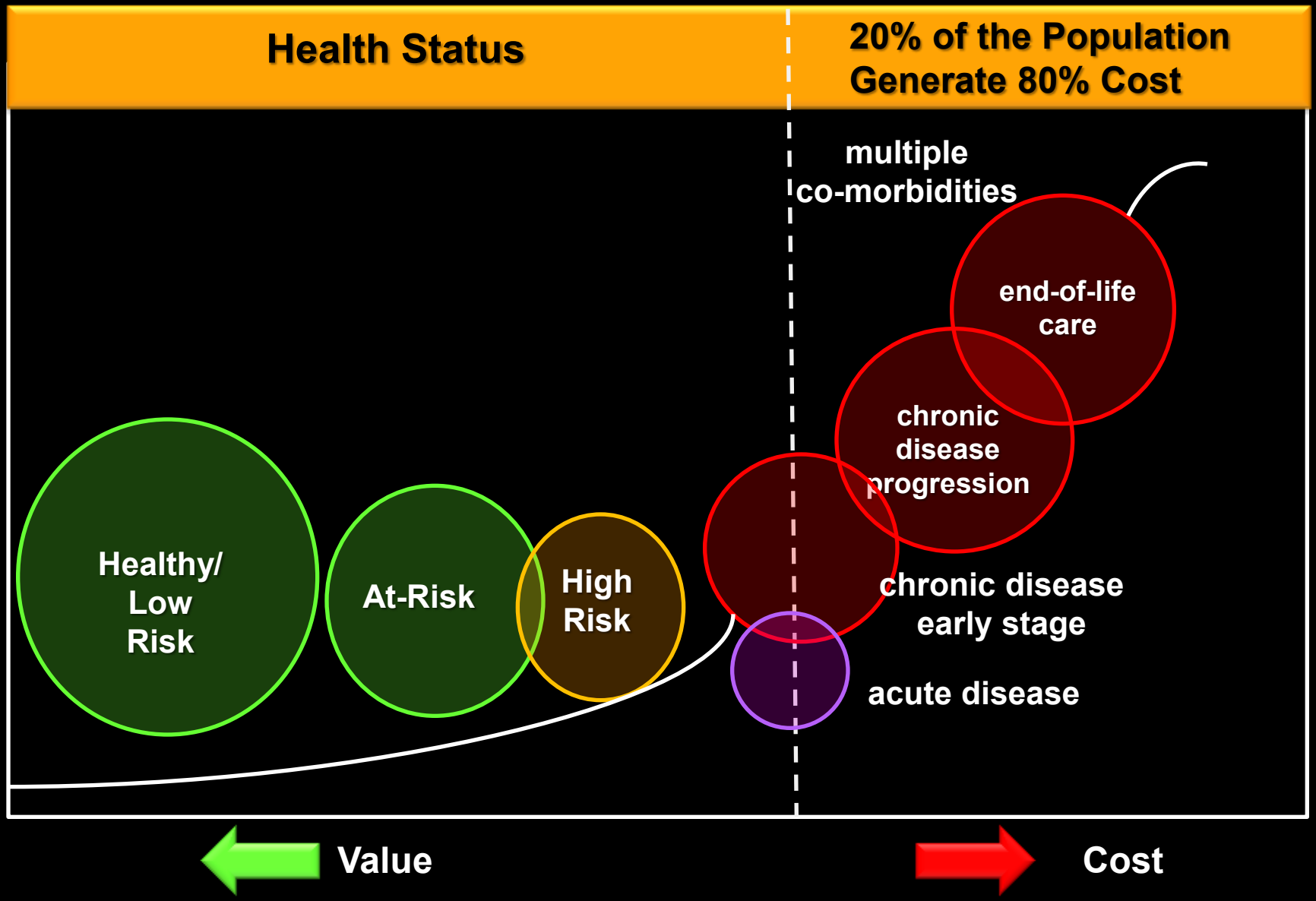
- extravagant microbial diversity
- complex microbe-microbe ecology and microbiota-host interactions
  - large scale sequencing and meta-analytics
- profiling the ‘healthy’ microbiome (age, gender (pregnancy) site-specific, geopgraphy, diet)
- dysbiosis and complex disease: cause or consequence?
  - characterization of role of emergent pathobionts (dysbiosis) in systemic disease
  - allergy/asthma, diabesity, metabolic disease, chromic inflammatory diseases?
  - aging and frailty?

Transition in Microbiota Composition from Community-(Left) to Long-Stay Facilities (Right) Mirrors Transition from Health (Green) to Fraility (Red)



From: M. J. Claesson et al. (2012) Nature 488, 182

The Economic, Social and Clinical Benefits of Proactive Mitigation of Disease Risk and Chronic Disease Co-Morbidities



Invasion of the Body Trackers

Individual Biosignature Profiling Via On Body: In Body (OBIB) Sensors and Devices

Remote Health Status Monitoring

M4: Making Medicine More Mobile

m.Health

Remote Health Monitoring and Chronic Disease Management

Lifestyle and Fitness

Information for Proactive Health Awareness (Wellness)

Increasing Engagement of Informed Consumers/Patients in Healthcare Decisions: Increased Personal Responsibility for Maintaining Health (Wellness)

Information Resources

- disease specific advocacy groups
- mass media
- web resources and social media
- mobile apps
- healthcare providers/ professionals

Optimizing Wellness and Risk Reduction

- "my profile"
- "my biorepository"
- "my health today"
- "quantified self"
- early alerts and risk mitigation
- virtual expertise network
- expertise locaters and clinical trial enrollment

Technology-Enabled Independent Living

“If I’d known I was going to live this long
I’d have taken better care of myself”

Eubie Blake, Musician on 100th Birthday 1983

The Wellness Premium

Greater Engagement and Incentivization of Consumers/Patients in Care Decisions and Sustaining Wellness

Social Spaces Become Quantifiable

- who knows why people do what they do?
- the fact is that they do!
- these actions can now be traced and measured with unprecedented precision
- with sufficient data, the numbers will reveal increasingly predictable rule sets for behavioral networks and individual risk
- major new business opportunities in multiple sectors including healthcare
- new ethical and legal issues

Proactive Engagement of Patient Communities in Investigational Clinical Trials and Observational Outcomes Studies

- Collate, Annotate and Curate Clinical Trial Data with Genomic Information from the Comparator Arms of Industry- and Foundation-Sponsored Clinical Trials
- Building a Site for Sharing Data and Models to evolve better Disease Maps.

Partnering for Cures | A FasterCures Meeting

CYCORE CYber-infrastructure for COmparative Effectiveness REsearch

CW PHS CENTER FOR WIRELESS & POPULATION HEALTH SYSTEMS

PURPOSE

To improve **cancer-related comparative effectiveness research** by better capturing data on physiological, behavioral and psychological status from research participants at home and as the go about their daily lives.

THE UNIVERSITY OF TEXAS MD Anderson Cancer Center Making Cancer History

Calit2

CYCORE

Interactive Participant-Centered Initiatives (PCI)

- new informed consent provisions
  - broad (future proof) versus narrow (explicit) investigation
  - flexibility to address personal preferences
- dynamic consent: e.consent tools and regular updating
  - EnCoRe, Indivo, PrivateAccess
- blurring of boundaries between research data and clinical records
- maintaining data privacy and public trust

Now Comes the Hardest Part of All!

Moving Downstream Beyond Discovery:
The Escalating Scale and Complexity of the Data Stream

Driving iOmics and Molecular Medicine and IT-Centric Capabilities Into Routine Clinical Practice

Silos, Turf and Cultural Conservatism as Barriers to Change

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

Current Era

- an opinion-rich, robust information content-poor world
- "silos" of research/clinical activities
- proliferation of poorly standardized and fragmented data, semantic anarchy and incompatible databases
- unacceptable levels of inaccurate diagnoses, fragmented care provision and flawed clinical decisions
  - highly variable treatment practices and erratic clinical outcomes
- extravagant waste and risk

Biomedical R&D and Clinical Medicine: An Unavoidable Yet Essential Transition to Date- and Computation-Intensive Processes

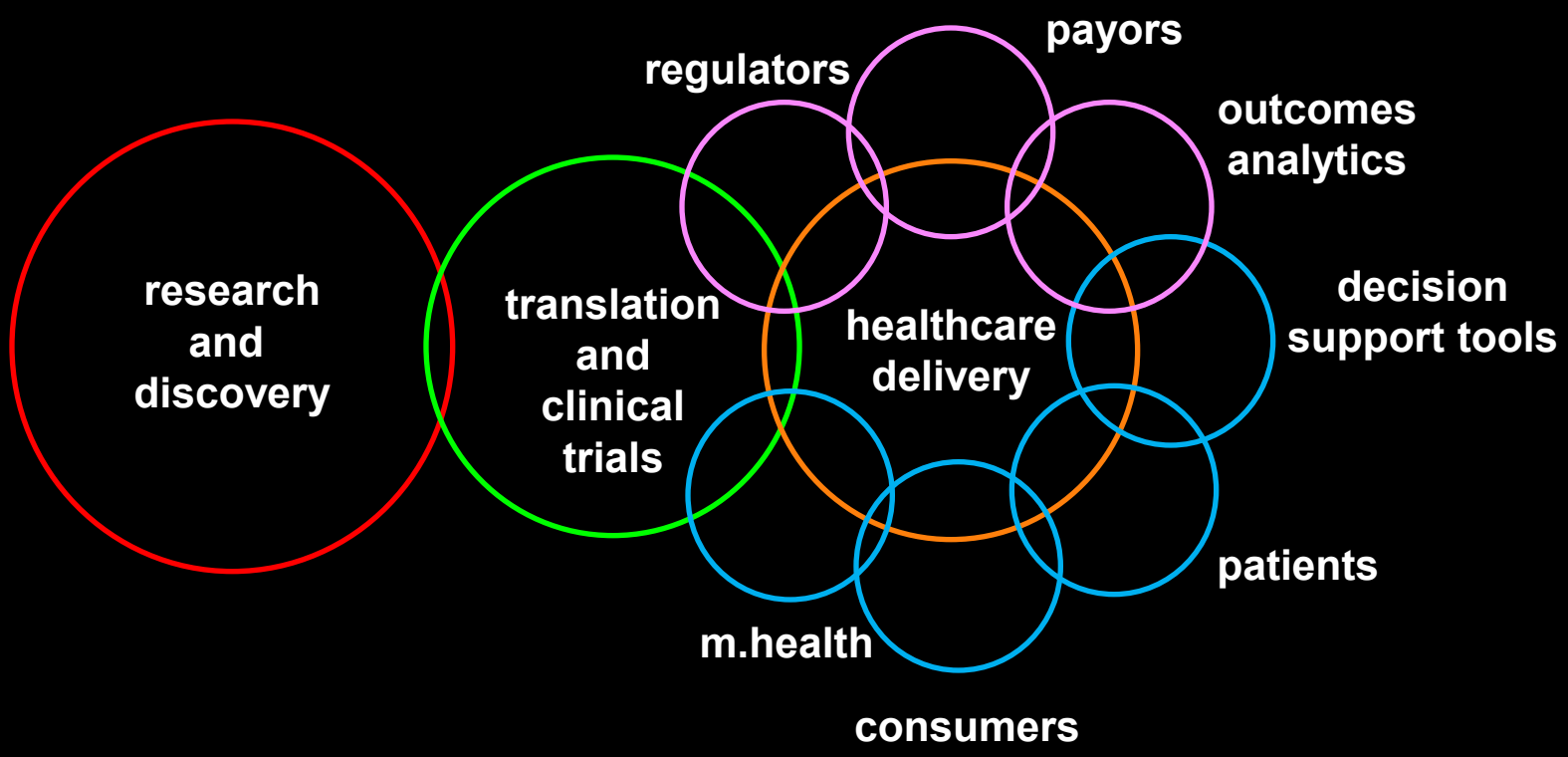
Pending Era

- massive data (big data)
  - V3: volume, velocity, variety
  - automated, massively parallel 'omics' profiling (research and clinical)
- diversification of data streams and cross-sector convergence
  - biomedicine, engineering, computing, telecommunications, social media
- new formats for data acquisition, validation and curation
- facile cross-disciplinary/cross sector dbase interoperabilities
- new machine-based analytics for management of mega-data, customized distribution and decision-support

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

Radical and Disruptive Changes Await!!!

The Need for Facile, Seamless Data Exchange Formats for Large Scale Biomedical Data Systems



Representation of Datasets and Abstractions

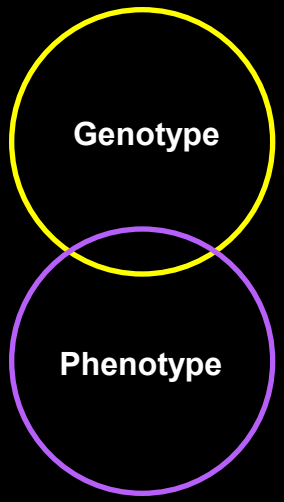
Discovery

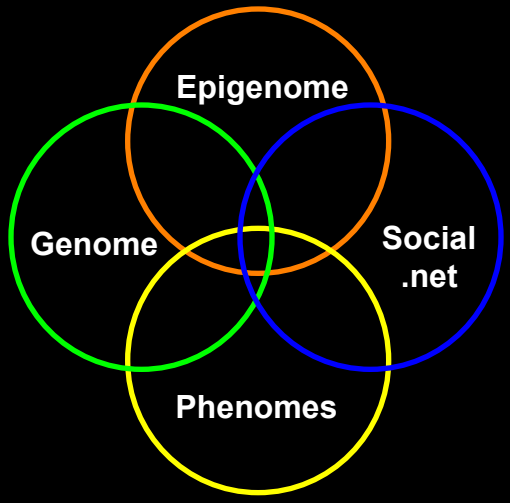
- controlled vocabularies and formal ontologies
- minimal information checklists and open source repositories
- algorithms and source code for analytical tools

Translation and Adoption in Routine Care

- exchange formats and semantic interoperability
- cross-domain harmonization/integration/migration/sharing
  - community-driven (eg. SMBL.org, BioSharing catalogue), industry-driven (eg. Pistoia Alliance), regulatory-driven (eg. CDISC), clinical (eg. HL7)
  - reimbursement (CPT, ICD) and HITECH EMR/MU
- consent, privacy, confidentiality, security
- meta-data tools
- machine-based natural language processing and decision support algorithms

Rich Data Will Drive Clinical Profiling to 'Interpreted Phenotypes'





Observed Phenotype	Interpreted Phenotype and Phenomes
• clinical annotation • EHR data mining	• clinical annotation • iPOPs + • large scale data analytics for "*robustness of match*" of observed clinical phenotype + iPOP profile + curated literature as a multi-dimensional matrix

What Is?
The Evolution of Computation Capabilities for Natural Language Q&A in Large Unstructured Datasets

Jeopardy 16 February 2011

- IBM's Watson
  - 2880 CPUs
  - natural language questions
- prelude to Q&A systems for biomedicine beyond keyword IR searches

WELLPOINT | Health.Care.Value.™

NUANCE®

The Pending Zettabyte Era

1,000,000,000,000,000,000,000

The Emergence of Large Scale, Integrated Data and Knowledge Networks:

Profound Consequences for Individuals, Enterprises, Infrastructure Investment and Governments

The Tianhe-BGI Bioinformatics & Computing Laboratory

- 14,336 Xeon X5670 Processors
- 7,168 Nvidia Tesla M2050 general purpose GPUs
- 2,048 FeiTeng 1000 SPARC-based processors
- 2.57 petaflops per second performance

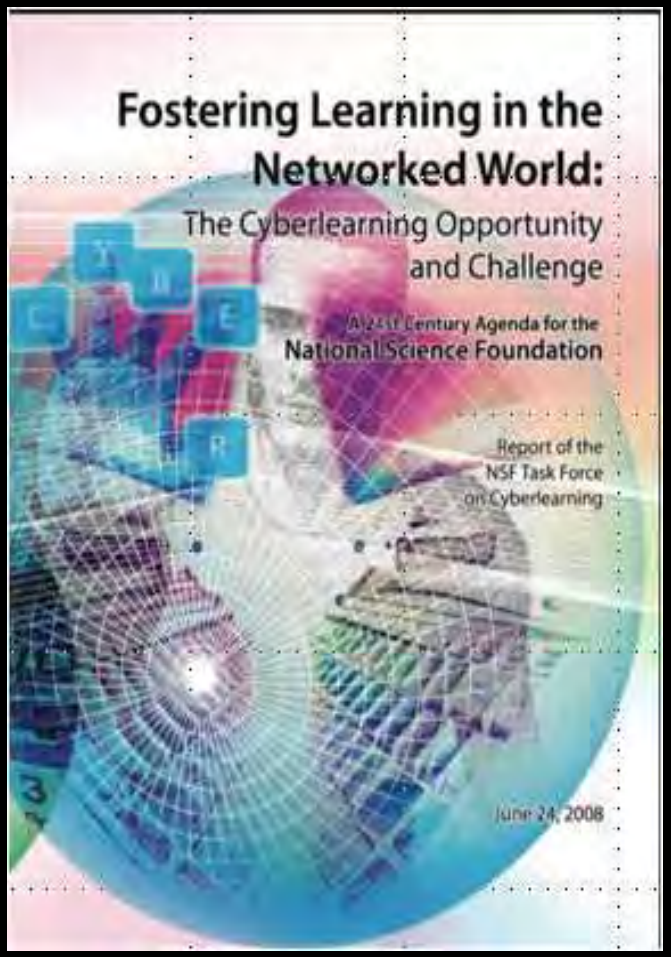
Key Principles in an Era of Data-Intensive Biomedical Computing

- increasing fraction of data is ‘born digital’
- more and more data is now networked and increasingly open
- ever larger data sets become increasingly unmovable with existing infrastructure
- modeling and simulations using big data and meta-analytics will amplify the data metaverse

Data-Driven Knowledge, Intelligence and Actionable Decisions

- changing the nature of discovery
  - hypothesis-driven versus unbiased analytics of large datasets (patterns, rules)
- changing the nature of explanation
  - statistical probabilities versus unitary values
- changing the cultural process of knowledge acquisition
  - large scale collaboration networks, open systems, social media
- changing the cognitive frameworks and intellectual competencies for knowledge-intensive competitiveness in multiple domains

Education and Training: The Looming Talent Gap



RESEARCH TRAINING
IN THE BIOMEDICAL, BEHAVIORAL,
AND CLINICAL RESEARCH SCIENCES

Committee to Study the National Needs for Biomedical, Behavioral, and Clinical Research Personnel

Board on Higher Education and Workforce Policy and Global Affairs

2011

- **by 2018 the US will need 160,000 more individuals with expertise in statistical methods and data analytics**

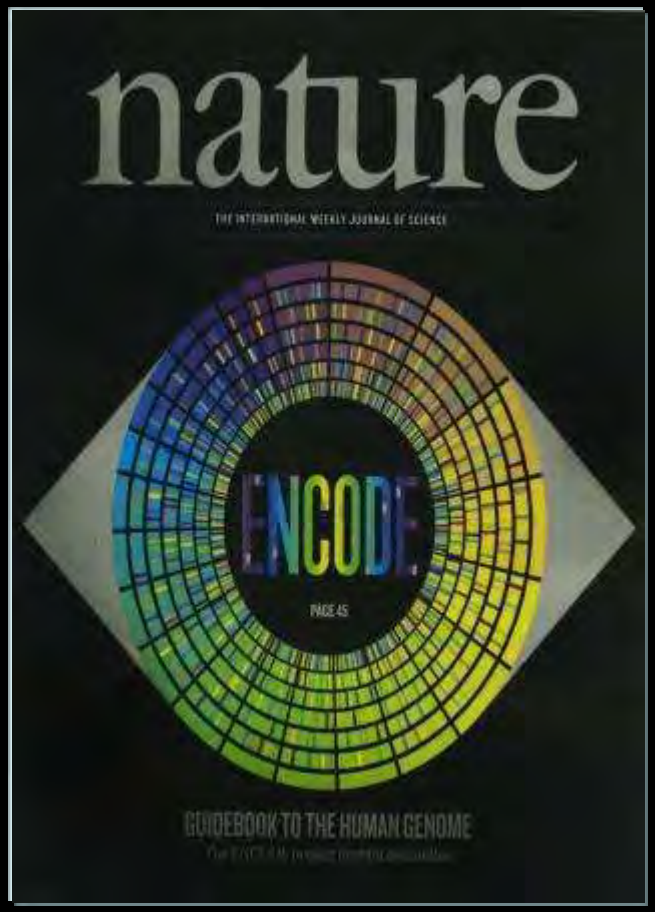
R.N. Rodriguez
President-Elect, American Statistical Association
Non-Clinical Biostatistics Conference, Boston 19 Oct. 2011

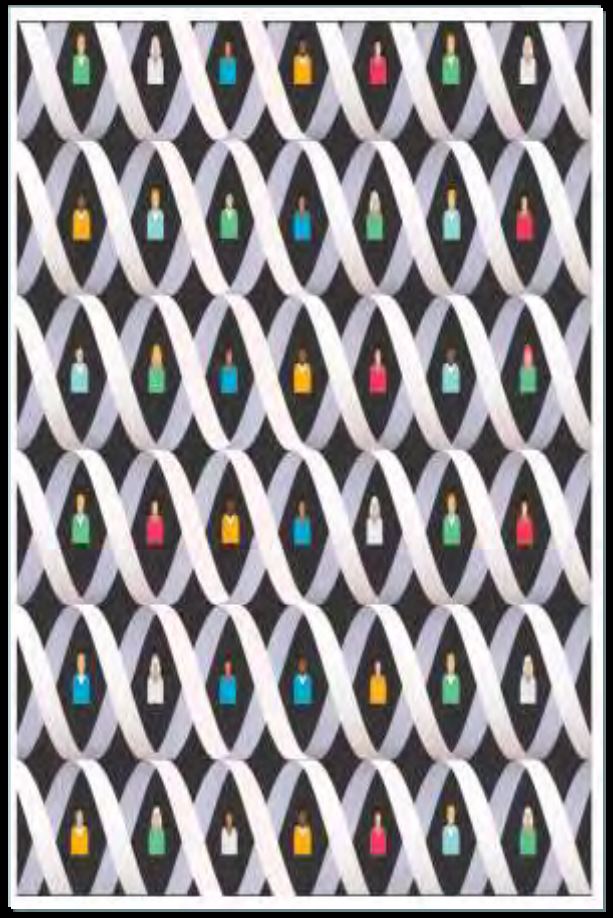
Consortium Science:Big Science

- biomedicine lags other fields of science and technology
  - engineering, materials science, computing, physics, astronomy, ecology, climate modeling
- big science antithetical to traditional organizational structures and career rewards in academic life sciences
- slow adaptation of public funding agencies to shift from individual-investigator to team-based science and enforce standards demanded by translational research
- '3M' projects: multi-investigator, multi-institution, multi-million
- increasing role of private public partnerships (3P) and pre-competitive consortia

Sustaining Relevance and Competitiveness: The Academy Must Engage With Real World Problems

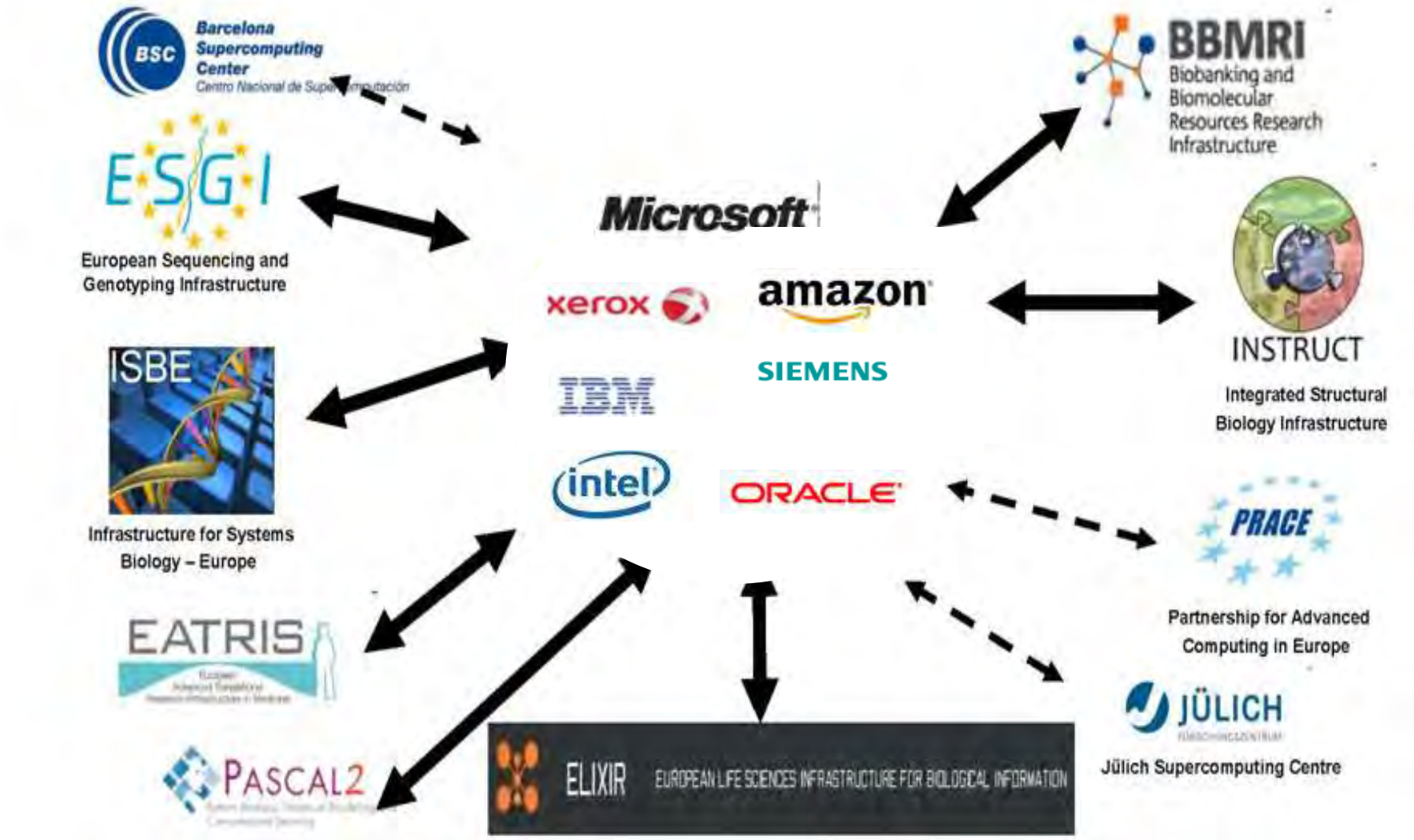
Big Data Projects in Omics: Logistical, Organizational and Cultural Challenges





- ENCODE
  - 10 year effort
  - $288 million
  - 442 Doctoral-level researchers in 2012 publications

IT Future of Medicine



BBMRI
Biobanking and Biomolecular Resources Research Infrastructure

EATRIS
European Infrastructure for Translational Medicine

ECRIN

elixir

BioMedBridges

will **move beyond** the scope of **the individual research infrastructures**...

...and **construct** the **technology** and **tools** needed to **connect** them

instruct
Integrating Biology

EMBRC
EUROPEAN MARINE BIOLOGICAL RESOURCE CENTRE

erinha
European Research Infrastructure on Highly Pathogenic Agents

EU-OPENSCREEN
ESFRI ROADMAP 2008

EURO-BIOIMAGING

The Sociology of Integration of Computational Science as a Core Component of Biomedical R&D

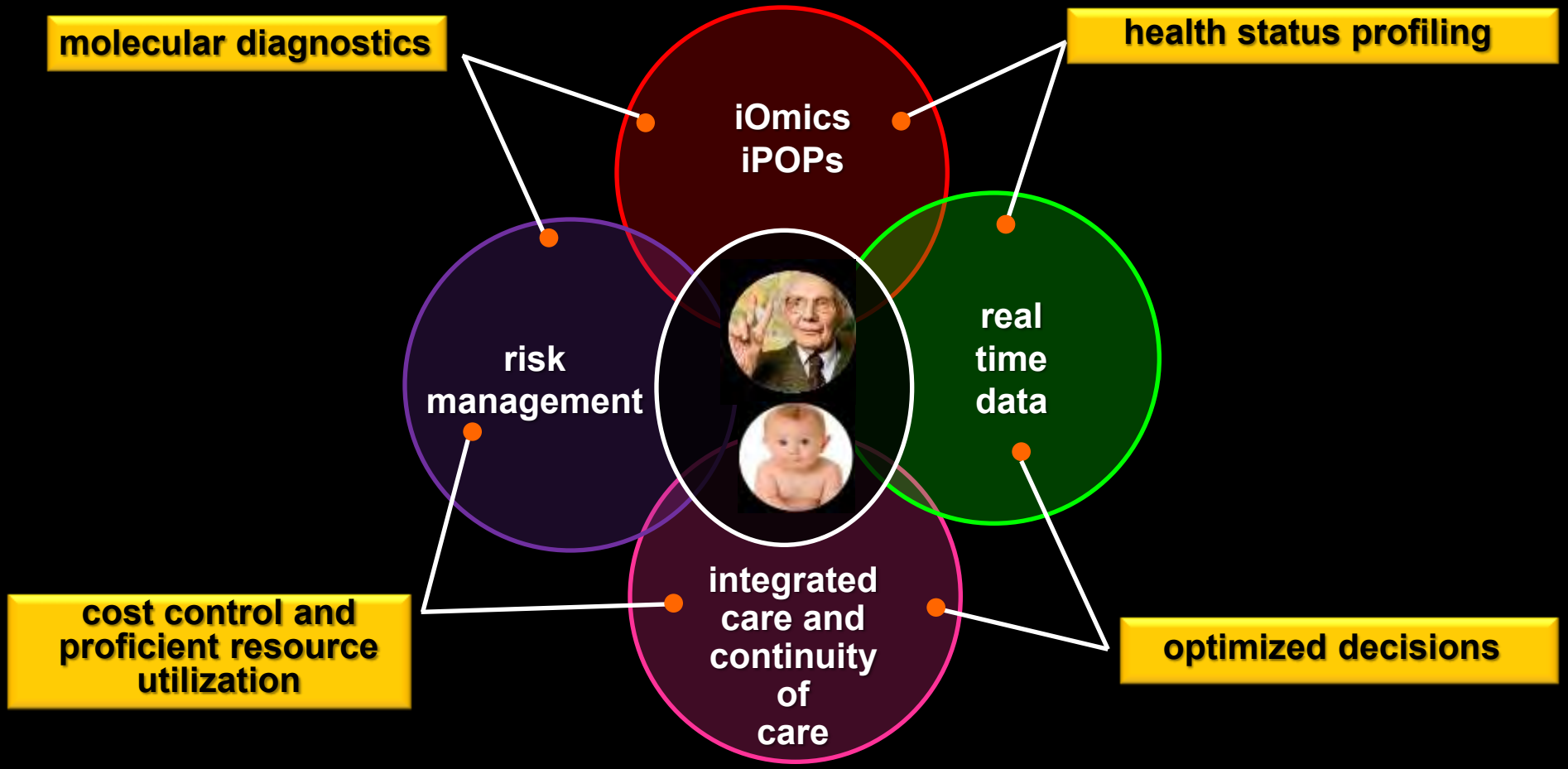
- bridging three cultures
  - biomedical specialities, software engineering and scientific computing
- new ‘hybrid’ competencies/specialities
- building sufficient expertise (individuals/communities)
  - training, funding, incentives, rewards
- designing workflows and interfaces for e.science and federated virtual research environments
- increasing dependence/contribution on open-source datasets
- mapping data provenance in large multi-source datasets
- life in ‘the perpetual beta’

- rethink
- recalibrate
- design

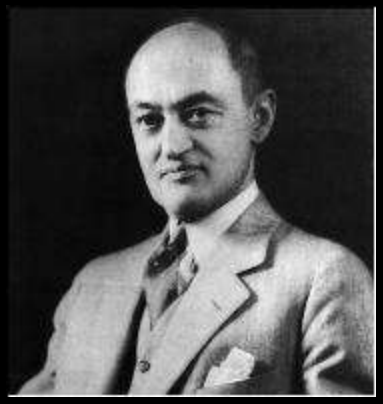
What is required?

What is sustainable?

Integrative Omics (iOmics) and Integrative Personal Omics Profiles (iPOPs) as the Core Foundational Elements for Molecular Medicine

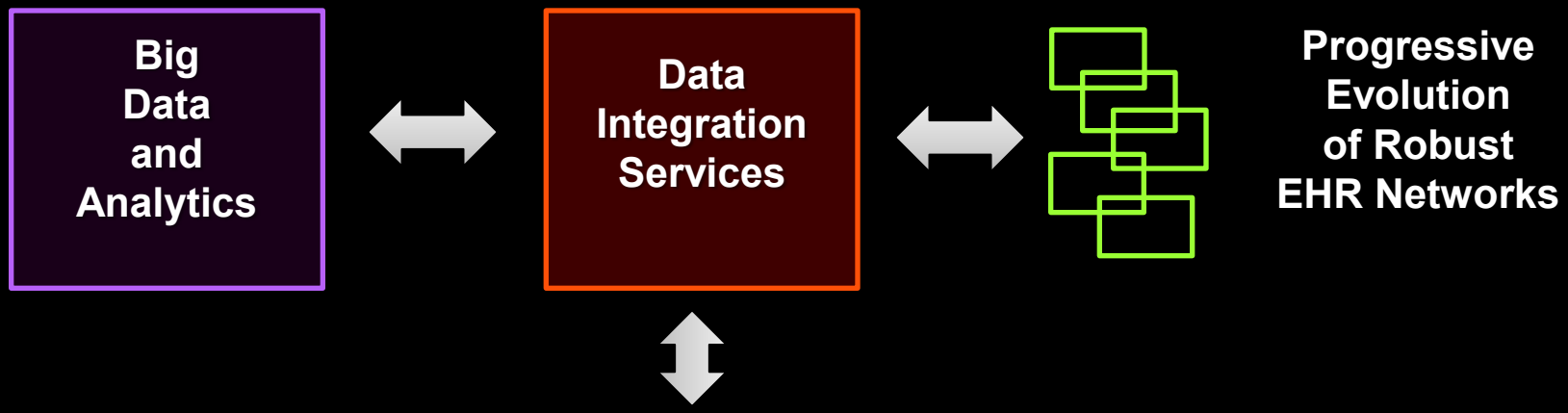


Disruptive Technologies and Creative Destruction




- arise at margins of existing fields
  or
  convergence/fusion interstices of previously separate technical domains/markets
- importance typically denied by KOLs and market leaders with often fatal consequences

iOmics and Computational Biomedicine as Disruptive Technologies for Radical Change in Life Sciences Research and Healthcare



Improved Clinical Risk Assessment, Management and Superior Outcomes

- robust, standardized research data and ontologies
- integrative personal omics profiles
- risk modeling
  - individual
  - population based
- precision diagnosis, monitoring and agile updating of best practice guidelines
- continuity of care
- pre-emptive care
- superior outcomes
- cost control
- improved care, QOL and wellness

Aligning Individual and Societal Needs for Optimum Health

Slides Available: http://casi.asu.edu/