ASU COMPLEX ADAPTIVE SYSTEMS
ARIZONA STATE UNIVERSITY

ASU SANDRA DAY O'CONNOR COLLEGE of LAW
ARIZONA STATE UNIVERSITY

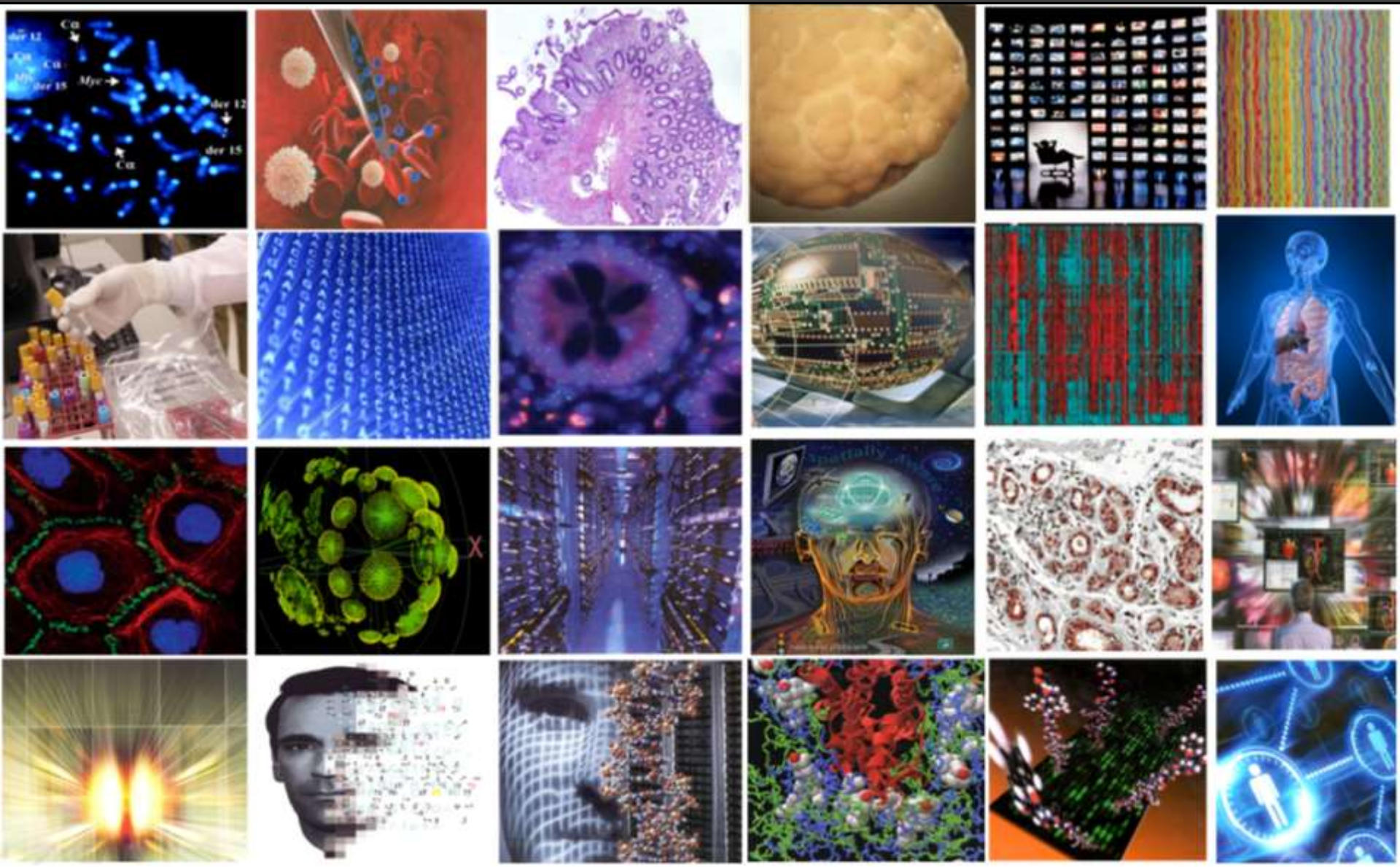
Genome Sequencing: A Status Report

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

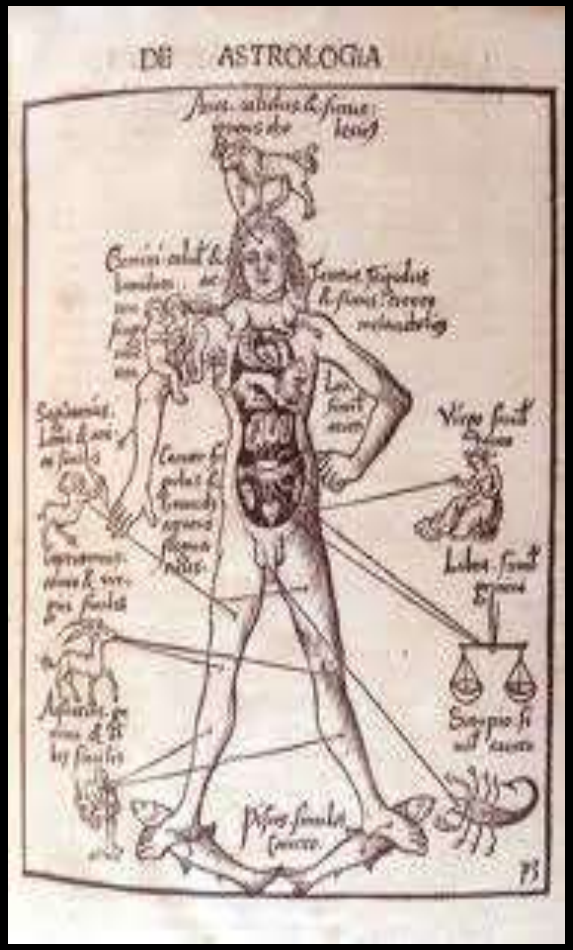
Presentation at the Workshop on
Whole Genome Sequencing: Regulatory and Reimbursement Issues

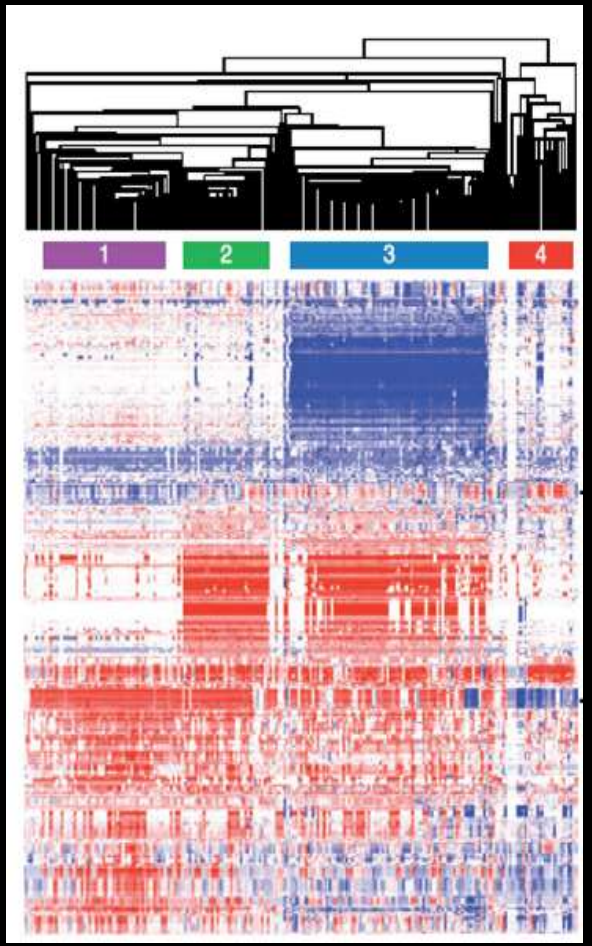
Arizona State University, Scottsdale
12 April 2013

Slides available @ http://casi.asu.edu/



Medical Progress:
From Superstitions to Symptoms to Signatures

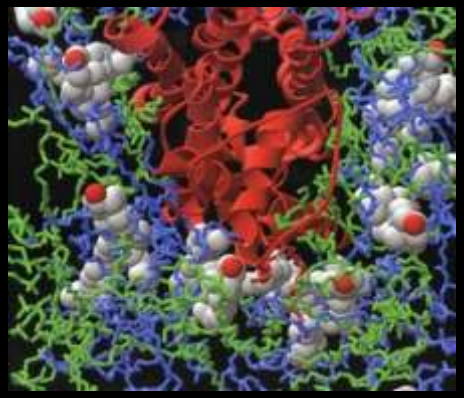




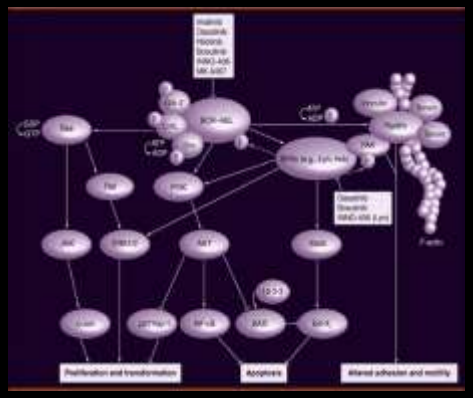
Mapping The Molecular Signatures of Disease: The Intellectual Foundation of Increased Diagnostic Precision and Rational Treatment Selection

Genomics

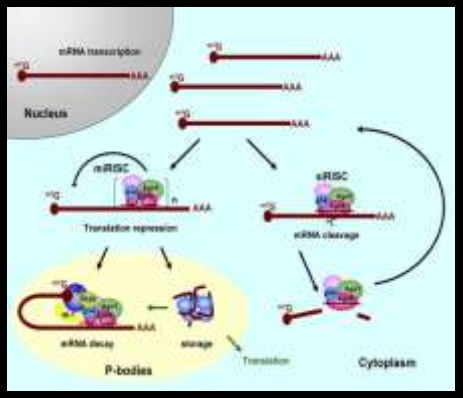
Proteomics

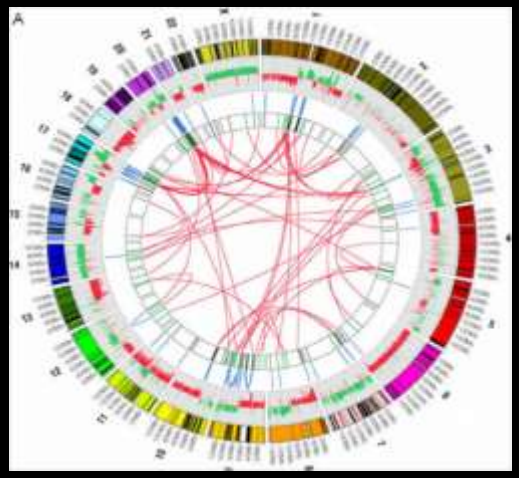


Molecular Pathways and Networks



Network Regulatory Mechanisms





ID of Causal Relationships Between Network Perturbations and Disease

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

Genome Sequencing

- targeted single gene or gene panels
- pharmacogenetics and polymorphisms in drug metabolism enzymes/HLAs
- whole exosome sequencing (WES)
- whole genome sequencing (WGS)
- transcriptome sequencing (over-or-under expression of genes and Rx selection)

Stakeholders in the Genome Sequencing Ecosystem: Fragmented Silos versus the Need for Integrated, Systems-Based Approaches

- research (discovery)
- translation (analytical and clinical validation)
- regulation (safety, efficacy, “fit for purpose”)
- reimbursement (utility, economics, value)
- clinical adoption (incentives)
- consumers/patients (personal access to WGS and engagement in care decisions)
- data ownership
- new ethical and legal issues

Industry Landscape is Crowded

Novel, Actionable Information: The Key Value Driver in Genome Sequencing

- immediately actionable
- known association/causation of disease but no Rx available
- informative (biological plausibility and likely causal but not actionable)
- unknown clinical significance

The Utility Matrix for Genome Sequencing (2013)

<table>
<tr><th>Decision</th><th>Single Genes</th><th>Gene Panels</th><th>WES</th><th>WGS</th></tr>
<tr><td rowspan="3">actionable variants</td><td>inherited and rare disorders</td><td>Rx targets: response/ resistance</td><td rowspan="3">late onset multi-genic diseases</td><td>microbes</td></tr>
<tr><td>carrier screening</td><td>PGx & Rx adverse events</td><td>rare diseases</td></tr>
<tr><td>NIPT</td><td>NIPT aneuploidy</td><td></td></tr>
<tr><td rowspan="2">informative but not actionable</td><td rowspan="2"></td><td>potential Rx targets but no Rx</td><td colspan="2">multi-genic neurodev. disorders</td></tr>
<tr><td></td><td colspan="2">GWAS</td></tr>
<tr><td>unknown significance</td><td></td><td></td><td></td><td></td></tr>
</table>

Will Low Cost Whole Genome Sequencing Change Everything?

- 1 million genomes x $1,000 = $1 billion

"It's not even a scary number anymore!"

Lander E. S. (2011) Nature 470, 187-197

WGS Sequencing Reported Costs
BioTechniques (2013) 54, 71

- Children's Mercy, Kansas City
  - $13,500 (STAT-Seq 50 hours)
- BGI
  - $12,000
- Partners Center for Personalized Medicine
  - $9000
- Washington University
  - $8000
- NHGRI
  - $7666
- Complete Genomics
  - $5000
- Illumina, Broad Institute
  - $4000

The $1000 (or less) Whole Genome Sequence (WGS)

The $? Interpreted WGS

The $? Reimbursed WGS for Clinical Use

Techno-optimism and the Seduction of New Technologies: The Omnipresent Risk of Hype and Herd Mentalities

The Tipping Point in the Clinical Adoption of WES/WGS

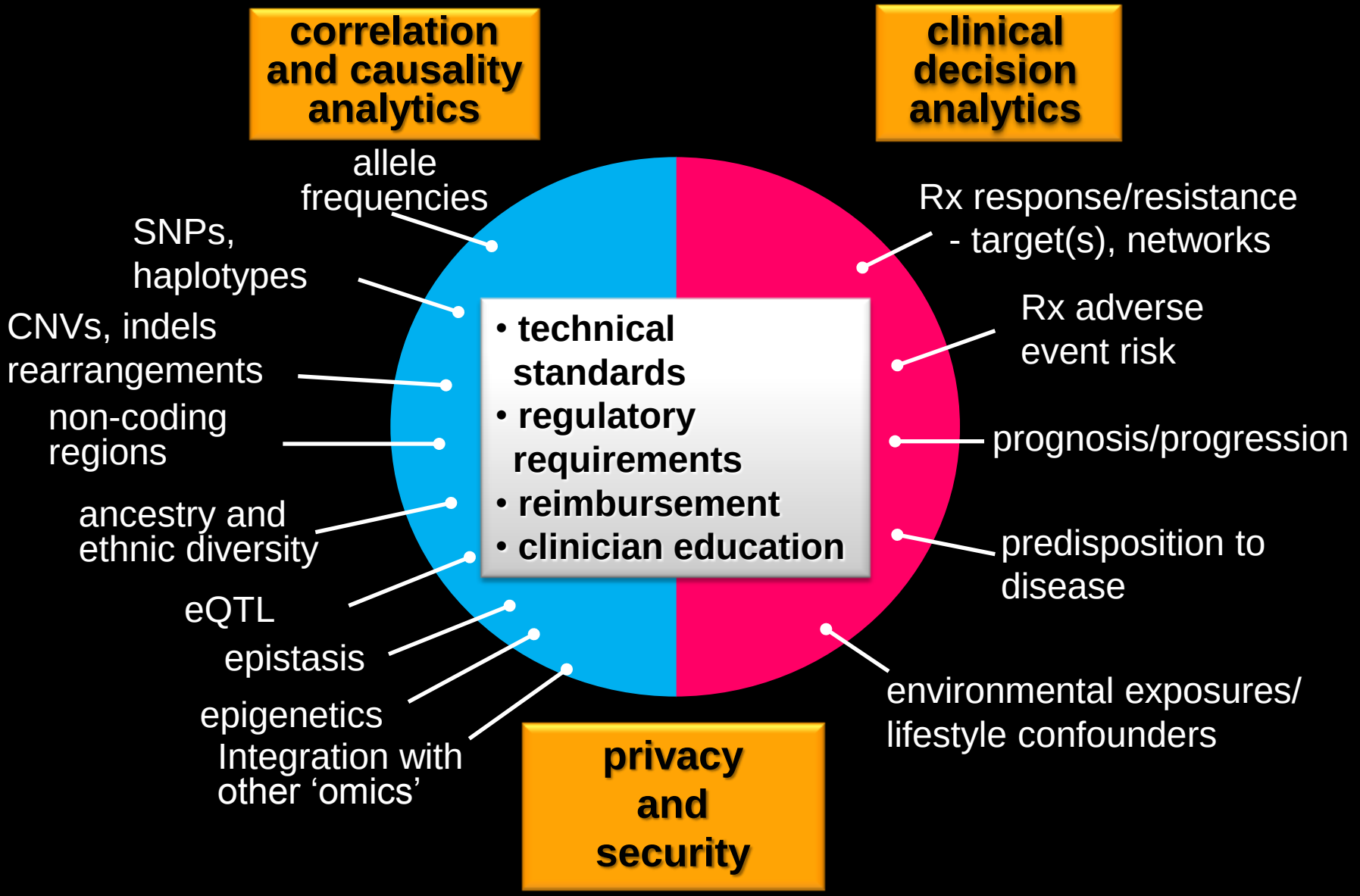
- threshold at which ready availability of *accurate* WES/WGS will be *cost-effective* and more *clinically useful* than targeted gene sequencing or *other molecular diagnostic tests*

From ‘Holes’ to ‘Wholes’ (Heidi Rehm)

- current technical limitations dictate that “hole” exomes/genomes is a more accurate description of WGS status today
- still major challenges in capturing ***complete*** and ***accurate*** WGS
- each individual has c.4 million variants

sequencing accuracy of 99.9%
=300,000 misreads per genome

Whole Genome Sequencing and Molecular Medicine



Whole Genome Sequencing and Analysis

- how often should individuals be resequenced?
  - lifespan epigenetic changes
  - disease dynamics (e.g. cancer)
  - intratumoral genomic heterogeneity in primary tumor and individual metastases

Lack of Reproducibility of SNV Profiling Using Different Sequencing Platforms*

- analysis of Khoisan individual (Namibia)
- three instrument platforms (Roche 454, SoliD3, HiSeq 2000)
- 3.3 million SNVs identified by all three
- significant false negative rates
  - between 71.5K and 443K variants (1.4-8.9%) detected by reads from different instrument platforms

*A. Ratan et al. (2013) PLoS One e55089 doi10.1371

Current Technical Challenges in the Production of a Complete and Accurate Human WGS Profile

- GC-rich exons
- copy number variants (CNVs)
- insertions and deletions (indels) and complex rearrangements
- tandem repeat expansions
- transposons, pseudogenes, satellite sequences, ribosomal sequences
- allele phasing and allele imbalance
- transcriptomic and metagenomic samples with nearly identical sequences

Regulatory Considerations for Genome Sequencing for Clinical Use

- LDT (CLIA) versus 510(k)/PMA review?
- sequencers as high complexity devices?
- obligate use of QSR-certified reagents?
- validation of analytical algorithms, software and source code(s) for sequence analysis and clinical records annotations?

Development of Quality Standard and Regulatory Frameworks for NGS (WES/WGS) in Clinical Laboratory Practice

- Next-Generation Sequencing: Standardization of Clinical Testing (Next-StoCT Workgroup) (CDC convener)
- College of American Pathologists (CAP) NGS checklist
- New York State NGS checklist
- FDA Sequencing Quality Control Project
- NIST/NIH Genome-in-a-Bottle NGS reference materials
- CDC Genetic Testing Reference Materials Coordination Program
- Association of Biomolecular Resource Facilities

*see Nature Biotechnol. (2012) 30, 1033

Sloppy Discovery Studies: Lack of Standardization and Inconsistent Methods

- 1000 Genomes Project stringent open-source protocols for variant calling in WGS
- survey of 299 papers from 2011 claiming protocol adoption
  - only 10 used the recommended tools
  - only 4 studies used full workflow for realignment and quality score recalibration

NGS Data Analysis: Transparency, Reproducibility and Accessibility

(Nature Rev. Genetics (2012) 13, 667)

- review of 50 papers published in 2011 using Barrows-Wheeler Aligner for mapping Illumina reads
- 43/50 did not provide even partial lists of the parameters used
- only 7 listed all necessary details
- 26/50 did not provide access to primary data
- 2/50 hyperlinked to own website but links inoperative

The Principal Challenge in Clinical Genomics

Inferring the Genetic Architecture of Disease and Disease Susceptibility With Incomplete Information

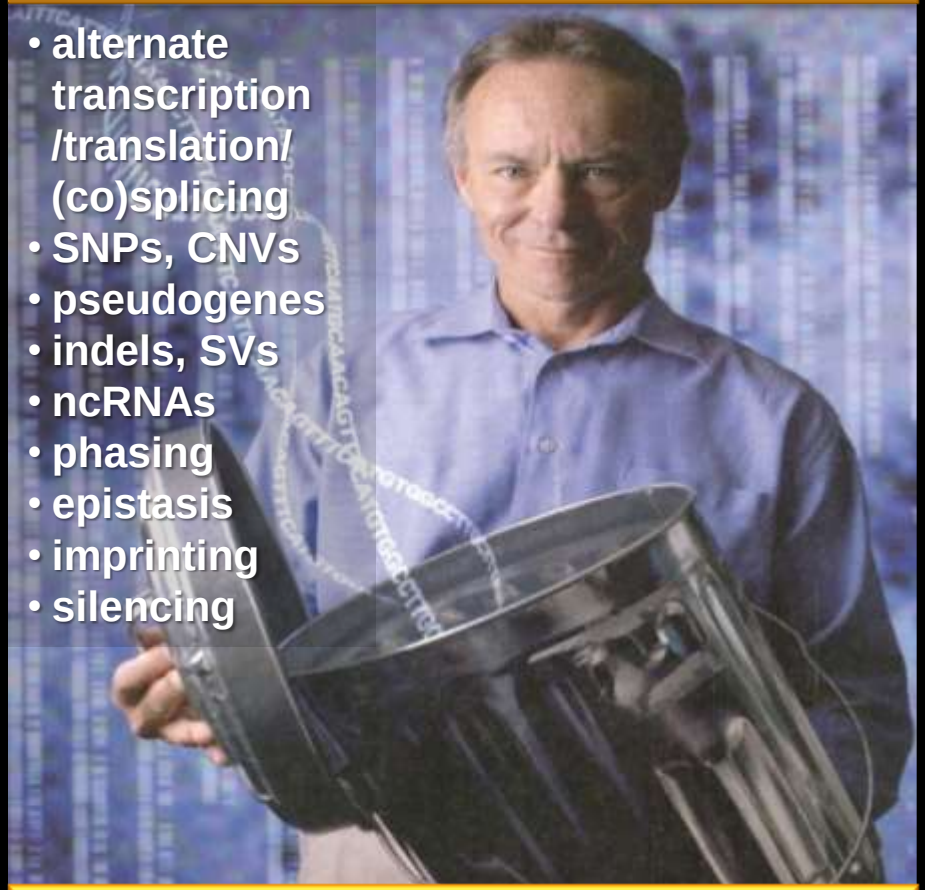
WGS and Claims Outstripping Current Analytical Capabilities: Disease Predisposition Risk Profiling (PDx)

Genes For …..

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

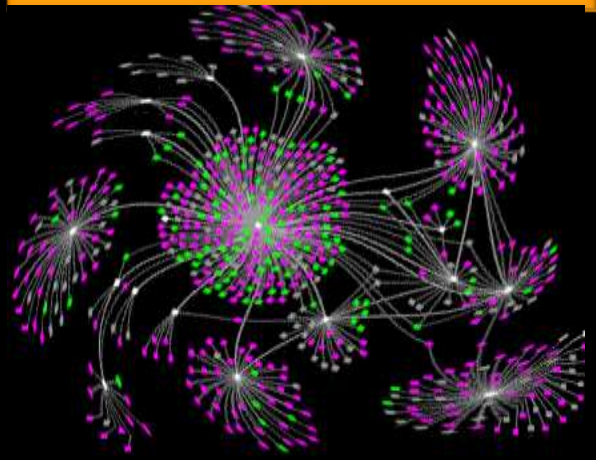
Individual Variation, Genome Complexity and the Challenge of Genotype-Phenotype Predictions

Junk No More: Pervasive Transcription

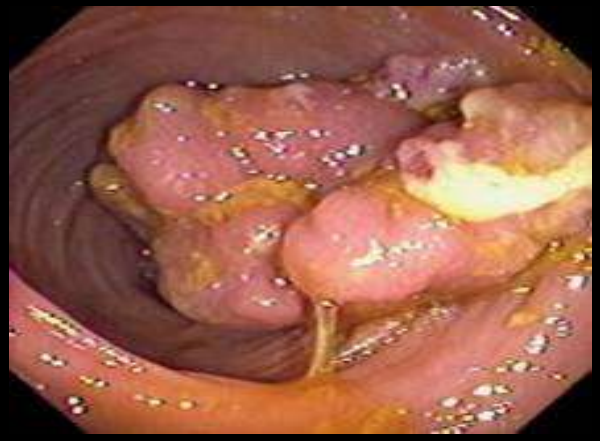


recognition of genome organizational and regulatory complexity

Cell-specific Molecular Interaction Networks



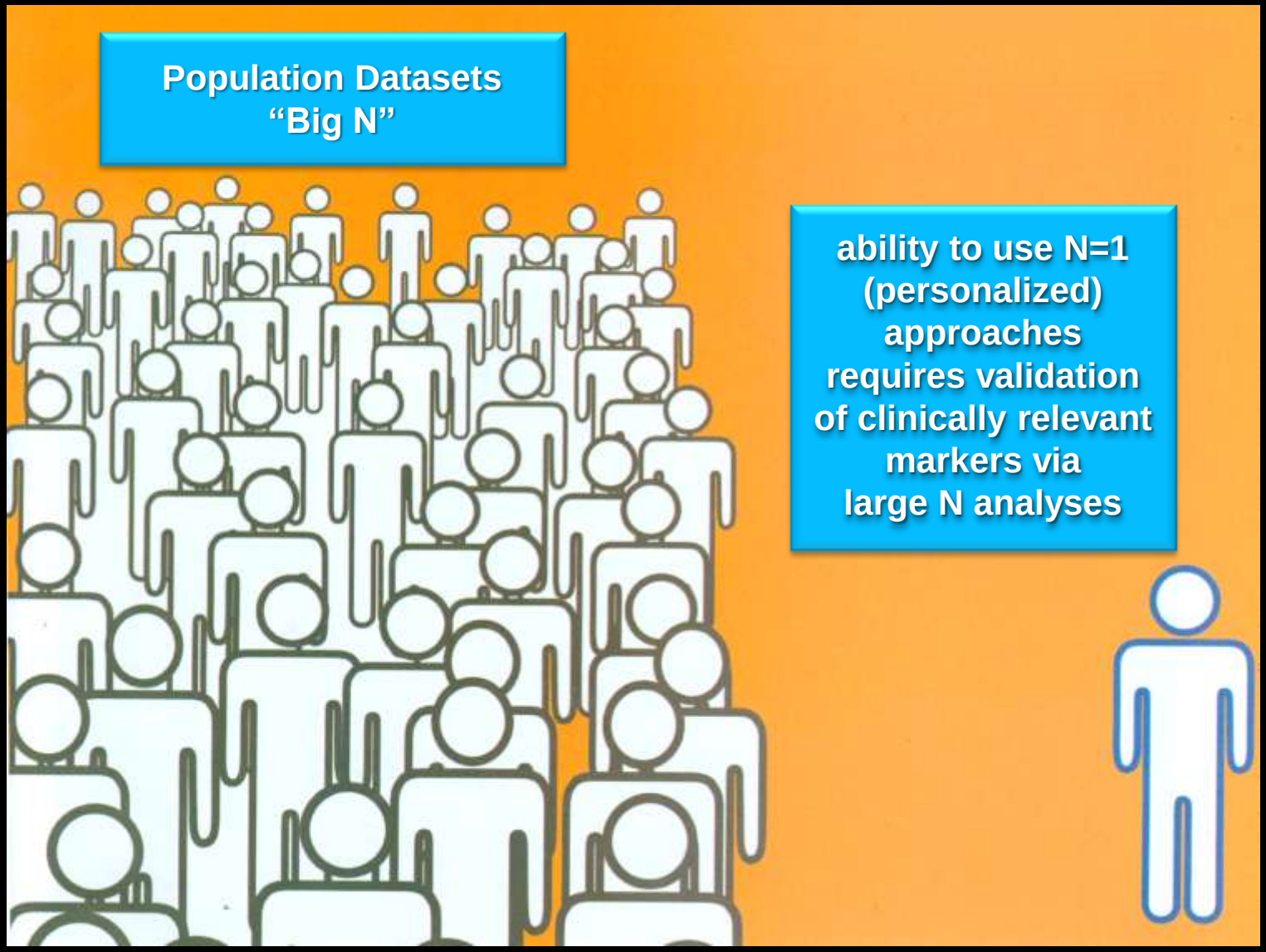
Perturbed Networks and Disease



Mapping Human Genome Variation and Identification of Causal Variants for Disease

- 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 ☑
- “known unknowns”
  - epigenetics and environmental confounders ☑

Precision Medicine: Evidentiary Standards for Dissecting the Correlation: Causality Matrix in "Omics" Profiling



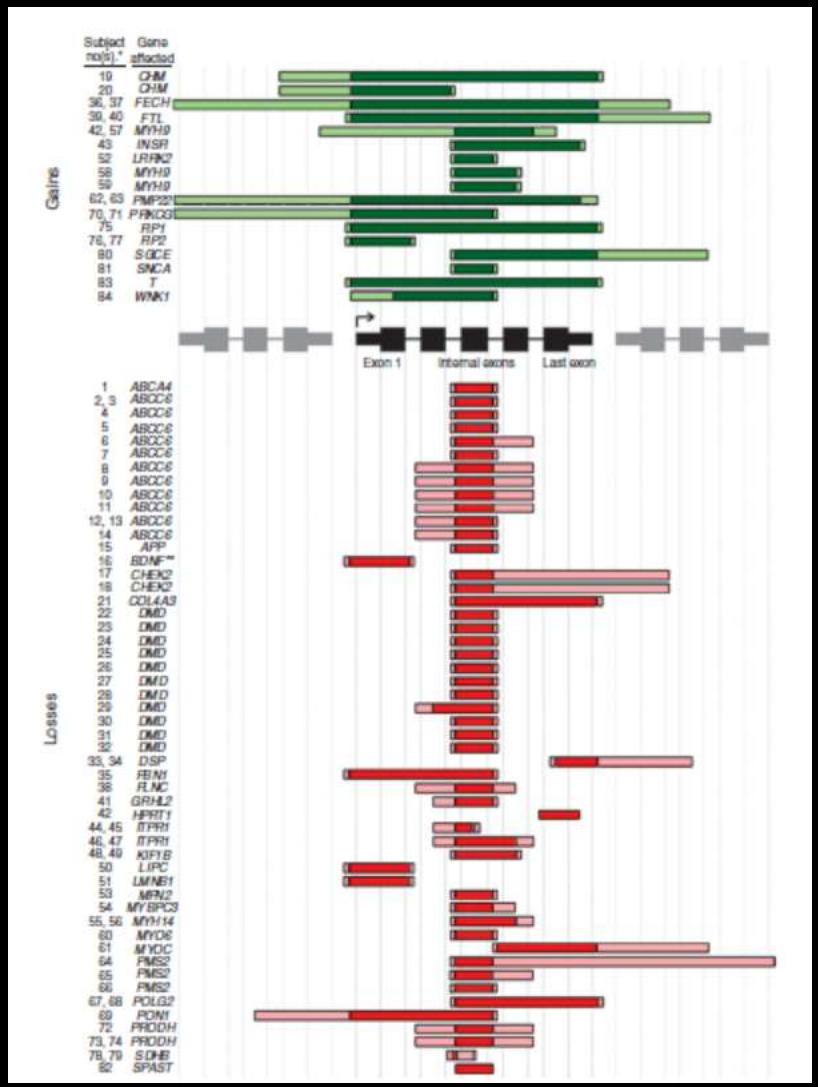
Classification of Genetic Variants Relative to Clinical Disease Risk(s)

- pathogenic (causal)
- likely pathogenic (plausibly causal)

- (currently) unknown significance

- likely benign (presumed non-causal but reported in disease)
- benign (neutral)

Incidental CNV Dominant-Adult Onset Disease Risk Variants Identified by Routine Genome Testing Unrelated to Initial Referring Diagnoses



- P.M. Boone et al. (2013) Genetics in Med. 15, 45

- ACGH profiling of CNV in 9,005 individuals
- SNV in 40 genes implicated in adult-onset disease
- 14 viewed as pre-disposing
- 25 likely benign
- 44 unknown significance (VUS)

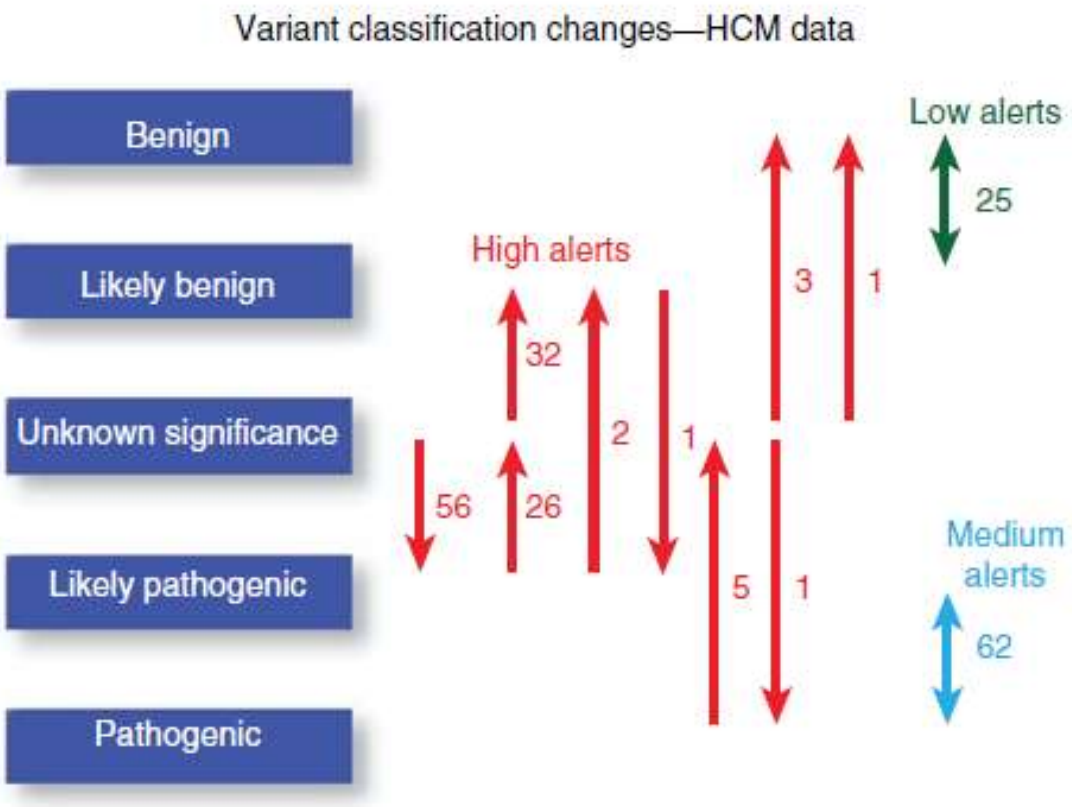
WES/WGS and Identification of 'Incidental' Potential Disease Risk(s)

- high likelihood of 'incidental' findings
- findings may also be relevant to family members
- pre-test counseling and opportunity to decline
- educational tools to assist patients in understanding test results

American College of Medical Genetics Guidelines on Incidental Findings

- approved March 2013 AGM
- 56 genes with certain variants for which individuals should be informed, irrespective of individual preference

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

Assessment of the Clinical Significance of Genome Sequence Variation

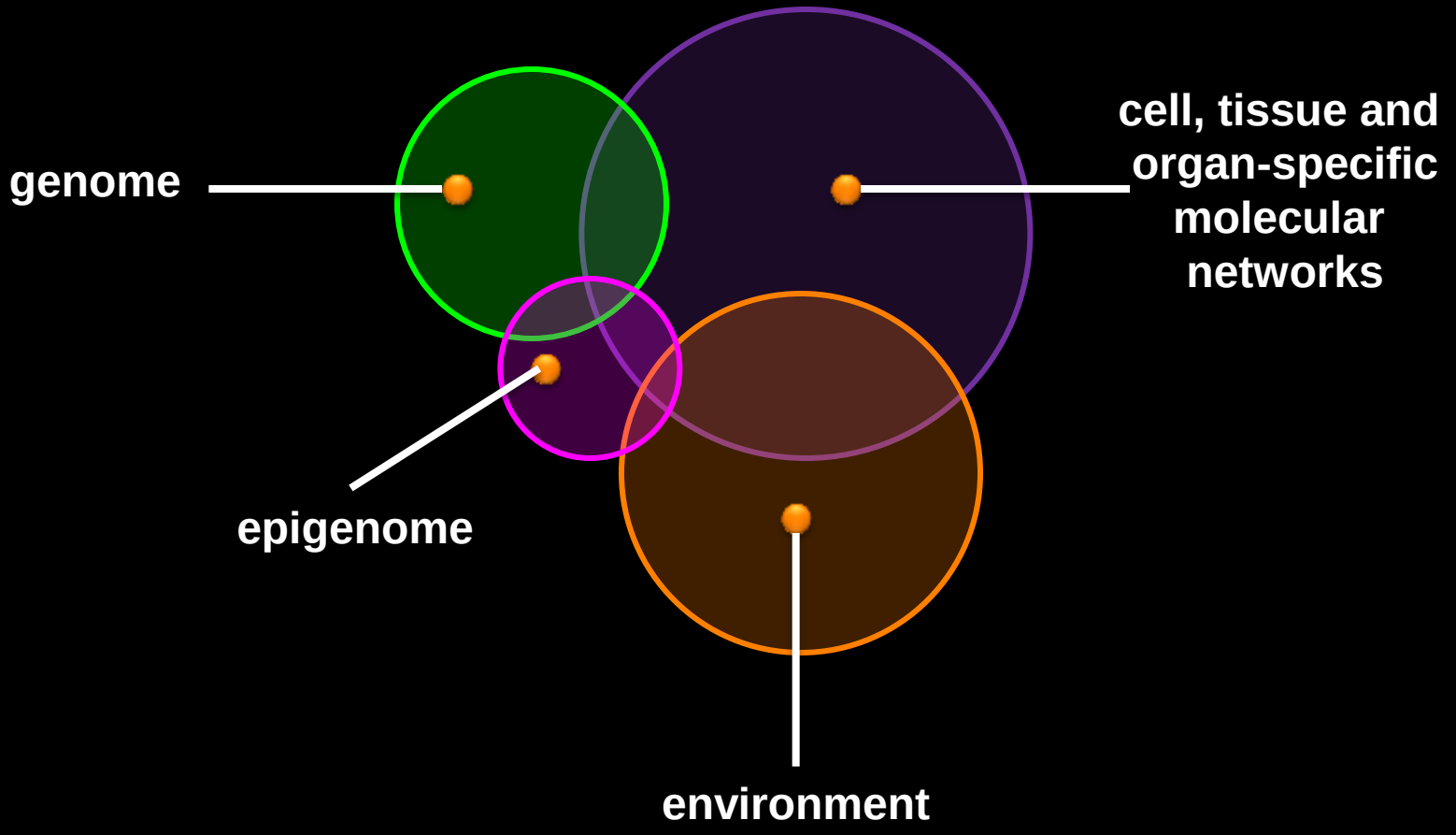
- availability of ever larger WGS databanks will allow greater precision in linking specific variants to disease risk, disease progression and Rx response
- evidentiary standards and who defines?
- logistics and cost of constant updating databanks of new actionable findings
- duties to inform individuals of new risk(s)?

Consumer Genomics

(DeCode)

(Navigenics)

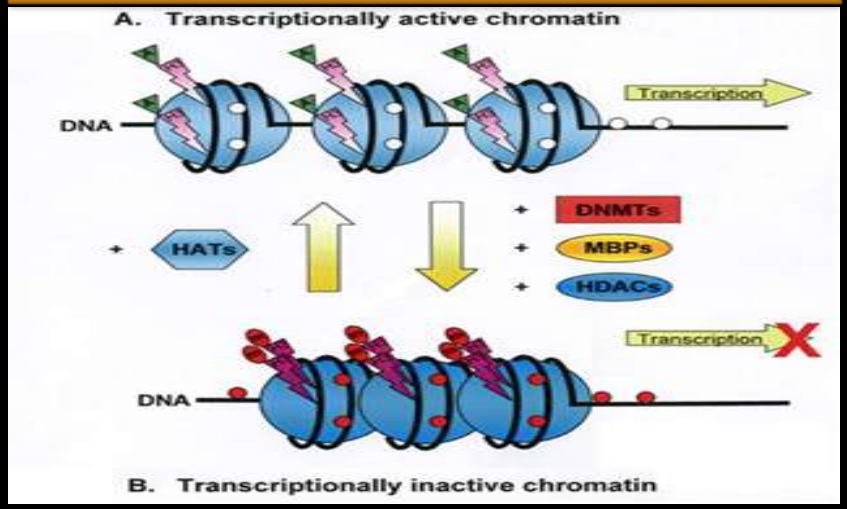
The Complex Interplay Between the Genome, Molecular Networks and Environmental Factors



The Epigenome

Effect of Maternal Diet/Stress/Rx exposure on Germ Line Genome (+ trans-three-generational?)

Modulation of Gene Expression/Regulation by Environmental Factors, Xenobiotics and Rx (The Exposome)



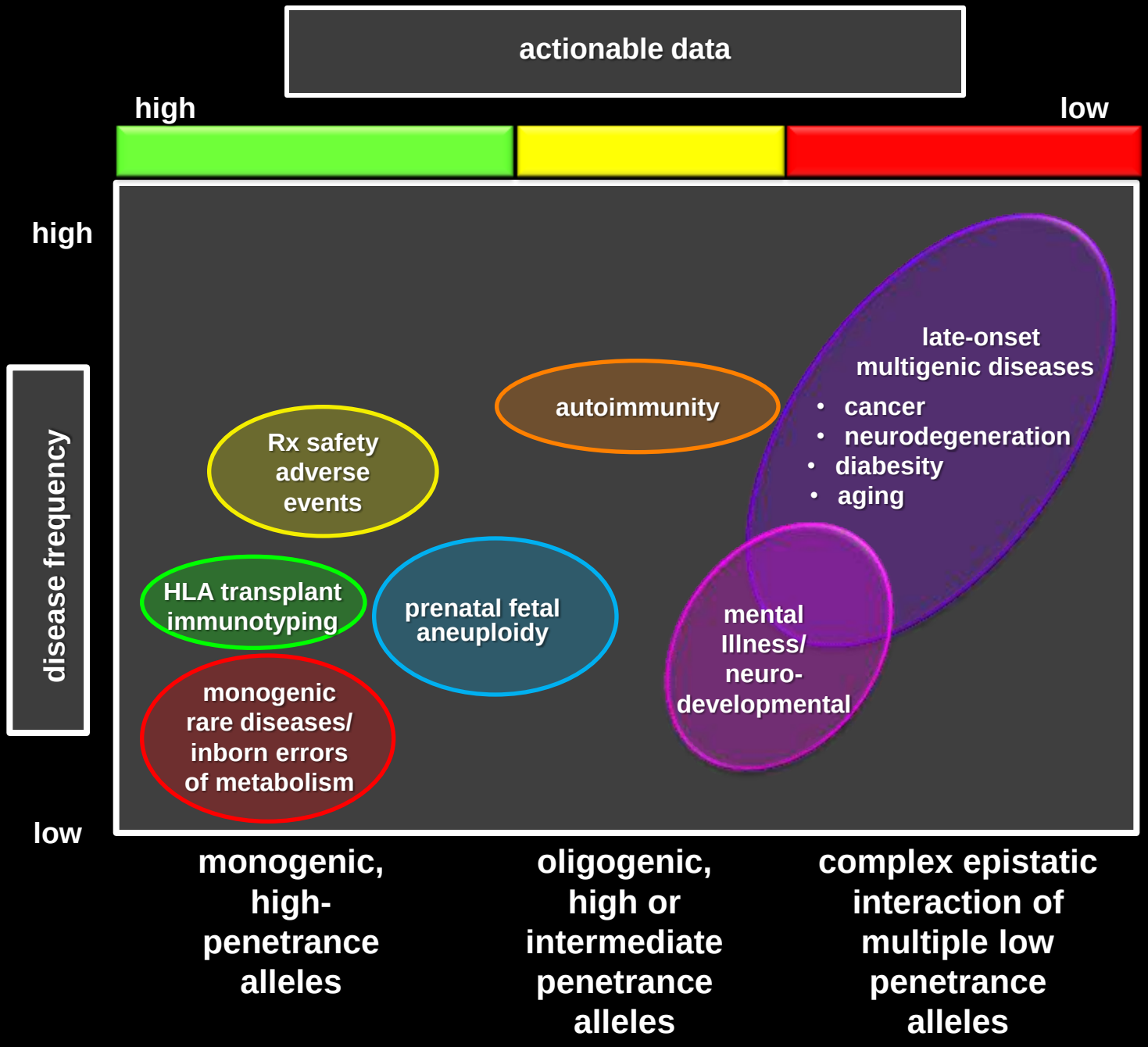
IHEC
International Human Epigenome Consortium

International Human Epigenome Consortium

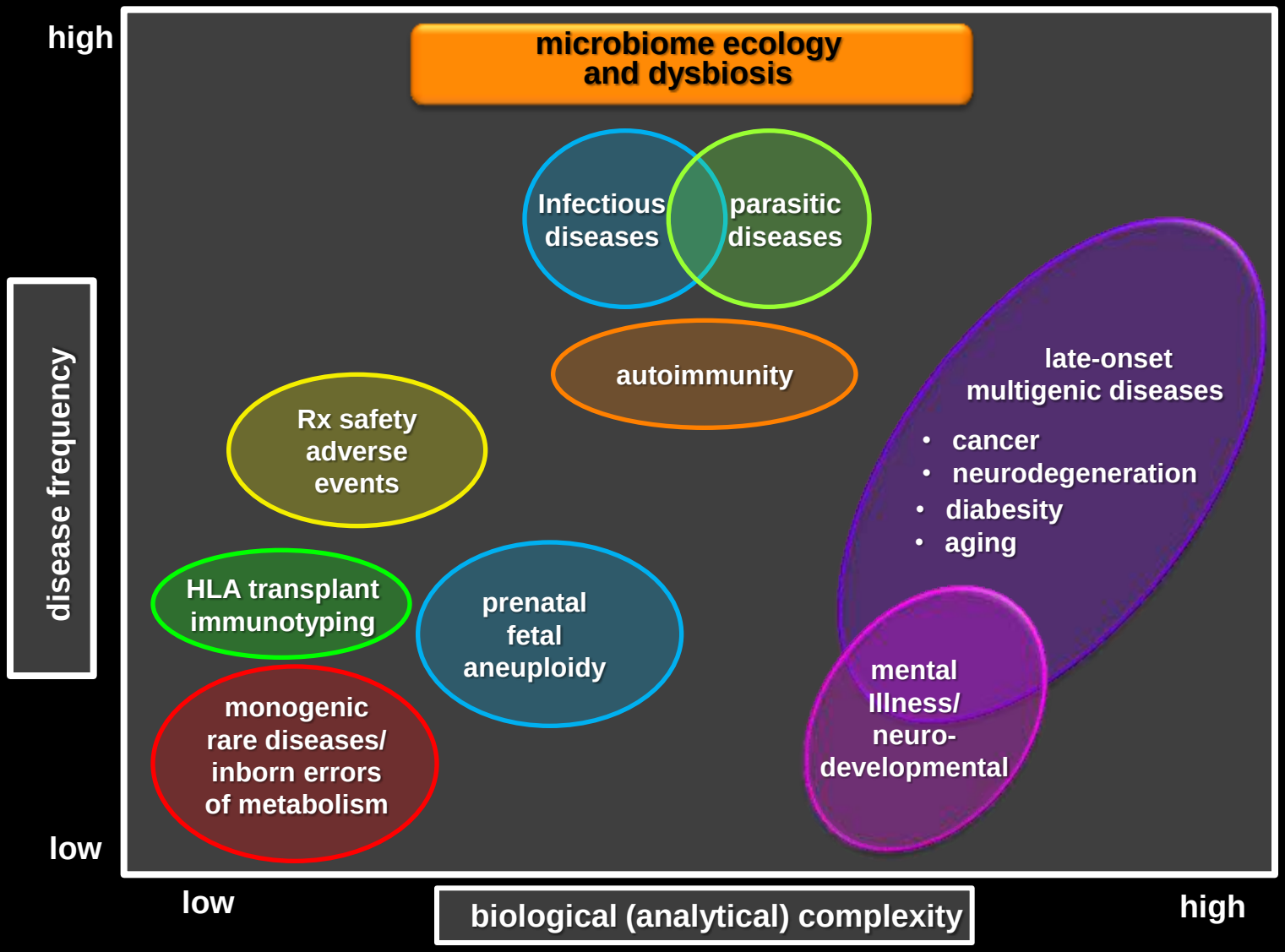
- **1000 reference genomes by 2020**

project blueprint

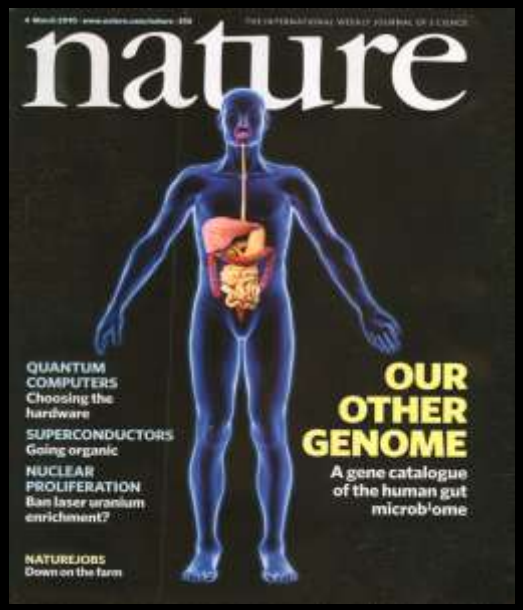
- **launch September 2011 with €30-million**
- **map epigenome in 60 human blood cell classes and neoplastic counterparts**

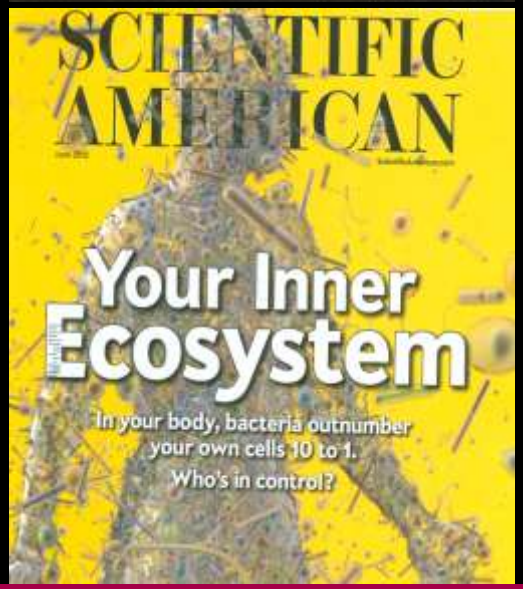

actionable data
high
low
high
disease frequency
low
Rx safety adverse events
autoimmunity
late-onset multigenic diseases
• cancer
• neurodegeneration
• diabesity
• aging
HLA transplant immunotyping
prenatal fetal aneuploidy
mental Illness/ neuro-developmental
monogenic rare diseases/ inborn errors of metabolism
monogenic, high-penetrance alleles
oligogenic, high or intermediate penetrance alleles
complex epistatic interaction of multiple low penetrance alleles

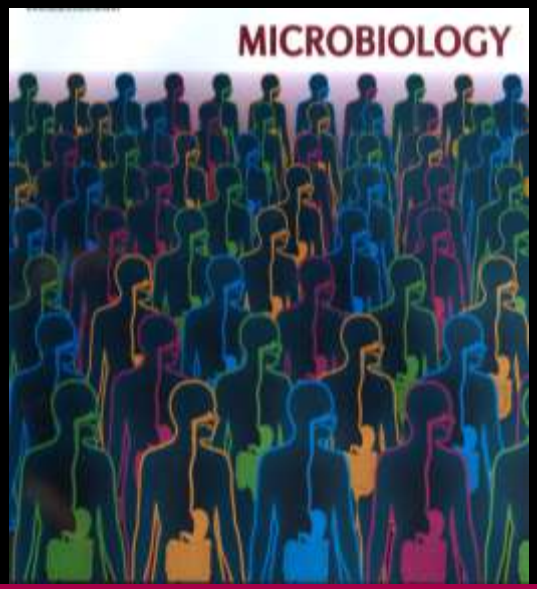
The Applications of Genome Sequencing



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

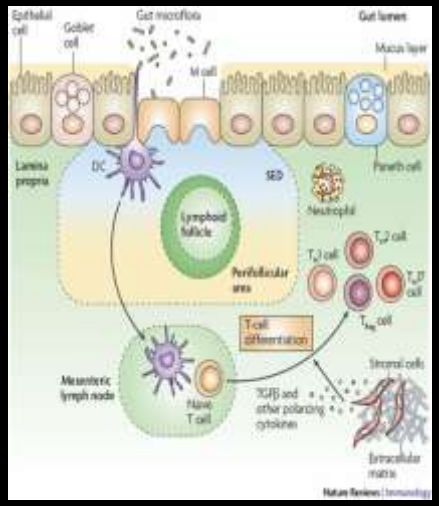


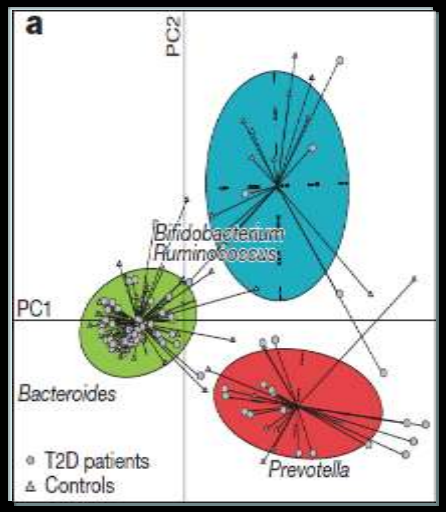


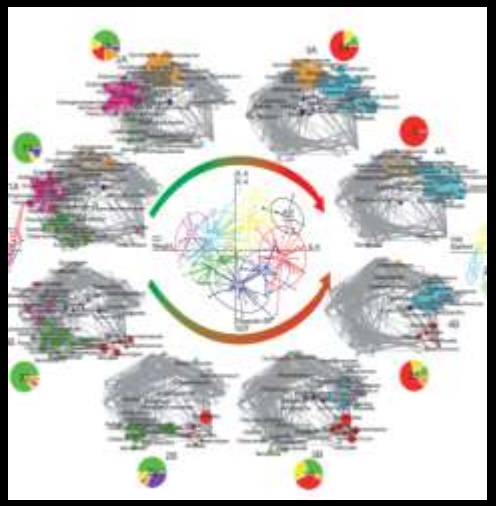


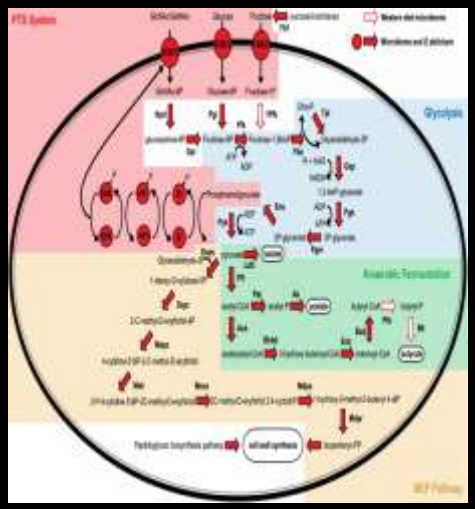
Commensal Microbiomes: The "Frenemy Within"
An Additional Dimension to Biomarker and Genome Profiling

Metagenome-wide Association Studies (MGWAS)



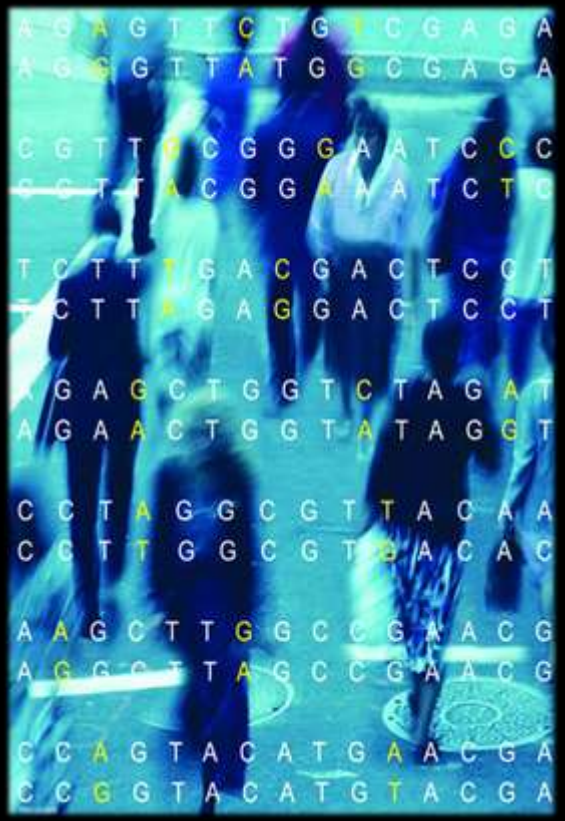






Immune-Mediated GI Diseases	Type 2 Diabetes Profile	Aging Metabolism and Fragility	Metabolic Activation of Carcinogens/ Pollutants

The Cost, Logistics and Infrastructure for Analysis and Management of Large Population WGS and 'Omics' Data and Integration Into Clinical Records



- big data
- big knowledge gaps

- big pipes
- big storage
- big bucks

- big payoffs?

The Omics Data Storage Challenge
(J. Starren et al. 2013 JAMA 309, 1237)

- typical EHR
  - 375 KB/patient
- radiologic picture archiving and communication system (PACS)
  - 104 MB/patient
  - x277 > EHR
- WGS
  - 3-10 million variants/individual
  - 5-10 GB/individual
  - x50 > imaging

Large Scale WGS, Big Data and Cyberinfrastructure:

EECS ELECTRICAL ENGINEERING AND COMPUTER SCIENCES
COLLEGE OF ENGINEERING UC Berkeley

A Million Cancer Genome Warehouse

David Haussler, David A. Patterson, Mark Diekhans, Armando Fox, Michael Jordan, Anthony D. Joseph, Singer Ma, Benedict Paten, Scott Shenker, Taylor Sittler and Ion Stoica

EECS Department
University of California, Berkeley
Technical Report No. UCB/EECS-2012-211
November 20, 2012

http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-211.pdf

- 1 million cancer patient WGS = 100 petabytes (after compression)
- not feasible to move such datasets
- not feasible to 'add on' to existing databases
- 'digital Darwinism': the prospect of stark separation between data-rich and data-poor enterprises

Omics Data and Healthcare Information Systems

- current paper records and most EHR/EMR formats ill-suited for facile integration of high dimensionality omics data
  - rapid ID of actionable data
  - ongoing dynamic alerts for altered individual risk from new discoveries

The Genome Is the Ultimate Personal Identifier

"….it will be hard for anyone
to find out anything about your personally
from this research"

Informed Consent Form for the 1000 Genomes Project (2008)

"profiling Y chromosome (Y-STRs),
recreational genetic geneology databases and
a combination of surname with other types of data,
such as age and state, can be used
to identify the target" (individual)

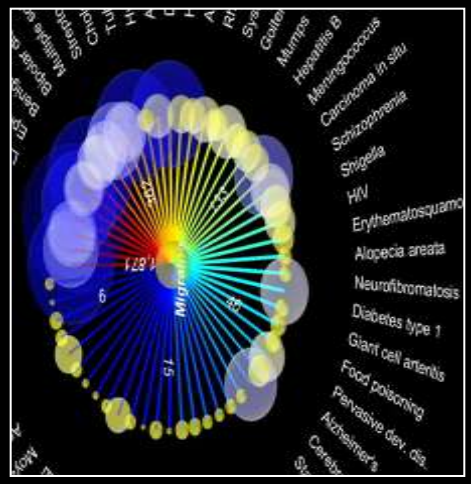
M. Gymrek et al. (2013) Science 339, 321

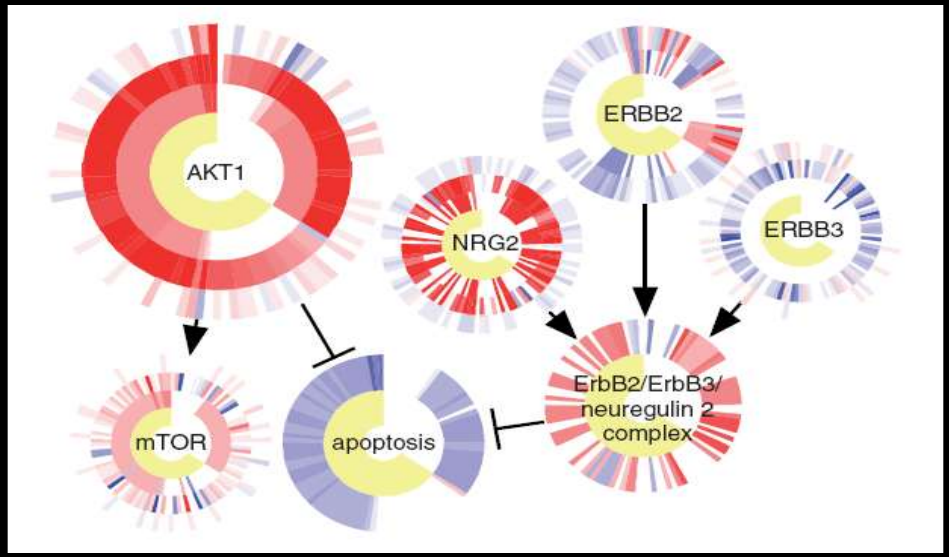
technology is outpacing physicians

technology is outpacing regulators

technology adoption is handicapped by outdated clinical guidelines (SOC) and reimbursement policies

Integration of iOmics Data Into Electronic Health Records and Clinical Decisions

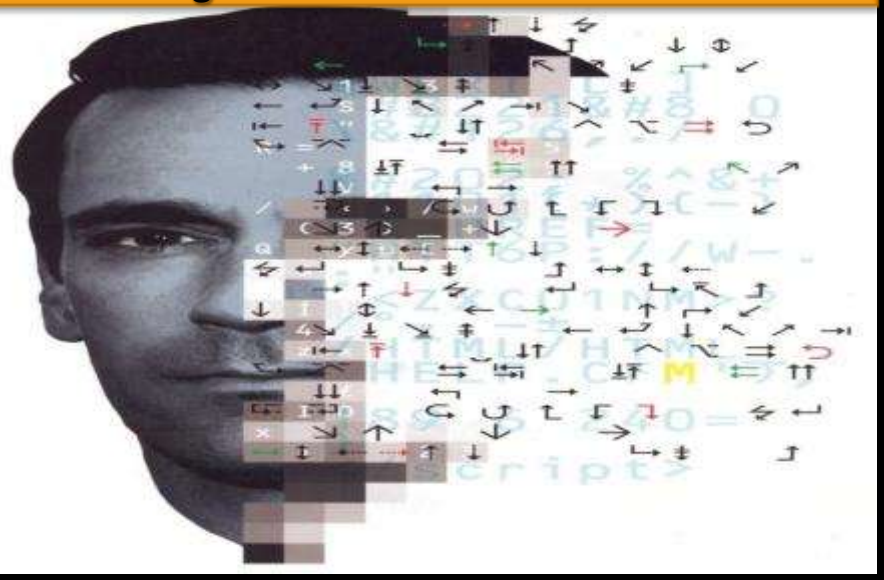




Technology Acceleration and Convergence: The Escalating Challenge for Professional Competency, Decision-Support and Future Education Curricula

Data Deluge

Cognitive Bandwidth Limits



Automated Analytics and Decision Support

Facile Formats for Actionable Decisions

Education in Molecular (Precision) Medicine and a Looming Skills Gap

- clinical specialists with genomic (omics) expertise
  - molecular genetics, pathology, genetic counseling
  - “molecular medicine 101” and CME for healthcare professionals
- informaticians and quantitative analytics for biomedicine
  - computational informatics, statistics and mathematics
  - database design and curation for optimized data flows for clinical decisions
  - data customization and visualization for different end-users
- social media and medical apps.

Reimbursement for Molecular Diagnostics

Reimbursement for Molecular Diagnostics and Related Omics Profiling Tests

- current payment policies based on earlier era of comparatively simple (low technical complexity) tests
  - time and materials used to conduct test
  - no premium for cost recovery for escalating test complexity/R&D investment for next-generation “Omics” tests
- failure of CPT coding to match pace of technical advances in MDx/WES/WGS
- inadequate HTA/reimbursement/business models for value-based pricing of next-generation diagnostic platforms

Summary

- balancing advances in cost:performance in sequencing versus time:cost and clinical utility of data generated
- current value propositions for immediate clinical utility of profiling single genes, gene panels, whole exomes and whole genomes
- single genes
  - monogenic inherited diseases and other rare disorders
  - polymorphisms in drug metabolism genes (CYP, HLA) and Rx adverse event risk
  - genotyping for transplantation
- gene panels
  - microbial diagnostics, biosurveillance and public health
  - cancer, diabetes, autoimmunity

- **WES and WGS**
  - **current paucity of clinically actionable information**
  - **risk/progression of major late-onset diseases determined by as yet largely unknown multi-gene interactions (epistasis), epigenetic modifications and other environmental confounders**

Setting Technical Standards and Keep Pace with Technology

- if it isn’t billable it won’t happen!
- urgent need for rigorous analytical standards, regulatory guidelines and value-based pricing to incentivize R&D investment, clinical validation and clinical adoption
- technology and production of sequencing data are outpacing clinical expertise and cogent regulatory and reimbursement policies

Making Precision Medicine a Reality: The Challenge of Escalating Complexity and Proficient Analysis, Management and Use of Massive Data Streams

"In God we trust,
all others must bring data"

W. Edwards Deming, Statistician

- validated data
- actionable data
- accessible data
- integrated data for knowledge-driven decisions

Slides available @ http://casi.asu.edu/