ASU Complex Adaptive Systems Initiative
Arizona State University

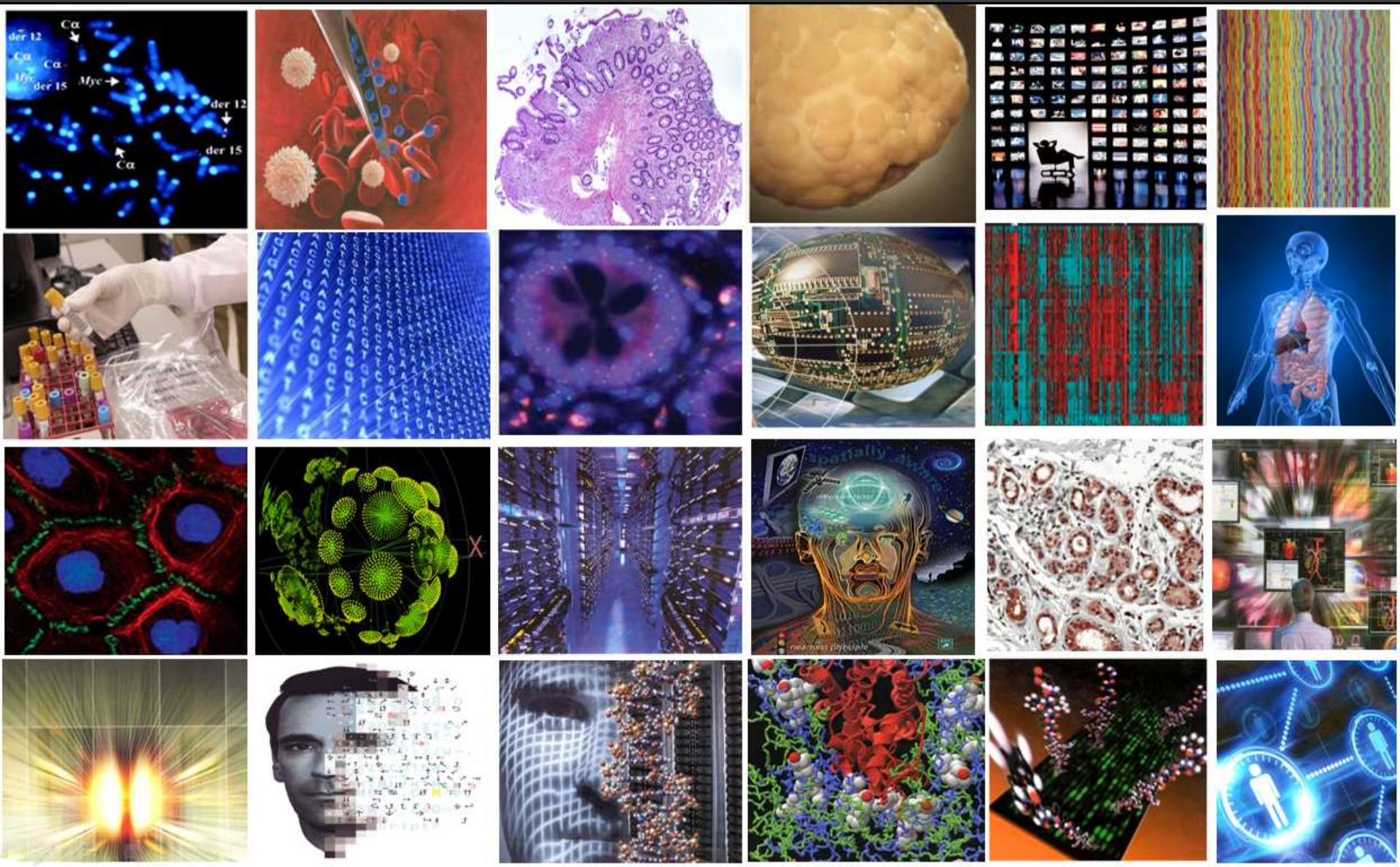
Burrill & Company
Life Sciences: Venture Capital, Private Equity, Merchant Banking, Media

New Value Propositions in Healthcare: Molecular Medicine and Large Scale Data Integration Services

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

Presentation at Burrill and Co. LP Meeting:
Beyond the Genome: Opportunities in Large Scale Sequencing
Cavallo Point Lodge, California ● 17 April 2012

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



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

New Value Propositions

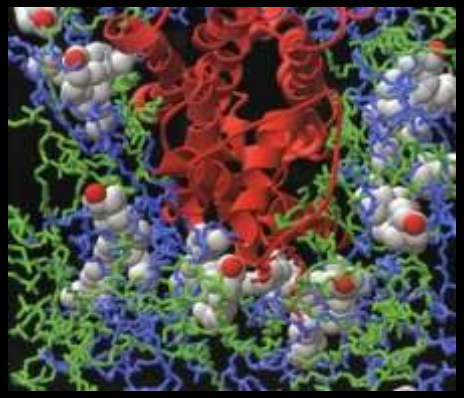
Emergence of a New Health Information Ecosystem and Business Models via Convergence of Molecular Medicine, Digital Networks and Social Media

Shift from Reactive, Incident-Centric Care to Proactive Engagement to Mitigate Individual Risk

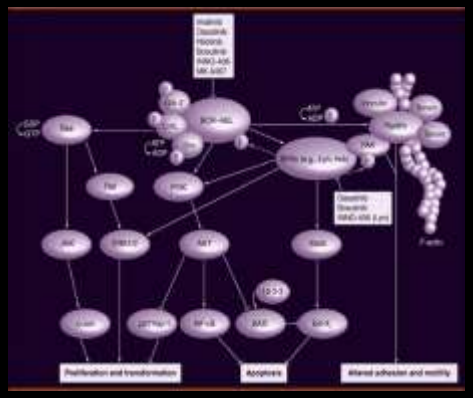
Mapping The Molecular Signatures of Disease: The Intellectual Foundation of Rational Diagnosis and Treatment Selection

Genomics

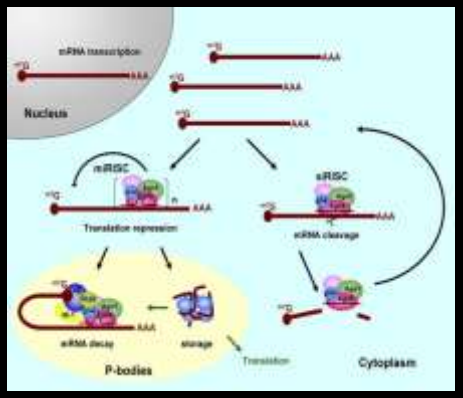
Proteomics

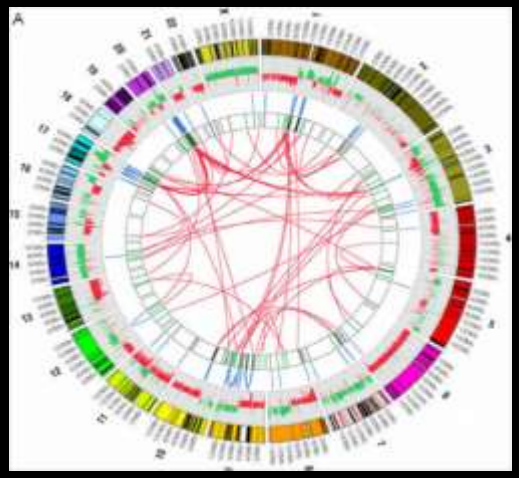


Molecular Pathways and Networks



Network Regulatory Mechanisms



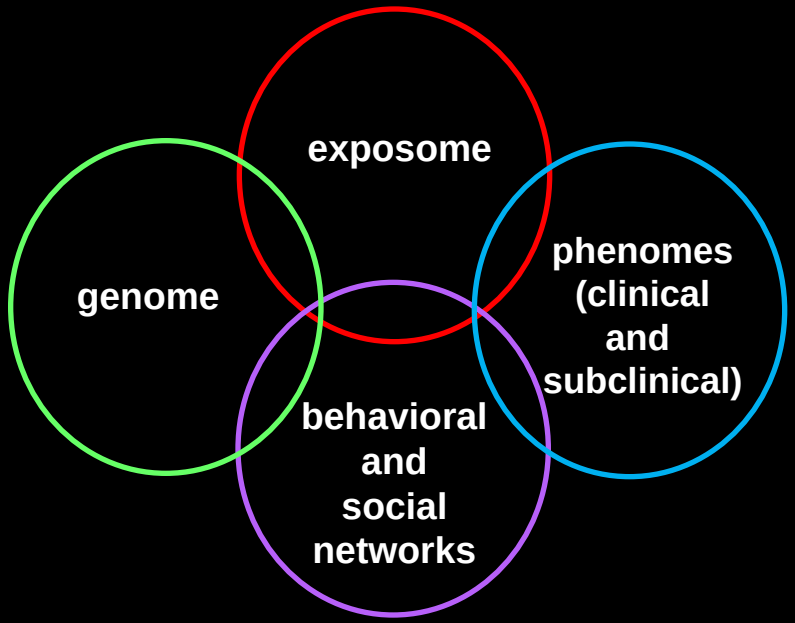


ID of Causal Relationships Between Network Perturbations and Disease

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

Information-Based Services for Healthcare and Wellness

Precision Profiling of Health Status



Population- and Individual-Datasets

facile integration and analysis of diverse datasets

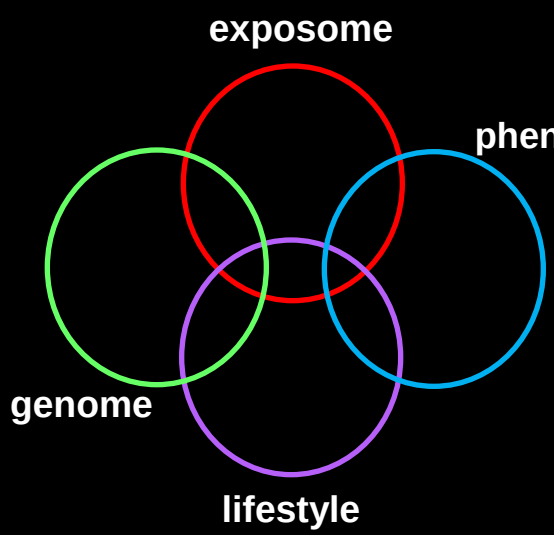
Actionable Information

risk identification and mitigation

Integrated Care and Wellness

VALUE

Information-Based Services for Healthcare and Wellness



data curation and analysis

risk identification and mitigation + decision-support

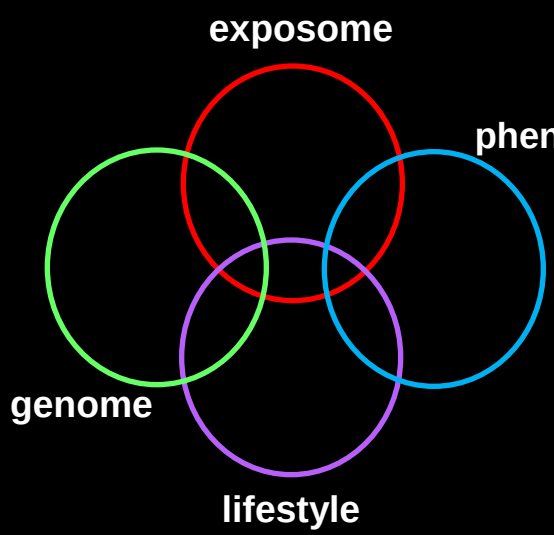
- earlier detection of disease
- rational Rx
- monitoring of health status
- predisposition risk

profile

analyze

act

Information-Based Services for Healthcare and Wellness



data curation and analysis

risk identification and mitigation + decision-support

- earlier detection of disease
- rational Rx
- monitoring of health status
- predisposition risk

profile

analyze

act

Will Low Cost Whole Genome Sequencing Change Everything?

- first 1TB USB
- $2K price tag
- store 350-400 human genomes plus annotation

Source: Davies, Kevin (2011). The Road to the 1000 Genome. PHT/SLA Spring Meeting

Systems Not Silos

Sequencing is a Silo

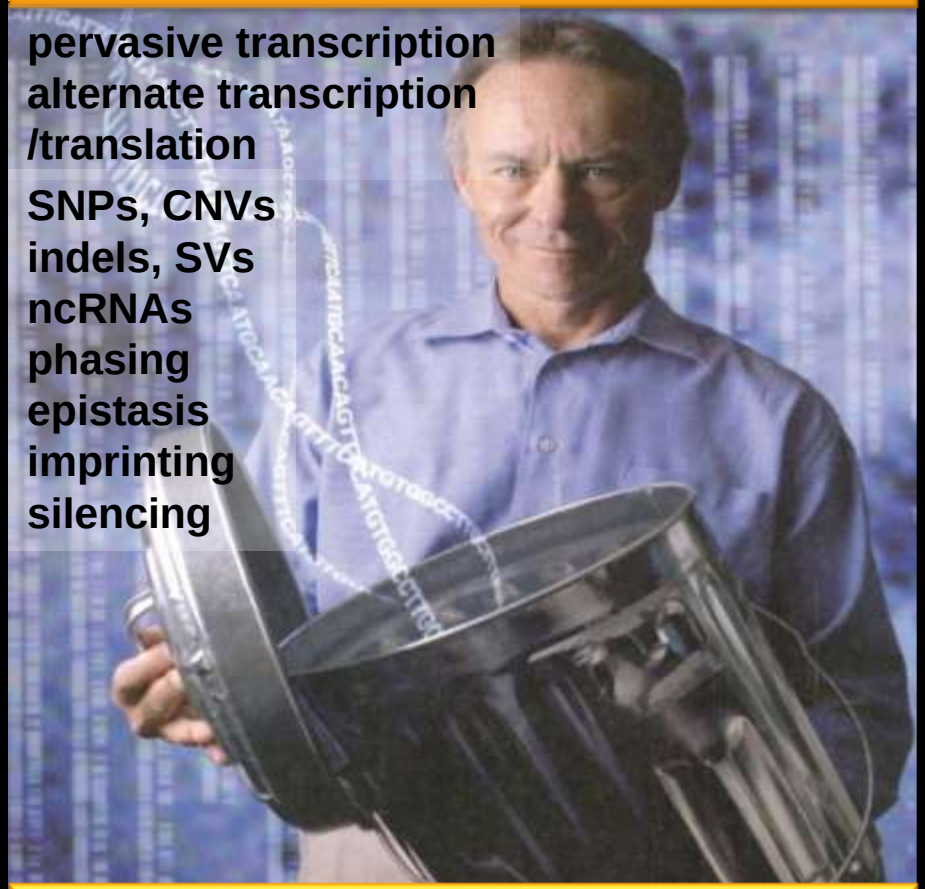
Beyond the Genome: Integration of Sequencing
With ‘other Omics, Environmental and Socio-Cultural Factors

Managing Massive Data and Imperative for High Quality Data

What is A Complete and Accurate Analysis of Genome Sequence, Architecture, Topology and Regulation?

The Complexity of Genome Organization
More Than a Linear Sequence

Junk No More!

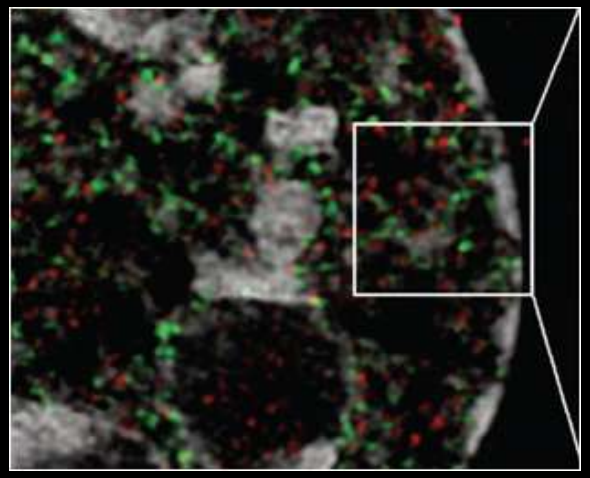


pervasive transcription
alternate transcription /translation

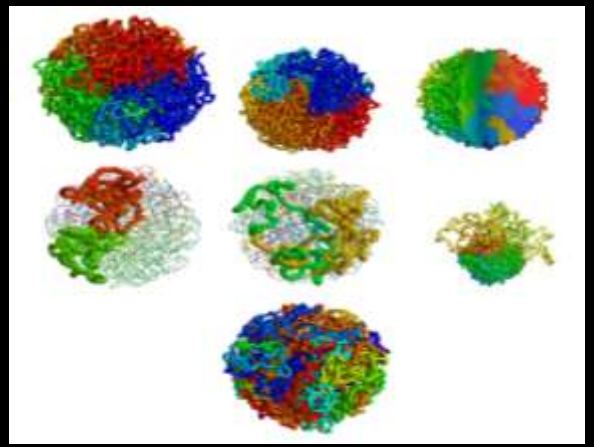
SNPs, CNVs
indels, SVs
ncRNAs
phasing
epistasis
imprinting
silencing

recognition of genome organizational and regulatory complexity

Chromatin Loop Domains



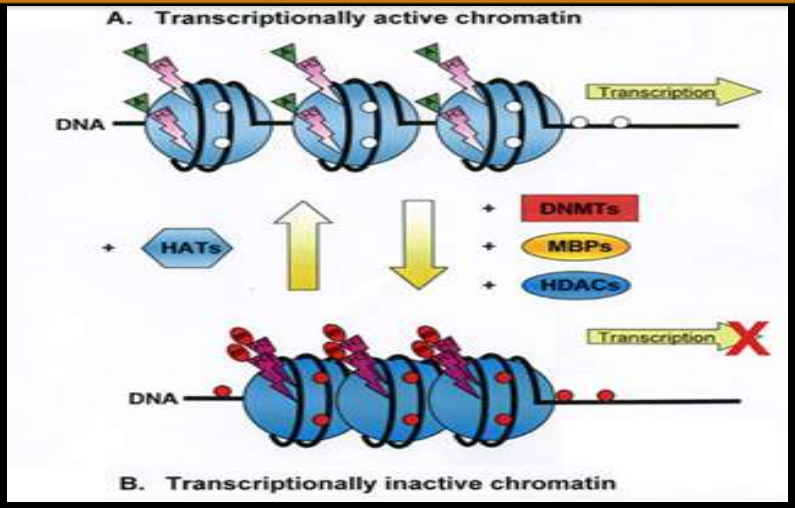
Modeling of Nucleosome Folding



The Epigenome

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

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



IHEC
International Human Epigenome Consortium

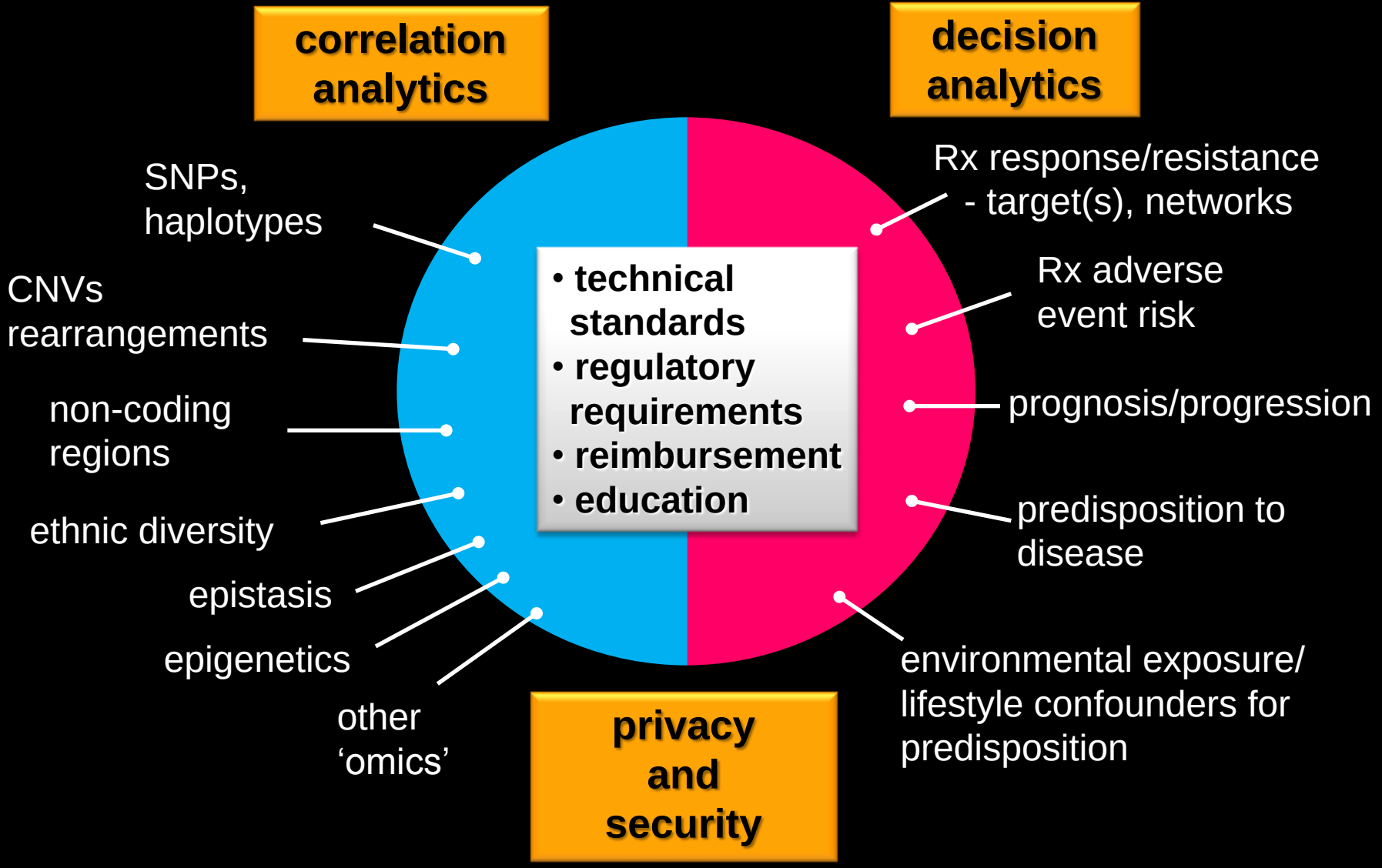
International Human Epigenome Consortium

- **1000 reference genomes by 2020**

project blueprint

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

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



Standards for Large Scale Profiling Technologies

Genome–Based Diagnostics: Clarifying Pathways to Clinical Use – Workshop Report

Released: March 20, 2012
Type: Workshop Report
Topics: Biomedical and Health Research, Public Health
Activity: Roundtable on Translating Genomic-Based Research for Health
Board: Board on Health Sciences Policy

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

Review of Validation Issues for Clinical Use of Genome Sequencing 23 June 2011

- criteria to assess platform accuracy
- minimum sequencing depth for reliable clinical decisions
- appropriate validation sample sets to evaluate platform accuracy
- metrics for quality of sequence assembly and alignment algorithms
- standardization of pre-analytical variables (e.g. preparation of libraries, extraction and quality control of nucleic acids, capture methods, amplification)

Regulatory Issues in Genome Sequencing for Clinical Decisions

- accuracy, depth of coverage, validation set, impact of pre-analytic/analytic variables
- CLIA/CAP facilities
- sequencers as Class III devices?
- RUO materials
- source computer code(s) for analytical algorithms
- performance thresholds and QA/QC requirements for error detection (instrumentation + analytics)

Performance Comparison of WGS Platforms
(H.Y.K. Lam et. al. 2012 Nature Biotechnol. 30, 78)

- sequencing of blood and saliva samples from same individual on Illumina and Complete Genomics Platforms at 76x coverage
- only 88.1% SNVs concordant ≡ 10,000s platform-specific calls in exons and intergenic regions
- need to supplement with exome sequencing to fill gaps in detection of coding variants
- only 26.5% indels concordant
- implications for use of WGS data for clinical decisions/regulatory submissions

The N of One: Large N Dilemma

N of One

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

Large N

- biomarker validation requires large statistically powered sample sets
  - high dimensionality markers (10^3 - 10^6)/WGS and small sample sets (10^1 - 10^2) and risk of overfitting
  - large N of 10^2 - 10^3 patients
  - logistics and cost of screening candidate pool for low frequency markers (e.g., ALK, ROS in NSCLC)

Statistical Sampling Powering Needs for Variant ID in Validation Studies: The Proportion of Theoretically Identifiable Variants in Different Population Sample Cohort Sizes (N)

Variant Frequency	N			
	100	200	500	1000
0.001	0.18	0.33	0.63	0.86
0.002	0.33	0.55	0.86	0.98
0.005	0.36	0.86	0.99	1.00
0.01	0.86	0.98	1.00	1.00

Adapted from: L. Bingsham and S.M. Leal (2009) PLoS Genetics 5, e1000481

“The Incidentalome”

- 2012 NIH proposal for screening exome-and WGS sequence data for findings of potential health or reproductive importance
- obligation to recontact/deidentify individuals in research studies
- criteria for “relevant” and “risk” returnable findings?
- requirement to reidentify original donor in deidentified samples?
- resources and cost to implement with anticipated rapid growth in datasets?
- why limit to genomic research using biobanks and archived data?
- if research participants are accorded duties why not all patients sequenced as part of clinical care?
- expanded IRB responsibilities and competencies

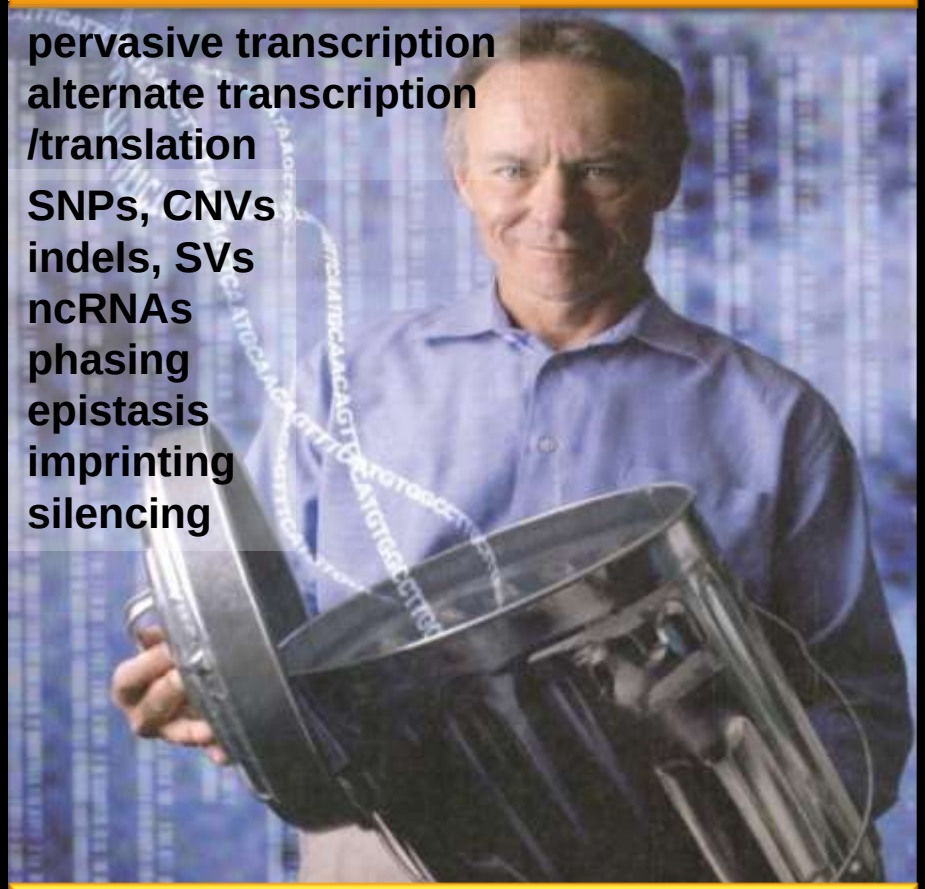
Presidential Commission for the Study of Bioethical Issues

Fed. Reg. 27 March 2012
Implications of Large Scale Human Genome Sequencing

- collection, use and governance of exome- and WGS information
  - genetic/genomic databases and biobanks
  - role of health IT
- privacy and access
- balancing of individual and societal interests
- access and use by law enforcement agencies

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

Junk No More!

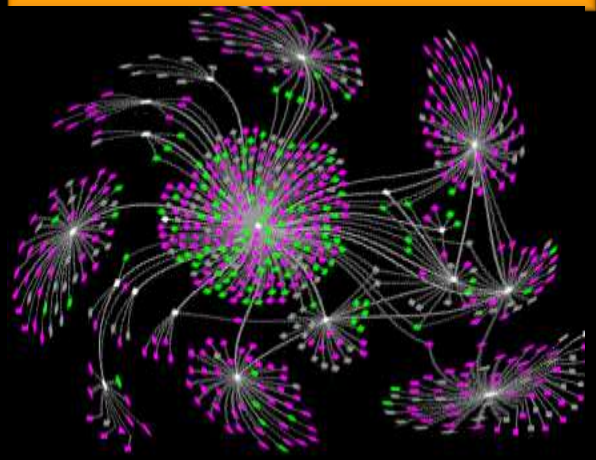


pervasive transcription
alternate transcription /translation

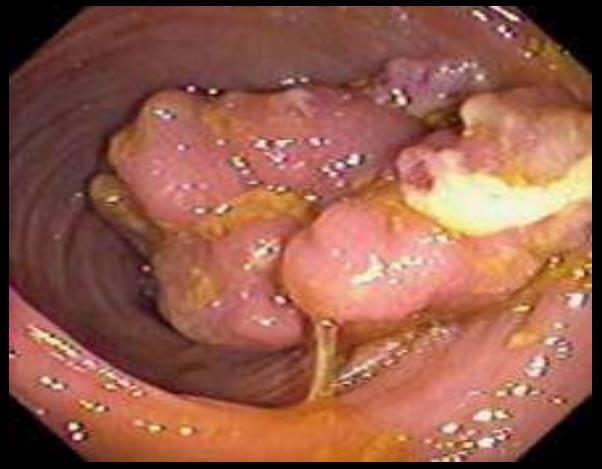
SNPs, CNVs
indels, SVs
ncRNAs
phasing
epistasis
imprinting
silencing

recognition of genome organizational and regulatory complexity

Cell-specific Molecular Interaction Networks

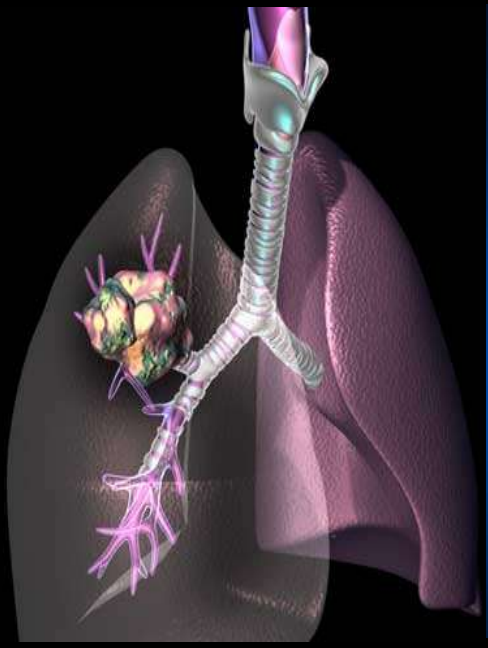


Disease Perturbations



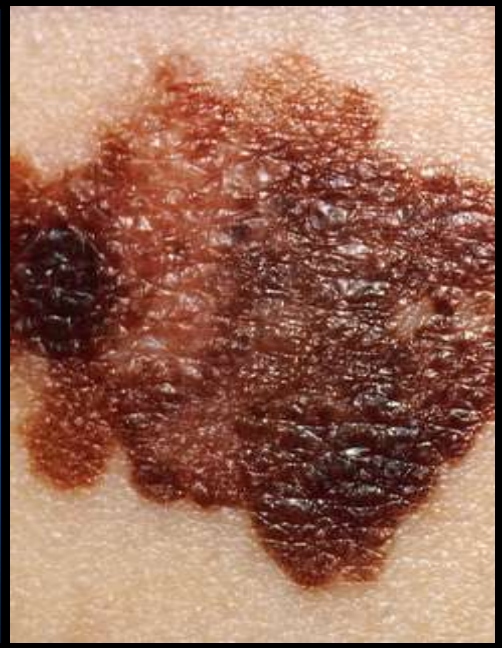
Mapping the Molecular Signatures of Disease, Disease Subtyping and Targeted Therapy: The Right Rx for the Right Disease (Subtype)

Her-2+
(Herceptin)



EML4-ALK
(Xalkori)

KRAS
(Erbitux)
(Vectibix)



BRAF-V600
(Yervoy)
(Zelboraf)

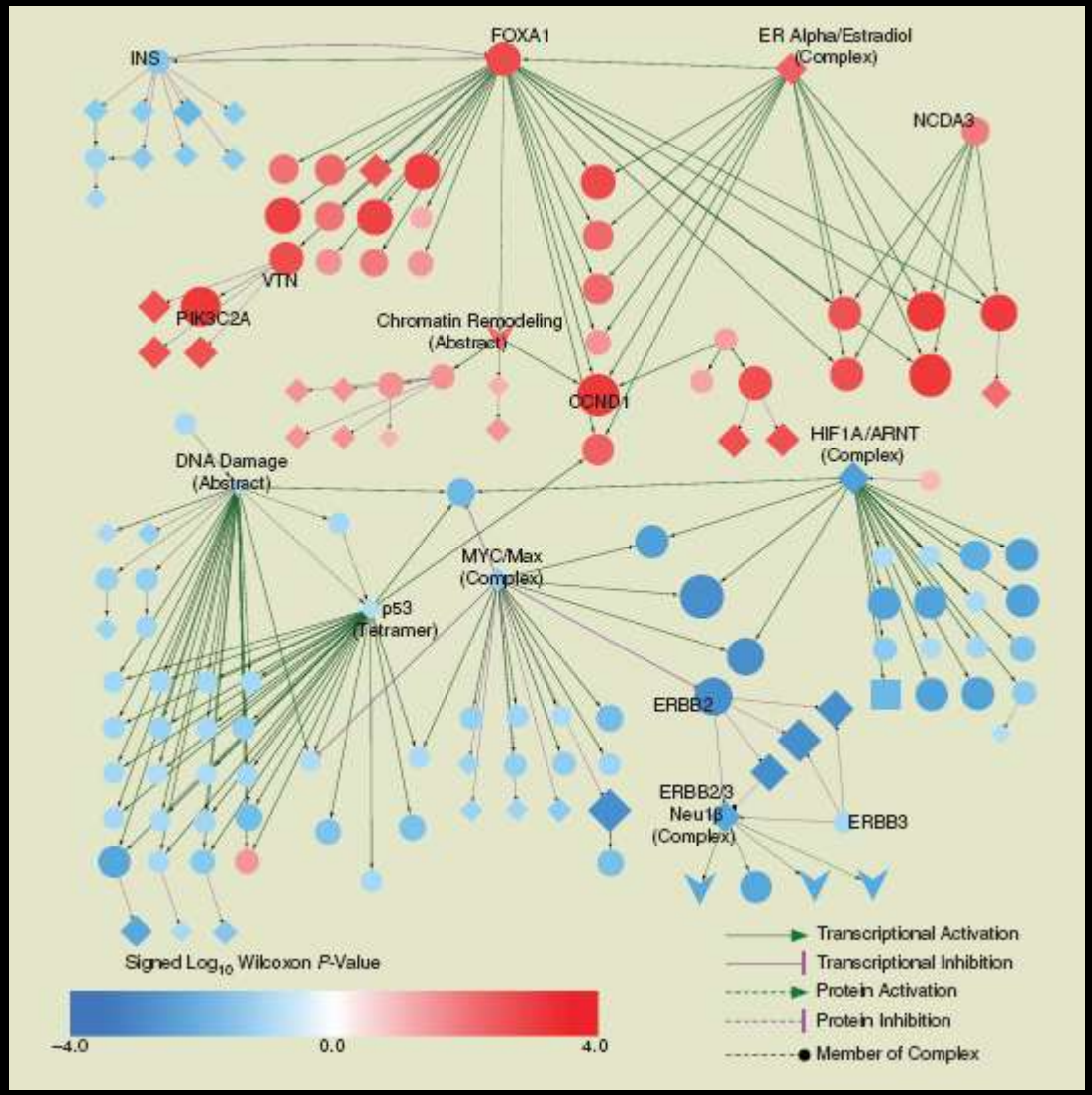
Reducing The Failure Rate of Investigational Drugs in Clinical Trials

- targeted therapies, YES!

but

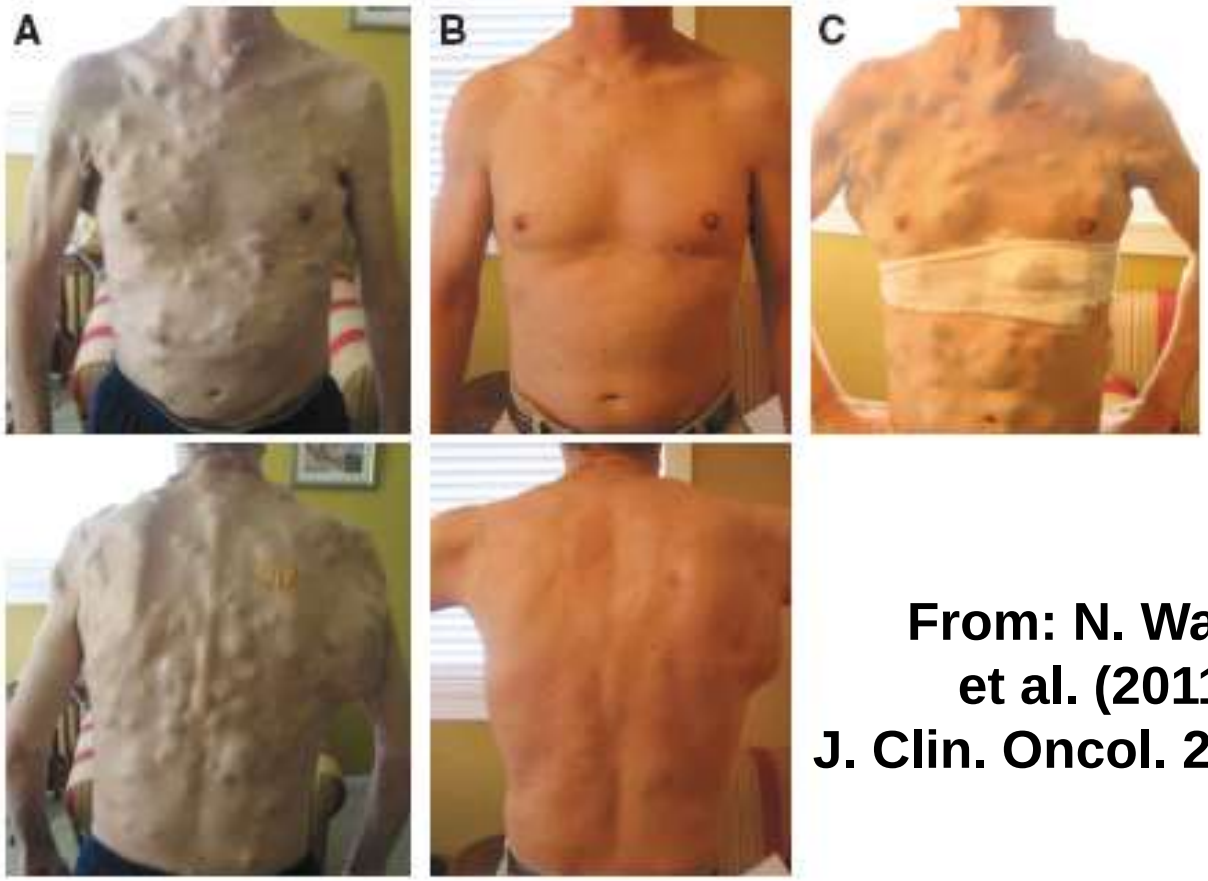
- improved success requires targeting network modules, pathways and subnetworks not single targets
- complexity of linked and overlapping modules and pathway “cross-talk”
  - long range pleiotropic effects
  - weak indirect effects

PARADIGM Modeling of Genetic Regulatory Networks in ER+ Breast Cancer: Up-regulation of ER and FOXA1 Networks and Down Regulation of HIF-1-Alpha p53 and MYC Networks



From: V. Varadan et al. (2012) IEEE Sig. Proc. 29, 43

Initial Response (A/B) of BRAF-V600 Positive Metastatic Miliary Melanoma After 15 Weeks Therapy with Vemurafenib (Zelboraf® - Roche) Followed by Rapid Recurrence of Rx-Resistant Lesions with MEKI C1215 Mutant Allele After 23 Weeks Therapy

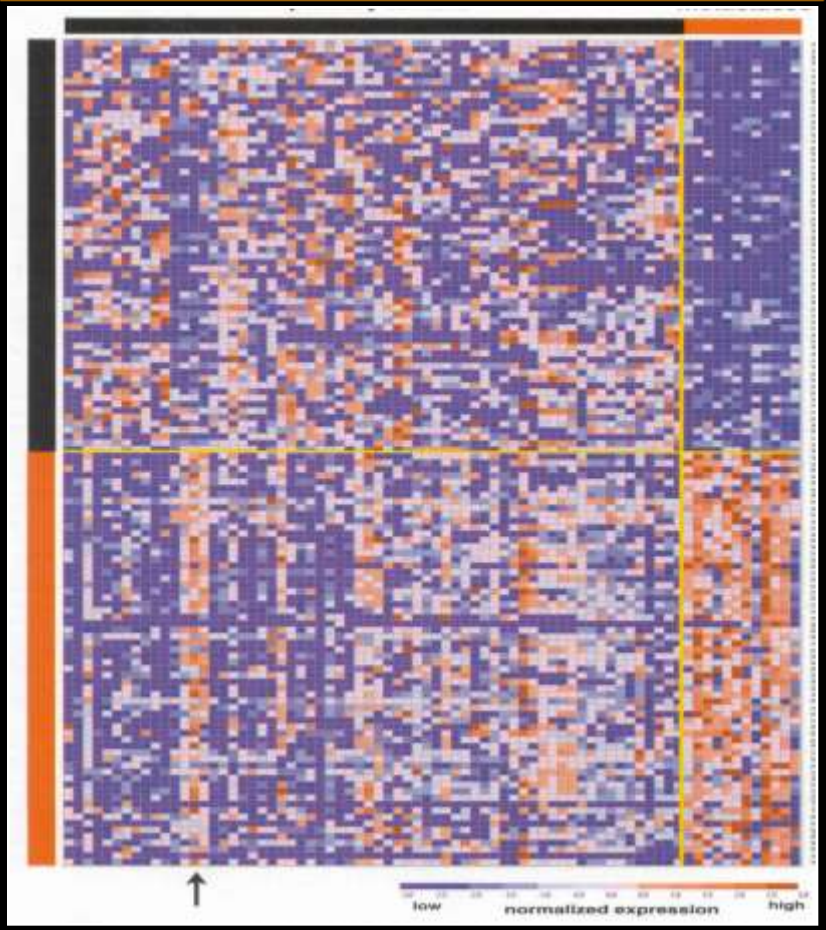


From: N. Wagle et al. (2011) J. Clin. Oncol. 29, 3085

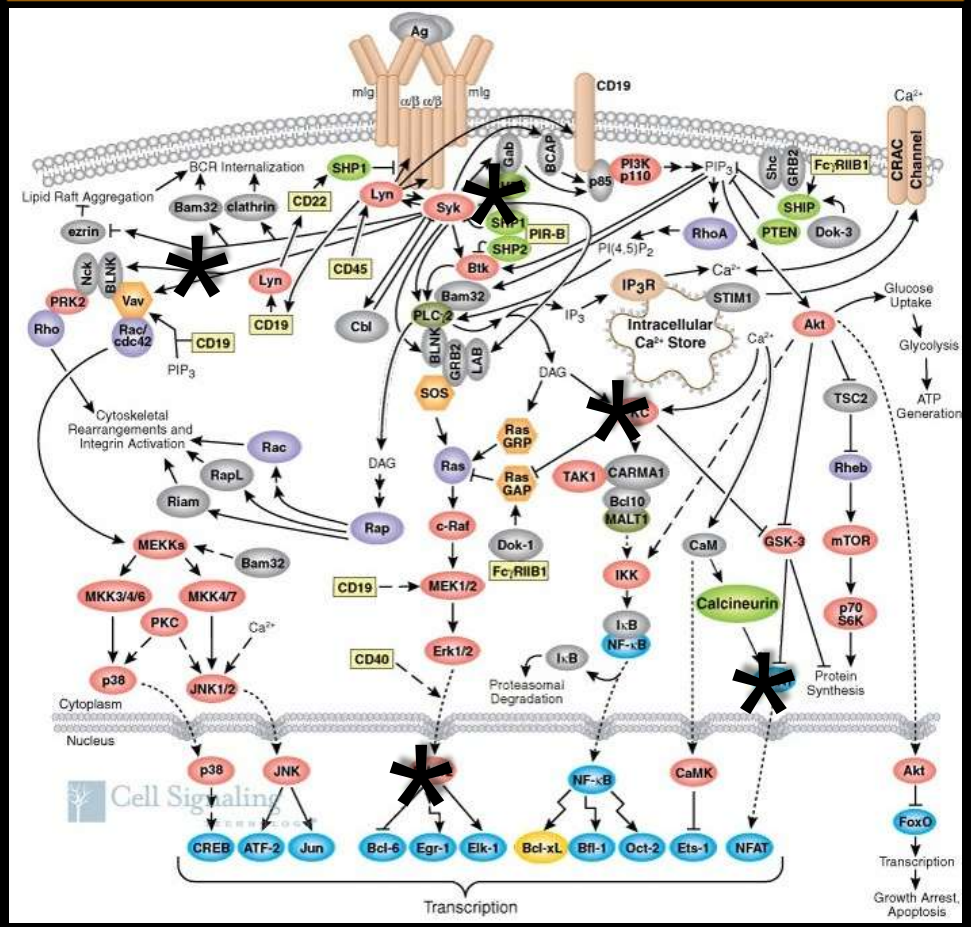
http://jco.ascopubs.org/content/29/22/3085.full.pdf+html

Understanding the Internal Circuit Diagrams of Cells and Identification of the Disruption(s) Caused by Disease

Disease Profiling to Identify Subtypes (+ or - Rx Target)



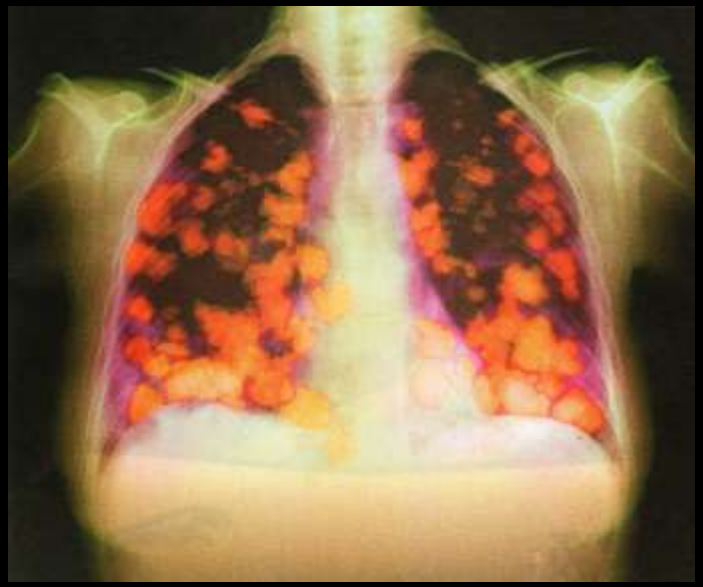
ID Molecular Targets for Rx Action and Blockade of Compensatory "By pass" Pathways



Network Pharmacology

- elucidation of definitive ‘chokepoints’ as optimum targets
  - subvert adaptive cellular options to use alternate compensatory pathways
- the design challenge for multi-target polypharmacology
  - multi-agent therapy (patient tolerance?)
  - controlled multi-target promiscuity in a single moiety
- does chronic progression in complex, multigenic diseases amplify module/subnetwork dysregulation?

Cancer: A Formidable Therapeutic Foe



- is the dysregulation of pathways/modules in advanced disease so extravagant that Rx 'homeostatic reset' is unlikely?

plus

- progressive genomic and phenotypic heterogeneity and intra- and inter-lesional heterogeneity

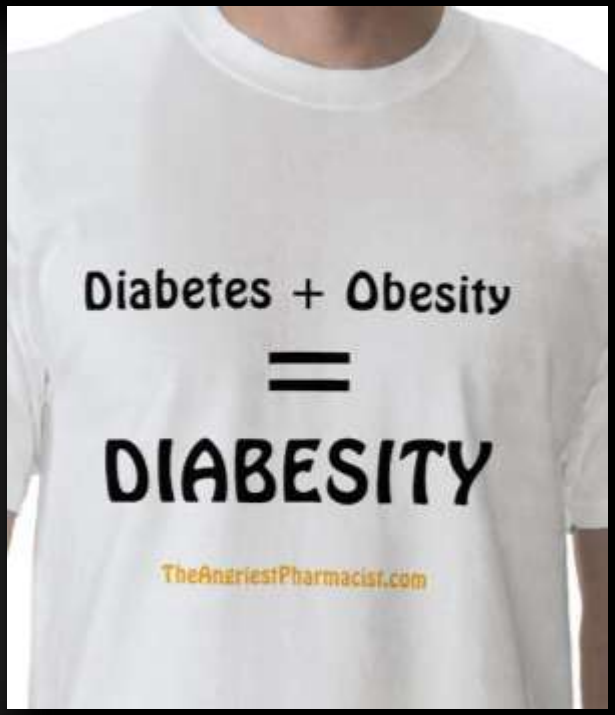
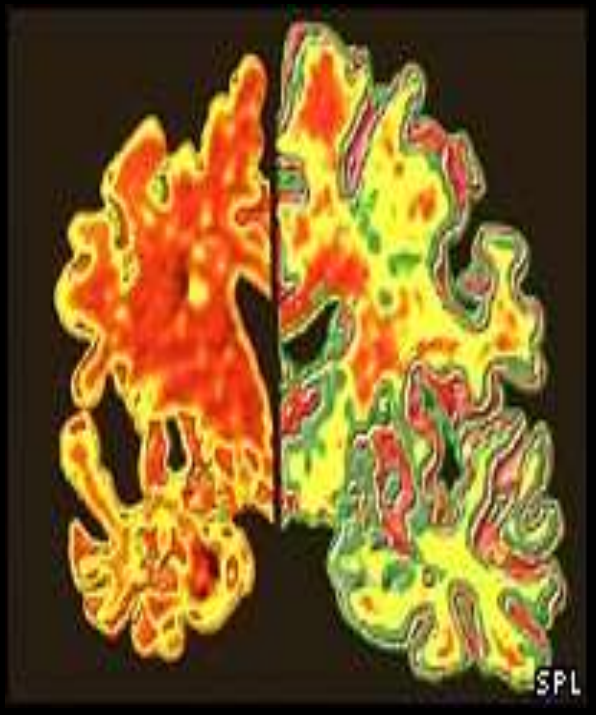
JOURNAL OF CLINICAL ONCOLOGY

ASCO SPECIAL ARTICLE

American Society of Clinical Oncology Provisional Clinical Opinion: The Integration of Palliative Care into Standard Oncology Care

Thomas J. Smith, Sarah Temin, Erin R. Alesi, Amy P. Abernethy, Tracy A. Balboni, Ethan M. Basch, Betty R. Ferrell, Matt Loscalzo, Diane E. Meier, Judith A. Paice, Jeffrey M. Peppercorn, Mark Somerfield, Ellen Stovall, and Jamie H. Von Roenn

Opportunities and Challenges for MDx for Ever Earlier Detection of Major Diseases

Cancer Detection Before Metastasis	Cardiovascular/ Metabolic Diseases	Neurodegenerative Diseases
		
Early Diagnosis and Curative Surgery	Lifestyle Changes and/or Rx to Limit Risk	The Dilemma of Early Diagnosis Without Rx

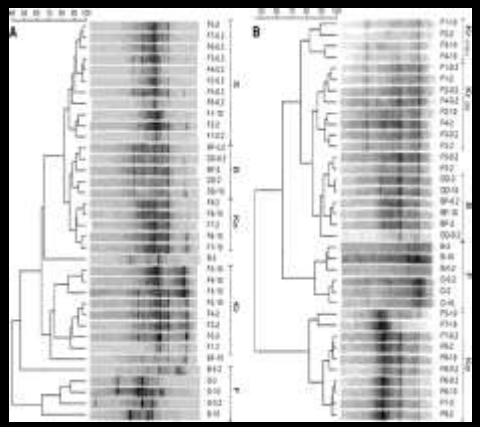
Biomedicine as a Data- and Computation-Intensive Exercise

BIG DATA, META-DATA AND META-KNOWLEDGE

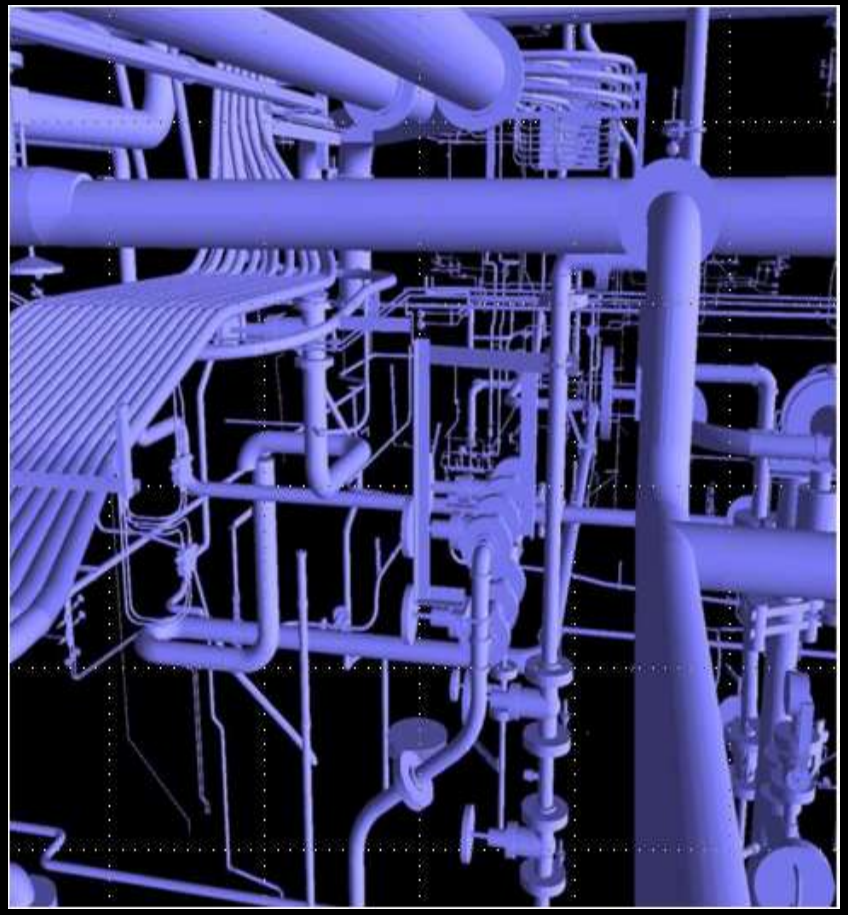
DATA STANDARDS, FORMATS AND FORMALISM

INFRASTRUCTURE, INVESTMENT, INTELLIGENCE

Data-Intensive Biomedical R&D and 'The Data Deluge'

Patient Stratification For Clinical Trials	Pharmacogenomics	m.Health	Monitoring Networks
			
			
Microbial Diagnostics	Biosurveillance and Public Health	Health IT and EMRs	Computing Infrastructure

Cyberinfrastructure and Movement of Big Data
Not All Pipes Are Created Equal



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

Biomedical Data in the Cloud

NewScientist Jobs

- research data (deidentified/anonymized)

  vs.

  clinical trial

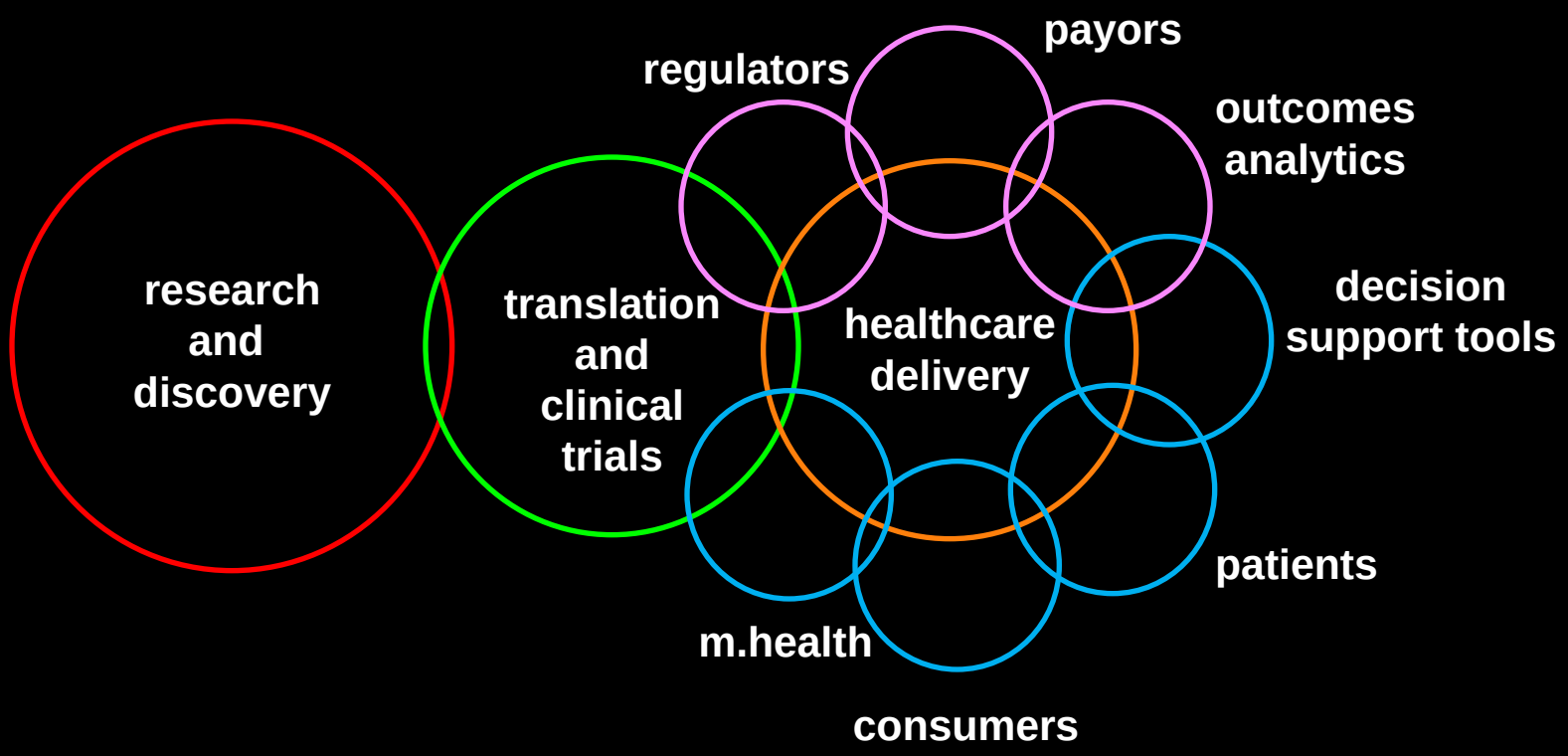
  and

  healthcare data

- confidentiality, privacy and security
- consent for facile integration of cloud-based services

21st Century Knowledge Networks
versus
20th Century Organizations

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

Mining EHRs to Identify Disease Correlations with Molecular Profiling Datasets and Improved Clinical Stratification (Phenotyping) of Patient Cohorts

- 18.688 million medical members
- 13.953 million dental members
- 10.410 million pharmacy members

- 966,000 healthcare professionals
- 543,000 primary care doctor specialists
- 5,200 hospitals

- 71 billion health records
- 75 TB storage (50% occupied)

From: Health Data Sept. 2011

Collaborative Clinical Connectivity and Open Source Technologies

PACeR (www.pacerhealth.org)
The Partnership to Advance Clinical Electronic Research

- use of aggregated EMR data to facilitate clinical research

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

- Collate, Annotate, Curate and Host 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®

CYCORE

caBIG cancer Biomedical Informatics Grid®

army of women

LOVE/AVON Going Beyond a Cure

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

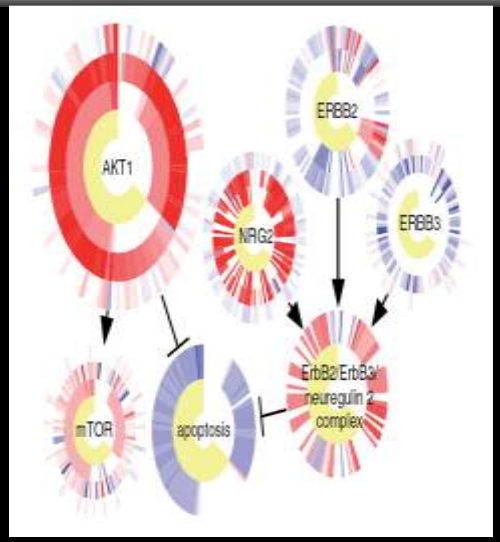
Jeopardy 16 February 2011

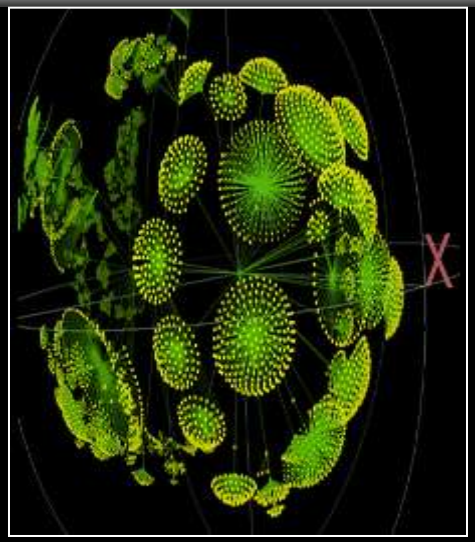
- IBM's Watson
  - 2880 CPUs
  - natural language questions
- prelude to Q&A systems for biomedicine beyond keyword IR searches

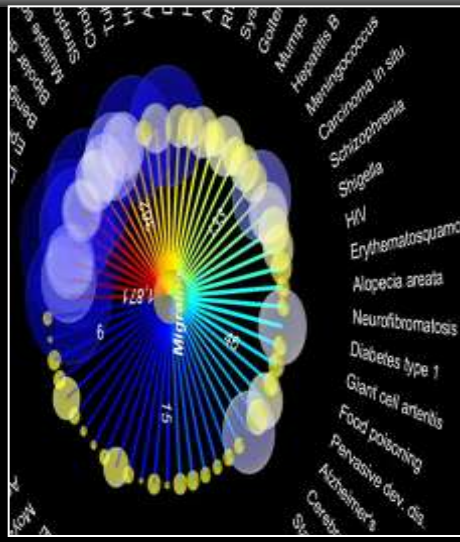
WELLPOINT | Health.Care.Value.™

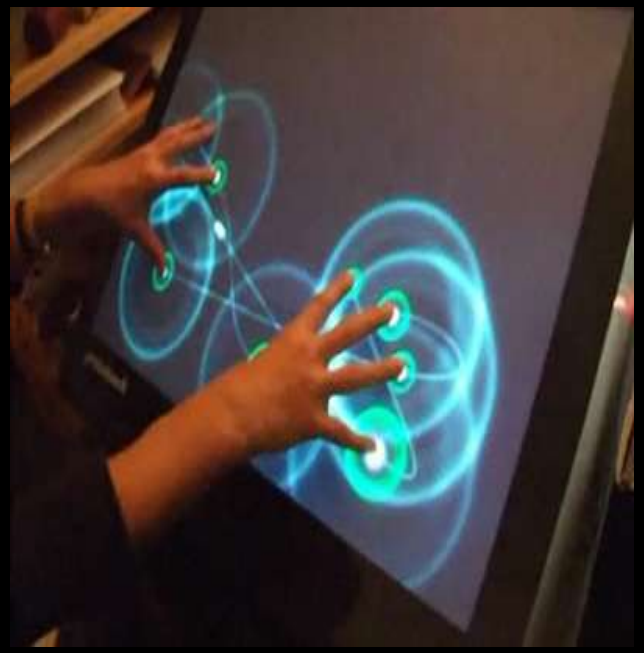
NUANCE®

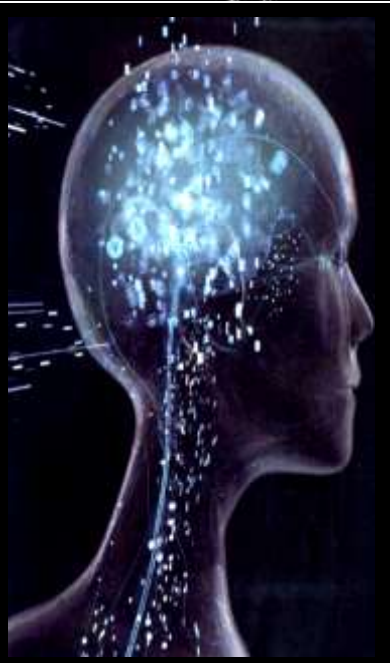
New Visualization Tools, Interactive Interfaces and Rapid Customization Formats



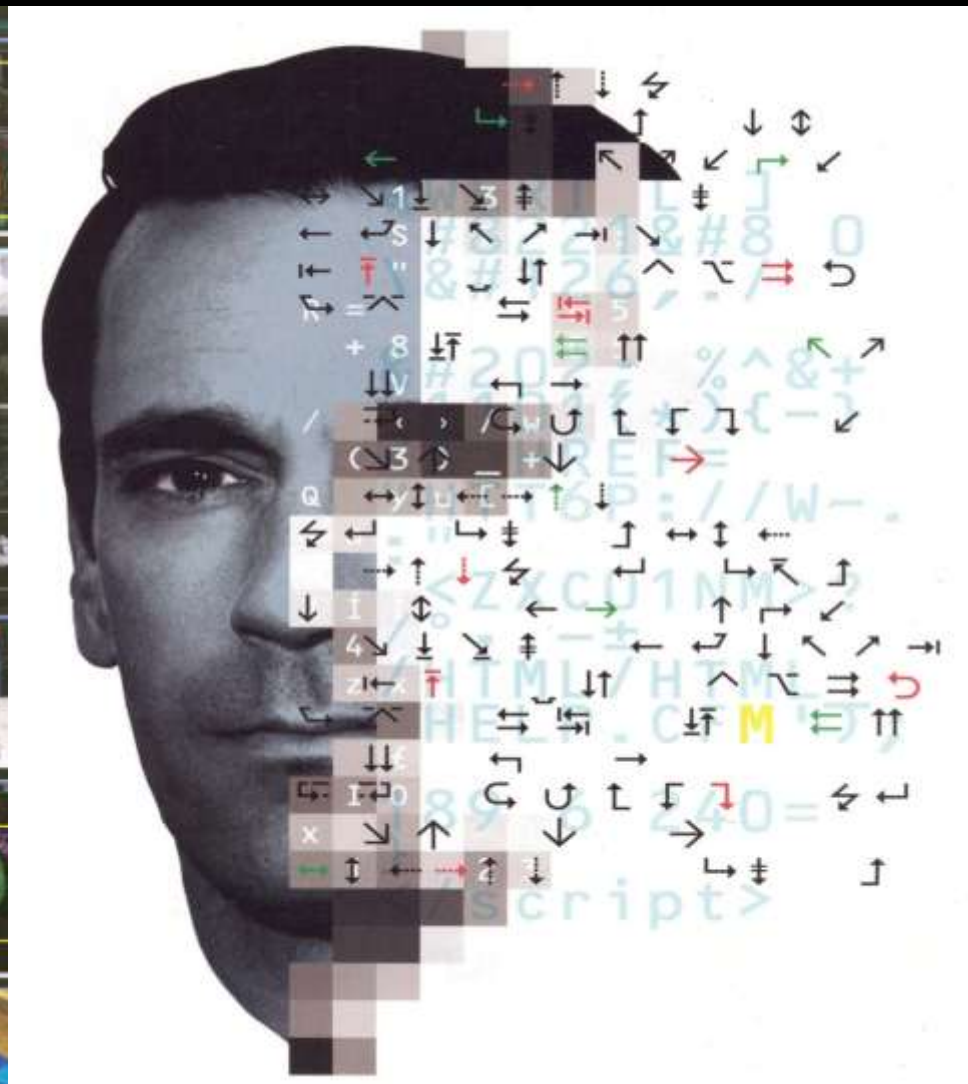








Technology Acceleration and Convergence: The Escalating Challenge for Professional Competency



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

If You Build It, Will They Pay?
Adoption of Disruptive Innovation

- new technology/service that simplifies a complex/costly problem
- business model that allows market adoption of the simplified solution at low(er) cost
- incentivized supply and demand to networks to reinforce the disruption

"If it isn't billable – it isn't going to happen"

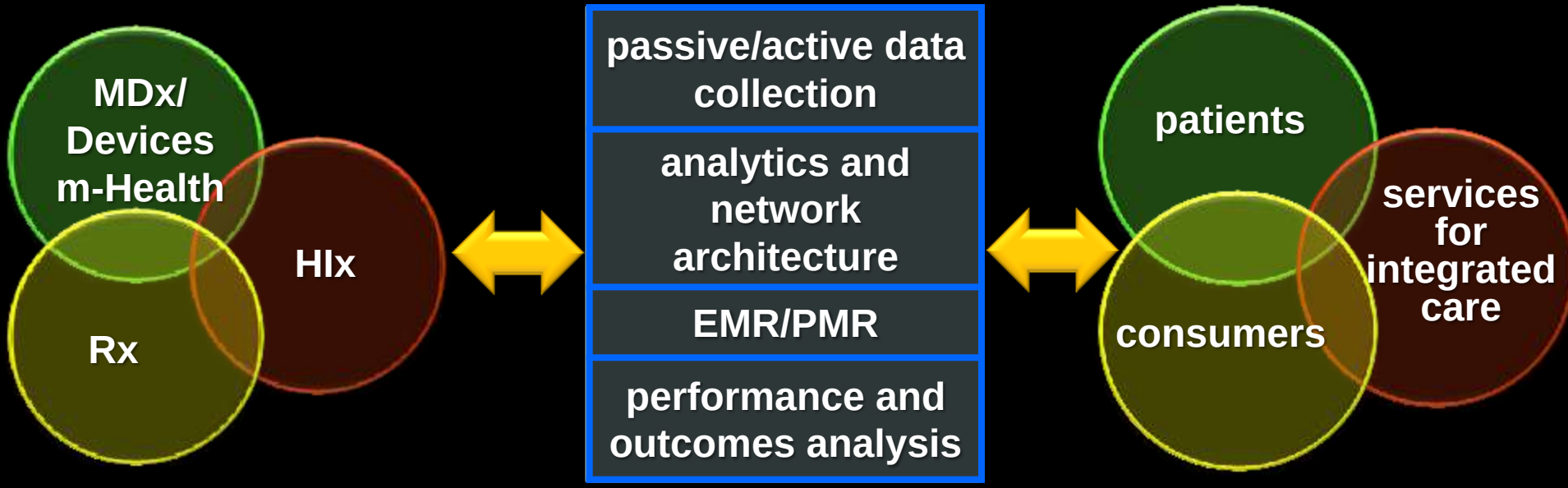
- value-based versus cost-based reimbursement
- new billing codes
- reimbursement for professional analysis of remote monitoring data streams

- rethink
- recalibrate
- redesign

What is required?

What is sustainable?

A New Healthcare Ecosystem Arising From Technology and Market Convergence



Integrated Technology Platforms for Comprehensive Profiling and Remote, Real Time Monitoring

Data Mining and Integration Services

Increasingly Targeted Care and Efficient Use of Finite Resources

Managing Massive Data and Driving New Value Propositions in Biomedical R&D and Healthcare Delivery

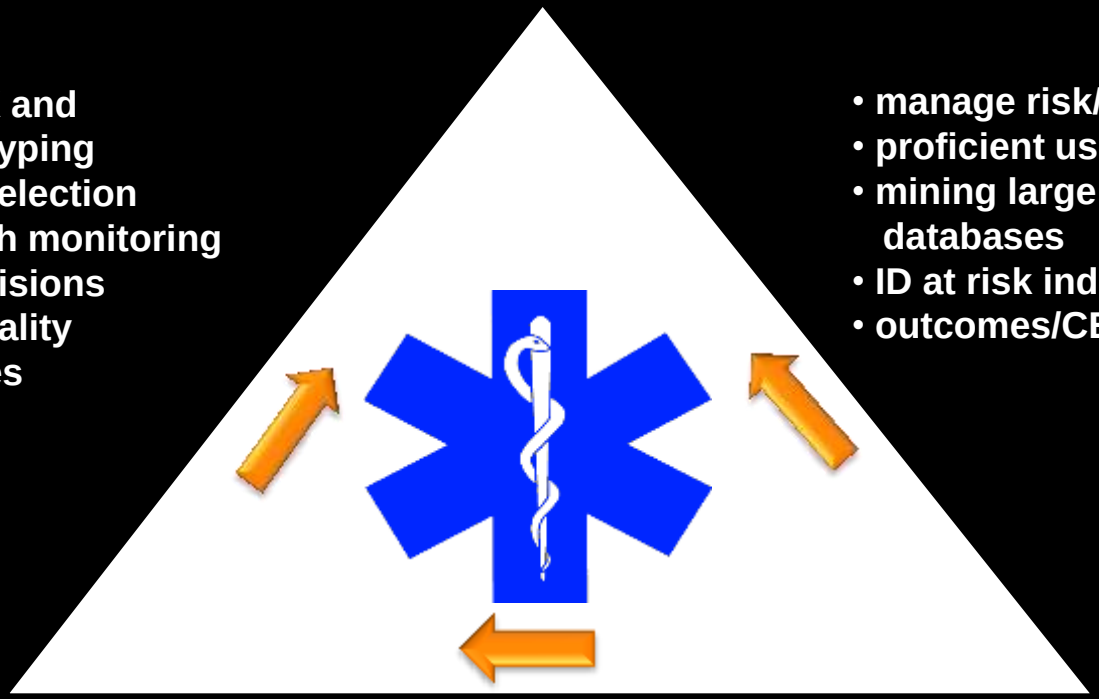
ACT

Individual

- precision DX and disease subtyping
- rational Rx selection
- remote health monitoring
- superior decisions
- improved quality and outcomes

H/Care System

- manage risk/cost
- proficient use of finite resources
- mining large population databases
- ID at risk individuals
- outcomes/CER



ANALYZE

- new services for data storage, mining, diagnostic algorithms

- molecular profiling (personalized medicine)
- global disease surveillance (public health)

PROFILE

- mapping dysregulation of biological networks in disease

Large Scale Molecular Profiling and Data Analytics as Foundational Technologies for Molecular Medicine

Cross-sector Convergence of Molecular Medicine, Digital Communication and Social Media Create Powerful Opportunities to Rethink, Recalibrate and Redesign Healthcare Delivery

New Value Propositions and Business Models for Identification and Mitigation of Risk Will Transition Healthcare Increasingly to a Consumer- and Payor-Centric Market Structure

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