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)
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
“Sorry For the Wait.
How Have You Been Since Your Last Visit?
Any stiffness?”
Integrative Omics (iOmics) and Personal Omics Profiling (the iPOP): Dynamic Tracking of Molecular and Medical Phenotypes

- 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

Cell (2012) 148, 1293
The Integrative Personal Omics File and Longitudinal Tracking of Health Status and Risk

Newborn Baseline Profile

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

- Profiling of molecular variants
- Evidentiary standards
- Clinical utility and adoption
- Value
- Data
Analytical and Clinical Validation of Molecular Determinants of Disease, Treatment Options and Predisposition Risk

- Reproducibility
- Validation
- Translation

- Robust geno-phenotypic correlations (iOmics)

- Profiling of molecular variants

- Clinical utility and adoption

- Evidentiary standards

- Data

- Value

- Relevance (disease severity)
- Treatability
- Workflow disruption
- MD/HCP education
- Longitudinal health monitoring and risk reduction

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

- Robust geno-phenotypic correlations
- Clinical utility and adoption
- Evidentiary standards
- Data
- Value

- Regulatory oversight
  - Pre- and post approval

- Reimbursement
- Coding
- Clinical guidelines
- Ethics
Data: The Foundation of Molecular Medicine, Risk Reduction and Optimizing Wellness

- Robust geno-phenotypic correlations
- Clinical utility and adoption
- Evidentiary standards
- Big data
- Health records and clinical-decision tools
- Automated analytics
- V3: volume, velocity, variety
- Standards
- Integration
- Curation
- Security
- 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

- Standardization
  - preanalytical
  - analytical
  - data formats

- Rigorous Subject Selection

- Pristine Biospecimens

- Study Design, Statistical Power and Analytics
## Lack of Standards and Shoddy Science: Increasingly Pervasive Problems in Academic Biomedical Research

<table>
<thead>
<tr>
<th>Poor Replication and Reproducibility</th>
<th>Slow Adoption of Standards</th>
<th>Failure of Academia to Work to Industry Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability of 'new drug target' claims called into question</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### The Small ‘N’ Problem

<table>
<thead>
<tr>
<th>Slow Adoption of Standards</th>
<th>Nature (2012) 485, 149</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inefficient Translation</td>
<td>Nature 5 April 2012</td>
</tr>
</tbody>
</table>

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

JAMA (2011) 305, 2200
# Genotyping Comprehensiveness and Tamoxifen Benefit by Exclusion of Metabolically Inactive CYP2D6 Isoforms in ER⁺ Breast Cancer

<table>
<thead>
<tr>
<th>Findings</th>
<th>Study</th>
<th>Number of SNPs in analysis</th>
<th>Alleles interrogated in analysis</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Okishiro et al. [28]</td>
<td>1</td>
<td>*10</td>
<td>*3 and *6 assayed but excluded</td>
</tr>
<tr>
<td></td>
<td>Toyama et al. [65]</td>
<td>1</td>
<td>*10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wegman et al. [59]</td>
<td>1</td>
<td>*4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stingl et al. [66]</td>
<td>1</td>
<td>*4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wegman et al. [12]</td>
<td>1</td>
<td>*4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nowell et al. [29]</td>
<td>1</td>
<td>*4</td>
<td>Quantitative bias analysis</td>
</tr>
<tr>
<td></td>
<td>Lash et al. [23]</td>
<td>1</td>
<td>*4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Park et al. [60]</td>
<td>3</td>
<td>*5, *10, *41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ramón et al. [67]</td>
<td>29</td>
<td>Roche AmpliChip™</td>
<td>Gene duplications included</td>
</tr>
<tr>
<td>Positive</td>
<td>Xu et al. [45]</td>
<td>1</td>
<td>*10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bijl et al. [30]</td>
<td>1</td>
<td>*4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thompson et al. [31]</td>
<td>29</td>
<td>Roche AmpliChip™</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: SNP, single nucleotide polymorphism.

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

- correlation and causality analytics
  - SNPs, haplotypes
  - CNVs
  - rearrangements
  - non-coding regions
  - ethnic diversity
  - epistasis
  - epigenetics
  - other ‘omics’

- decision analytics
  - Rx response/resistance
    - target(s), networks
  - Rx adverse event risk
  - prognosis/progression
  - predisposition to disease
  - environmental exposure/lifestyle confounders for predisposition

- privacy and security
  - technical standards
  - regulatory requirements
  - reimbursement
  - education
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)
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

**Cell-specific Molecular Interaction Networks**

**Perturbed Networks and Disease**

recognition of genome organizational and regulatory complexity
Genotype-Phenotype Relationships: Integrative Personal Omics Profiling

- Monogenic
- Multigenic
- Epistasis
- Epigenetics

Homeostasis (physiology)

Dysregulation (predisposition risk/overt pathology)

Phenotype (phenomes)

iPOP
complex multigenic traits

homeostasis
dysregulation

complex developmental disorders
unigenic Mendelian disorders
late onset ‘common diseases’
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% genome = 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
● 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

● role of environmental and epigenetic influences
The Relationship Between Allele Frequency and Effect Size and The Lack of Intermediate Frequency Associations

- **Effect Size**
  - Large
  - Modest
  - Small
  - (Very) Rare
  - Rare
  - Intermediate
  - Common

- **Allele Frequency**
  - <0.0001
  - 0.001
  - 0.02
  - 0.5

- **Linkage Studies**
  - common alleles with large effects unlikely to exist
  - negative selection of deleterious variants

- **WGS?**
  - complex traits
  - combinatorial effects of rare variants

- **GWAS**
  - intermediate frequency associations

- **GWAS**
  - genetic associations
  - common alleles
  - large effects unlikely to exist
  - negative selection of deleterious variants

- **WGS?**
  - complex traits
  - combinatorial effects of rare variants

- **GWAS**
  - genetic associations
  - common alleles
  - large effects unlikely to exist
  - negative selection of deleterious variants

- **WGS?**
  - complex traits
  - combinatorial effects of rare variants

- **GWAS**
  - genetic associations
  - common alleles
  - large effects unlikely to exist
  - negative selection of deleterious variants
Are Humans Still Evolving?
Human Genetic Diversity and Evolutionary History

- The paleo-populations
- Population and genetic bottlenecks

- Reduced historical population and genetic bottlenecks
- Increased migration and inter-breeding

- Distal (paleo-) ancestors

- Population expansion and progressive reduction of geographic isolation
  
  - Anthropogenic-relaxation of selection intensity (food, sanitation, Rx)
  - Natural selection pressures (famine, pestilence)

- Population and individual genetic diversity
Defining Genetic Variation in Human Populations: The Skewing of the Allele Frequency Spectrum Towards Rare and Private Genetic Variants

- Distal (paleo-) ancestors
- Common variants that segregate within large populations
- Rare-to-intermediate frequency variants arising in recent history (extended pedigree)
- Inherited rare variants arising in (G1, G^n) parental lineages
- De novo mutations
Defining Genetic Variation in Human Populations: The Skewing of the Allele Frequency Spectrum Towards Rare and Private Genetic Variants

- **Allele Frequency**
  - common
  - intermediate to rare
  - rare to private

- **Period of Negative Selection Pressure for Removal of Deleterious Alleles**
  - long
  - short (except extreme/lethal phenotypes)

- **Distal (paleo-) ancestors**
- **clan**
- **pedigree**
- **individual**

Clan: An individual whose rare genetic variation has been passed down through generations within a specific lineage or community.

Pedigree: An individual whose rare genetic variation has been passed down through generations within a specific family or broader social group.

Individual: A rare genetic variation that has been passed down from an individual to their offspring, or has arisen de novo in an individual.

Distal (paleo-) ancestors: A long period of negative selection pressure for removal of deleterious alleles, extending back to distant ancestors of the human species.
Mapping The Spectrum of Human Genetic Variation: The Under-Explored Epigenome and the Chromatin Landscape

- **Distal (paleo-) ancestors**
- **clan**
- **pedigree**
- **individual**

**epigenetics**
- 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
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

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
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.
- Consent vs. non-consent 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

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

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

<table>
<thead>
<tr>
<th>Target</th>
<th># Patients Screened</th>
<th># Eligible Patients</th>
<th># Centers</th>
<th># Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>EML4 ALK*: lung cancer*</td>
<td>1500</td>
<td>82</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>HER2*: gastric cancer**</td>
<td>3803</td>
<td>549</td>
<td>122</td>
<td>24</td>
</tr>
</tbody>
</table>

* 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 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
90% of Americans lack confidence in their clinicians' ability to understand and use genetic information


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
“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)
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, geography, diet)
- dysbiosis and complex disease: cause or consequence?
  - characterization of role of emergent pathobionts (dysbiosis) in systemic disease
  - allergy/asthma, diabesity, metabolic disease, chronic 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)

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

Health Status

Healthy/Low Risk
At-Risk
High Risk

20% of the Population Generate 80% Cost

multiple co-morbidities
end-of-life care
chronic disease progression
chronic disease early stage
acute disease

Value
Cost
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
mHealth

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

Now Comes the Hardest Part of All!

Driving iOmics and Molecular Medicine and IT-Centric Capabilities Into Routine Clinical Practice
"We’re ready to begin the next phase of keeping things exactly the way they are."
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 Data- and Computation-Intensive Processes

- 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

Pending Era
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

- research and discovery
- translation and clinical trials
- healthcare delivery
- m.health
- consumers
- payors
- outcomes analytics
- decision support tools
- patients
- regulators
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’

- clinical annotation
- EHR data mining

Observed Phenotype

- clinical annotation
- iPOPs

Interpreted Phenotype and Phenomes

- 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

- IBM’s Watson
  - 2880 CPUs
  - natural language questions

- prelude to Q&A systems for biomedicine beyond keyword IR searches

Jeopardy 16 February 2011
The Pending Zettabyte Era

1,000,000,000,000,000,000,000,000
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
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
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
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
will move beyond the scope of the individual research infrastructures...

...and construct the technology and tools needed to connect them.
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’
What is required?
What is sustainable?

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

- Molecular diagnostics
- Health status profiling
- Risk management
- Real-time data
- Integrated care and continuity of care
- Cost control and proficient resource utilization
- Optimized decisions

Integrative Omics (iOmics) and Integrative Personal Omics Profiles (iPOPs) form the core foundational elements for molecular medicine, integrating advanced diagnostic technologies with personalized health profiling to enhance cost control, resource utilization, decision-making, and overall health outcomes.
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
iOomics and Computational Biomedicine as Disruptive Technologies for Radical Change in Life Sciences Research and Healthcare

- 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