ASU Complex Adaptive Systems Initiative
Arizona State University

NBDA | National Biomarker Development Alliance

Mapping Complex Biological Information Networks to Develop Robust Biomarkers and Define Phenotypes

Dr. George Poste
Chief Scientist, Complex Adaptive Systems Initiative
and Regents Professor of Health Innovation
Arizona State University
george.poste@asu.edu
www.casi.asu.edu

Presentation at NBDA Workshop:
Big Data, Artificial Intelligence and Biomarkers
Phoenix, Arizona
21 May 2018

Biomarkers

- the core conceptual and functional components in understanding disease risk, pathogenesis and making precision medicine a reality

- molecular, histologic, radiographic or physiologic characteristics
- new classes of digital markers and complex algorithms
- “not an assessment of how a patient feels, functions or survives” (???)

Biomarkers:
Representation(s) of Biological 'System States' at Different Hierarchical Levels of Biological Order

spatial

- (epi)genome to organism

temporal

- embryogenesis to end-of-life

dynamic

- reconfiguration of biological (information) processes across the health to disease continuum
- robustness, redundancy, adaptation, emergence, evolution

information

- encoded/encrypted determinants of the architecture and dynamics of molecular signaling networks and structural assemblies
- content, compartmentalization, context, and concurrency
- pleiotropy

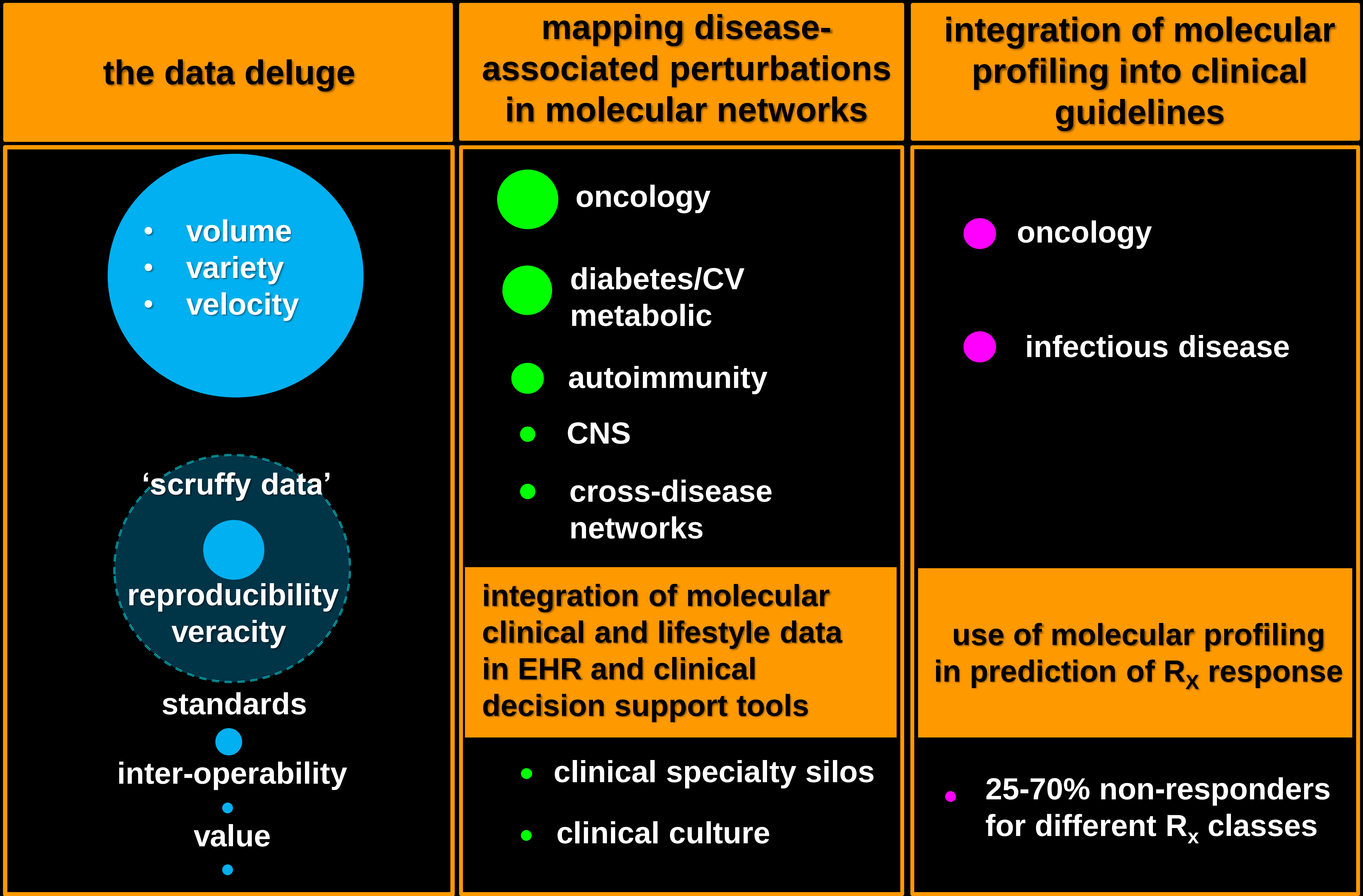
The Challenge (Complexity) of Biomarker Discovery, Validation and Clinical Adoption

- solutions
  - asking the right question
  - intended use
  - fit for purpose
- standards
  - too much sloppy science (poor reproducibility)
  - publish and vanish
  - processes, platforms
  - data, databases, algorithms
- seduction
  - embrace of each new “shiny object” technology platform in isolation from the problem or the prospect of clinical adoption

The Challenge (Complexity) of Biomarker Discovery, Validation and Clinical Adoption

- specimens
  - stringent QA/QC
  - biobanks and pre-analytical methods
- segmentation (stratification) of profiled cohorts
  - new clinical trial designs, including RWE
  - deep phenotyping: integration of multi-Omics, clinical, lifestyle and environmental exposure datasets
- scale, scalability and sharing
  - cohort sizes for statistical robustness
  - big data
  - cross-disciplinary integration
  - multi-institutional collaboration/consortia
  - infrastructure, integration logistics, cost

The Status of Biomarker Data Generation and Use In Biomedical Research and Clinical Care



the data deluge

- volume
- variety
- velocity

‘scruffy data’

reproducibility
veracity

standards

inter-operability

value

mapping disease-associated perturbations in molecular networks

- oncology
- diabetes/CV metabolic
- autoimmunity
- CNS
- cross-disease networks

integration of molecular clinical and lifestyle data in EHR and clinical decision support tools

- clinical specialty silos
- clinical culture

integration of molecular profiling into clinical guidelines

- oncology
- infectious disease

use of molecular profiling in prediction of R_x response

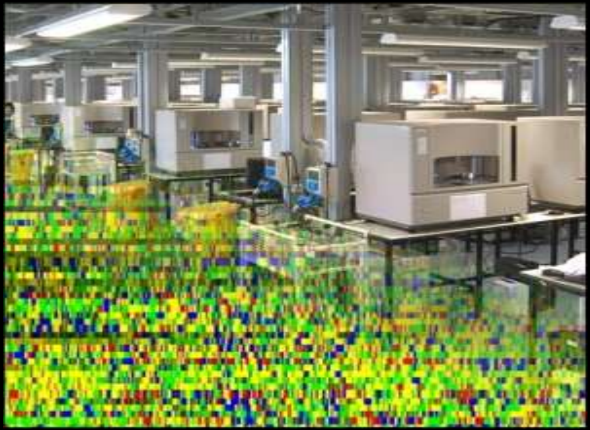
- 25-70% non-responders for different R_x classes

Silos Subvert Solutions - The History of Biomarker R&D

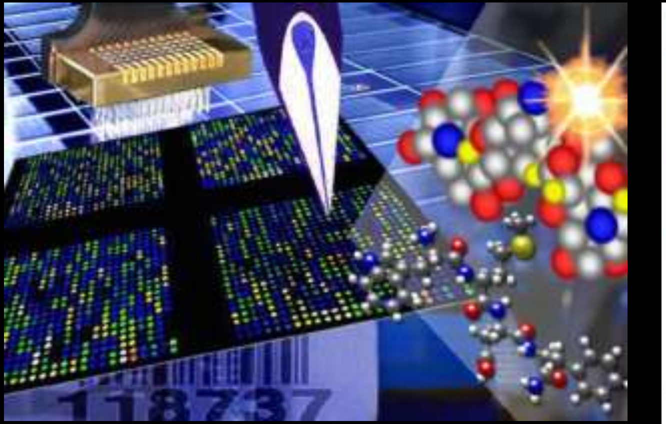
The Imperative for Systems-Based Approaches

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

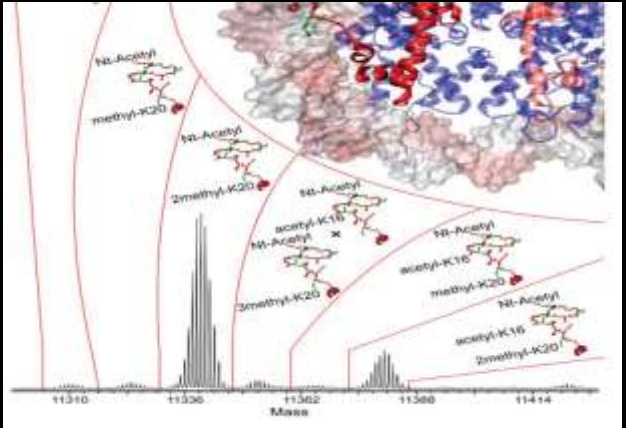
(Epi)Genomics

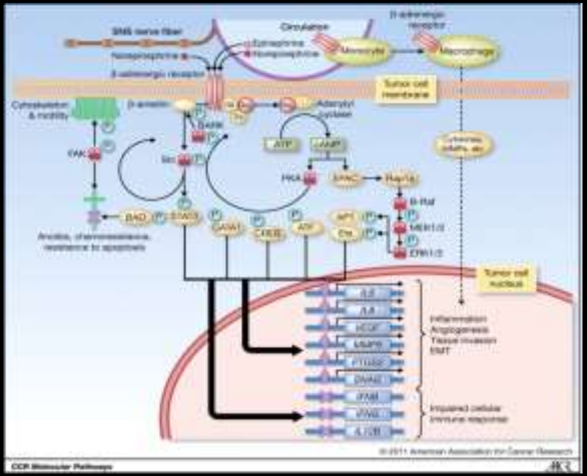


Transcriptomics

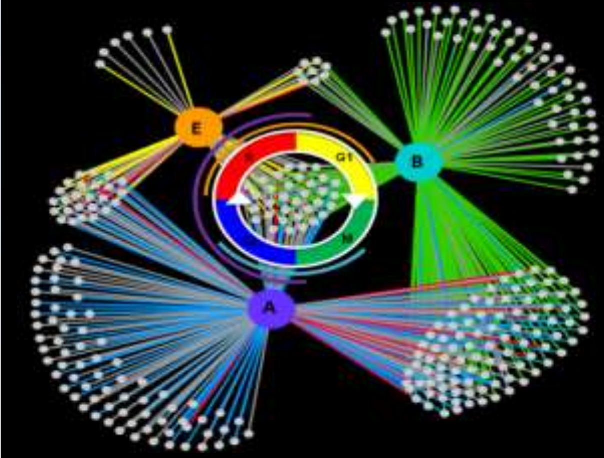


Proteomics

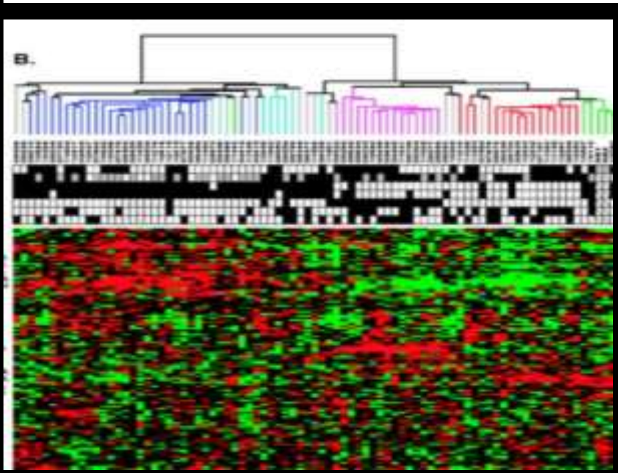




Molecular Pathway Analysis



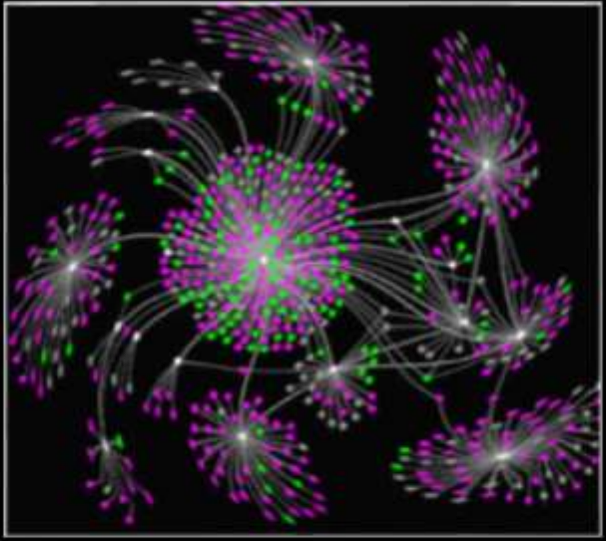
Network Topology and Architecture



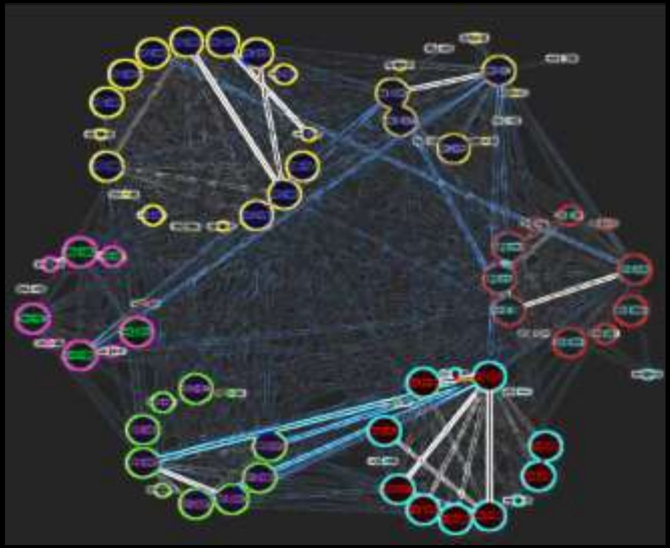
Network Perturbation(s) and Disease Subtypes

Precision Medicine

Causal Relationships Between Disruption of Molecular Signaling Networks and Disease



Disease Comorbidity Patterns



Patient-Specific Signatures of Disease or Predisposition to Disease

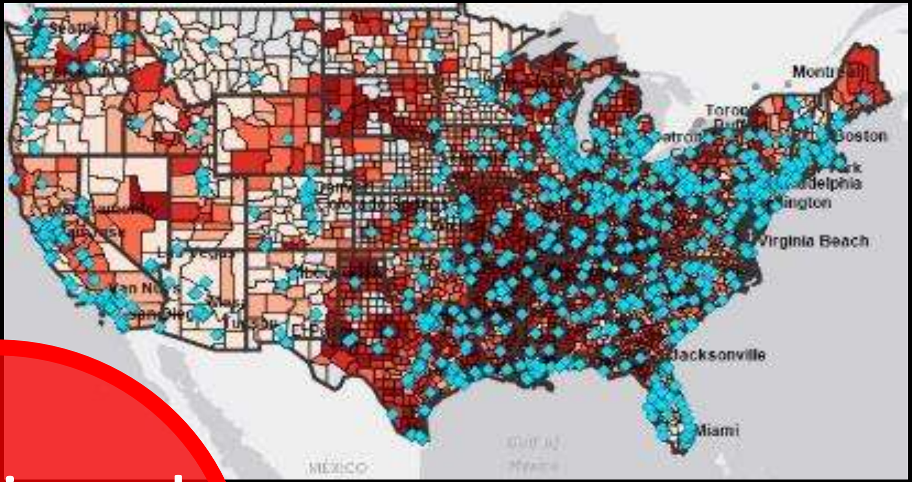
- terabytes per individual
- zettabyte – yottabyte population databases

Big (Messy) Data

Precision Medicine: “Computed Phenotypes” and ‘Digital Siblings’

Individual Data

Population Databanks



integration and analysis of large scale, diverse data categories

“matching” individuals to ‘best match’ cohorts using data on similarities of deep phenotyping profiles and treatment outcomes

JAMA (2018) 319, 1981

VIEWPOINT

In the Era of Precision Medicine and Big Data, Who Is Normal?

Arjun K. Manrai, PhD
Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts.

Chirag J. Patel, PhD
Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts.

John P. A. Ioannidis, MD, DSc
Stanford Prevention Research Center, Stanford University, Stanford, California; and Department of Medicine, Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California.

Supplemental content

The definition of "normal" values for common laboratory tests often governs the diagnosis, treatment, and overall management of tested individuals. Some test results may depend on demographic traits of the tested population including age, race, and sex. Ideally, laboratory test results should be interpreted in reference to a population of "similar" "healthy" individuals. In many settings, however, it is unclear exactly who these individuals are. How much population stratification and what criteria for healthy individuals are optimal? In particular, with the evolution of medicine into fully personalized or "precision" medicine and the availability of large-scale data sets, there may be interest in trying to match each person to an increasingly granular normal reference population. Is this precision feasible to obtain in reliable ways and will it improve practice?

There are limited systematic analyses of baseline variation across demographically diverse population strata (including race/ancestry, gender/sex, age, and socioeconomic strata of the population) for even widespread clinical laboratory tests. Even after decades of routine use, it may be that reference standards should

strumentation? How can multiplicity across myriad population strata be overcome as the normal population becomes more precise and personalized?

It is essential to answer these questions for widely used clinical laboratory tests such as complete blood cell count and blood chemistries before delving into more rare tests. Such tests are a routine entry point for invasive and expensive follow-up tests and procedures, yet remain poorly characterized across strata. Data sets sufficiently capacious to study stratified variation at scale include select research cohorts, electronic health records, and insurance claims data sets. Although some data sets may be queried with relative ease (eg, electronic health records at an investigator's institution or public claims data), how generalizable findings are to other clinical settings is unclear.[4]

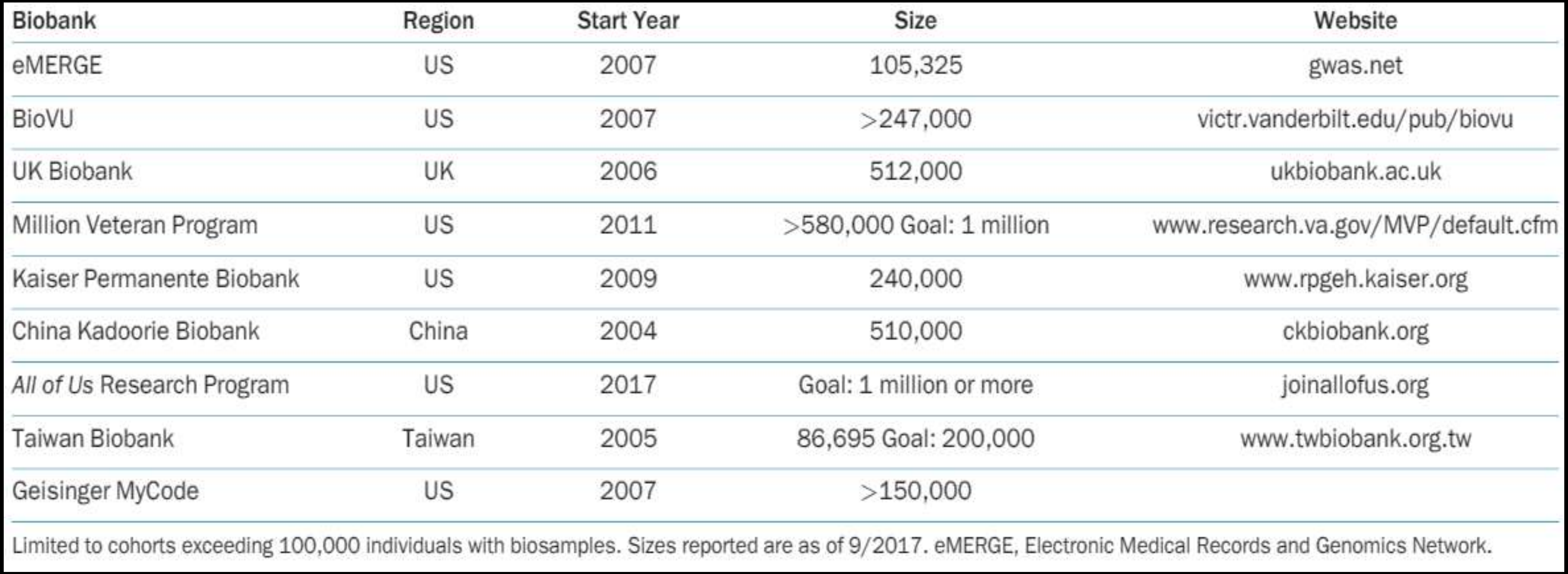
Challenges of Precision Medicine and Big Data

Defining Normality

The first challenge to ensuring precise application of clinical laboratory testing is defining a "healthy" population to estimate the normal range of variation across population strata. A set of criteria for normality (eg, absence of chronic disease) may appear reasonable but substantial differences can result from 2 sets of equally reasonable criteria. More specifically, the Clinical and Laboratory Standards Institute (CLSI) guidelines state that 120 "reference individuals" should be used to establish reference intervals for labora-

> However, with the proliferation of large data sets emblematic of precision medicine, it is becoming feasible to study stratified variation and clinical outcomes at scale.

Large Population Cohorts for Molecular Profiling



Biobank	Region	Start Year	Size	Website
eMERGE	US	2007	105,325	gwas.net
BioVU	US	2007	>247,000	victr.vanderbilt.edu/pub/biovu
UK Biobank	UK	2006	512,000	ukbiobank.ac.uk
Million Veteran Program	US	2011	>580,000 Goal: 1 million	www.research.va.gov/MVP/default.cfm
Kaiser Permanente Biobank	US	2009	240,000	www.rpgeh.kaiser.org
China Kadoorie Biobank	China	2004	510,000	ckbiobank.org
All of Us Research Program	US	2017	Goal: 1 million or more	joinallofus.org
Taiwan Biobank	Taiwan	2005	86,695 Goal: 200,000	www.twbiobank.org.tw
Geisinger MyCode	US	2007	>150,000	

Limited to cohorts exceeding 100,000 individuals with biosamples. Sizes reported are as of 9/2017. eMERGE, Electronic Medical Records and Genomics Network.

Adapted from: J.C. Denny et al. (2018) Clin. Pharm Therap. 103, 409

Large Scale Genome Sequencing Projects:
The Dangers of Reductionism and Ignoring Biological Complexity

NIH National Institutes of Health
All of Us Research Program

All of Us


Biden Focuses on Goals at Wrap-Up of the Cancer Moonshot
from The Cancer Letter

precisionFDA

U.S. Department of Veterans Affairs

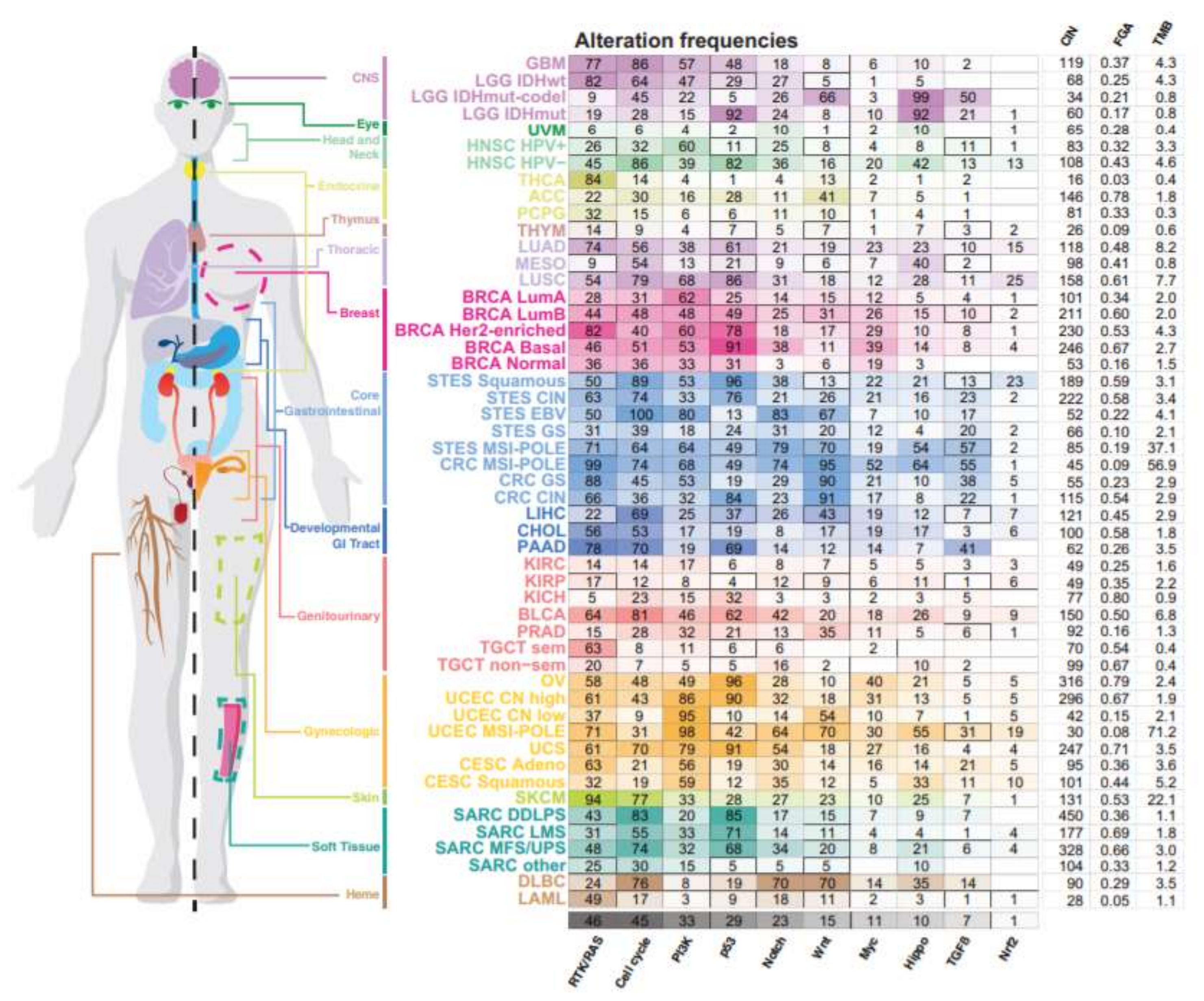
华大基因 BGI

Genomics england

GENOME denmark

GRAIL

Fraction of Tumor Samples with Alterations in 10 Curated Signaling Pathways



Alteration frequencies

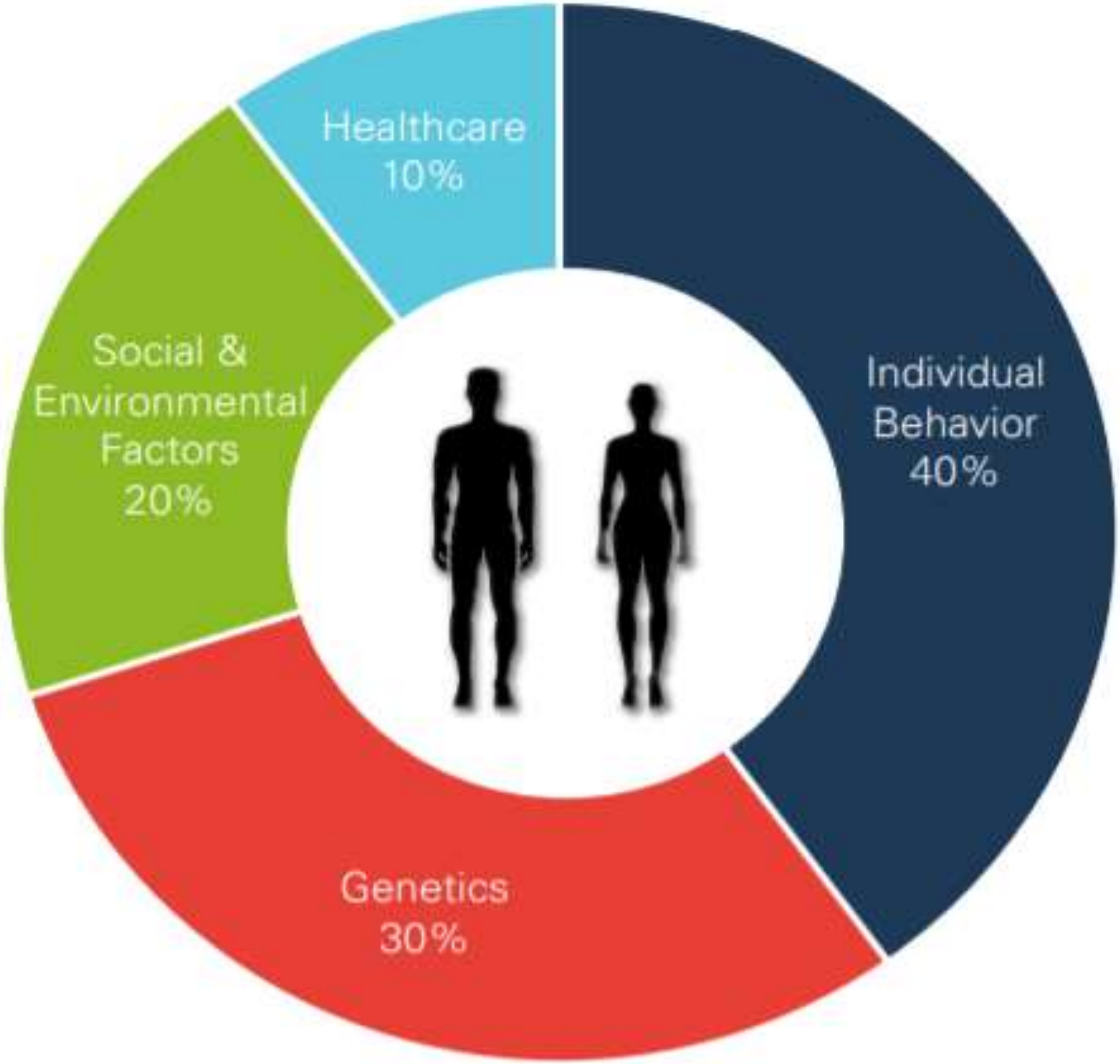
	RTK/RAS	Cell cycle	PI3K	p53	Notch	Wnt	Myc	Hippo	TGFB	Nrf2	CIN	FGA	TMB
GBM	77	86	57	48	18	8	6	10	2		119	0.37	4.3
LGG IDHwt	82	64	47	29	27	5	1	5			68	0.25	4.3
LGG IDHmut-codel	9	45	22	5	26	66	3	99	50		34	0.21	0.8
LGG IDHmut	19	28	15	92	24	8	10	92	21	1	60	0.17	0.8
UVM	6	6	4	2	10	1	2	10		1	65	0.28	0.4
HNSC HPV+	26	32	60	11	25	8	4	8	11	1	83	0.32	3.3
HNSC HPV−	45	86	39	82	36	16	20	42	13	13	108	0.43	4.6
THCA	84	14	4	1	4	13	2	1	2		16	0.03	0.4
ACC	22	30	16	28	11	41	7	5	1		146	0.78	1.8
PCPG	32	15	6	6	11	10	1	4	1		81	0.33	0.3
THYM	14	9	4	7	5	7	1	7	3	2	26	0.09	0.6
LUAD	74	56	38	61	21	19	23	23	10	15	118	0.48	8.2
MESO	9	54	13	21	9	6	7	40	2		98	0.41	0.8
LUSC	54	79	68	86	31	18	12	28	11	25	158	0.61	7.7
BRCA LumA	28	31	62	25	14	15	12	5	4	1	101	0.34	2.0
BRCA LumB	44	48	48	49	25	31	26	15	10	2	211	0.60	2.0
BRCA Her2-enriched	82	40	60	78	18	17	29	10	8	1	230	0.53	4.3
BRCA Basal	46	51	53	91	38	11	39	14	8	4	246	0.67	2.7
BRCA Normal	36	36	33	31	3	6	19	3			53	0.16	1.5
STES Squamous	50	89	53	96	38	13	22	21	13	23	189	0.59	3.1
STES CIN	63	74	33	76	21	26	21	16	23	2	222	0.58	3.4
STES EBV	50	100	80	13	83	67	7	10	17		52	0.22	4.1
STES GS	31	39	18	24	31	20	12	4	20	2	66	0.10	2.1
STES MSI-POLE	71	64	64	49	79	70	19	54	57	2	85	0.19	37.1
CRC MSI-POLE	99	74	68	49	74	95	52	64	55	1	45	0.09	56.9
CRC GS	88	45	53	19	29	90	21	10	38	5	55	0.23	2.9
CRC CIN	66	36	32	84	23	91	17	8	22	1	115	0.54	2.9
LIHC	22	69	25	37	26	43	19	12	7	7	121	0.45	2.9
CHOL	56	53	17	19	8	17	19	17	3	6	100	0.58	1.8
PAAD	78	70	19	69	14	12	14	7	41		62	0.26	3.5
KIRC	14	14	17	6	8	7	5	5	3	3	49	0.25	1.6
KIRP	17	12	8	4	12	9	6	11	1	6	49	0.35	2.2
KICH	5	23	15	32	3	3	2	3	5		77	0.80	0.9
BLCA	64	81	46	62	42	20	18	26	9	9	150	0.50	6.8
PRAD	15	28	32	21	13	35	11	5	6	1	92	0.16	1.3
TGCT sem	63	8	11	6	6		2				70	0.54	0.4
TGCT non−sem	20	7	5	5	16	2		10	2		99	0.67	0.4
OV	58	48	49	96	28	10	40	21	5	5	316	0.79	2.4
UCEC CN high	61	43	86	90	32	18	31	13	5	5	296	0.67	1.9
UCEC CN low	37	9	95	10	14	54	10	7	1	5	42	0.15	2.1
UCEC MSI-POLE	71	31	98	42	64	70	30	55	31	19	30	0.08	71.2
UCS	61	70	79	91	54	18	27	16	4	4	247	0.71	3.5
CESC Adeno	63	21	56	19	30	14	16	14	21	5	95	0.36	3.6
CESC Squamous	32	19	59	12	35	12	5	33	11	10	101	0.44	5.2
SKCM	94	77	33	28	27	23	10	25	7	1	131	0.53	22.1
SARC DDLPS	43	83	20	85	17	15	7	9	7		450	0.36	1.1
SARC LMS	31	55	33	71	14	11	4	4	1	4	177	0.69	1.8
SARC MFS/UPS	48	74	32	68	34	20	8	21	6	4	328	0.66	3.0
SARC other	25	30	15	5	5	5		10			104	0.33	1.2
DLBC	24	76	8	19	70	70	14	35	14		90	0.29	3.5
LAML	49	17	3	9	18	11	2	3	1	1	28	0.05	1.1
	46	45	33	29	23	15	11	10	7	1			

From: F. Sanchez-Vega et al. (2018) Cell, 173, 321-337 the myopic, reductionist uni-dimensional focus on (epi)genome sequencing

necessary but not sufficient

it’s the phenotype (phenomes), stupid!

Impact of Different Factors On Disease and Premature Death



Source: Beyond Health Care: The Role of Social Determinants in Promoting Health and Health Equity. Kaiser Family Foundation, 2015.

Most Events That Affect Our Health Occur Outside of the Healthcare System And Are Not Monitored

Need for Continuity of Care Record: From Womb to Tomb

Behavior

Environment

Consortium for Exome Sequencing of 500,000 UK Biobank Samples by 2020 (Launched Jan. 2018)

REGENERON

biobank uk

abbvie

AstraZeneca

Alnylam PHARMACEUTICALS

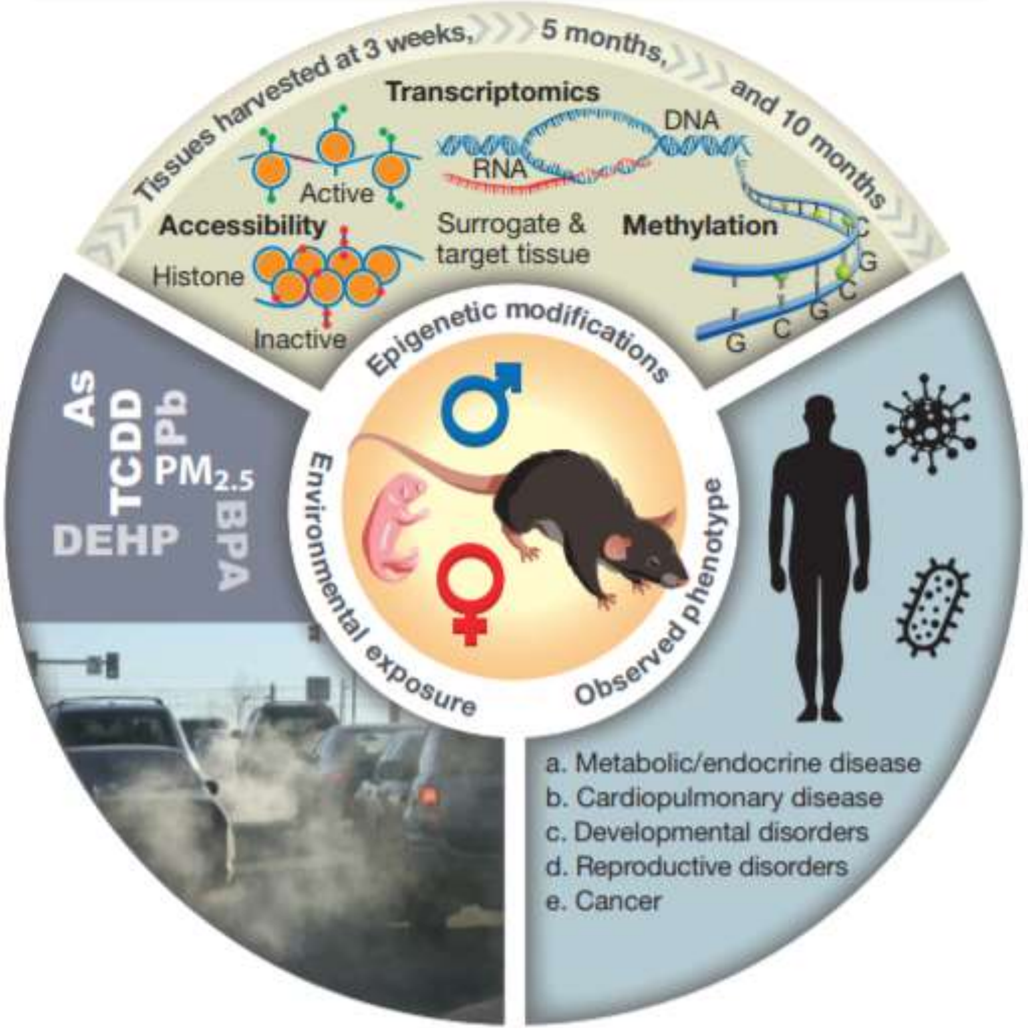
Biogen

Pfizer

IBM

- integration with medical records, lab test data and psychological assessments

The NIEHS Toxicant Exposures by Genomic and Epigenomic Regulators of Transcription (TaRGET) Consortium



T. Wang et al. (2018) Nature Biotechnology 36, 226

"People Analytics"
Social Activities and Behavior 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 reveal increasingly predictable behavior, individual risk patterns and health events
- the confessional of social media
- the blurring of private and public spaces
- complex ethical and legal issues
  - consent, privacy, security, surveillance

Major Investments in Digital Health by Major Corporations From Within and Outside of Traditional Healthcare Services

amazon

Google verily

IBM®

facebook

Alibaba.com

Tencent 腾讯

Pfizer

Johnson & Johnson

gsk GlaxoSmithKline

NOVARTIS

Takeda

AstraZeneca

Roche

MERCK

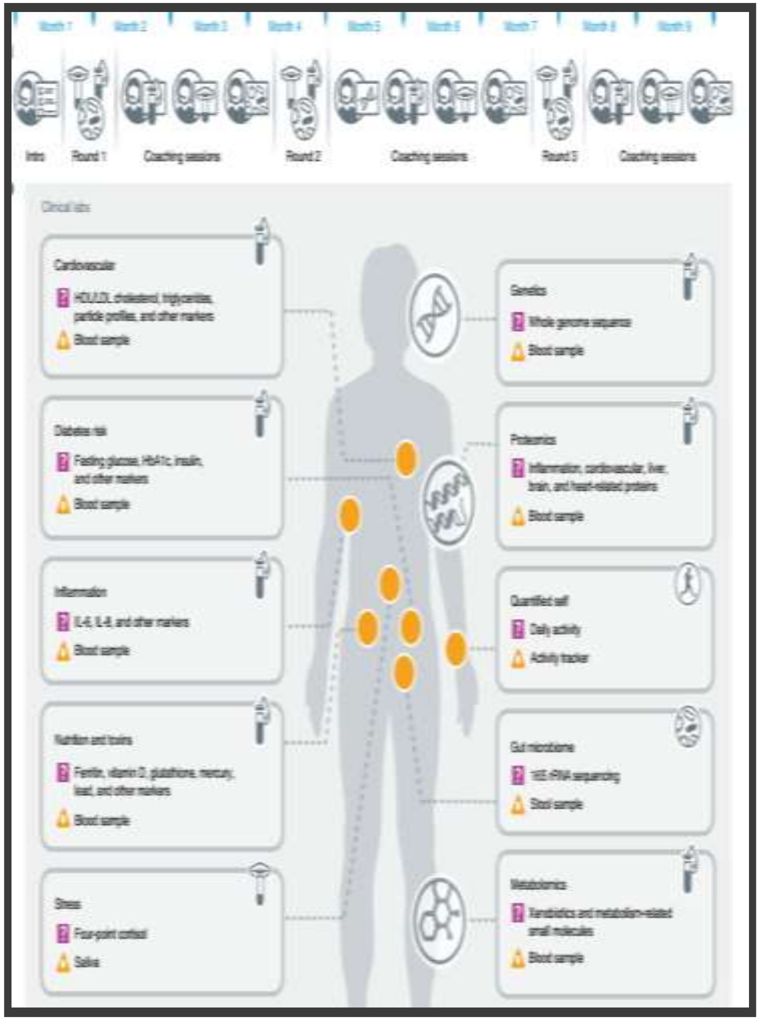
Digiceuticals: Software as Therapy

"We envision empowering individuals
with digital therapeutic solutions
that address underlying motivational and technical deficits
by deciphering neural pathways
that support motivation, decision-making and reinforcement
to prompt health."

Dr. Ben Wiegand
Global Head, Janssen R&D
World Without Disease Accelerator
PharmaVoice 2017

Personal, Dense, Dynamic Data Clouds: Comprehensive Profiling of Health Status of 108 Individuals Over 9 Months

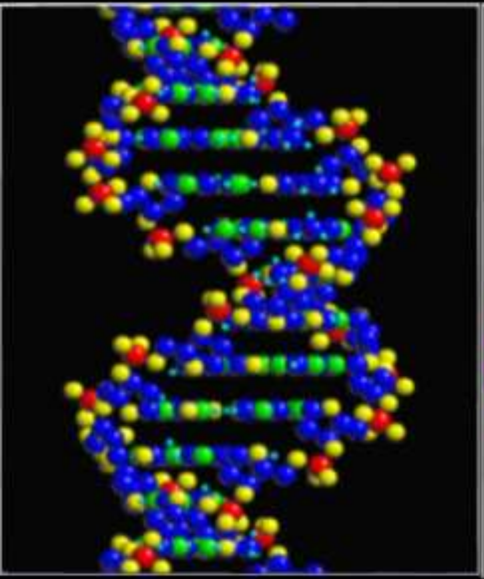
N. D. Price et al. (2017) Nature Biotechnology 35, 747



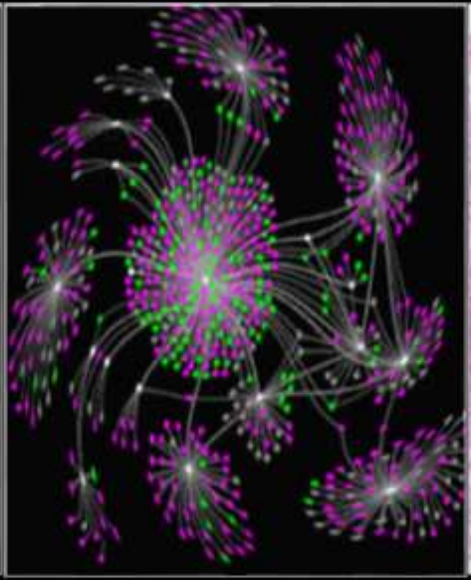
- WGS
- daily physical and sleep activities
- 3 month blood, saliva, urine and stool analysis
- 643 metabolites
- 262 proteins

- cost
- scalability
- data interpretation
- clinical utility

Precision Medicine: Understanding the Disruption of Molecular Information Networks in Disease

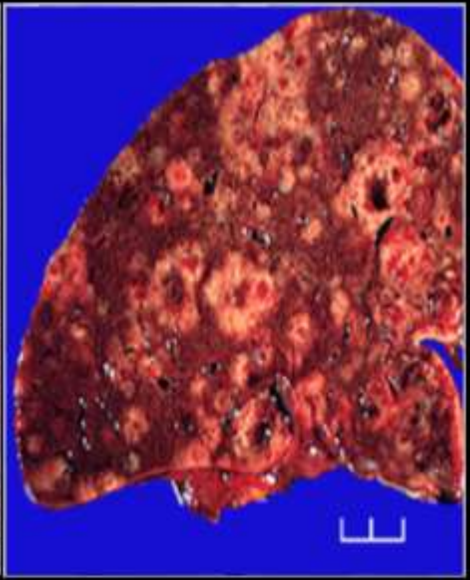


encoded information and expression as cell-specific signaling networks



patterns of information flow within signaling networks (network topology)

stable networks and information fidelity (health)



dysregulated networks and altered information patterns (disease)

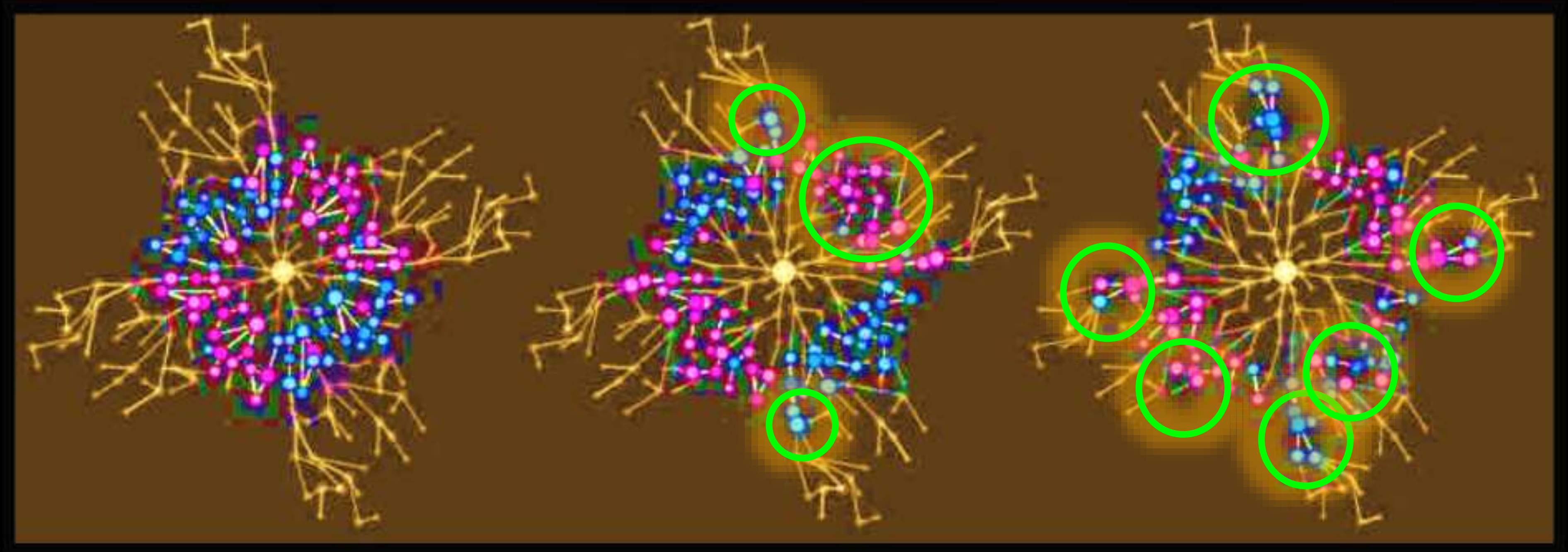
Large Scale Profiling of Molecular Biomarkers and the Design Architecture and Dynamics of Molecular Signaling (Information) Networks

‘ground truth’ source data for the deconvolution of biological “system states” in the health to disease continuum

major knowledge gap and impediment to making precision medicine a reality

new conceptual models, analytical platforms and computational tools to map non-deterministic, non-linear networks

Understanding System State Shifts (Phenomes) and Emergent Perturbations in Molecular Signaling Networks in the Health to Disease Continuum



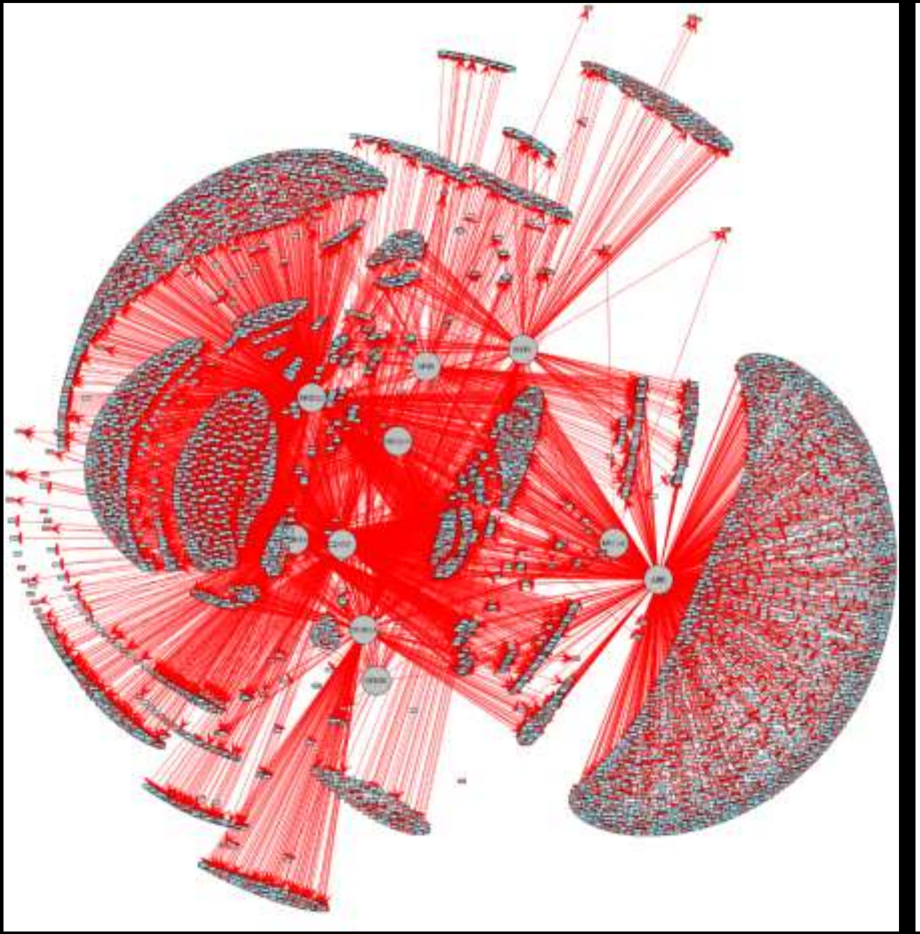
$T_{1(n)}$ $T_{2(n)}$ $T_{3(n)}$

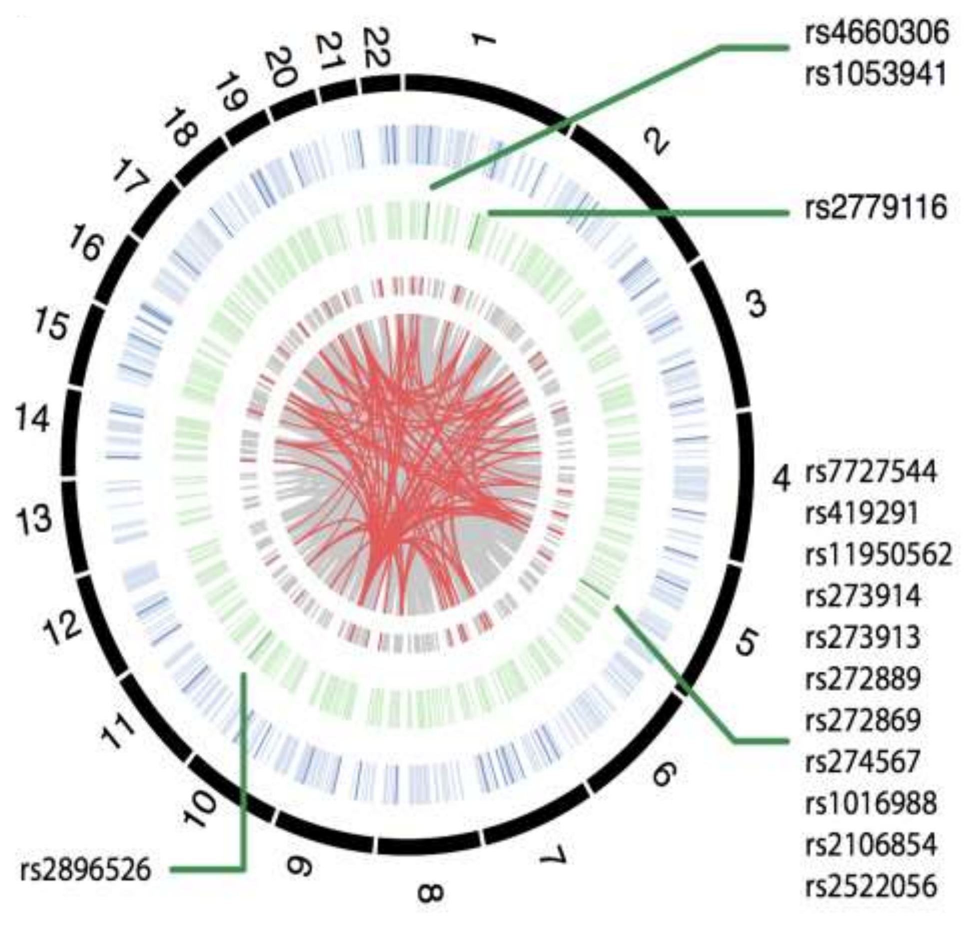
health

subclinical disease

overt disease

Integrative Gene Expression Network Models and Classification of Functional Modules (Communities) That Span Multiple Chromosomes





Courtesy of Dr. J. Quackenbush, Dana Farber Cancer Center

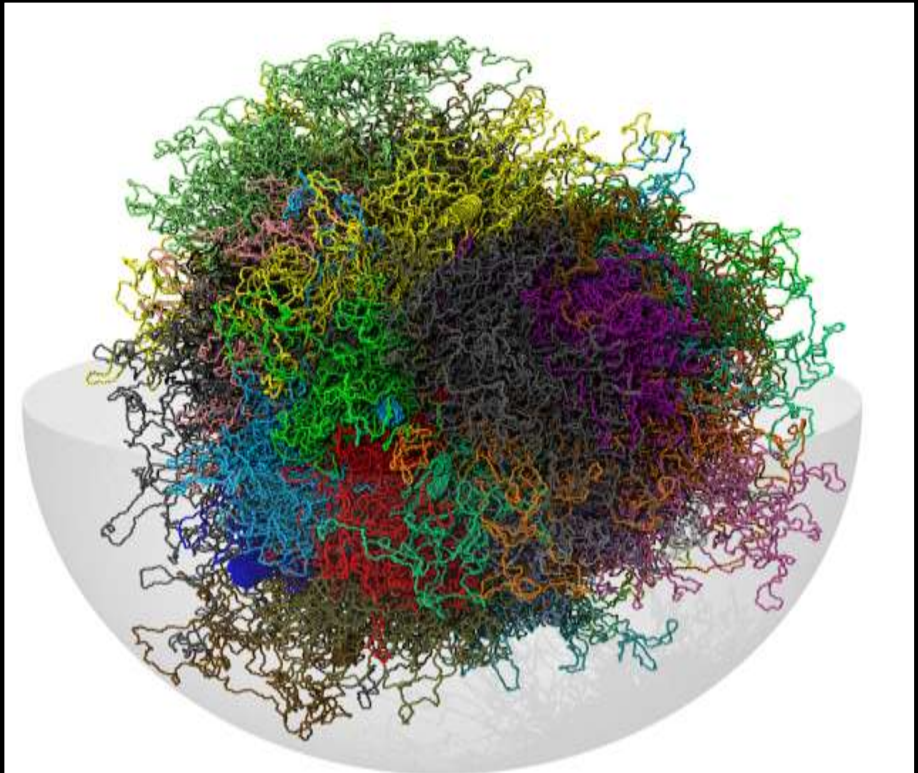
Characterization of Gene Interaction Networks and Identification of New Rx Targets for Genome-Dependent Vulnerabilities

- loss-of-function markers and essentiality mapping/synthetic lethals
- predicted shared biological functions via ‘guilt-by–association’ pulldown of protein complexes
- large scale pair-pair knockouts in cell lines
  - RNAi
  - CRISPR/Cas 9 screens (in vitro and in vivo)

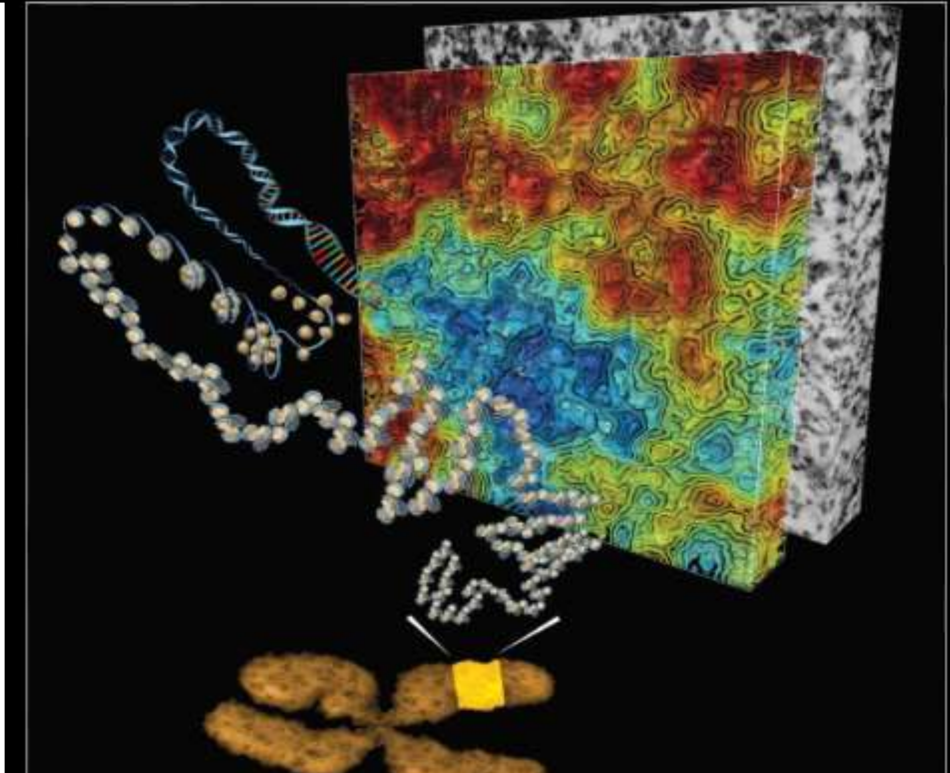
Defining Short- and Long-Range Cis- and Trans- Regulation of Gene Networks

Chromosomal Neighborhoods: Understanding the 3-D and 4-D Genome

ChromEMT Mapping of Chromatin Ultrastructure and DNA Packing



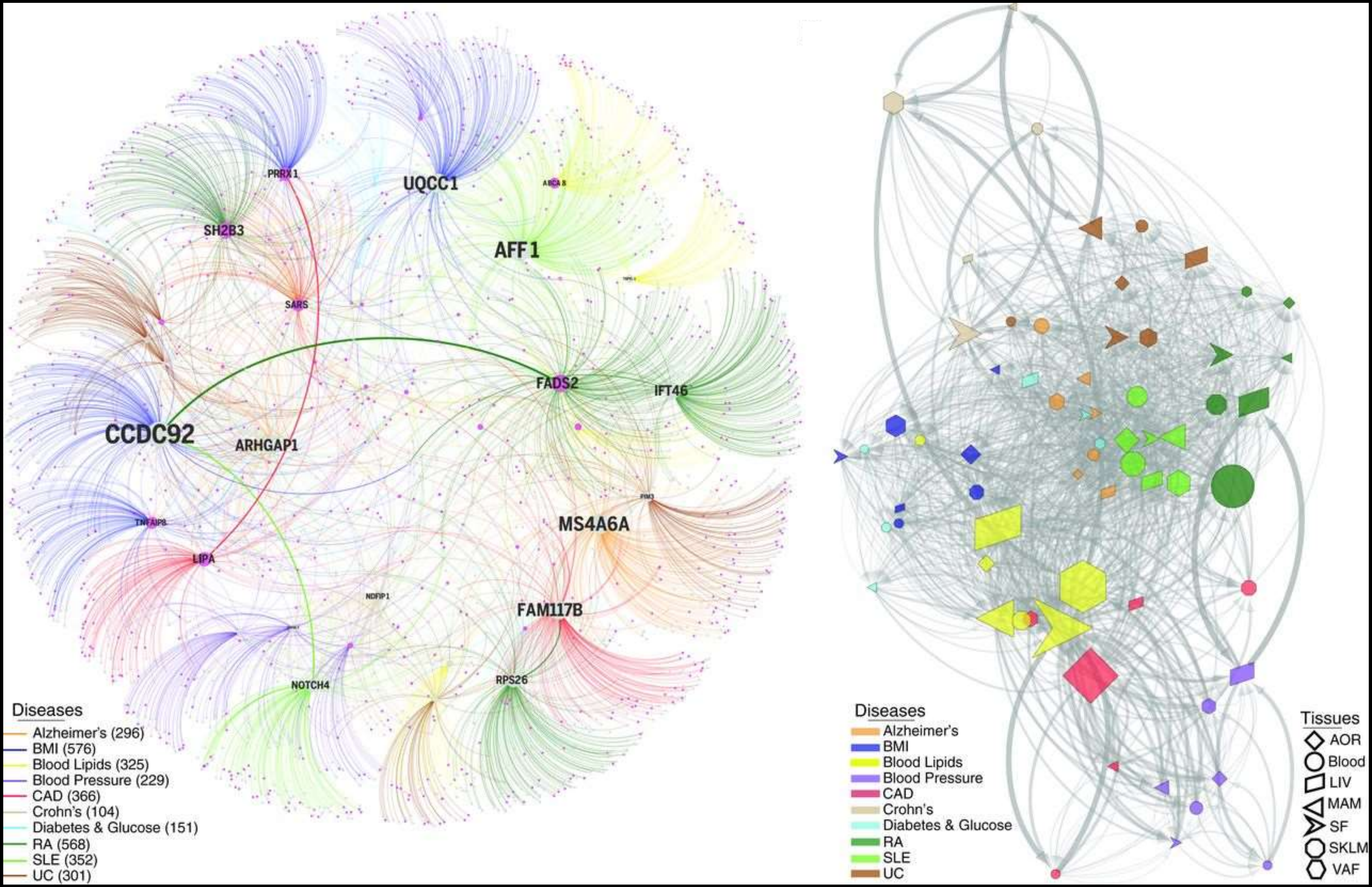
From: International School of Advanced Studies (SISSA) [October 26, 2016]



From: H. D. Ou et al. (2017) Science eaaag.0025

- **spatial and temporal regulation of topological association domains (TADs)**
- **intra and inter-chromosomal cis- and trans- juxtaposition of TFs, promoters and enhancers**

Cis- and Trans-gene Regulation Across Risk Loci for Cardiometabolic Disease and Alzheimer’s Disease

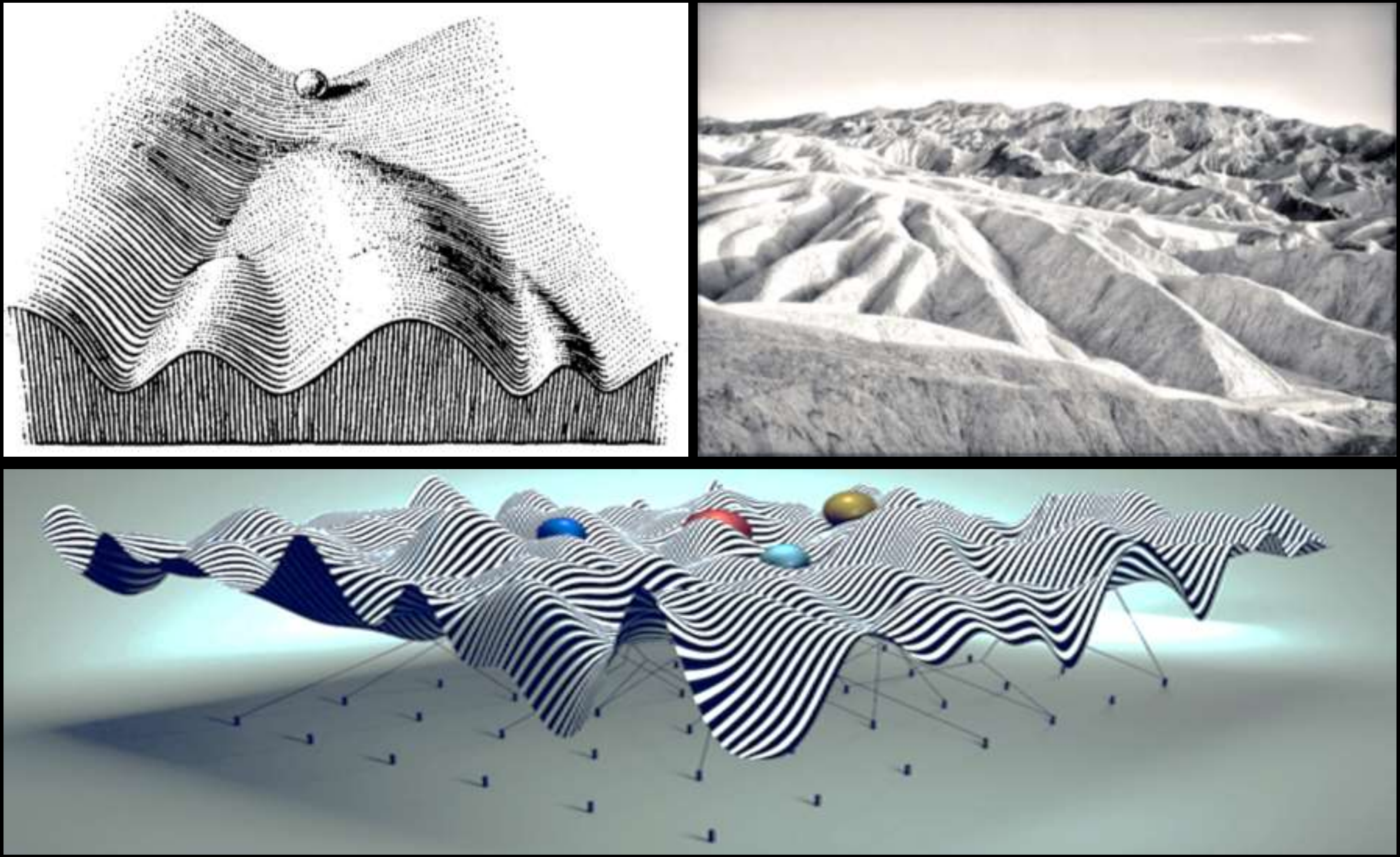


From: O. Franzén et al. (2016) Science 353, 827

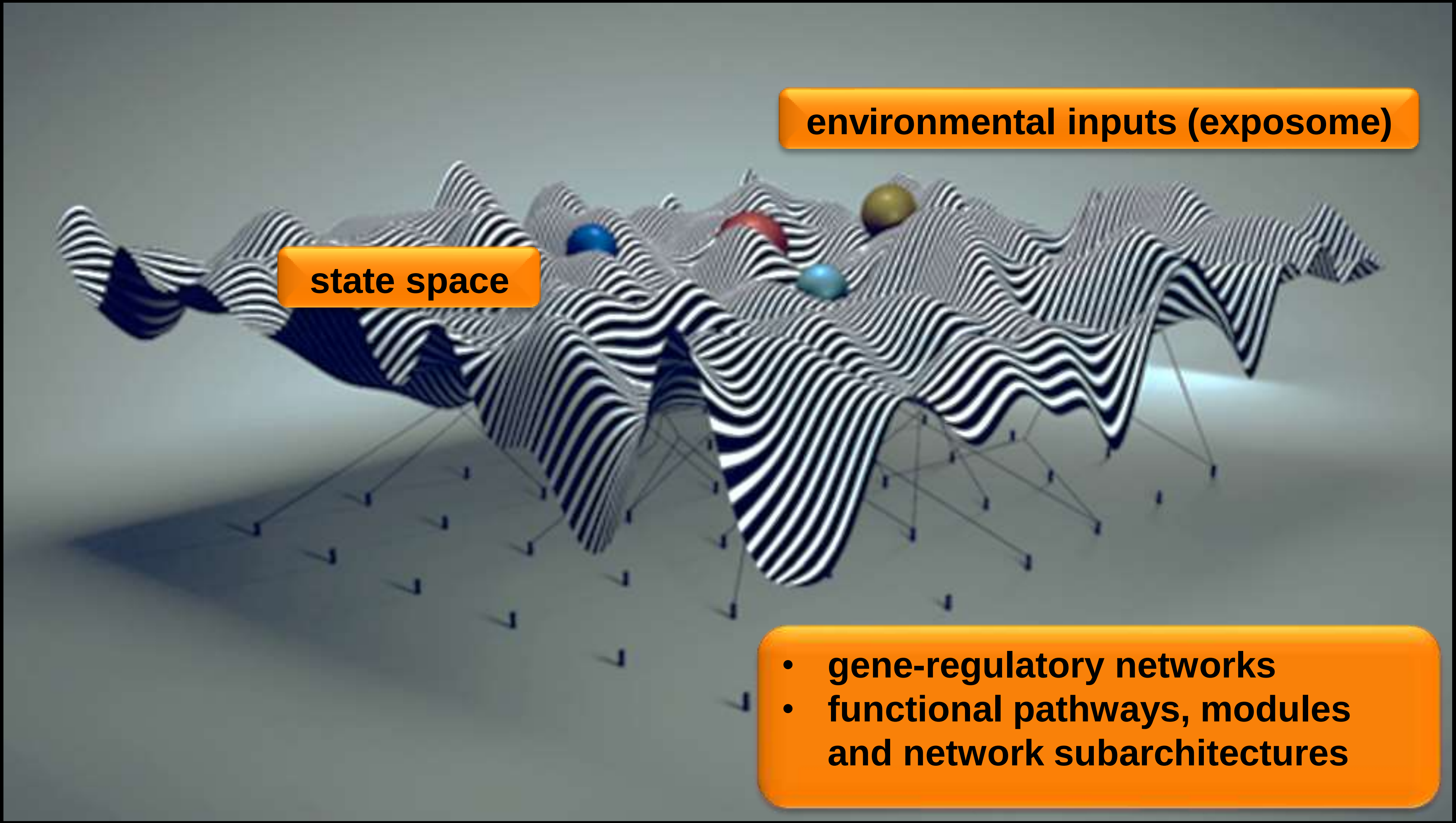
Complex Adaptive Systems (CAS)

- comprise diverse components/agents
- non-linear interactions between lower level components/agents define the systems global properties
  - ‘more than the sum of its parts’
- interaction of lower-level components themselves may exhibit CAS properties (sub-architectures)
- subarchitectures are more than hierarchical
  - entangled spatio-temporal patterns
- overall system state(s) defined by complex high-dimensional attractors

Multi–Attractor Landscapes and State Space Occupancies in CAS (After Haldane)



Multi–Attractor Landscapes and State Space Occupancies in CAS



environmental inputs (exposome)

state space

- gene-regulatory networks
- functional pathways, modules and network subarchitectures

Complex Adaptive Systems (CAS)

- **self-organized criticality**
  - **generation of ordered system states with defined ‘excursional’ boundaries**
  - **operate at far-from-equilibrium states**
- **system states/boundaries are highly robust (resilient; anti-fragile) to commonly encountered perturbations (selection pressures)**
- **novel and rarely encountered perturbations can elicit major shifts in system states and new properties (emergence)**
  - **fragility triggers and mechanisms often only recognized retrospectively**

System Robustness

- (apparent) phenotypic robustness may still involve perturbations in internal physical states that predispose to fragility to other perturbations
- potential disease-causing variants may be masked (cryptic) in robust networks
- mapping of ‘protective’ genomic features

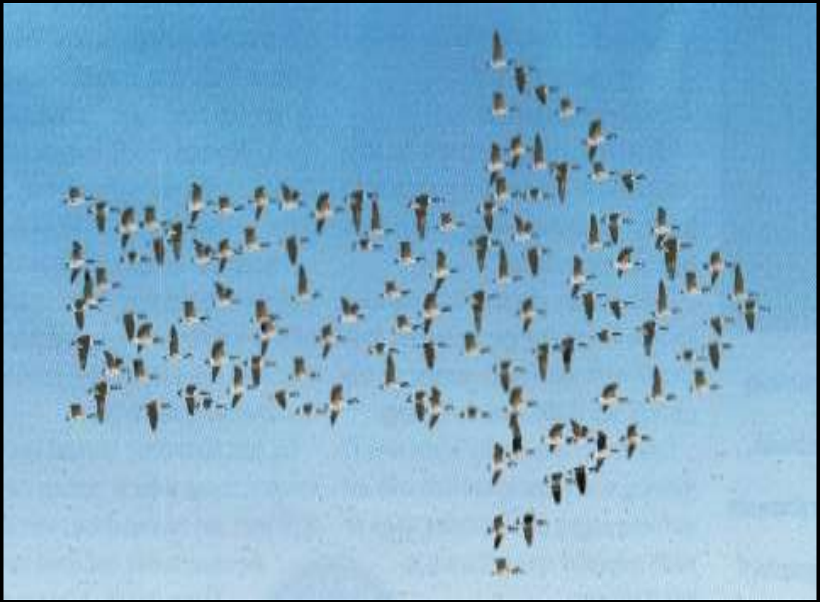
  - human gene knockouts

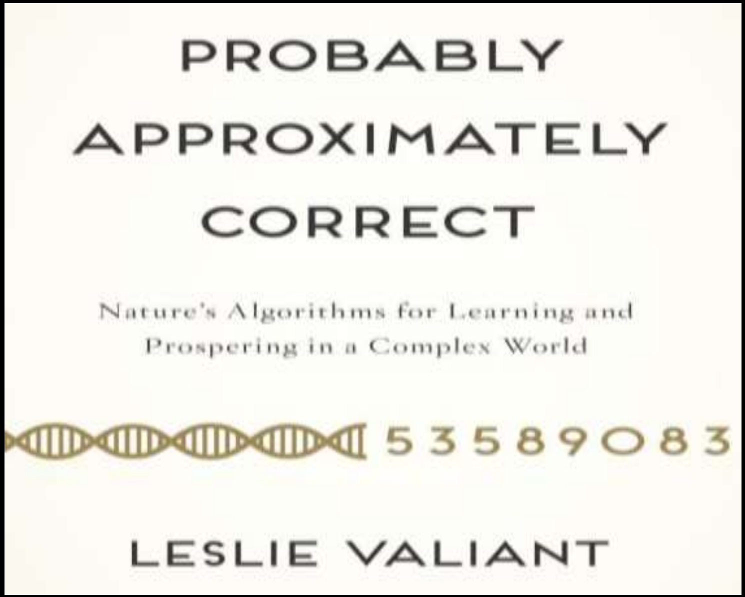
  - presence of mutations for serious Mendelian conditions that no disease manifestation

What Is Information in Biological CAS?

- data (descriptions of input/output relationships)
- entropy (potential information)
- information (confers predictive decision trajectories that generate outcomes greater than by chance)

New Analytics, Algorithms, Models and Simulations for Non-Deterministic, Non-Linear and Stochastic 'Systems Spaces'



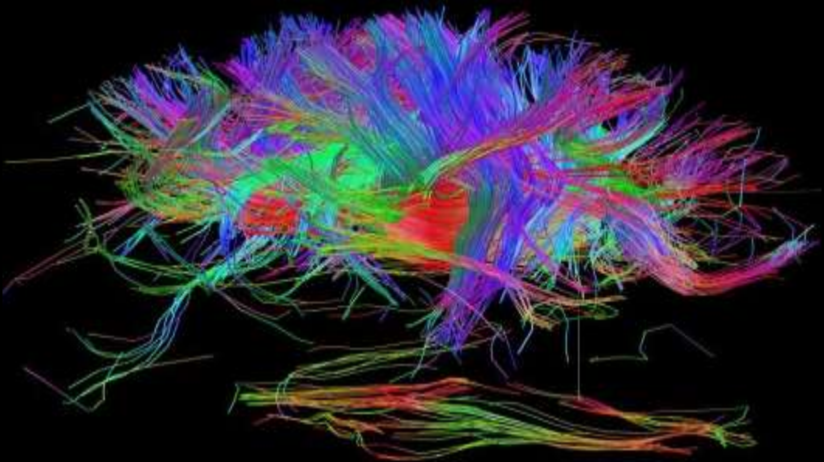


PROBABLY
APPROXIMATELY
CORRECT

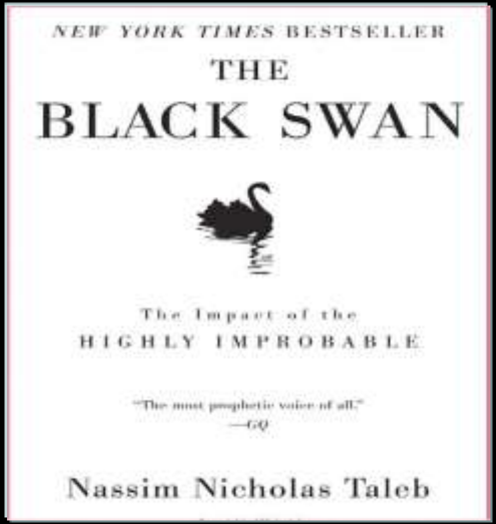
Nature's Algorithms for Learning and
Prospering in a Complex World

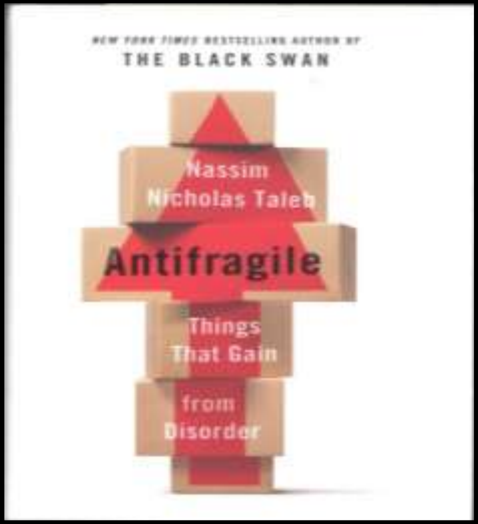
5 3 5 8 9 0 8 3

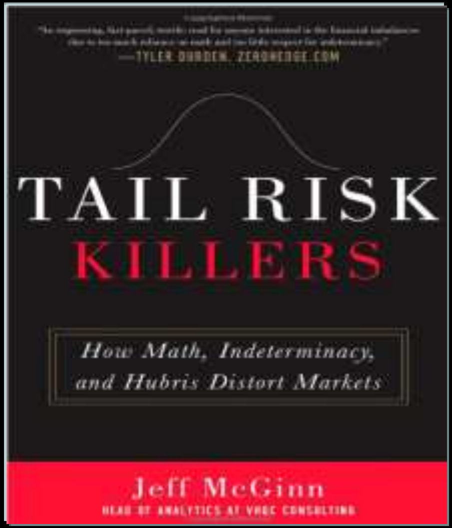
LESLIE VALIANT

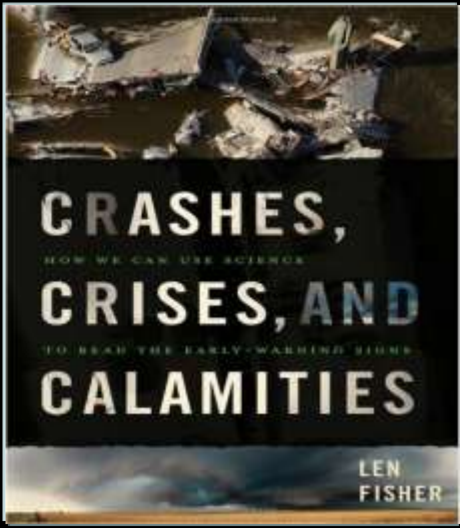


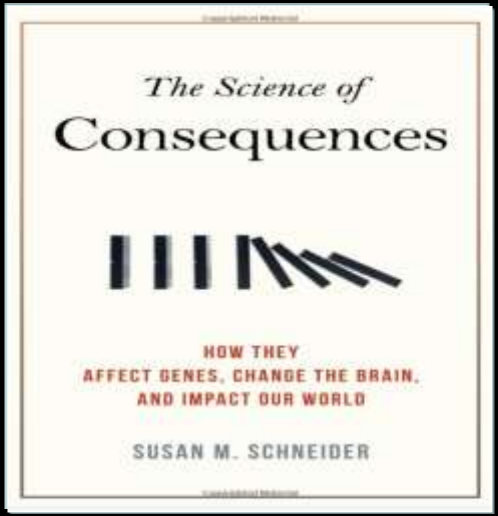
Managing (Taming) Complexity and Uncertainty

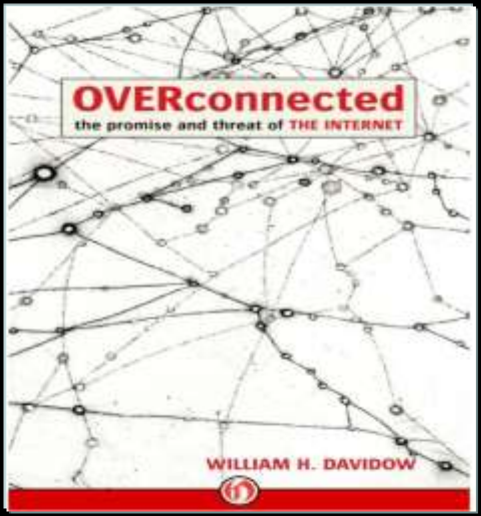


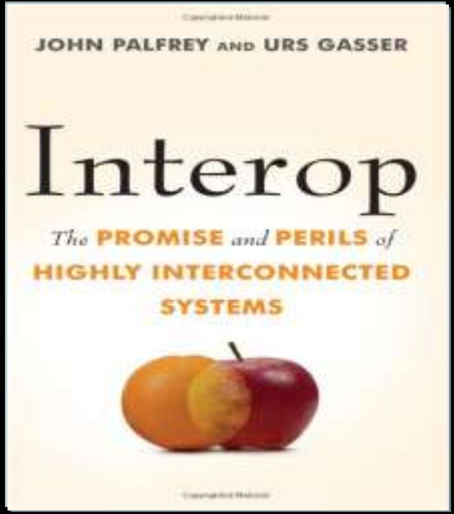








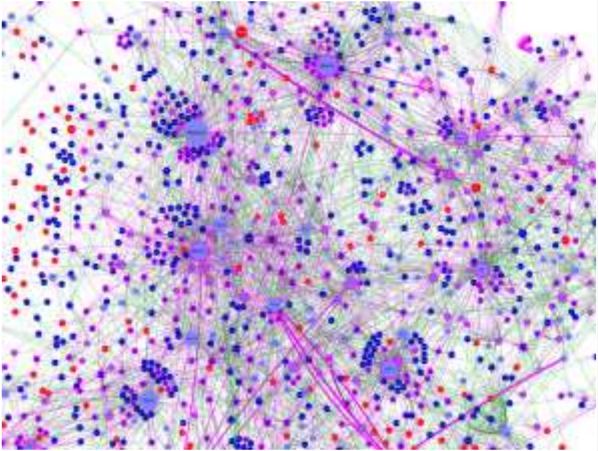




Dynamic Modeling of Signaling Pathways and Networks in Complex Systems

- what parts of the system are the most/least sensitive to perturbation ?
- what parts of the network are most/least influential on the rest of the network when perturbed ?
- identification of R_x targets and R_x resistance

Is Unifocal R_x Modulation of Complex Network Dysregulation in Advanced Chronic Diseases Feasible?



- multi-node/multi-module/ multi-subnetwork dysregulation
- "too disrupted to restore"? (homeostatic reset)
- low feasibility of multi-R_x intervention against multiple dysregulated targets
- even lower feasibility of design of promiscuous multi-target single R_x

“I don’t think any physician today should be practicing without artificial intelligence assisting in their practice.

It’s just impossible otherwise to pick up on patterns, to pick up on trends to really monitor care.”

Bernard J. Tyson
CEO, Kaiser Permanente
Cited in Forbes: The Future of Work
1 March 2017

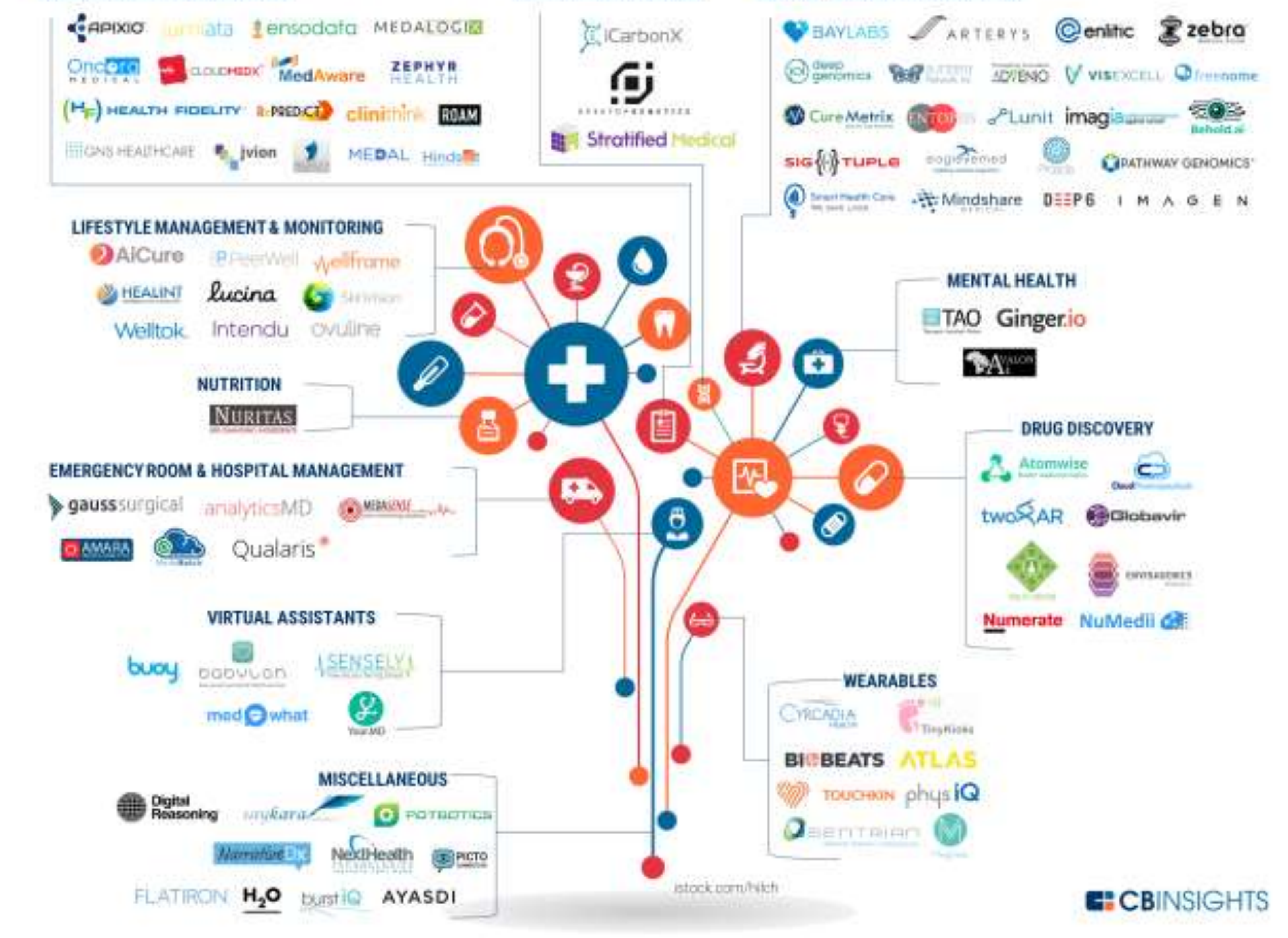
“By far the greatest danger of Artificial Intelligence is that people conclude too early that they understand it.”

Eliezer Yudkowsky

O’Reilly Artificial Intelligence

May 2017

90+ Startup AI Companies in Healthcare

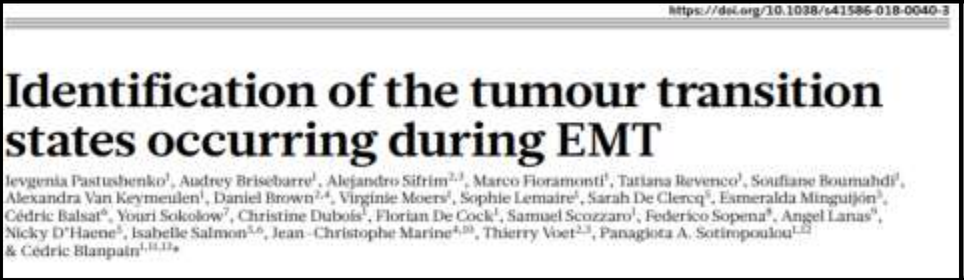


Deep Phenotyping and Dynamic Phenotyping

Longitudinal Phenotype Progression
in the Health to Disease Continuum (Phenomes)

Single Cell Analytics and Deep Multi-Omics Phenotyping

Nature (2018) 556, 463



https://doi.org/10.1038/s41586-018-0040-3

Identification of the tumour transition states occurring during EMT

Ievgenia Pastushenko[1], Audrey Brisebarre[1], Alejandro Sifrim[2,3], Marco Fioramonti[1], Tatiana Revenco[1], Soufiane Boumahdi[1], Alexandra Van Keymeulen[1], Daniel Brown[2,4], Virginie Moers[1], Sophie Lemaire[1], Sarah De Clercq[5], Esmeralda Minguijón[3], Cédric Balsat[6], Youri Sokolow[7], Christine Dubois[1], Florian De Cock[1], Samuel Scozzaro[1], Federico Sopena[8], Angel Lanas[9], Nicky D'Haene[5], Isabelle Salmon[5,6], Jean-Christophe Marine[4,10], Thierry Voet[2,3], Panagiota A. Sotiropoulou[1,12] & Cédric Blanpain[1,11,12]*

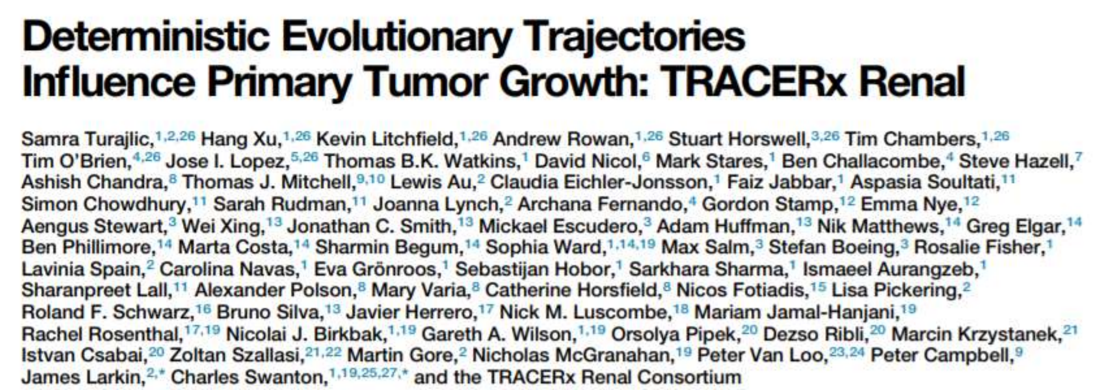
Nature (2018) 556, 457

https://doi.org/10.1038/s41586-018-0024-3

Intra-tumour diversification in colorectal cancer at the single-cell level

Sophie F. Roerink[1,13], Nobuo Sasaki[2,11,13], Henry Lee-Six[1,13], Matthew D. Young[1], Ludmil B. Alexandrov[1,4,5], Sam Behjati[1,6], Thomas J. Mitchell[1,7], Sebastian Grossmann[1], Howard Lightfoot[1], David A. Egan[8,12], Apollo Pronk[9], Niels Smakman[9], Joost van Gorp[10], Elizabeth Anderson[1], Stephen J. Gamble[1], Chris Alder[1], Marc van de Wetering[2], Peter J. Campbell[1], Michael R. Stratton[1]* & Hans Clevers[2]*

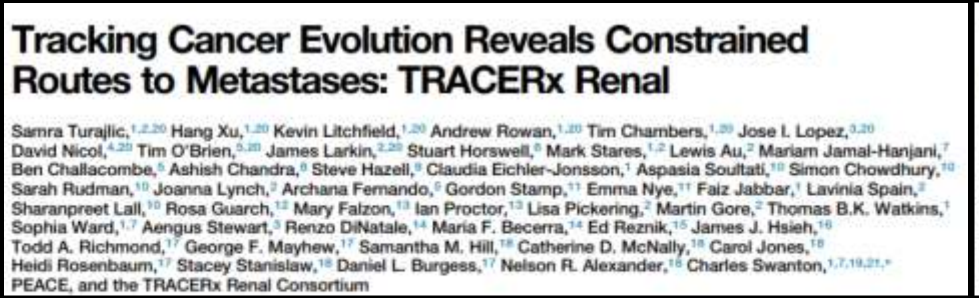
Cell (2018) 173, 595



Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal

Samra Turajlic,[1,2,26] Hang Xu,[1,26] Kevin Litchfield,[1,26] Andrew Rowan,[1,26] Stuart Horswell,[3,26] Tim Chambers,[1,26] Tim O'Brien,[4,26] Jose I. Lopez,[5,26] Thomas B.K. Watkins,[1] David Nicol,[6] Mark Stares,[1] Ben Challacombe,[4] Steve Hazell,[7] Ashish Chandra,[8] Thomas J. Mitchell,[9,10] Lewis Au,[2] Claudia Eichler-Jonsson,[1] Faiz Jabbar,[1] Aspasia Soultati,[11] Simon Chowdhury,[11] Sarah Rudman,[11] Joanna Lynch,[2] Archana Fernando,[4] Gordon Stamp,[12] Emma Nye,[12] Aengus Stewart,[3] Wei Xing,[13] Jonathan C. Smith,[13] Mickael Escudero,[3] Adam Huffman,[13] Nik Matthews,[14] Greg Elgar,[14] Ben Phillimore,[14] Marta Costa,[14] Sharmin Begum,[14] Sophia Ward,[1,14,19] Max Salm,[3] Stefan Boeing,[3] Rosalie Fisher,[1] Lavinia Spain,[2] Carolina Navas,[1] Eva Grönroos,[1] Sebastijan Hobor,[1] Sarkhara Sharma,[1] Ismaeel Aurangzeb,[1] Sharanpreet Lall,[11] Alexander Polson,[8] Mary Varia,[8] Catherine Horsfield,[8] Nicos Fotiadis,[15] Lisa Pickering,[2] Roland F. Schwarz,[16] Bruno Silva,[13] Javier Herrero,[17] Nick M. Luscombe,[18] Mariam Jamal-Hanjani,[19] Rachel Rosenthal,[17,19] Nicolai J. Birkbak,[1,19] Gareth A. Wilson,[1,19] Orsolya Pipek,[20] Dezso Ribli,[20] Marcin Krzystanek,[21] Istvan Csabai,[20] Zoltan Szallasi,[21,22] Martin Gore,[2] Nicholas McGranahan,[19] Peter Van Loo,[23,24] Peter Campbell,[9] James Larkin,[2,*] Charles Swanton,[1,19,25,27,*] and the TRACERx Renal Consortium

Cell (2018) 173, 581



Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal

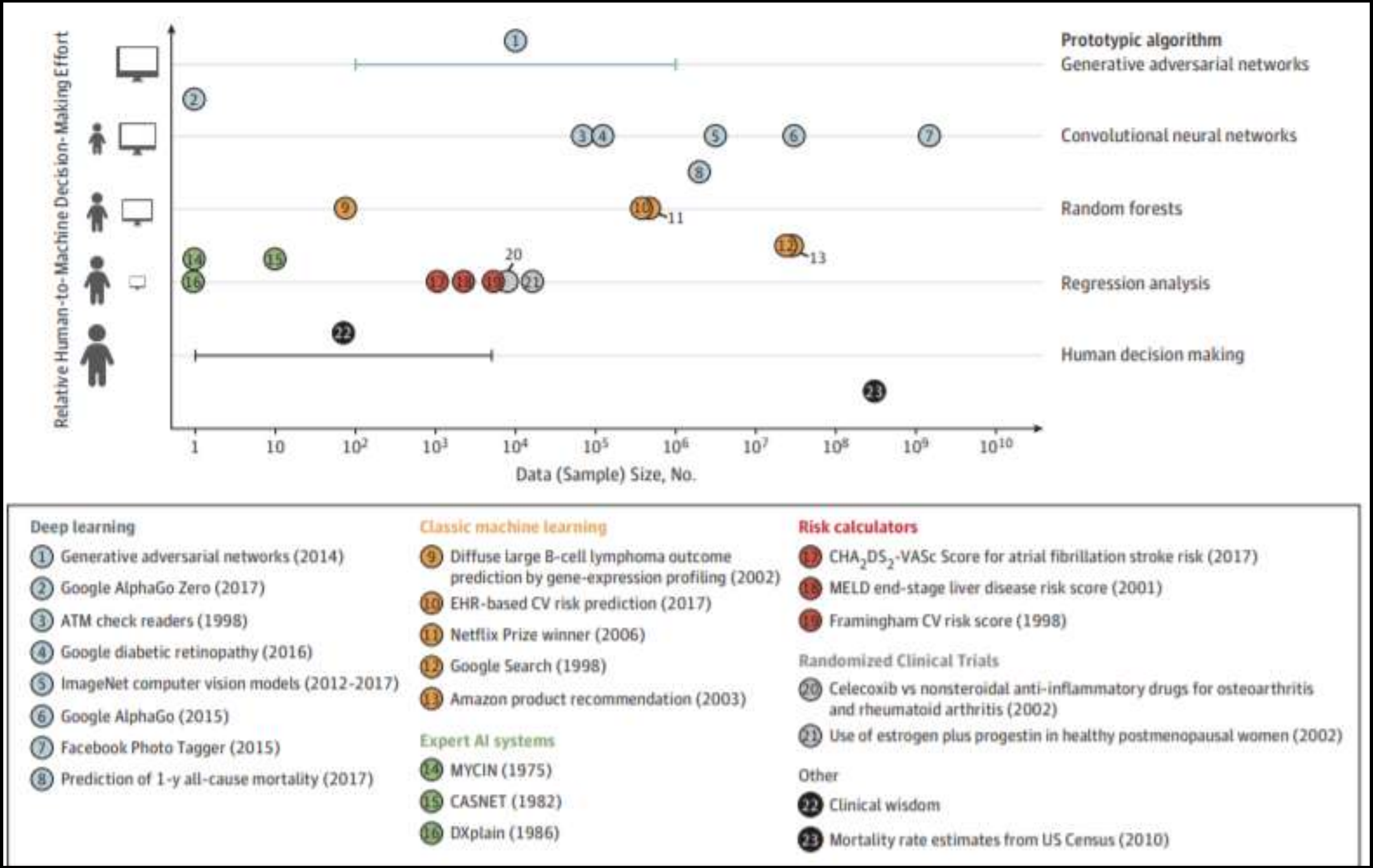
Samra Turajlic,[1,2,20] Hang Xu,[1,20] Kevin Litchfield,[1,20] Andrew Rowan,[1,20] Tim Chambers,[1,20] Jose I. Lopez,[3,20] David Nicol,[4,20] Tim O'Brien,[5,20] James Larkin,[2,20] Stuart Horswell,[6] Mark Stares,[1,2] Lewis Au,[2] Mariam Jamal-Hanjani,[7] Ben Challacombe,[5] Ashish Chandra,[8] Steve Hazell,[9] Claudia Eichler-Jonsson,[1] Aspasia Soultati,[10] Simon Chowdhury,[10] Sarah Rudman,[10] Joanna Lynch,[2] Archana Fernando,[5] Gordon Stamp,[11] Emma Nye,[11] Faiz Jabbar,[1] Lavinia Spain,[2] Sharanpreet Lall,[10] Rosa Guarch,[12] Mary Falzon,[13] Ian Proctor,[13] Lisa Pickering,[2] Martin Gore,[2] Thomas B.K. Watkins,[1] Sophia Ward,[1,7] Aengus Stewart,[6] Renzo DiNatale,[14] Maria F. Becerra,[14] Ed Reznik,[15] James J. Hsieh,[16] Todd A. Richmond,[17] George F. Mayhew,[17] Samantha M. Hill,[18] Catherine D. McNally,[18] Carol Jones,[18] Heidi Rosenbaum,[17] Stacey Stanislaw,[18] Daniel L. Burgess,[17] Nelson R. Alexander,[18] Charles Swanton,[1,7,19,21,*] PEACE, and the TRACERx Renal Consortium

Cell (2018) 173, 611

Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal

Thomas J. Mitchell,[1,2,21] Samra Turajlic,[3,4,21] Andrew Rowan,[3,21] David Nicol,[4,21] James H.R. Farmery,[5] Tim O'Brien,[6] Inigo Martincorena,[1] Patrick Tarpey,[1] Nicos Angelopoulos,[1] Lucy R. Yates,[1,4] Adam P. Butler,[1] Keiran Raine,[1] Grant D. Stewart,[2] Ben Challacombe,[6] Archana Fernando,[6] Jose I. Lopez,[7] Steve Hazell,[4] Ashish Chandra,[8] Simon Chowdhury,[6] Sarah Rudman,[6] Aspasia Soultati,[6] Gordon Stamp,[9] Nicos Fotiadis,[4] Lisa Pickering,[4] Lewis Au,[4] Lavinia Spain,[4] Joanna Lynch,[4] Mark Stares,[4] Jon Teague,[1] Francesco Maura,[1] David C. Wedge,[10] Stuart Horswell,[11] Tim Chambers,[3] Kevin Litchfield,[3] Hang Xu,[3] Aengus Stewart,[11] Reza Elaidi,[12] Stéphane Oudard,[12] Nicholas McGranahan,[3,13] Istvan Csabai,[14] Martin Gore,[4] P. Andrew Futreal,[15] James Larkin,[4] Andy G. Lynch,[5,16] Zoltan Szallasi,[17,18] Charles Swanton,[3,13,19,*] Peter J. Campbell,[1,20,22,*] and the TRACERx Renal Consortium

Machine Learning and Big Data

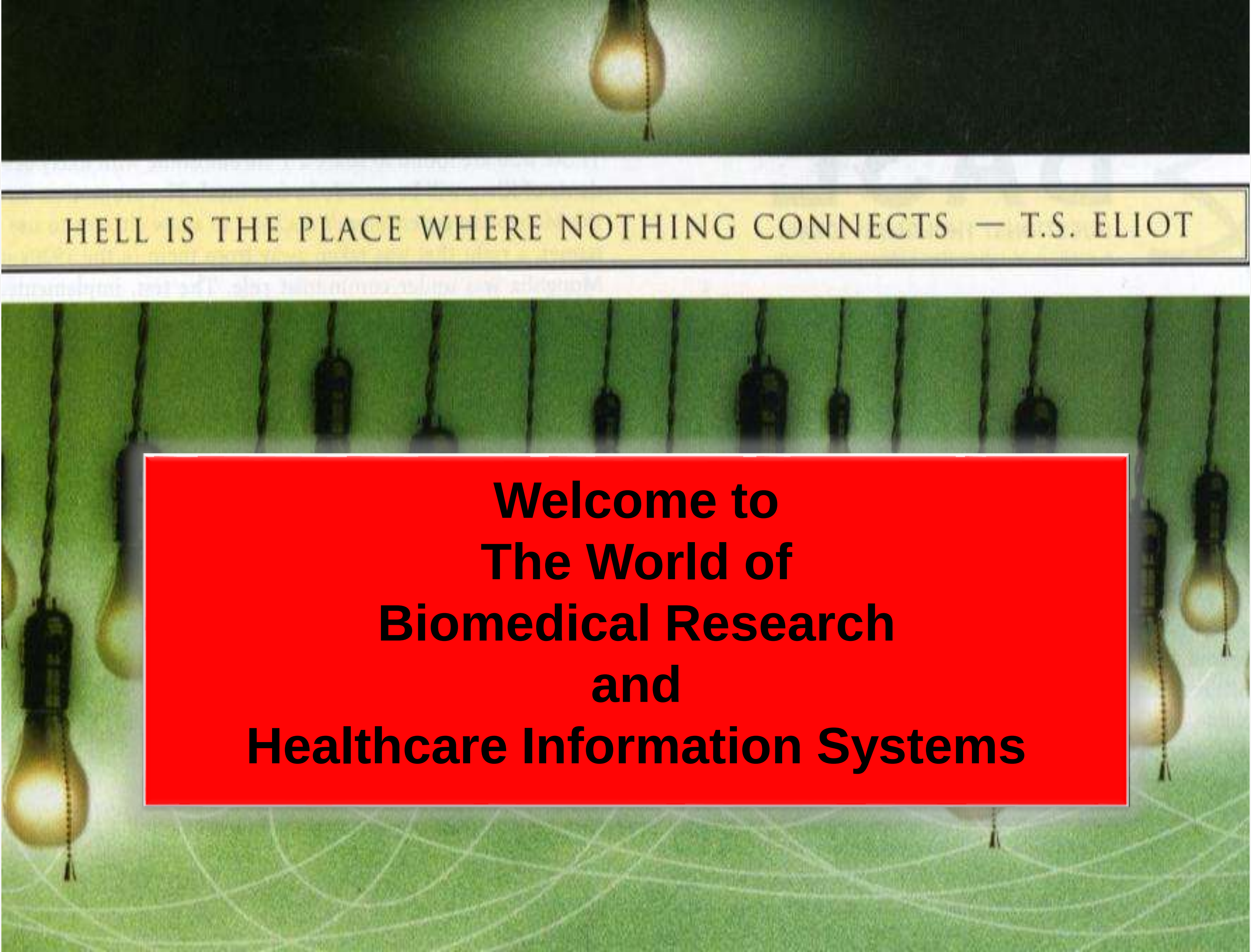


Adapted from: A.L. Beam and I.S. Kohane (2018) JAMA 319,1317

Now Comes the Really Hard Part!

Integration of Complex Network-Based Biomarker Profiling with Other Diverse Data Categories Relevant to Disease Risk and Optimum Clinical Care Decisions

The Problem with Real World Data is the Real World



Welcome to
The World of
Biomedical Research
and
Healthcare Information Systems

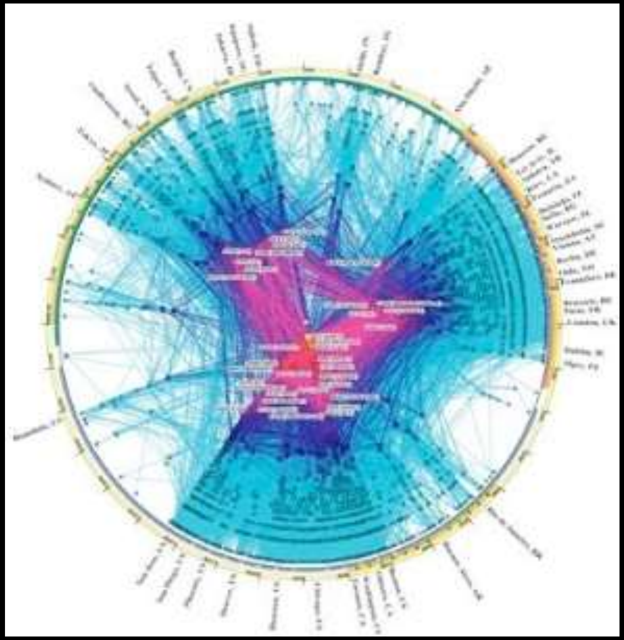
Integration of Molecular Profiling, Clinical and Social Datasets for Computable Disease Phenotypes

- need for generalizable computational infrastructure for diverse deep phenotyping data classes
  - HL7 Fast Healthcare Interoperability Resources (FHIR)
  - integration of cTAKES, SMART, SHARP, TIES, OBO
- ONC requirements for EHR interoperability
- payer requirements for RWE
  - new trial protocols and registries

The Emergence of Big Data Changes the Questions That Can Be Asked

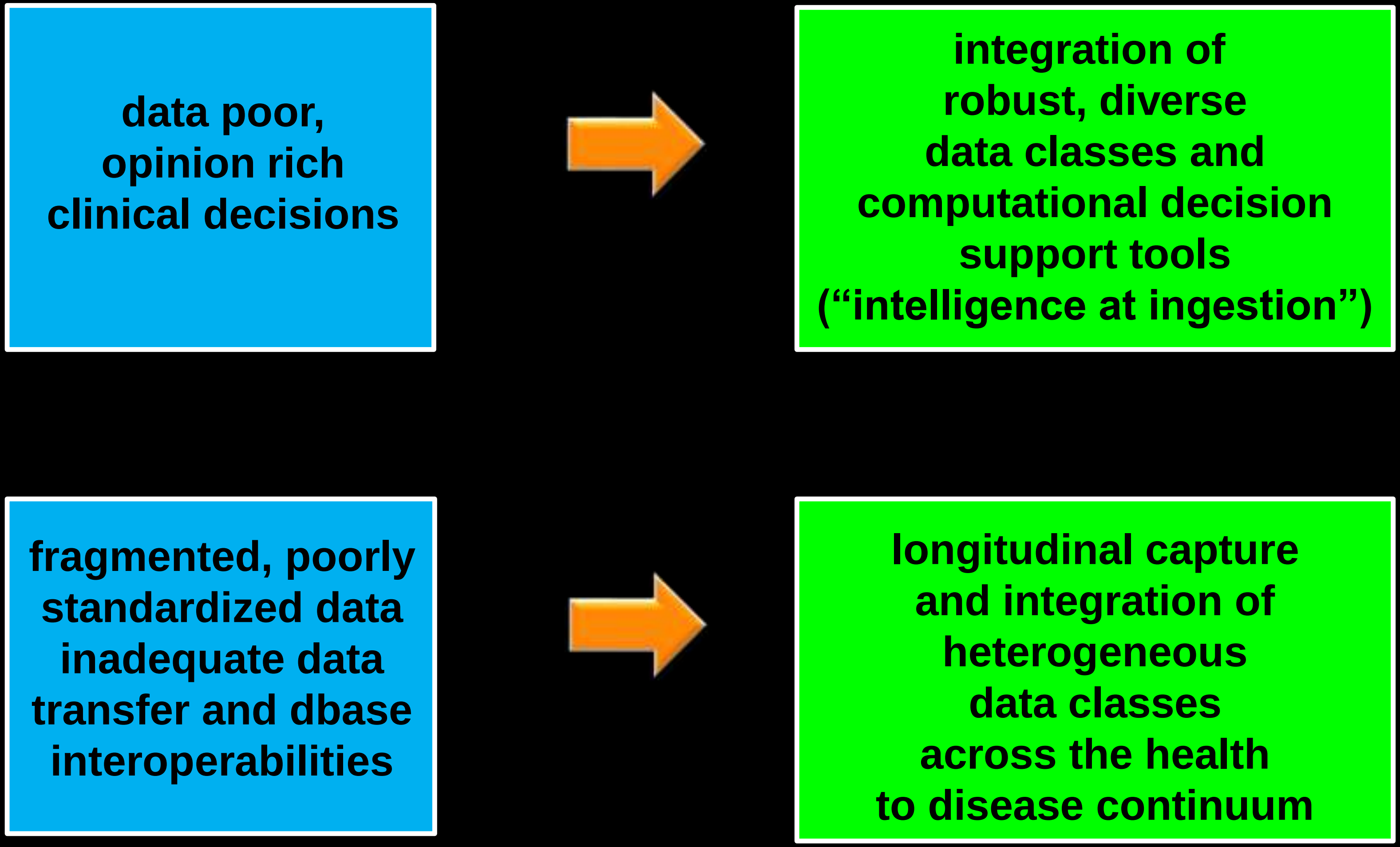
Isolated Data

Complex Networked Data



Complex Computational Data

Precision Medicine and Digital Medicine: Obligate Inter-Dependencies



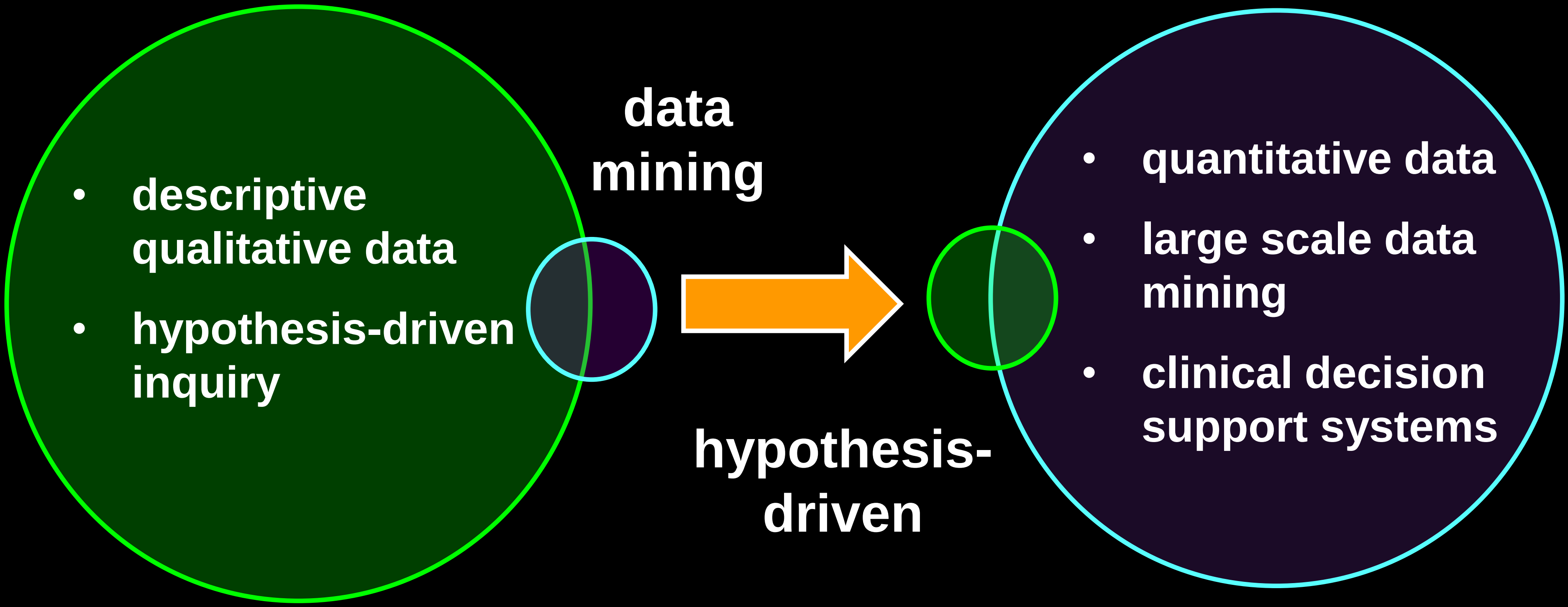
data poor,
opinion rich
clinical decisions

integration of
robust, diverse
data classes and
computational decision
support tools
(“intelligence at ingestion”)

fragmented, poorly
standardized data
inadequate data
transfer and dbase
interoperabilities

longitudinal capture
and integration of
heterogeneous
data classes
across the health
to disease continuum

A Pending Transition in Biomedical Research and Clinical Care Decisions



- descriptive qualitative data
- hypothesis-driven inquiry

data mining

hypothesis-driven

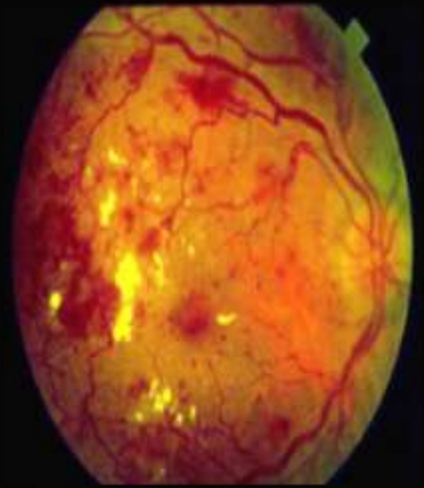
- quantitative data
- large scale data mining
- clinical decision support systems

Machine Learning and Image Analysis in Clinical Medicine

pathology

radiology

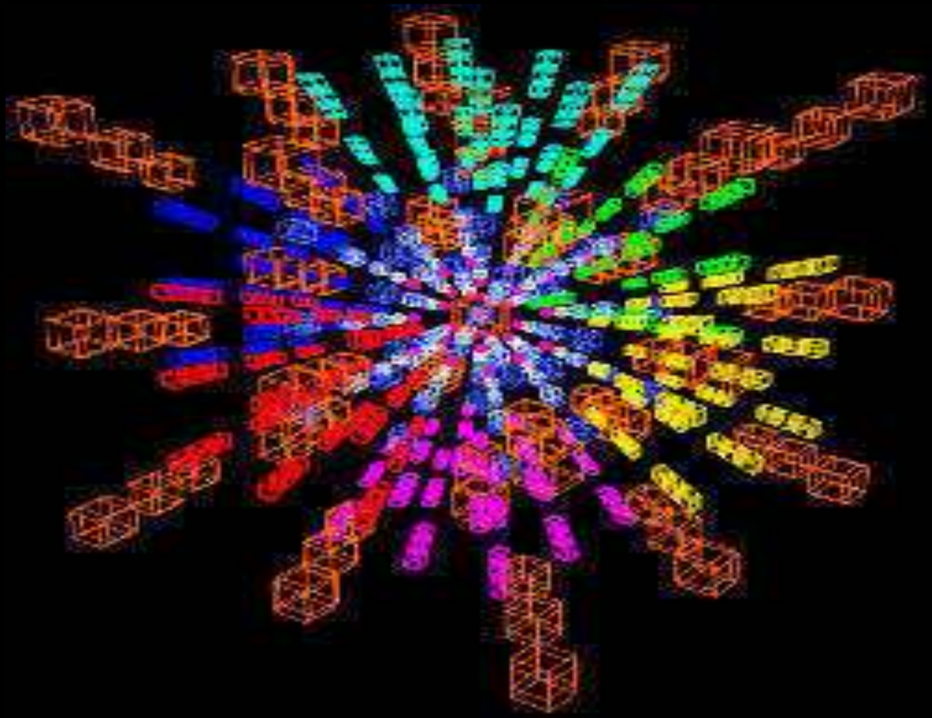
dermatology

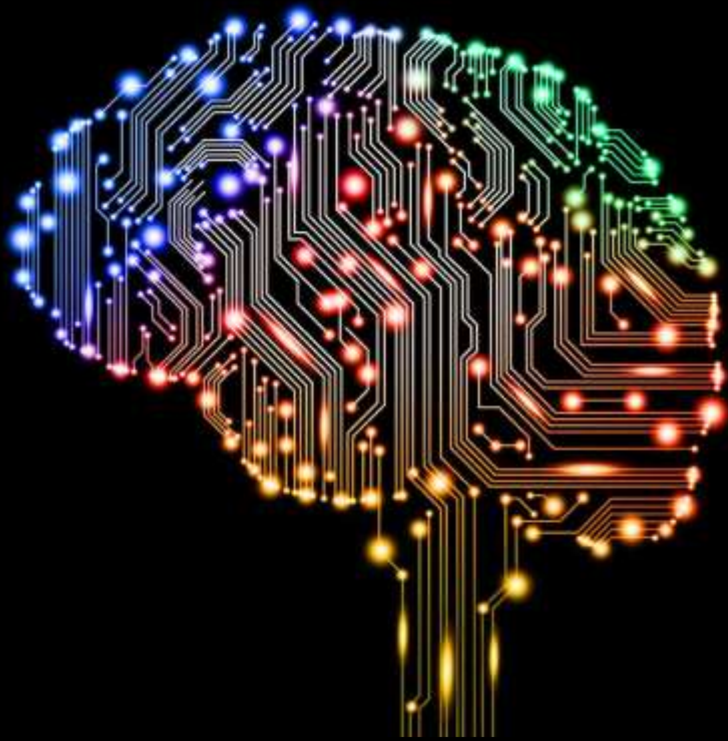


ophthalmology

- large scale training sets and classification parameters
- standardized, reproducible and scalable
- 260 million images/day for $1000 GPU

Automated Context: Data Finding Data
“Intelligence at Ingestion”





Feature Extraction and Classification

→

Context Analysis

↕

Persistent Context

→

- **Relevance Detection**
- **Learning Systems**

→

- **Situational Awareness**
- **Rapid, Robust Decisions**

Data-Driven Knowledge, Decision Support Systems

- changing the nature of discovery
  - hypothesis-driven versus unbiased analytics of large datasets (patterns, rules)
- changing the cultural process of knowledge acquisition
  - large scale collaboration networks, open systems versus individual investigators and siloed data
- changing knowledge application
  - increased quantification and decision-support systems
- changing the cognitive and intellectual competencies for data-intensive decisions
- changing education, training and research

Inside the Black Box of Algorithms

Data is Inherently Dumb

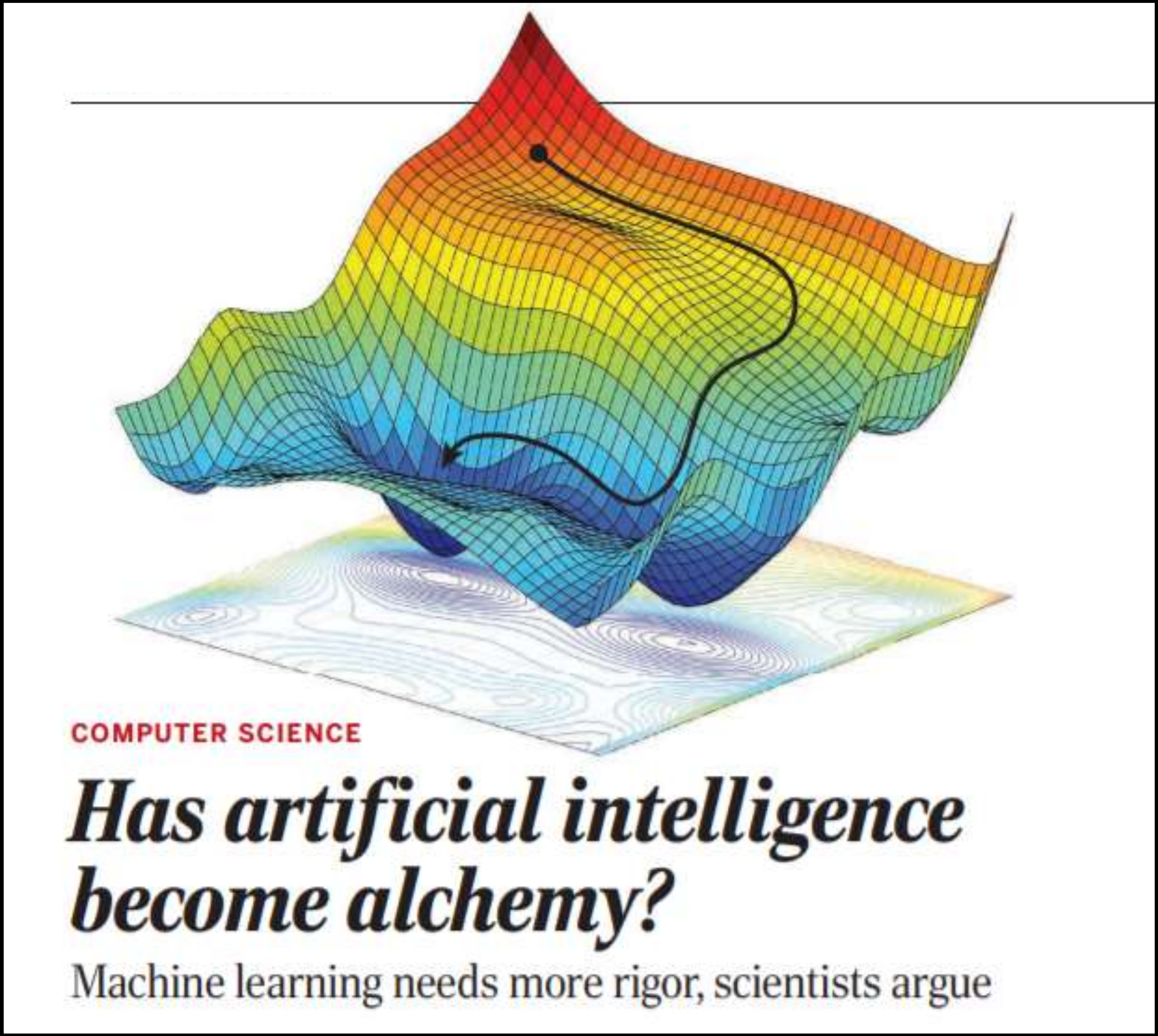
Algorithms Define Action and Value

Algorithms with Agendas

Risk, Regulation and Responsibility

Critical Questions in the Application of ML/AI Platforms in Profiling Large Scale Biomedical Data

- overfitting and bias in datasets used in training
  - error propagation versus automated recognition and exclusion
- scale and layered datasets
  - impact of accretion by incorporation of legacy systems of uncertain quality/provenance?
- “black box” effects versus “explainable AI”
  - algorithm evolution neither predicted nor understood by original coders?
  - generative adversarial networks (GANs)



COMPUTER SCIENCE

Has artificial intelligence become alchemy?

Machine learning needs more rigor, scientists argue

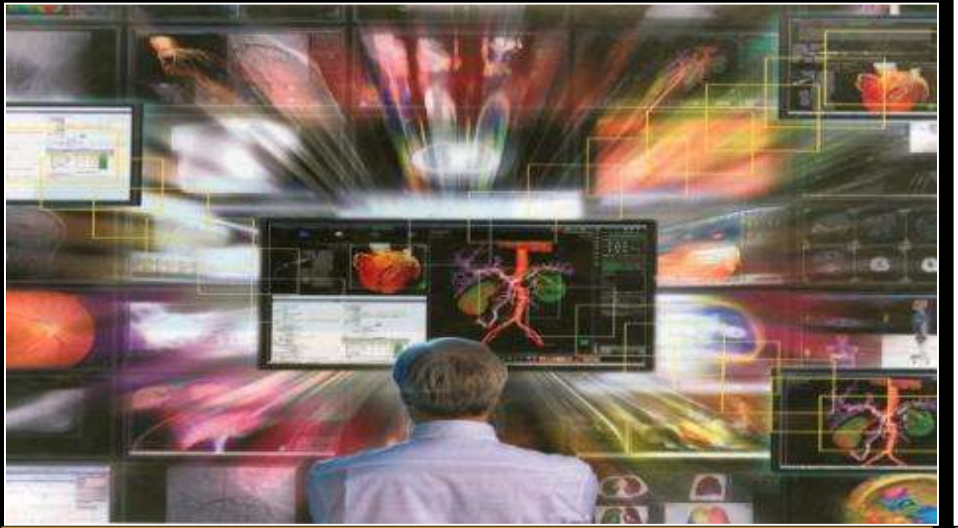
M. Hutson (2018) Science 360, 478

Black Box Algorithms

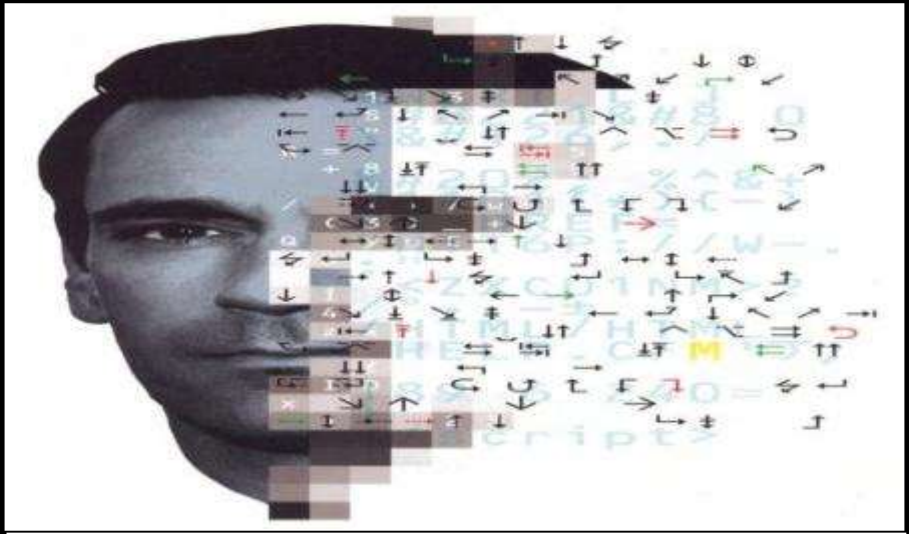
- input-output function may accurately model an actual system but internal structure may not
- convergent phenotypes can arise from different genotypes by complex mechanisms hidden from black box statistical approaches
- algorithm validation/transparency should best predict the state space as optimized by adaptive evolution not by computation
- should biomarker ID/phenotype classification used in clinical decisions require that system/subsystem architecture modeled *in silico* approximate its true state *in vivo*?

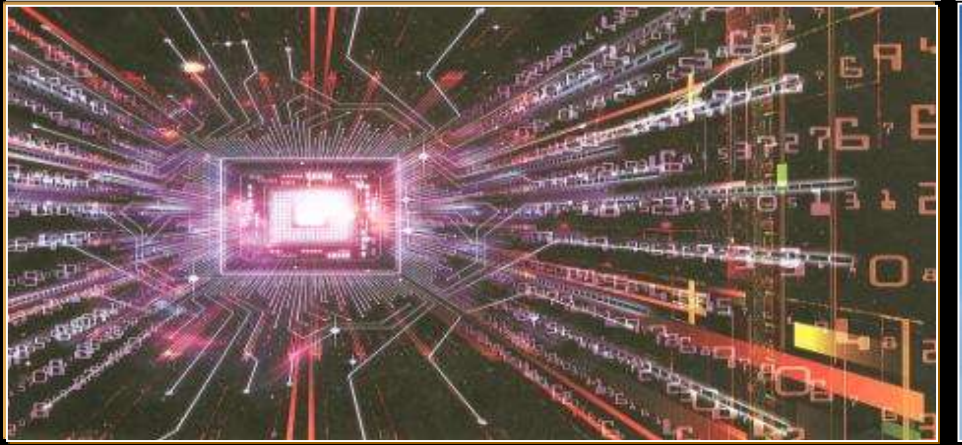
Technology Acceleration and Convergence: The Escalating Challenge for Professional Competency, Decision-Support and Future Medical Education

Data Deluge



Cognitive Bandwidth Limits





Automated Analytics and Decision Support

Facile Formats for Actionable Decisions

Living in a World Where the Data Analytics and Interpretation Algorithms Are Obscure to the End User

- ceding decision authority to computerized support systems
- culturally alien to professionals in their claimed expertise domain but they accept in all other aspects of their lives
- who will have the responsibility for validation and oversight of critical assumptions used in decision tree analytics for big datasets in biomarker-centric algorithms
    - regulatory agencies and professional societies?
    - humans?
    - machines?

Artificial Intelligence (AI) and Healthcare

- how will AI algorithms/decision analytics be validated/regulated ?
- how will AI be integrated into current work flows or will extensive reorganization/re-training be required ?
- how will AI platforms alter payment schemes ?
- what new malpractice liabilities will emerge by failure to use/interpret AI platforms ?
- legal liabilities for AI-based clinical decision software ?

The Integrator’s Dilemma

The Imperative for Holistic End-to-End Approaches For Biomarker-Based Profiling of the Health to Disease Continuum

The Integrator's Dilemma
The Imperative for Systems-Based Approaches

- conceptual
  - asking the right question, selection of models and methods
- analytical
  - multi-dimensional, biological network architectures and the taxonomy of dynamic state spaces
- computational
  - algorithms for modeling and analysis of non-deterministic, non-linear systems
- scale
  - the V7 big data challenge: volume, variety, velocity, veracity, virtualization, visualization and value

- validation and regulation
  - fit for purpose

The Integrator's Dilemma

- conceptual
  - asking the right question, selection of models and methods
- analytical
  - multi-dimensional biological network architectures and the taxonomy of dynamic state spaces
- computational
  - algorithms for modeling and analysis of non-deterministic, nonlinear systems
- scale
  - the V7 big data challenge: volume, variety, velocity, veracity, virtualization, visualization and value
- validation and regulation
  - fit for purpose

value

- new clinical care paradigms
- improved health outcomes
- societal benefits beyond the individual

Precision Medicine = Biomarker-Centric Medicine

Precision Medicine = Deconvolution of Architecture and Dynamics of Complex Biological Adaptive Systems

The Intellectual Grand Challenge for Biomedical Research and Healthcare Delivery

"So, as you can see, health care is so complicated you may never get well."

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