21st Century Medicine: Big Data, Wellness, Disease

Lee Hood, MD, PhD

Senior VP and Chief Strategy Officer
Institute for Systems Biology, Seattle

Senior VP and Chief Science Officer
Providence St. Joseph Health,
Seattle

Measuring Man

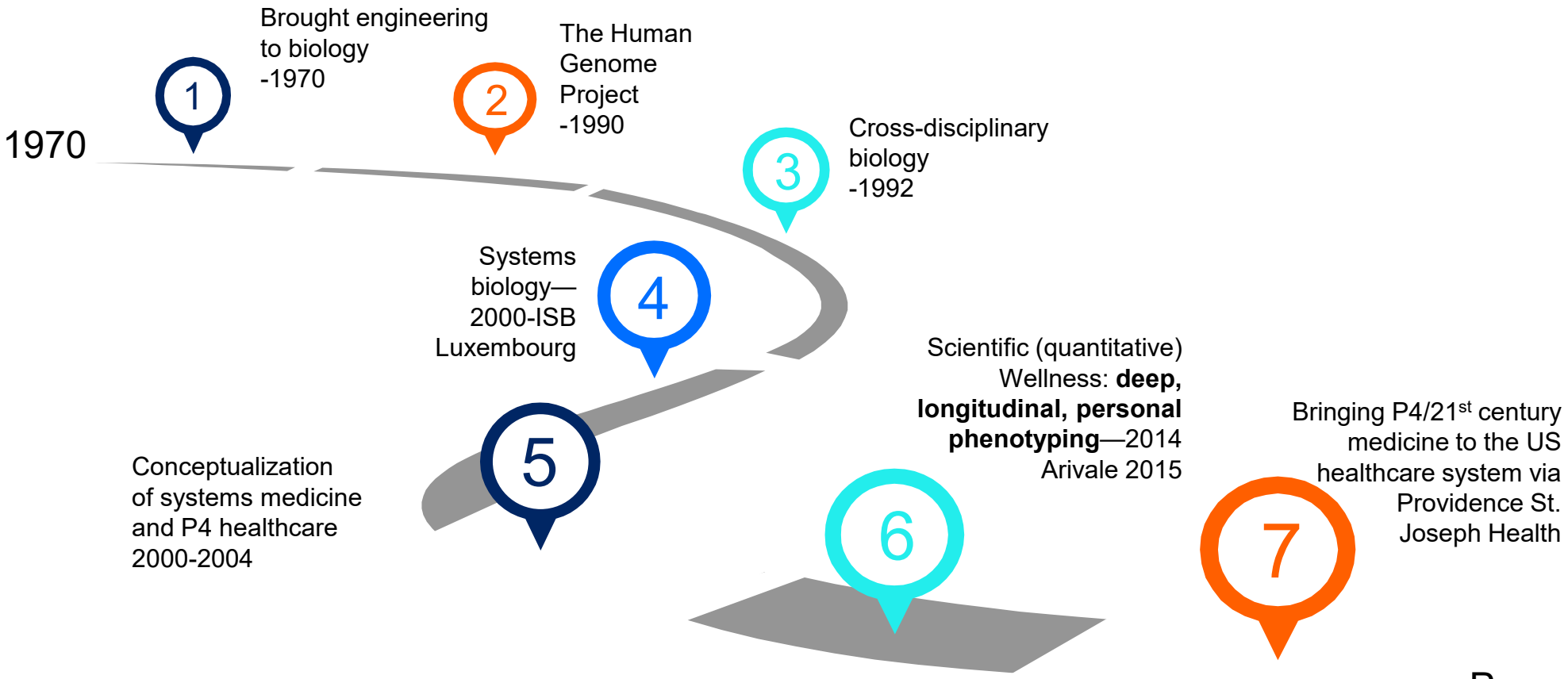
October 10, 2019

The grand challenge for biology and medicine:

Deciphering biological complexity

(1970, Assistant Professor Caltech)

I Participated in Seven Paradigm Changes in Biology Dealing with Complexity Which Led to My View of 21st Century Medicine



Two Paradigm Changes in the Last 100 Years Framing US Medicine

1900s

Germ Theory

Chemistry

Physiology

Pathology

Physics

Precision medicine

Find It
Fix It

Understanding Disease

20th Century Medicine

2000s

Systems medicine

P4 healthcare

Scientific wellness

Genomics

Deep phenotyping of individuals N=1 (Omics)

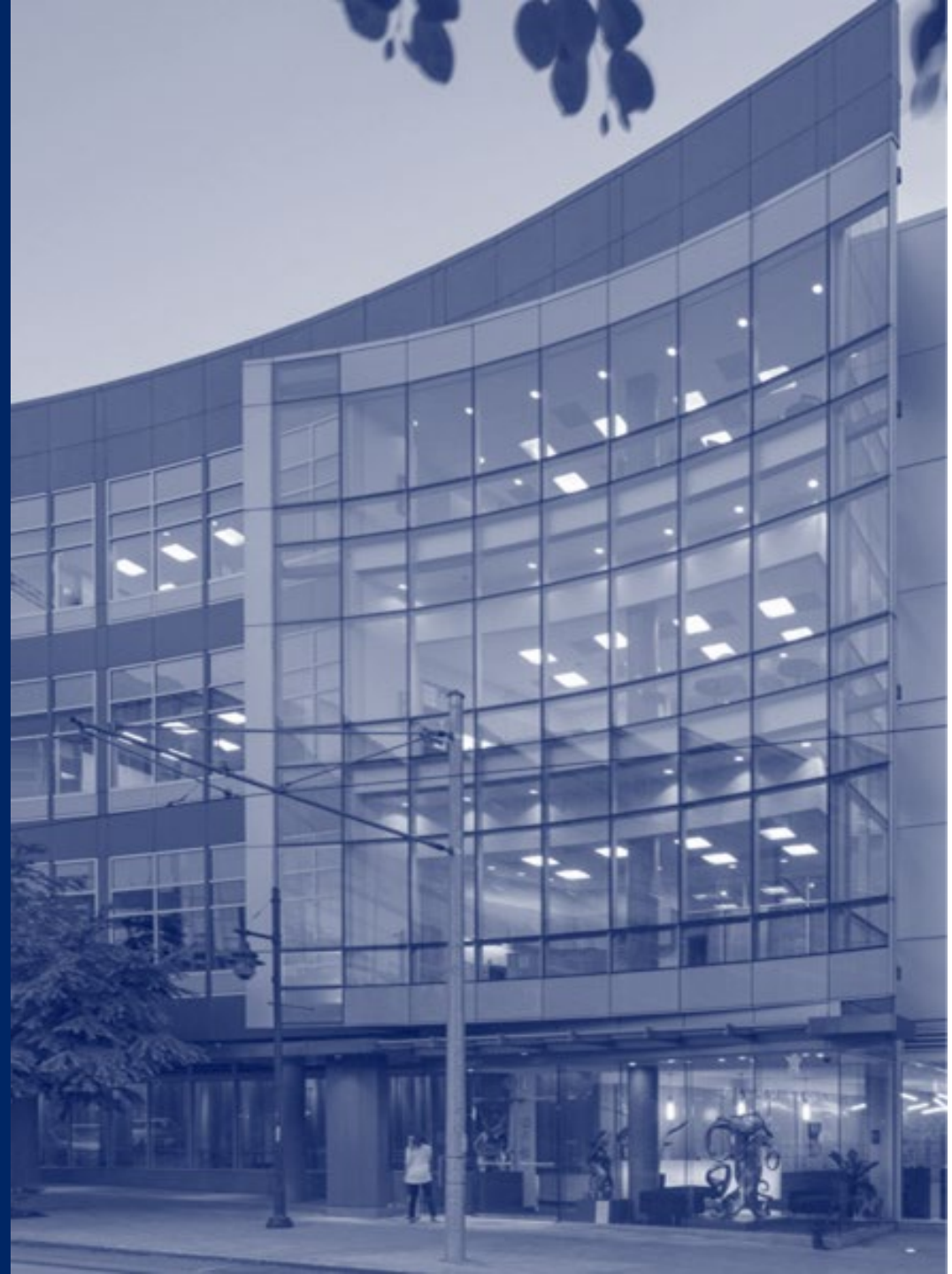
Digital self measurements

Microbiomes

Predict It
Prevent it
Personalize It
Participatory

Systems Approaches to Understanding Wellness, Disease, and their Transitions

Systems-Driven 21st Century Medicine



Institute for Systems Biology

Non-profit scientific research organization founded in 2000

2019

- *12 faculty, 220 staff*
- *$40 million annual budget*

Holistic
Longitudinal deep phenotyping/complexity
Networks and hierarchy
Dynamics
Integrative
Discovery vs. mechanistic hypothesis data generation

Inventing the Future with Integrative Systems Biology at ISB

20th Century Biomedicine

- **Reductionist**

  Analyzing one gene and one small problem at a time

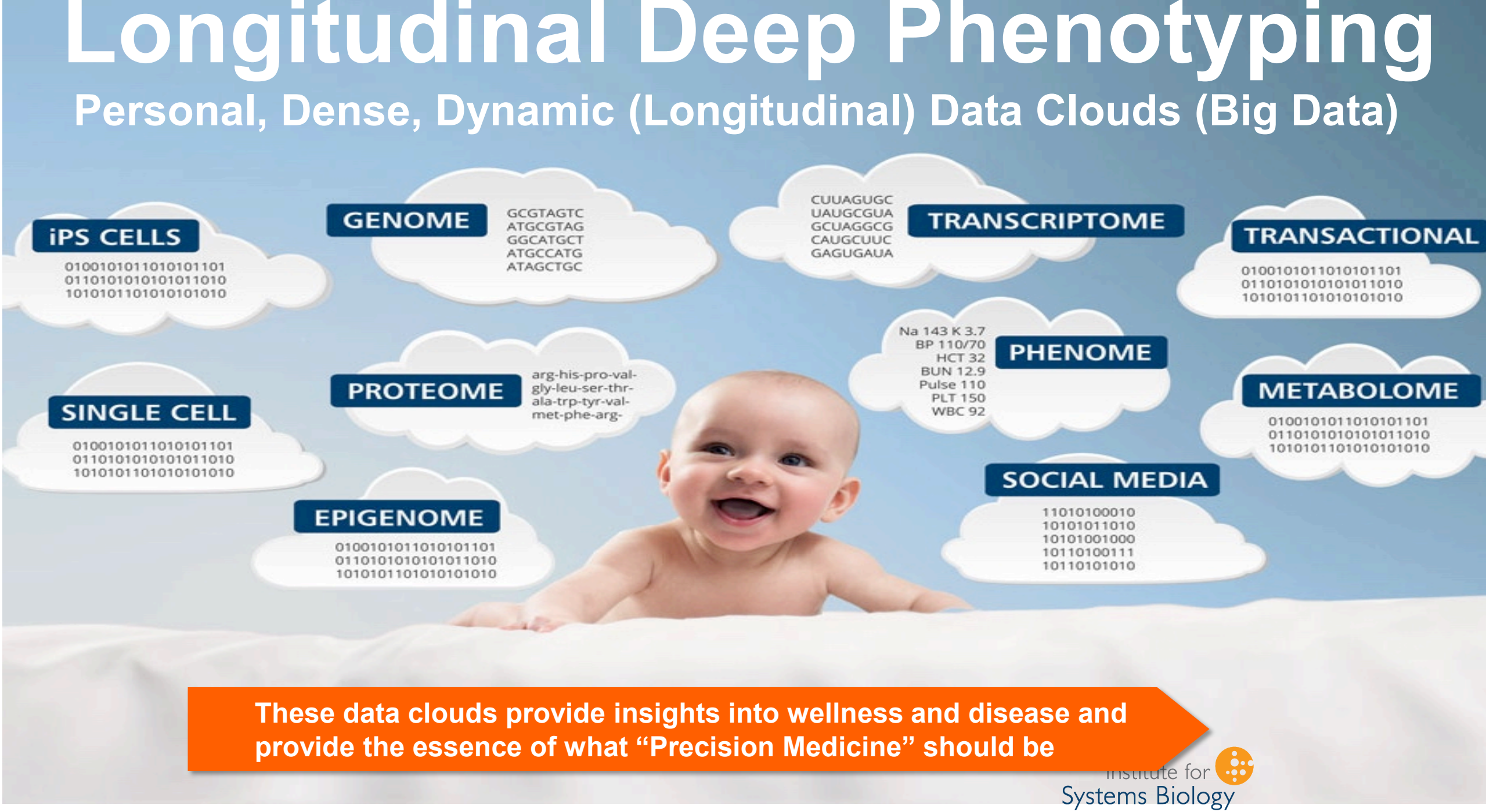
ISB FOUNDED in 2000

21st Century Biomedicine-ISB

- **Integrative Biological Systems Analysis**
- **Systems-Driven Technology and Systems-Driven Strategies**
- **Pioneering Computational Tools**
- **Cross-Disciplinary Environment**
- **Employs Big Science to Attack Large-Scale Integrative Biological Problems**
- **Strategic Partnerships**
- **Transferring Knowledge to Society – education & start ups**

Longitudinal Deep Phenotyping

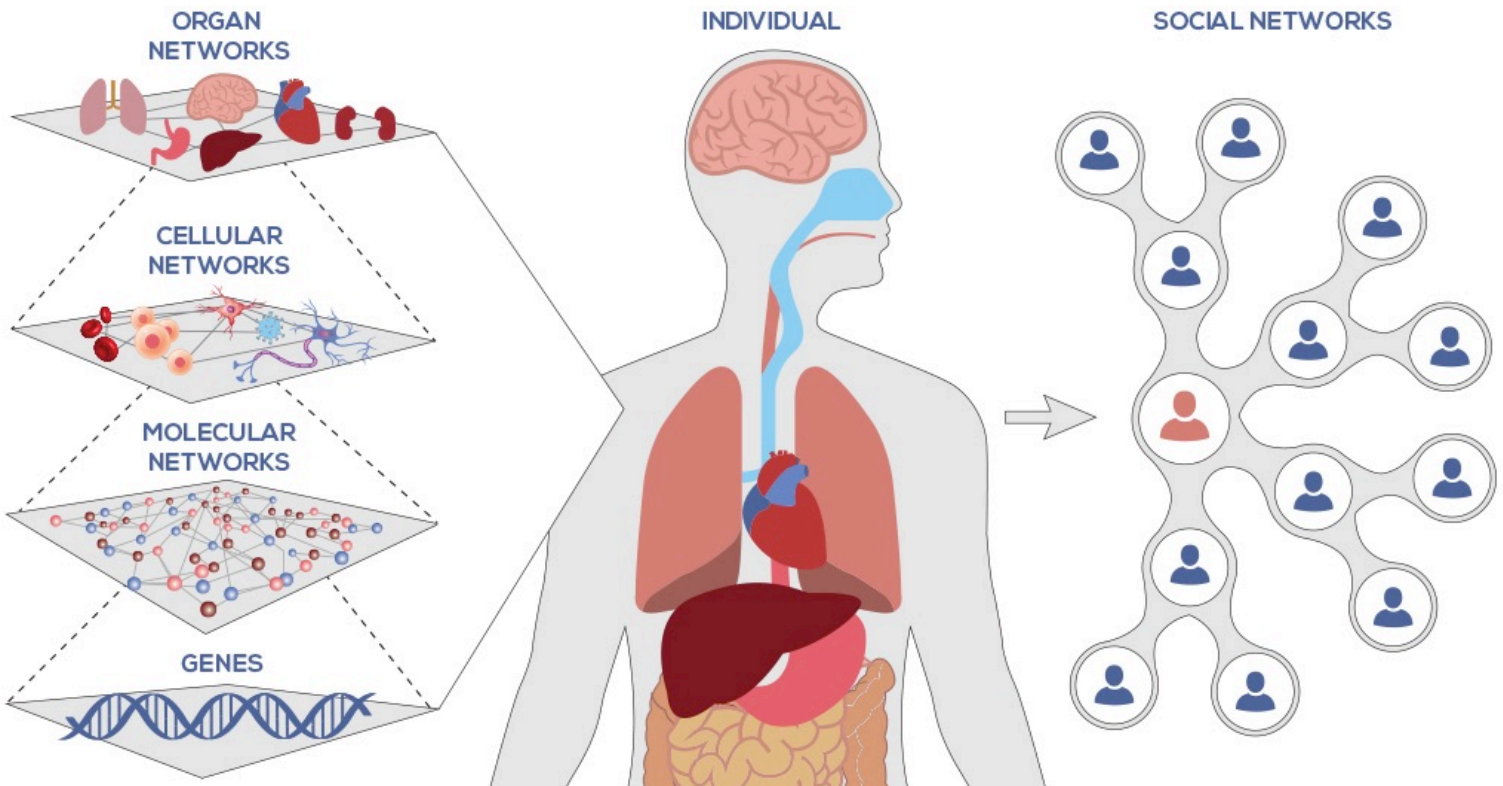
Personal, Dense, Dynamic (Longitudinal) Data Clouds (Big Data)



These data clouds provide insights into wellness and disease and provide the essence of what "Precision Medicine" should be

Institute for Systems Biology

The Network of Networks is Hierarchal in Nature

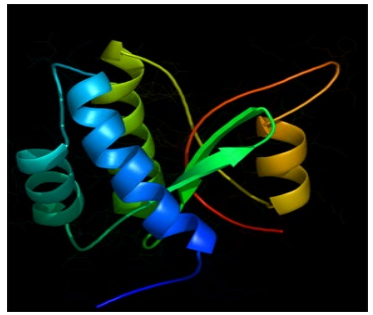
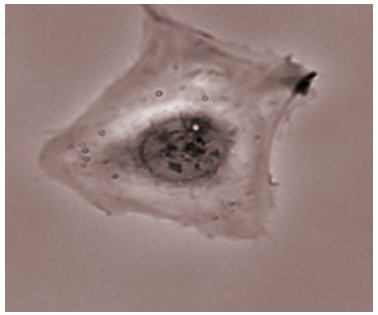
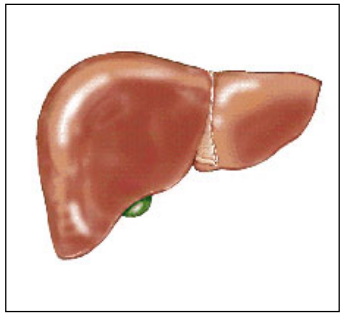


Human Biological Information Is Quantized and Hierarchical

Analysis of single molecules, single cells, single organ and single individuals—quantized units of information

Technologies for measuring each quantized unit are needed

Biology must be attacked at the level of each single quantized unit follow by an "information integration" of their data types

Single molecule	Single cell	Single organ	Single individual
			
3rd generation DNA sequencing	Analyze 40 proteins In each of 10,000 T cells	Organ-specific Blood proteins	Personal, dense, dynamic Data clouds

Systems-Driven Technologies and Strategies

Technologies

- 3rd generation DNA sequencing ($100 genome)—single molecule sequencing
- Targeted and SWATH proteomics (blood biomarkers)—SRM Atlas
- Peptide protein-capture agents (replace antibodies as diagnostics and drugs)
- Single-cell analyses (deciphering biological complexities)
- Digitalized measurements of many features of self

Strategies

- Family genome sequencing (identify disease genes and compare 1000s of genomes)
- Animal model disease dynamics (identify earliest disease-perturbed networks)
- MS-based proteomics blood biomarker (protein) discovery (cancer, preterm birth, PTSD, liver disease)
- Organ-specific blood proteins assess health of many organs simultaneously
- Deep phenotyping--dense, dynamic, personal data clouds to analyze wellness and disease
- Use analysis of disease-perturbed networks to identify drug target candidates
- Synthetic biology for new drug generation and high throughput screening
- Blood is a window into the dynamics of human biology and disease (separate/analyze molecules of blood, vesicles, cells)

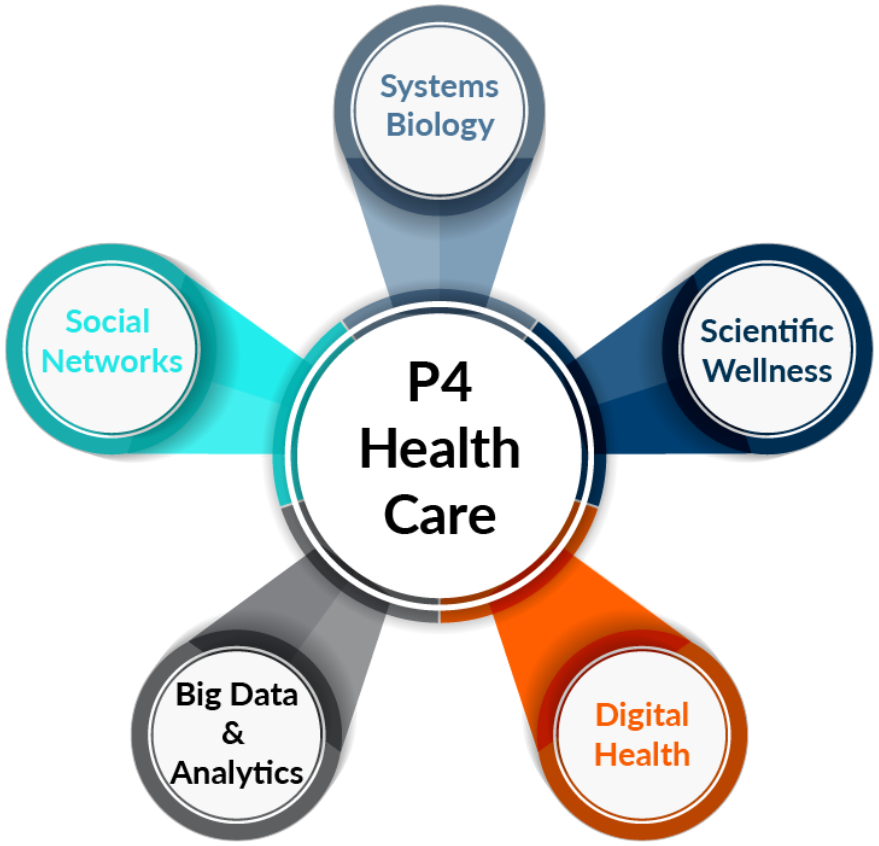
Analytics

- AI—expert systems Newco
- Machine learning
- Pattern recognition
- Imaging
- Integration

The Emergence of P4 Medicine in 2014

Predictive, Preventive, Personalize, Participatory

Converging Megatrends



P4 Medicine

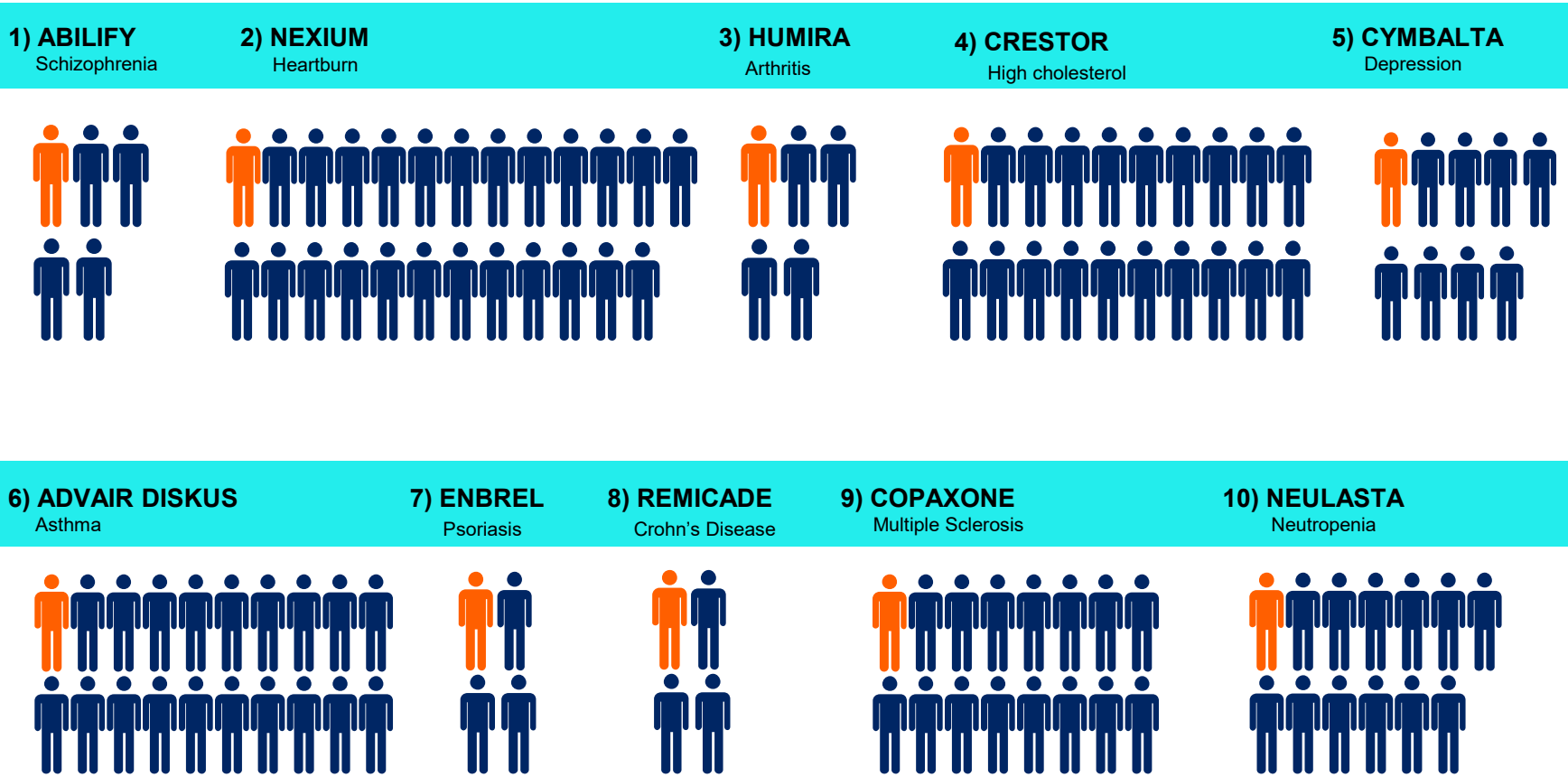
- Proactive
- Individual N=1 medicine
- Wellness & Disease
- Deep phenotyping and personalized data clouds
- Personalized data clouds for clinical trials (N=1 experiments)

Contemporary Medicine

- Reactive
- Population
- Primarily Disease
- Averaged patient populations
- Averaged patient populations for clinical trials

Imprecision Medicine:

Time for N=1 Drug Trials to Stratify Disease Subtypes, Responders and Toxicities



For every person in the US that the 10 highest grossing drugs do help (orange), they fail to improve the conditions of between 3 - 24 people (blue).

Schork, Nicholas. Time for one-person trials. Nature. Vol 520. April 2015

Genomics and Deep Phenotyping

2014- The 108 Person Scientific Wellness Pilot Project

(Pioneers)

Principal Investigators:
Lee Hood and Nathan Price

Using longitudinal deep phenotyping and genomics

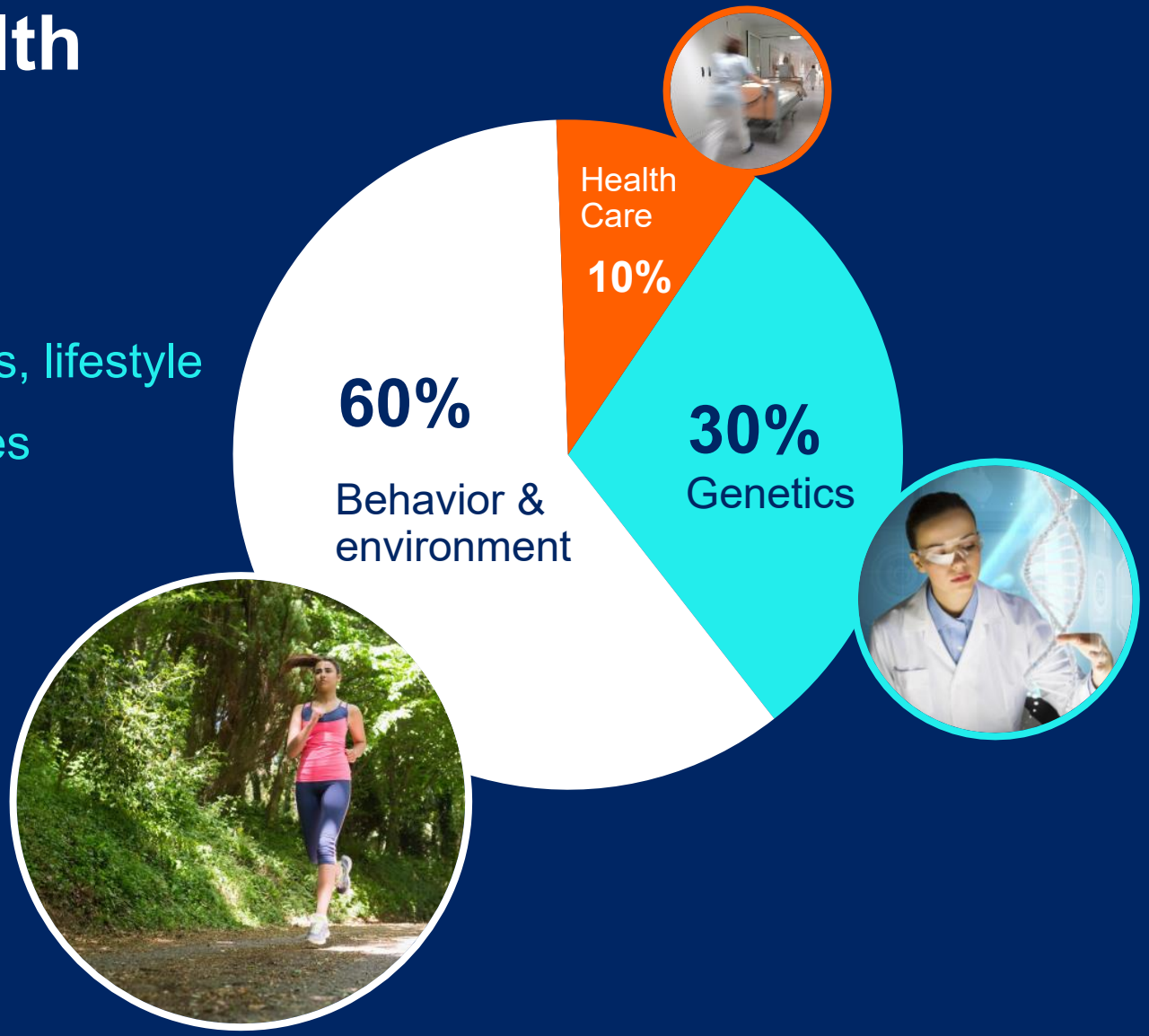
IRB approved study: **Price, Magis, Earls, Hood et al, Nature Biotechnology, 2017**

Determinants of Health in the U.S.

Deep phenotyping will assess

the integration of individual genetics, lifestyle

and adverse environment exposures



Deep Phenotyping of 108 Scientific Wellness Pioneers

Creating dense, dynamic, personal data clouds

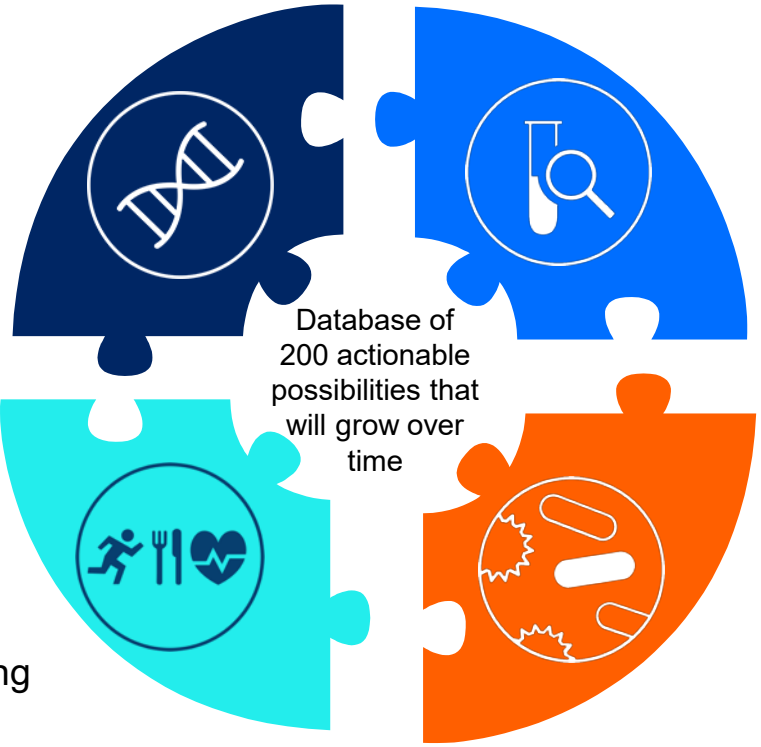
Price, Magis, Earls…Hood,
Nature Biotechnology, 2017

GENOME

Whole Genome Sequencing
SNPs Millions

LABS

Detailed lab tests 3x (blood, urine, saliva)
Clinical chem. 150
Metabolites 700
Proteins 400
Almost 1200 analytes



Database of 200 actionable possibilities that will grow over time

SELF-TRACKING

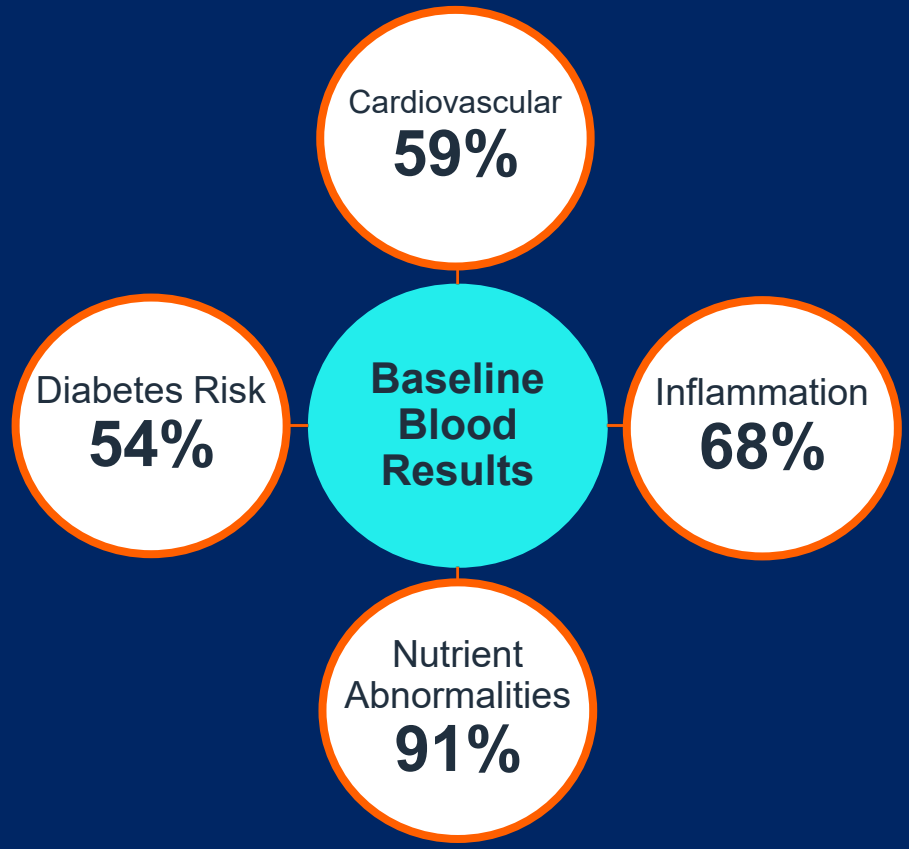
Continual self-tracking of physiology & lifestyle monitoring

MICROBIOME

Gut Microbiome
Ration of microbial species 3x

Initial Clinical Labs Discovery:
High Rate of Actionable Clinical Results

- The 108 “well” participants had a high rate of initial abnormal lab results
- **100%** of the participants had multiple actionable recommendations from their blood results



Some of Lee Hood's Actionable Possibilities

- Weight—lost 20 pounds—within 5 lbs college football weight\
- Exercise—balanced and more extensive—resting pulse rate changed from about 55/min to 41/min
- Continuous glucose monitoring—2 weeks—N=1 perturbations—optimize diet
- Intermittent fasting—lower blood glucose
- Carotid artery ultrasound analysis—detect athrosclerosis
- Assessment of distribution of body fat—ultra sound imaging
- On statin--complications—muscle atrophy, diabetes (genetic markers indicate susceptibility)
- Corrected 5 nutritional deficiencies—supplements and vitamins
- Vitamin D extremely low—need mega-doses—genome vitamin D uptake blocking variants
- High mercury (eliminate tuna sushi)
- Inflammation—control with diet
- Biological age is 15 years younger than chronological age and decreasing
- **Realized wellness is my responsibility; acquired deep insights about my personal wellness and act on them (participatory)**

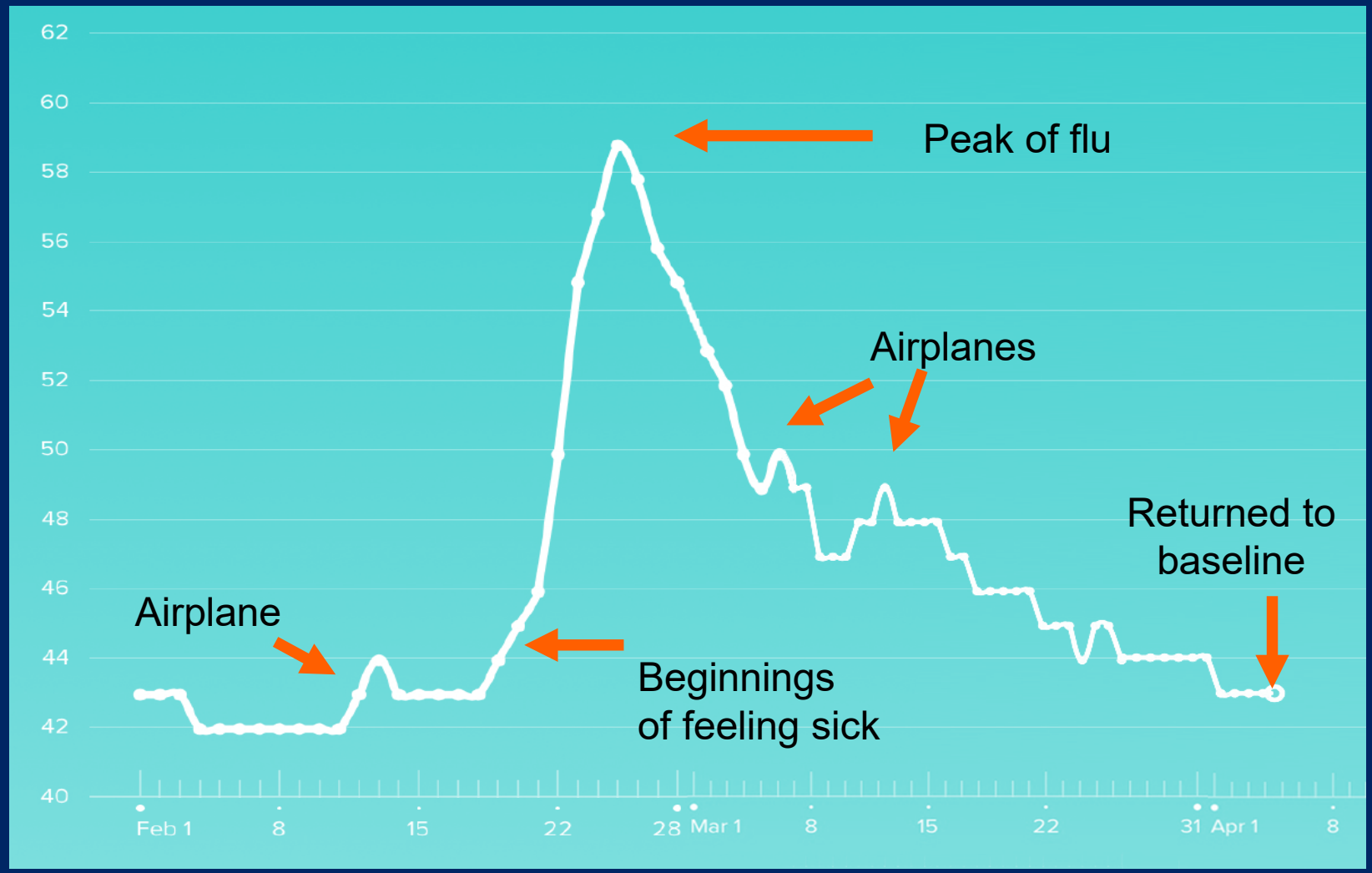
Lee Hood Pulse During a Bout of the Flu

Digital Health

- Pulse
- Voice
- Typing
- Blood—1200 analytes
- Saliva—n analytes
- Urine –m analytes
- CSF—x analytes
- Electric signals: brain, heart
- Sleep

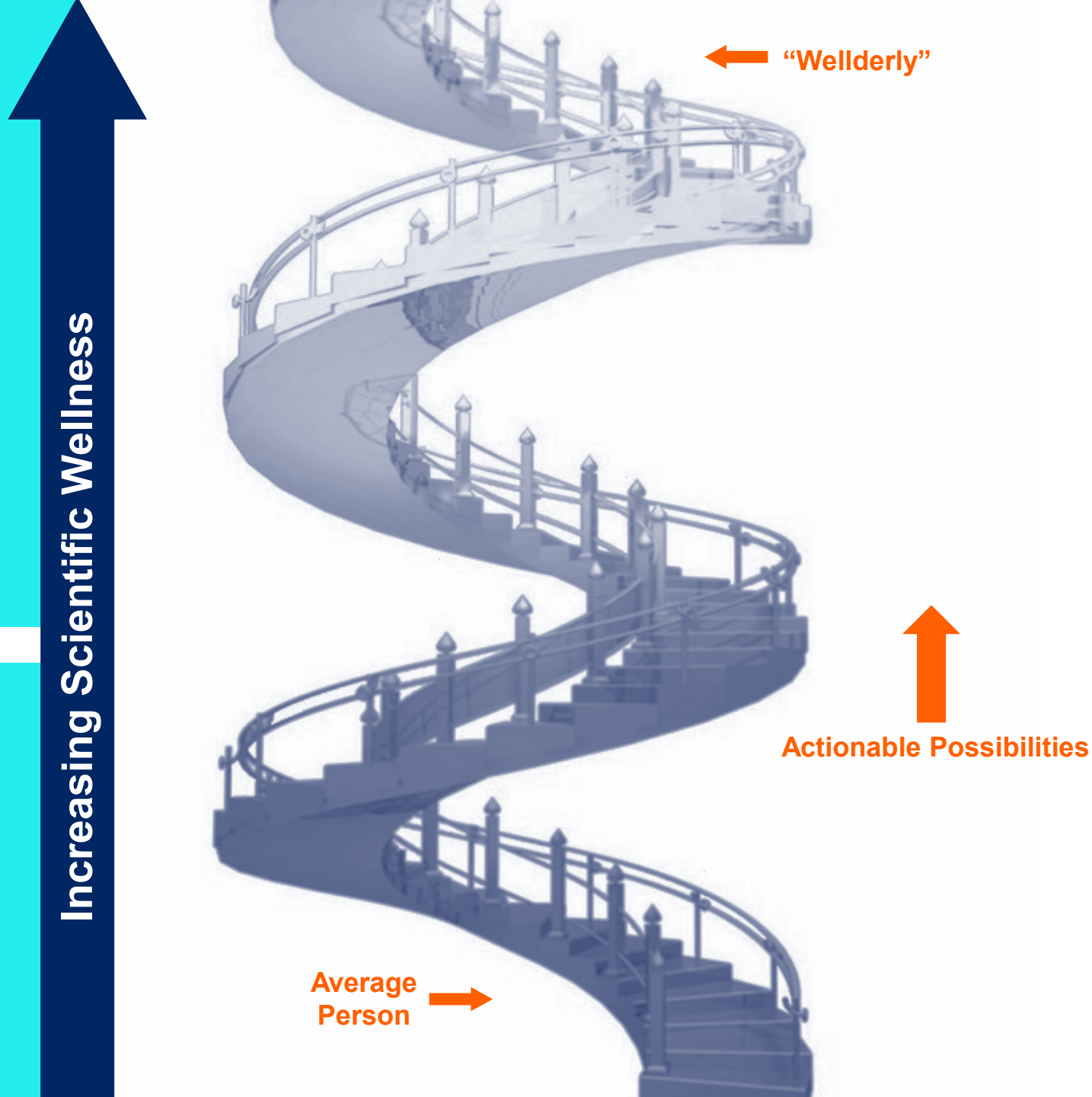
Heart rate variability

Cognitive brain assays and exercises



Where Do You Reside On The Wellness Staircase?

Scientific wellness is a life long journey to healthy aging—mentally physically active—90s



Creation of a Consumer-Based Scientific Wellness Company

Arivale

Your scientific path to wellness

2015 LAUNCH

6000 clients

100 wellness to disease Transitions

2019 Shut Down

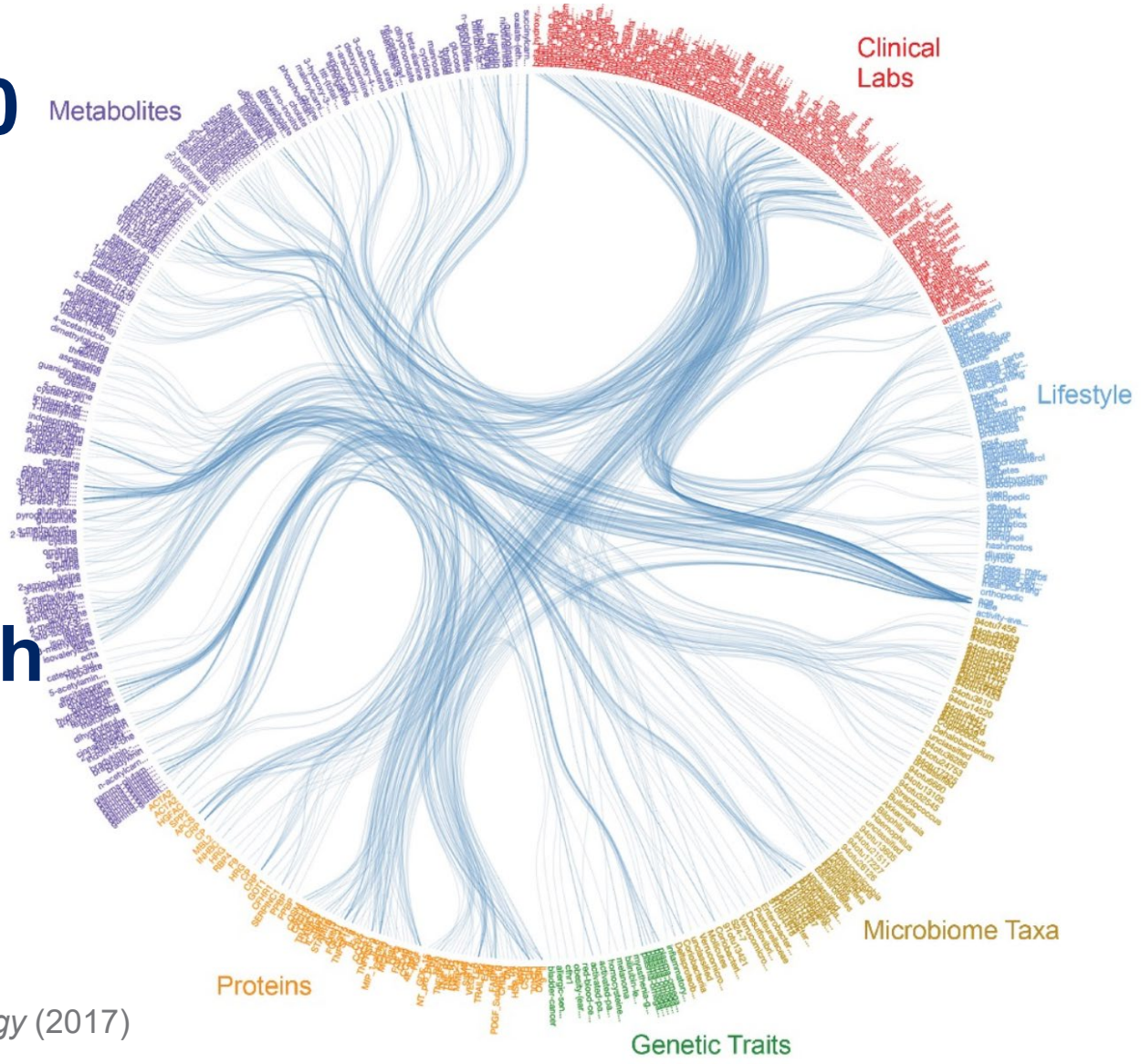
P4 Medicine Clinic--Seattle

Personal Dense, Dynamic Data Clouds: Probing the Dark Matter of Wellness And Disease

The Hubble Telescope allows us to probe the dark matter of the universe just as dense and dynamic personal data clouds allow us to probe the dark matter of human biology and disease.

Statistical Correlations

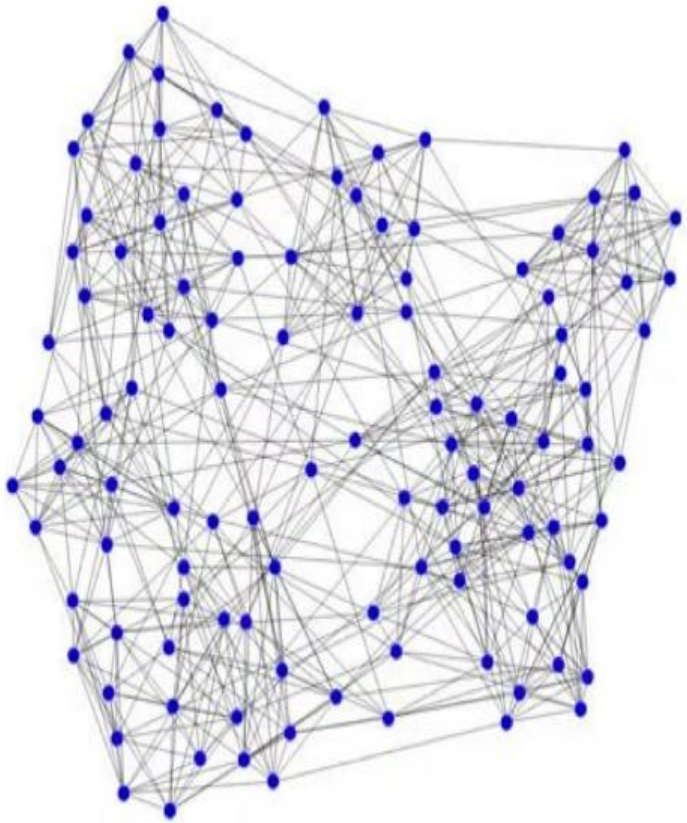
Deriving Insights from Data: New Frontiers—3500 Statistical Correlations with 1200 analytes

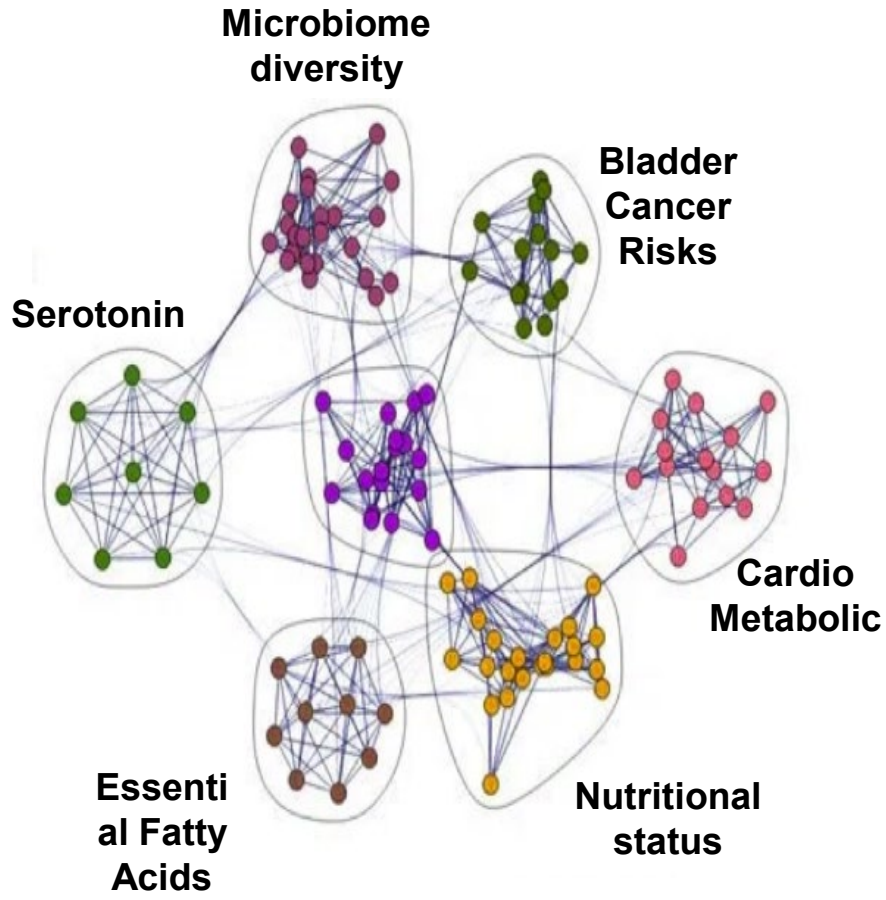


– 35,000 Correlations with 5000 More Proteins

Price, Magis, Earls , Hood, et al, Nature Biotechnology (2017)

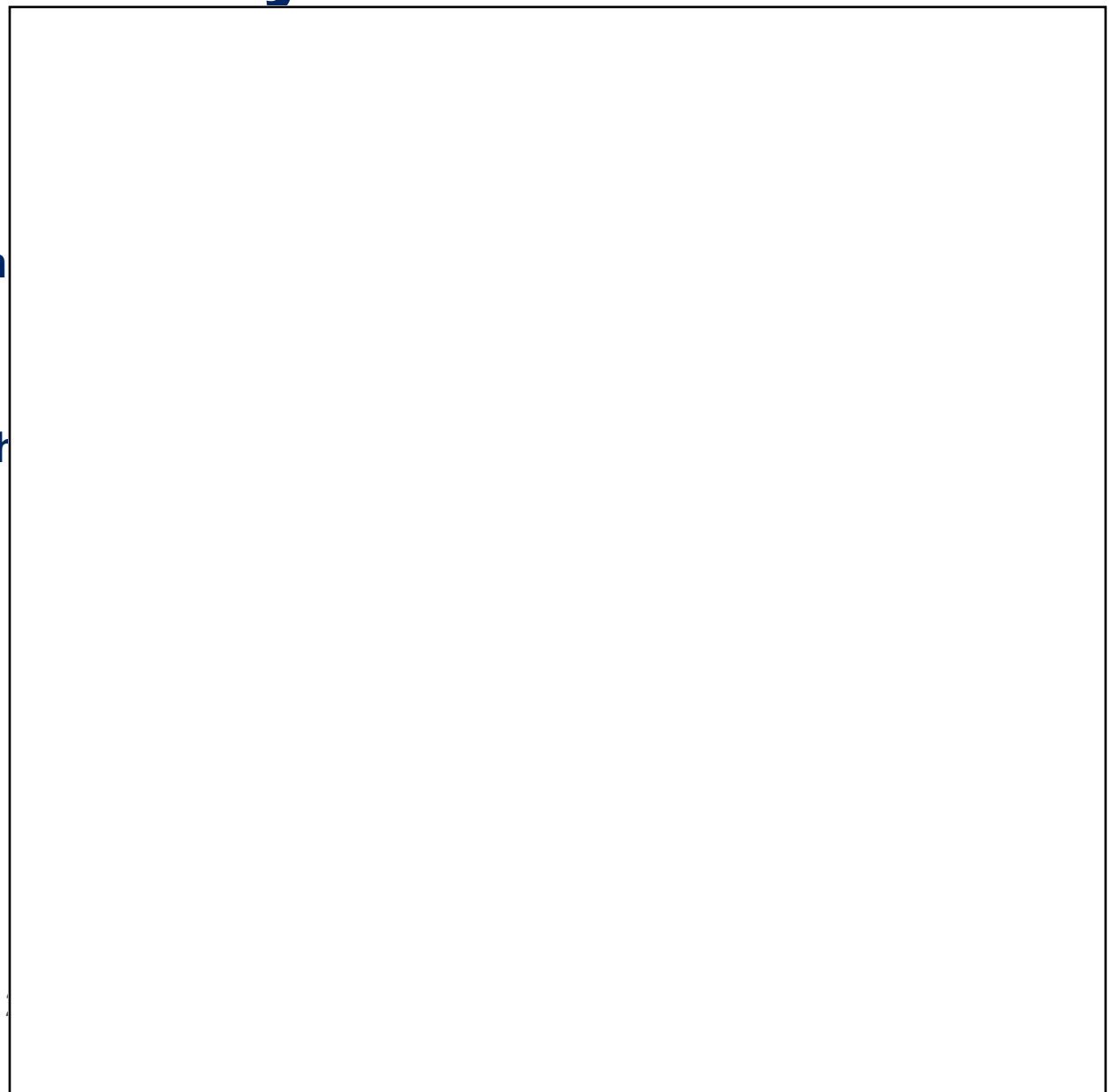
Identification of 70 multi-omic functional communities (modules) in the correlation network.





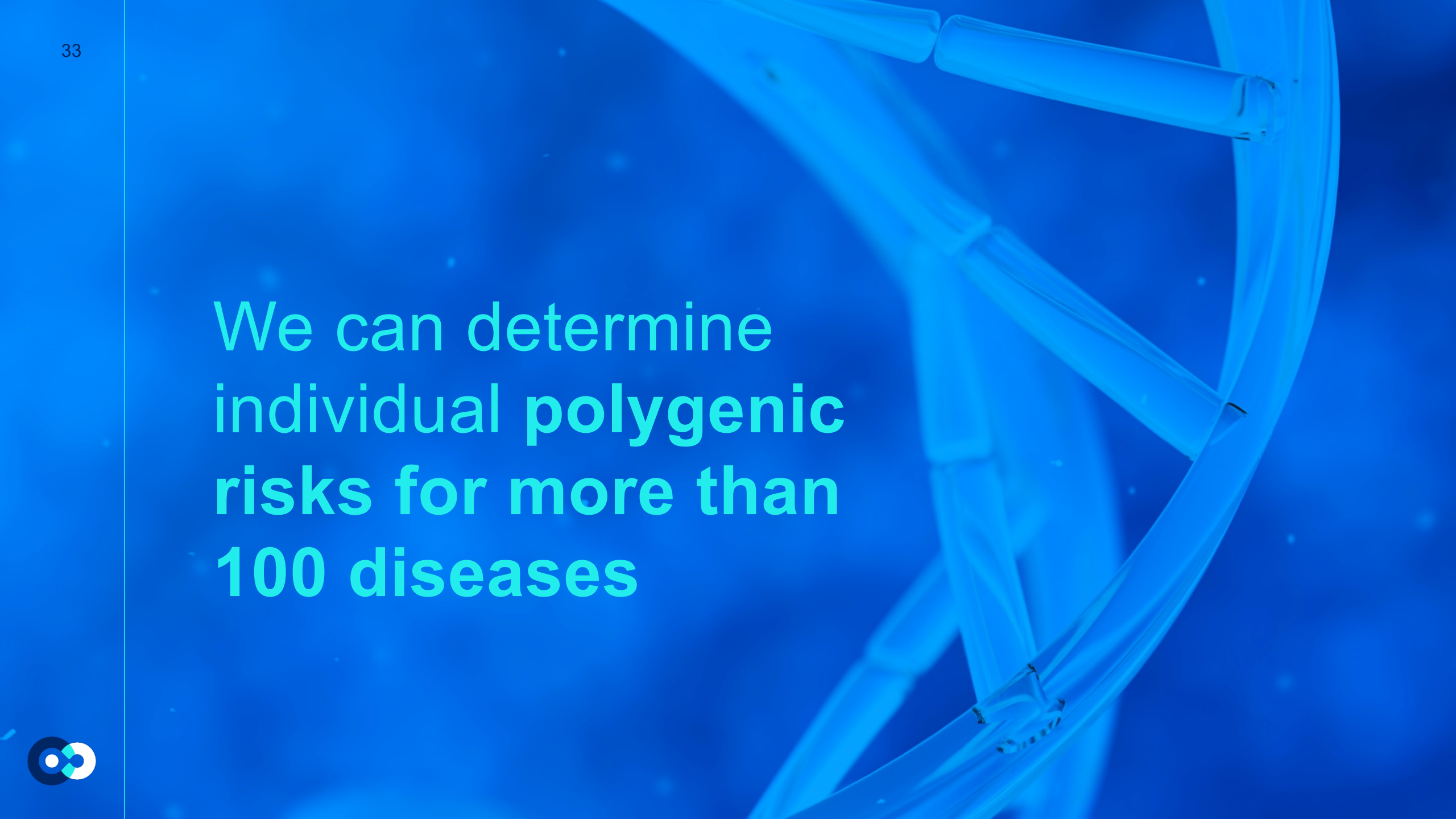
Total cholesterol community as one of 70 communities

- Cholesterol is **positively associated with alpha-tocopherol** (Vitamin E)
- Cholesterol is **negatively associated with endogenous thyroxine**
- A beneficial side effect of **the drug thryroxine** (Synthroid) is lowering LDL cholesterol



Price, Magis, Earls, Hood et al, *Nature Biotechnology*,

We can determine individual **polygenic risks for more than 100 diseases**



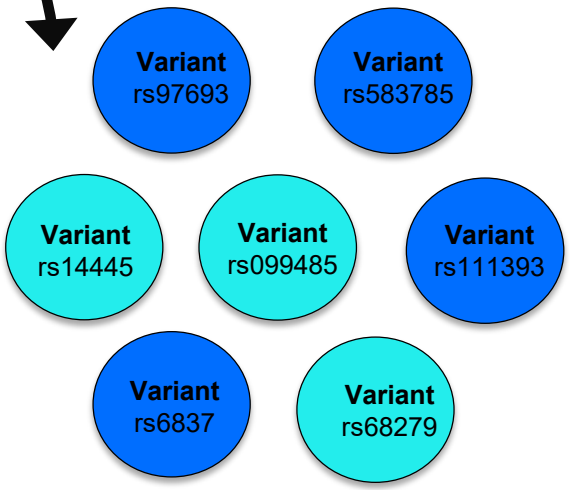
Calculation of 127 Polygenic Scores from WGS and GWAS Data

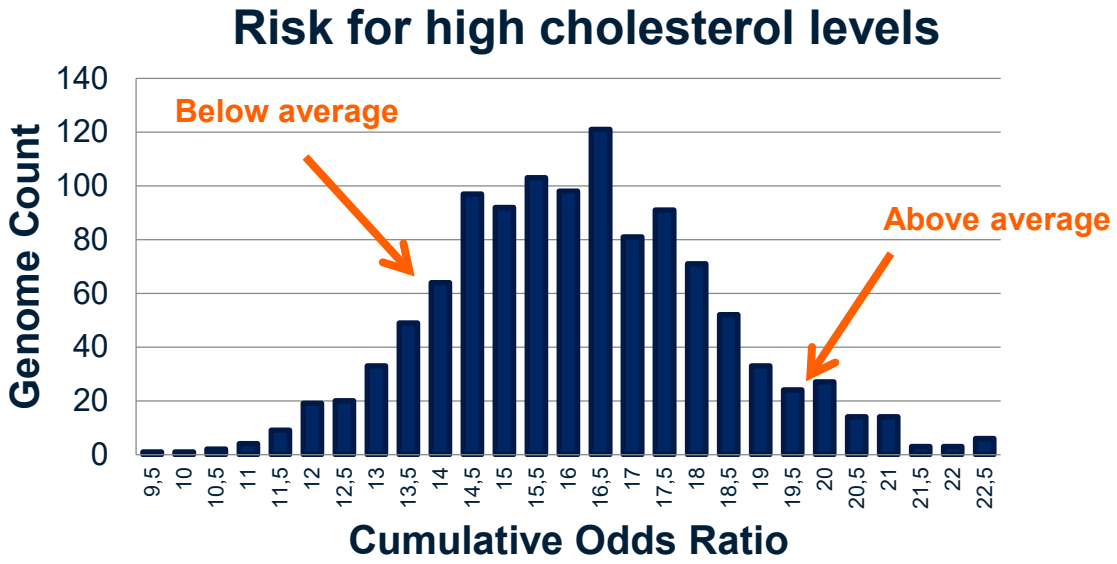
Phenotype	$N_{samples}$	$N_{variants}$	Reference
Asthma	35,083	8	Hirota et al. 2011
Bilirubin levels	9,937	14	Kang et al. 2010
Bladder cancer	65,308	8	Rothman et al. 2010
Body Mass Index	249,796	32	Speliotes et al. 2010
Coronary Artery Disease	109,124	23	Dichgans et al. 2013
Inflammatory Bowel Disease	77,064	110	Jostins et al. 2012
LDL cholesterol	188,577	56	Willer et al. 2013
Omega 6 PUFAs	8,631	8	Guan et al. 2014
Type 2 Diabetes	187,590	61	Mahajan et al. 2014

1. Welter, D. et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*, *42*(Database issue), D1001–6.

Cumulative sum for an individual can give an estimate of risk relative to a population of 2000 normal individual genomes

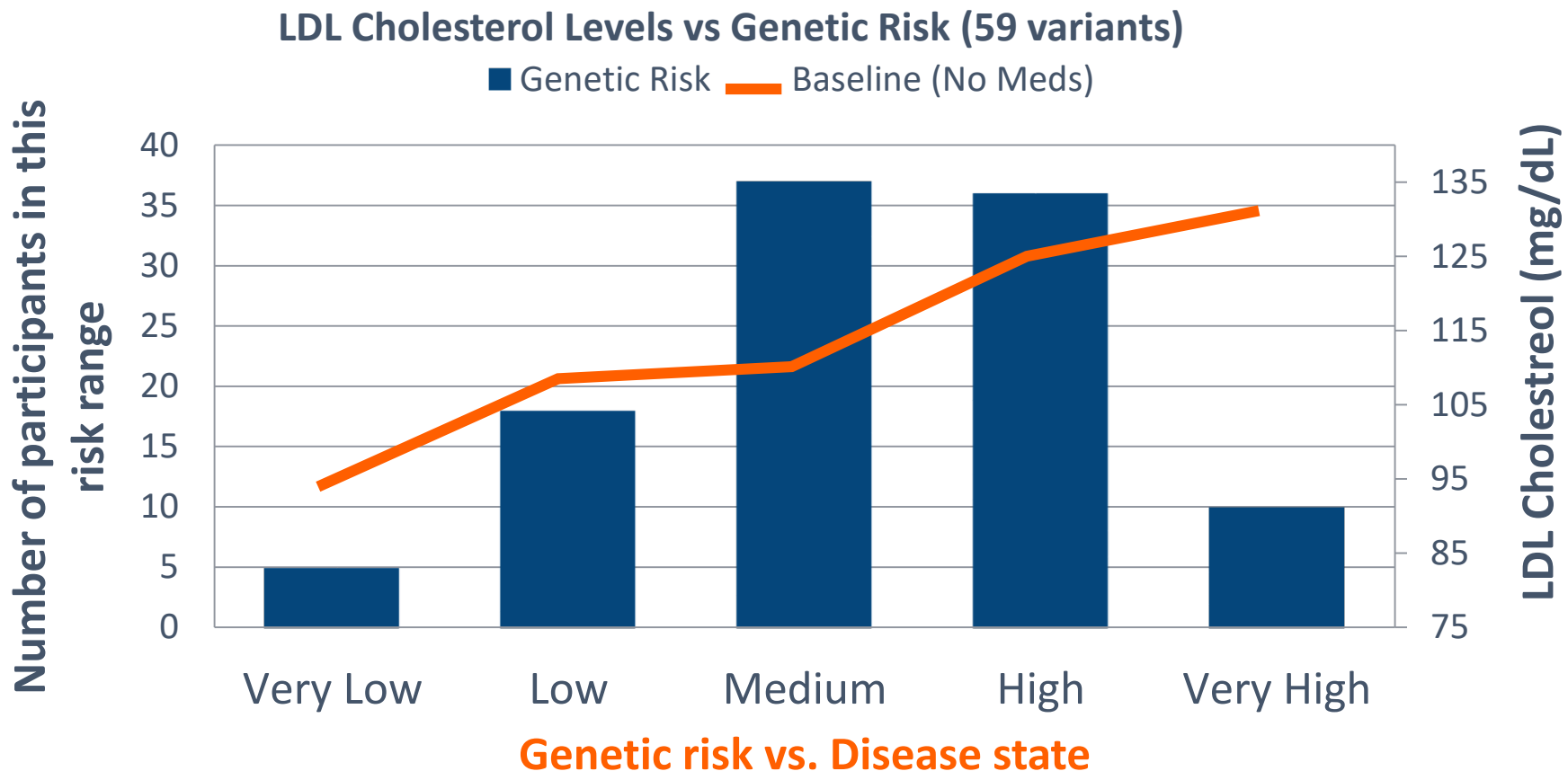
Calculate the effects of all common variants within an individual



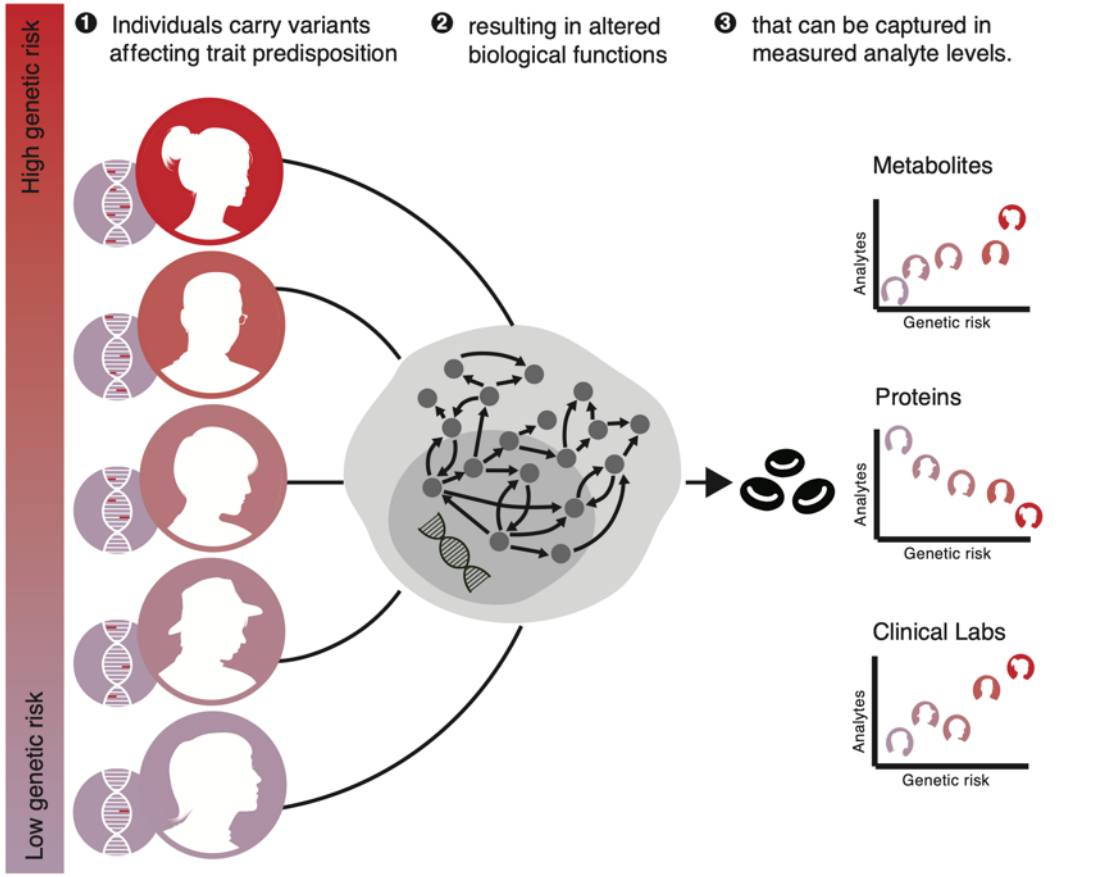


Estimate the risk for high cholesterol relative to a population of 2000 normals.

LDL cholesterol in Participants Shows Monotonic Relationship with ‘Genetic Risk’



Reflections of polygenic risks in the blood

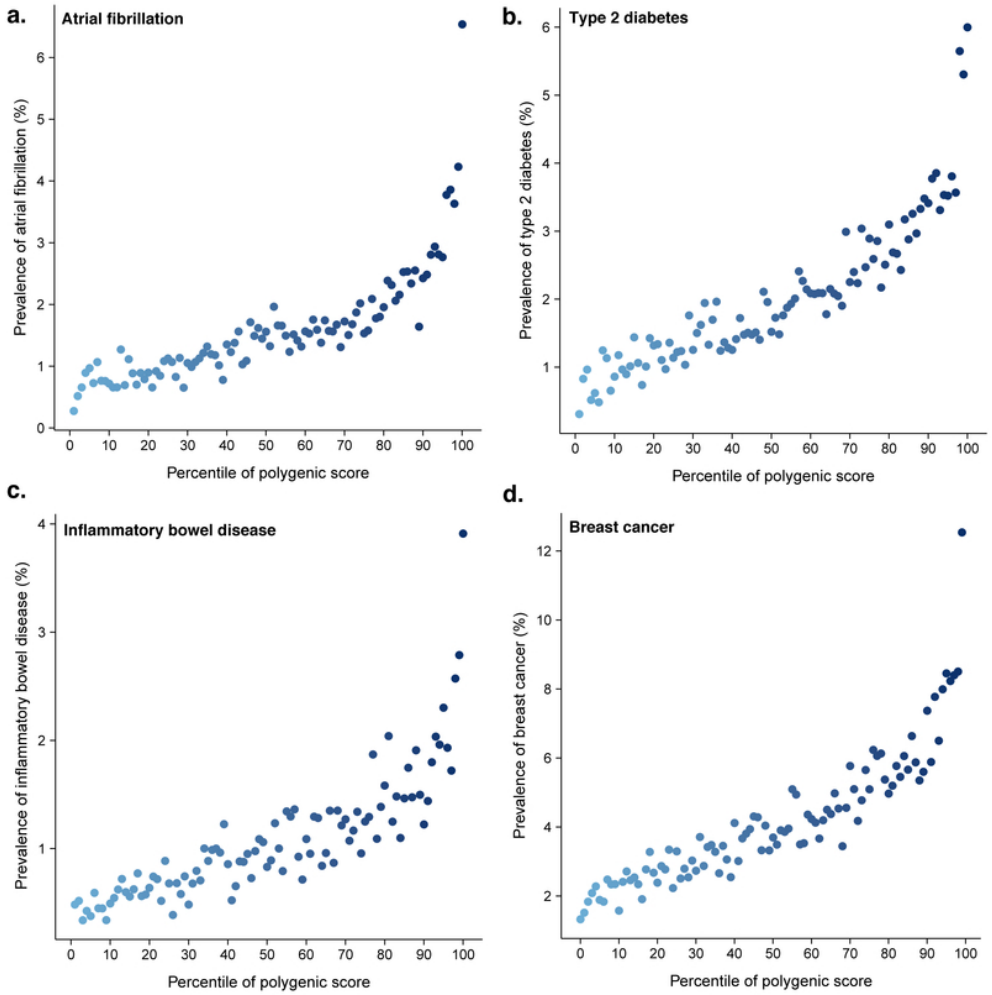


The multi-omic blood manifestations of genetic disease risk

Michael Wainberg, Andrew T. Magis, John C. Earls, Jennifer C. Lovejoy, Nasa A. Sinnott-Armstrong, Gilbert S. Omenn, Leroy Hood*, Nathan D. Price*

*Nature, submitted.

Polygenic risk scores (PRS) predict genetic risk as a weighted sum of risk alleles found by genome-wide association studies (GWAS)



Summary

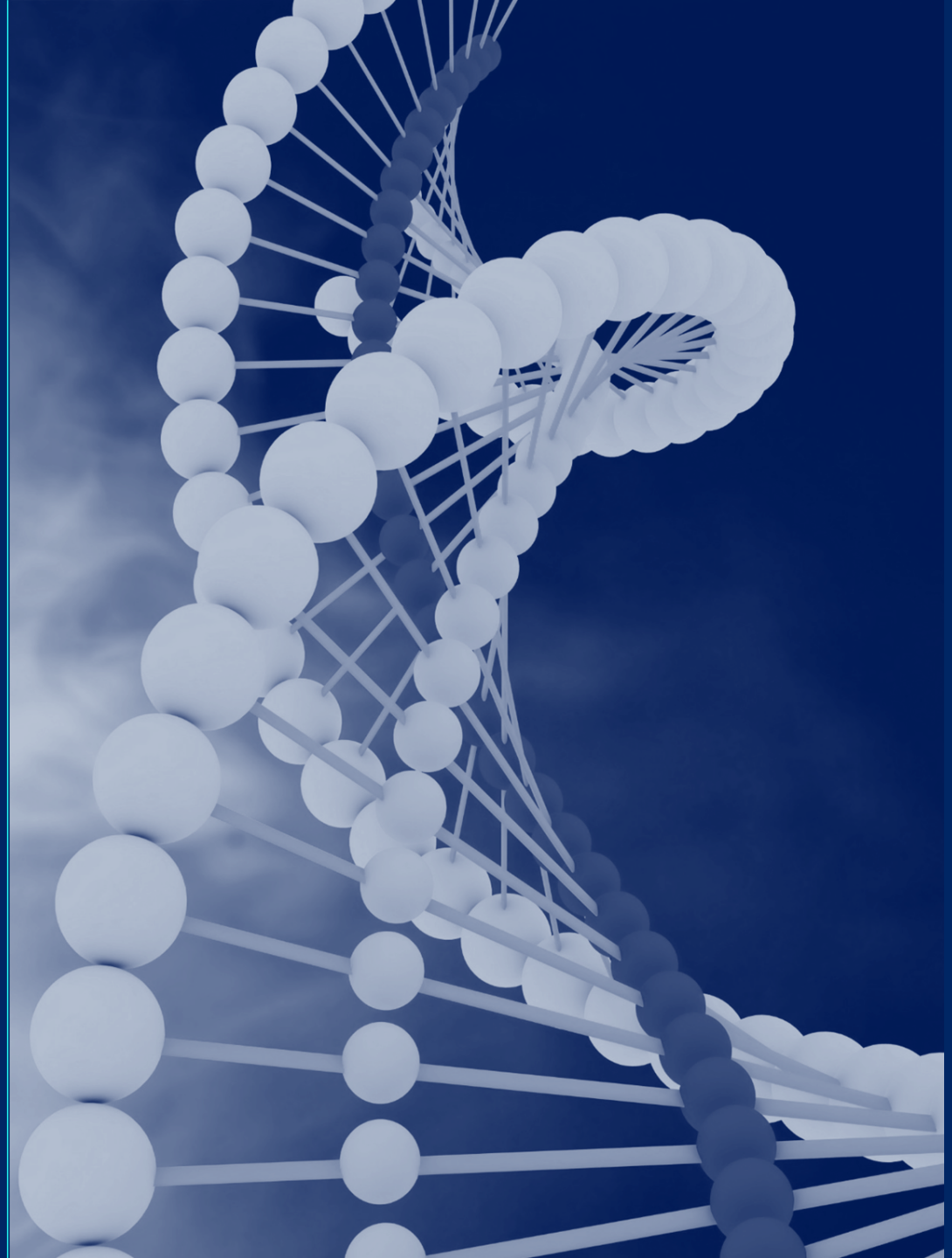
Genetic risk for **54 traits is associated with 858 detectable alterations** in plasma proteomic, metabolomic and clinical laboratory measurements.

These alterations provide insights into molecular pathophysiology, and some suggest **therapeutic strategies** and protection from aging.

Individuals at high genetic risk for a **trait display dysregulation** in many of the same analytes that are **dysregulated in frank disease**, and this signature of dysregulation is frequently detectable in the blood.

Our results emphasize that genetic risk scores, far from being a mere statistical tool for disease risk stratification, also **reflect underlying disease biology—**treat at this stage? See clearly initial disease stages.

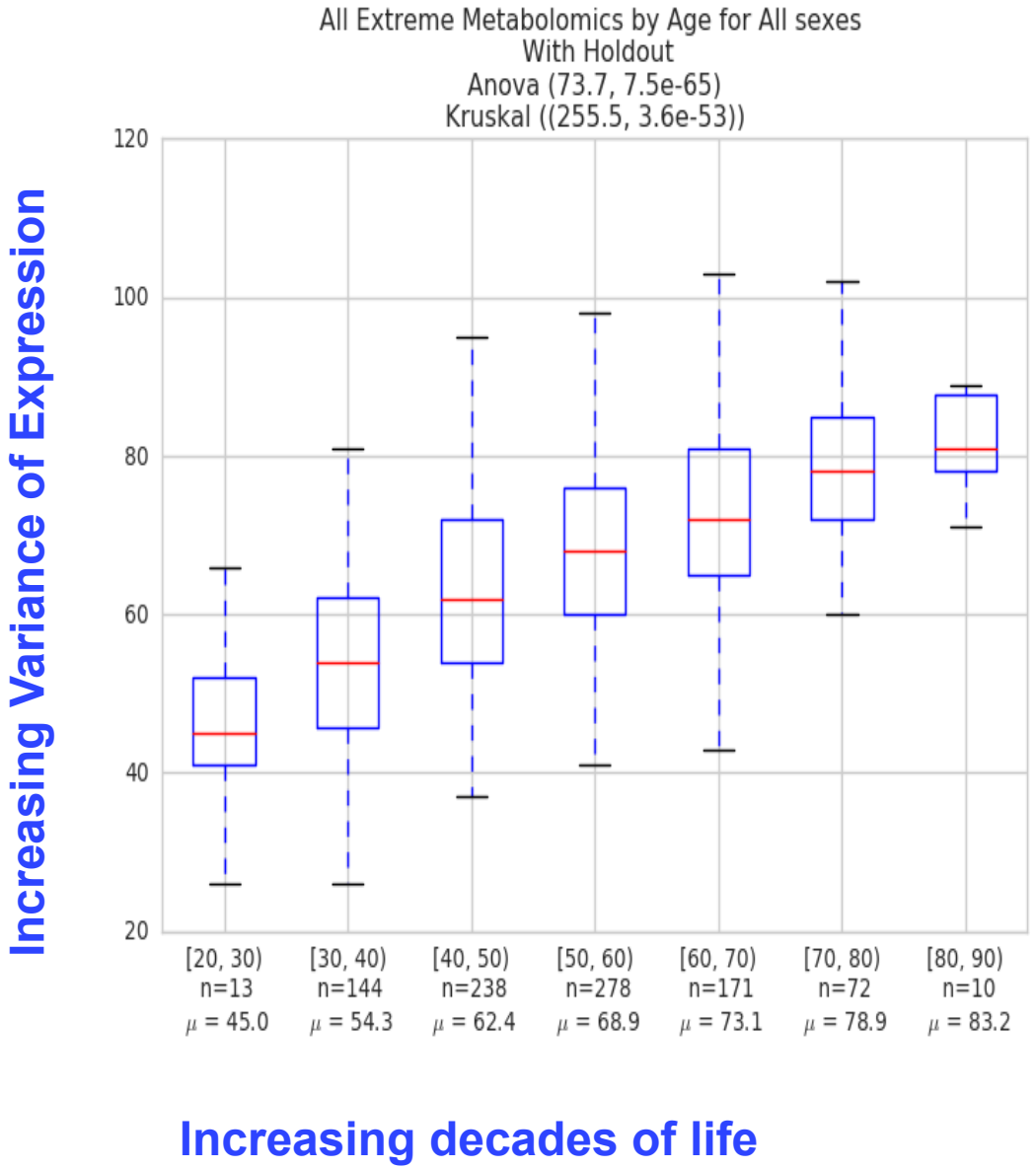
Using **genetic risk as a proxy for prodromal disease** can substantially broaden the insights gained from population-scale multi-omic cohorts. Learn about disease from wellness populations.



Analyte Expression Levels Diverge with Age and Permit the Determination of Your Biological Age

The age your body says you are rather than your birthdate—your chronological age

Loss of Control of Analyte Expression with Increasing Age



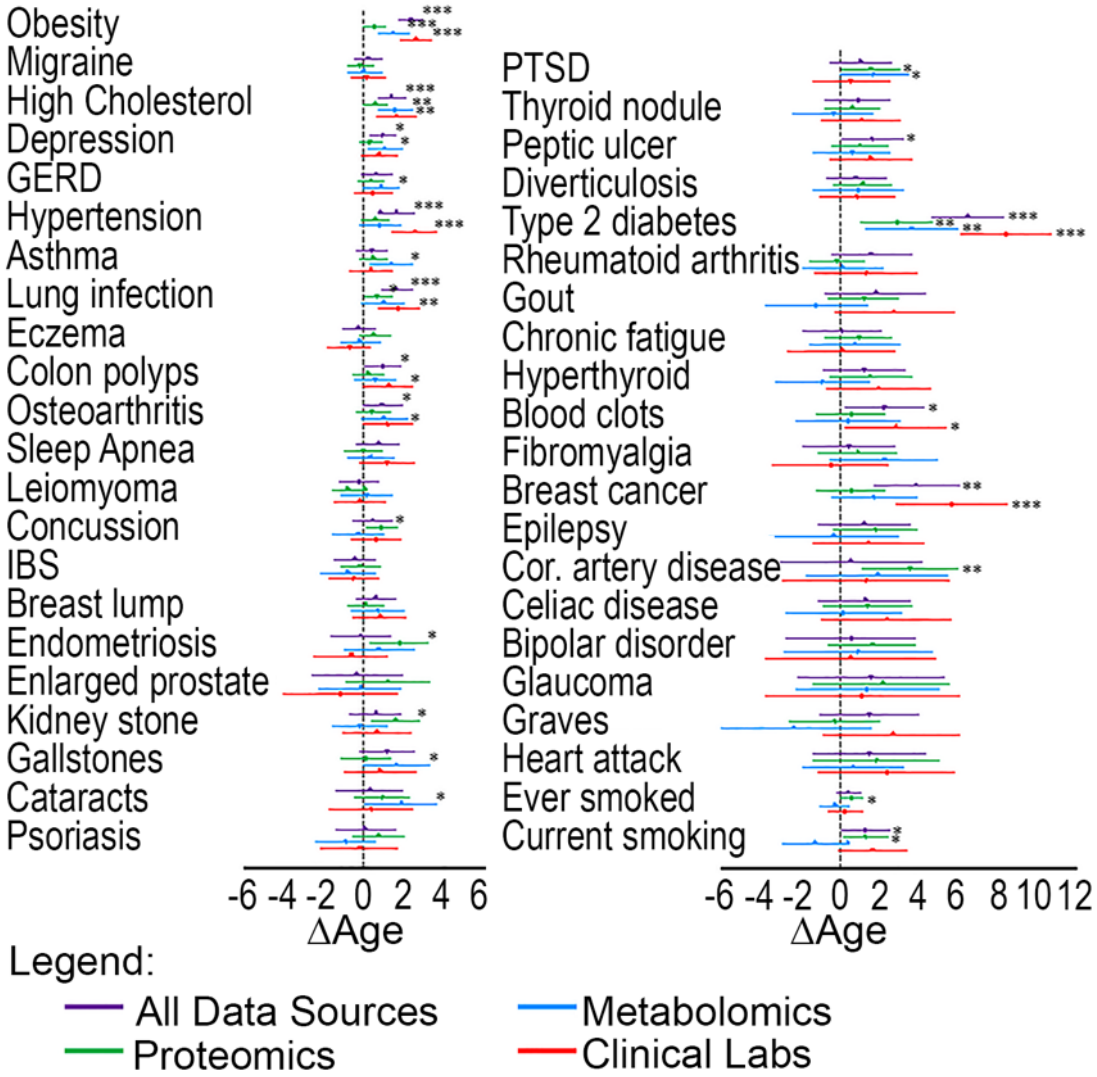


Figure and analysis by Laura Heath

Multi-omic biological age estimation and its correlation with wellness and disease phenotypes: A longitudinal study of 3558 individuals

John C. Earls, Noa Rappaport, Laura Heath, Tomasz Wilmanski, Andrew T. Magis, Nicholas J. Schork, Gilbert S. Omenn, Jennifer Lovejoy, Leroy Hood, Nathan D. Price

Journal of Gerontology

in press.

Employing Scientific Wellness for Individuals is Key to Healthy Aging

More than half of all children born in 2007 in developed countries can expect to celebrate their 100th birthday.

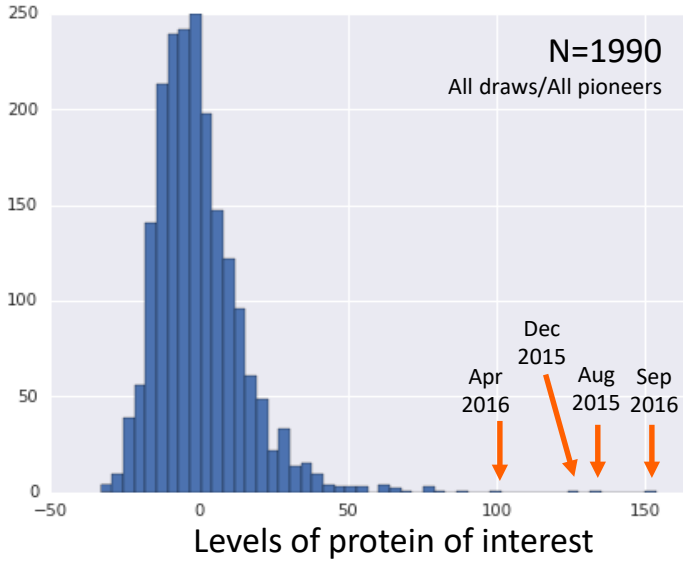
Christensen, Ageing Populations: The Challenges Ahead, Lancet , 2009

Microbiome to blood

- 11 metabolites allow us to predict the alpha diversity of the gut microbiome
- T. Wilmanski, et. al Nature Biotechnology Sept. 2019


ISB
State Transitions :
Wellness to Disease

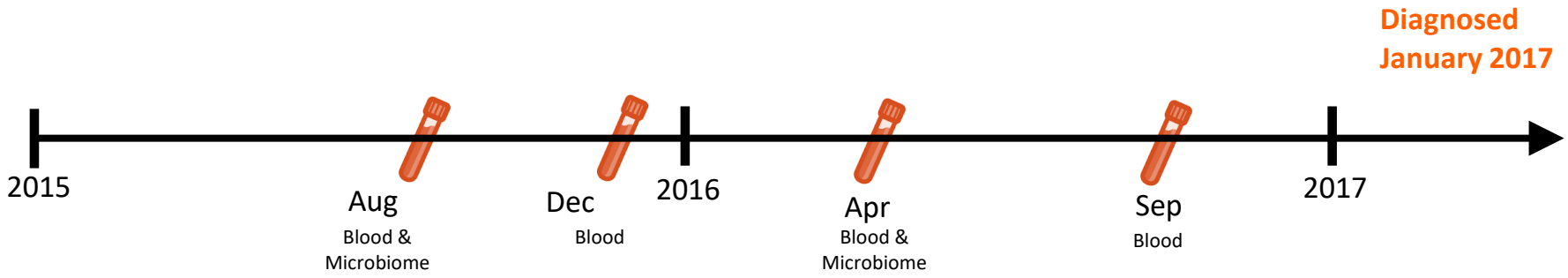
Example – identifying outlier for pioneer with pancreatic cancer



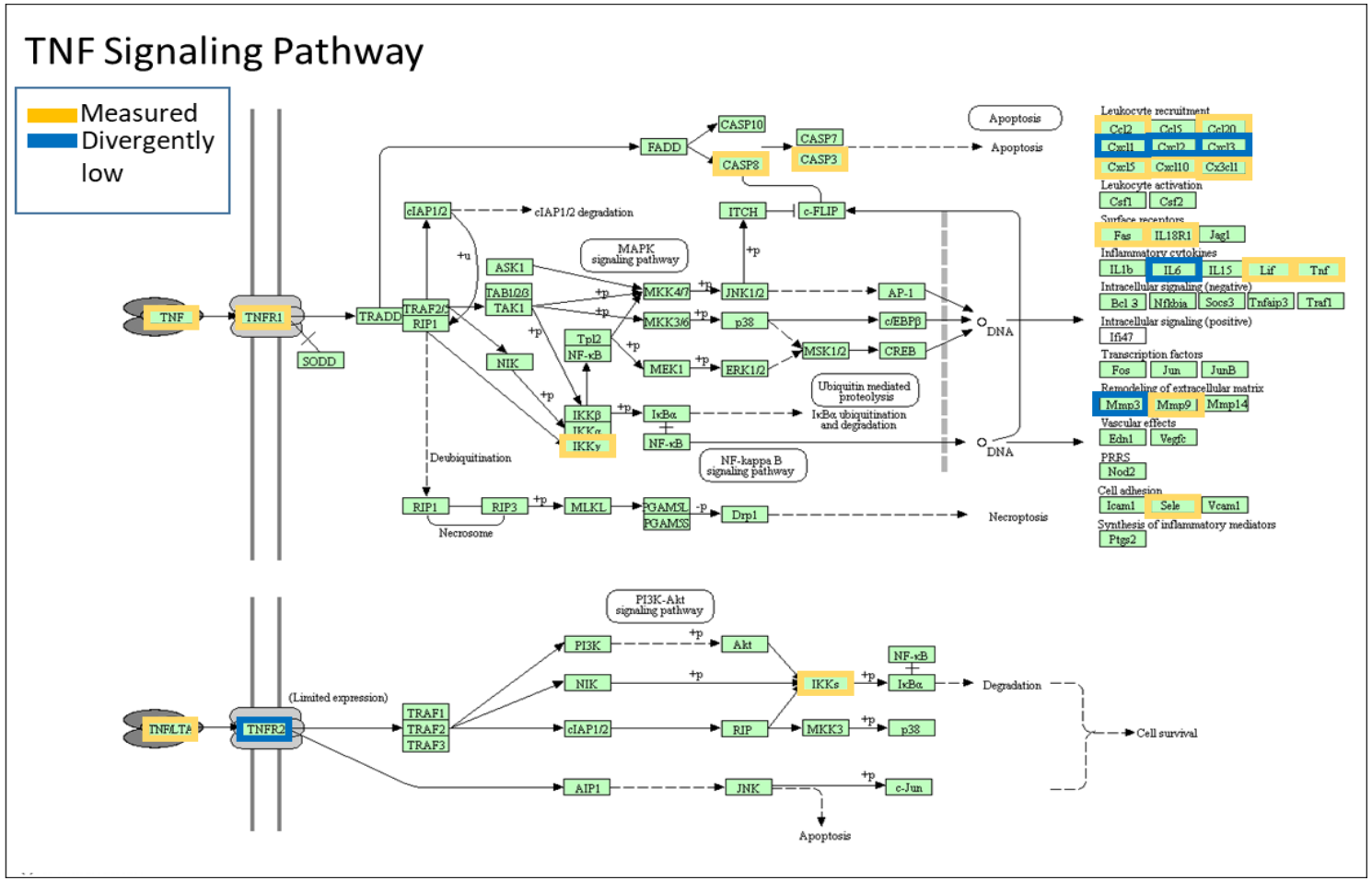
Over the course of the four blood draws, this pioneer has **consistently** been the **largest outlier for the protein of interest.**

A 2010 study by Bert Vogelstein estimates at least 5 years between parental, non-metastatic founder cell and metastatic ability in pancreatic cancer[1].

1. Yachida, S et al. (2010). Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*, *467*(7319), 1114–1117.



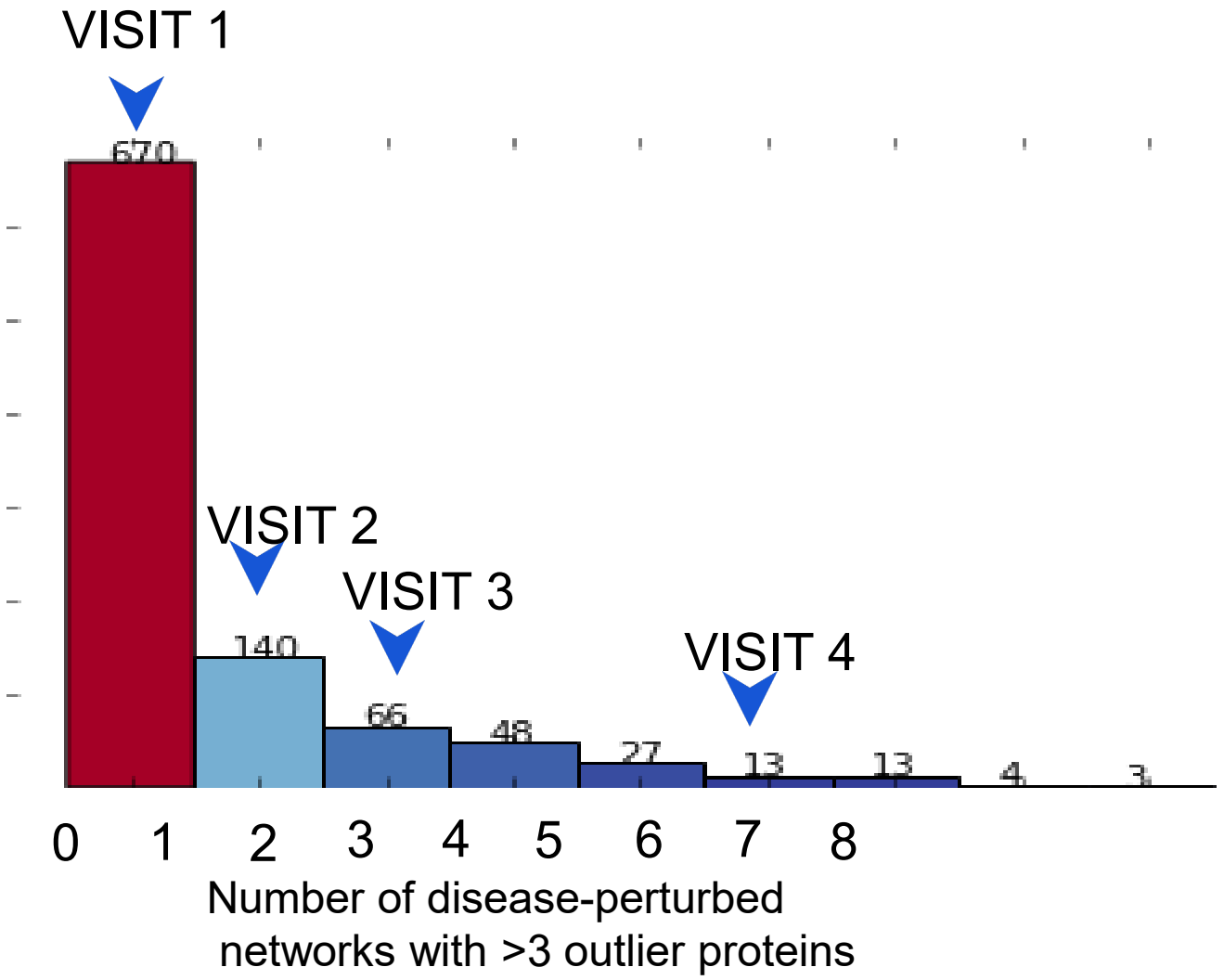
Identifying disease-perturbed networks pre-diagnosis: Early example for pancreatic cancer

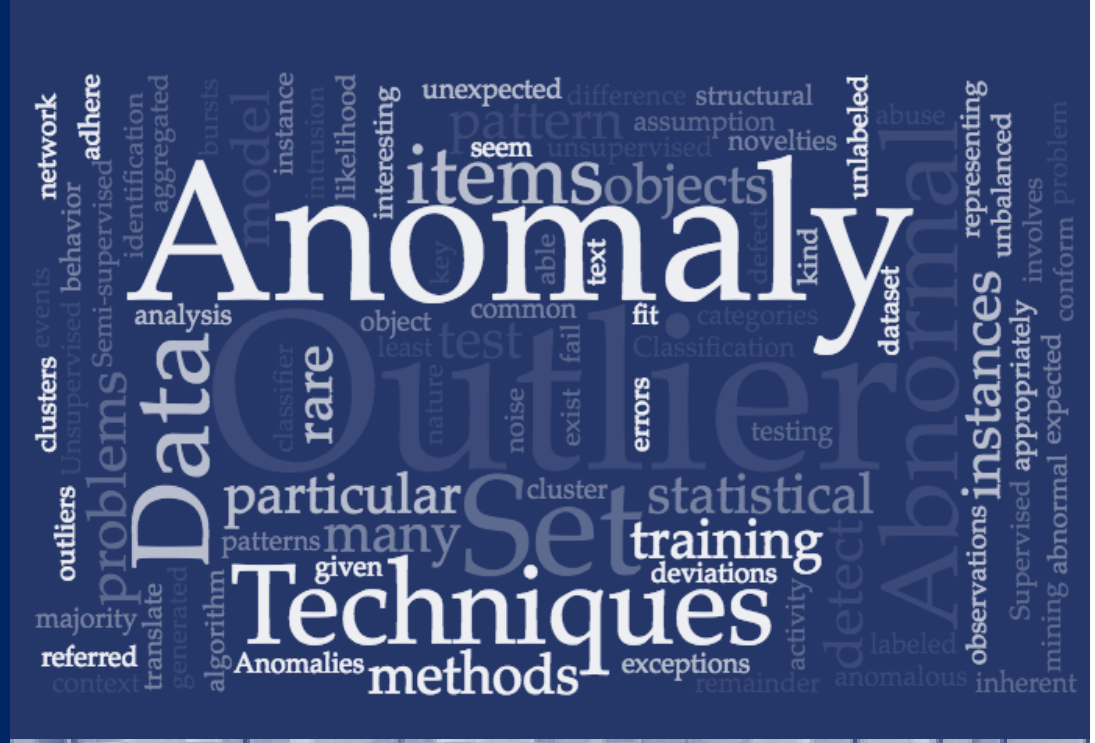


Distribution of # of significantly low p-values

948 clients/how many of 9 potential cancer-related networks perturbed

Eve

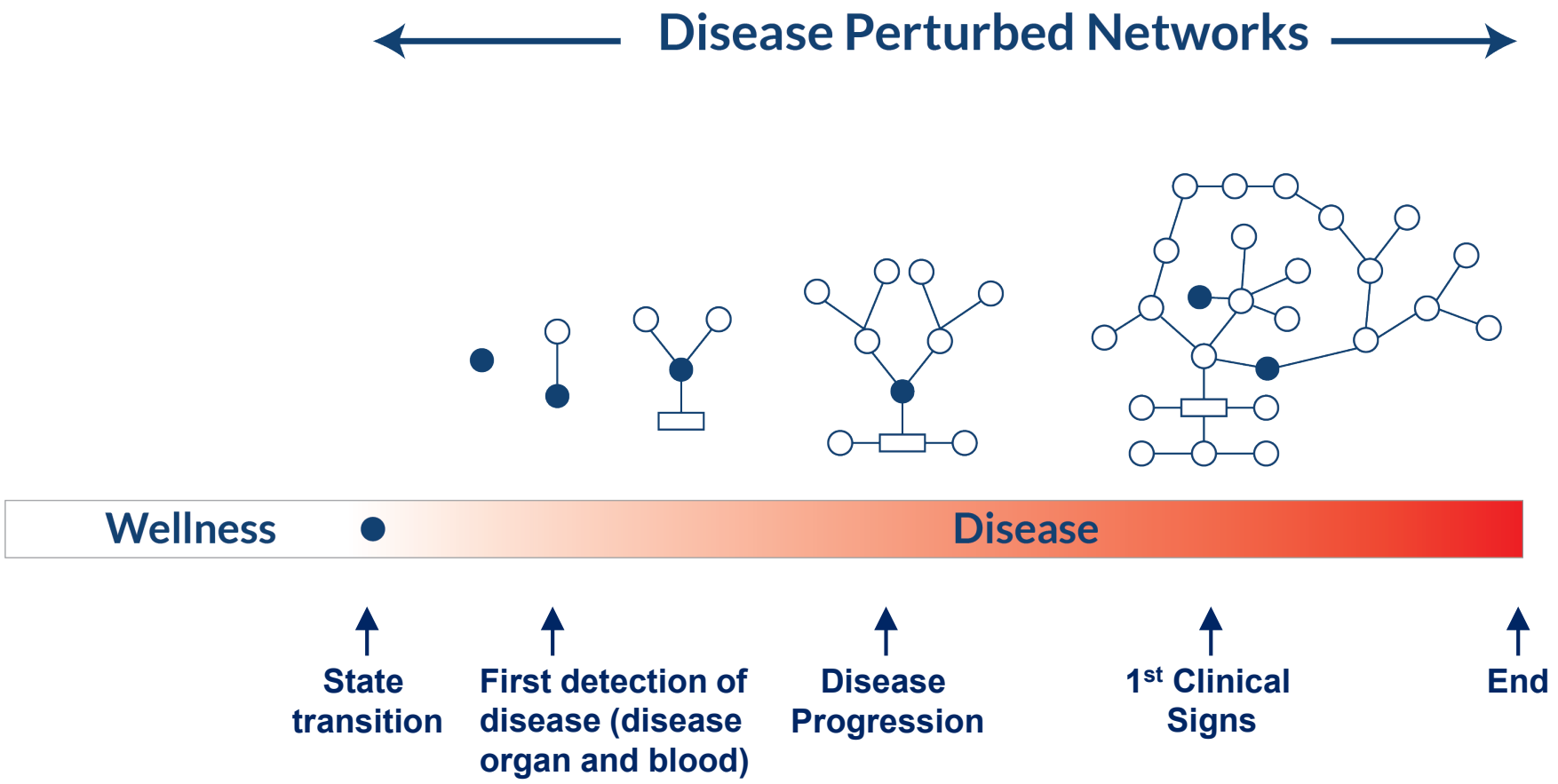




Building computational systems for truly personalized diagnostics

- Define wellness state(s) from individual data cloud
- Identify divergent (outlier) values *for each individual—analytes, networks, correlations and dynamics—at each blood draw*
  - Informed by dynamics
- Functional analysis of *individual divergences—signal to noise*
  - Deep context learning to interpret meaning
  - AI/Machine Learning/Systems Biology
- Informing next steps of individual assessments
  - Identify earliest wellness to disease transitions for each individual for all chronic diseases
  - Leads to individual diagnostics and ultimately individual therapeutics & prevention

A New N=1 Approach to Disease



Early Reversal of Chronic Diseases: Preventive Medicine of the 21st Century

- In following about 6,000 or more patients over an extended time period, we have started to see more than 100 wellness to earliest disease transitions for all common diseases (as measured by blood analytes).
- Use data clouds to develop blood biomarkers for the earliest transitions for each disease and disease-perturbed network biology analyses to identify drug candidates/life style changes for therapies to reverse each disease at its earliest transition.
- Thus individuals will have diseases reversed before the diseases manifest themselves as a chronic disease phenotype—an approach to eventually eliminating chronic diseases--preventive medicine of the 21st century—note that 86% of the healthcare budget spent on chronic diseases

ISB & Providence St. Joseph Health Affiliation

Alaska, California, Oregon, Montana, Washington, New Mexico, Texas

States served	7
Hospitals	50
Physicians	7500
RNs	36,000
Unique patients served each year	8 million
Total Assets	$23 billion

Third largest not-for-profit healthcare system in the US
Integrated Medical Electronic Health Records for 30 million patients

Systems-Driven Clinical Trials

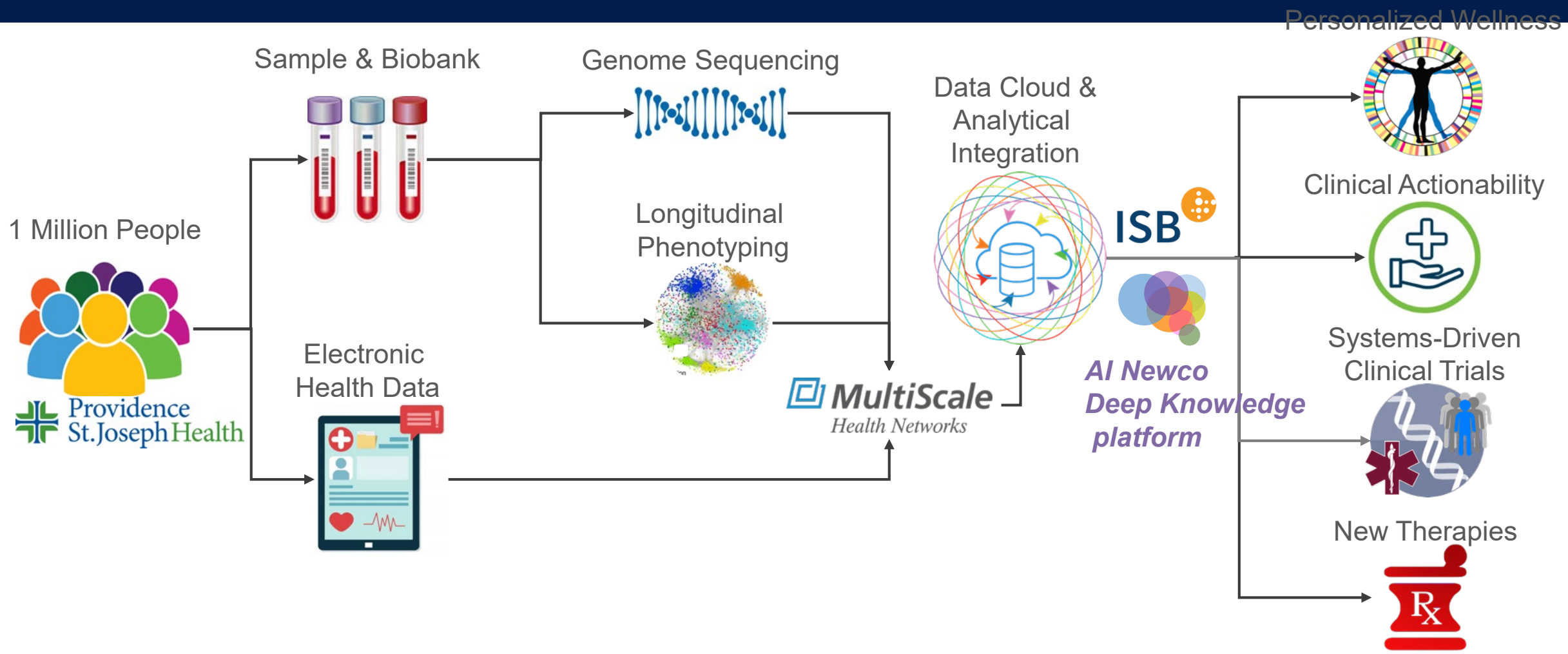
1. **Scientific wellness**
2. **Alzheimer's (3 trials)**
3. **Multiple sclerosis**
4. Wellness for breast cancer survivors
5. Lyme disease
6. Immunotherapy for cancer (vaccine)
7. Pregnancy
8. Dental caries
9. Sepsis

Strategies

Systems-driven technologies
and strategies
Longitudinal, deep phenotyping

One-Million Longitudinal Patient Genome/Phenome Project

1 Million Genomes Bring PSJH Into 21st Century Medicine



Sample Pipeline

LabCorp BioBank

Genotyping/Phenotyping

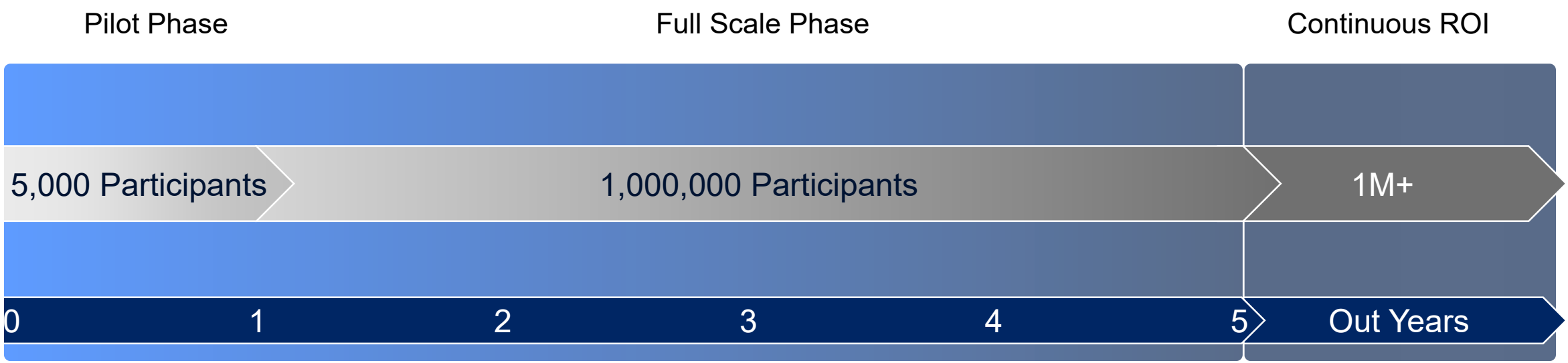
illumina METABOLON olink PROTEOMICS Veritas The Genome Company LabCorp

Informatics Technology

Microsoft IBM Google

Pharma--
3 companies

Implementation of Health 2.0 Population Screening – 1 Million Genome/Phenome (Gx/Px) Project



Project Timeline (Years)

Pilot Program
Wellness—Family Practice
Cardiovascular
Diabetes/Obesity
Cancer
GI Diseases

Longitudinal Deep Phenotyping

- Blood proteins—O-link
- Selected clinical chemistries—Lab Corp
- Blood metabolites--Metabolome
- Gut microbiome—CoreBiome
- Wearables and digital health applications
- Cognitive Brain metrics and training
- Epigenetics*
- DNA sequencing of WBC mRNA and miRNA*
- Every 6 months

*Future analysis of biobanked samples

Overview of the Arivale Platforms at ISB

- The **Arivale production platform** is designed to **manage hundreds of thousands of individuals** through blood draw collection, sample shipping, and clinical data processing.
  - Automated requisitions and clinical data ingestion
  - Member tracking and clinical decision support
  - Saliva and microbiome kit shipping and tracking
  - Online lifestyle and health history assessment collection
  - Member notifications (critical for data collection syncing)

- The **Arivale research platform** is designed to **enable research on longitudinal multi-omic datasets**.
  - Ingest and process genetic, microbiome, proteomics, and metabolomics data
  - Automated batch correction for metabolomics and proteomics
  - Automated computation of polygenic scores
  - Consolidated datasets designed for immediate scientific analysis through Jupyter notebook interface

- The **Arivale clinical platform** consists of **clinical and scientific curation processes** that provide a strong evidence base for our approach to data integration and automated recommendations.

Clinical Benefits for Physicians and Patients: Identify verified Actionable Possibilities and Discover Thousands of New Actionable Possibilities

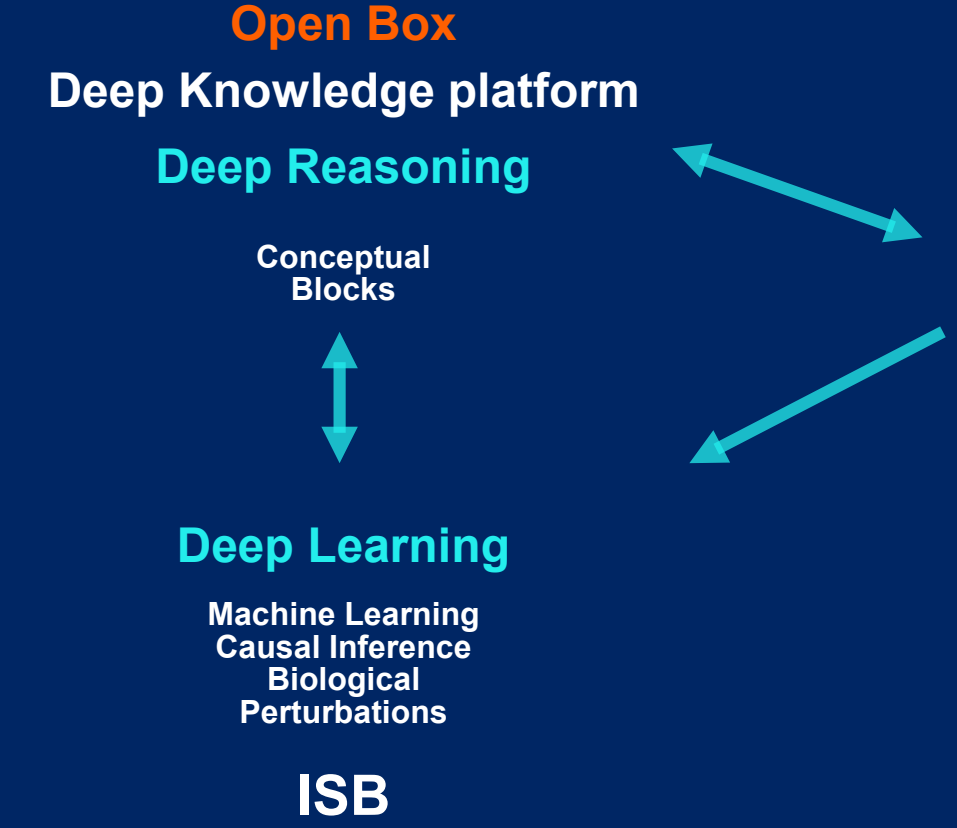
Genome

- Assess genetic risks for 125 diseases (hundreds more in the future)—polygenic profiles (GWAS-based)
- 59 actionable variants (ACHG)
- Pharmacogenomics 25/top 100 prescribed drugs
- 7500 recessive Mendelian diseases (e.g. hemochromatosis)
- Convert the two lower levels of actionable possibilities in ClinVar to high level possibilities—discovery of hundreds of new actionable possibilities
- Far most sophisticated assays of immunological function (HLA typing by DNA sequence)
- Nutrigenomics
- Athletic injury susceptibility
- Optimize diet and exercise according to your genomic constraints
- Rare diseases
- Pediatric Emergence room

Phenome

- Determine biological age to facilitate healthy aging (be mentally and physical functional in 90s)
- Identify wellness to disease transitions for all chronic disease and reverse before the disease manifests itself
- Identify new biomarkers and potential drug target candidates by statistical correlations
- 100 polygenic scores to determine how to treat patients and follow high risk patients for earliest transitions and learn to reverse them

AI—Expert Systems Newco--the Deep Knowledge platform creates the continuous self-learning healthcare system that is open box—the rationale is visible



The platform brings together deep reasoning, deep learning, and deep clinical practices (domain expertise).

It enables best medical thinking to be applied individually to patients anywhere on the globe. N=1 treatments.
Democratization of medicine.
Deliver actionable possibilities as notes that can directly be placed in clinical database

Doubling of Medical Knowledge

Time for medical knowledge to double

- **1950 50 years**
- **2010 3.5 years**
- **2020 estimated 73 days**

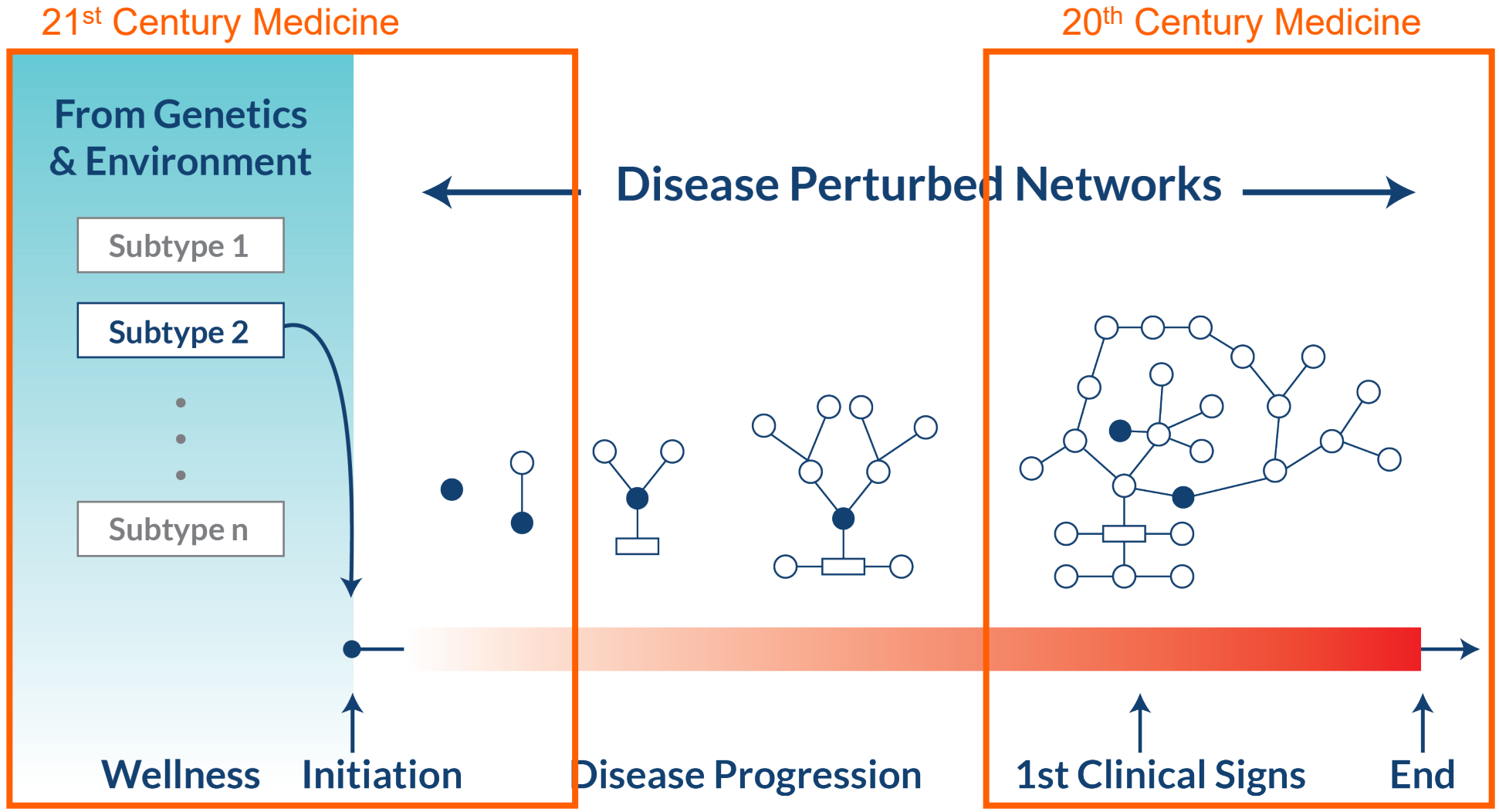
Assume

- **Doubles every 2 years—in 50 years 33.5 million times today's medical information**
- **Doubles every 78 days—in 50 years 10^{75} times the information**

Unique Features of the 1-million person Gx/Px project

- **Longitudinal deep phenotyping unique**—longitudinal data—N=1—establish baselines and follows wellness to disease transitions (and reverse)—unique new actionable possibilities
- **Discovery—**generate 1000s of new actionable possibilities
- **Converting data to actionable possibilities** to improve physician treatment and optimize patients' health. Follow with **outcomes.**
- **Integrated computational platforms and a deep AI knowledge platform** to integrate these data (Gx, Px and EHRs) and convert them into **actionable possibilities**—directly deliverable to physicians and hence to patients—these integrated computational platforms make possible a "**continuous learning system**" and **open box** explanation of rationale for resulting diagnosis and therapy
- **Unique strategic partnerships**—because of scale and immediacy of initiation
  - DNA sequencing companies
  - Phenomic vendors
  - Technology companies
  - Pharma companies
  - Biobank company?
  - Open AI both deep knowledge and deep learning
- **Focus on both wellness and disease--**CV, Cancer, GI, Diabetes/Obesity, Wellness (family practice)

The wellness to disease spectrum and 21st century medicine vs. 20th century medicine



I. Deep Phenotyping and 21st Century Medicine Will:

- **Optimizing individual wellness--**this will help avoid disease
- **Identifying and reversing wellness to disease transitions at their earliest point** for many chronic diseases—preventive medicine of the 21st century
- **Identifying through longitudinal deep phenotyping populations at high genetic risk for 100 or so diseases** and follow them closely to respond to earliest transition and find early transitions.
- **Identifying high risk individuals with prodromal disease—and reversing it before it every transitions into frank disease. Understanding early disease.**
- **Understanding human biological complexity by N=1 approach**—human nutrition
- **Pioneering new approaches to blood biomarker and drug target discovery—**analyte correlations (static and dynamic), analyte correlations with increasing genetic risk, systems approaches to biology at transition points, **multimodal therapies** for complex diseases, etc.

II. Deep Phenotyping and 21st Century Medicine Will:

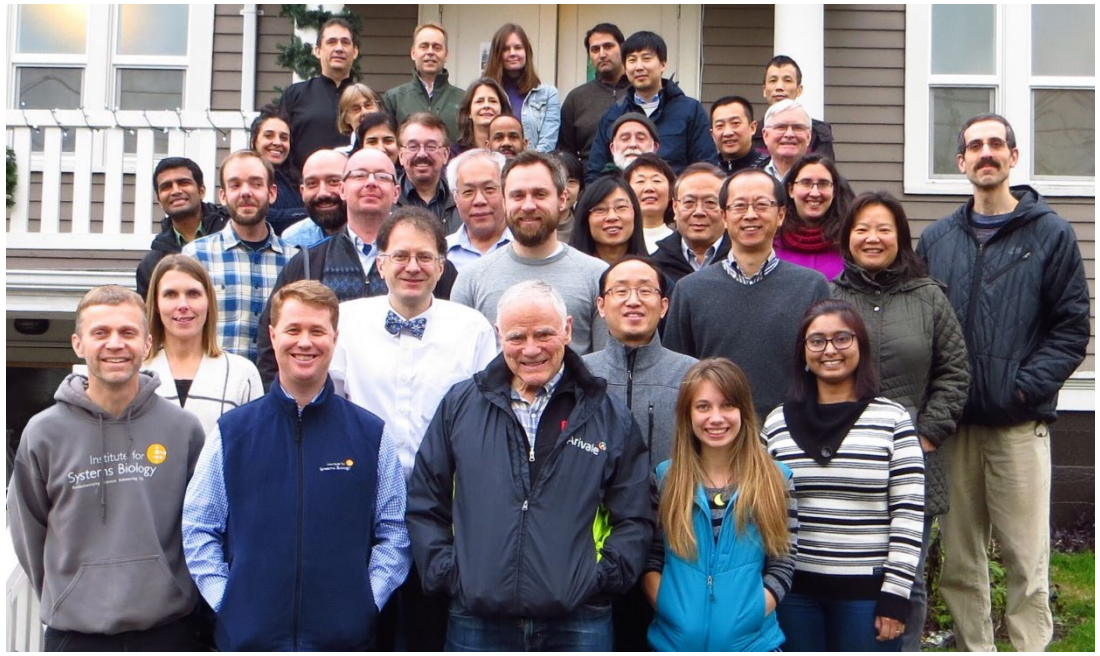
- **Healthy aging**—bring individuals into their 90s mentally alert and physically active. Biological age metric key to optimization of healthy aging. Identify other aging metrics
- **Drug discovery with two 50 patient trials** (1st trial--identify biomarker for responders and 2nd trial include only responders).
- **AI-expert systems Newco (continuously learning knowledge-based, open-box system for diagnosing and treating disease)** will make each physician an expert in the relevant disease—democratization of medical expertise and dealing with biological complexity.
- **Managing complex diseases of individuals by N=1.**
- **Saving the healthcare system substantial expenses** (e.g. Alzheimer's) and drugs (dual N=50 clinical trials)
- **Migrating wellness healthcare** from big healthcare systems to individual clinical practices and ultimately the home.

What Will 21st Century Medicine Achieve?

- Improve wellness for the individual, healthy aging and functional into 90s
- Reverse diseases at their earliest transitions--eliminate many chronic diseases
- Reverse and strikingly decrease the ever-escalating healthcare costs
- Scientific wellness will move healthcare to the home--tricorder

Integrated Lab for Systems Biomedicine

PIs: Lee Hood and Nathan Price



Principal Scientists

Gustavo Glusman, PhD

Kai Wang, PhD

Senior Research Scientists

Richard Gelinas, PhD
Inyoul Lee, PhD
Shizhen Qin, PhD
Jared Roach, PhD
Lee Rowen, PhD
Arian Smit, PhD
Qiang Tian, PhD
Kathie Walters, PhD

Research Scientists

Anjalee Bheda-Malge, PhD
Cory Funk, PhD
Alison Paquette, PhD
Noa Rappaport, PhD
Max Robinson, PhD
Vangelis Simeonidis, PhD
Kalliopi Trachana, PhD
Yong Zhou, PhD

Postdoctoral Fellows

Rhishikesh Bargaje, PhD
Priyanka Baloni,PhD
Vikas Ghai, PhD
Laura Heath, PhD
Neda Jabbari, PhD
Minyoung Lee, PhD
Matt Richards, PhD
Martin Shelton, PhD
Tomasz Wilmanski, PhD

Graduate Students

John Earls

Jocelynn Pearl (*now at Altius*)

Undergraduate Researchers

Brendan King (U Washington)
Noah McClean (U Chicago)

Research Associates

David Baxter
Shannon Fallen
Kelsey Scherler
Li Tang

Senior Software Engineers

Paul Shannon
Robert Hubley

Senior Research Engineer

Chris Lausted

Bioinformatics Scientists

Xiaowei Yan
Xiaogang Wu
Taek-Kyun Kim

Lab Senior Program Manager

Simon Evans, PhD

Program Manager

Mary Brunkow, PhD

Executive Assistants

Alicia Levesque
Sheryl Suchoknand

Collaborators

***Andrew Magis, PhD**, and **Jennifer Lovejoy, PhD**,* Arivale

***Gilbert Omenn**, M.D.,* U Michigan

Funding

National Institutes of Health
Department of Defense
gapps
NSF
Robert Wood Johnson Foundation
Department of Energy
NIH National Institute on Aging
Bill & Melinda Gates Foundation
M.J. Murdock Charitable Trust