

# Data-Driven Individual Health Will Lead to a Healthcare that Optimizes Wellness, Prevents Disease, Extends Healthspan and Employs Peptides as Rapidly Generated & Effective Drugs

**Lee Hood MD, PhD**

**Chief Executive Officer**  
*Phenome Health, Seattle, WA*

**Co-founder, Professor**  
*Institute for Systems Biology, Seattle WA*

**Chief Innovation Officer, Professor**  
*Buck Institute for Aging, Novato, CA*

**N=1 Healthcare**



# I Participated in Eight Paradigm Changes in Deciphering Human Complexity

## Realizing Paradigm Shifts Through Small and Agile New Organizations

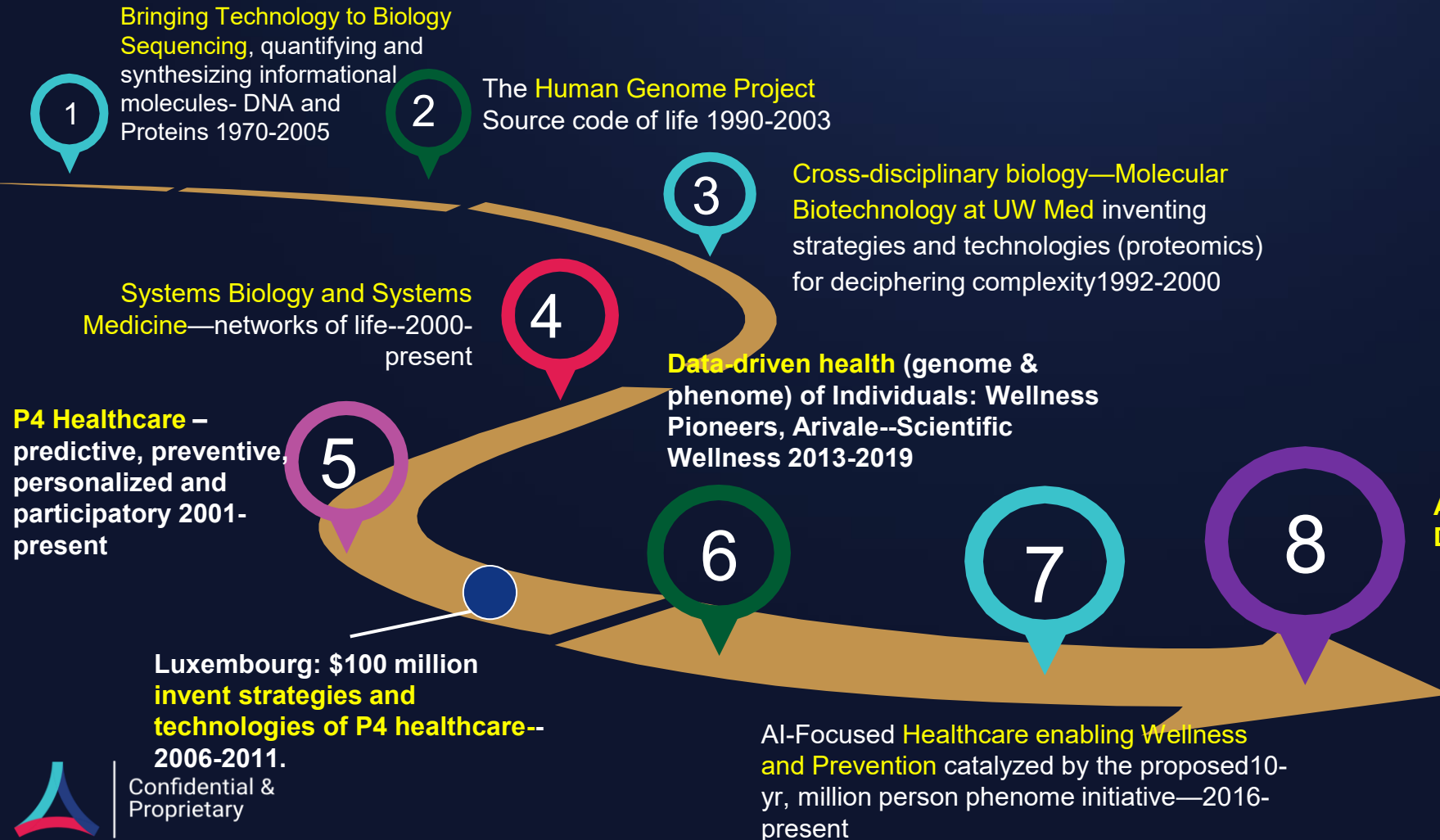
### Non-Profit Initiatives:

- UW Molecular Biotechnology
- P4 Health Institute
- Institute for Systems Biology (ISB)
- Phenome Health

### For-Profit Ventures:

- Applied Biosystems
- Amgen
- Rosetta Informatics
- NanoString
- Arivale
- Phenomia.ai
- P4Bios

**AI-Driven, Information-Based Peptide Drugs—P4Bios--2024 – present**



Confidential & Proprietary

# Data-Driven Individual Health: the Genome and the Longitudinal Phenome

Genotype 10%

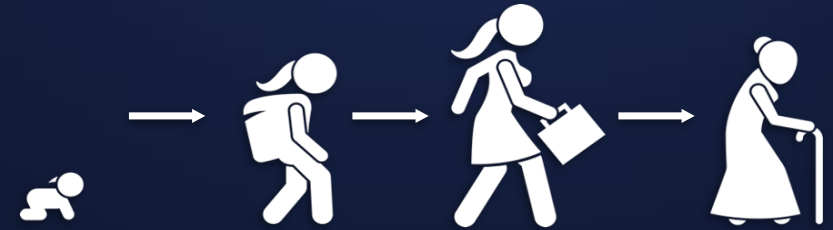


1 Person = 1 Genome

Human Genome Project (HGP)



Phenotype 90%

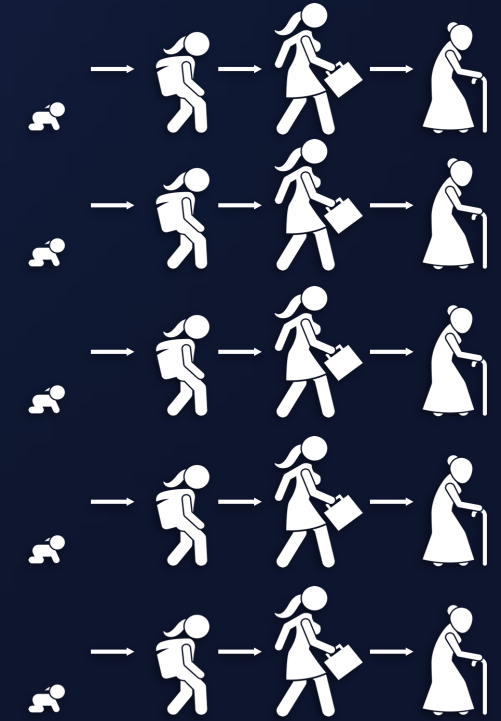


1 person = Infinite phenomes across one's longitudinal life trajectory  
One's *behavior* and *environment* perturb the genome-dictated health trajectory through out life.

The phenome allows an assessment of one's always changing phenotype by quantifying *blood analytes, gut microbiome and brain and body digital health assays.*

*Arivale Wellness Company—5000 individuals 2x over 4 years*

1 Million Person Phenome Initiative over 10 years



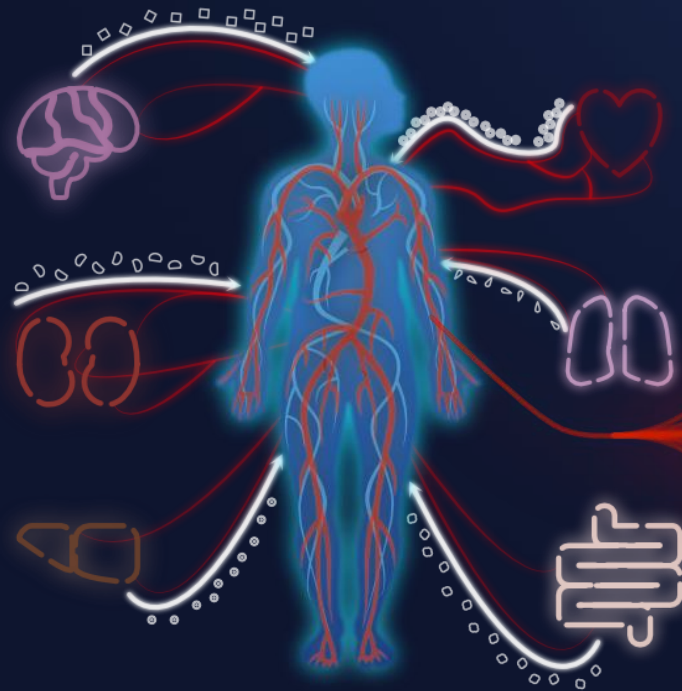
**Validate**  
Individual wellness  
Early detection and reversal of disease  
Technology development  
Return results/ invent future  
Trillion \$s cost savings

# How can Big Data be Analyzed: Systems Biology Approaches

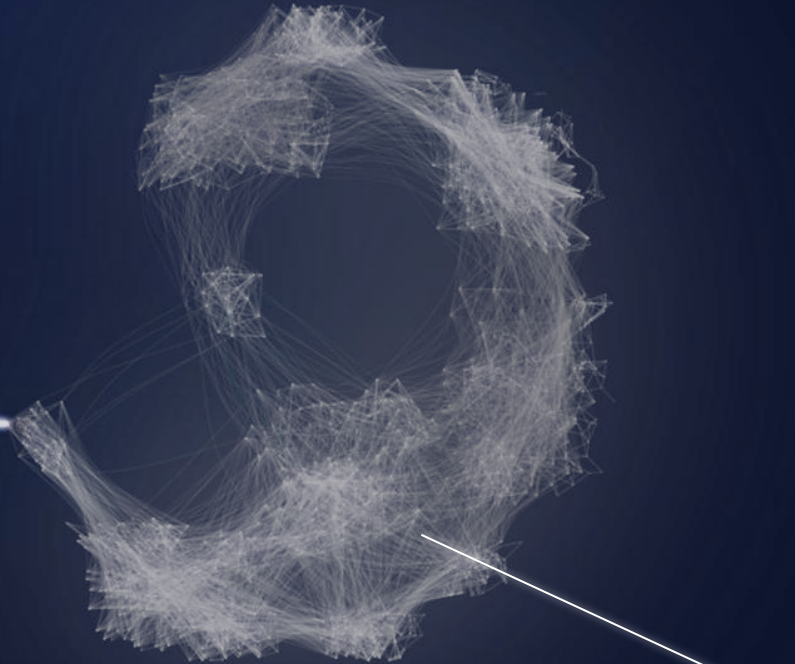
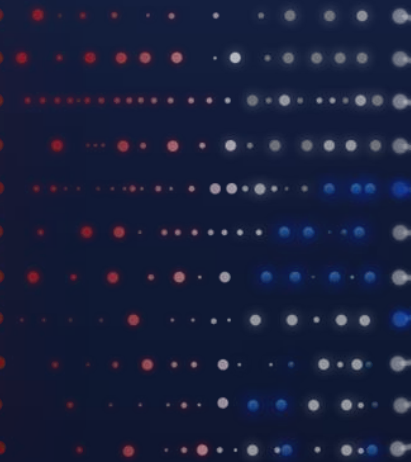
Holistic  
Global  
Dynamical  
Hierarchical  
Integrative

## Dynamic Biological Networks and Communities that Underpin Wellness and Disease

Blood is a **window into wellness and disease**



Integration of phenomic measurements



Quantify Informational Molecules:  
DNA, RNA, Proteins, Metabolites

Derived from 5,000 individuals in the Arivale cohort

- > Machine Learning (ML) & Artificial Intelligence (AI)
- > Knowledge Graphs for connecting medical knowledge
- > Digital twins for predictions and perturbations of health trajectories
- > Large Language Models (LLMs) for deriving actionable recommendations

# Lessons from the Arivale 5000 Person Wellness Program

# Arivale, Data-Driven Health and Emergence of **Scientific Wellness**

## Wellness Cohort

- 5000 individuals from Arivale Wellness Company (2015-2019) x2 per year
- Data generation lead to **Scientific Wellness Lake to Actionable Possibilities**

## **Six data features**

- Genome variants
- Proteins
- Metabolites
- Clinical chemistries
- Gut microbiome
- Physiology--fitbit

## **Three Types of Actionable Possibilities**

Optimize wellness

Prevent disease

Early detection and reversal of disease

Price, Magis, Earls...Hood, *Nature Biotechnology*, 2017; more than 25 published papers



# Seven Classes of Actionable Genomic Possibilities (AP)

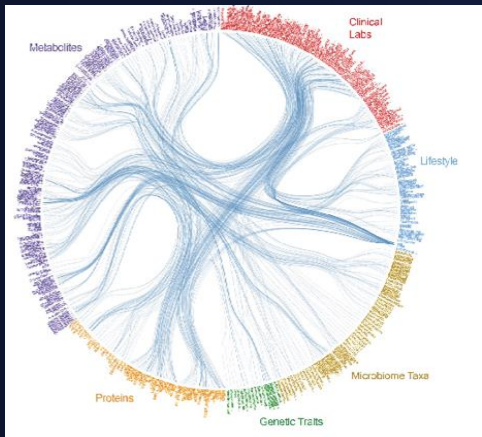
- American College of Medical Genetics  
81 actionable variants
- Pharmacogenomic variants--100s AP
- Polygenic wellness- and disease- phenotype scores—100s APs
- Rare diseases—7000+ APs
- Mendelian recessive diseases—7000+ APs
- Lifestyle genetics—activity, nutrition, stress, sleep, athletic injuries—100s
- Ancestry—ancestry-specific disease propensities
- DNA analysis of oral and gut microbiomes

Thousands of actionable possibilities



# Statistical and Community Data Analyses Demonstrate 100s of Potential Drug Targets

Deriving Insights from Data: New Frontiers—3500 Statistical Correlations in 108 patients

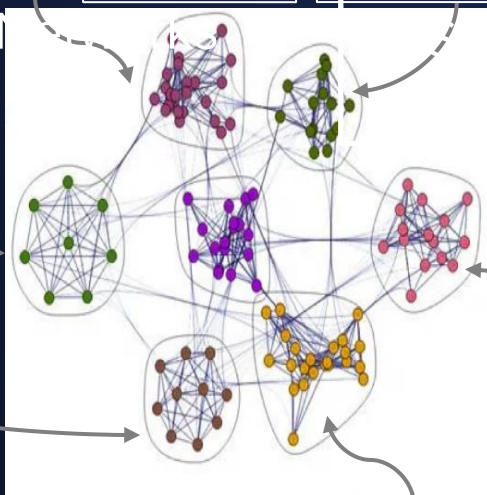


Systems Biology & Biochemical

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

- Protein
- Metabolic
- Transcriptional
- Microbial

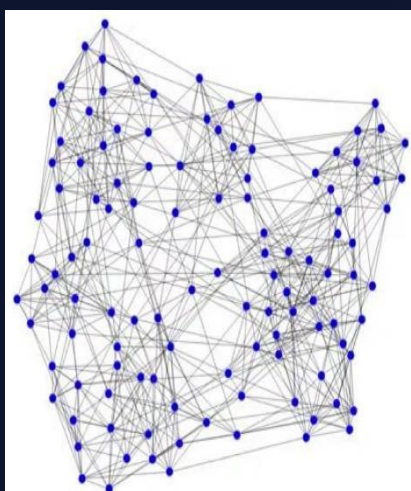
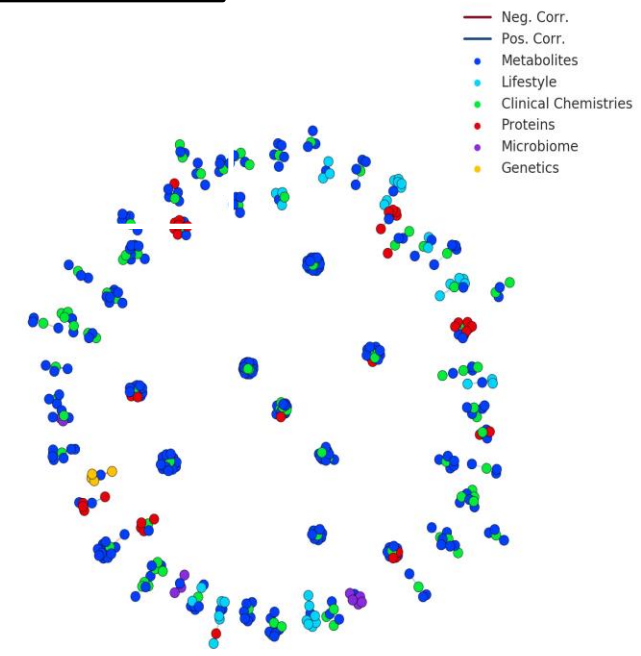
Bladder Cancer



Synthetic & Degradation Networks

- Cholesterol is negatively associated with endogenous thyroxine
- A beneficial side effect of the drug thyroxine (Synthroid) is lowering LDL cholesterol

Total cholesterol community as one of 70 communities



Networks & Communities are core of Scientific Wellness Data Lake

Price, Magis, Earls, Hood et al, *Nature Biotechnology*, 2017

500 proteins give 3500 correlations; 5000 proteins give 35,000 correlations and 100s communities



- Multidimensional algorithmic metrics for distinct physiological trajectories—optimizing wellness in multidimensional space—going far beyond traditional wellness
- - Biological/Organ Age—blood analytes
  - Biological BMI--metabolites
  - Frailty—blood analytes & physiology
  - Immune function—single-cell analyses 5000 WBCs/sample
  - Organ-specific blood proteins to follow organ transitions
  - Brain cognition and autonomic nervous system—brain HQ

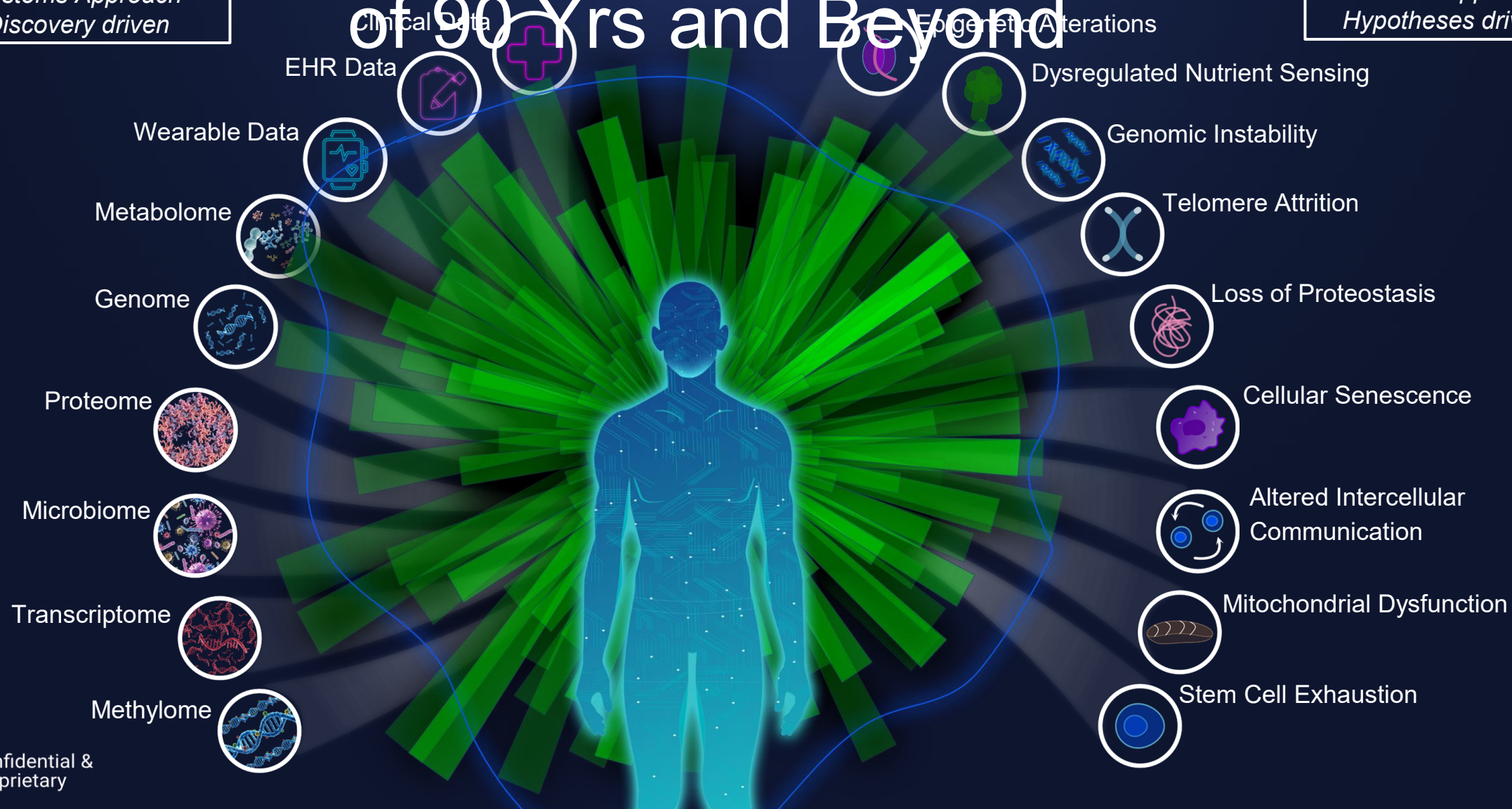


Integration by Knowledge Graphs and Transformers

# Optimized Healthspan of 90 Yrs and Beyond

Phenomics  
Blood is a Window  
Systems Approach  
Discovery driven

Gero-Science  
Hallmarks of Aging  
Focused Approach  
Hypotheses driven



# Data-Driven P4 Healthcare: Early Detection and Reversal

## Big Data and Systems-Biology Analyses of Individuals and Disease

Conventional  
Medicine

One

First detectable  
clinical marker

Wellness

Transition Interval

Onset of disease

Transition Interval

Disease progression

Transition Interval

Clinical Diagnosis

First detectable  
phenomic changes

Targeted therapy

Arivale 5000 4 yrs  
167 transitioned  
30 to cancer  
10 analyzed

Phenomic  
Medicine

Health has 3 stages and we can optimize each:  
wellness, transitions, disease

Return to wellness

# Actionable Possibilities include...

- **Optimize wellness** & aging—clinical chemistries, lifestyle optimizations, health trajectories
- **Ameliorate or reverse genome and/or phenome encoded deleterious phenotypes**
- Utilize **algorithms for novel health trajectories**—age, metabolism, frailty, immune function, cognition,
- **Integrate wellness, gero-science and Traditional Chinese Medicine** into the **Scientific Wellness data lake** to generate new health trajectories
- **Detect early disease detection** and reverse disease phenotypes
- **Enable prevention**
- Deal more effectively with the complexities of diagnosed chronic diseases—**multimodal clinical trials**—including behavior change, balancing blood chemistries and drugs, avoiding or modifying deleterious environments
- Identify **candidate biomarkers and their drug targets for wellness, prevention and disease** from the Scientific Wellness data lake



# Where Phenome Health is taking data-driven health?

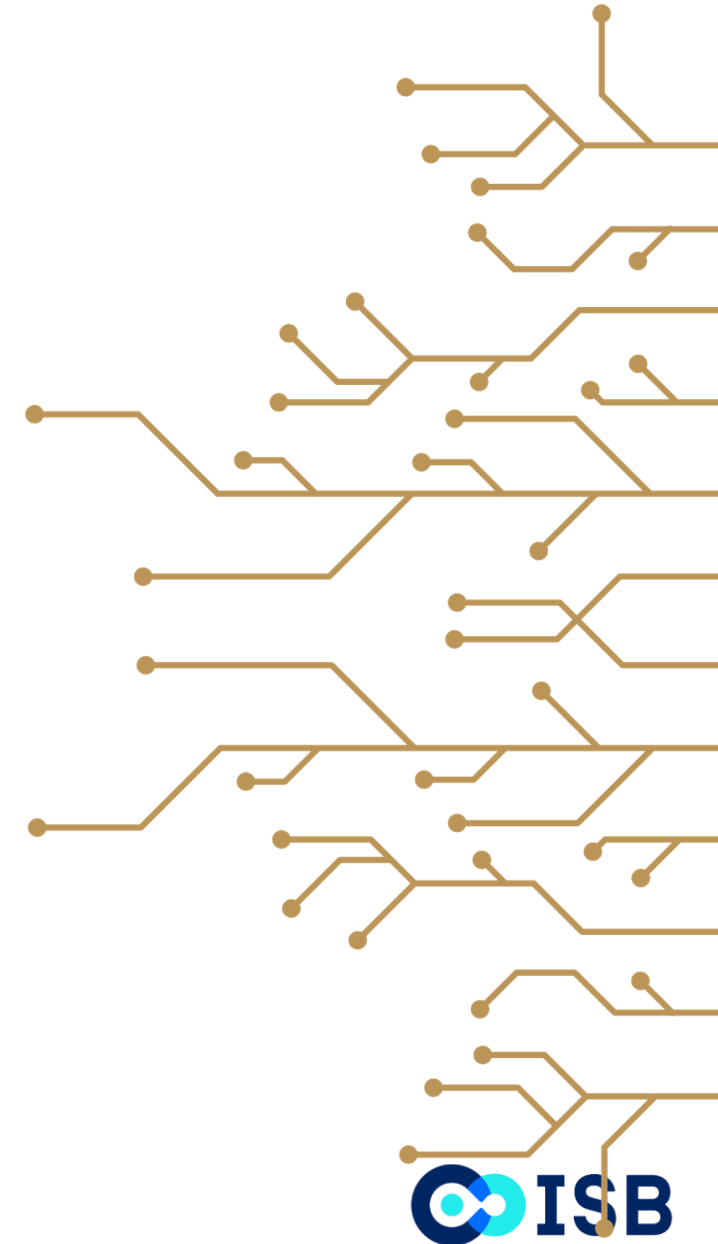
# Phenome Health's Objectives

**Generate and aggregate large-scale genome/longitudinal genome/phenome data into the Scientific Wellness data lake for individuals both normal and diseased**—connect to large federated data lakes from partners and bring the analytic tools to lakes to extract relevant meta-data

**Harmonize, integrate & analyze individual genome and longitudinal phenome data for normal and diseased individuals**—yielding actionable possibilities, stratification of individuals and diseases, candidate biomarkers and drug targets, causal explanations of state transitions and IP

**Invent the data-driven health of the future**—with new technologies, new data types, intense analyses of existing data sets, AI, biobanks that capture in samples the history of disease and physiology cohorts for multiomic analyses and P4 education

**Identify partners** that provide patients, resources, new technologies, interesting cohorts and with an interest in joining the phenome initiative.



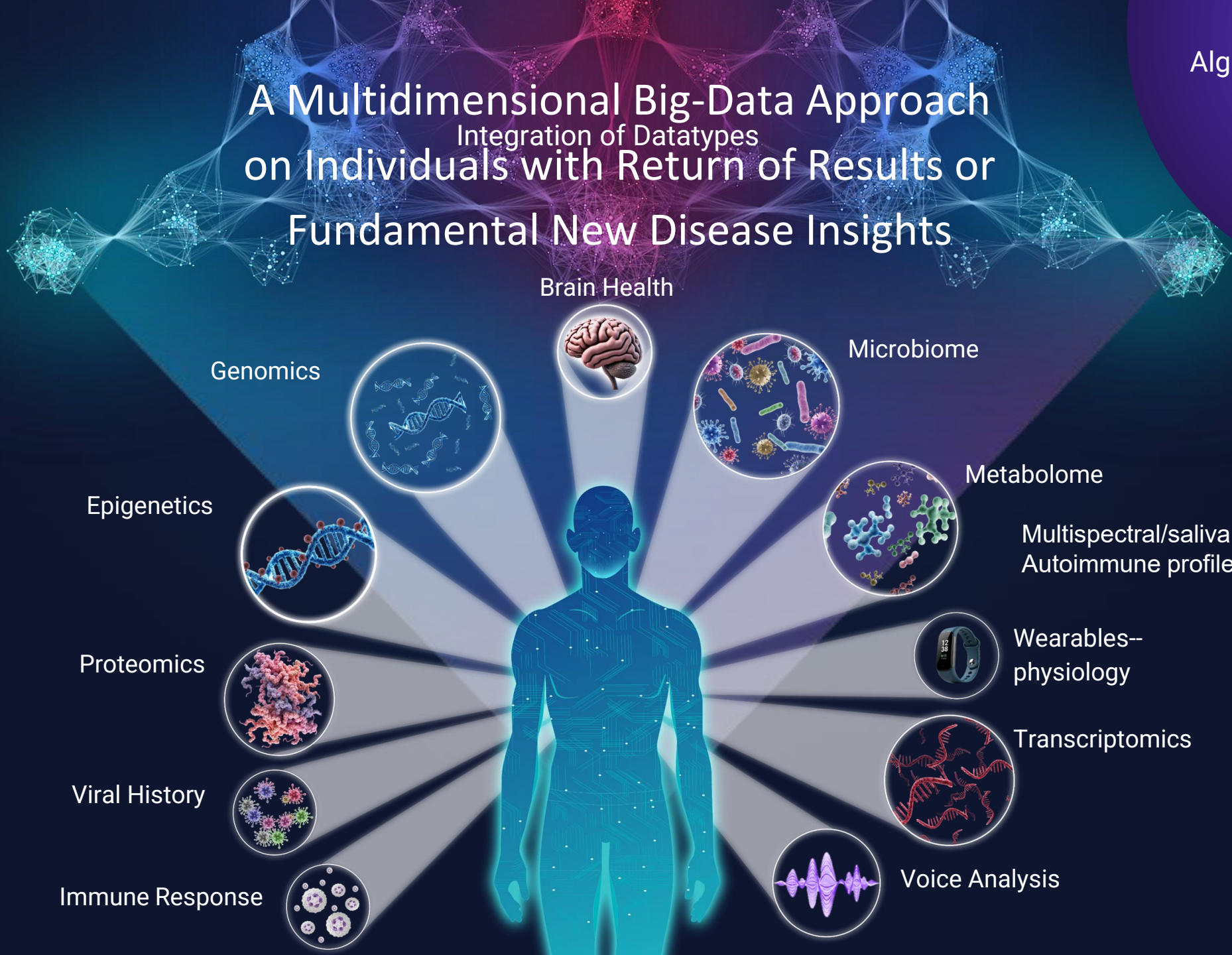
# Five New Data Features from Long-Read Genomic Sequencing

1. Epigenetics all methylated Cs across all human chromosomes
2. Genomic indels (e.g., insertions, deletions and/or rearrangements)
3. Accurate sequencing through GC rich regions—HLA locus, pharmacogenomic genes
4. Phasing of individual chromosomes (maternal and paternal contributions)—key to data-driven family studies
5. Epigenetics assessing signals from covalent DNA modifications other than methylation of Cs



# A Multidimensional Big-Data Approach Integration of Datatypes on Individuals with Return of Results or Fundamental New Disease Insights

Algorithms for health  
trajectories



Confidential &  
Proprietary

# Science Foundation Being Built



KRAKEN—18 integrated knowledge graphs



KESTREL—a tool that allows agentic agents to traverse across KRAKEN knowledge graphs



AGENTIC LAYER—specific AI tools



Multi-omic data platform—ingest, integrate and analyze phenomic data (systems approaches)



AI-driven biological reasoning and the delivery of APs to patients and physicians



# Phenome Health's Computational Core

*Delivering Knowledge from Data*



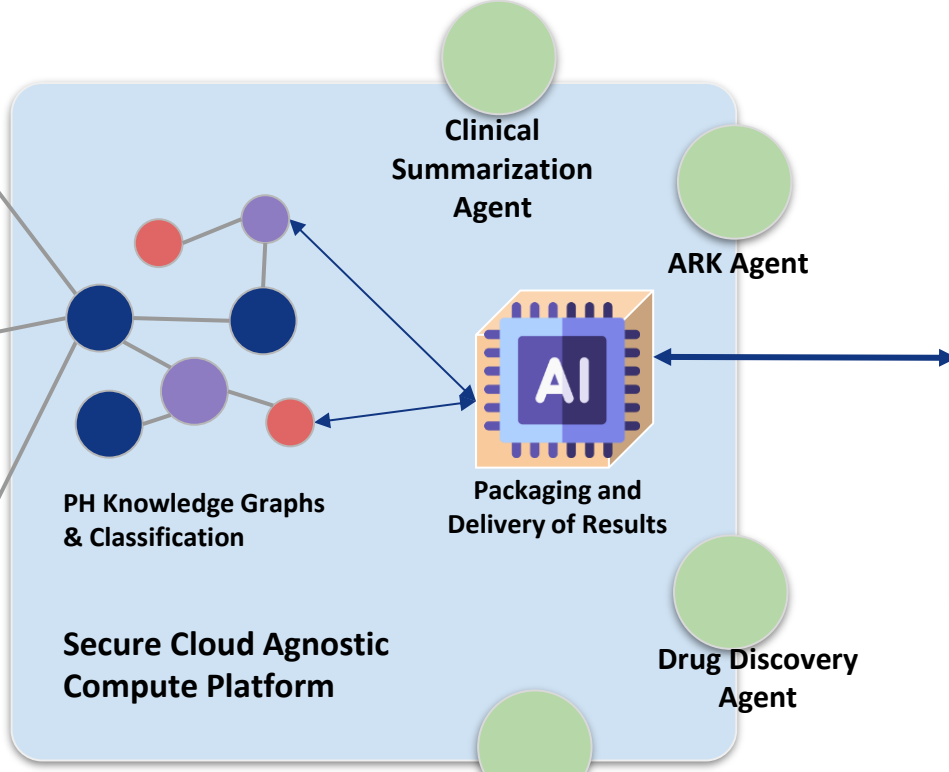
**Phenome Health Algorithms**  
Biological Age, Biological BMO, Frailty Index, Cognitive Brain Transitions, Organ Specific Markers (25 organs), Hallmarks of Aging, Traditional Chinese Medicine



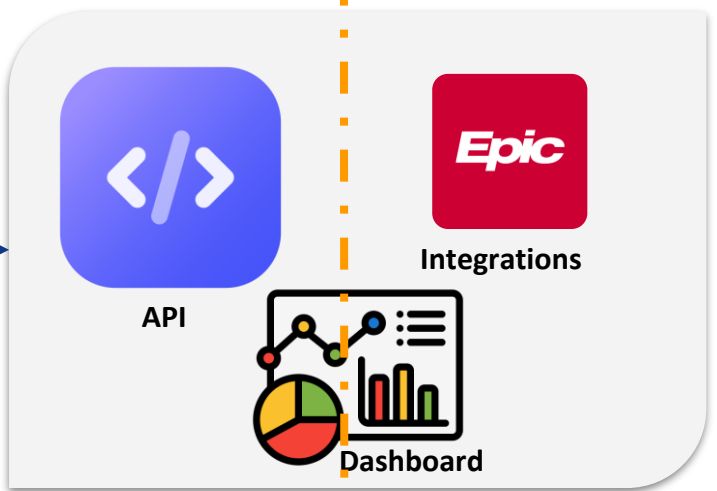
**Phenome Health ARK**  
Actionable Recommendation Knowledge



**Arivale + other multiomic data sets**  
5000+ subjects of proprietary longitudinal data



**IN**  
Personalized Subject Data



Actionable Recommendations



# Multidimensional, big-data, longitudinal, broad-spectrum, 5-yr clinical trial

Broad-spectrum clinical trial—5 patient groups

High risk with normal blood chemistries

Prediabetic

Diagnosed patients divided into 3 groups

Early

Middle

Late

Omic classification

Early detection

Subtypes and their transitions to chronic disease features

Identification of transition biomarkers and possible drug targets

Causal insights into disease mechanisms



# Two general classes of genomic/phenomic opportunities— individual and disease/physiology

- **Individual genome/phenome analyses** of large populations and **return of relevant actionable possibilities** to individuals (and their physicians)
  - An international and a US million person phenome projects
  - Rural health—with Montana, Alabama--\$50 billion Trump initiative for states
- **Multidimensional, big-data, longitudinal, broad-patient-spectrum** with individual observational clinical trials for the longitudinal analyses of normal development/physiology (e.g. pregnancy) and four major chronic diseases



# Active Projects for Phenome Health

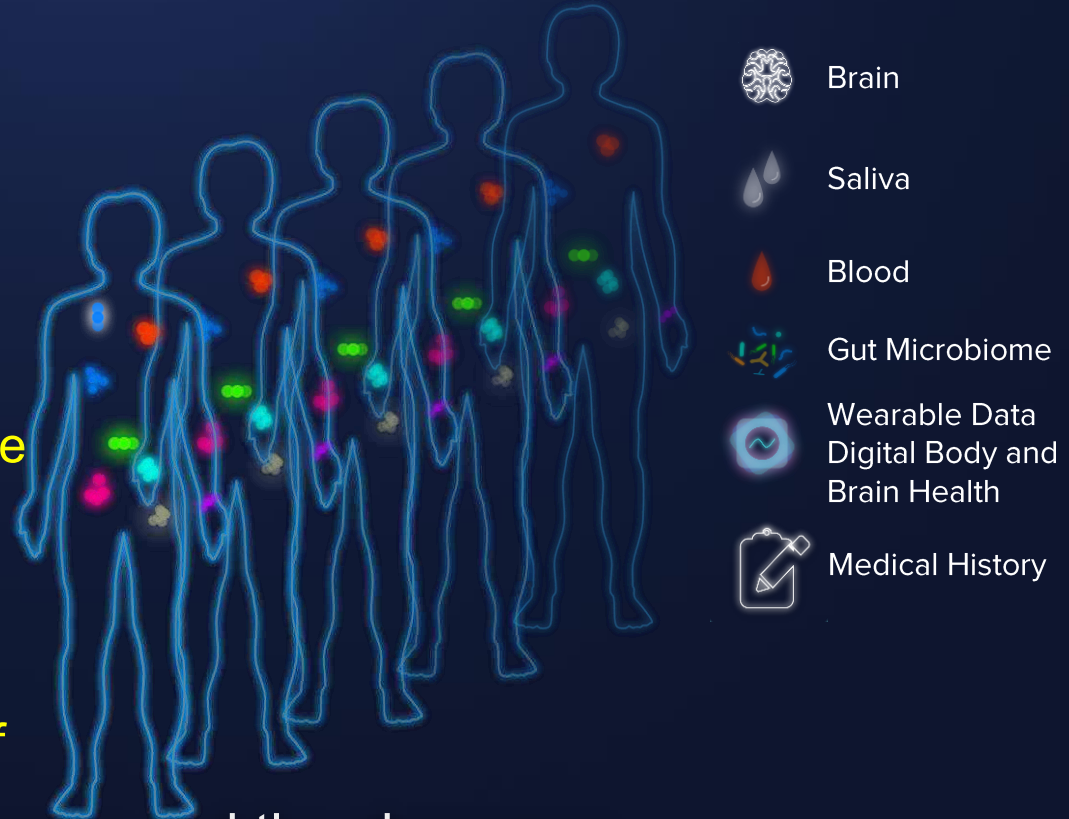
- ARPA-H grant--\$52 million/4yrs—**5 phenomic projects**—with the Buck
- South Carolina and Guardian Research Network—**type 2 diabetes** of 1000 subjects plus controls--\$21 million two years
- Alabama with Hudson/Alpha on 50 patients with **Alzheimer's disease--\$1 million**
- **Heart failure**—deep omic analyses of 60 patients with high risk for heart failure—Tampa General Hospital
- **Proposed Rural health\***—Montana and Alabama—Trump Administration \$50 billion for rural health
- **American Specialty Health\***—tier 3 phenome analyses in diverse populations
- **Pregnancy**—deep omic analyses x 5 across pregnancy—436 patients
- **International million person phenome project**—Turkey, South Korea-, Europe-
- **Proposed US 10 yr, million person project\***—extension of the Human Genome Project—return of APs to individuals—validate data-driven health!



# Extension of the Human Genome Program—the US Human Phenome Initiative

Validation of data-driven health: wellness and prevention

- Enhanced personal health
- Optimizing Wellness--10s of 1000s APs
- Est. 200,000 wellness to disease transitions
- Reduced healthcare costs—reduce trillions \$s
- Catalyzing Innovative data-driven technologies at home and new data types—democratization
- Generating tens of thousands of new actionable possibilities and drug targets for wellness and prevention
- Enormous ROI (return on investment)
- Leadership healthcare & inventing the future of medicine



The whole-genome sequences and the phenomes  
for **one million** individuals

Tracked longitudinally over 10 years 2x times/yr

Return of relevant actionable possibilities to individuals

# About us

Phenome Health is a 501(c)(3) nonprofit organization

Mission: Developing the Tools & Insights that Power Data-Driven Personalized Health



**Lee Hood, MD, PhD**  
CEO



**Simon Evans,**  
MBA, PhD  
Operations



**Noa Rappaport,**  
PhD  
Data Science



**Jennifer Lovejoy,**  
PhD  
Translational Science



**BJ Yurkovich,**  
MS  
Cybersecurity



**Tim Yeatman,**  
MD  
Medicine



**Clayton Lewis**  
BA  
Partnerships

*Predict Health. Empower Wellness. Prevent Disease*



# What are the paradigm changes of data-driven health?



Shifts healthcare from current disease focus to **wellness and prevention and ultimately N=1 medicine** with 10s of thousands of APs and a multiplicity of new drug targets for wellness and prevention

**Democratization of healthcare** to bring data-driven health to all individuals by new technologies at home mediated through a cell phone



**AI/Physician Partnerships**—AI analyses of individual patients' data and bring the resulting actionable possibilities to physicians (and patients) dramatically expanding physician domain expertise —enabling superb patient care



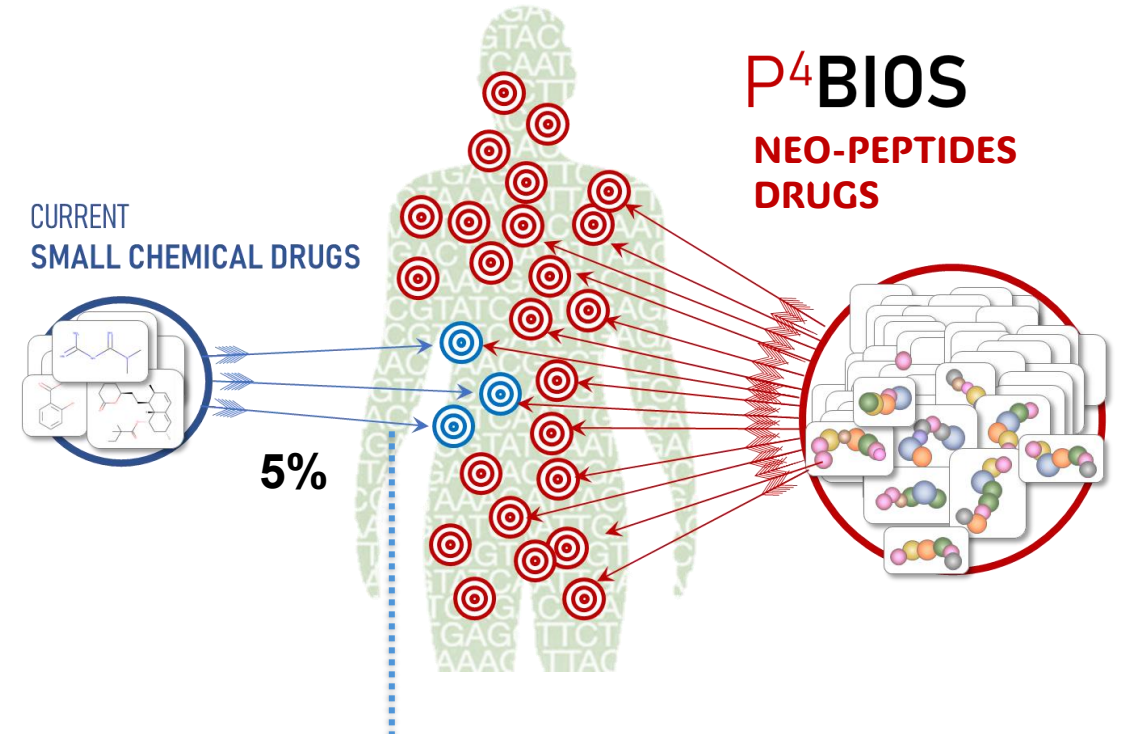
## WHAT IS THE PROBLEM ?

Current **small chemical drugs** (= majority of FDA approved drugs) target only **5% of all human proteins**.

The rest, among which are >2000 pathogenesis-related targets, notably transcription factors, are currently **“undruggable”** by small molecules

## WHY PEPTIDES - WHY NOW?

- **Neo**-peptides will be transformative more so than mRNA and CRISPR: **can be made “drug-like”**
- We can harness the rapid **progress in peptide drug delivery** (oral and into cells)
- Peptide drugs are a **fast growing multi-billion \$ market** with well understood generalizable **pharmacokinetics**



Current approaches to expand small molecule drugs into the undruggable target domain:

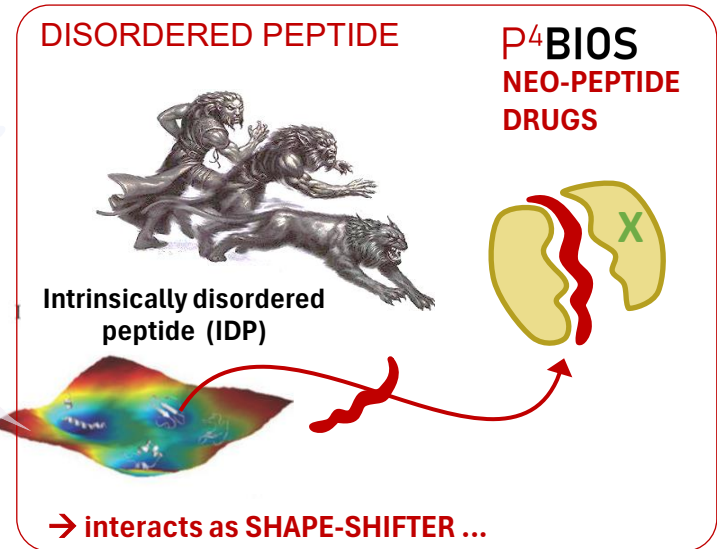
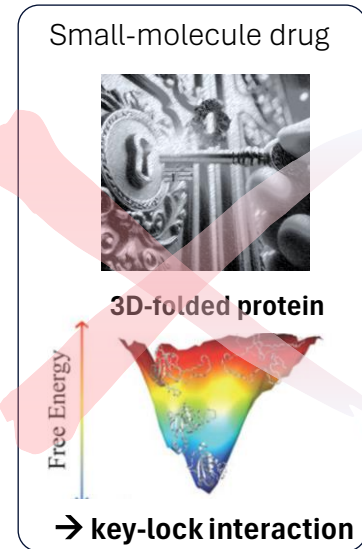
- **PROTAC** (targeted protein degrader)
- **Molecular glues**

→ *Drawback for discovery:* lengthy, poorly-scalable, only target-based rationale design, unsuited for phenotype based HT screening...

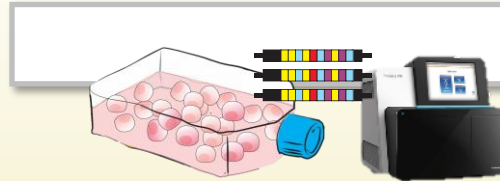
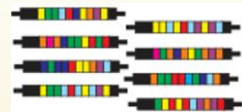
# WHAT IS OUR SOLUTION?

## DISORDERED PEPTIDES AS DRUGS ("NeoPeptides")

- can target flat surfaces and **disrupt protein complexes** beyond reach of small chemicals drugs, including target **transcription factor complexes**
- act like the natural **intrinsically disordered proteins (IDP)** ("shape-shifter) that have **pivotal regulatory activity**



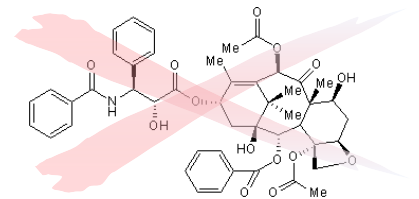
Uniquely diverse, biased  
disordered peptide gene library



**NEO-PEPTIDE DRUGS  
(+ NOVEL TARGETS)**

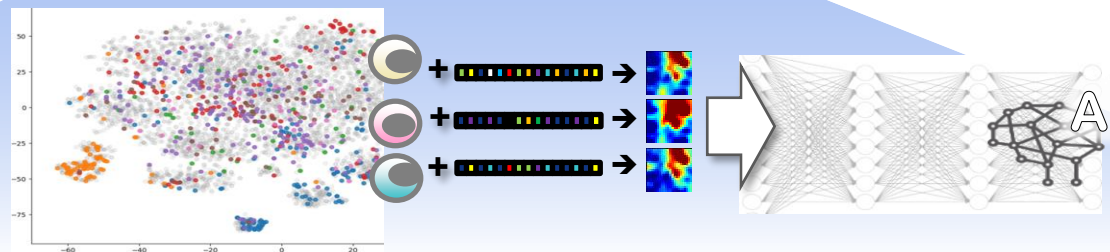


Beyond small molecules



Target- or phenotype-based screening  
in millions of single-cells offers:

- statistics and parallelism ("one-shot screen")
- speed – good for evolutionary iterations
- massive data for AI training to uncover target



# The uniqueness of P<sup>4</sup> Bios neo-peptide discovery platform: A Paradigm Change for Pharma

**Scale of molecular diversity** □ artificial gene synthesis of billions of Intrinsic Disordered peptides (IDP) that can target undruggable targets (flat surfaces and disordered regions)

**Phenotypic screening** □ harness the power of expression cloning in live cells to select candidate neo-peptides for desired phenotype alteration – not achieved by screening for *target-binding in vitro*

**Speed** □ “one-shot“ massively-parallel screening of hundreds of millions of neo-peptides in pools of cells at single cell level in a few days, without robotics

**Iterative selection & Mutational optimization** □ speed of one-shot screening allows the hill-climbing scheme for boosting phenotype effect, not just target-binding; the peptide sequence allows for systematic exploration of neighboring variants by mutagenesis

**Modality and Range** IDP interacts potentially with all proteins and as well as regulatory (non-coding) RNAs—informational drugs

**Screening combined with single-cell transcriptomics** □ millions of neo-peptide directly coupled to millions of cellular transcriptomic shift for *high information density discovery*.

**Transcriptomes connected to biomedical knowledge graphs (BiomedKGs)** □ hundred of pathways contained in the transcriptomes can be linked to biomedical knowledge graph(BiomedKGs) to identify disease-driven pathways

**Integration of BiomedKGs with deep phenotype profiling and clinical metadata** □ allows for immediate insights in disease-driving pathways helps interpreting and prioritizing screening hits.

**Generation of data to train new AI models** □ AI models that will predict properties of and designs de novo IDP neo-peptide drugs

**Close partnership with Phenome Health and Karolinska Institute** (via P4BIOS founders) □ identify target pathways for screening and subsequent selection to develop a new class of biologics for wellness, healthy aging, and early reversal of disease

# The Five Major Initiatives Catalyzing Integrated Paradigm Changes in Healthcare and the Pharmaceutical Industry to N=1 Healthcare

**Scientific Wellness**  
*Data Lake*

Gero-Science

Traditional Chinese  
Medicine

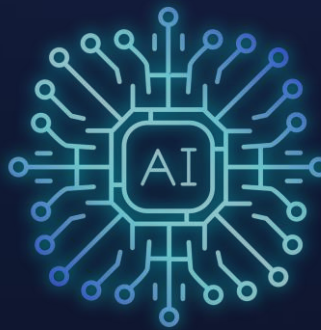
Digital Twins

**Millions of  
Human  
Genomes**

Source code of life for  
each individual

**Human  
Phenome  
Initiative**

How behavior & environment  
influence the individual source  
code readout



AI

Knowledge Graphs

**P4BIOS**



Informational peptide drugs to optimize  
wellness, aging, and disease prevention for  
each individual. Disease drugs.



Confidential &  
Proprietary



# Phenome HEALTH

[www.phenomehealth.org](http://www.phenomehealth.org)  
[collaborations@phenomehealth.org](mailto:collaborations@phenomehealth.org)



Confidential &  
Proprietary

P<sup>4</sup>BIOS



ISB



**Buck**

Live better longer.