DATA+AI SUMMIT 2022

Turning Big Biology Data into Insights on Disease – The Power of Circulating Biomarkers

ORGANIZED BY databricks

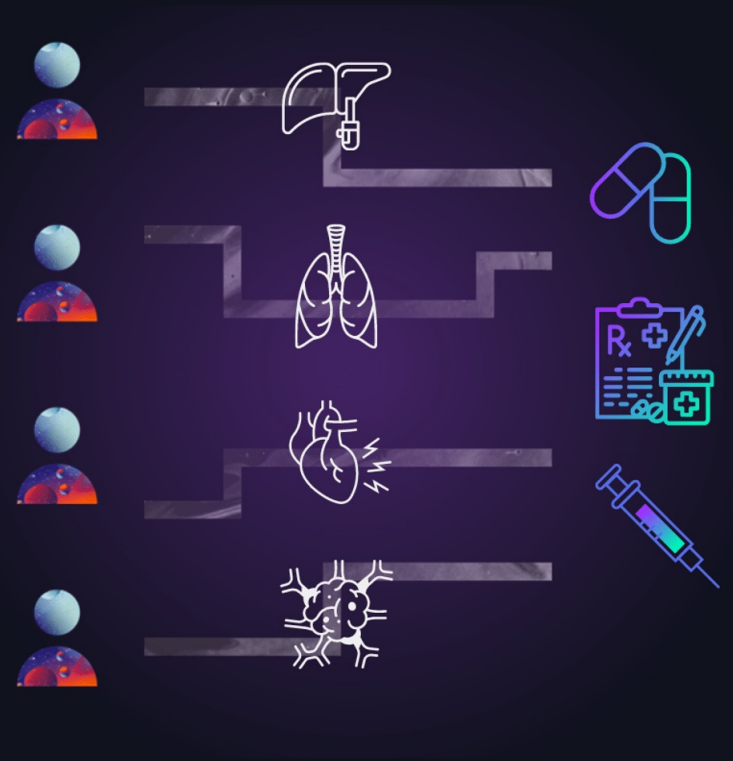
Tao Long, PhD, MBA
Co-Founder & Head of Data Science, Sapient

Why is drug development so challenging?

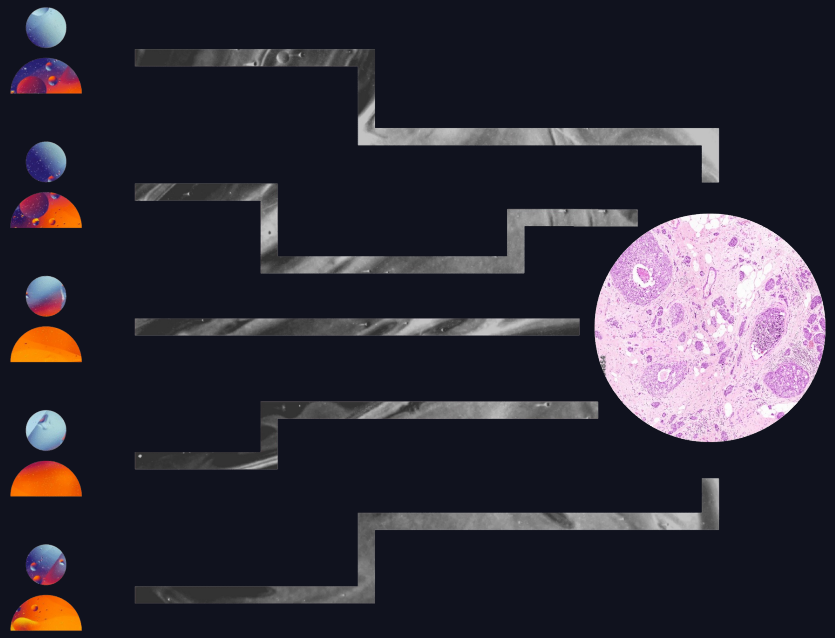
Drug development is associated with a >90% failure rate.

Even FDA-approved agents in clinical studies are only effective in a minority of individuals.

This is due to a fundamental misalignment among patients, disease, and therapies that must be addressed to transform **drug development** and **diagnostics** and improve **health quality** while reducing **cost of care**.



Why is drug development so challenging?



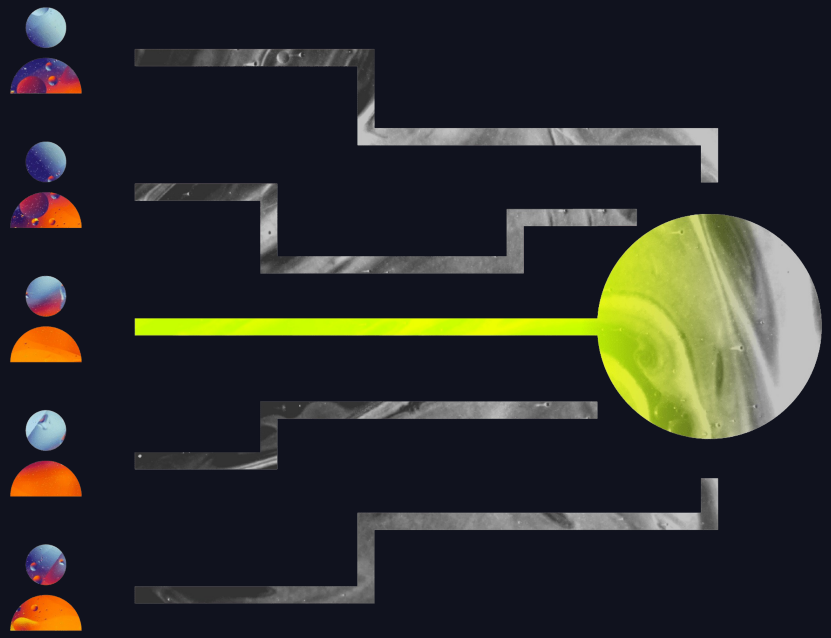
The Fundamental Mismatch

Diseases are singularly defined based on end pathology, but most represent **diverse groupings of contributing factors**.

Disease heterogeneity significantly impacts:

- **EFFICACY**
  Leads to over-treatment of non-responder populations
- **SAFETY**
  Non-responders absorb all the risk of drug without benefit
- **COST**
  From sizing trials to include enough responder patients

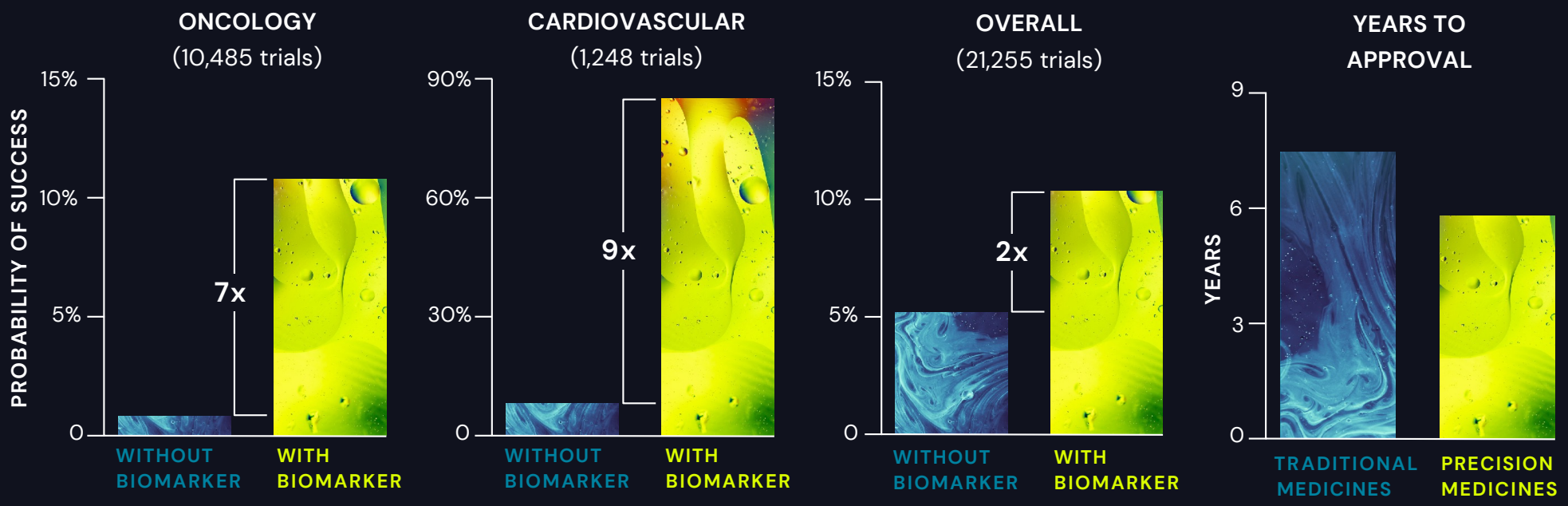
How can we overcome this challenge?
With biomarkers.



Circulating biomarkers align:

- **THE RIGHT PATIENT**
  To identify those most likely to respond to therapy
- **THE RIGHT DISEASE**
  To elucidate contributing factors & biological pathways
- **THE RIGHT THERAPY**
  To optimize dosing and timing of treatment

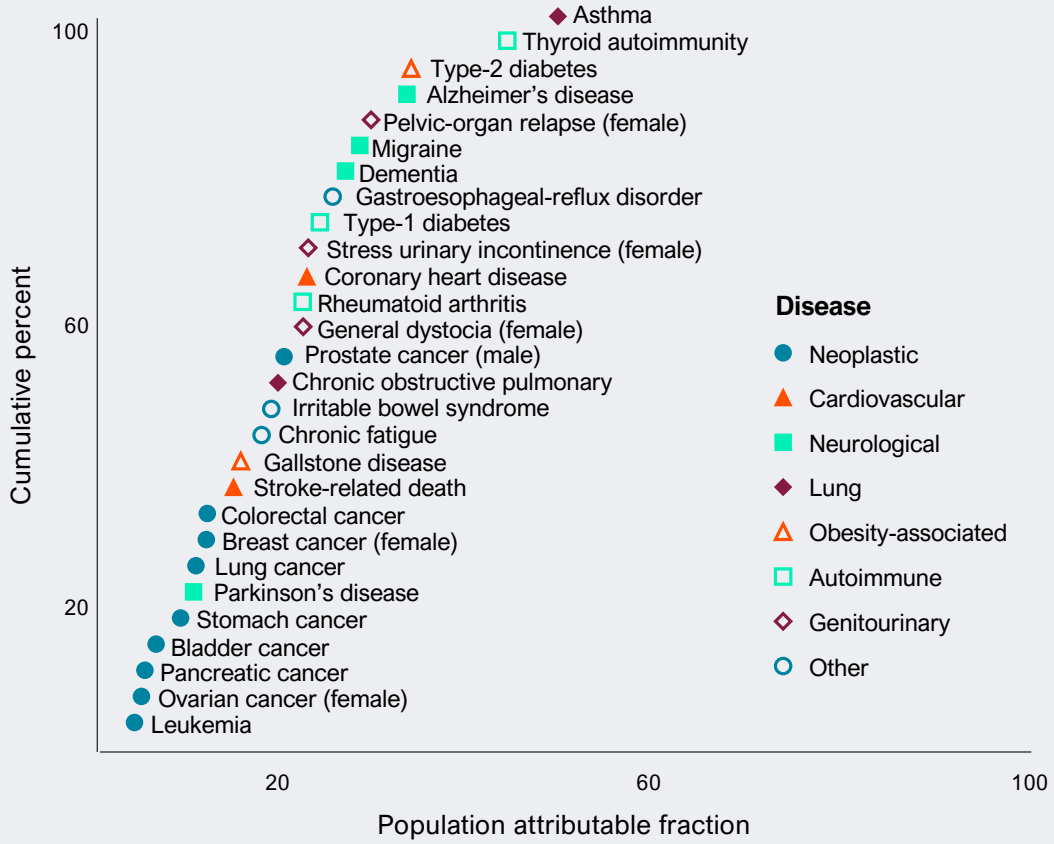
Biomarkers increase FDA approval rates >2–10x & drive 22% faster time to approval.



Wong et al. (2019). Estimation of clinical trial success rates and related parameters. *Biostatistics,* 20(2), 273–286. https://pubmed.ncbi.nlm.nih.gov/29394327/

Underlying risks for most human diseases are only partially encoded in genomes.

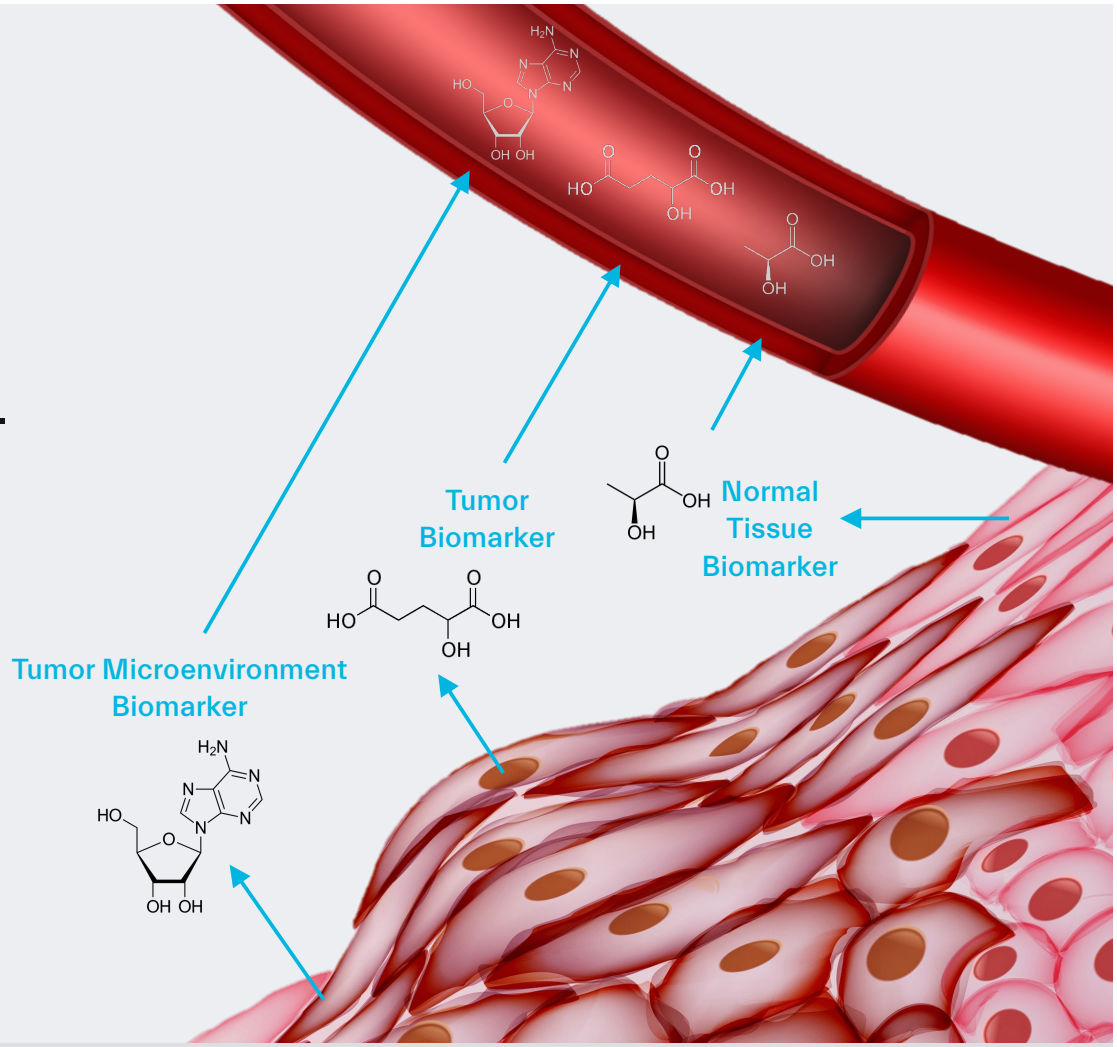
>85% of risk factors and biomarkers for disease are **still unexplored.**



Rappaport (2016). Genetic factors are not the major causes of chronic diseases. *PLoS One* 11, no. 4 (2016): e0154387. https://pubmed.ncbi.nlm.nih.gov/27105432/

Small molecule biomarkers in blood read out genetic and non-genetic factors that contribute to human disease.

- **REFLECT** real-time disease processes
- **RELEASE** from tissue into central circulation for non-invasive capture

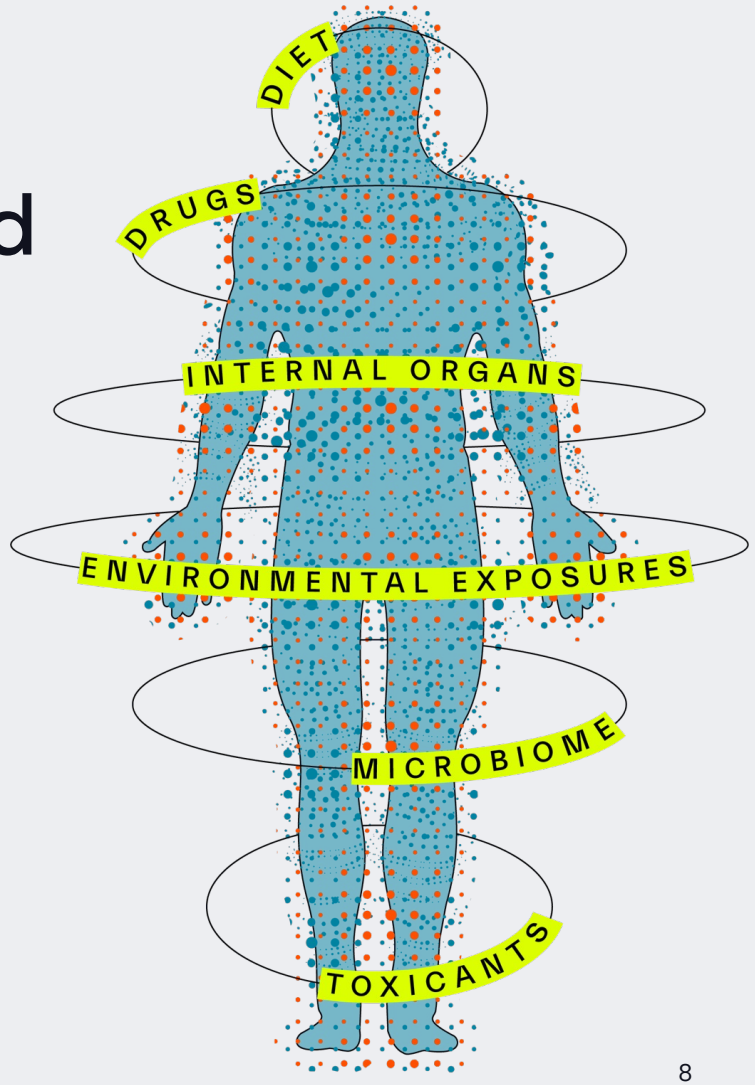


Human blood contains thousands of yet unmapped small molecule biomarkers.

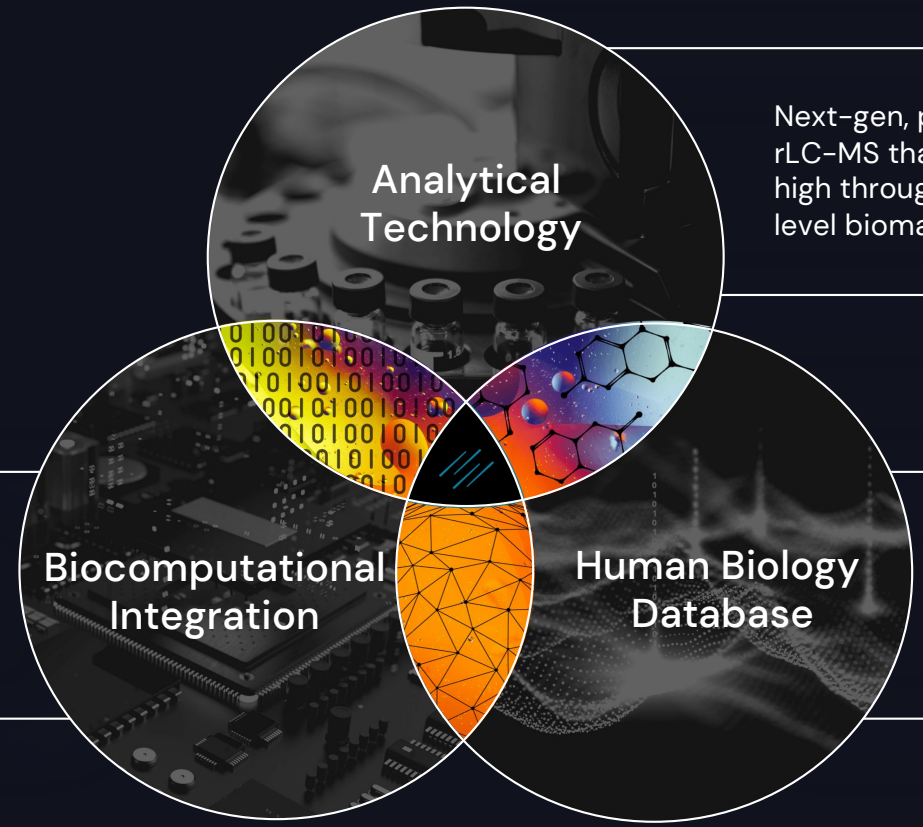
They hold the potential to elucidate human health and disease across populations, informing on:

- **Host** organ level physiology
- Local **disease** processes
- Effects of **exposures** stemming from both lifestyle and environment

Herein we describe a new approach to map the chemistry of blood at population scale.



Our platform integrates 3 core components.



Next-gen, proprietary rLC-MS that enables ultra high throughput, population-level biomarker discovery

Expert data science team that applies statistical and ML tools to identify key biomarkers that drive actionable insight

Encompassing data from >100,000 human biosamples for cross-validation of biomarkers

Analytical Technology

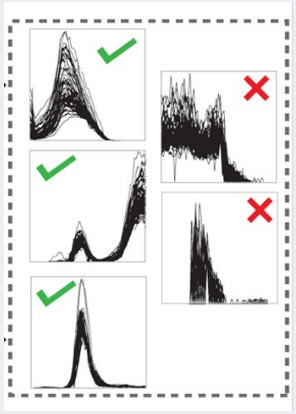
Rapid liquid chromatography-mass spectrometry (rLC-MS)

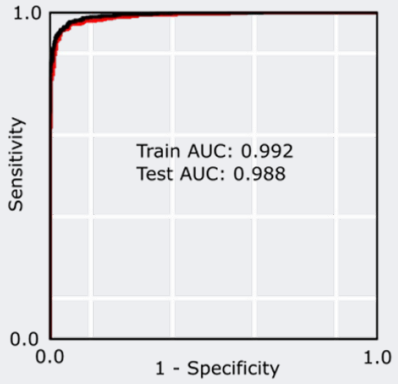
- **Sensitivity** to measure >11,000 small molecule biomarkers
- **Speed** with analytical cycle time of <1 minute per sample
- **Capture** of broad chemistries
- **Capacity** to handle thousands of biosamples

Biocomputational Integration

Using statistical and machine learning tools

- **Integration of high-dimensional data:**
  - Preclinical models
  - Genomics
  - Human biology
  - Clinical outcomes





Powered by a Proprietary Mass Spectrometry Data Extraction Pipeline

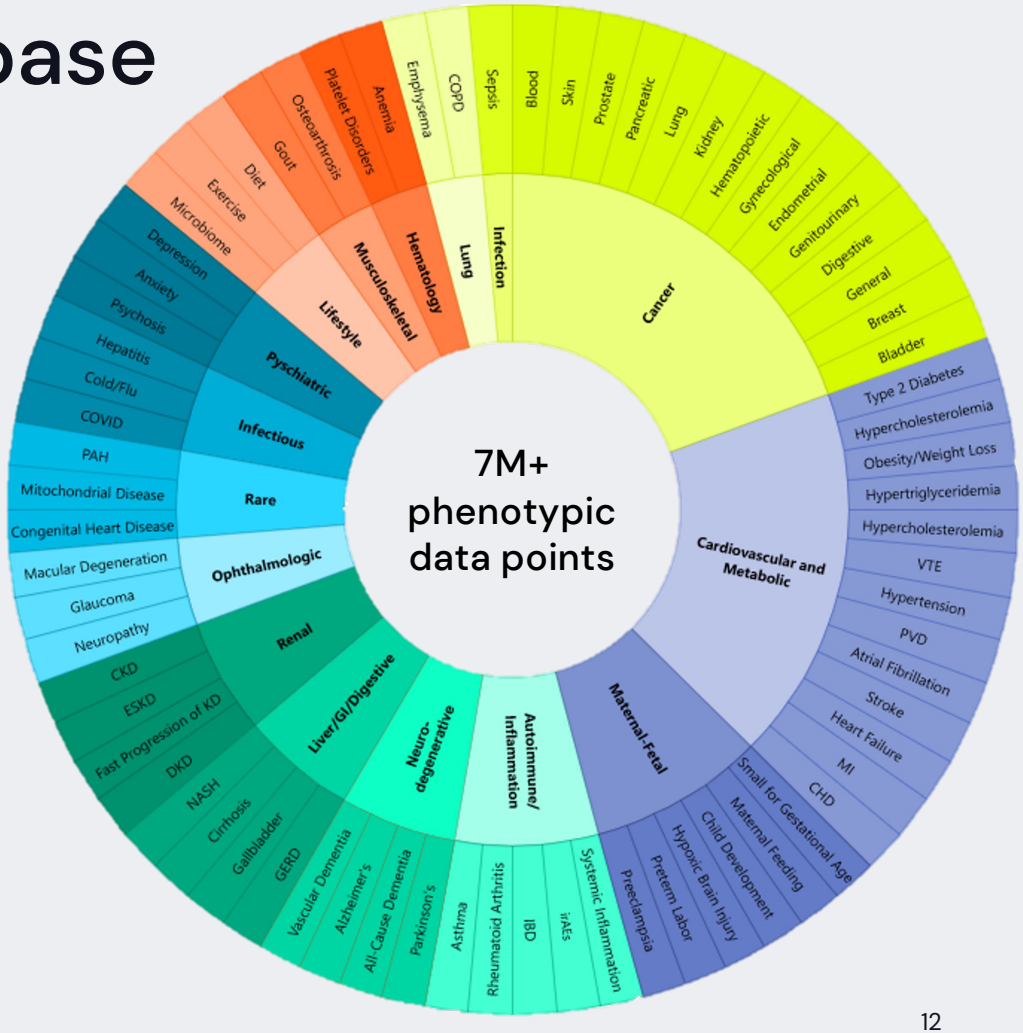
- **Scalable** computing clusters
- Machine learning to remove **up to 90%** of false spectral peaks without reducing true signals
- **Neural network**-based pattern recognition
- **Distributed** image processing

Kantz et al (2019). Deep Neural Networks for Classification of LC-MS Spectral Peaks. *An Chem* 91 (19), 12407-12413. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7089603/

Human Biology Database

Built from molecular profiling of samples from >100,000 individuals

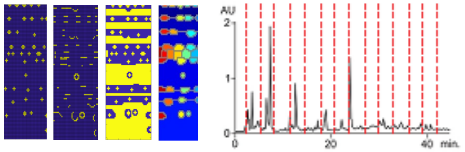
- Cohorts across **diverse** backgrounds, demographics, geographies, lifestyles, diseases, and medication regimens
- **Decades of follow-up** across patients with data on diet, exercise, mental health, clinical outcomes, etc.
- **Genetic** information for a subset of subjects and **microbiome** sequencing data



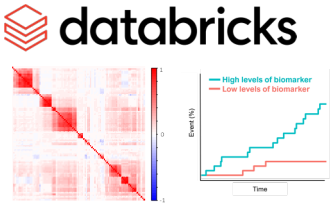
Embedded in an integrated discovery platform

Sample Intake, LIMS, and rLC-MS Analysis to Measure >11,000 Small Molecule Biomarkers per Blood Sample

ETL from On-Premise to AWS Cloud



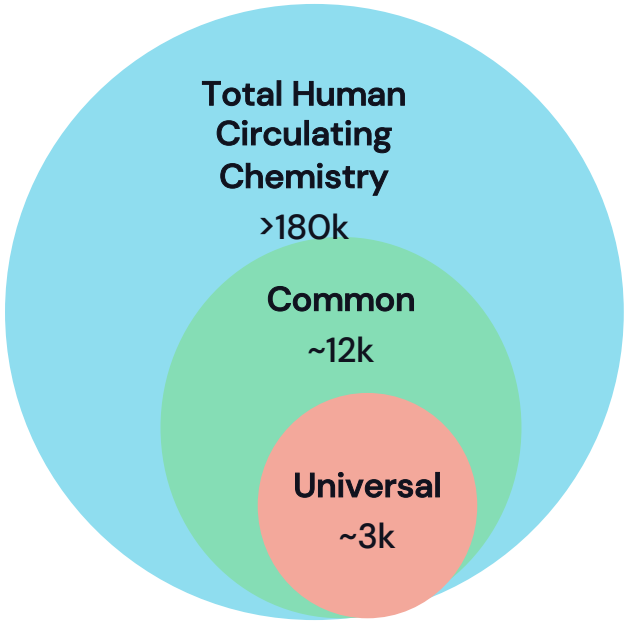
ML and Distributed Computing-based Feature Extraction

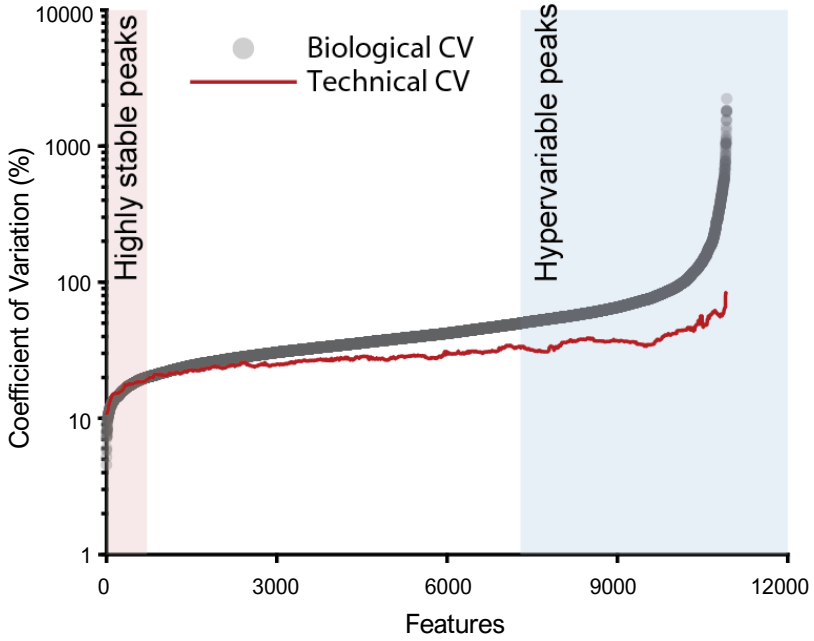


Databricks Data Science Environment and Lakehouse to Derive Insight

Learnings from large-scale biomarker profiling

Widespread human chemical diversity



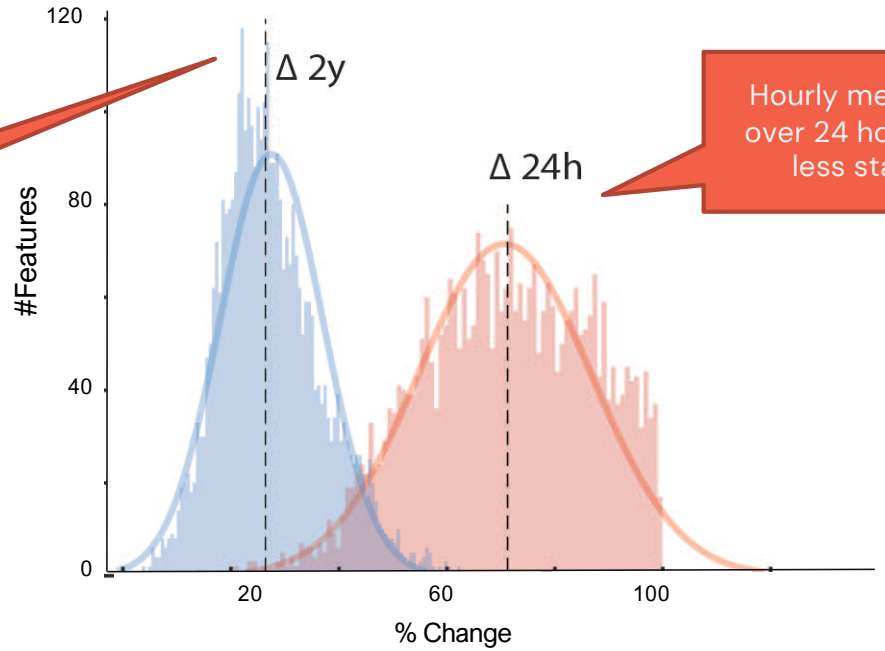


Learnings from large-scale biomarker profiling

Features vary more over a day than over years

Stability of features in 412 individuals

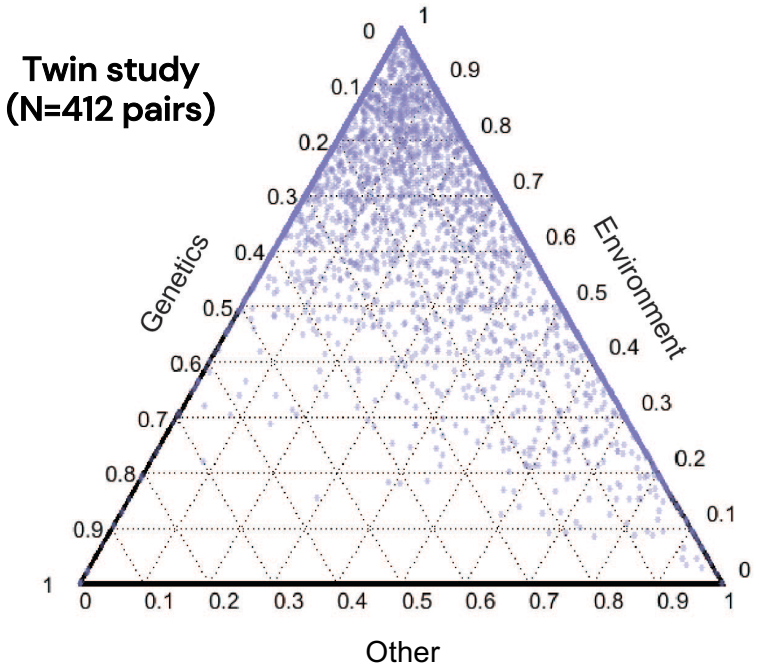
Fasting monthly measures over 24 months are stable

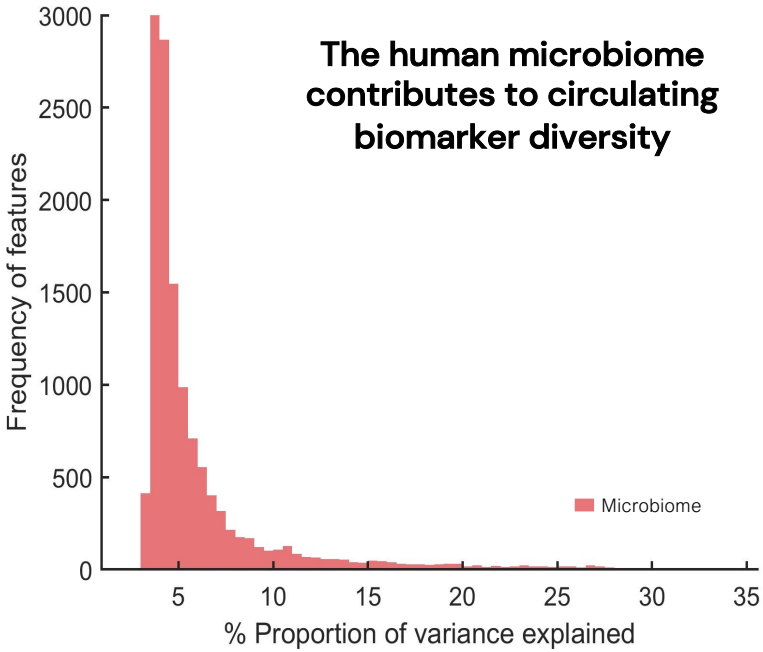


Hourly measures over 24 hours are less stable

Learnings from large-scale biomarker profiling

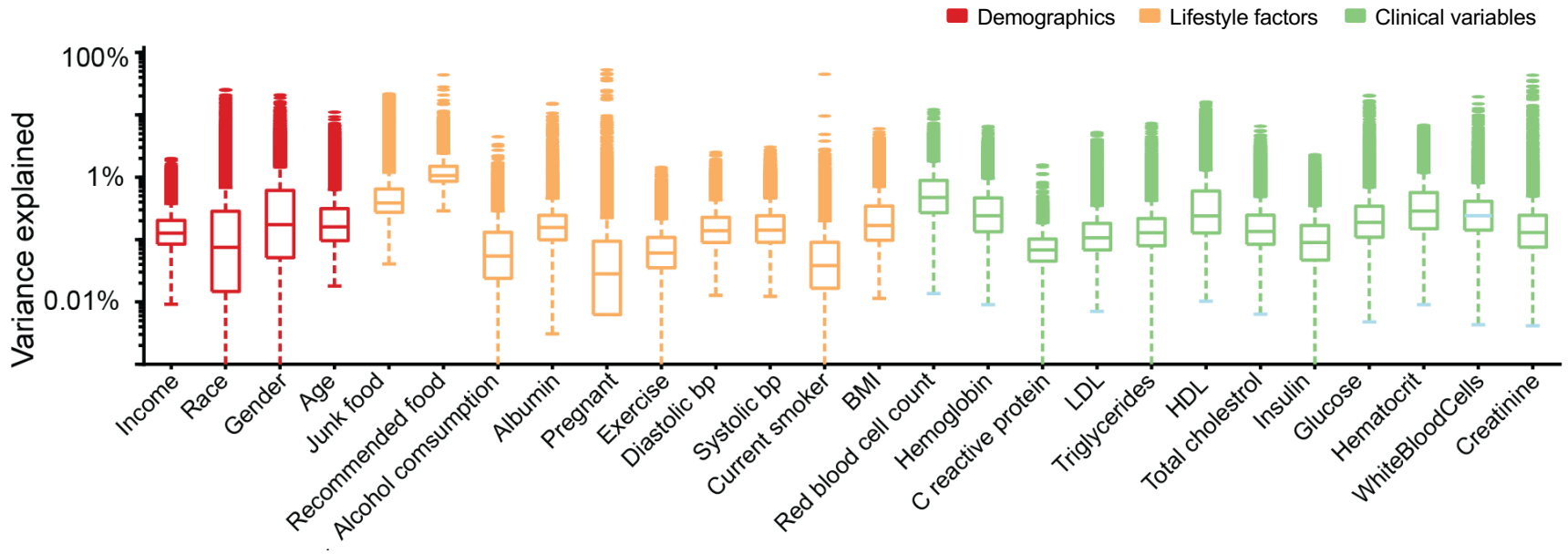
Most variation is attributable to environmental factors





Learnings from large-scale biomarker profiling

Demographic, lifestyle, and clinical variables contribute to overall circulating biomarkers



Case Study

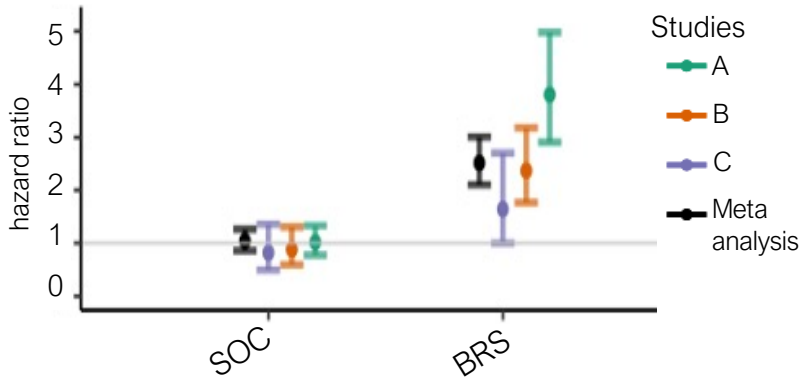
Discovery of a biomarker for disease prognosis

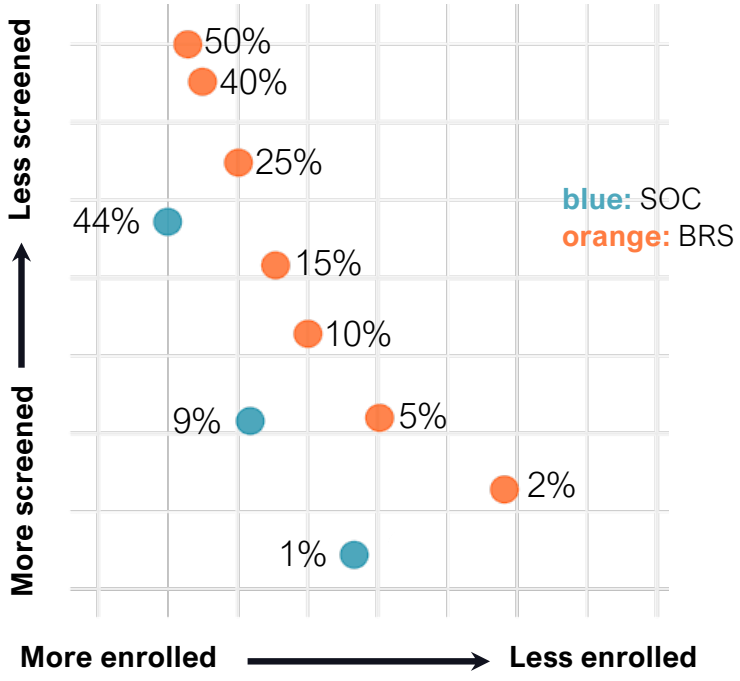
Use biomarker risk score for clinical trial enrichment

Selecting individuals at high risk of disease-related event

SOC: standard of care

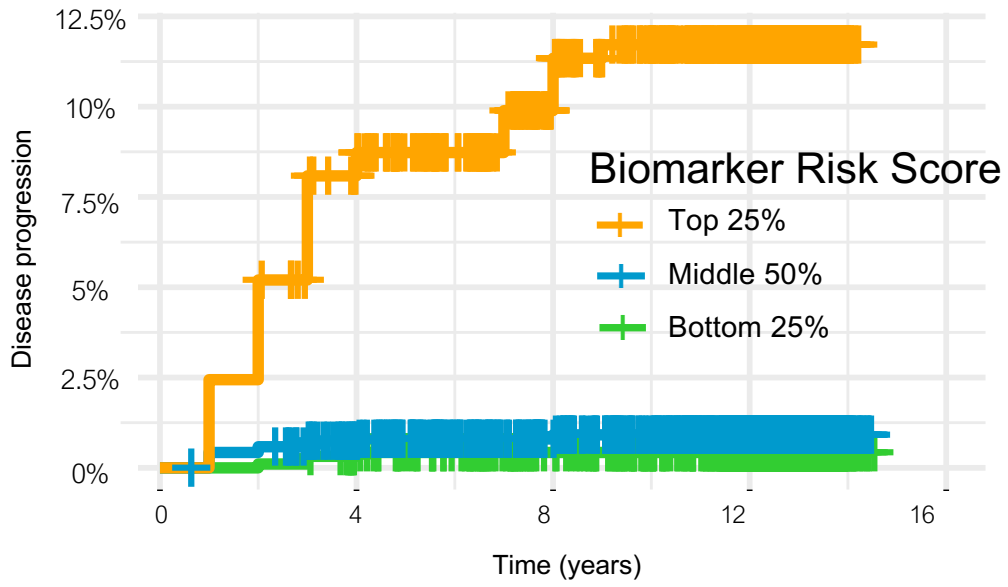
BRS: Biomarker Risk Score





Longitudinal human database discovery

Leveraging Sapient's Human Biology Database



Patient Stratification

Across multiple diverse populations, Sapient's Human Biology Database confirms the key circulating biomarkers predict individuals at risk for rapid disease progression.

DATA+AI SUMMIT 2022

Thank you!

Any questions?

Tao Long, PhD, MBA
Co-Founder & Head of Data Science, Sapient