

CALIBIR Bootcamp #3:

translational bioinformatics

Nicholette (Palmer) Allred

Professor, Department of Biochemistry

Nicholette.Allred@AdvocateHealth.org →
Nicholette.Allred@wfusm.edu



You

Explain Translational Bioinformatics



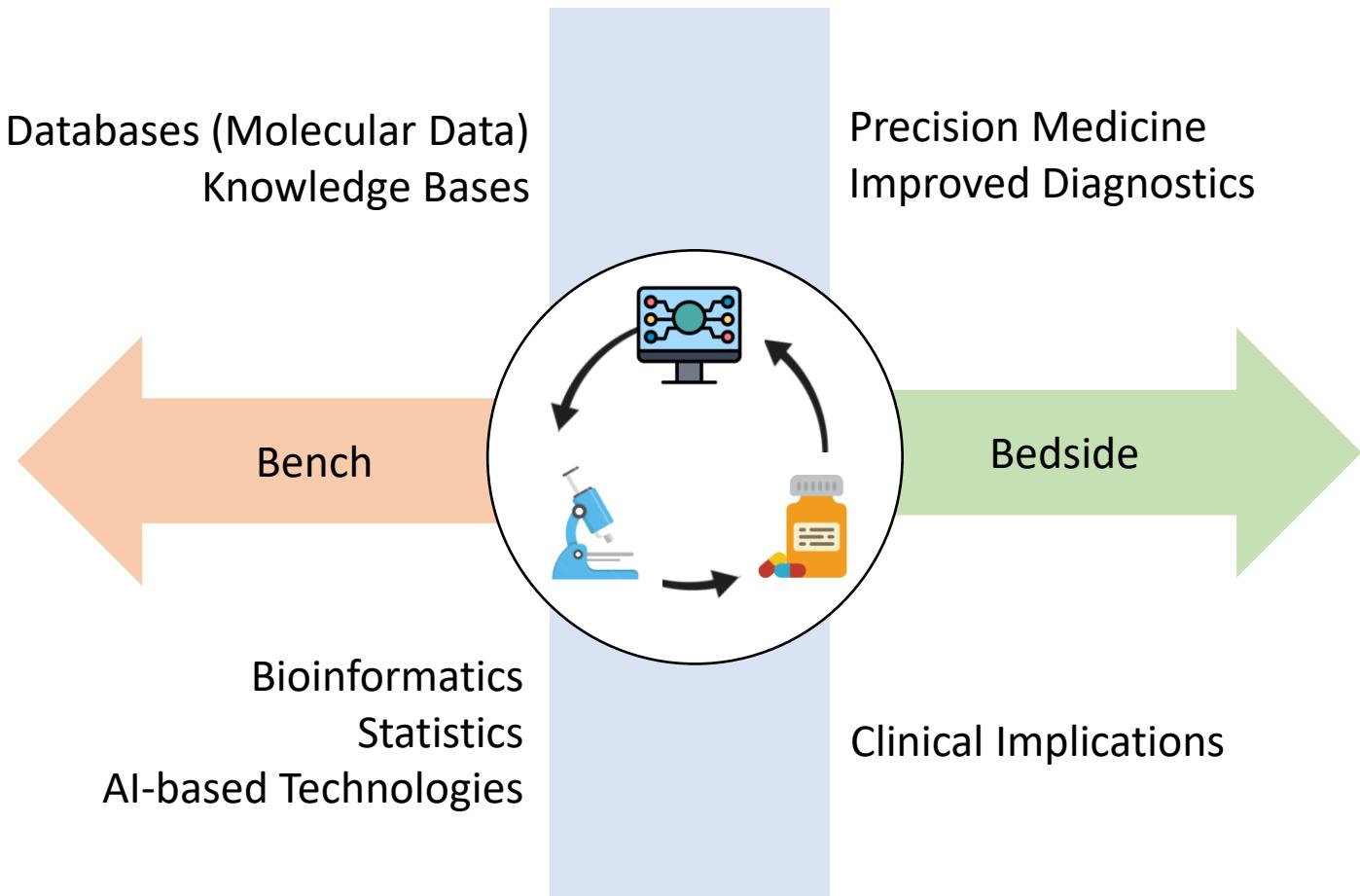
ChatGPT

Translational bioinformatics is a **multidisciplinary field** that applies computational techniques to biomedical data with the goal of translating biological findings into clinical applications. It involves the **integration of biological and clinical data using computational methods** to accelerate biomedical research and improve patient care. Here's a breakdown of its key components:

1. **Biomedical Data Integration**: Translational bioinformatics involves integrating various types of biomedical data, including **genomic, proteomic, transcriptomic**, and clinical data. This integration enables researchers and clinicians to gain comprehensive insights into disease mechanisms, biomarkers, and potential therapeutic targets.
2. **Data Mining and Analysis**: **Computational techniques** such as machine learning, data mining, and statistical analysis are applied to analyze large-scale biomedical datasets. These methods help identify patterns, correlations, and associations within the data, facilitating the discovery of novel biological insights and clinical applications.
3. **Biological Interpretation**: Translational bioinformatics aims to translate complex biological data into actionable knowledge that can inform clinical decision-making. By integrating computational analysis with **biological knowledge and expertise**, researchers can decipher the underlying mechanisms of diseases and identify potential interventions or personalized treatment strategies.
4. **Clinical Applications**: One of the primary objectives of translational bioinformatics is to bridge the gap between basic biomedical research and clinical practice. This involves developing computational tools and algorithms that can assist clinicians in **diagnosis, prognosis, treatment selection, and patient management** based on individual genetic profiles and clinical data.
5. **Precision Medicine**: Translational bioinformatics plays a crucial role in advancing precision medicine initiatives by leveraging genomic and clinical data to tailor medical interventions to the specific characteristics of individual patients. This personalized approach to healthcare aims to **improve treatment outcomes, minimize adverse effects, and optimize patient care**.

Overall, translational bioinformatics serves as a powerful framework for **translating biological knowledge into clinical practice**, ultimately contributing to the development of more effective diagnostics, therapeutics, and preventive strategies for various diseases.

Translational Bioinformatics



Components of Translational Research

- Big Data



Health Care

- Omics
- Publicly available repositories

Bootcamp #1: Healthcare Informatics
Drs. Rezapour & Lu

Bootcamp #2: Clinical Informatics
Drs. Ma

- Analytical Tools

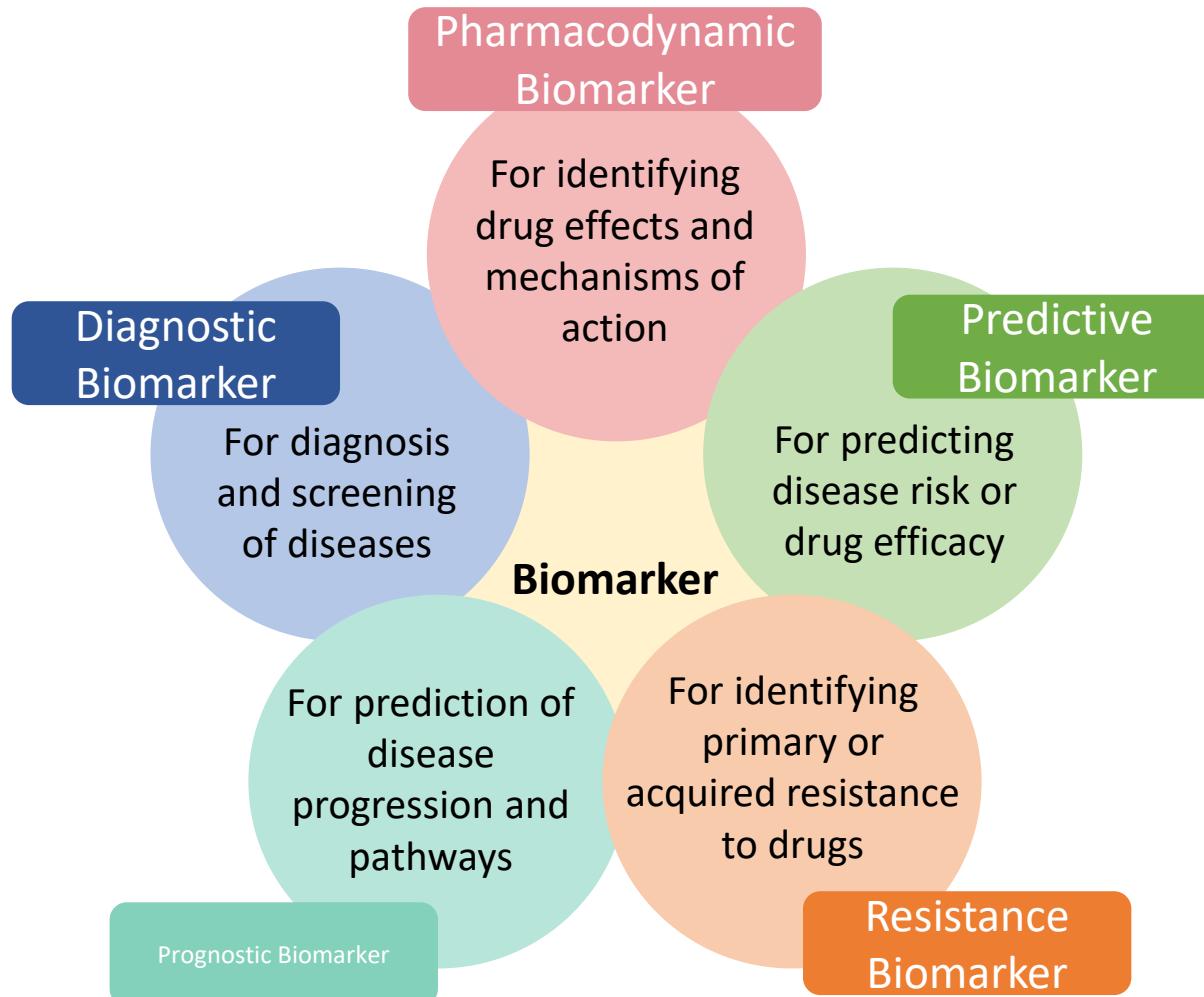
- Biostatistics



Artificial intelligence-based technologies

Bootcamp #4: Artificial Intelligence
Drs. Movaghari, Lu & Koyun

Translational Bioinformatics



How do we identify biomarkers?

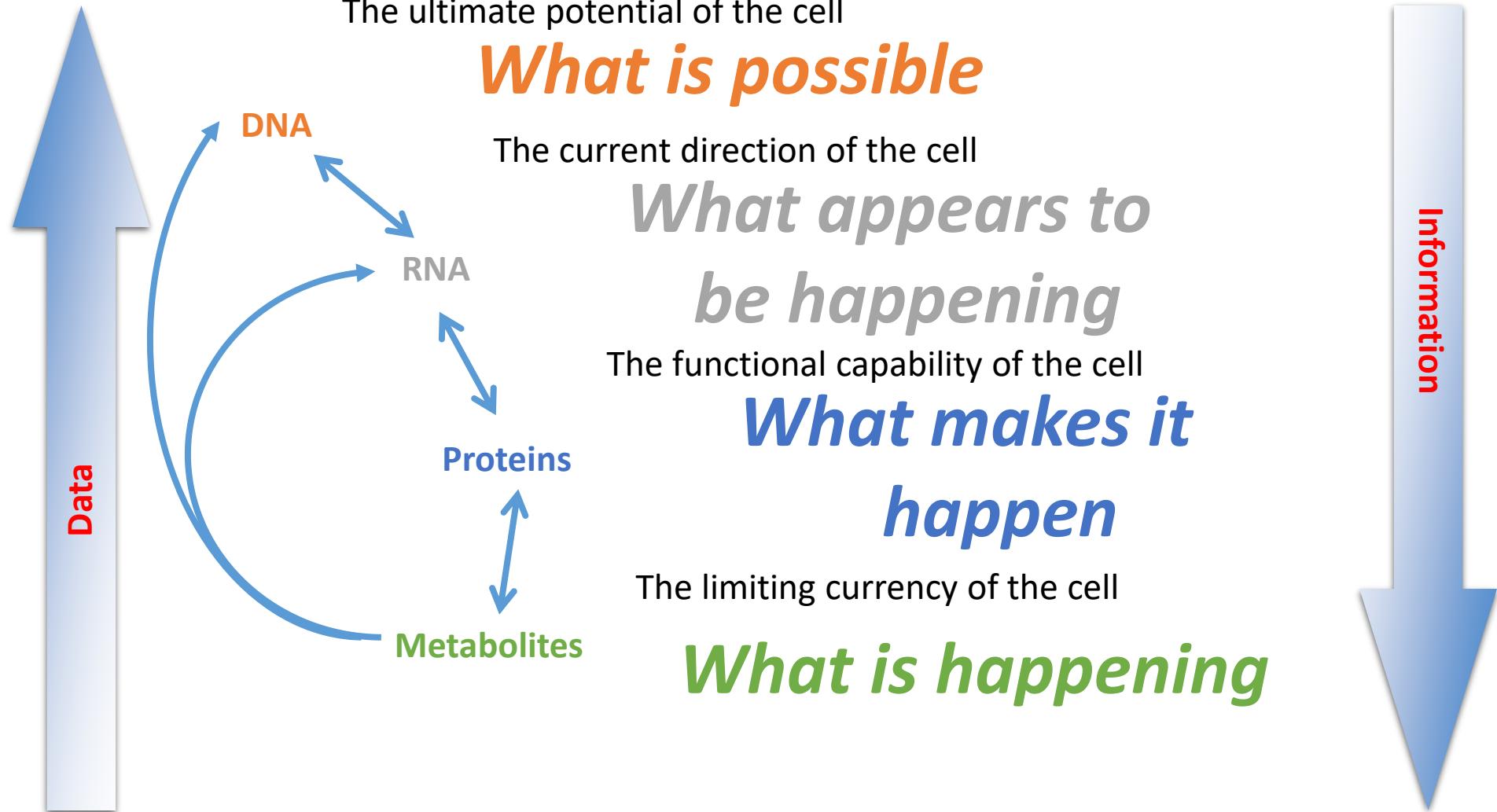


Omics is a rapidly evolving, multi-disciplinary, and emerging field that encompasses genomics, epigenomics, transcriptomics, proteomics, and metabolomics.

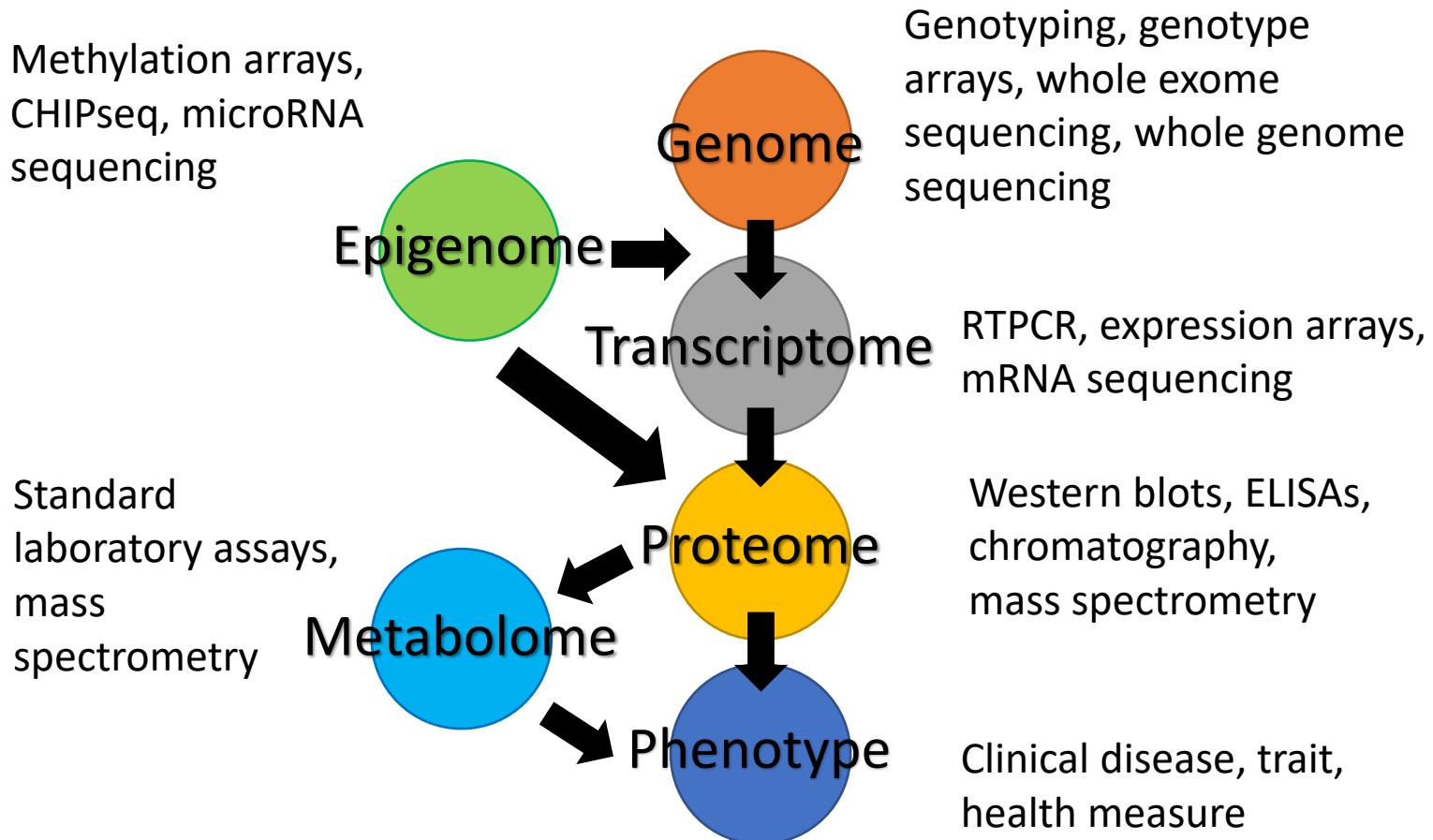
Each omics technology offers the possibility to understand and view biology from a global perspective.

High-throughput methodologies can rapidly provide a global picture of the processes within cells at multiple levels, allowing for accelerated discoveries in health and disease.

Omics



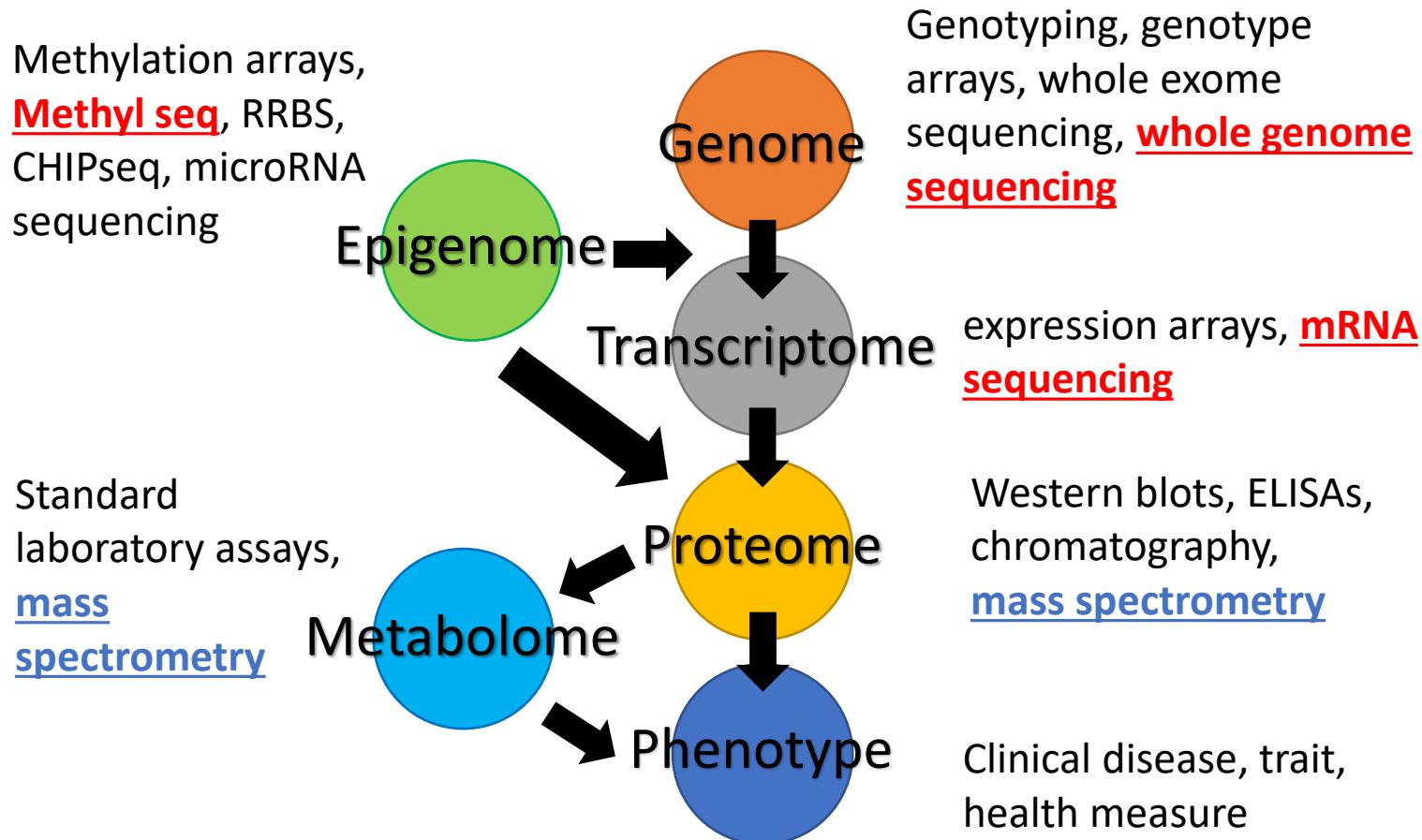
The Central Dogma of Biology



Utility for Understanding Disease

- DNA
 - Largely stable over time and across tissues
 - Highly predictive for monogenic disorders, less so complex adult-onset traits
- Transcripts, protein levels, epigenetics, metabolomics
 - More challenging to collect
 - Directionality of effect unclear
 - Tissue and cell-type specific
 - Changes over time and in response to environment

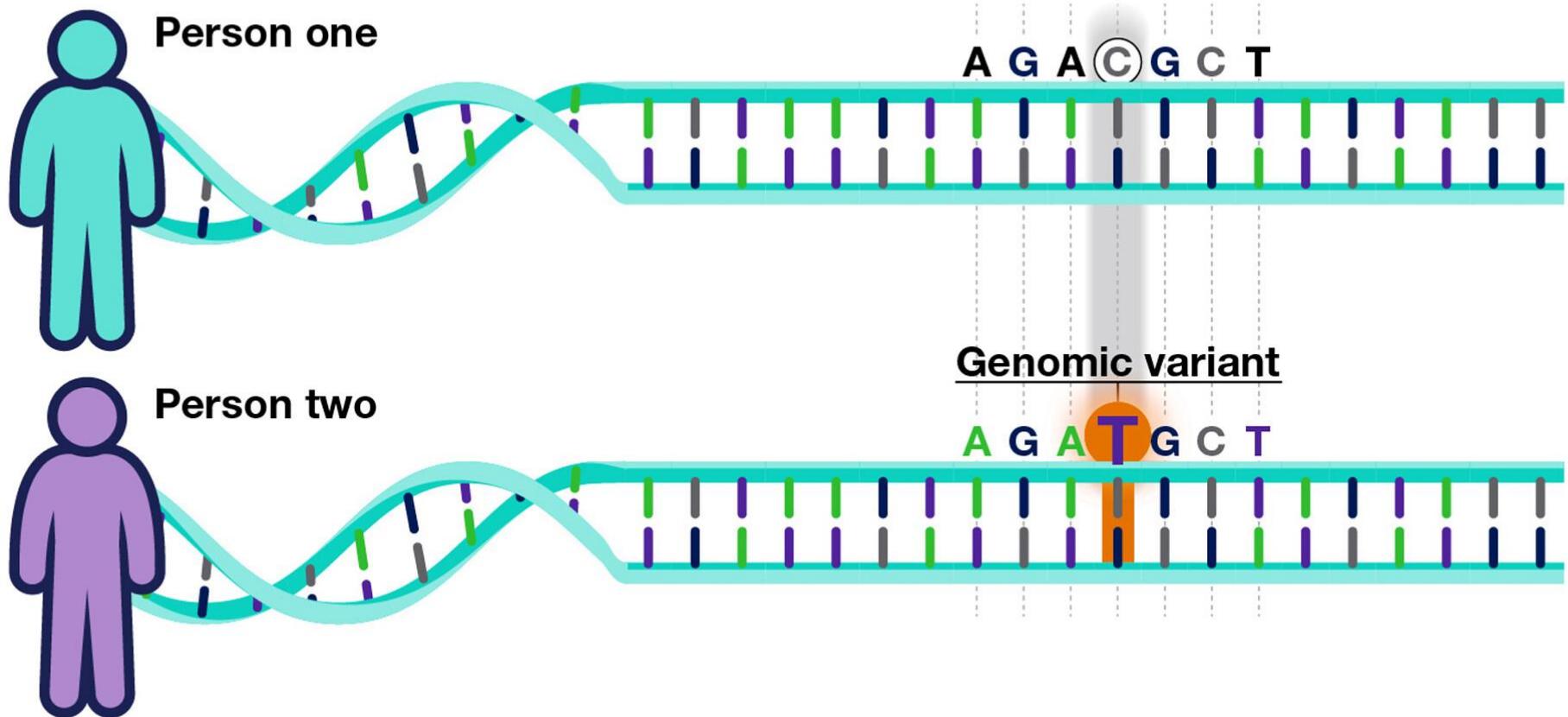
The Central Dogma of Biology



DNA Sequence



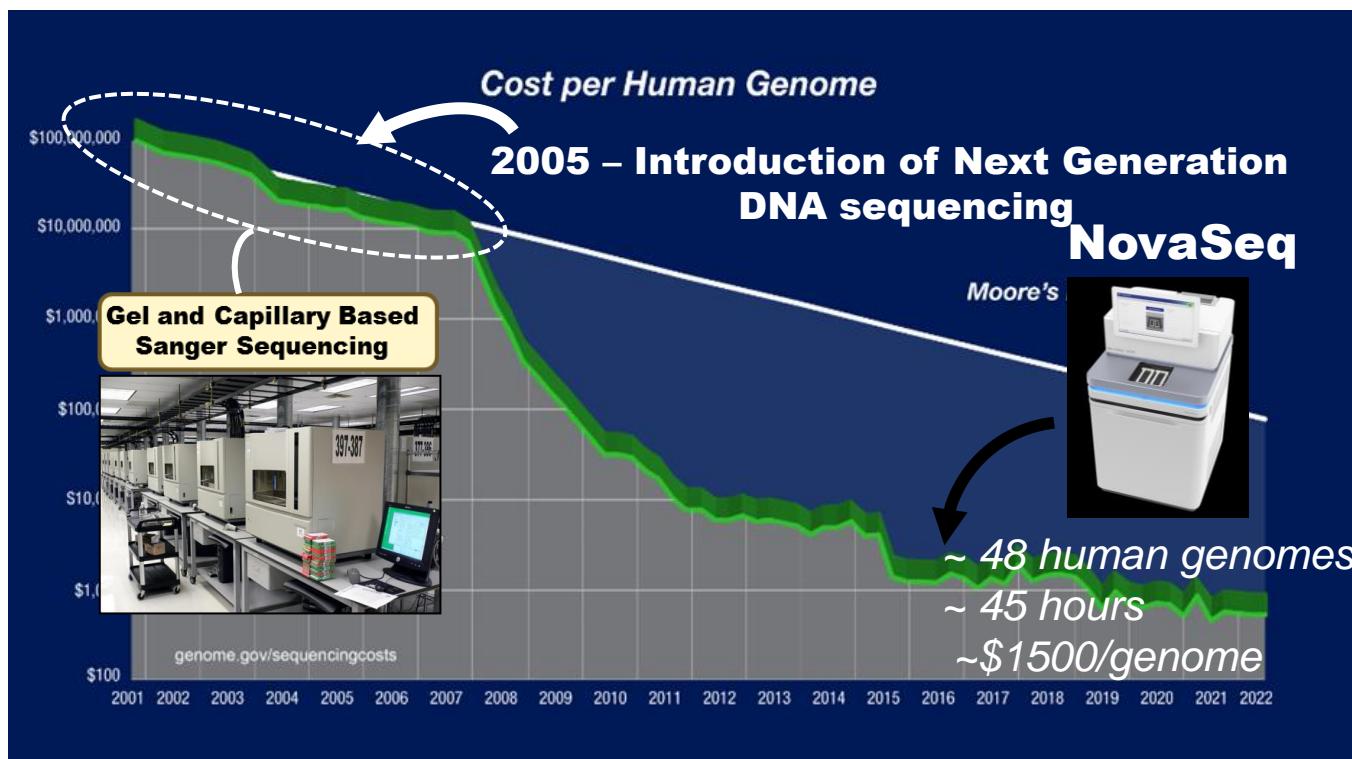
DNA Sequence



History

Sequence-based Omics

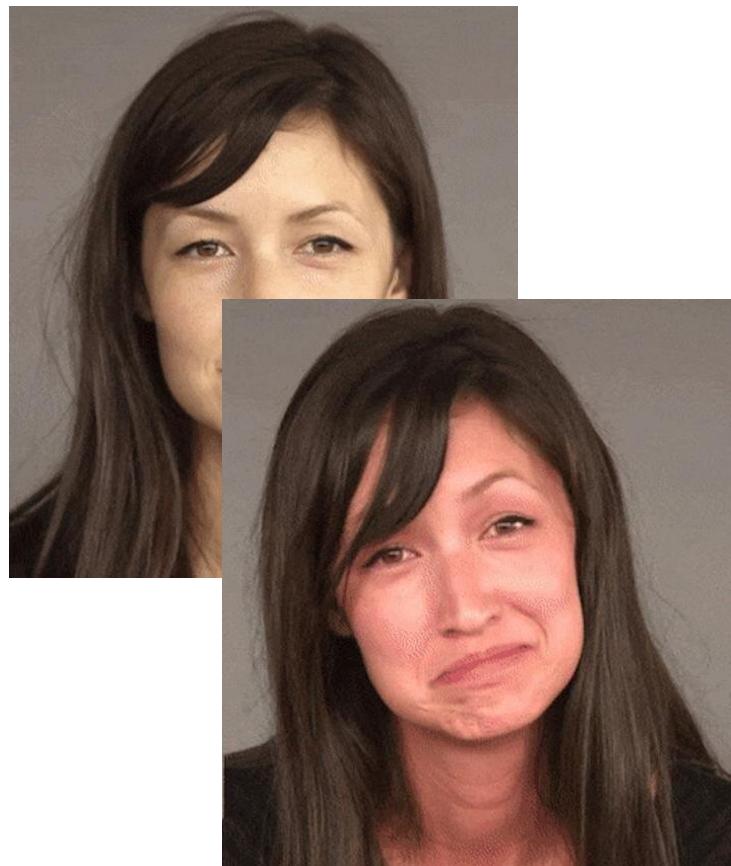
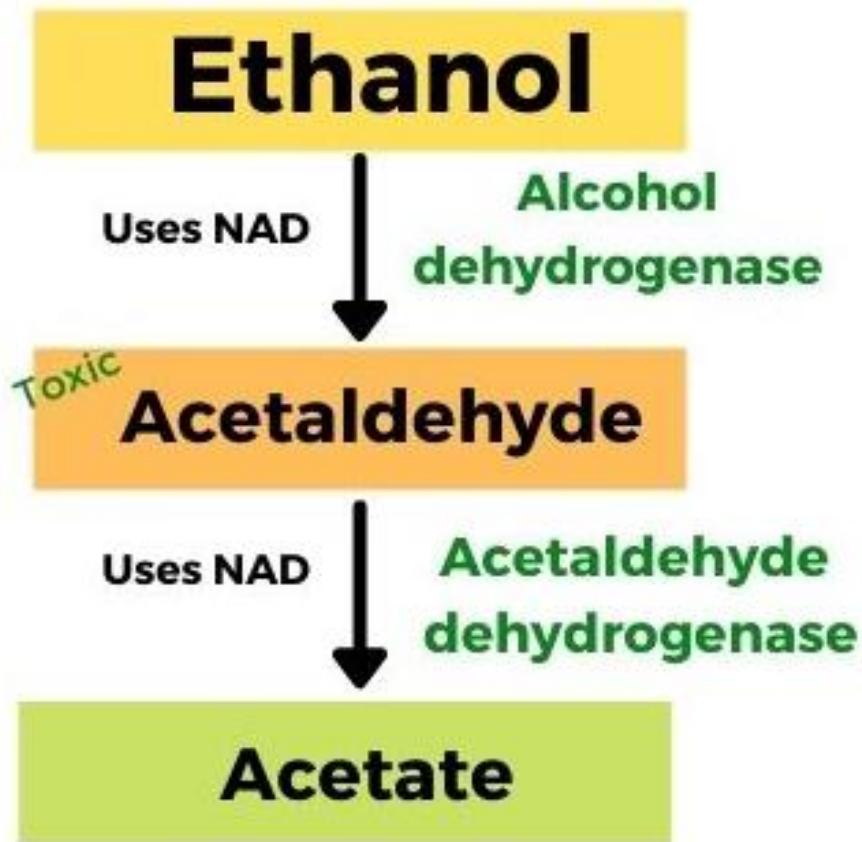
- Cost over time: First human genome sequence draft took ~10 years (1990-2000)*, \$2.7 billion and involved 20 genome sequencing centers to complete



DNA Sequencing

- Sample
 - blood, skin, buccal swabs, saliva, tissue; relatively stable
- Cost
 - 30X coverage >\$700/sample (germline applications)
- Data
 - Raw data >1Tb (.bcl2, .fastq; cloud computing)
 - Processed data: Mb-Gb (.bam, .vcf)
 - Annotation files: Mb (annovar)
- Repositories
 - GenBank – hundreds of organisms sequenced
 - dbGAP – interaction of genotype-phenotype

Alcohol Flush



How does

Global Genetic Testing Work?

Drivers

- Increase in prevalence of genetic disorders
- Surge in demand for personalized medicine

By Test Type

Prenatal Testing

- Dominant market share in 2022

By Service Provider

Hospital-based Laboratories

- Relatively high CAGR during the forecast period



□ Hospital-based Laboratories
■ Diagnostic Laboratories
■ Speciality Clinics
■ Others

23andMe

My Home

Inbox

Health

- Clinical Reports
- Research Reports
- Health Labs

Ancestry

- Maternal Line
- Paternal Line
- Relative Finder
- Ancestry Painting
- Global Similarity
- Ancestry Labs

Sharing & Community

- Compare Genes
- Family Inheritance
- 23andMe Community

23andWe

- My Surveys (30)
- Research Initiatives

Search

clinical reports

Show results for Paul Stamatou

Print summary of elevated risks

Disease Risks (12) ?

- Celiac Disease
- Venous Thromboembolism
- Prostate Cancer ♂
- Crohn's Disease
- Psoriasis
- 1 locked report

See all 12 risk reports...

Carrier Status (21) ?

Condition	Status
Alpha-1 Antitrypsin Deficiency	Variant Absent
Bloom's Syndrome	Variant Absent
Canavan Disease	Variant Absent
Connexin 26-Related Sensorineural Hearing Loss	Variant Absent
Cystic Fibrosis	Variant Absent
1 locked report	

See all 21 carrier status...

Traits (10) ?

Trait	Description
Alcohol Flush Reaction	Does Not Flush
Bitter Taste Perception	Unlikely to Taste
Earwax Type	Wet
Eye Color	Likely Brown
Lactose Intolerance	Likely Tolerant

See all 10 traits...

Drug Response (8) ?

Drug	Response
Clopidogrel (Plavix®) Efficacy	Unknown
Warfarin (Coumadin®) Sensitivity	Increased
Alcohol Consumption, Smoking and Risk of Esophageal Cancer	Typical
Response to Hepatitis C Treatment	Typical
Abacavir Hypersensitivity	Typical

See all 8 drug response...

In emerging economies

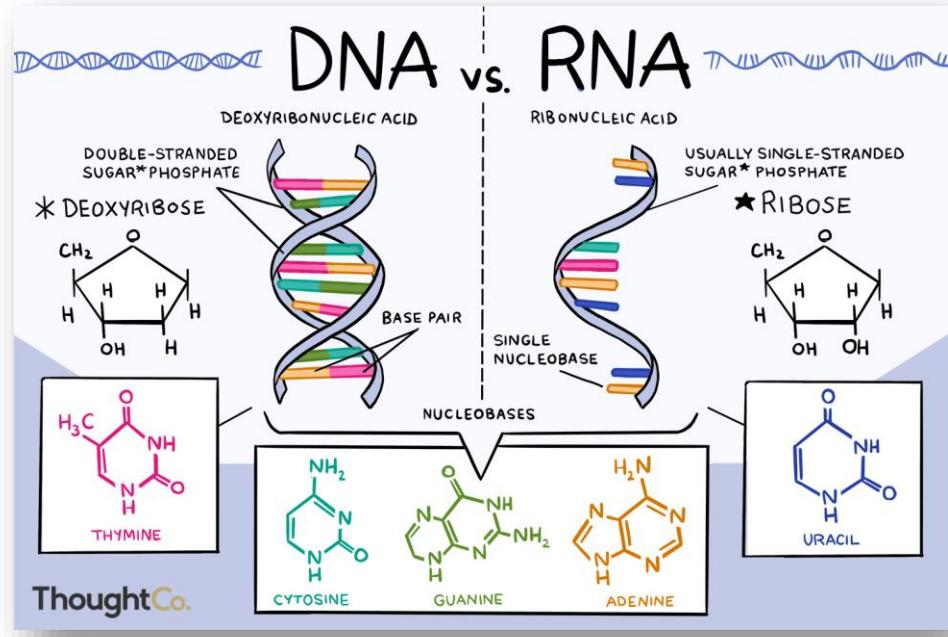
Key Players



Asia Pacific

- Relatively high CAGR from 2023 to 2031





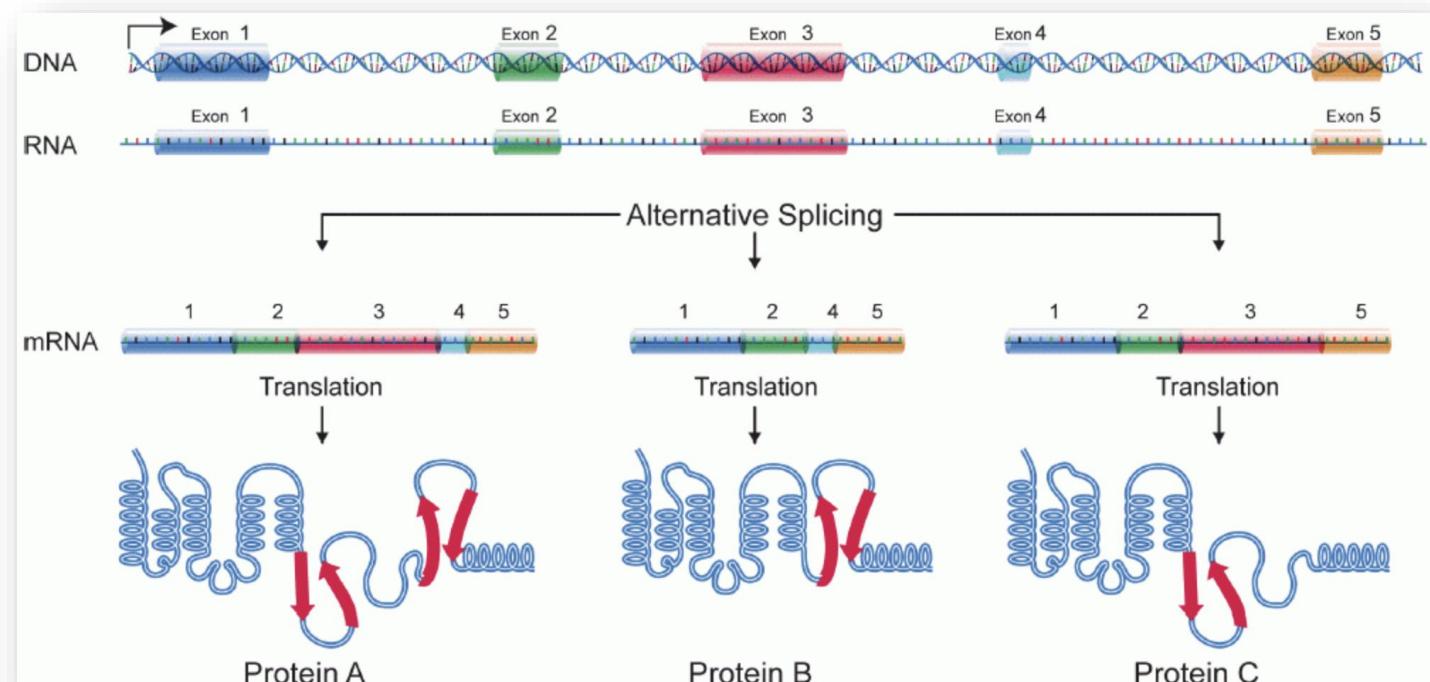
RNA Sequencing

RNA Sequencing

- Sample
 - blood, skin, buccal swabs, saliva, tissue; relatively unstable
- Cost
 - 20M-35M reads - \$350 - \$450/sample
- Data
 - Raw data >1Tb (flow cell dependent; .bcl2, .fastq; cloud computing)
- Gene Expression Data
 - GEO, ArrayExpress

Transcript and Transcript Splice Variant Synthesis is Dynamic

DNA variants
Tissue
Environment
Time



Why analyze the transcriptome?

- Snap-shot of entire transcriptome
- Cell/Tissue specific
- Evidence of genetic regulation
 - Transcript start site variants
 - Splice variants
 - RNA editing (with very high coverage + genome)
- Very sensitive, quantitative
- Accurate identification of transcripts based on sequence identity

Limitations of transcriptome analysis

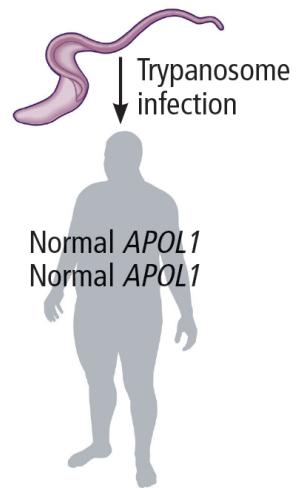
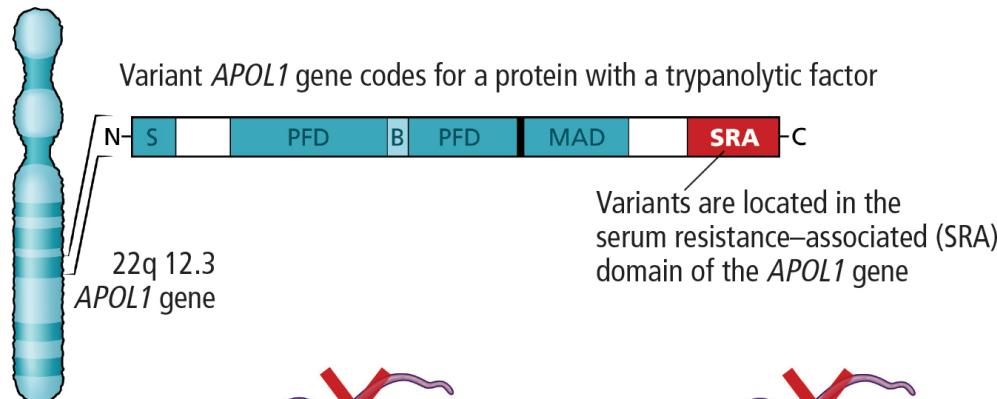
- Transcript abundance does not indicate transcript translational rates
- Transcript abundance does not always correlate with protein abundance
- Protein abundance or protein post-translational modifications may be more biologically relevant than transcript abundance
- There are currently no good algorithms to predict transcript-protein correlations

GIGO

- RNA is finicky and unstable - stability differs by transcript



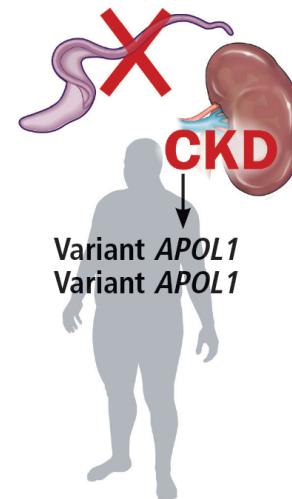
Application – Kidney Disease



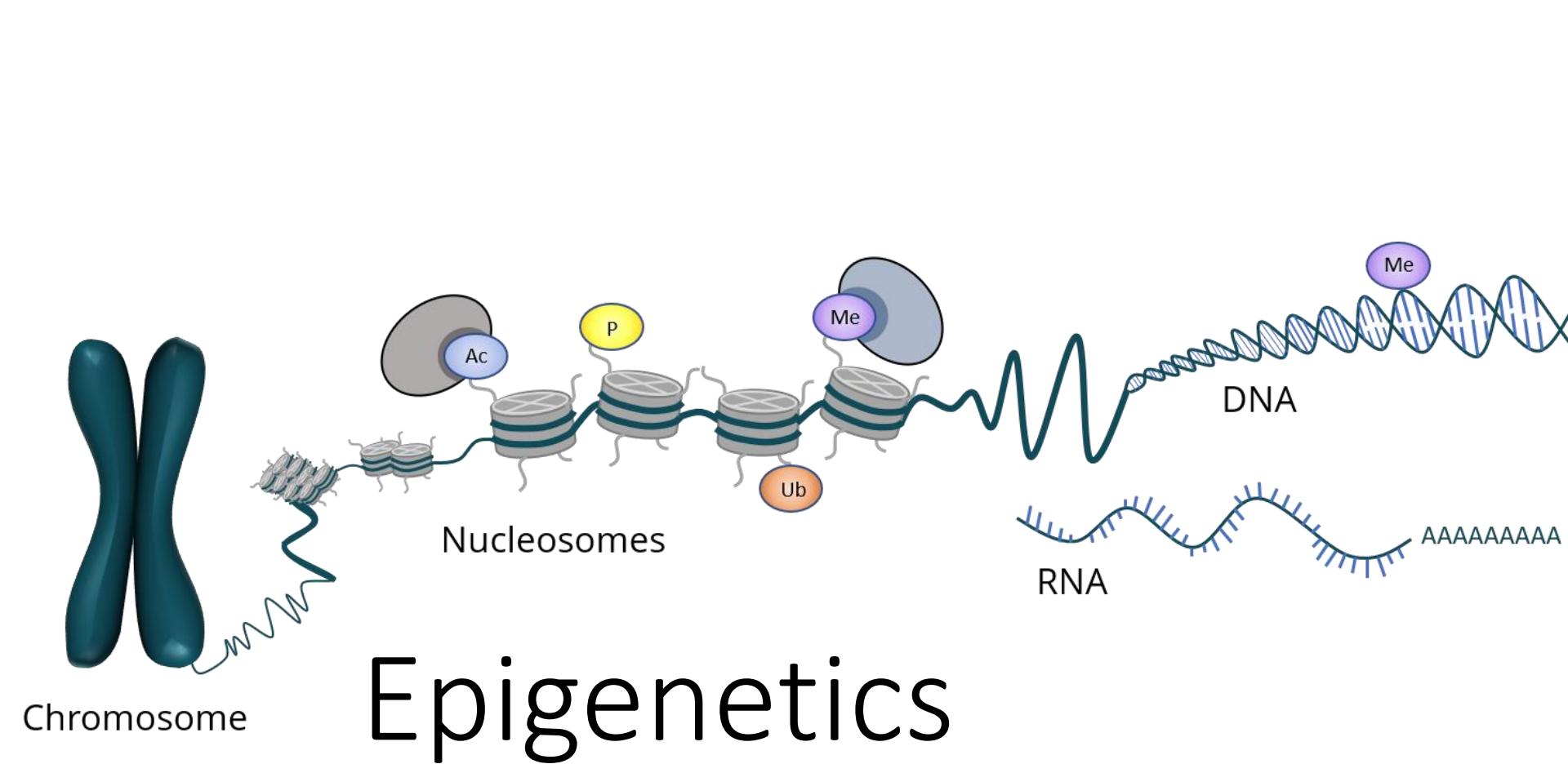
Individuals with 2 normal copies of *APOL1* can be infected by the parasite.



Having 1 variant allele confers protection against trypanosomiasis without leading to kidney disease.

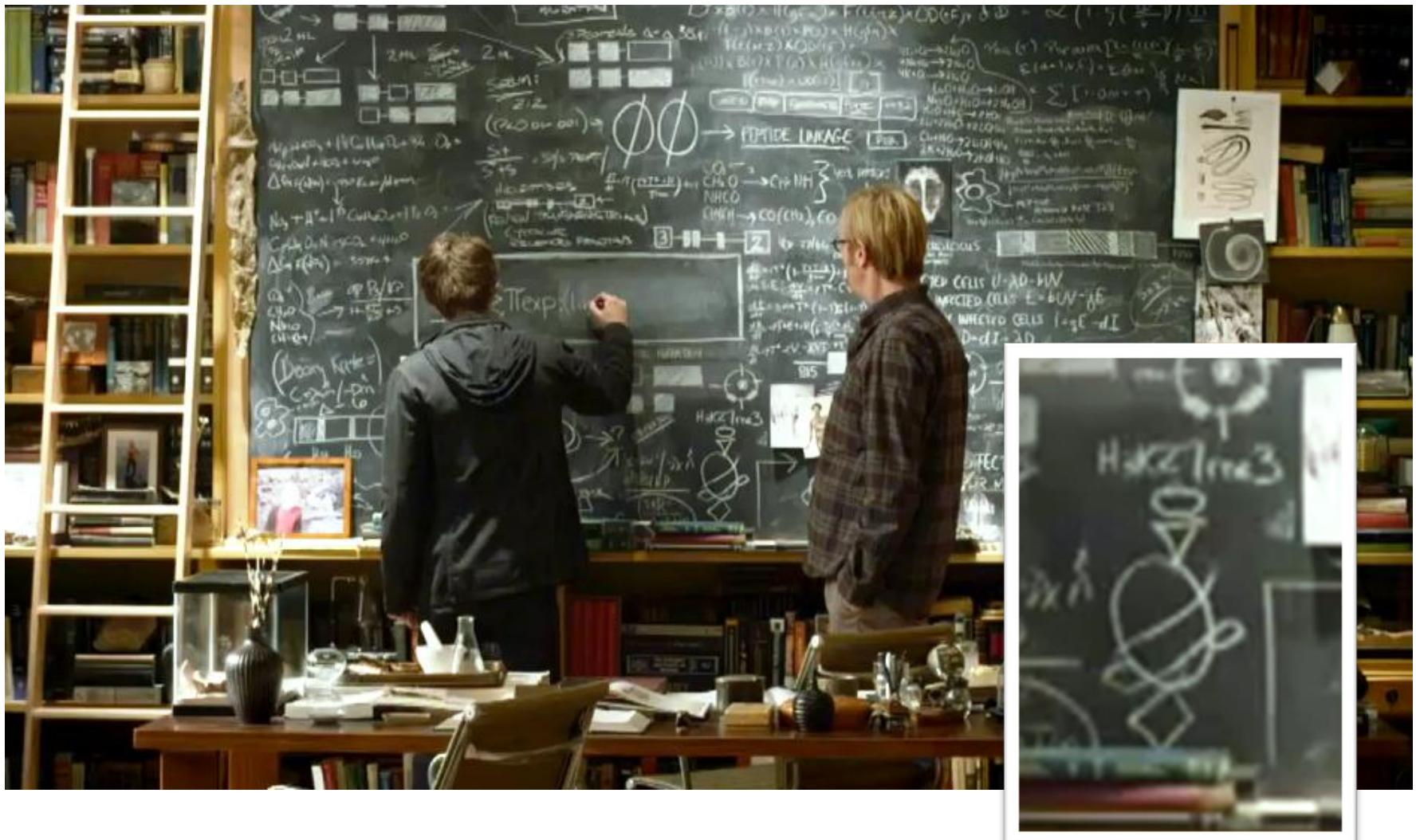


Having 2 variant alleles confers protection against trypanosomiasis but increases the risk of chronic kidney disease (CKD).



Epigenetics

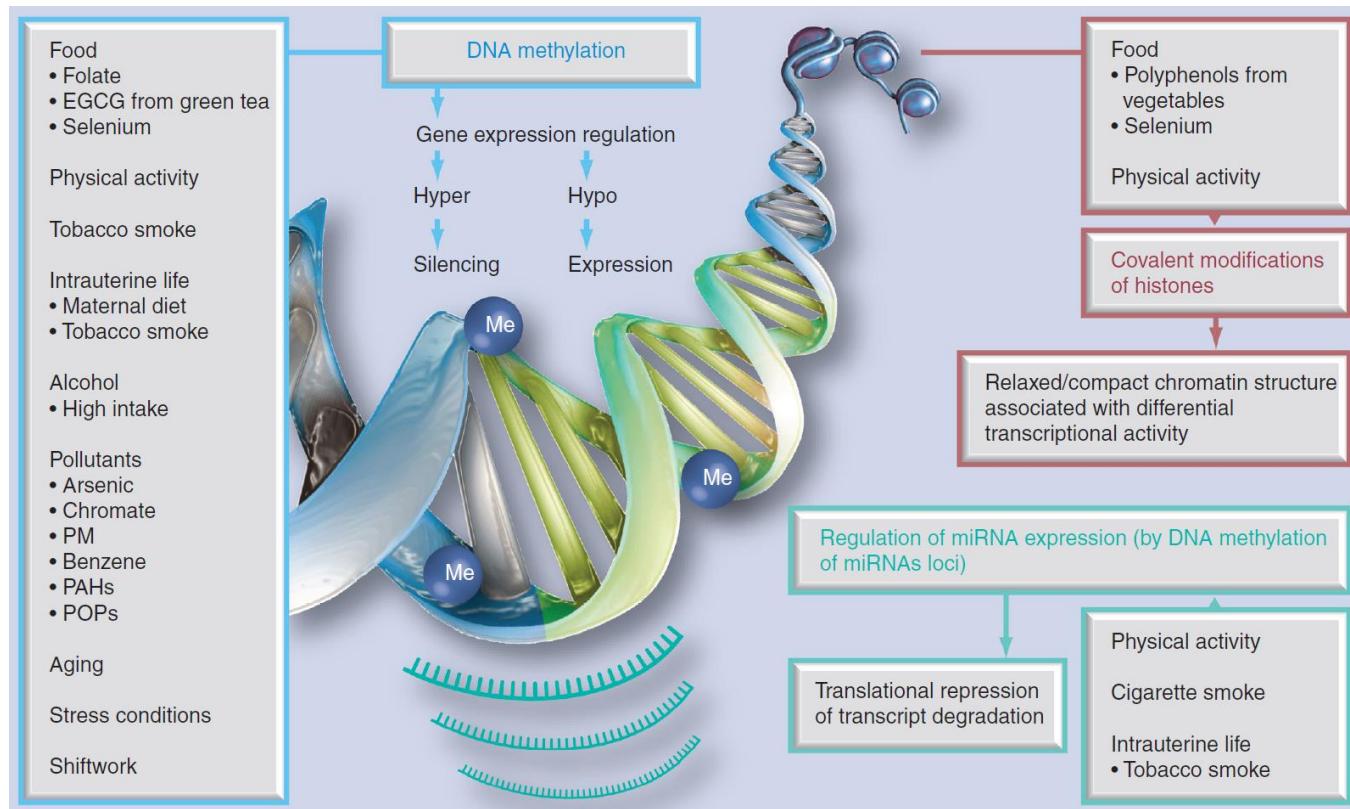
Epigenetics in Hollywood



Nucleosome with H3K27me3

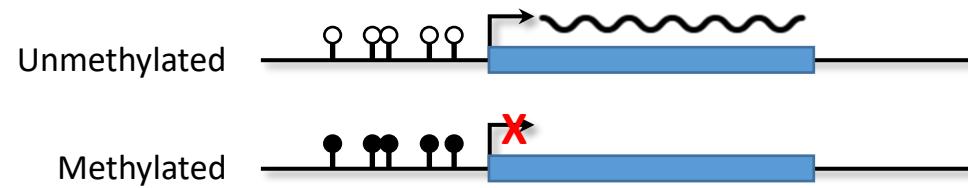
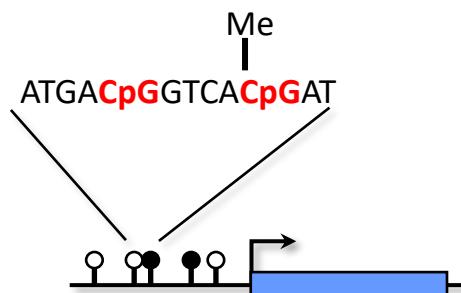
Epigenetics

- The study of heritable changes in gene function that occur *without a change in the sequence of the DNA (A, C, G, T)*.
- Some changes are transient and can be regulated by environmental factors (diet, chemicals, behaviors, etc.)



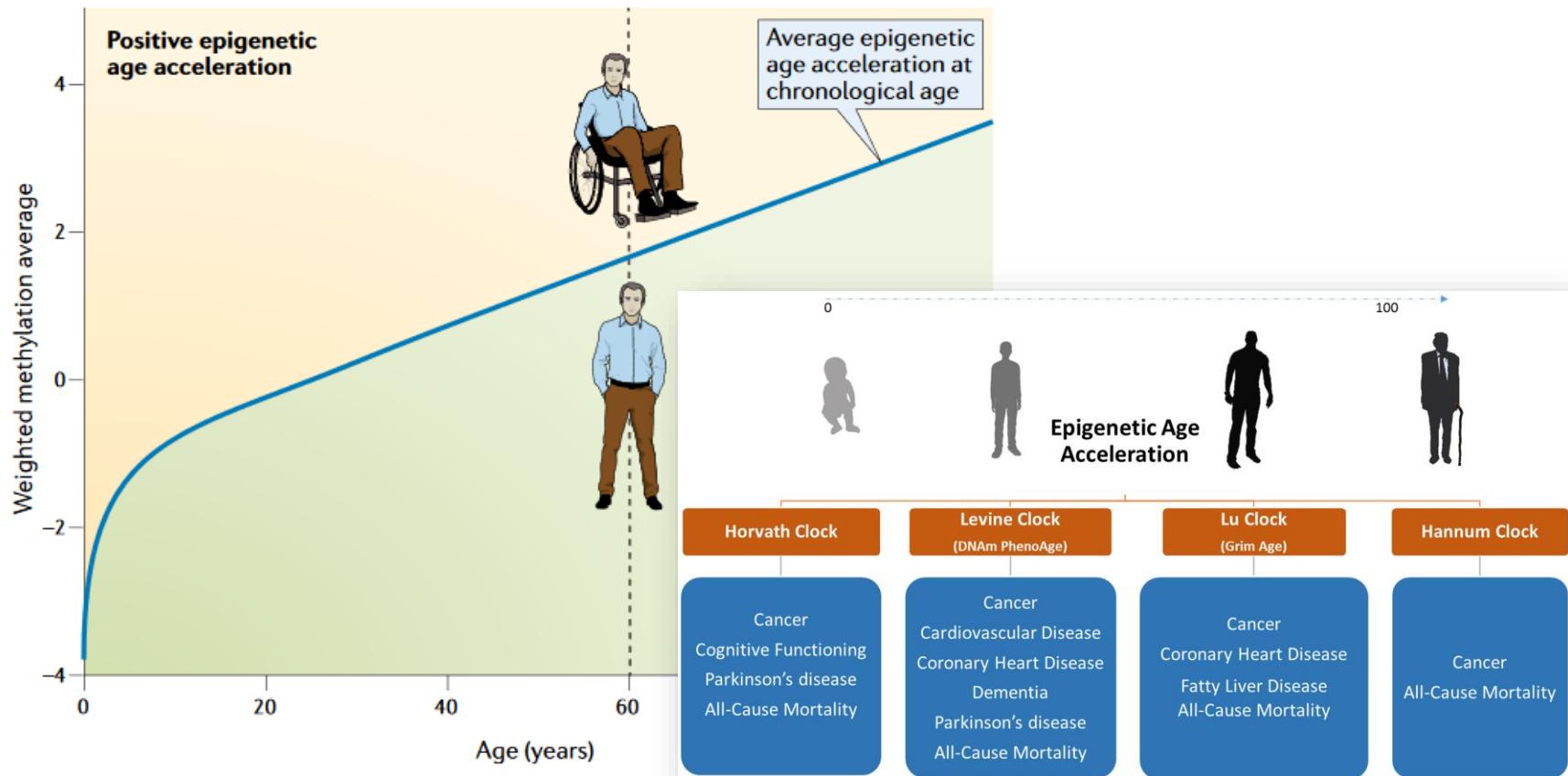
DNA methylation

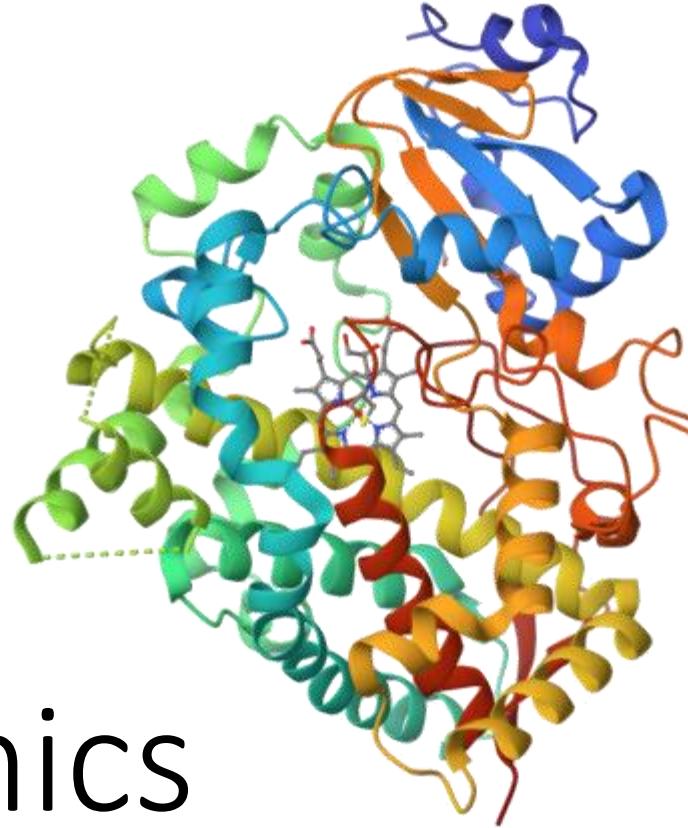
- C's of CpGs can be methylated and potentially affect gene transcription



Application - Methylation Clock

- Do some behaviors or exposures affect **healthy aging**, based on changes in DNA methylation?





Proteomics

Why Proteins?

- The majority of cellular processes are mediated by proteins (transcription factors, enzymes, receptors)
- Abundance and activity of proteins are key drivers of biological processes
- The overall correlation between RNA transcript levels and protein abundance is ~0.5
 - A LOT of genes/proteins (and the resulting cellular processes) are not regulated by transcriptional regulation, but by post-translational regulation (protein synthesis and degradation)
- Protein activity and cellular localization are additional key factors in regulating cellular processes

Untargeted Proteomics Analyses

- Identification of proteins present in complex samples (any species, any tissue)
- Quantitative proteomics (cells, tissues, mitochondria, exosomes, microbiome)
- Label-free quantification, isotopic labeling
- Analysis of posttranslational modifications

**You cannot select which proteins you analyze
Analysis is restricted to a subset of reasonably
abundant proteins (1000-3000)**

Application

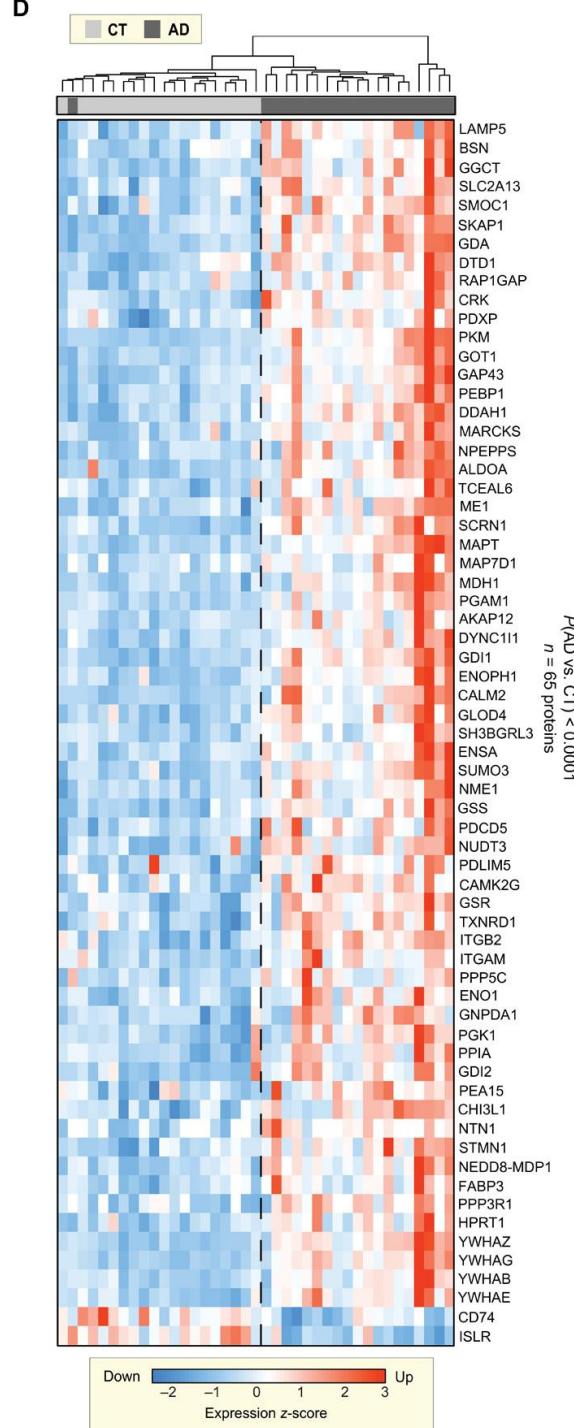
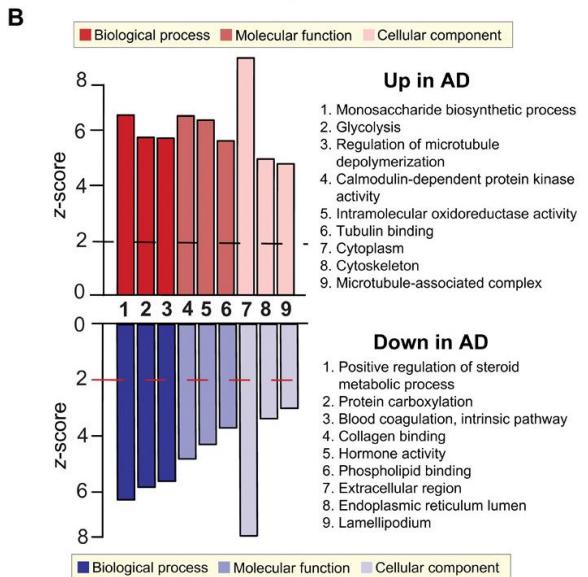
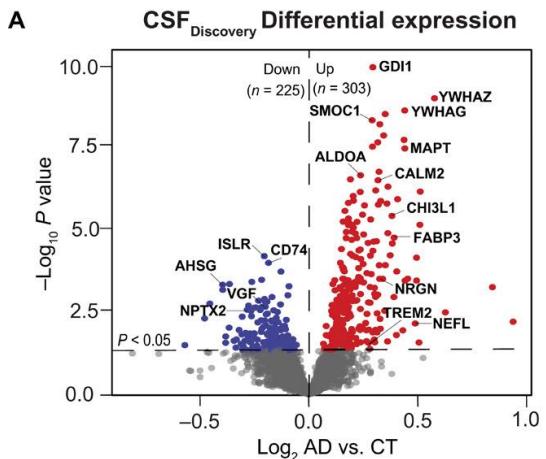
Alzheimer's Disease

SCIENCE ADVANCES | RESEARCH ARTICLE

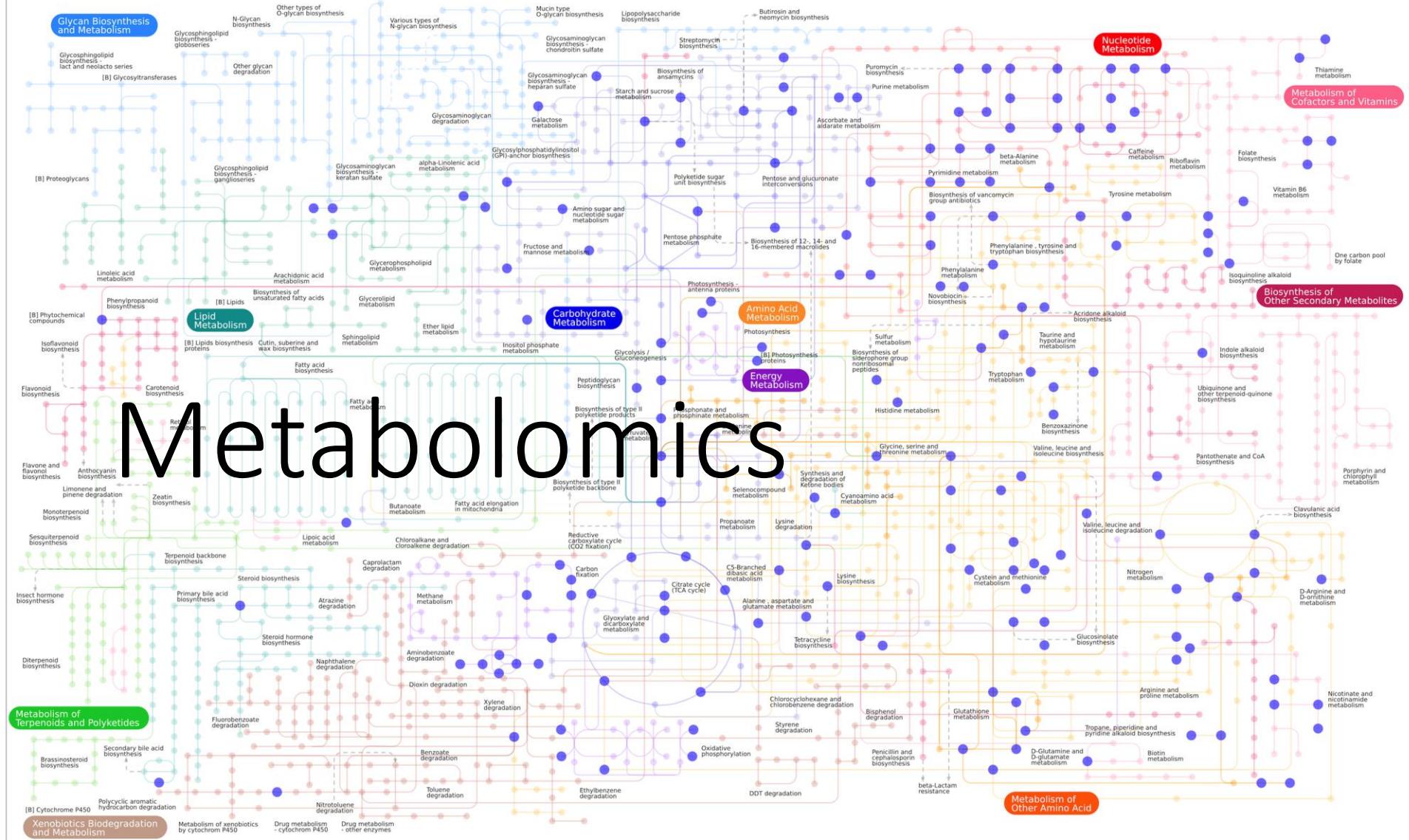
NEUROSCIENCE

Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic Alzheimer's disease

Lenora Higginbotham^{1,2,3*}, Lingyan Ping^{2,3,4*}, Eric B. Dammer^{2,3,4}, Duc M. Duong^{2,3,4}, Maotian Zhou^{2,3,4}, Marla Gearing^{1,3,5}, Cheyenne Hurst², Jonathan D. Glass^{1,2,3}, Stewart A. Factor¹, Erik C. B. Johnson^{1,2,3}, Ihab Hajjar^{1,3}, James J. Lah^{1,2,3}, Allan I. Levey^{1,2,3†}, Nicholas T. Seyfried^{1,2,3,4†}



Metabolomics



What is Metabolomics?

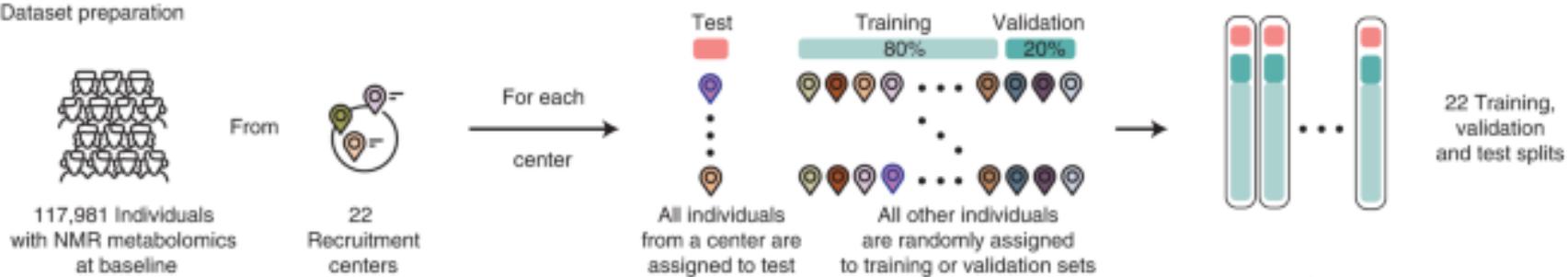
- The study of small molecules, metabolites, found in biofluids, bio-volatiles, cells, tissues or organisms
- Collectively called the metabolome
- Human Metabolome Database
 - 18,557 metabolites detected and quantified
 - 3,271 metabolites detected in but not quantified
 - 82,274 metabolites expected
- Provides a molecular phenotype or signature through a metabolic snapshot of your sample
- Differs from other Omics in that the metabolites and their concentration are a direct reflection of the biochemical activity and metabolic state

Limitations

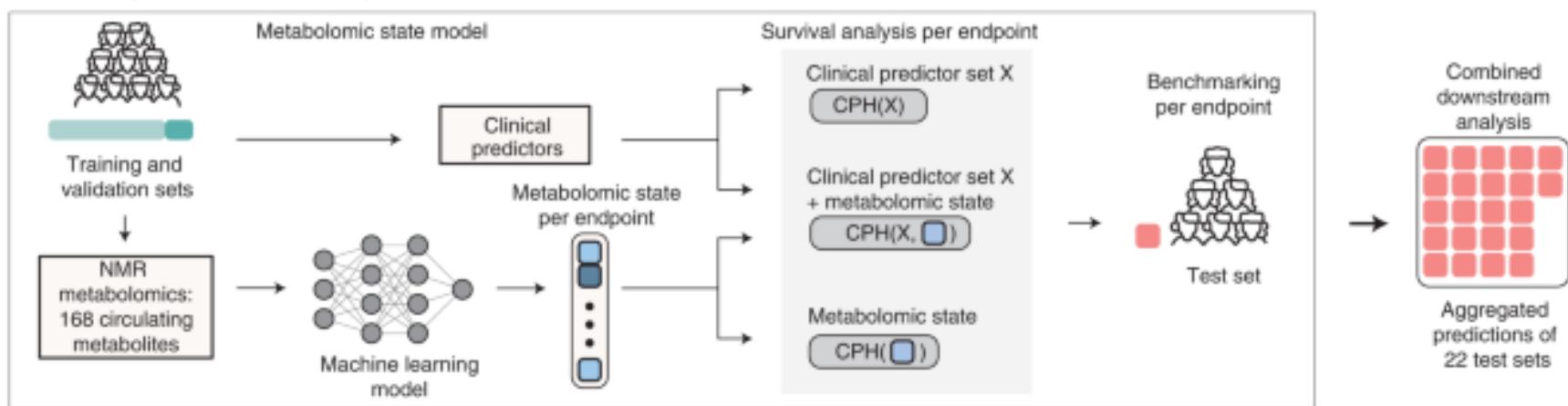
- There is no one size fits all method to detect all metabolites
- Within areas of instrumentation, there is no one method to capture all metabolites present.
- Limit of detection → missing data

Application – Disease Prediction

a Dataset preparation

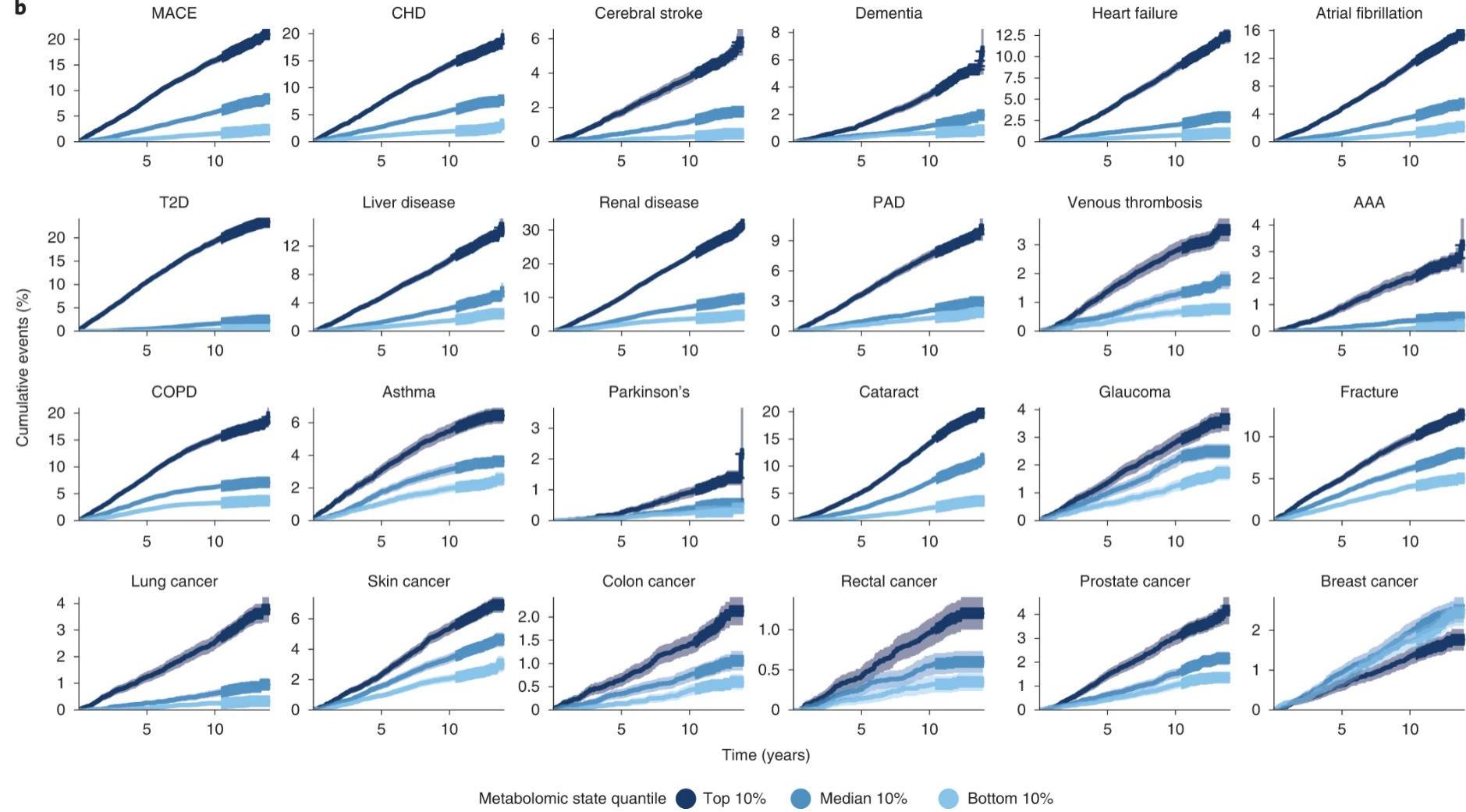


b Model development for each of the 22 partitions

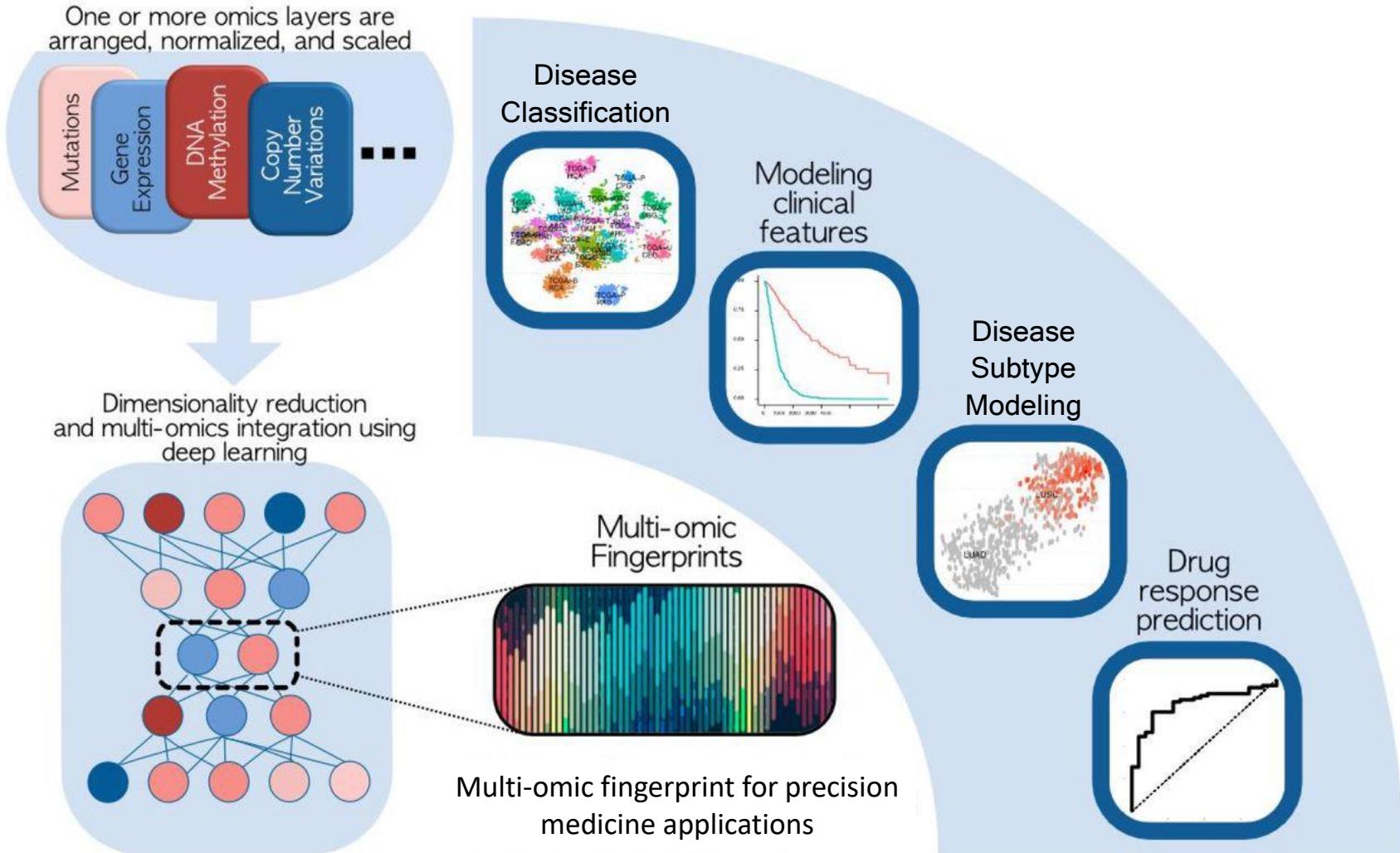


Application – Disease Prediction

b



Integrating Results



Where are we going?

- Multiomics
 - NHLBI TOPMed (>180K participants)
 - UKBiobank (500K participants)
 - All Of Us (1,482,000 participants)
 - Geisinger Health System (MyCode) - sequence the exomes of >354,000 patients
- Challenges
 - Data Standards - understanding, navigating, and using existing standards
 - Inter-interdisciplinary coordination and collaboration - medicine, biology, and computer science

Questions

