

9.5b From Clinical Genetics and Genomics to Precision Medicine

10x10 – Introduction to Biomedical Informatics & Artificial Intelligence
William Hersh, MD, FACMI, FAMIA, FIAHSI
Copyright 2026, All Rights Reserved



From Clinical Genetics and Genomics to Precision Medicine

Genetic disorders

Genetic testing

Pharmacogenomics

Gene expression

Cancer genomics

Genomewide association studies

Polygenic risk scores

Biobanks

Consumer genomics

Precision medicine

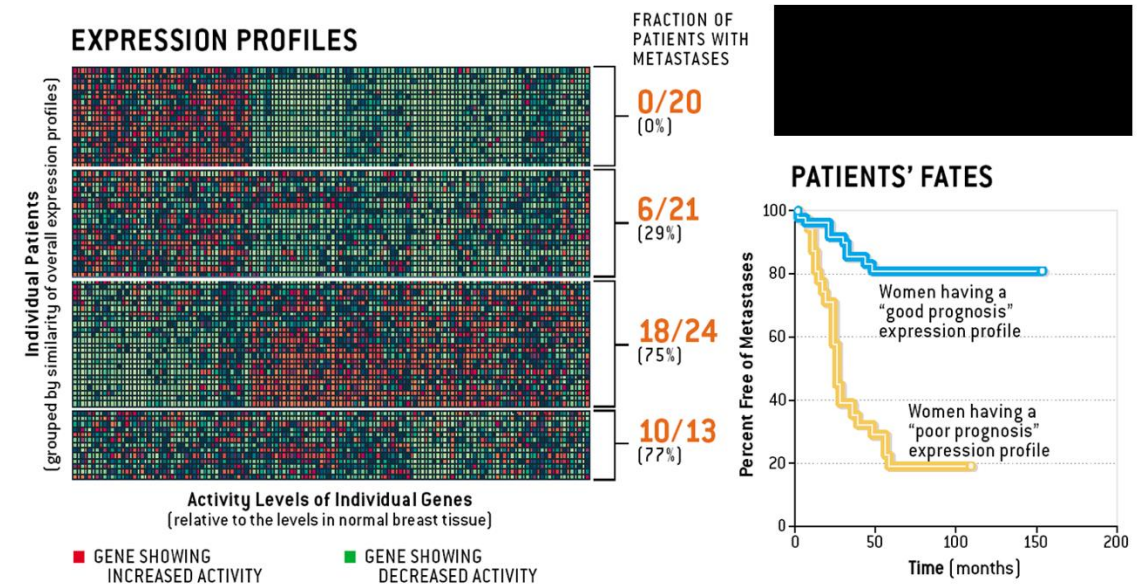
Gene expression

Measures actual expression of genes, and not just presence in DNA

- Initial technology was microarrays (Williams, 2007), which have mostly been supplanted by RNA-Seq (Stark, 2019)

Early success was in predicting prognosis in breast cancer (van de Vijver, 2002)

- Developed into commercial test, Mammoprint (Glas, 2006)
- Other tests developed as well, e.g., Octotype DX (Györfy, 2015)



Uses of gene expression

Impact on expression – in older adults, 596 genes with expression related to aging were found to have profile of aging “reversed” by resistance exercise (Melov, 2007)

Use of Mammoprint identified women with breast cancer at high clinical but low genomic risk and found could safely forego chemotherapy (Cardoso, 2016)

In patients with advanced (Stages II-III) colon cancer, non-expression of CDX2 gene associated with better response to adjuvant chemotherapy (Dalerba, 2016)

Site-specific therapy guided by a 90-gene expression assay versus empirical chemotherapy in patients with cancer of unknown primary origin (Liu, 2024)

Growing development of tools and data for use in variety of diseases (García-González, 2026)

Cancer genomics understanding enhanced by genomics advances

“Hallmarks” papers describe how cancer cells proliferate and evade death by a number of mechanisms (Hanahan, 2000; Hanahan, 2011; Hanahan, 2026)

Whole genome sequencing of tumor cells shows they undergo genomic changes, evolving the ability to proliferate and metastasize (Chin, 2011)

Cancer gene mutations are heterogeneous within tumors (Gerlinger, 2012)

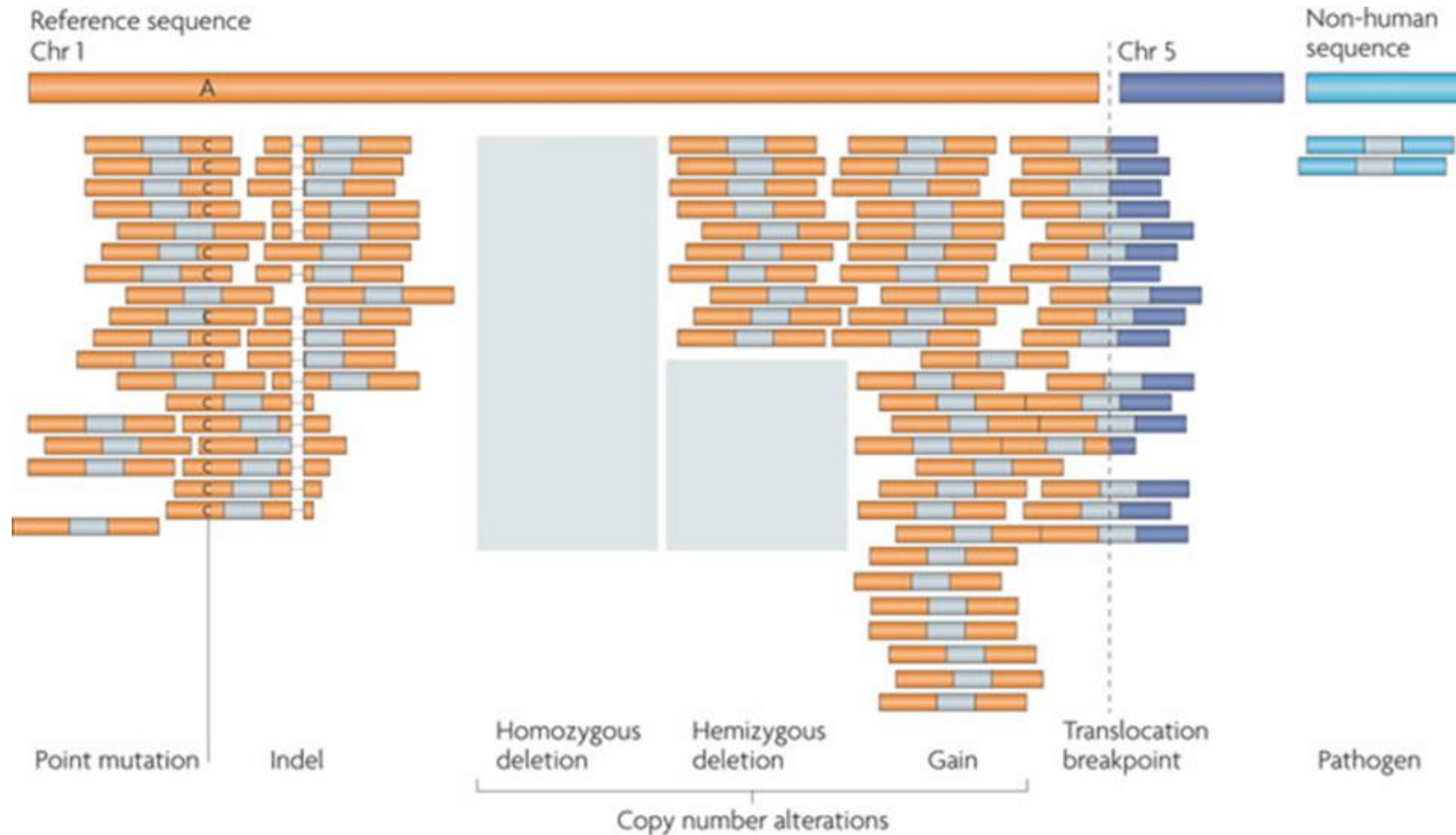
Some genomic changes representing common tumor types occur in different cancer locations (Hoadley, 2014; Lu, 2015)

DNA and protein products in blood facilitate early detection for simpler and safer treatment (Schrag, 2023)

- Detectable up to 3 years prior to clinical diagnosis (Wang, 2025)

Change names of cancers from disease of origin to describe genomic and molecular mechanisms? (André, 2024)

Types of genomics changes that occur in cancer cells (Meyerson, 2010)



Acting on genomic knowledge in cancer

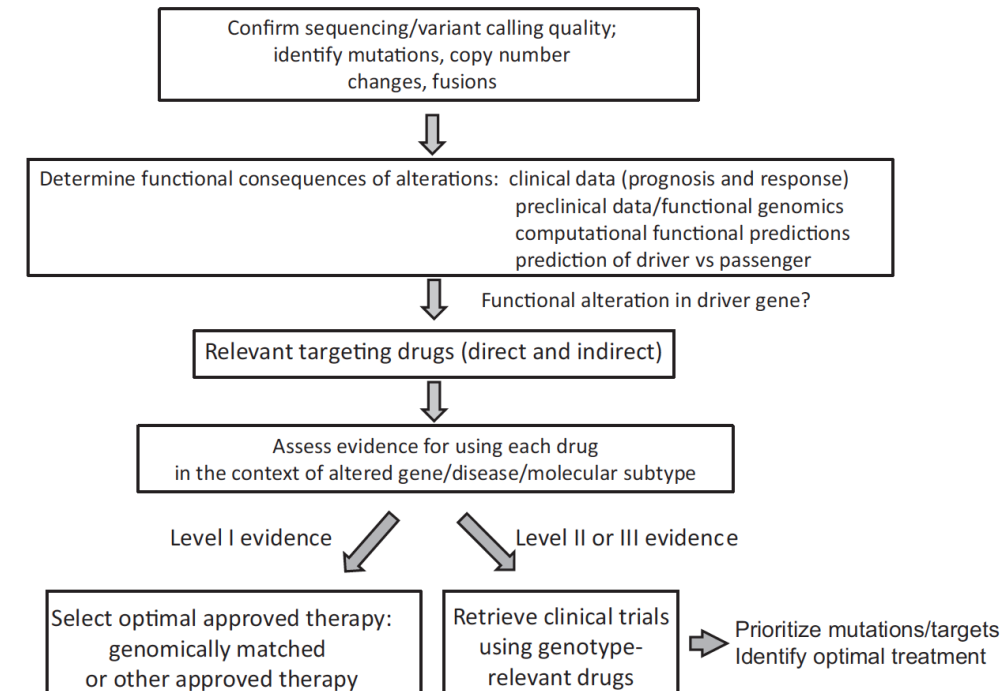
Need clinical decision support to match genomic findings to best treatment (Meric-Bernstam, 2015)

Requires identification of clinical trials involving molecular-focused drugs in resources such as ClinicalTrials.gov (Xu, 2016; Dickson, 2020)

Deep learning and AI methods

- Discover genomic drivers of cancer (Sherman, 2022)
- Predict outcomes from whole-slide pathology images and molecular features (Chen, 2022)
- Development of vaccines to augment immunotherapy (Pail, 2025)
- LLM for precision oncology recommendations (Jun, 2026)

All facilitating data-driven cancer research and care (Gomez, 2024)



Genomewide association studies (GWAS)

Next generation sequencing technologies allow high-throughput detection of genotype and its variation (e.g., SNPs, CNVs)

Growing collection of electronic clinical data allow increasing measure of phenotype, setting the stage for genome-wide association studies (GWAS) that associate genotype and phenotype (Price, 2015; Buniello, 2019)

Early and well-known effort to use EHR data for GWAS came from [Electronic Medical Records and Genomics \(eMERGE\) Project](#) (Crawford, 2014; eMERGE, 2019)

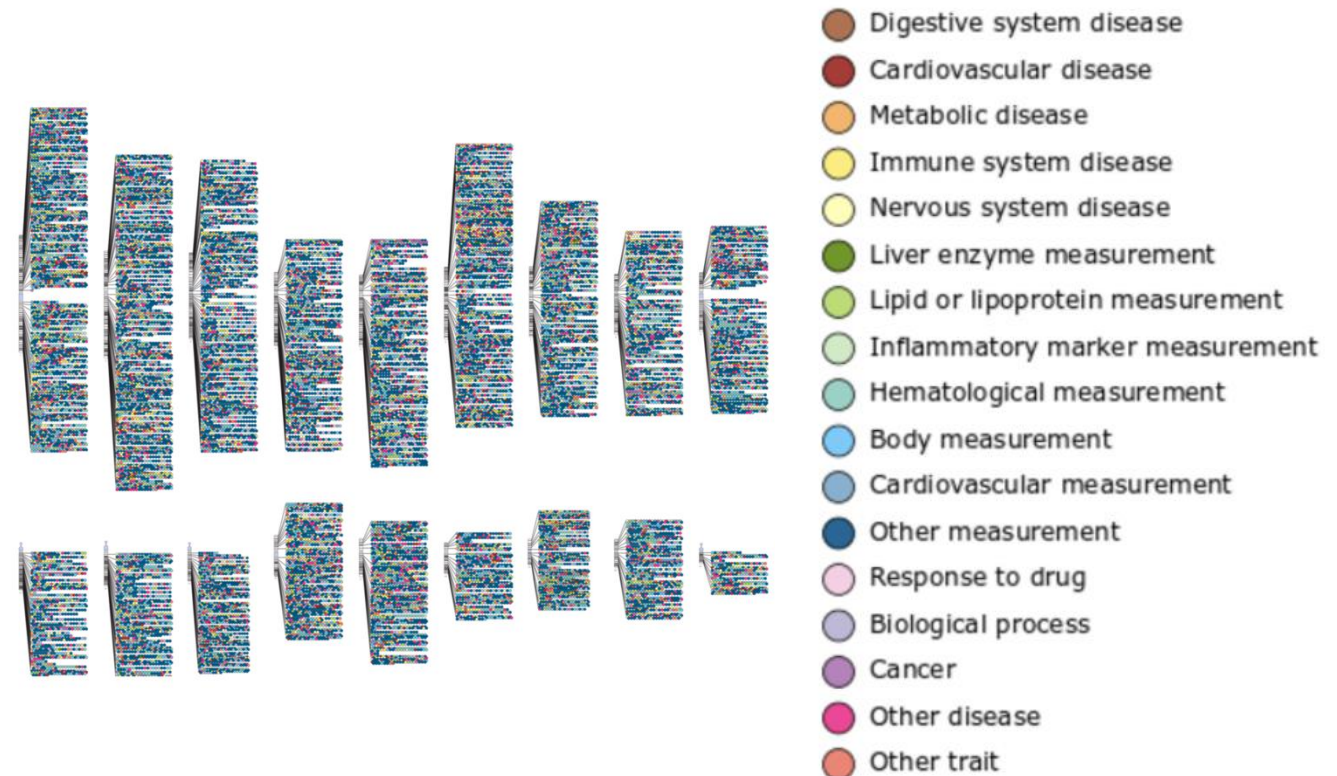
Results of GWAS

[NHGRI-EBI Catalog](#) of published GWAS

- Over 45,000 published GWAS across over 5000 human traits (Sollis, 2023)

In cancer, nearly all common malignancies studied, with over 450 genetic variants associated with increased risks identified (Sud, 2017)

Most female reproductive health diagnoses have heritable component, with varying degrees of polygenicity and discoverability (Pujol Gualdo, 2025)



Concerns about GWAS

Old adage: Correlation/association does not imply causality

Predictive ability of GWAS may not improve upon conventional clinical approaches, e.g., in coronary artery disease (de Vries, 2014), diabetes (Langenberg, 2014; Vassy, 2014)

When attempted to validate GWAS observations in real-world setting, previous associations with cardiac conduction abnormalities (Ritchie, 2013) were not found (van Driest, 2016)

Most complex traits are “omnigenic,” with many genes playing roles and interacting, so value of single associations may be of limited value (Boyle, 2017)

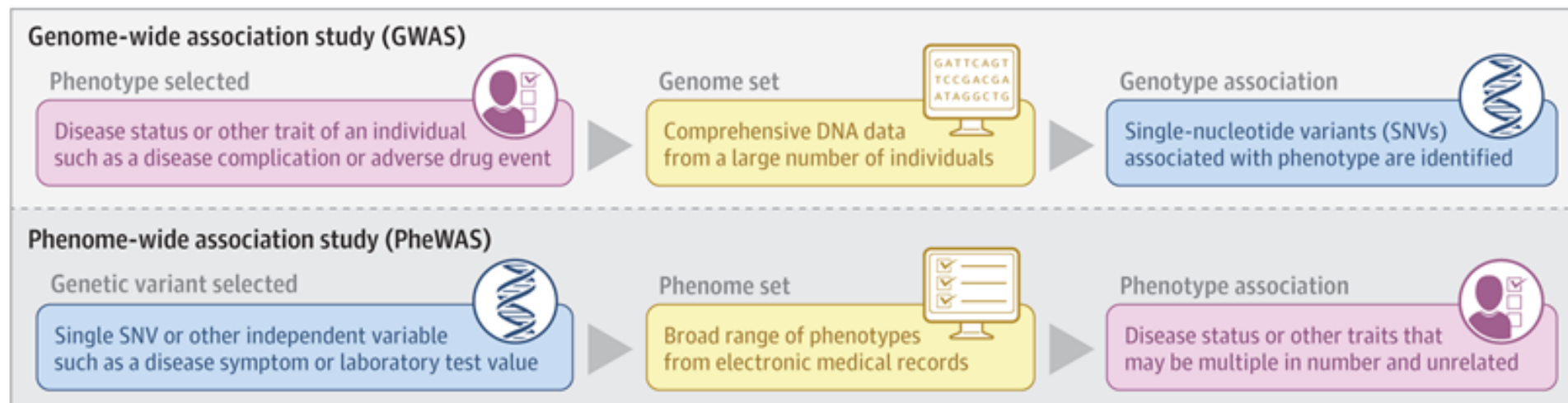
Difficulty translating findings into clinical care (Tam, 2019)

Another approach: phenomewide association studies (PheWAS)

Associate phenotypes with genes or other predictors

- Initial work focused on genes in eMERGE (Denny, 2010; Hebbring, 2014)
- Enabled by growing quantity of phenotype data in EHRs and other sources (Denny, 2016; Verma, 2018; Bastarche, 2022)

Catalog of results



Polygenic risk scores (PRS)

Aim to calculate clinical risk based on scores for presence of genes and variants (Kullo, 2025)

May be combined synergistically with other data, such

- Family history (Mars, 2022)
- Blood cell traits (Xu, 2022)
- Multi-ancestry data (Patel, 2023)

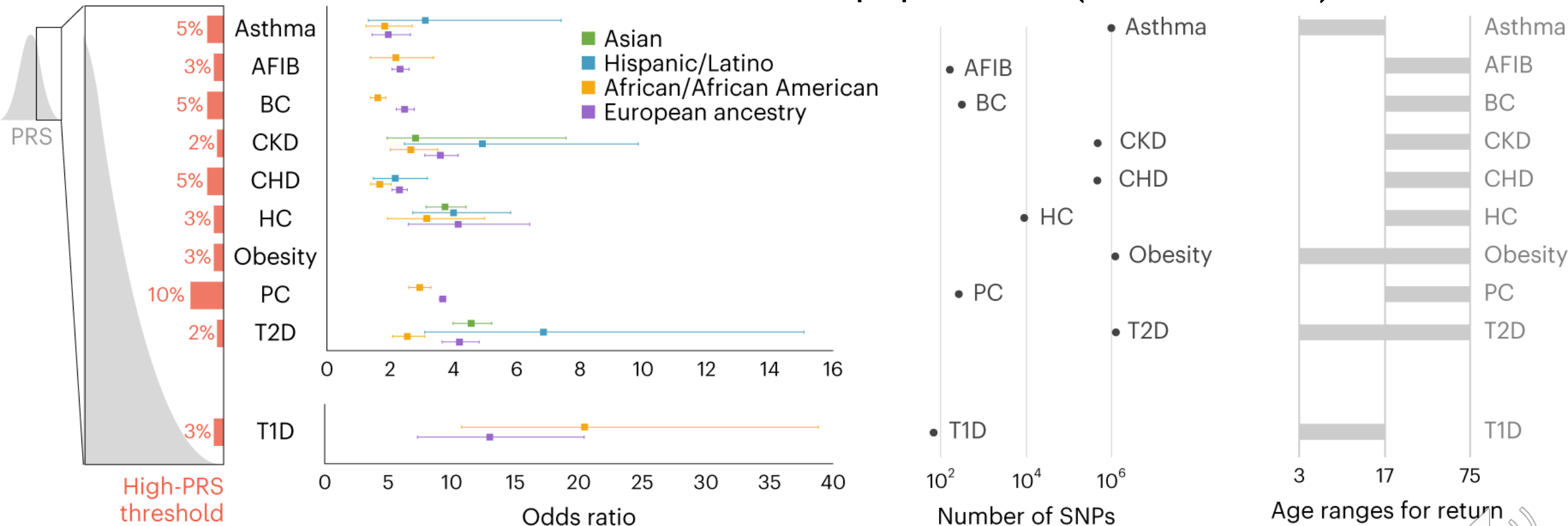
Have been developed for many clinical conditions

- Modifiable risk factors for dementia (Luo, 2024)
- High genetic predisposition to obesity increases risk of obesity-related morbidities but healthy lifestyle prevents morbidities, even in individuals with high genetic risk (Kim, 2024)
- Genomic drivers of coronary artery disease that predict risk of future outcomes after coronary angiography (Supriami, 2025)

Applications of PRS

[Polygenic Score Catalog](#) (Lambert, 2021; Lambert, 2024)

PRS for 10 chronic conditions in diverse US populations (Lennon, 2024)



Challenges for PRS

Role in clinical care, including actionability, to be determined (Hunter, 2019)

- Challenges remain in presenting results to physicians and patients (Lewis, 2022) and in development of efficient and effective workflows (Hao, 2022)

To predict from EHR, need precise and accurate data (Li, 2020)

Diversity of patients studied important, since results may not generalize across all ethnic groups (Liu, 2021)

Development and evaluation of PRS reports (Hanley, 2025)

Landscape analysis of professional group perspectives (Purvis, 2025)

Biobanks

Growing number allowing omics and clinical data analysis

One of earliest and largest, with most clinical results, is [UK Biobank](#) (Sudlow, 2016)

- Established in 2006 and recruited 500K participants aged 40-69 over several years
- Followed longitudinally, with ongoing collection of genotypic and phenotypic data, including via newer technologies
- Adopting data standards, e.g., OMOP (Papez, 2023)
- Developing new tools for phenotyping – POPDx (Yang, 2023)

Now has nearly half-million people with whole genome sequencing (UK Biobank, 2025)

- Many findings uncovered in analysis of first 150K genomes, such as genomic variants and ethnic clusters (Halldorsson, 2022)

Concerns

- “Participant bias,” i.e., patients in UK Biobank not representative of entire population, distorting results (Schoeler, 2023)
- Disagreement between EHR data and patient surveys, often due to information not typically included in EHR (Sulieman, 2022)

Biobanks (cont.)

Some clinical findings (not all genomic)

- Coffee and tea consumption associated with lower risk of stroke and dementia (Zhang, 2021)
- Rare genome variants contribute to common diseases (Wang, 2021)
- Telomere shortening associated with risk of many diseases and mortality (Schneider, 2022)
 - Associated with stroke, dementia, and late-life depression but lessened with modifiable risk factor profiles (Kimball, 2025)
- Moderate consumption of unsweetened and sugar-sweetened coffee associated with lower risk for death; less so for artificially-sweetened (Liu, 2022)
- Vitamin D deficiency associated with increased mortality (Sutherland, 2022)
- Higher step counts associated with reduced incidence of several diseases (Master, 2022)

Future directions for biobanks

- Meta-analysis across 24 global biobanks with over 2.2M people (Zhou, 2022)

Consumer genomics

Era of “consumer-driven genomics” arrived (Evans, 2010)

- Or should it be called “recreational genomics” (Dan Masys, MD)?

Most prominent company with clinical focus has been [23andMe](#)

- Saliva sample sent for analysis of [about 640,000 SNPs and risk reported for 250+ diseases](#)

Concerns include

- Knowing genetic risk does not improve health or healthy behaviors (Burke, 2016)
- Approach to patients who present with results can be challenging (Artin, 2019)
- Current offerings may be both too broad (many non-actionable results) and too narrow (not enough detail for action) (Kilbride, 2020)
- What happens to data when company files for bankruptcy (Hernandez, 2025)

Precision medicine – promises

Understanding

- Acute myeloid leukemia (AML) can be characterized into molecular subgroups (Papaemmanuil, 2016; Tyner, 2018)
- Coronary artery disease has ~60 genetic variants associated with various presentations (Khera, 2017)

Results

- 12 disparate types of cancer responded to pembrolizumab (Le, 2017; Lemery, 2017)
- Geisinger Health System results from genotyping 50K+ patients under routine care (Williams, 2018)
 - 3.5% had actionable variants for hyperlipidemia (Dewey, 2016) and could benefit from treatment (Abul-Husn, 2016)
 - Using EHR data and scaling up to other diseases (Buchanan, 2020)
- In AML, targeted therapy led to longer survival (Burd, 2020)

Precision medicine – mixed results

First trial of targeted therapy outside of approved indicators (SHIVA) showed no improved outcomes (Le Tourneau, 2015)

Use of high-throughput genomics in hard-to-treat advanced cancers (MOSCATO 01) showed benefit for 7% of patients (Massard, 2017)

RCT of WGS in routine care found 22% had identified variants but treatment was of uncertain value (Vassy, 2017)

[NCI Molecular Analysis for Therapy Choice \(NCI-MATCH\) trial](#) evaluated cancer treatment based on genomics (Lih, 2017; Flaherty, 2020) – results showed about 25% of patients had improved outcomes (O'Dwyer, 2023)

Precision medicine – concerns

Challenges with large-scale gene sequencing – results impacted by

- Inaccuracy (Goldfeder, 2016) and DNA damage (Chen, 2017)
- Gene name errors widespread in literature (Ziemann, 2016)
- “Annotation” bias (Haynes, 2018)

Small number of explicit targets in cancer and even smaller number of treatments for those targets (Prasad, 2016; Joyner, 2019)

- Requires studies not only of individual treatments but also overall paradigm (Khoury, 2017)
- Small numbers also problem for rare diseases (Hong, 2024)

Need to close gap in use and benefit (Schilsky, 2022)