

9.4 Translational Bioinformatics

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



Translational bioinformatics

Terminology and definitions

Central dogma

Precision medicine

Phenotyping

Privacy issues

Role of generative AI

Informatics challenges

Bioinformatics and related terminology

Bioinformatics

- Usually defined as informatics applied to cellular and molecular processes, including genomics

Translational bioinformatics (Altman, 2012, Tenenbaum, 2016)

- “Translational bioinformatics research integrates information about molecular entities (DNA, RNA, proteins, small molecules, and lipids) with information about clinical entities (patients, diseases, symptoms, laboratory tests, pathology reports, clinical images, and drugs) to improve patient care.” (Altman, 2012)

Genetics – study of inheritance patterns

Genome – the genetic material in an organism

Other -omes and -omics

Proteome – proteins in an organism

Transcriptome – DNA transcribed

Metabalome – molecules involved in metabolism

Phenome – traits of an organism

Interactome – biomolecular pathways and interactions of proteins

Microbiome – microorganisms inhabiting an individual

Exposome – everything to which an organism exposed

Bibliome – the literature of science

Other terminology related to bioinformatics

Computational biology – application of computational methods to discovering biological principles

Biomarker – measurable indicator of molecular, histologic, radiographic, or physiologic characteristics used to assess normal or pathologic process

- [Glossary of related terms](#)

Digital twin – virtual representation of patient and their data (Coorey, 2022)

Bioinformatics – more information

Books

- Bioinformatics – Baxavenis, 2020; Ahman, 2025; Agostino, 2026
- Medical/clinical genomics – Wünschiers, 2022; Lesk, 2025; Shashikant, 2026
- Public interest – Ridley, 1999; Collins, 2010; Topol, 2012; Mukherjee, 2016

Web sites

- [National Human Genome Research Institute \(NHGRI\) Bioinformatics](#)
- [International Society for Computational Biology \(ISCB\)](#)
- [European Bioinformatics Institute \(EBI\)](#)
- [Bioinformatics.org](#)

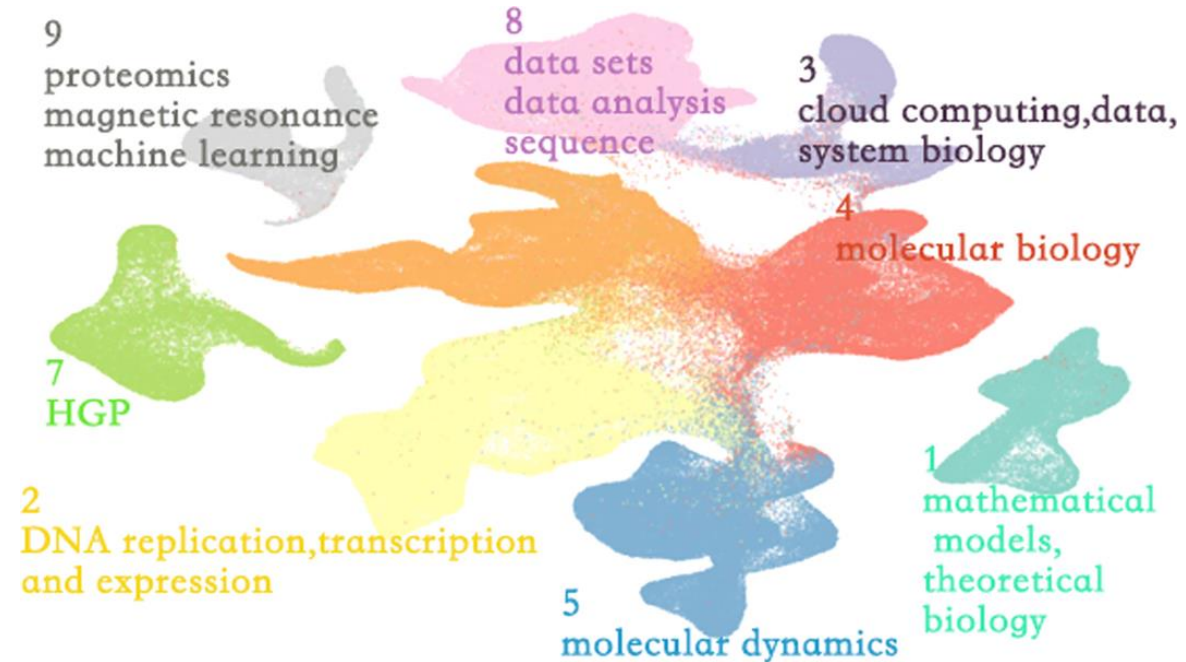
Bioinformatics – more information (cont.)

Visualization of areas of bioinformatics
(Wang, 2022)

“Owner’s guide” to human genome
(Pritchard, 2023)

Grand challenges in bioinformatics
education and training (Işık, 2023)

[Curriculum guidelines](#)

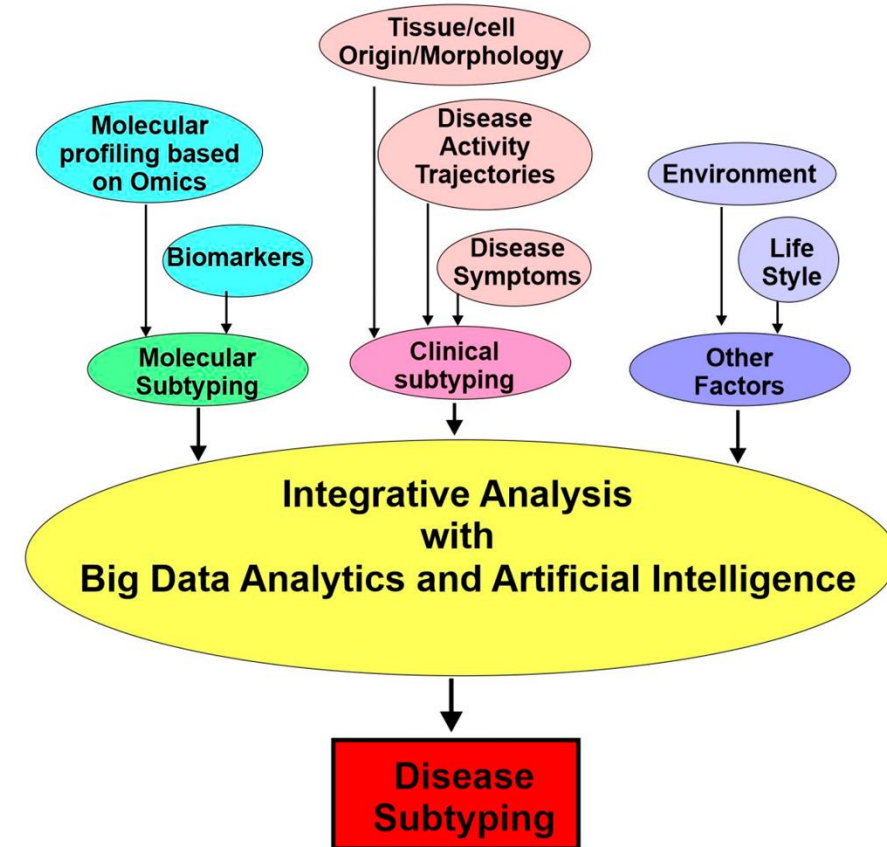


Precision medicine

Diagnosis and treatment targeted to individual patients on the basis of genetic, biomarker, phenotypic, or psychosocial characteristics that distinguish them from other patients with similar clinical presentations (Wang, 2023)

Precision Medicine Initiative (PMI) of NIH (Collins, 2015; Denny, 2019; Bakken, 2024)

- Renamed to [AllofUs Research Program](#)
- Major initial goal is [building cohort of 1M Americans with genotyping and clinical data](#)
- Initial data set of patients and associated tools [available](#) (Ramirez, 2022)



Promise (?) of precision medicine

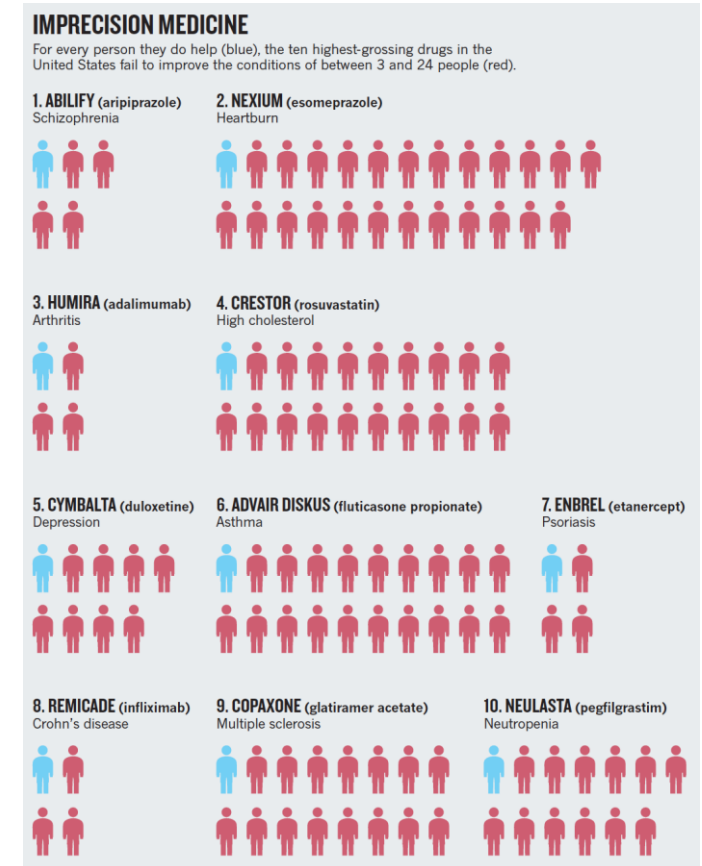
Current medicine very imprecise, e.g., high NNTs (Schork, 2015)

Increasing amount of biology is being discovered through “high-throughput” technologies

Growing amount of patient and consumer data in electronic and personal health records (the “phenome”)

Potential to reduce unnecessary overtreatment, especially in cancer (Katz, 2018)

Example of precision medicine paradigm is pembrolizumab, which treats an abnormality in 12 different types of cancer that inhibits DNA repair (Le, 2017)

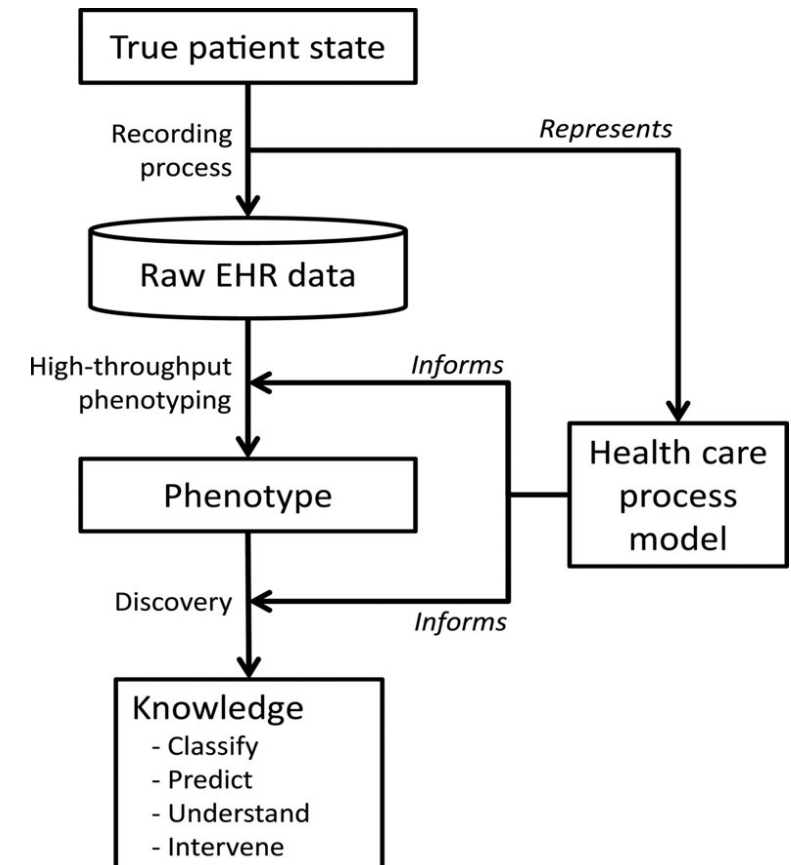


Phenotyping

Phenotype is observable traits of organisms; human phenotype ascertained from many sources of data

- [EHR](#) (Hripcsak, 2013; Robinson, 2018)
- Sensors from phones, wearables, etc. that measure activities, behaviors, etc. (Insel, 2017)
- Future also includes biobanks, continuous monitoring, and observational studies (Glicksberg, 2019)

Computable phenotyping computes phenotype from these data sources for various use cases, e.g., genotype-phenotype associations or patient cohort identification (He, 2023)



Phenotyping – ongoing efforts

[Human Phenotype Ontology](#) (HPO; Köhler, 2021)

- Mapping to standard terminologies, e.g., LOINC (Vasilevsky, 2017)
- Phenopacket schema for exchange of computable longitudinal case-level phenotypic information (Jacobson, 2022)

Pheocodes derived from ICD-9/10-CM support EHR phenotyping (Bastarche, 2021)

- Used with HPO to link rare and common disease vocabularies (McArthur, 2023)

Additional efforts to automate and validate, e.g.,

- [eMERGE](#) (eMERGE Consortium, 2021)
- Phenotyping – PheNorm (Yu, 2018) and EHR Phenolyzer (Son, 2018)
- Using natural language processing (NLP) (Savova, 2017; Pacheco, 2023)
- Computational drug discovery (Sadybekov, 2023)
- Leveraging wearables, especially in mental health disorders (Oudin, 2023; Choi, 2024; Jung, 2025)
- Advancing exposomics (Beans, 2025)

An area of concern is genetic privacy

Genetic discrimination is continued threat

- Is genome the ultimate “personal identifier?”

Recognized early on (Gostin, 1995)

New technologies and availability of data make individual recognition even easier

- Concerns about genetic discrimination, inferences from phenotypes, and re-identification (Wang, 2017)
- Risk not only with genotypes but also sequence data and metadata (Shi, 2017)
- Recent analysis found ability to re-identify ~80% of individuals from genotype and phenotype data in ancestry database (Lippert, 2017)

Efforts to protect privacy

[Genetic Information Nondiscrimination Act of 2008 \(GINA\)](#) (Green, 2015)

ONC data security principles for PMI (2016)

Statistical methods to increase privacy (Simmons, 2016; Beaulieu-Jones, 2017; Jagadeesh, 2017)

Bring software to the data – researchers do not see data and only obtain results (Guinney, 2018)

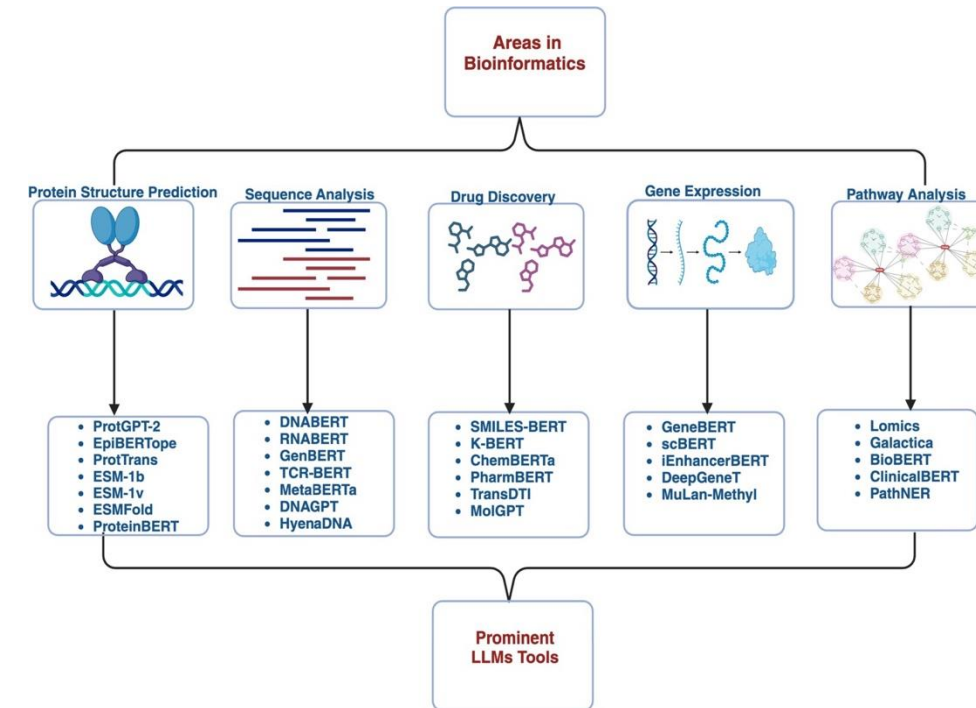
Apply sociotechnical approaches, i.e., not only technologies but guiding people to do right things (Wan, 2022)

Role of generative AI

Large language models (LLMs) showing promise in many pattern identification and predictive tasks, mostly encoder transformers (Sarumi, 2024)

Specific applications include

- Genomic language models (gLMs) to provide insights (Benegas, 2024)
- GeneWhisperer to assist annotation of genomes (Li, 2025)
- Foundation model to predict functional impacts of genetic variation (Brix, 2026)



Informatics challenges

Biology has become an “information science” (Insel, 2003)

Need informatics solutions for many challenges (Kolker, 2014; Tenenbaum, 2016)

- Consent
- Specimen, protocol tracking
- Standards and interoperability
- Data provenance
- Optimal use of EHRs
- Precision medicine knowledge capture and application
- Patient/consumer engagement

Also challenges for metadata sharing from omics studies from “perceptual and technical barriers” (Huang, 2025)

- Lack of uniform standards
- Privacy concerns
- Study design limitations
- Insufficient incentives
- Inadequate infrastructure
- Shortage of trained personnel

Informatics challenges (cont.)

Standards leader is [Global Alliance for Genomics and Health \(GA4GH\)](#)

- Address rising need for tools and standards to facilitate international data sharing across institutions (Siu, 2016; Rhem, 2021)

Patient/citizen/consumer engagement critical (Petersen, 2019)

- Including engagement via social media, e.g., Genes for Good (Brieger, 2019)

Challenges for precision medicine clinical decision support tools (Johnson, 2024)

- Including need for standards, user-centered design, alternative forms of interventions, and evaluation of patient outcomes

