Optum

Clinicogenomics: How new, linked data is advancing life sciences research

The study of clinicogenomics across large, diverse populations can help researchers better understand disease pathways, discover novel therapeutic targets, identify variation in disease onset and progression, and evaluate drug safety and efficacy.

Since the human genome was first sequenced, life sciences researchers have sought to use this information to improve health. They soon discovered, though, that DNA sequence data alone has limited value when trying to understand the effects of genetic variation on health.[1] Today, researchers are increasingly linking genomic data to rich, longitudinal clinical information – a practice that yields **clinicogenomic** data. The field of clinicogenomics is unlocking novel insights to better understand diseases and to guide drug discovery, development and commercialization.

Read on to learn more about:

- The foundations of genomics-driven research
- The sources of clinicogenomics data: opportunities and limitations
- The potential to fuel discovery and development of safer, more effective drugs
- The ways that clinicogenomics data can enhance clinical trials

Building on the foundations of genomics-driven research

Genomic data have fueled scientific discovery for decades, providing valuable insights about **genetic variants.** All humans have genetic variants, which can be inherited or occur as a result of environmental factors. Many are harmless and determine traits like hair or eye color.

Genetic variants are also the root cause of many diseases. Studying them leads to discoveries about genomic biomarkers, epigenetic changes and treatment responses. Such discoveries are critical to pharmaceutical and biotech companies' ongoing efforts to develop new drugs and precision treatments.

Genomic biomarkers are measurable DNA and/or RNA variants that indicate normal biologic processes, pathogenic processes, and/or responses to therapeutic or other interventions.[2] They are useful for researchers to understand because they may inform disease risk, progression, outcomes or treatment response.

The study of genetic variants has also led to discoveries in **epigenetics.** This field of study explores how behaviors and environmental factors can impact gene function, including which genes are turned on and off. Epigenetic changes can be inherited, are reversible and do not impact a person's DNA sequence. But they can lead to abnormal gene activity or inactivity, and they may cause or influence certain diseases.

Lastly, genomic research can help predict **treatment response.** This knowledge can enhance efforts to repurpose drugs for increased efficacy, develop drugs with fewer or less severe side effects, and better understand drug-drug and drug-gene interactions. Such discoveries can help improve medication adherence and reduce medication misuse.

High-quality clinicogenomic datasets create new possibilities for drug development, approval and appropriate use, including:

- **More comprehensive understanding of disease.** Researchers can better discern factors associated with disease risk, onset and progression.
- **Increased efficiency in early-stage drug discovery and development.** Teams can strategically zero in on potential targets and identify more high-quality drug candidates.
- **Access to diverse patient populations.** Researchers can optimize drug development and dosing guidelines for diverse populations by using characterized cohorts that support multi-racial, multi-ethnic recruitment efforts.
- **Faster, more cost-effective clinical trials.** Linked datasets can yield higher-quality study candidates through more efficient recruitment and enrollment efforts.
- **Higher FDA approval rates.** A recent study indicated that drugs with genetically supported targets were more likely to achieve FDA approval.[3] A better understanding of a drug or biologic's associated trait and drug interaction can also inform indication selection — another lever that can influence approvals.
- **More opportunities to develop companion diagnostics.** Researchers can identify patients most likely to benefit from a specific therapy based on their phenotypic and genetic profile.
- **Drugs that are safer and more effective, and produce fewer side effects.** Clinicians can tailor prescriptions based on the patient's genetic make-up.
- **Reduced trial-and-error prescribing.** Prescribing the right drug faster increases patient satisfaction, helps improve outcomes and minimizes payer costs.

Sources of clinicogenomics data: Opportunities and limitations

Genomic research is opening new frontiers in disease diagnosis, treatment and prevention. To realize the full potential of genomics-driven drug discovery, development or commercialization, researchers can combine genomic data with detailed, longitudinal phenotypic data. This becomes an "enriched" or "linked" clinicogenomics dataset. However, linked datasets vary in terms of quality and utility.

Current sources of clinicogenomics data fall into four distinct categories. While each option offers benefits, none completely meets the needs of today's researchers.

1. **Public health and research entities** offer clinicogenomics data for research purposes. The data they collect and make available may be influenced by the demographics of their local geographies and, thus, may not reflect the diversity of the general population. Life sciences companies may also find it difficult to derive meaningful competitive advantage from these datasets because they are meant to support the public good rather than answer questions specific to development pipelines.
2. **Private enterprises** have begun to offer some clinicogenomics data and related services, generally focused on specific clinical areas like oncology or prenatal screening. Even within those areas of specialization, sourcing and linking both clinical and genomics data can prove difficult. Complicating matters further, industry merger and acquisition activity can threaten any entity's long-term viability as a data source.[4]
3. **Partnerships with hospitals or health systems** can yield genomics data that is coupled with rich, longitudinal patient health records. But the scale and composition of cohorts may be limited by the size and demographics of the local population. This can become especially problematic when it hinders exploration of potential treatments for rare disease or when researchers seek greater diversity in their studies.
4. **Consumer-focused DNA testing companies** have begun to offer datasets that can inform drug discovery and development efforts.[5] However, these data often reflect genetic panel results instead of higher-quality whole genome sequencing (WGS) or whole exome sequencing (WES). The phenotypic data tied to these analyses typically depend on patient-reported information that may reflect inconsistencies, bias or inaccuracies. The population contributing to the dataset may also not be fully representative from an age, racial, ethnic or other demographic perspective, hindering the data's utility in large-scale research.

What is high-quality data?

High-quality genomic data includes whole exome sequence (WES) and whole genome sequence (WGS) data from large, diverse patient populations. High-quality genomic data is often derived from clinical-grade sequencing conducted with a high degree of coverage, meaning that a nucleotide at a specific locus is analyzed many times. The more reads, the higher the confidence in the output.

High-quality clinical data reflects detailed, longitudinal information across the continuum of care. This information includes discrete data, like vital signs or lab results, as well as more free-form data captured in electronic clinical notes, like a history of disease progression or qualitative observations.

Together, these types of data can help researchers pinpoint specific drug targets and potentially develop treatments around them.

Potential to fuel discovery and development of safer, more effective drugs

Linked datasets can be applied to many types of research, including functional genomic studies, **genome-wide association studies (GWAS), phenome-wide association studies (PheWAS)** studies, and systems biology and network analysis.

Statistical models can derive information from these integrated data sources and provide a labeled dataset. **Machine learning** or **artificial intelligence** algorithms can then ingest the labeled data to model the interactive and dynamic networks that regulate gene expression, cell differentiation and cell cycle progression. When understood in the context of linked clinical data, discoveries about disease-associated variants provide a critical foothold for generating testable hypotheses.

High-quality clinicogenomic data has the potential to be the new driving force in drug discovery and development across nearly every area of biomedical research. Research in this field creates the foundation for many applications of genomics, such as **pharmacogenomics** and other forms of precision medicine. It also can inform consumer segmentation strategies, enable the pursuit of new indications and augment safety monitoring.

For life sciences companies, these advancements could mean fewer resources deployed toward unsuccessful research and development, faster time to market and enhanced guidelines for appropriate use. Clinicogenomic data can also aid in the development of better-tailored companion diagnostics (and thus their associated revenue streams). Patients also stand to benefit with access to drugs that are safer, more effective and with fewer side effects when prescribed based on their genetic makeup.

Clinicogenomics and technology experts are critical partners in this endeavor. Their expertise can help life sciences companies leverage this emerging field of research by providing secure data storage architecture, analytics platforms, patient privacy safeguards and, most importantly, access to large-scale, real-world patient cohorts at scale.

Oncology researchers: Early adopters of clinicogenomics

Oncology researchers have relied on clinicogenomics in the diagnosis and treatment of cancer for several years. **For example, in 2017, the FDA approved pembrolizumab (Keytruda®) for adult and pediatric patient use based on the presence of a biomarker rather than based on the cancer site.**[6] This decision stemmed from a growing body of evidence that tumors sharing critical molecular aberrations may share profiles of clinical response to therapies, regardless of solid tumor type.

The Keytruda example also points to an additional application of clinicogenomics-driven research – **the development of companion diagnostics.** In 2020, a life sciences company received FDA approval for a test to identify patients who could benefit from Keytruda when standard of care therapies fail.[7]

Spotlight: Leveraging clinicogenomics data to enhance clinical trials

In the last few years, life sciences and technology companies have started to deploy real-world data — namely, clinical data derived from existing electronic health records (EHRs) and/or insurance claims — in efforts to re-imagine clinical trial designs and processes. These data can be particularly useful in identifying patients for clinical studies. After identifying a cohort, researchers may approach individuals about study participation. For patients whose electronic medical records (EMR) do not contain comprehensive, high-quality genomic sequencing information, researchers may ask those patients to submit biospecimens that are analyzed using WES and WGS methods.

Researchers can now link vast databases containing rich longitudinal clinical data to high-quality genomic sequencing at the individual patient level. Consistent, high-quality, deep coverage genomic data allows for direct, reliable comparison of genomic data points.

Linking clinical data derived from EHRs and insurance claims provides researchers with a richer view into patients' health care journeys. Clinical data can provide phenotypic information related to:

- Patient demographic details, including age, ethnicity and sex
- Disease risk factors, including socioeconomic, environmental and lifestyle factors
- Disease progression
- Care across the health care continuum
- Laboratory tests ordered and diagnostic test results
- Response to therapy and outcomes data
- Therapeutic interventions
- Health care utilization

Because EHR databases are large and include diverse populations, life sciences companies can target highly characterized cohorts and support their pipeline-specific recruitment needs. In this way, researchers may be able to better reach populations whose members are underrepresented in clinical trials. These may include elderly patients, patients with less access to transportation, those without access to medical research centers and those who have difficulty managing daily activities.

Warfarin dosing guidelines and the importance of diversity in research

Warfarin dosing considerations demonstrate the importance of using a broad, large, diverse genomic data set. Warfarin is a well-known and accepted treatment for the prevention of certain cardiac and pulmonary conditions. Every year, an estimated 2 million Americans with certain cardiac conditions start taking warfarin.[8] Dosing is critical for warfarin administration: too high, and the patient may experience excessive bleeding; too low, and potentially life-threatening clots may occur.

African American and Latino populations have greater variability in warfarin dose requirements and are at greater risk of experiencing warfarin-related adverse events than individuals with European ancestry.[9] The evidence suggests that individuals of non-European ancestry may benefit from improved warfarin dose estimation. [10]

However, most of the literature on genotype-guided warfarin dosing, including data from prospective randomized trials, is based on those of European ancestry.[11]

Connecting data to tackle complex problems

The questions that the life sciences research community must confront will only grow in complexity as regulatory agencies become more sophisticated, payers demand more clearly articulated value propositions, and patients and providers look to advance the field of precision medicine.

Armed with clinicogenomic data, researchers can begin to unpack the genetic and environmental interdependencies that influence an individual's health and open up new avenues for treatment and prevention of disease. GWAS and PheWAS studies can yield insights and connections that in turn can inspire both drug development strategies and clinical programs.

Now is the time for our industry to look at how these rich, longitudinal clinicogenomic datasets can address the challenges we'll face in the coming years and to partner in building these incredibly important sources of information.

GWAS research and its impact on drug development

Genome-wide associated studies (GWAS) contribute to our understanding of many diseases and have led to novel target discoveries. Let's examine the research on age-related macular degeneration (AMD), a leading cause of severe vision loss in elderly populations.

Using GWAS methods, researchers discovered that a mutation in the complement factor H (CFH) gene is associated with AMD progression.[12] This unexpected discovery cleared the way for new therapeutic strategies targeting the complement factor pathway. Many treatments are currently under development, including a recombinant CFH protein carrying a protective version of the gene.[13]

In related work, the National Eye Institute conducted the Age-Related Eye Disease Study (AREDS) and the follow-up AREDS2 to study cataract and AMD.[14]

Researchers were able to risk-stratify individuals based on phenotypic factors (like age or smoking status) and genomic insights, and then test whether taking nutritional supplements could prevent or slow these diseases. Clinical data also allowed the researchers to examine disease progression and treatment response. The formulations tested in the trials are now sold as the AREDS and the AREDS2 formulas.

Researchers confirmed that the AREDS formula's effectiveness depends on the individual patient's risk status related to the genes CFH and ARMS2.[15] They also found that one genotype group (GTG2) is likely harmed by treatment with the AREDS2 formulation, while another group (GTG3) substantially benefits.[16]

Ashley Brenton, PhD

Vice President, Real-World Evidence and Genomics

Bachelor of Arts, The Johns Hopkins University
Doctorate, University of California, Davis

Dr. Ashley Brenton is the vice president of Real-World Evidence and Genomics for Optum Life Sciences. Ashley is responsible for leading the data strategy, development, business development and operations for real-world data with a focus on genomics. In this role, she bridges the gap between clinical genomics and product development to leverage the wealth of Optum data into commercial datasets and strategic alliances.

Prior to joining Optum, Dr. Brenton was the chief science officer at Mycroft Bioanalytics, where she led corporate valuation and sales strategy of the intellectual property portfolio, including the world's largest clinicogenomic biobank in chronic pain. Previously, Dr. Brenton developed and commercialized a number of precision medicine tests and algorithms while building clinicogenomic datasets. In addition to her product development in the space, she has established herself as a thought leader, publishing several manuscripts that provide support for the integration of clinicogenomics in health care.

Before entering the industry, Dr. Brenton was a molecular biologist in academia. She earned a bachelor's degree in public health studies from Johns Hopkins, a doctorate from the University of California, Davis and a viral pathogenesis fellowship from the Scripps Research Institute.

Rich, longitudinal clinicogenomics data can fuel all stages of drug development.

Learn more at
optum.com/clinicogenomics

Key terms

Artificial intelligence (AI)

The collection of technologies that perform "smart" tasks often associated with the human mind, such as learning and reasoning.

Clinicogenomics

The field of study that links genomic data and clinical data to understand disease processes and identify and develop therapeutic targets.

Epigenetics

The field of study that explores how behaviors and environmental factors can impact gene function, including which genes are turned on and off. Epigenetic changes can be inherited, are reversible and do not impact a person's DNA sequence.

Genetic panel

A test (or assay) that analyzes small slices of DNA to examine specific mutations.

Genetic variants

The mutated DNA nucleotides in a cell.

Genome-wide association studies (GWAS)

The field of study that explores associations between genetic variations and the presence or absence of disease.

Genomic biomarkers

Measurable DNA and/or RNA variants that indicate normal biologic processes, pathogenic processes and/or responses to therapeutic or other interventions.

Genomics

The study of all genes in an organism and their interrelationships.

Machine learning

A form of artificial intelligence that has self-learning or self-improving capabilities based on its analysis of historical and real-time data.

Pharmacogenomics

The study of how an individual's DNA affects the way they respond to drugs.

Phenome-wide association study (PheWAS)

The study of phenotypic details compared to a single genetic variant.

Phenotypic data

Information about a person's phenotype, that is, the observable characteristics and clinical traits of diseases and organisms (a subset of clinical information).

Treatment response

How a medication or therapy affects a patient's prognosis.

Sources:

1. National Institutes of Health. U.S. National Library of Medicine. Systems genetics for drug target discovery. Last updated October 1, 2012. Accessed March 14, 2022. ncbi.nlm.nih.gov/pmc/articles/PMC3185183/
2. European Medicines Agency. Note for Guidance on Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories. Last updated November 2007. Accessed March 14, 2022. ema.europa.eu/en/documents/scientific-guideline/ich-e-15-establish-definitions-genomic-biomarkers-pharmacogenomics-pharmacogenetics-genomic-data_en.pdf
3. PLOS Genetics. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. Last reviewed December 12, 2019. Accessed March 14, 2022. journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008489
4. Roche. Roche and Foundation Medicine reach definitive merger agreement to accelerate broad availability of comprehensive genomic profiling in oncology. Last updated June 19, 2018. Accessed March 14, 2022. roche.com/media/releases/med-cor-2018-06-19.htm
5. Bloomberg Businessweek. All those 23andMe spit tests were part of a bigger plan. bloomberg.com/news/features/2021-11-04/23andme-to-use-dna-tests-to-make-cancer-drugs. November 4. 2021. Accessed March 14, 2022.
6. Health Affairs. Real-world evidence in support of precision medicine: clinico-genomic cancer data as a case study. Last updated May 2018. Accessed March 14, 2022. healthaffairs.org/doi/pdf/10.1377/hlthaff.2017.1579
7. Foundation Medicine. FoundationOne® CDx receives FDA approval as the first companion diagnostic to identify advanced cancer patients with solid tumors that are tumor mutational burden-high (TMB-H) and appropriate for immunotherapy treatment with KEYTRUDA® (pembrolizumab). Last updated June 17, 2020. Accessed March 14, 2022. foundationmedicine.com/press-releases/2bac198e-d31b-4e95-bfbf-28a55093b8e8
8. National Institutes of Health. U.S. National Library of Medicine. Warfarin pharmacogenomics in diverse populations. Last updated September 6, 2017. Accessed March 14, 2022. https://pubmed.ncbi.nlm.nih.gov/28672100/
9. Ibid
10. Ibid
11. Ibid
12. Source 1 (Systems genetics for drug target discovery)
13. Ibid
14. National Institutes of Health. National Eye Institute. AREDS/AREDS2 clinical trials. Last updated November astudies-aredsareds2/about-areds-and-areds2
15. Healio. Study: Genetics largely determines treatment response to AREDS. Last updated January 30, 2018. Accessed March 14, 2022. healio.com/news/optometry/20180130/study-genetics-largely-determines-treatment-response-to-areds
16. Ibid

Optum

11000 Optum Circle, Eden Prairie, MN 55344 | optum.com