Rapporteurs: Lisa Chadwick, Heather Colley, Sara Currin, Chris Wellington
Division of Genome Sciences, NHGRI

Identifying Research Priorities to Accelerate Genetic Diagnosis

Tuesday, April 16th, 2024 – Wednesday, April 17th, 2024
Bethesda, MD

Executive Summary

The Identifying Research Priorities to Accelerate Genetic Diagnosis workshop was hosted by the National Human Genome Research Institute (NHGRI) April 16th - 17th, 2024 in North Bethesda, MD.

The NHGRI gathered insights from the scientific community regarding the major gaps, challenges, and opportunities for advancing the understanding of the genetic basis of Mendelian diseases and improving the proportion of cases with a precise molecular diagnosis for rare diseases (the “solve rate”). Next generation sequencing approaches such as whole-exome (WES) and whole-genome sequencing (WGS) are the current state of the art for identifying causal variants underlying Mendelian diseases. Despite the transformative impact of next-generation sequencing over the past decades, these methods are unable to identify the causal variant for a significant proportion of individuals with suspected genetic disease.

The two-day meeting was primarily offered in an in-person format for invited participants, but was also livestreamed via Zoom for public viewing. Twenty-six in-person attendees participated in the workshop by offering their expertise in a series of panel discussions focused on distinct aspects of genetic diagnosis. The panel topics included:

- Tapping into emerging technologies
- Data sharing is caring
- Genetics, it’s complicated
- Effectively linking variants to function
- Computational tools to enable genetic diagnoses

These sessions were complemented by two additional sessions. The first was focused on additional topics of interest to participants that were not included in the formal agenda. The second was focused on developing recommendations about resources of research projects needed in this area. These were designed to solicit additional insights from meeting participants about key topics that may not have been included in the original agenda, and to obtain specific recommendations about next steps that NHGRI could take to drive innovation and significantly advance the genetic diagnosis of rare disease.

The full agenda, including speakers and panel discussants can be found on the workshop webpage: Identifying Research Priorities to Accelerate Genetic Diagnosis.

Session Highlights

Session One: Tapping into Emerging Technologies

The first session of the workshop explored the range of currently available molecular technologies for studies of rare disease. This included discussion of potential applications, challenges that prevent existing technologies from being used to their full capacity, emerging molecular technologies and needs for further technology development. Participants noted that technologies are at different stages of development. Some are ripe for standardization, while others need a stronger focus on innovation. Overall, there was enthusiasm for identifying novel ways to approach to genetic diagnosis that diverge from the current state-of-the-art.

Meeting participants identified several research and technology development needs that could facilitate the use of emerging technologies in clinical practice, including:

- Systematic evaluation of the use of new and emerging molecular technologies in obtaining genetic diagnoses, and concrete recommendations for how to most effectively implement them
- Technological improvements that allow -omic technologies beyond genomics need to be better enabled at scale
- Methods that allow integration of multi-omic data from a single sample collection
- Comprehensive reference data that establishes the range of normal values across populations
- Democratized and accessible technologies that can be directly implemented in diverse settings

Participants identified several key questions to consider within the space of emerging technologies:

- Is the “genomics first” approach stale?
- Does a specific causal genetic variant always need to be identified, or could a causal molecular signature be sufficient for diagnostic purposes?

Session Two: Data Sharing is Caring

The second session of the workshop focused on data sharing to advance genetic diagnosis of Mendelian conditions. Panelists discussed the current state of data sharing, effective data sharing practices, pertinent data categories and infrastructure needs for data sharing. They also provided examples of challenges that currently stifle data sharing, as well as examples of successful data sharing. There was a general emphasis on developing ways to leverage these successes to shape future endeavors.

Meeting participants stressed that individuals with rare disease should be empowered in the sharing of their own data. Several related topics arose from this broad topic, including:

- Resources are needed to further enable this activity
- While a majority of individuals with rare diseases want their data to be shared, some do not. There is a need for methods that accommodate these differences in

data sharing preferences. This may include more informed and proactive consents with options that align with individuals' preferences regarding data sharing
- Legacy samples were often collected with consents that preclude broad data sharing, and there is often no way to reconsent these patients for data sharing. Given the substantial volume of data that is currently being generated in clinical laboratories, panelists questioned whether there was still a need to try to resolve issues related to sharing data from legacy samples

The following topics were also emphasized in this session:
- Moving the field from two-sided to one-sided to zero-sided data sharing would have a substantial impact. Data sharing should not only focus on genomics and -omics data. It is also critical that the associated phenotype information is shared with appropriate standards and meta-data
- A key roadblock preventing effective data sharing is the amount of hands on time that it requires. In order to facilitate the sharing of genomic and phenotype data from existing clinical practice, there is a need for tools that seamlessly integrate into existing Electronic Health Records (EHRs) workflows
- Efforts should be directed towards dismantling data silos, particularly through enhanced data federation
- Clear guidelines are needed regarding the types of data that can and should be shared and that are most useful when shared. ClinVar was highlighted as a resource that could facilitate the implementation of data sharing on variants.
- Standardization protocols and metadata frameworks are essential for cohesive data sharing practices

Participants identified a question to consider within the space of data sharing:
- What is needed to incentivize and actualize widespread data sharing?

Session Three: Genetics, it's Complicated
The third session of the workshop focused on scrutinizing the assumptions that underlie current approaches to identifying causal variants, and identifying where these assumptions may need to be revisited. Panelists discussed how these assumptions may lead to blind spots in analytical models, the value of incorporating diverse data types in analyses, and the importance of considering contextual factors and ancestral diversity in analytical frameworks.

Key points raised in this session included:
- Phenotypic expression is influenced by a combination of genetic factors, environmental influences, and stochastic processes. We need to have analytical methods that take all of these factors into account, rather than focusing on only one aspect
- Longitudinal follow-up of individuals can be informative, particularly in cases where there is not a conclusive genetic diagnosis
- High-throughput genome-wide approaches must be complemented by deep studies of specific genes and phenotypes

- Keep clinically relevant use cases (such as how data might be used in ClinGen panels) in mind even in more basic or mechanistic genomics research, in order to maximize impact
- Variation in previously excluded regions of the genome (like centromeres, telomeres, sex chromosomes, repeat regions, etc.) needs to be better understood and incorporated in analytical frameworks
- Acknowledgement of the significance of contextual factors in genetic analyses

Session Four: Effectively Linking Variants to Function

The fourth session of this workshop focused on identifying best practices and new approaches for using functional information to identify and prioritize variants and conducting functional validation of candidate variants. The participants discussed the current landscape of the field and the limitations for assessing the functional impact of different types of potentially relevant variants, and the challenges associated with modeling the function of variants within specific biological and cellular contexts. Animal models and their role in functional validation were also considered.

Key points raised by meeting participants included:

- While there have been significant advancements in the development and use of technologies to assess the functional impact of single nucleotide variants, there is a need for continued development of technologies to investigate the functional impacts of noncoding SNPs, structural variants, and other variant types
- It is particularly challenging to predict or assess the functional impact of non-coding variants; we need more informative technologies, readouts and methods in this area
- Current approaches often focus on one variant at a time, and it remains challenging to investigate the functional impact of multiple SNPs in combination
- A need for more precise delineation of the roles of functional assays across various model systems, including cells, animals, and organoids
- Although a considerable amount of relevant functional data has been generated, this data is not well organized. Much of the data exists in silos that can be difficult to search or integrate in useful ways. Furthermore, negative results, which are useful for assessing candidate variants, are often not published or shared
- The field would benefit from enhanced communication channels between geneticists, biologists, and other researchers involved in functional modeling

Participants identified a question to consider within the space of functional validation:

- Can Artificial Intelliegence (AI) be leveraged to integrate all available functional data (molecular, cellular, human and model organism), predict the effect of a given variant, and design models for functional validation?

Session Five: Computational Tools to Enable Genetic Diagnosis

The fifth session was dedicated to exploring current applications and future advancements in analytical methods for genetic diagnosis. Participants noted the increasing use of artificial intelligence (AI) and machine learning (ML) and the need to both assess their potential benefits and to address hurdles impeding their integration

into the field. Productive conversations were held on the use of continuous variant scores, as opposed to categorical classification of variants. Participants agreed that there is a need for a more refined definition of variants.

Key suggestions in this section fell in two primary categories:

- Improving the development of variant scores, this encompasses:
    - Ongoing innovation in methodologies for generating variant scores
    - Evaluation of the comparative merits of continuous versus categorical variant scoring systems, and their application in Bayesian frameworks
    - Understanding who uses variant scores and how they are used, and ensuring these needs are considered in the development of tools
    - Ensuring alignment of variant scoring methodologies and variant score reporting with the requirements of the broader community. Making sure that users understand the assumptions and limitations of scores, and how to integrate them into downstream uses and analyses
- Enabling AI/ML in rare disease diagnosis, including:
    - Improved sharing, standardization and harmonization of data and metadata in machine readable formats
    - Ensuring the independence of benchmarking and test data from training datasets
    - The creation of a “model zoo” of pre-trained, transparent and interpretable models that people can share, interrogate and use for prediction

Meeting Outcomes

After the discussion sessions, participants were asked to make recommendations of activities that NHGRI, or other entities, could pursue that would address some of the challenges and gaps identified throughout the meeting. They were asked to consider short term or small scale activities as well as long term or large scale activities. Recommended activities fell into a number of broad categories, some of which are highlighted here:

Advance the use of multi-omic and other emerging technologies in molecular diagnostics

- Conduct a structured and systematic evaluation of the utility of multi-omic or other emerging molecular technologies in rare disease discovery and diagnosis
- Generate comprehensive multidimensional datasets that can be used for validation, model development
- Create core technology sites to run new technologies or functional assays, develop standards for emerging technologies, and facilitate the uptake of promising technologies in clinical labs
- Develop multidimensional reference datasets in clinically meaningful cell types
- Support the development of computational tools that facilitate visualization and integrative analysis of these data in the context of rare disease discovery and diagnosis

- Invest in the development of scalable technologies for generating multi-omic data from clinical biospecimens

Develop new analytical approaches and computational tools for prediction, discovery and diagnosis

- Develop user friendly tools that leverage artificial intelligence to bring together disparate datasets that enable the entire genome to be fully utilized in clinical genetics
- Invest in the development of computational tools that enable long-read and other whole genome data to be analyzed to the fullest extent possible
- Build a next-generation discovery platform that brings together and integrates the universe of published and publicly available data into a knowledge graph, and provides user friendly entry points to query and assist with variant prioritization and interpretation.
- Develop an algorithm to identify the right genomic test for any individual based on their phenotype or electronic health records
- Conduct pilot experiments to rigorously evaluate the potential and effectiveness of model-driven experimental design for functional discovery

Prioritize data sharing and collaboration

- Fund infrastructure to facilitate innovative data sharing through platforms like Matchmaker Exchange, including variant-level matching and aggregation of full genomic datasets for joint analyses
- Engage with the international rare disease research community for enhanced data sharing and collaboration
- Host grand challenges for VUS interpretation across all disease causing genes, that engage clinical laboratories, technology developers, and patient communities
- Bridge the divide between clinical labs, affected individuals, and research labs by creating a streamlined pathway to share appropriately consented genomic data and clinical metadata into a larger research and data sharing ecosystem
- Provide a common infrastructure that enables patient organizations to build and contribute to registries, share genomic and phenotypic data, and engage as central partners in research

Comprehensive investigation and cataloging of variant/function relationships

- Create a unified database that allows a user to easily discover all known functional and phenotypic information about a gene or variant of interest (“a one stop shop”)
- Invest in curation and annotation of functional data of all kinds
- Provide a centralized repository for functional data, including negative data
- Measure the effect of every possible variant in all clinically meaningful genes in at least one functional assay
- Perturb every nucleotide in every promoter and every conserved enhancer

- Incorporate early developmental cell types into large scale variant to function studies

Scientific Planning Committee:

- Ali Crawford, Ph.D., Illumina Inc.
- Nara Lygia De Macena Sobreira, M.D., Ph.D., Johns Hopkins University School of Medicine
- Aaron Quinlan, Ph.D. The University of Utah

NHGRI Program Staff:

- Lisa Chadwick
- Heather Colley
- Sara Currin
- Chris Wellington