Rapporteurs: Alexander Arguello, Jake Baroch, Sara Currin, Ismail Safi
Division of Genome Sciences, NHGRI

Advances in the Genetic Architecture of Complex Human Traits

Thursday, November 16th, 2023 – Friday, November 17th, 2023
Bethesda, MD

Executive Summary

The National Human Genome Research Institute (NHGRI), in consultation with other NIH institutes and the National Science Foundation (NSF), organized and hosted the workshop "Advances in the Genetic Architecture of Complex Human Traits" in Bethesda, MD in November 2023. The goal of the workshop was to evaluate current conceptual and analytic approaches for disentangling how genetic and non-genetic factors contribute to human diseases and traits. With the guidance of a Scientific Organizing committee composed of leaders in the field, NHGRI program staff invited researchers with expertise in evolutionary, population, quantitative, and statistical genetics as well as those in the behavioral and social sciences to participate. Through a series of presentations, panels, and roundtables, attendees provided informative and fruitful discussions underscoring areas of scientific consensus and those needing further development. Overall, the meeting was well-received by participants who appreciated the diversity of disciplinary perspectives and the compelling research project from NHGRI's 2020 strategic vision to "elucidate the genetic architecture of the majority of human diseases and traits."

Meeting objectives

- Adopt a historically grounded and forward-looking approach to genetic architecture.
- Assess and build on the current state of the field of complex trait genetics.
- Identify gaps and emerging opportunities for research.
- Explore future research directions in the genetics of complex traits across biological levels.
- Facilitate a theoretically informed, holistically inspired, and integrated systems approach for understanding complex trait architecture.

Scientific overview

A major goal of genetics and genomics is to decompose the sources of trait variation into constituent causal factors. The hope is to better understand both the evolutionary forces that have shaped the number, frequency, and interactions of alleles influencing traits, and to identify potential levers of intervention to better predict or influence those traits. Although the foundation of quantitative genetics was laid over a century ago, there continue to be disagreements, both philosophical and empirical, about how to interpret the interplay of genetic and non-genetic factors on trait variation. These issues have recently risen to the fore as large-scale genome-wide association studies (GWAS) reveal differences in apparent genetic effects across families and pupations with increasing interest in identifying the sources of that heterogeneity. It is now clear that association studies may be confounded by interdependencies across scales of biological, social, and ecological organization that limit their biological interpretations, especially for the most complex traits. The growing availability, however, of large-scale human genetic, genomic, and phenotypic data affords opportunities to revisit our existing approaches, develop new ones, and leverage complexity to improve our understanding of phenotypic variation in people.

The workshop was broadly structured around the theme of biological levels of organization, and this theme cut through many of the issues that arose throughout the meeting. Most fundamentally, this was seen through conceptions and definitions of genetic architecture itself. As one definition, genetic architecture encompasses the number, frequencies, effects, and interactions of trait-associated alleles. This statistical and population-level view considers genetic effects in the aggregate across individuals. As an alternative definition, genetic architecture may refer to the intervening molecular and cellular effects that link genetic variation to a trait (i.e., the components of a genotype-phenotype map). This mechanistic and biological view considers isolated genetic effects within individuals. These views mirror debates from the early twentieth century between the Mendelians and biometricians who emphasized the contribution of many small effect versus few large effect

alleles, respectively, that was reflected by participants who emphasized the biological tractability of rare over common alleles.

The importance of considering the level of biological organization also emerged for one of the most prominent themes of the meeting – context. At lower levels, context refers to the cell types, cellular states, or ancestral genetic background in which a genetic effect may be expressed. At higher levels, context may refer to the developmental period or physiological state an individual is in or the ancestral environment a population evolved. These alternate views of context revealed important unresolved questions in the field. Recent work in functional genomics indicates ubiquitous context-specificity for gene regulation and expression, and this specificity better resolves the effects of disease-associated alleles. A large literature in quantitative genetics, however, suggests context-specific effects (e.g., GxE) are rare for human traits. This discrepancy may be reconciled by considering that these contexts are found either within or across individuals and are thus subject to different evolutionary pressures. For example, extrapolating genetic interactions seen in model or agricultural organisms (both GxE and GxG) to humans may fail when not accounting for the unique life histories experienced and ecological niches inhabited by different species. It remains to be seen whether accounting for greater types of contexts at molecular and cellular levels or at the individual and population levels will help reconcile the "missing regulation" that underlies many unresolved disease and trait-associated variants.

In addition to acknowledging the mediating effects of context in interpreting genetic effects, the differences among types of traits became important to consider. There was a general understanding that partitioning traits according to whether they reflect biological, behavioral, or social features may not be the most conducive path forward. Rather, different types of traits may be subject to unique forms and degrees of confounding in genetic association studies. Developing study designs and methods that account for different biases may better resolve differences in genetic architecture as well as pleiotropy across traits. Towards that end, diverging views emerged on whether existing theoretical and analytic approaches developed over the past century of quantitative genetics are sufficient to account for the full variety of traits expressed by people across different contexts. Conspicuously absent were in-depth discussions about polygenic scores, which trace their origins to genomic prediction and selection methods in agriculture, and whether quantitative genetics can inform theoretical limits to their utility. The limits of these approaches can be quickly reached when contexts and traits depend on how societies organize themselves across time and place. This is where the experience and expertise of those working in the social sciences may be helpful for designing instruments that better capture different social contexts or environmental variables for genetic association studies.

Future directions

While there were active, thoughtful, and illuminating discussions, there were no research areas that emerged as needing an influx of further investment and support. Rather, there were empirical questions that remain to be answered mainly relating to how far existing methods and approaches will get us toward a more complete understanding of genetic architecture across a variety of human traits. As such, NHGRI is encouraging novel methods development in this space through standing funding announcements. We will monitor progress and reassess how well these approaches are pushing the field forward in the coming years.

Rapporteurs: Alexander Arguello, Jake Baroch, Sara Currin, Ismail Safi
Division of Genome Sciences, NHGRI

Workshop Summary

Workshop format

The two-day meeting was a hybrid meeting with both in-person and virtual participation, including attendees from the public via webinar. Discussions were structured into three sessions covering levels of biological organization. Each session included a panel to discuss outstanding questions in the field. Additional large-group roundtables addressed related questions in a format conducive to wider audience participation. The session and roundtable topics included:

- Session One: Genetic architecture in the context of cells, tissues, and organs
- Session Two: Genetic architecture in the context of anatomy, physiology, and behavior
- Session Three: Genetic architecture in the context of families, populations, and societies
- Session Four: Research needs and opportunities
- Roundtable One: What's missing from our understanding of genetic architecture across levels of biological organization?
- Roundtable Two: Quantitative vs. qualitative differences in the genetic architecture of biological and social traits

The full agenda, including speakers and panel discussants, can be found on the workshop webpage: Advances in the Genetic Architecture of Complex Human Traits.

In preparation for the workshop, participants were asked to consider these two overarching questions:

- Biology spans temporospatial scales of organization from cells in individuals, populations across geographies, and from phylogeny and ontogeny. How do we incorporate developmental contingencies and evolutionary histories in our understanding of genetic effects on trait variation?
- Data can be consistent with multiple, sometimes mutually exclusive, models. How do we distinguish among competing models? How do we balance the need for further model refinement and additional data generation?

Session Highlights

Session One: Cells, Tissues, and Organs

The first session explored the relationship between genetic variation and "lower level" molecular and cellular traits. Context specificity emerged as a major theme of the session. Several suggestions from the panel and audience are included below:

- Context specificity must be considered when attempting to link trait-associated genetic variation with molecular/cellular genetic effects.
- Effects on gene regulation should be carefully considered in small-scale, in-depth studies to guide large-scale, wider-breadth studies.
- Two competing perspectives emerged about reconciling population-level genetic effects that average out across individuals and those genetic variants that might be most relevant to any given individual with their unique combination of variants.
- Additional molecular traits beyond those well-covered by existing genomic datasets (e.g., GTEx), such as regulatory features, should be included in future QTL mapping efforts.
- Experimental and computational methods need to be developed to explore genomic perturbations efficiently and infer the most relevant contexts for genetic effects.
    - Establishing guidelines for rigorous study designs including number of contexts, how to handle joint-contexts, and how to collect data is imperative.
- Disease processes and trajectories should be considered as a context in studies.

- Questions arose about how to reconcile the apparent ubiquity of context-dependent genetic effects for molecular features and the limited amount of environmental-dependent genetic effects (i.e., GxE) observed for most disease traits.

The audience and panel also posed questions, including:

- Are we entering the realm of "candidate environments"? To identify the most relevant contexts/ environments, how do we narrow the search space in a rigorous manner and control error rates for false associations?
- Are we expecting environmental effects to fall through the same "hourglass" that genetic effects fall through leading to specific categories of diseases and traits? In other words, what is the specificity of environmental effects?
- Can artificial intelligence, or other methods for predictive modeling, be leveraged to integrate the different layers of phenotype ranging from gene expression to biochemical processes?

Session Two: Anatomy, Physiology, and Behavior

The second session focused on how genetic variation impacts traditional organismal-level traits. Several suggestions from the panel and audience are listed below:

- For any given trait, genetic architecture may be shaped by an organism's need to balance two conflicting features of complex systems – stability and flexibility in response to changing environmental conditions.
- Pleiotropy will impact attempts to systematize research efforts into genetic architecture across traits, especially considering the lack of knowledge surrounding the number of causal variants that comprise the molecular basis of a trait.
- The role of allelic heterogeneity in traits should be considered.
- The conflict between population and individual level effects emerged again – how to link the statistical properties of genetic architecture (e.g., the joint distribution of allelic effects) with the biology of a trait.
- Multi-omic tools and measures should be considered as a complement to genetic association studies, for their ability to biologically assess both inherited and modifiable risk factors.
- Polygenicity needs to be well defined and measured and traits should be considered accordingly.
- Dimensional reduction across multiple sources of genetic data should be considered as a way to meaningfully define biological traits.
- Fitness components should be considered as a measure of the biological importance of well-defined traits.
- The relationship between mutation rate at a given locus and how much the resulting genetic variation at that locus contributes to trait heritability is a readily addressable question.
- Care should be given when stating 'every variant affects everything' in the context of the infinitesimal model since there are situations where certain traits have biomarkers and known underlying pathways that indicate some level of specificity. Given this, it is still important to consider the mutational target size for different traits.
- Rare variants also compose genetic architecture and are often more biologically interpretable and tractable than common variants.
- Future biological experiments, including CRISPR perturbation experiments, may offer the ability to alter multiple genes simultaneously and assess polygenic models of disease architecture.

Session Three: Families, Populations, and Societies

The meeting's third session was focused on how different study designs and methods could be used to study genetic architecture across families and populations in the context of societal features, including social determinants of health. Several suggestions from the panel and audience are listed below:

- There was a large focus on the differences between family-based association studies and standard population-based studies.
- Family-based studies can control for much of the genetic and environmental confounding as well as indirect effects that plague population-based studies.
- Family-based studies may, however, be answering different types of questions than population studies. In sampling a different subset of genetic variation, environment, and genetic background, family-based studies may identify a different set of genetic effects.
- Although family-based studies can be useful depending on the trait, they need further development to determine what precisely they are measuring.
- To move forward and capture all relevant effects, the field needs more diversity in samples across geography, socio-economic groups, life stages, and ages.
- Data from efforts guided by diversifying environmental or social context is largely missing.
- As demographic variables are used to make decisions about inclusion criteria, discussions should encompass the potential impacts of the resulting scientific findings, such as how the implementation of polygenic scores without calibration across demographic groups can affect health equity.
- Using polygenic scores to study moderating and mediating roles of environmental variables and genetic confounding may introduce technical issues and bias.
- The field of human genetics should focus on mechanisms as they relate to context dependencies that may nonetheless illuminate more generalizable answers. A cohesive theory of how genetic effects depend on context at the molecular, cellular, environmental, and social levels is missing.
- So far, the evidence for pervasive GxE effects in humans is limited. We need to think in deeper theoretical and statistical terms to identify evidence of interactions and piece out what are environmental effects or not. These effects will depend on the complexity of the trait and will be especially pertinent for behavioral traits. It is under these conditions that family-based studies may be most useful.
- Data from efforts guided by diversifying environmental or social context is largely missing.
- Evidence in animal models shows that GxE is only important if you make a massive change in the environment.
- Large exogenous shocks in humans are needed to see GxE interactions (e.g., economic shocks, changes in educational policy reforms, etc.). Modeling these shocks as instrumental variables will make it difficult to isolate multiple interrelated effects. Many of the traits in humans are inherited phenotypes, rather than exogenous environmental changes.
- Given their expertise, social scientists should have more central roles in study designs aimed at identifying and measuring relevant environments showing GxE effects. There should be more meaningful collaboration between experts in the social sciences or ethics and geneticists. ELSI experts should not be positioned as the assistant to the scientist. There is often a disconnect between the goals of the two groups.
- The quantitative genetics approach for modeling GxE effects might be underpowered to detect some types of environmental interactions. Finding better ways to model and analyze GxE (e.g., by adopting alternative frameworks for the effects of sex on traits) can reveal cryptic effects.
- Collaborations with social scientists could be fruitful when population stratification is considered as a type of trait rather than a confound.

<u>Roundtable One: What's missing from our understanding of genetic architecture across levels of biological organization?</u>

The first roundtable of the meeting focused on critical gaps in the comprehension of genetic architecture across biological levels. Several suggestions from the panel and the audience are listed below:

- In the next ten years, the field of human genetics may experience a shift away from searching for more trait-associated variants. Since many traits have identified variants already, the next step is to measure the variance of allelic effects or study trait variance within the same genotype.
- Care should be taken when integrating complexity at different layers in biological experiments.

- Genetic architecture may be considered similar to urban architecture. A city cannot be understood through its component parts, and as such, independent biological experiments for individual genetic effects may not be informative with regard to the system as a whole.
- Sampling larger-scale, diverse environments is important to moving forward as a field. The environments that are currently being assayed are distinct from the environments the human body has seen through evolutionary time and are widely confined to the Global North.
- Genetics should be used to test phenotypes to determine the nature of the disease and the accuracy and validity of assigned phenotypes.
- Context needs to be better defined before machine learning can be introduced in this space.
- The omnigenic model suggests core and peripheral regulatory pathways and may be a promising way to look at perturbed disease states.
- A focus on the cell and tissue-level phenotypes should be considered given the difficulty in connecting trait-associated genetic variation to specific cellular and tissue-level outcomes.
- There is a need for cellular information from individuals with varying degrees of genetic risk.
- A multidisciplinary approach should be considered to combine experimental techniques and innovative analytical frameworks so the field can better address challenges in unraveling genetic architecture.

The audience and panel also posed several questions on this topic, including:

- With more support in research into genetic architecture, what type of investment should be prioritized?
    - Should the field consider more research into cellular assays, screenings across environments, and/or large-scale perturbation experiments?
    - Are we currently limited by data and/or theory?
    - Are there resources missing?
    - How do we support research on integrated views that prioritize recognizing the system outside of the component parts?
- Are findings biased towards well-annotated portions of the genome, and is this leading to a positive feedback loop?
- How much of polygenicity is really just time? If most traits develop over the course of many years, how much of our study of polygenicity is just capturing the accumulation of time-specific effects?

<u>Roundtable Two: Quantitative vs. qualitative differences in the genetic architecture of biological and social traits</u>

The second roundtable of the meeting focused on genetics in the context of social and cultural influences and the need for a nuanced and context-aware approach to genetic research. Several suggestions from the panel and the audience are listed below:

- It may not always be useful to distinguish between biological and social traits when thinking about phenotypic variation. Researchers need to consider environmental effects on both biological and social traits. Entanglements between biological and social traits do not necessarily exhaust the way we think about behavior. For example, identifying genetic effects for socially constructed traits that vary across space and time (e.g., sexual orientation or disability) may not reveal a fundamental understanding of human conditions.
- It is useful to think about the extent to which GWAS are confounded by indirect genetic and environmental effects. One source of this confounding is due to traits being shaped by how individuals, groups, families, and societies organize themselves and how environments are transmitted across generations.
- A model that may be worth exploring is that of GWAS on housemates. If the housemates are the target person's siblings, you'll pick up direct genetic effects on the target person, as well as indirect genetics effects from the housemates.
- Larger structural forces from economic and political systems should be taken into account as they may be informative contexts in which to interpret biological and social aspects of trait variation.

- The following three questions should drive us to think about different aspects of study design: How do we learn about biology from GWAS? How do we do personalized prediction for medicine? How do we use genetics in other fields (e.g., behavioral genetics, sociology, etc.)?
    - For personalized genetics, we need to think more broadly about how individuals are situated in their environment and communities, how these things intersect with "race" and genetically defined ancestry, and many other factors.
- GWAS can be a more powerful approach if we beyond the idea that it's solely a biological tool. It can be useful as a social science tool, for example by using genetic correlation of the same trait across time as an indicator of medically important social changes.
- When large biobanks like the UK Biobank were constructed, some social factors, like educational attainment, were conceived as social determinants, not as health outcomes themselves. At what point did the social determinants become medicalized? The medicalization of traits has always had a social component, and historically this process has been used to reinforce or justify existing disparities and inequalities.
- Population stratification is not a trait in and of itself, it's a limit on our ability to do causal inference. It has been shown in various ways to be a limitation on GWAS data, in particular.
- There is not enough benchmarking when genetic correlation is being applied in studies (e.g., estimates of genetic correlation between datasets of the same disorder as well as different disorders can be interpreted as an ancestry effect when it is actually a phenotype effect).
- Genetic findings have real implications for how we understand humanity, how we treat each other, and how we structure our societies. When executed properly, genetic studies may help identify sociocultural conditions that are detrimental to human health and well-being.
- Behavioral traits aren't necessarily more prone to measurement error as studies in animals have shown. But imprecision in measurements (i.e., coarse graining of traits) may be useful for tracking similar and related traits that would otherwise be considered distinct traits across varying contexts.
- Geneticists should embrace and acknowledge complexity and leverage it to forge a path forward that allows a diverse breadth of experts to address different questions.

Scientific Organizing Committee

Nancy Cox, Ph.D., Vanderbilt University
Lorin Crawford, Ph.D., Microsoft Research
Molly Przeworski, Ph.D., Columbia University
Peter Visscher, Ph.D., University of Queensland

NHGRI Program Staff

Alexander Arguello
Jake Baroch
Sara Currin
Carolyn Hutter
Ismail Safi

Meeting webpage:
https://www.genome.gov/event-calendar/advances-in-the-genetic-architecture-of-complex-human-traits