National Human Genome Research Institute

NHGRI Strategic Planning Meeting *January 22-24, 2019*

From Genome to Phenotype:
Genomic Variation Identification, Association, and Function in Human Health and Disease

Guidance to Breakout Session Participants

Breakout sessions have two assigned co-chairs and assigned participants. *Please participate in your assigned session.*

Breakout co-chairs and participants are asked to propose and discuss specific ideas for what NHGRI should do in the next 5-10 years. These ideas can be in the form of knowledge to be gained, resources and/or capabilities to be created, and/or specific projects.

- For each, consider: What is its goal? Why is it important? Why should NHGRI undertake it?
- If NHGRI has an existing commitment in this area already, should that change? How (e.g., in the way we approach the question)? Why?
- If the proposed idea(s) are likely to connect to topics of other breakout sessions, how will they connect, and where are the synergies between them?
- If relevant, how should NHGRI's efforts fit into the larger ecosystem of other national and international efforts?

Some questions above will apply more to some breakout topics than others.

We ask co-chairs to come prepared to begin the discussion. They should review the ideas the meeting participants provided prior to the meeting, and solicit additional ideas and explanations, justifications, areas of synergy, etc. from the participants.

Following the breakout session, co-chairs will have 10 minutes per breakout group to report back to all the workshop participants with a summary and highlights of the discussion including, e.g., key ideas and recommendations; items of consensus; areas of disagreement. Following reports from each of the 3 breakout groups, the discussion moderators should lead a broader discussion to identify overlapping ideas across the breakout sessions and ideas of high interest to the group, and also make sure to ask the workshop participants if any relevant items appeared not to be covered.

The Day 1 Breakouts (sessions 1,2,3) will cover a set of potentially related topics on variant discovery:

1) How much more sequencing, if any, is needed to study Mendelian and common inherited disease, both in general and in the context of diverse populations? How should these activities be organized? *What should NHGRI do in this area? Why?*

2) How and why to approach structural variation and other "hard to measure" variation? What technologies and approaches are needed?

3) How and why to approach new complex features - e.g., GxE, epistasis?

The Day 2 Breakouts (sessions 4,5,6) will cover a set of topics on assessing genome function each from a different perspective:

4) What is needed to identify all regulatory elements (enhancers, promoters, insulators, RNA stability, etc.) as well as genes (including protein-coding isoforms, lncRNAs, smORFs)? How can we characterize the function of all genes and regulatory elements in different biological contexts?

Topics may include:

- Technologies, strategies, and tools to do this; opportunities to advance these.
- Which are important and why?
- What should "characterizing" genes entail?
- What should "characterizing" regulatory elements entail?
- How much more is there to do? Can this task be "finished"? Is there a definable stopping point?
- Should comparative genomics be used to address these questions?

5) What is needed to determine the functional consequences of variants, both individually as well as combinations of variants and, ultimately, all variants in a genome?

- Experimental systems:
    - Experimental systems for different biological contexts. What makes them appropriate?
    - What needs to be developed to understand variant function for known disease variants (e.g., model organism assays)? For understanding the potential effect of any variant (e.g., MPRA assays in organoids)?
    - How can we bring the relevant biological context to the assessment of variant function (e.g., cell type or tissue type)?
    - What are the best systems for understanding both coding and noncoding variation?
- Experimental and computational approaches
    - What new capabilities, data, and tools are needed, especially to move beyond one variant at a time?
- How do we solve the problem that this experimental space is potentially so large? (e.g., brute force development of ultra-high throughput methods; sensible prioritization; better fundamental understanding?

6) What is needed (experimental data and computational analysis) to accurately predict the consequences of genetic variants on biological structure and function? How do network perspectives fit into this? What is the role of data integration?

Topics may include:

- How do we understand the action of coding and noncoding variants in combination (cis and trans)?
- How do we ultimately relate findings about genes, regulatory elements and variants in them to biological phenoptypes?
- Computational modelling; AI/ML; sparse data
- What data types are needed for what purpose? What are the most useful data types, and how do we validate them?