NHGRI Cross-Consortia Day 2023 Meeting Report

Wednesday, September 13th, 2023

Washington University, School of Medicine
St. Louis, MO

Cross Consortia Meeting Report

Overview

Meeting Goal: To achieve synergies in research and analysis efforts and enhance communication across several consortia (the Atlas of Variant Effects (AVE) Alliance, the Clinical Genome (ClinGen) Resource, the Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR), the Human Pangenome Reference Consortium (HPRC) and the Impact of Genomic Variation on Function (IGVF)).

Objectives:

- Identify projects (and leaders) we can start now that will facilitate the goals or leverage the resources of one or more of the consortia.
- Identify areas of high interest requiring further consideration to determine barriers and possible solutions to help each other.

Discussion was organized by three broad themes:

- Advancing variant interpretation
- Functional genomics and pangenome
- Clinical genetics and the pangenome

A one day, in person meeting was held including members of the above consortia and the funding agency. Opportunities for working together were identified and are described below under “Ongoing Collaborations” and “Proposed Collaborations”.
Appendix 1 includes the agenda and consortia descriptions;
Appendix 2 is a list of attendees;
Appendix 3 is a matrix of the goals, offers, and needs of the participating consortia.

Complete Report

A one-day cross-consortia meeting was convened 13 September 2023 at Washington University, St. Louis to identify common goals, learn from each other, empower ongoing collaborative work and find new opportunities to work together, thus achieving synergies across 5 consortia (see Appendix 3). This included a grass-roots effort, the Atlas of Variant Effects (AVE) Alliance, as well as 4 consortia funded and programmatically managed, at least in part, by NHGRI: the Clinical Genome (ClinGen) Resource, the Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR), the Human Pangenome Reference Consortium (HPRC), and the Impact of Genomic Variation on Function (IGVF). This report first describes ongoing collaborations, proposed collaborations, and future considerations that were discussed and identified as outcomes at the meeting (Part 1). These are subdivided by the broad themes of the meeting, 1) advancing variant interpretation, 2) functional genomics and pangenome, and 3) clinical genetics and the pangenome. The report also summarizes the meeting organization, including the rationale, planning, and structure (Part 2).

Part 1: Meeting outcomes

Ongoing collaborations

Participants summarized collaborative efforts that were already under way between the various groups. Some of these resulted from the collaborative planning sessions leading up to the meeting, as described in Part 2 below. Broadly, the ideas underlying these collaborations included modifying experimental plans to leverage complementary science, and standardizing variant names and federating resources.

Towards the theme of advancing variant interpretation, GREGoR and ClinGen had previously suggested genes and variants for IGVF to functionally characterize. GREGoR had also nominated genes to be included in an IGVF study characterizing the effects of gene perturbations on transcriptional programs across several cell types, as well as. genes and variants to be characterized by the IGVF coding variant groups. ClinGen asked that IGVF work with with variant curation expert panels (VCEPS) for future coding variant studies. The goals of these suggestions are to generate IGVF data which might help improve understanding of the role of variants in these genes in disease (especially variants that are today classified as Variants of Unknown Significance), and to benchmark these IGVF assays for the ClinGen and GREGoR communities. Additionally, IGVF engagement with clinically facing consortia like ClinGen and GREGoR will allow those consortia to benchmark IGVF assays and better prepare to ingest noncoding experimental data.

Also, in the theme of advancing variant interpretation, ClinGen, GREGoR, AVE and IGVF had agreed to work together to register variants tested by IGVF with ClinGen Allele Registry IDs, which are standard, unique identifiers used for variants by the ClinGen community. This will include registering both coding and non-coding variants and will include variants that have been observed in people as well as previously unseen, synthetic variants characterized in IGVF functional assays. Standard names are expected to increase ease of use by different communities. On the database side, AVE is developing a bidirectional interface from MaveDB,

their open-source platform for data on variant effects, to the ClinGen Linked Data Hub, which aggregates and delivers structured information about variant effects from literature curation and other resources. Finally, AVE recently held a workshop "Curating the Clinical Genome" on the clinical application of data from multiplexed assays of variant effect (MAVEs) data.

In the theme of functional genomics and the pangenome, pilot studies have been initiated by HPRC to layer existing transcriptomic and epigenomic data onto pangenomes. Tools have already been developed to layer short read data onto pangenomes. Pilot work using ENCODE data indicates that use of the pangenome reference enables identification of additional genomic features, including features only found in structural variants that are absent from the GRCh38 reference. This holds promise for using related data from IGVF.

Proposed collaborations

Participants identified multiple new areas for collaboration. One cross-cutting idea was to identify working groups across consortia that oversee related topics and find a way for them to keep each other appraised of their efforts. This idea, which is focused on improved communication, could be important as new opportunities emerge. Another cross-cutting idea is to improve coordination of generating variant effect data by funding agency-led consortia and community grassroots efforts.

There were several proposals for advancing variant interpretation focused on standards/nomenclature and experimental design. As a big picture goal, the group recognized it was important to consider unified and unifying standards and implementation, ideally in collaboration with leading international standards bodies such as the Global Alliance for Genomics and Health (GA4GH). A specific example of this was to test the effectiveness of two ClinGen products, the Allele Registry and Linked Data Hub, to facilitate variant naming and data linking. If successful, these proposals could enhance the experience of the community when using these resources, as well as increase the efficiency of cross-consortia work. It was proposed that variant characterization experiments should be designed to include positive and negative controls for variant pathogenicity, thus allowing validation of the functional assay and calibration of the data for use in variant interpretation workflows. It was also proposed that IGVF consider testing whether assays distinguish particular diseases from each other (not merely distinguishing pathogenic v. benign variants). Perhaps one source of test cases could be GREGoR suggestions of genes with phenotypic expansion beyond ClinGen annotations.

Under the theme of functional genomics and the pangenome, proposed collaborations considered approaches both forannotating pangenomes with functional data as well as for generating reference genomes of use to the functional genomics community. One proposal was to layer existing and new functional data on the draft pangenome. Examples of data sources included ENCODE, IGVF and GENCODE. At this time, Tools for aligning short reads are in place, sufficient for most of the existing data; however, some newer data require tools for long read alignment that are in development. There was also interest in adding and displaying annotations from predictive models to pangenomes, as well as adapting predictive models to use pangenomes as inputs. As noted in ongoing collaborative activities described above, a pilot effort suggests this would add value to the data and the reference. Another proposal was to generate complete (nearly telomere to telomere) reference genomes of interest to the

functional genomics community. Examples included commonly used cell lines (e.g. K562, WTC11, H1, H9) and hIPSC panels that encompassing population diversity. There is also a need from IGVF for a mouse pangenome for commonly used mouse strains and crosses. A related proposal from discussion at the meeting was to generate complete reference genomes for the ENCODE EN-TEx samples--there are epigenomic and transcriptomic data on approximately 30 organs and tissues from 4 individuals through EN-TEx. It was noted that EN-TEx was a collaborative effort between GTEx and ENCODE, and serves as a good example of cross-consortium collaboration.

A variety of proposals emerged on the theme of clinical interpretation and pangenomes. Perhaps the most concrete proposal was to extend the ClinGen Allele Registry to support pangenomes; this would provide unique identifiers to variants in a different context., to define variants and provide annotations. The HPRC has already proposed to host a cross-consortium working group for knowledge sharing, through an iterative feedback-design-build process. HPRC also proposed to design pangenome reference resources and tools to serve the clinical genetics community. Although some of this HPRC work was already planned, it had not been discussed at a cross-consortium level, and cross-consortium involvement could advance and enhance the goals of those efforts. GREGoR proposed to use the draft pangenome to address unsolved cases. As part of this idea, GREGoR would benchmark the solve rate using the pangenome against the current reference, including benchmarking the solve rate using the pangenome for cases that are annotated solved v. unsolved using the current reference. If use of the pangenome increases the solve rate, that might encourage further uptake of the pangenome in other settings. Novel disease-associated variants might emerge from this work.

Future considerations

Several additional ideas surfaced during planning or the meeting, ranging from implementation guidelines to ideas that might become proposals after further study.

For the theme of advancing variant interpretation, the issue of federating, coordinating or perhaps merging resources was raised. The resources discussed included the IGVF Catalog, AVE MaveDB, ClinGen Allele Registry, Linked Data Hub, GREGoR AnVIL repository.

Under the theme of functional genomics and the pangenome, a few potential projects emerged. IGVF, ENCODE and investigator-initiated projects are producing data on mouse crosses that could immediately benefit from a publicly available, high quality mouse pangenome, although that will require additional collaborations with groups producing this type of data. Any successful combination of functional genomics and the human pangenome could be a roadmap, demonstration, or encouragement for clinical efforts; perhaps a project tailored to variant interpretation could accelerate clinical adoption. Finally, infrastructure to enable version-controlled updates of pangenomes and their annotations, rather than rebuilding de novo for each iteration, could help both the user community and the reference builders.

Guidance on the theme of functional genomics and the pangenome also emerged. One call was for consideration of personal genomes and reference-free annotations, which in some studies may be more powerful or more appropriate than pangenomes. Another call was to avoid confusion when using reference genomes from cell lines; while it is important to have easy access to genomes of historical cell lines for functional genomics today, it would be

problematic if users studying human variation were misled into thinking cell line abnormalities were found in actual people. Another point was that, while there is enthusiasm for annotating haplotype-pangenomes with long read functional data, most existing functional data are short read, and tools and methods for the pangenome may need to consider this. Employing diverse genetic backgrounds, including the types of diversity captured in the HPRC efforts, in functional characterization studies could enhance their power and generalizability. Finally, there was a call to Identify a slightly larger group of consortium partners for collaboration, perhaps adding in MorPhiC, SMaHT and AnVIL.

Guidance on the theme of clinical interpretation and pangenomes called for caution. First, it was noted that transition to a new reference is not trivial for medical evidence boards, clinical testing companies and insurers. Specificity is often more important for clinical needs, while sensitivity can be more important for research needs. There is currently a lot of clinical genomics resources and infrastructure that use older builds that would need to be lifted over. Clinical testing labs are likely to require both guidance and incentives as to why transitioning to a new reference genome would be beneficial.

The theme of cross-cutting Ethical, Legal and Social Implications (ELSI) related to the work of all the consortia was raised during the discussion, and is critical to current, proposed and future collaborations. Issues related to how to define diversity, inclusion of indigenous populations in large data resources, data sharing across multiple domains, and the downstream impact of processes such as reclassification within the clinic are common to many of these projects.

Part 2: Meeting organization

Initially, the rationale was to identify ways for NHGRI consortia to work together on the broad theme of variant interpretation, from basic science to the clinic. NHGRI-funded consortia with overlapping or related goals have often pursued opportunities to collaborate. In some cases, the collaborative spirit was explicit in the funding agency concepts. Some of these collaborations have been initiated by the funded investigators, including those who were funded to work on more than one of these consortia. These collaborations have also been supported by NHGRI program staff, who, as part of managing their programs and advancing the field, look for opportunities for productive collaborations. In this case, there was recognition of a common thread running between consortia that aim to understand and characterize variant effects and those that aim to understand Mendelian and common diseases to those that provide resources supporting those activities.

Two working groups were formed, an internal NHGRI working group, and a cross-consortia working group that included funded investigators from AVE, ClinGen, GREGoR and IGVF. Initial discussions focused on how the consortia might help each other. Each participating consortium was asked to provide information on their Goals, “Offers” of what they could provide, and “Needs” from other consortia, including stating what collaboration is already taking place, and what are the barriers to collaboration. This began to surface concrete ideas. (This matrix of consortia goals, offers and needs can be found in Appendix 3.) This potential for productive interactions—around issues ranging from data interoperability to complementary

designs such as validating disease variants or using common target gene lists, spurred the Cross-consortium Meeting. It soon became clear that was also an opportunity to discuss adoption of reference genomes for improving functional analyses, leading to inclusion of HGRP, and the development of the broad themes related to the pangenome.

As output from the working groups, ambassadors from each consortium provided overviews to other consortia during their regularly scheduled meetings. (Descriptions of the consortia, including their high-level goals and links to their resources, can be found in Appendix 1.) During these discussions it became apparent that collaboration might benefit from an in person meeting for exchange of ideas; as a result, a meeting planning group was formed. Members of the Cross-Consortia Planning Committee are listed in Appendix 1.

The meeting was structured along a few principles. An early choice was a goal of identifying practical efforts that were likely to be achieved (rather than a list of ideas placed on a shelf and forgotten). Attendees were invited with this same idea in mind, so that at the end of the meeting ideas might have leaders. It was decided to have representatives from NHGRI and external scientists work together. It was also decided to include people across career stages who might have different vantage points, such as students, postdocs, and program analysts, to PIs and program officers. Three consortia then planned their annual meetings at the same location, at similar times, so that members might attend a joint, in person meeting.

The meeting was structured around a core of broad theme discussions: advancing variant interpretation, functional genomics and pangenomes, as well as clinical genetics and the pangenome. Discussion leads introduced the landscape (including ongoing work) and in some cases introduced proposals, followed by discussion of what efforts might benefit the consortia. Discussion leads were PIs from the relevant consortia, while the moderators were graduate students. To set the stage for these discussions, the meeting began with an NHGRI Town Hall, led the Dr. Eric Green, Institute Director, which aimed to frame the discussion in the larger context of the NHGRI 2020 Strategic Vision, followed by short presentations from each consortium about their collaboration goals. After the broad theme discussions, the meeting ended with summaries from the funding agency, the individual consortia, and a discussion session. The agenda can be found in Appendix 1.

Attendees included about 220 in person and about 70 virtual attendees. Of those attending in person, about 100 listed IGVF as an affiliation, about 70 listed HPRC, about 70 listed GREGoR, about 20 listed ClinGen, and about 15 listed AVE (attendees were permitted to list more than one affiliation). A list of attendees can be found in Appendix 2.

Sources

This meeting report quotes and paraphrases presentations and discussions at the meeting, as well as notes from attendees.

These individuals from NHGRI worked on the meeting, meeting planning, or report:
Sarah Anstice, Alexander Arguello, Afia Asare, Zo Bly, Lisa Chadwick, Sara Currin, Jyoti Dayal, Adam Felsenfeld, Daniel Gilchrist, Nicole Lockhart, Joannella Morales, Stephanie Morris, Erin Ramos, Alessandra Serrano Marroquin, Maya Vanzanten, Christopher Wellington.

Appendix 1: Meeting Booklet

Agenda, Descriptions of Consortia, Meeting Planning Committee, Logistics

Appendix 2: Meeting attendees

Appendix 3: Cross-consortia goals, offers and needs table

Meeting logo: