2024 Clinical Atlas of Variant Effects meeting summary

Writing credit (Appendix 1), correspondence to: dfowler@uw.edu
DOI: 10.5281/zenodo.13900941

We remain largely unable to interpret the simplest single-nucleotide genetic changes, even in the genome's best-understood (protein-coding) regions. These clinically uninterpretable variants end up as Variants of Uncertain Significance (VUS), which cannot be used to diagnose or treat disease. But a confluence of experimental and computational technologies have enabled determination of single nucleotide variant effects at scale. Applying these technologies comprehensively would generate an 'atlas of variant effects', with profound implications for understanding human biology and diagnosing and managing disease.

On July 23rd, 2024 researchers, clinicians, funders and other stakeholders met to develop a vision for a draft atlas of variant effects (attendees, Appendix 2; agenda, Appendix 3). The primary goal of the meeting was to develop recommendations that will enable realization of a draft atlas by 2030, with a focus on resolving VUS. The meeting was highly interactive and included discussion of the types of measurements and predictions that should be made, integration of experimentally determined variant effects from a variety of types of assays executed by different labs and production centers, and the critical role that will be played by AI in guiding experiments and predicting variant effects. Thus, we also explored the advantages and disadvantages of investing resources in different ways and drafted recommended actions for the community to take.

Vision and goals for an atlas in 2030

We discussed three straw proposals (Appendix 4) designed to stimulate discussion of the vision and goals for a draft atlas in 2030. **Proposal A** prioritised eliminating VUS quickly by scaling up data production and translation; **Proposal B** prioritised technology development to understand variant pathogenic mechanisms and yield mechanism-aware predictive models; and **Proposal C** prioritised testing variants in a wider range of genes and cell types to understand gene-cell type relationships and yield cell type-aware predictive models. We reached the following conclusions:

- The draft atlas should be focused on addressing the clinical imperative, especially VUS, while acknowledging that building and sharing an atlas will empower diverse applications and further investigations focused on deepening understanding of variant biology
- Most effort and resources should be focused on applying and scaling existing assays and predictive algorithms to evaluate variant effects in clinically-relevant, high-need genes. The remaining effort should be focused on (i) developing and deploying novel assays to interrogate additional high-need genes and (ii) models that yield a deeper understanding of how variants perturb molecular and cellular function
- An atlas effort should be tightly integrated with other large collaborative efforts that can either inform the deployment of assays and models or add value to the variant effect data generated. Examples of such efforts include wwPDB (to provide protein structural information and integration with the structural biology community); Human Protein Atlas (to map human proteins in different cells and tissues); DepMap and HCA (to identify relevant cell types and genetic contexts); and the AVE Alliance and IGVF (to develop and deploy variant effect mapping technology)
- Tight integration of AI and experiments is urgently needed, both to optimize allocation of experimental resources and to deliver high-quality variant effect information
- Data coordination, standards and dissemination will be essential for ensuring the clinical impact of the atlas.
- A draft atlas can best be achieved through close engagement and collaboration across academic and diverse industry sectors (experimental technology, AI, pharma, diagnostic).

Strengths, weaknesses and opportunities

We explored the strengths, weaknesses and opportunities of the current atlas community and efforts, as well as identifying threats that may be encountered in the coming years:

Strengths

- The Atlas of Variant Effects community is engaged, strong, and collaborative
- Existing data infrastructure in enabling data sharing and coordination (e.g. MaveDB, MaveRegistry, etc)
- The current generation of experimental technologies and predictive models effectively map variant effects with sufficient accuracy to have transformative clinical impact
- The existing community has a demonstrated record of success in spanning all the necessary stages from technical development to clinical impact

Weaknesses

- While data standards for both experiments and predictions exist, they are currently incomplete. And, there is poor adherence to and supporting resources for applying these standards within the broader research community
- We currently lack variant effect experimental data for most genes
- Current experimental and predictive technologies are not applicable to all genes, regulatory elements, cellular and molecular phenotypes, and cell types
- Additional generalizable experimental technologies that are scalable and information-rich are needed
- The paucity of clinical control variants limits quality control and benchmarking of functional data for many genes
- The cost limitations of experiments do not yet match available funding and desired timelines
- The current funding landscape is fragmented and is not sufficiently coordinated or at the scale needed to lead to realization of the goals/vision by 2030

Opportunities

- Technology advances will reduce costs, increase scale and give richer phenotype information
- Connecting and collaborating with a broader community of stakeholders, especially the clinical, industry and AI communities, could inform more innovative and effective strategies
- Supporting development of data standards and associated tools and educational resources will improve interoperability and reusability of MAVE data for downstream applications. Key applications include AI-assisted experimental decision making and prioritization, as well as integration into clinical variant classification frameworks
- Better alignment of experimental and modeling efforts with clinical studies and priorities would increase impact on human health
- Comprehensive variant effect measurements and predictions under diverse conditions could help address the representation and diversity gaps in genomic research and medicine
- Engagement and coordination of funders in this area could transform the funding landscape

Threats

- The silo effect and typical academic and industrial pressures to hoard data and not invest resources in robust sharing approaches could limit needed data sharing and international coordination
- The shift of industry away from open source models towards proprietary code
- Failing to think outside the box may lead to missed opportunities

Draft roadmap to 2030

Based on discussions at the workshop, we are developing a set of recommendations for the community in five key areas: technology, infrastructure and standards, clinical translation and data production and coordination. These recommendations build on key strengths, address key weaknesses and exploit the opportunities identified during the meeting. A cross-cutting and very high priority recommendation is to develop clear pathways for AI researchers to engage with atlas efforts in all four areas. We intend to make public a roadmap to 2030, comprising our recommendations for the community. To accomplish this goal, the organizing committee, joined by interested meeting attendees, will expand upon and detail the summary recommendations below. The expanded recommendations and roadmap will then be opened for public comment and a final draft will be released.

- **Technology**
  - Improve scalability and reduce costs of existing assays
  - Develop new experimental assays and predictive models for rich phenotypes (e.g. biophysical properties, imaging, -omics) and diverse contexts (e.g. cell and tissue type-aware, presence of “environmental stressors” e.g., drugs)
  - Explore and exploit synthetic biology approaches to dissect gene regulation and protein structure/function
  - Build a platform to share libraries, protocols and expertise
  - Develop tight integration of experimental and predictive approaches

- **Infrastructure and standards**
  - Create a global data coordinating center to drive data standardization and dissemination
  - Engage with global clinical and data standards organizations for cross-community alignment
  - Standardize data analysis workflows to enhance data interoperability
  - Garner adequate funding for needed infrastructure and data curation/standardization
  - Guide journals to enforcing appropriate data sharing standards

- **Clinical translation**
  - Convert experimental and predictive data into a format that clinicians can easily use
  - Give guidelines for use of experimental and predictive data in variant classification and disease association
  - Provide training to clinicians regarding the use of experimental and predictive data
  - Develop a rubric for disease, gene and variant-level priorities for data production

- **Data production**
  - Develop infrastructure to build large-scale production centers and empower small/medium scale producers
  - Ensure that data generation is equitable with respect to different populations and communities
  - Create a framework for coordination of data production, building upon existing infrastructure
  - Engage a broad community of potential funders, including within industry, in a manner that allows them to focus on their particular mission.