International Data Sharing Efforts: Lessons from Canadian Open Genetics Repository (COGR)

Shana White [1], Samantha Baxter [1], Michael Oates [1], Chet Graham [1], Matthew S Lebo [1], Jordan Lerner-Ellis [2]

[1] Laboratory for Molecular Medicine, PCPGM; Brigham and Woman's Hospital, Harvard Medical School; [2] Mt Sinai Hospital, University of Toronto, Ontario Institute for Cancer Research

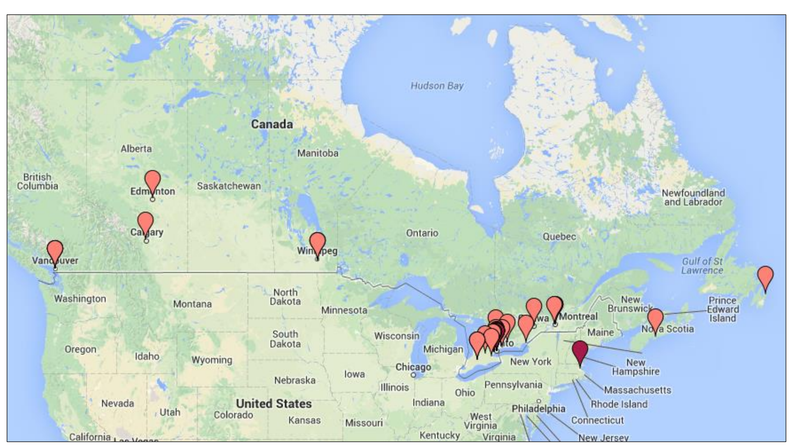
Background

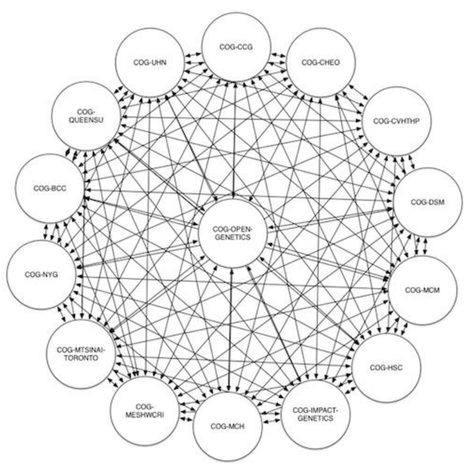
Individual laboratory knowledge and siloed data sets impede our knowledge of variants and prevents clinicians from receiving the most accurate interpretations for variants found in their patients. This in turn may prevent patients from receiving the most appropriate care.

The Canadian Open Genetics Repository is a collaborative effort to create a unified, open-access, clinical-grade database of variants reported by medical diagnostics laboratories across Canada. Utilizing GeneInsight® as a common platform, labs will collect, store and share variant information in real time.

As clinical laboratories adopt modern genomics technologies, the need for this type of collaborative framework is increasingly important.

Participating Centers and Data Sharing Model





Participating COGR Centers with a GeneInsight Instance

- Mt Sinai Hospital
- Impact Genetics
- Memorial University
- Sick Kids Hospital
- Western University
- McGill University Health
- United Health Network
- University of Manitoba
- Alberta Children's Hospital
- British Columbia Cancer Agency
- Children's Hospital of Eastern Ontario
- British Columbia Children's Hospital
- North York General Hospital
- Hamilton Health Sciences-McMaster University
- Credit Valley Hospital-Trillium Health Partners
- Canada's Michael Smith Genome Sciences Centre
- Edmonton Molecular Diagnostics-University of Alberta
- Research Molecular Genetics Lab-Women's College Hospital
- Kingston General Hospital- Queen's University

Experiences

Our progress thus far has highlighted the critical need for robust and sustainable IT infrastructure. Many laboratory systems don't validate variants against a reference sequence, resulting in errors during manual entry. In addition, variants, their classifications and their text interpretations are not always saved in one unified system. This creates significant barriers to accessing essential information for future interpretation or data sharing.

Differences between laboratories in regards to category names and discrepancies in the variant assessment process has been accentuated as a result of variant sharing. Our goal is that continued use of a common variant assessment tool and real-time variant sharing will increase standardization and discussion.

Aims

Activity 1. Design freely available and consistent variant assessment procedures.

- The goal is to facilitate the process of transforming individual variant data holdings into a unified format, while eliminating discrepancies, omissions and duplicated efforts. Enabling multiple stakeholders to assess variant significance in a systematic, comprehensive, and consistent manner will foster knowledge aggregation from different individuals, institutions, and areas of expertise.

Activity 2. Data extraction and transfer.

- The project team will devise methods and optimize operating procedures to support the extraction of disease, gene and variant data currently held within participating laboratories in Canada. Our bioinformatics team will work with each laboratory to ensure that their data are transmitted safely and efficiently to a common platform that will allow ongoing curation and sharing efforts to occur.

Activity 3. Data access and dissemination.

- Methods will be developed to sure the data holdings are extremely accurate and readily accessible. Data will be presented to different stakeholder groups in appropriate levels of summary. Individual laboratories see their variants with high levels of specific detail while the wider diagnostic community views a more general summary that is mindful of privacy and confidentiality.

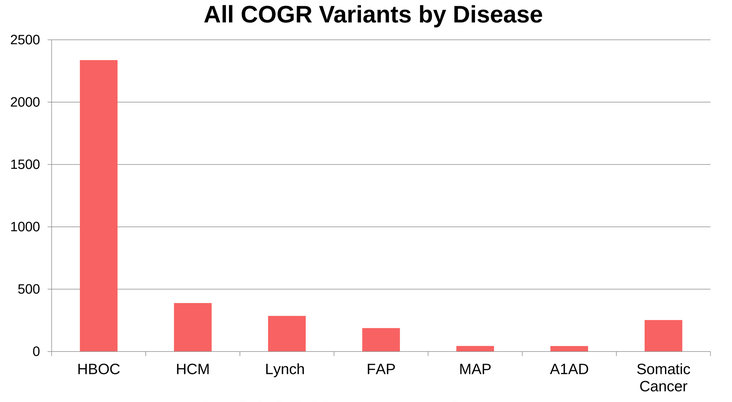
Current Variant Sharing

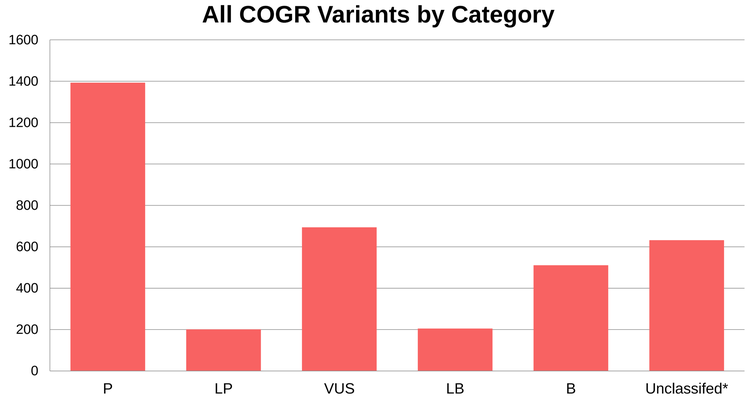
Participating laboratories are currently sharing:

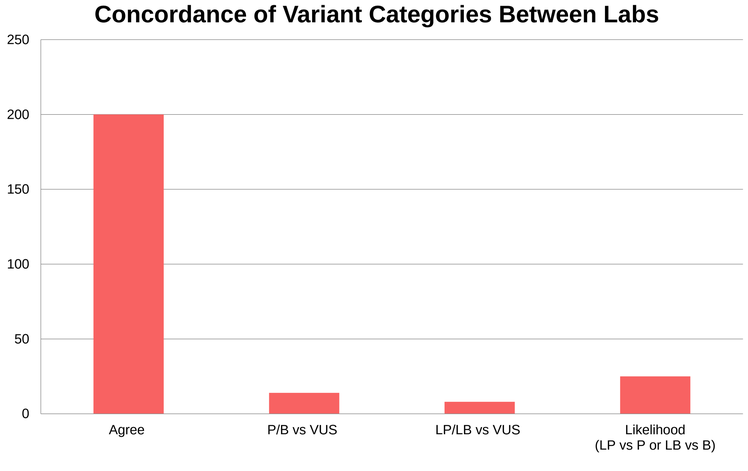
- 7 Diseases: HBOC, HCM, Lynch syndrome, FAP, MUTYH-Associated Polyposis (MAP), Alpha-1 Antitrypsin Deficiency (A1AD), Somatic Cancer
- 52 Genes
- >3,000 variants

- Individual laboratories maintain their own variant classifications in a local site that can be shared with other laboratories and the central hub in real time.
- Laboratories can view external laboratories' categories and evidence for that category.
- *Unclassified variants are under review, pending classification.

- Over 3,000 variants are being shared
  - 391 have been reported by 2 different labs
  - 50 have been seen by 3 different labs
- The majority of variants categorized by multiple labs have concordant categories.







Future Directions

The COGR project is positioned to help remove some of the chief obstacles to advancements in personalized healthcare including the lack of standardized resources and protocols for interpreting the ever-increasing volumes of patient data being generated by clinical labs.

Continuing initiatives include:

- Establishing committees of disease area experts to create consensus interpretations for variants categorized by multiple laboratories.
- Make consensus variant interpretations available to stakeholder groups with different but appropriate levels of summary.
- Continue to collaborate with other international data-sharing efforts including ClinVar.

In summary, the COGR serves as a focal point for the collaboration of Canadian laboratories with themselves and other countries in the development of tools and methods that take full advantage of laboratory data in diagnosing, managing and treating genetic diseases. Furthermore, as more laboratories share data, knowledge will improve and ultimately lead to better patient care.

Canadian Open Genetics Repository
Email: cogr@opengenetics.ca
Web: **http://opengenetics.ca**